Sparsity and Nonnegativity in Least Squares Problems and Matrix Factorizations

Public PhD Defense

Nicolas Nadisic

21 April 2022

Université de Mons, Belgium

Introduction

Our motivation

General motivation for data science: extract useful knowledge and meaningful information from data.

High-level motivations of this thesis:

- Extract underlying structures in data
- Better leverage a priori knowledge, notably nonnegativity and sparsity, to improve models
- Develop algorithms that are both guaranteed and computationally tractable

Starting point: linear models

Focus of this thesis: linear models of the form

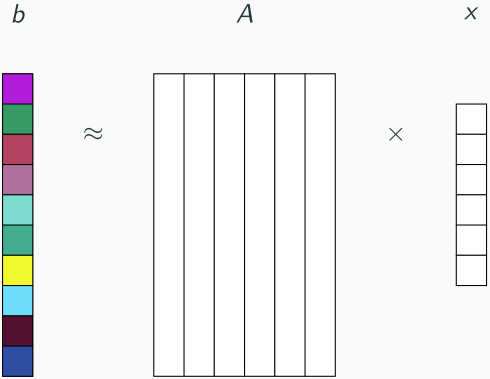
$$Ax \approx b,$$

where

- $x \in \mathbb{R}^r$ is a signal or information vector,
- $b \in \mathbb{R}^m$ is the data vector, representing measures or observations,
- $A \in \mathbb{R}^{m \times r}$ is a coeficient matrix, called dictionary, representing features, atoms, or components.

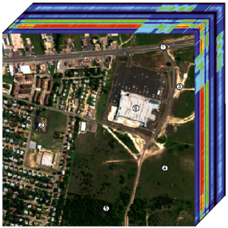
Linear models

$$b \approx A \times x$$



One application — Hyperspectral imaging

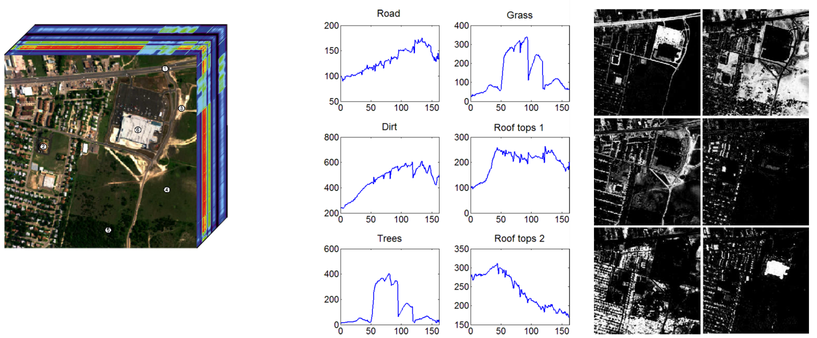
$\underbrace{\hspace{3em}}_{\text{spectral signature of one pixel}}^{b}$



Images from Bioucas Dias and Nicolas Gillis.

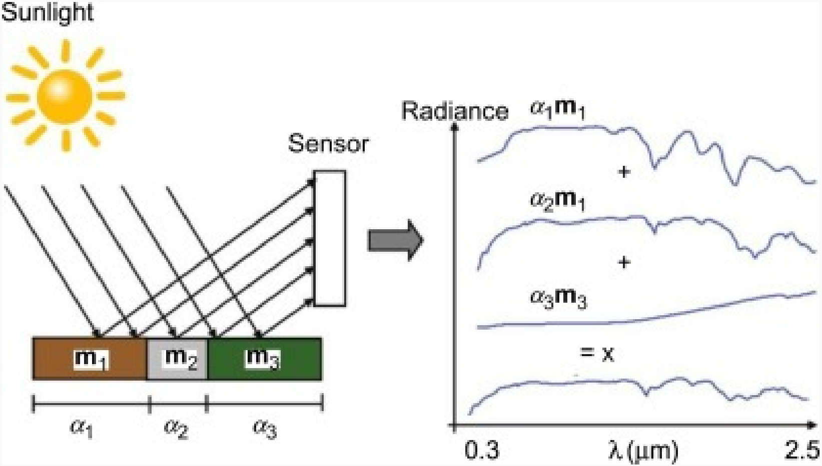
One application — Hyperspectral imaging

$$\underbrace{b}_{\substack{\text{spectral signature of}\\ \text{one pixel}}} \approx \underbrace{A(:,p)}_{\substack{\text{spectral signature of}\\ \text{p-th material}}} \underbrace{x(p)}_{\substack{\text{abundance of p-th material}\\ \text{in one pixel}}}$$



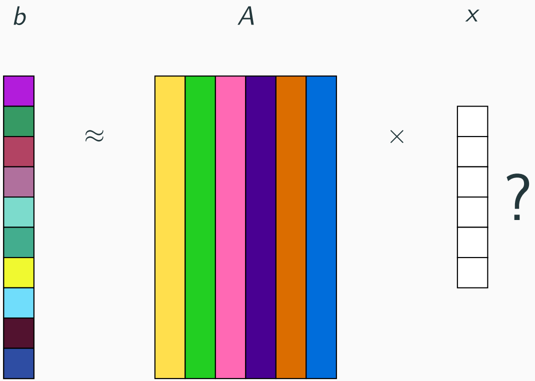
Images from Bioucas Dias and Nicolas Gillis.

Linear mixing model



Our starting point: Linear inverse problem

Given b and A, find x



Least squares problem

How to recover x given A and b, in the presence of noise?

Least squares problem

How to recover x given A and b, in the presence of noise?

Choose a data fidelity measure.

Least squares problem

How to recover x given A and b, in the presence of noise?

Choose a data fidelity measure.

Here we choose the squared ℓ_2-norm, $\|v\|_2^2 = \sum_i v_i^2$, leading to a least squares problem

$$\min_x \|Ax - b\|_2^2.$$

Least squares problem

How to recover x given A and b, in the presence of noise?

Choose a data fidelity measure.

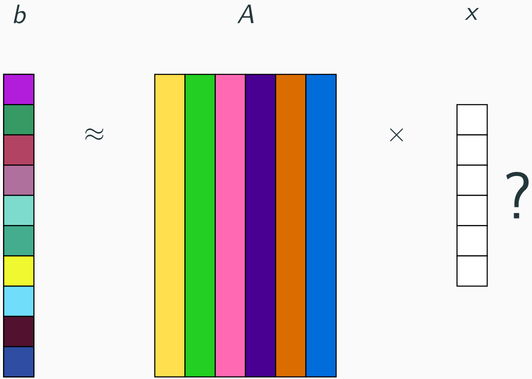
Here we choose the squared ℓ_2-norm, $\|v\|_2^2 = \sum_i v_i^2$, leading to a least squares problem

$$\min_x \|Ax - b\|_2^2.$$

- In most cases, ill-posed problem.
- When the data is noisy, the solution x may not represent well the reality.

How to improve the model?

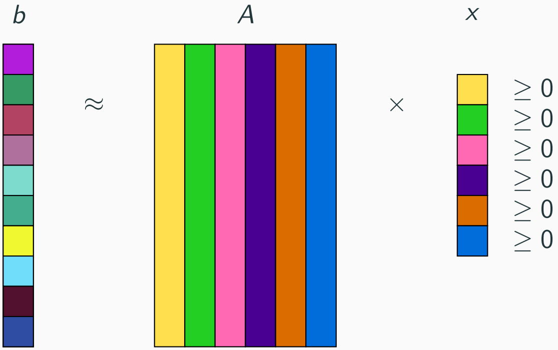
Leverage a priori knowledge or assumptions on the structure of the solution.



Assumption 1: nonnegativity

Nonnegativity of x

$\Rightarrow$ data comes from an additive combination of features



Nonnegative least squares

- Nonnegative least squares (NNLS)

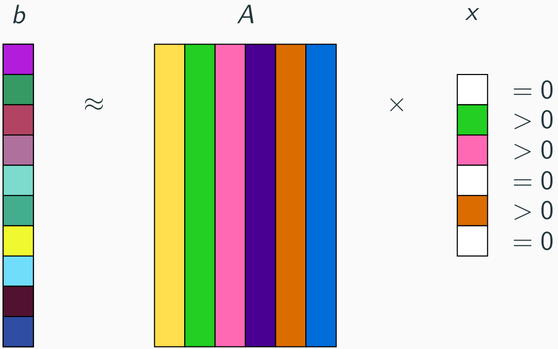
$$\min_{x} \|Ax - b\|_2^2 \text{ s.t. } x \geq 0$$

- More interpretability
- Natural in many applications

Assumption 2: sparsity

Sparsity of x $\Rightarrow$ few non-zero entries

$\Rightarrow$ data comes from a combination of few features



How to enforce sparsity?

- A natural sparsity measure: ℓ_0-"norm"
  $||x||_0 = |\{i : x_i \neq 0\}|$ (number of nonzero entries of x).

How to enforce sparsity?

- A natural sparsity measure: ℓ_0-"norm"
  $||x||_0 = |\{i : x_i \neq 0\}|$ (number of nonzero entries of x).
- With a ℓ_0 constraint, k-sparse NNLS

$$\min_{x \geq 0} \|Ax - b\|_2^2 \text{ s.t. } \|x\|_0 \leq k$$

How to enforce sparsity?

- A natural sparsity measure: ℓ_0-"norm"
  $||x||_0 = |\{i : x_i \neq 0\}|$ (number of nonzero entries of x).
- With a ℓ_0 constraint, k-sparse NNLS

$$\min_{x \geq 0} \|Ax - b\|_2^2 \text{ s.t. } \|x\|_0 \leq k$$

- Intuitive formulation:
  a data point is generated from at most k features
- Hard to solve: combinatorial problem with $\binom{r}{k}$ possible supports (set of nonzeros entries)

Generalization of NNLS: MNNLS

A generalization of NNLS with multiple columns:

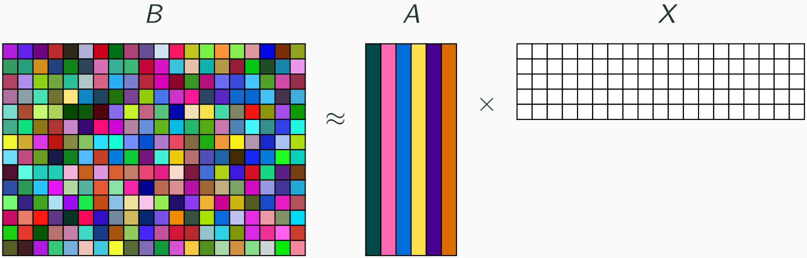
Multiple Nonnegative Least Squares (MNNLS)

$$\min_{X \geq 0} \|B - AX\|_F^2,$$

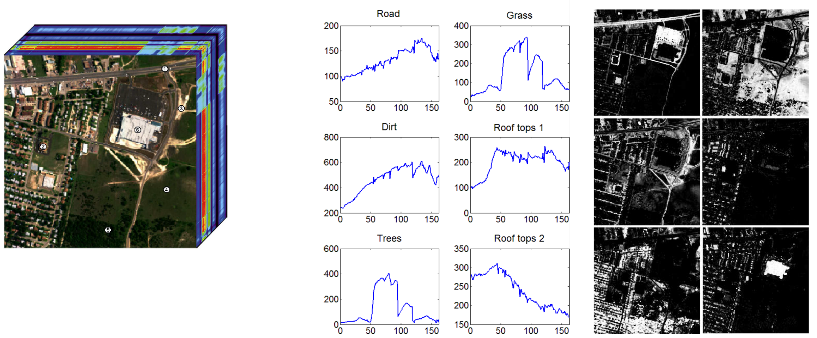
with $B \in \mathbb{R}^{m \times n}$, $A \in \mathbb{R}^{m \times r}$, and $X \in \mathbb{R}^{r \times n}$.

Can be divided in n independent NNLS subproblems

Multiple Nonnegative Least Squares (MNNLS)



$$\underbrace{B(:,j)}_{\substack{\text{spectral signature of}\\ \text{j-th pixel}}} \approx \sum_p \underbrace{A(:,p)}_{\substack{\text{spectral signature of}\\ \text{p-th material}}} \underbrace{X(p,j)}_{\substack{\text{abundance of p-th material}\\ \text{in j-th pixel}}}$$



Images from Bioucas Dias and Nicolas Gillis.

Generalization: Nonnegative matrix factorization

If A is also unknown?

Generalization: Nonnegative matrix factorization

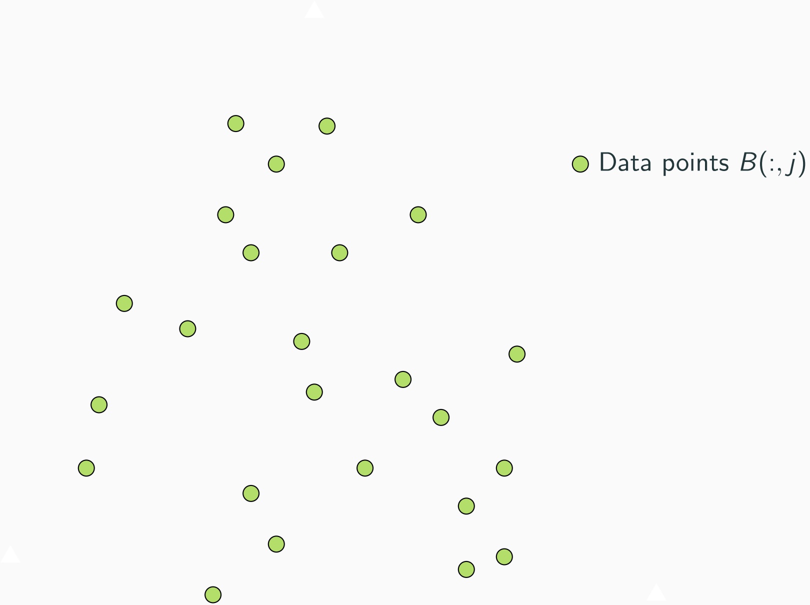
If A is also unknown?

Given $B \in \mathbb{R}_+^{m \times n}$ and $r \in \mathbb{N}$, find $A \in \mathbb{R}_+^{m \times r}$, and $X \in \mathbb{R}_+^{r \times n}$,

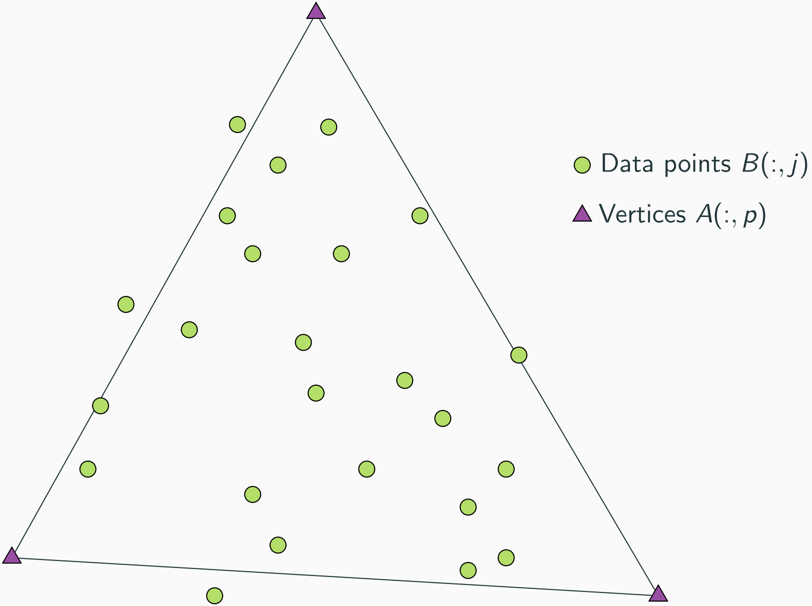
Nonnegative matrix factorization (NMF)

$$\min_{A \geq 0, X \geq 0} \|B - AX\|_F^2$$

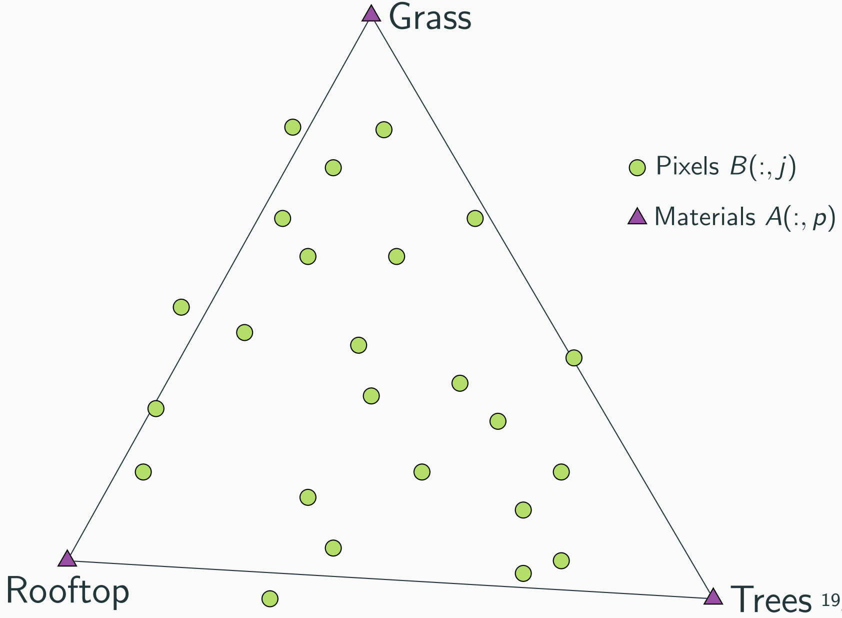
NMF Geometry ($B \approx AX$)



NMF Geometry ($B \approx AX$): cone / convex hull



NMF Geometry ($B \approx AX$): cone / convex hull



NMF $\longrightarrow$ MNNLS $\longrightarrow$ NNLS

- NMF:

$$\min_{A \geq 0, X \geq 0} \|B - AX\|_F^2$$

- Optimizing one factor while fixing the other is a multicolumn nonnegative least square (MNNLS) subproblem

$$\min_{X \geq 0} \|B - AX\|_F^2,$$

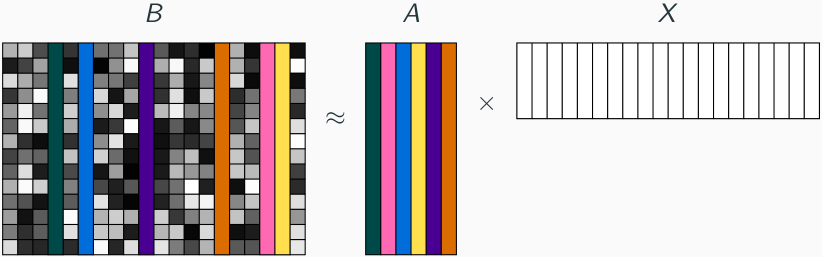
- that can be decomposed into n nonnegative least squares (NNLS) subproblems

$$\min_{x \geq 0} \|Ax - b\|_2^2,$$

where $X(:, j)$, A, and $B(:, j)$ correspond respectively to x, A, and b.

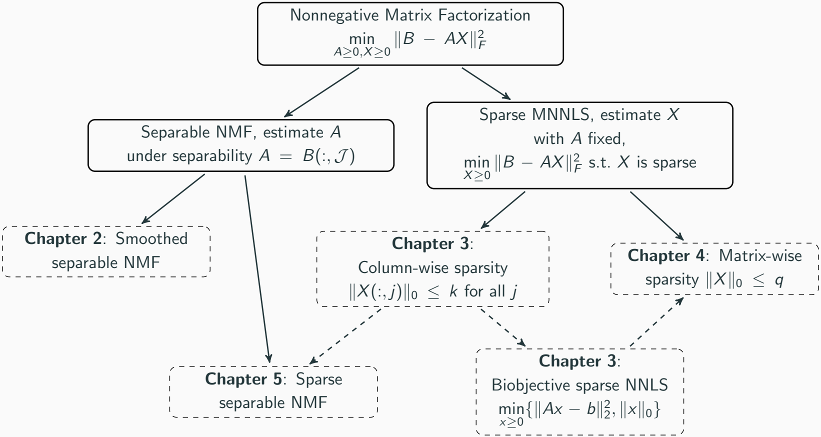
Another assumption: separability

For each vertex, there exist at least one data point equal to this vertex
$\Leftrightarrow$ pure-pixel assumption



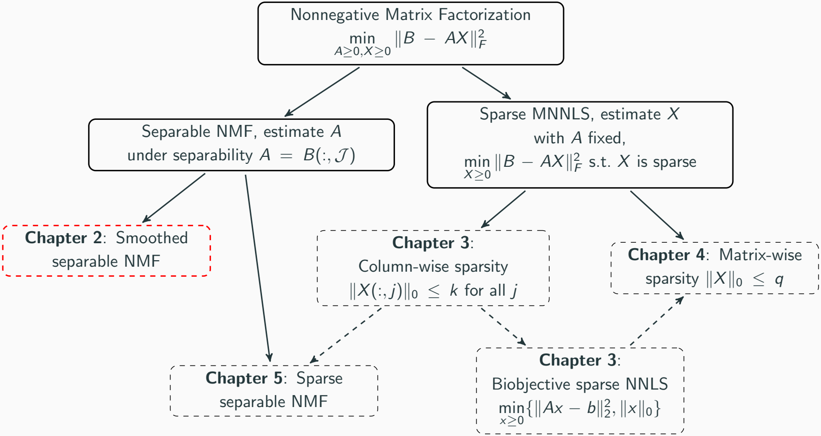
$\Leftrightarrow$ There exists an index set $\mathcal{J}$ with $|\mathcal{J}| = r$ such that $B \approx B(:, \mathcal{J})X$

Overview of contributions



Smoothed separable nonnegative matrix factorization

Overview



Nonnegative Matrix Factorization
$\min_{A \geq 0, X \geq 0} \|B - AX\|_F^2$
Separable NMF, estimate A
under separability $A = B(:, \mathcal{J})$
Sparse MNNLS, estimate X
with A fixed,
$\min_{X \geq 0} \|B - AX\|_F^2$ s.t. X is sparse
Chapter 2: Smoothed
separable NMF
Chapter 3:
Column-wise sparsity
$\|X(:, j)\|_0 \leq k$ for all j
Chapter 4: Matrix-wise
sparsity $\|X\|_0 \leq q$
Chapter 5: Sparse
separable NMF
Chapter 3:
Biobjective sparse NNLS
$\min_{x \geq 0} \{\|Ax - b\|_2^2, \|x\|_0\}$

Smoothed separable NMF

Chapter 2 of the thesis. Presented in the article:

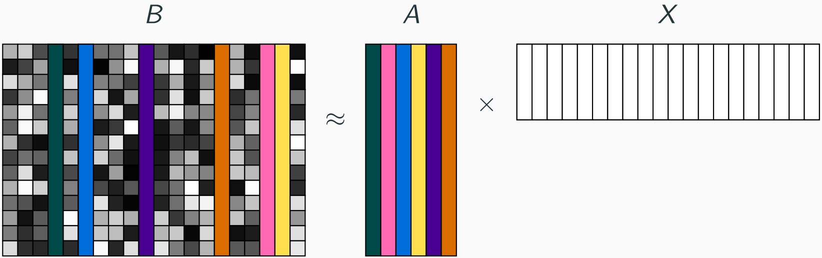
NN, Nicolas Gillis, and Christophe Kervazo (2021). "Smoothed separable nonnegative matrix factorization". In: *preprint arXiv:2110.05528*.

Why? Separable NMF is popular and powerful but algorithms do not leverage the presence of multiple pure data points (only one does so, and it has limitations)

What? Two smoothed separable NMF algorithms that outperform the state of the art

Model 1: Separable NMF

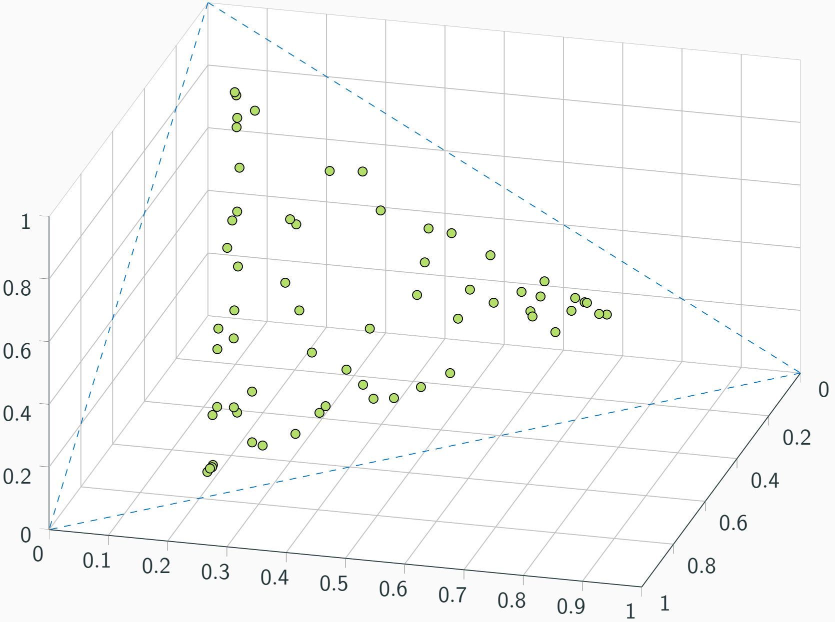
- NMF is NP-hard in general.
- Under the separability assumption, it is solvable in polynomial time.



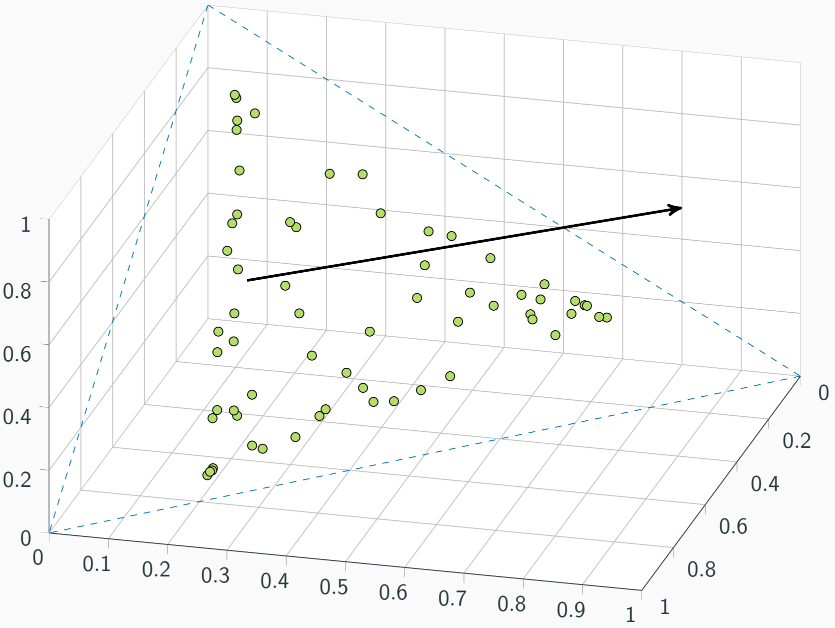
Algorithms: we focus on two greedy algorithms

- VCA: Vertex Component Analysis (Nascimento et al. 2005)
- SPA: Successive Projection Algorithm (Araújo et al. 2001)

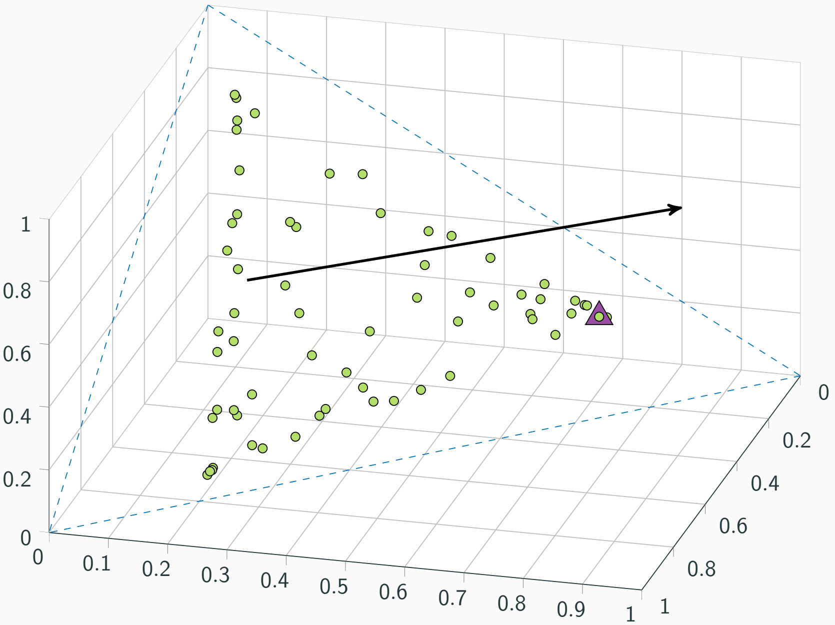
VCA — Animation



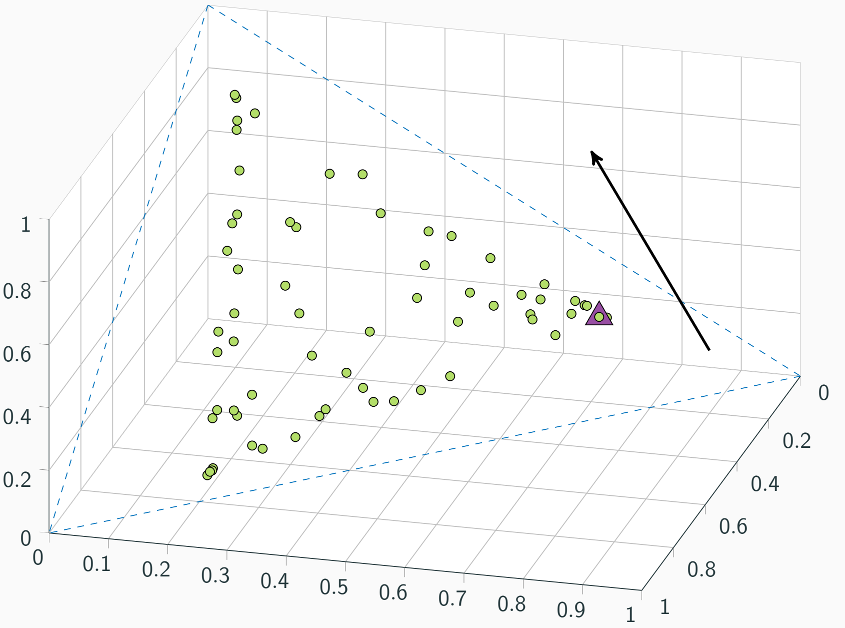
VCA — Animation



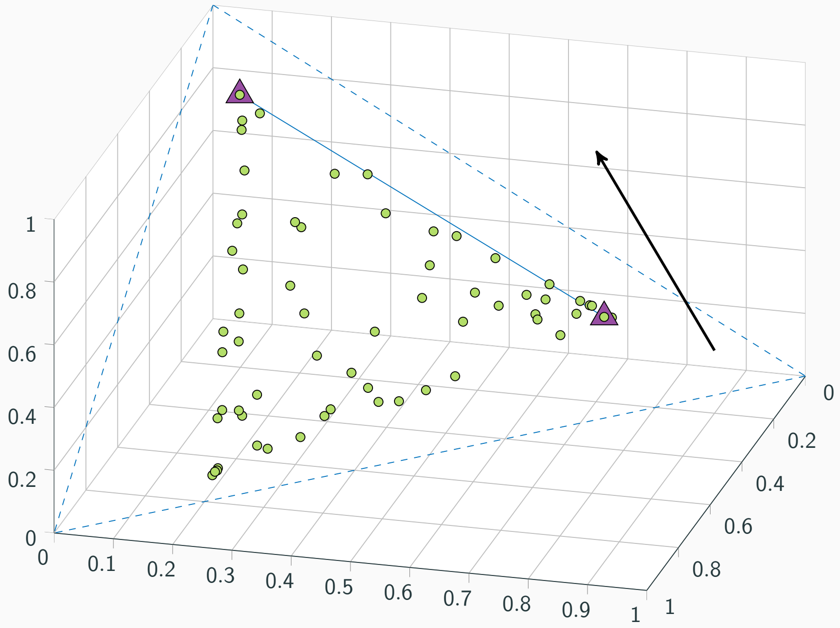
VCA — Animation



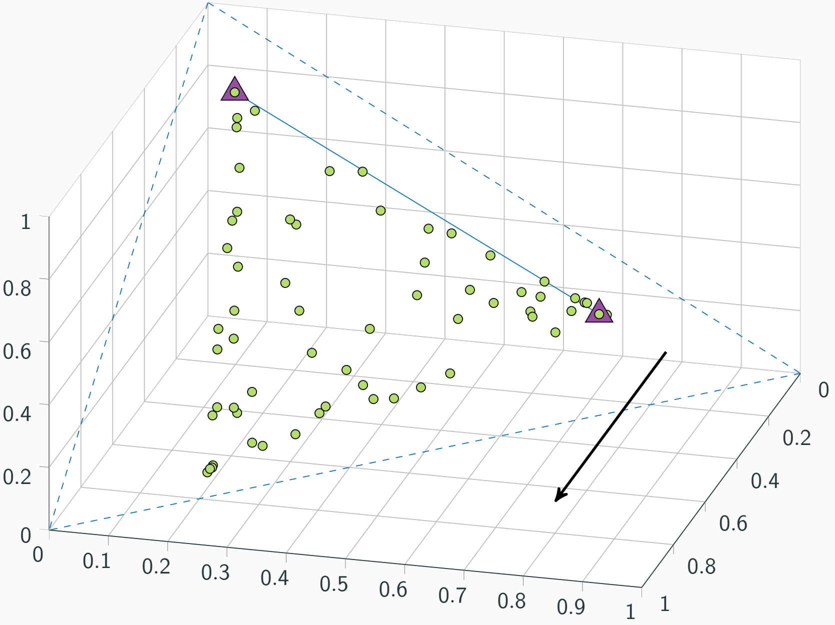
VCA — Animation



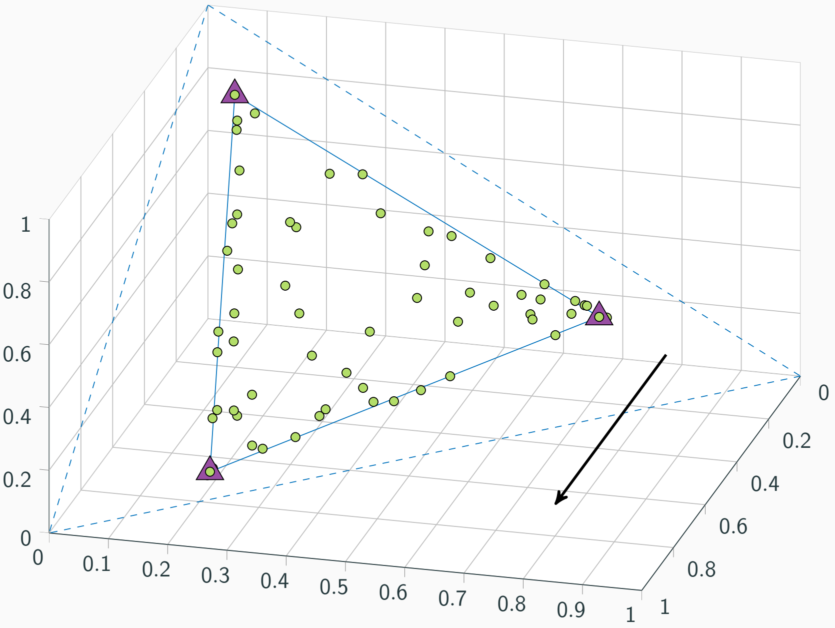
VCA — Animation



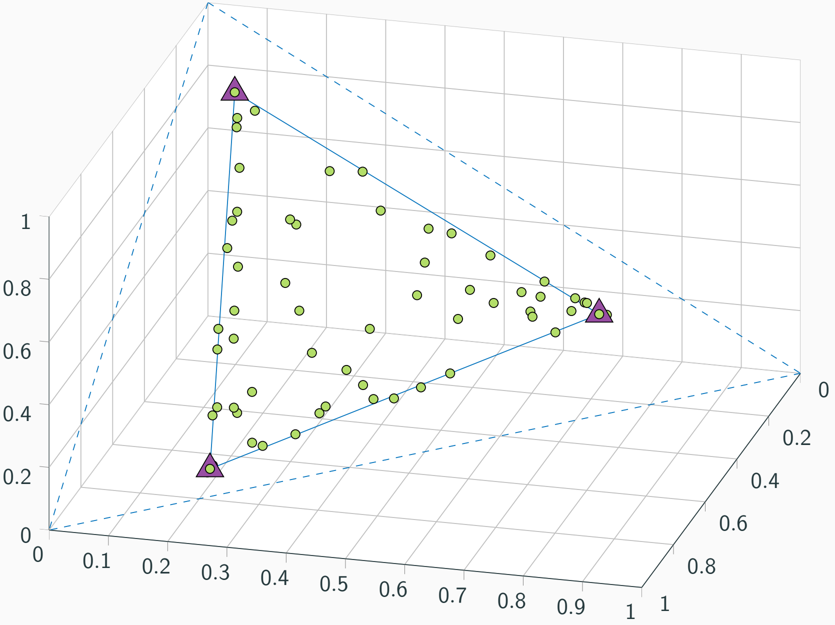
VCA — Animation



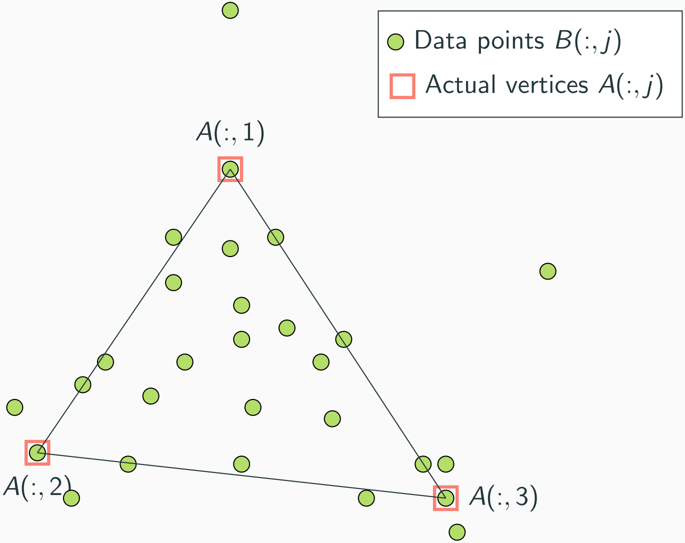
VCA — Animation



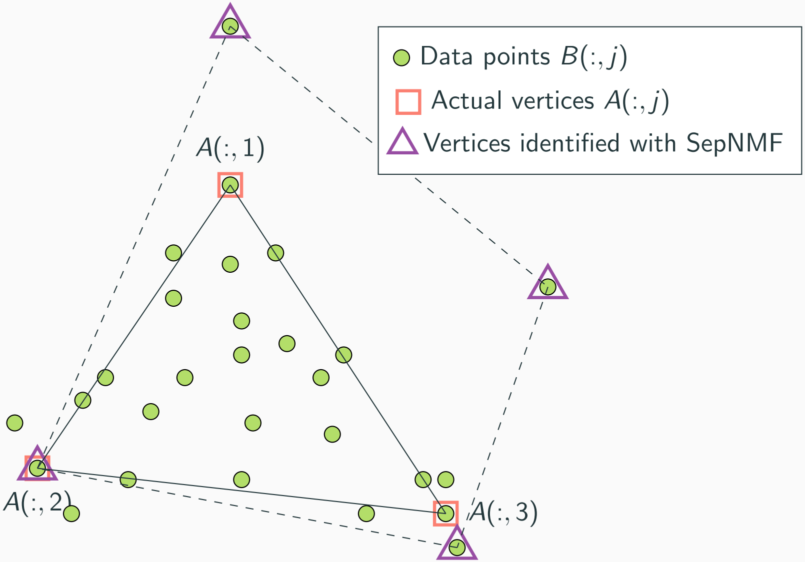
VCA — Animation



Issues of Separable NMF: outliers, extreme points

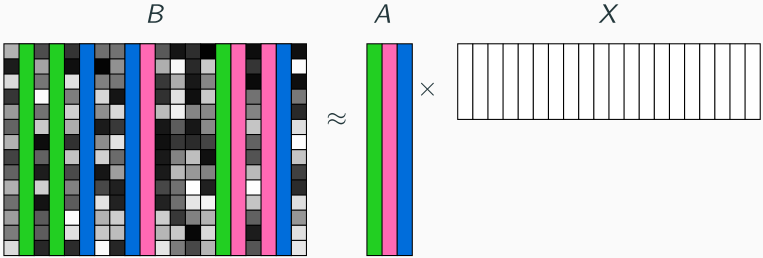


Issues of Separable NMF: outliers, extreme points



Model 2: Proximal latent points (Bhattacharyya et al. 2020)

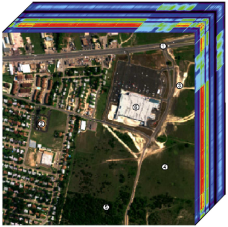
Interpretation: Each vertex has at least p data points close to it.

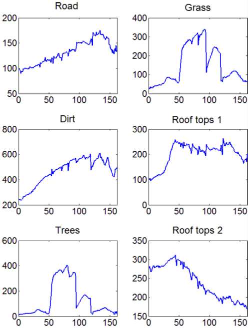


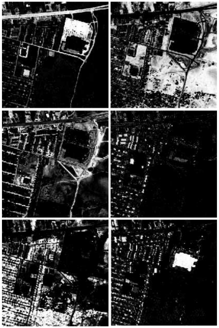
Model 2: Proximal latent points (Bhattacharyya et al. 2020)

- Assumption is stronger than separability, but it allows more noise, and is realistic in practice.
- The proposed Algorithm to Learn a Latent Simplex (ALLS) has practical issues.

$$\underbrace{B(:,j)}_{\substack{\text{spectral signature of}\\ \text{j-th pixel}}} \approx \sum_p \underbrace{A(:,p)}_{\substack{\text{spectral signature of}\\ \text{p-th material}}} \underbrace{X(p,j)}_{\substack{\text{abundance of p-th material}\\ \text{in j-th pixel}}}$$





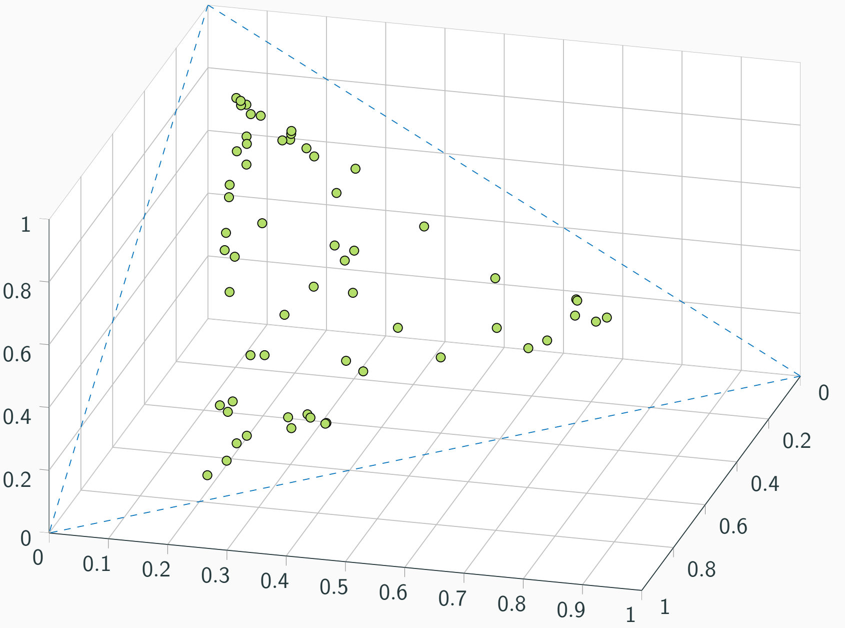


Images from Bioucas Dias and Nicolas Gillis.

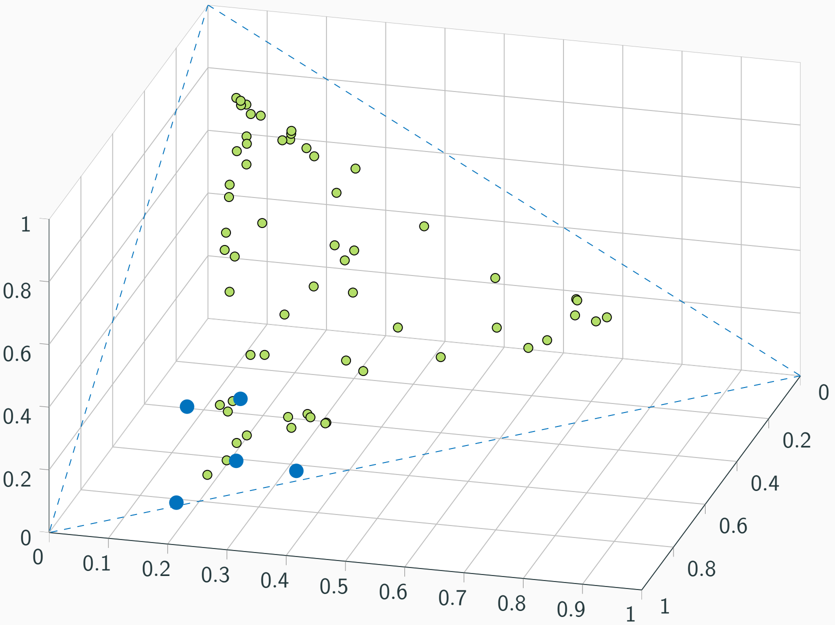
Our contribution

- Smoothed variants of algorithms VCA and SPA that leverage the proximal latent points assumption $\Rightarrow$ SVCA and SSPA
- Aggregates p data points to find each vertex
- Best of both worlds

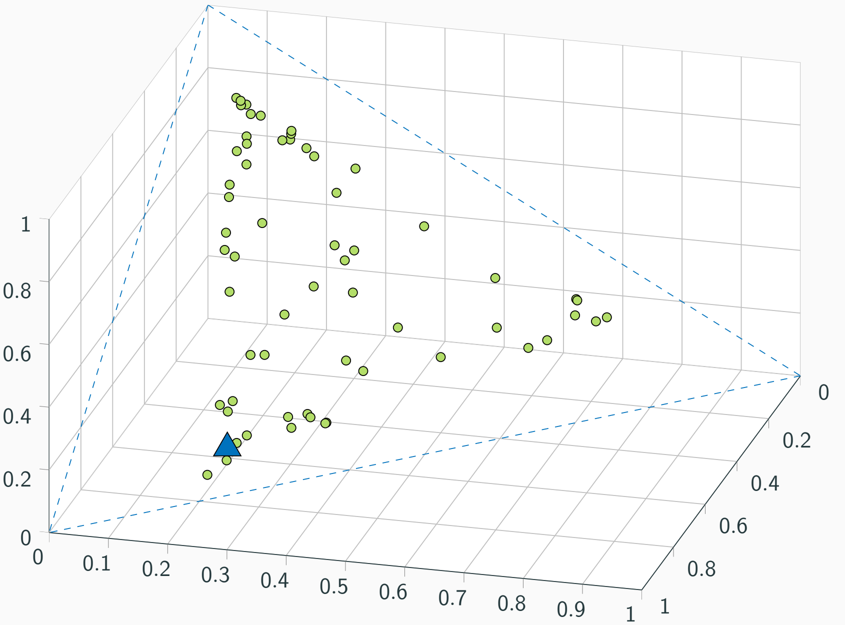
Smoothed VCA — Animation ($p = 5$)



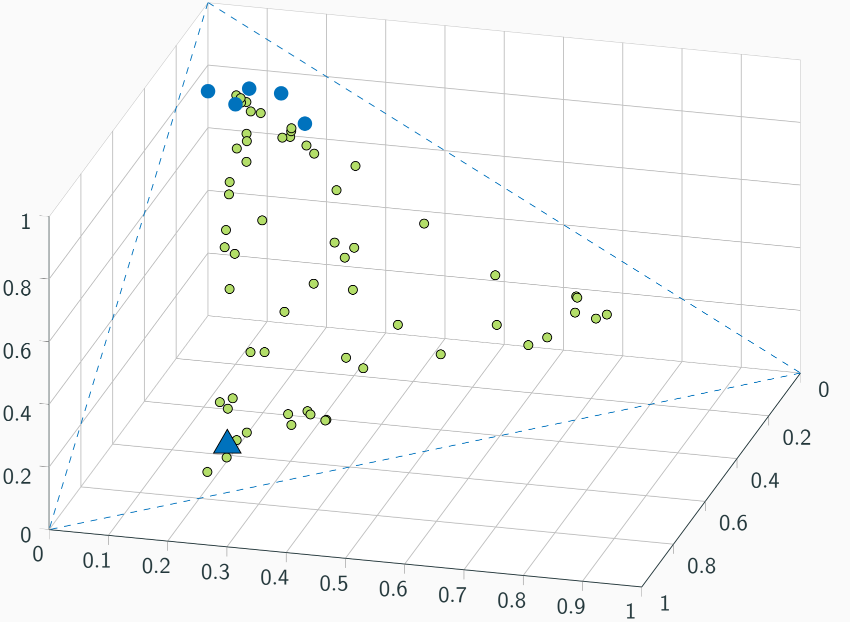
Smoothed VCA — Animation ($p = 5$)



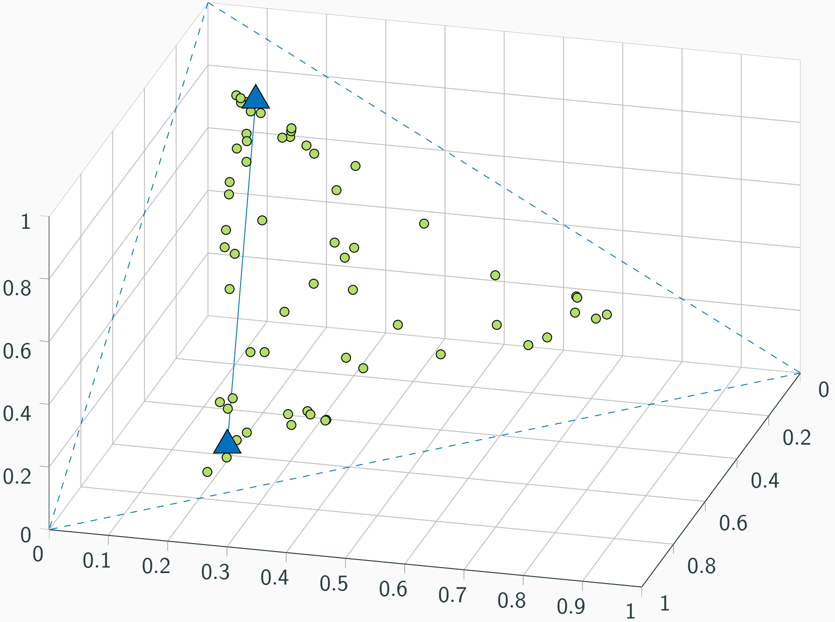
Smoothed VCA — Animation ($p = 5$)



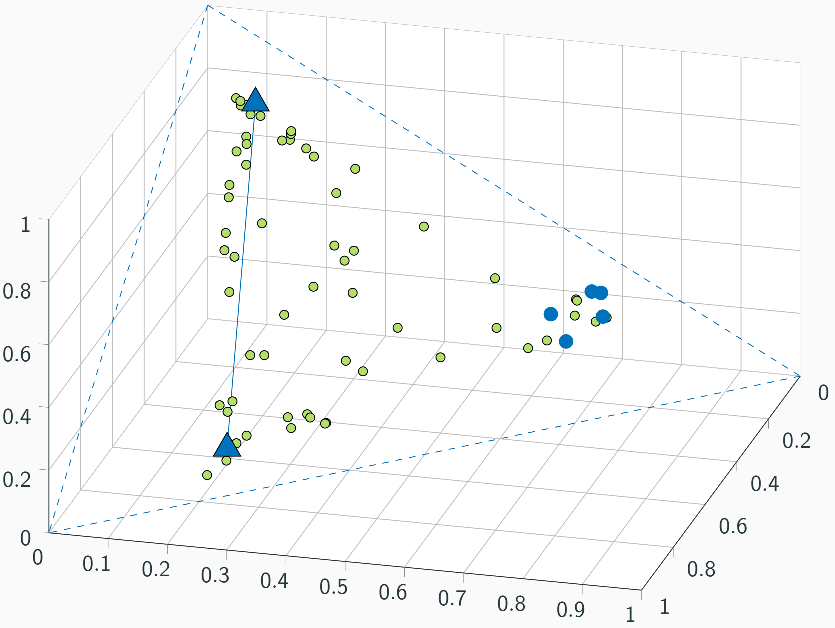
Smoothed VCA — Animation ($p = 5$)



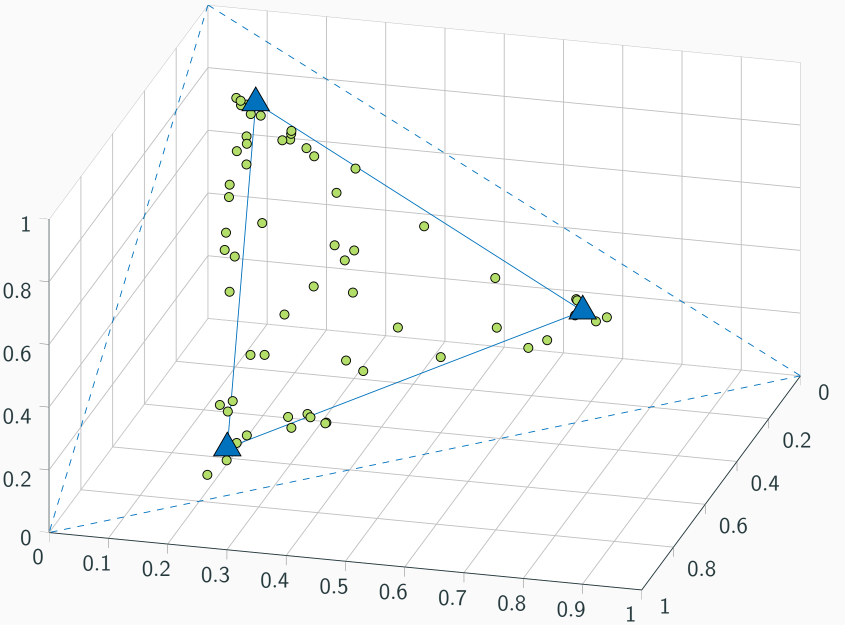
Smoothed VCA — Animation ($p = 5$)



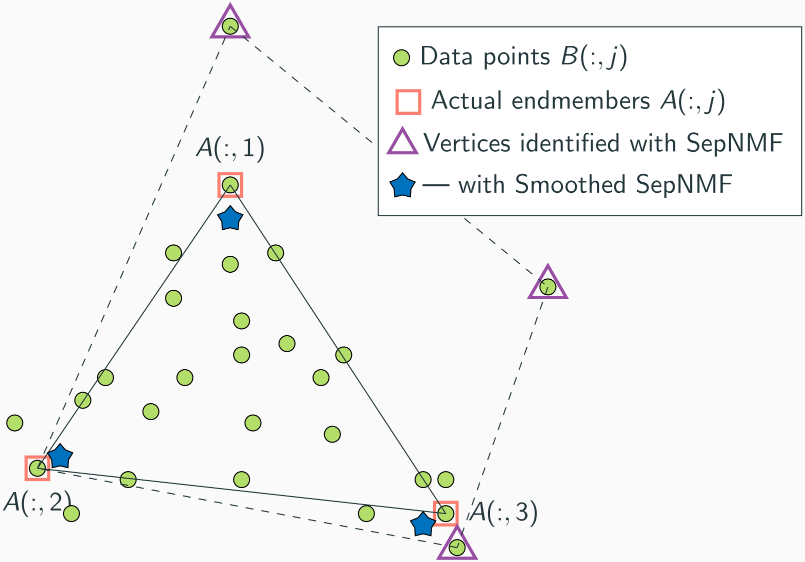
Smoothed VCA — Animation ($p = 5$)



Smoothed VCA — Animation ($p = 5$)



With smoothed separable NMF



Experiment: unmixing of hyperspectral image Urban

a VCA, error= 6.24%

b SVCA p=200, error= 5.24%

Conclusion

- Empirically, smoothed algorithm perform better than VCA, SPA, and ALLS
- More robust to outliers
- More robust to noise
- Good way to handle spectral variability.

Exact sparse nonnegative least squares

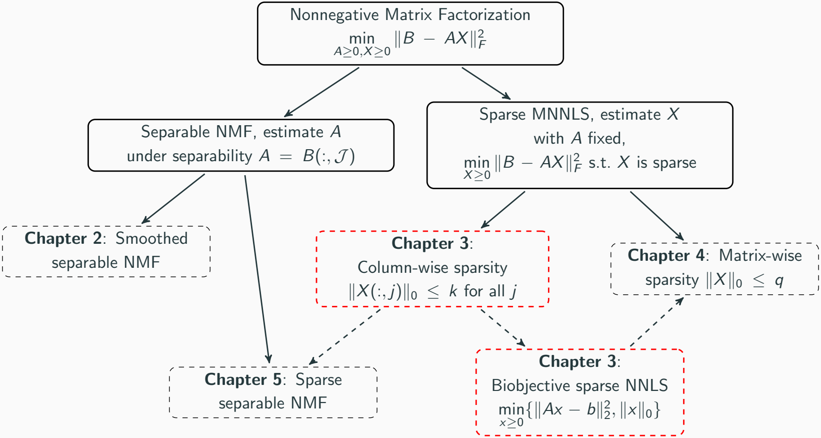
Exact sparse NNLS

Chapter 3 of the thesis. Presented in the articles:

NN, Arnaud Vandaele, Nicolas Gillis, and Jeremy E Cohen (2020). "Exact sparse nonnegative least squares". In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5395–5399.

— (2021). "Exact biobjective k-sparse nonnegative least squares". In: *29th European Signal Processing Conference (EUSIPCO)*, pp. 2079–2083.

Overview


Nonnegative Matrix Factorization
$\min_{A \geq 0, X \geq 0} \|B - AX\|_F^2$
Separable NMF, estimate A
under separability $A = B(:, \mathcal{J})$
Sparse MNNLS, estimate X
with A fixed,
$\min_{X \geq 0} \|B - AX\|_F^2$ s.t. X is sparse
Chapter 2: Smoothed
separable NMF
Chapter 3:
Column-wise sparsity
$\|X(:,j)\|_0 \leq k$ for all j
Chapter 4: Matrix-wise
sparsity $\|X\|_0 \leq q$
Chapter 5: Sparse
separable NMF
Chapter 3:
Biobjective sparse NNLS
$\min_{x \geq 0} \{\|Ax - b\|_2^2, \|x\|_0\}$

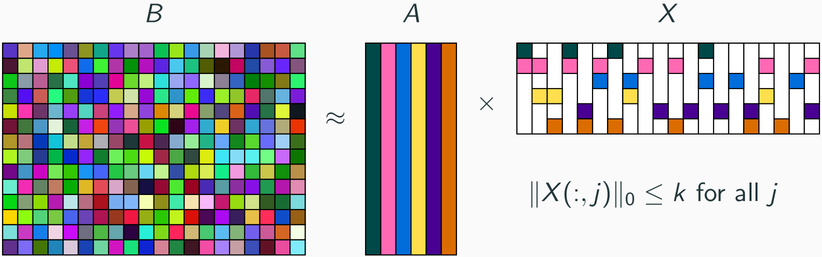
First contribution: exact algorithm

k-sparse NNLS: $\min_{x \geq 0} \|Ax - b\|_2^2$ s.t. $\|x\|_0 \leq k$

Intuitive formulation: each data point is a combination of at most k components

Why? No dedicated exact algorithm
What? Branch-and-bound algorithm

k-sparse NNLS in a multi-column problem



$\|X(:,j)\|_0 \le k$ for all j

Exact Sparse Nonnegative Least Squares

- k-sparse NNLS
$$\min_{x\geq 0} \|Ax - b\|_2^2 \text{ s.t. } \|x\|_0 \leq k$$
is a combinatorial problem
- Reduces to find the best support of cardinality k
- $\binom{r}{k}$ possible supports

Can we do better than brute-force?

Pruning the search space

How can we exploit the problem's structure to prune safely the search space?

- Branch-and-bound
- Idea: when adding constraints to a problem, the optimal solution can only worsen (or stay the same)
- Our algorithm: arborescent[1]

[1]arborescent Realizes a Branch-and-bound Optimization to Require Explicit Sparsity Constraints to be Enforced in NNLS Tasks

Illustration of arborescent, $r = 5$ and $k = 2$

root node, unconstrained

$\boxed{\mathrm{X} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5]}$ $k = r = 5$

Illustration of arborescent, $r = 5$ and $k = 2$

root node, unconstrained

$\boxed{\mathrm{X} = [x_1\ x_2\ x_3\ x_4\ x_5]}$ $k = r = 5$

Illustration of arborescent, $r = 5$ and $k = 2$

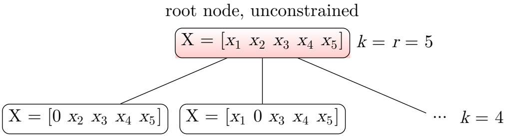


Illustration of arborescent, $r = 5$ and $k = 2$

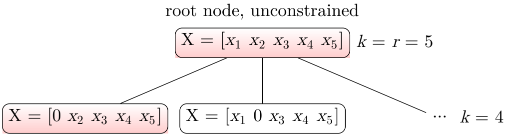


Illustration of arborescent, $r = 5$ and $k = 2$

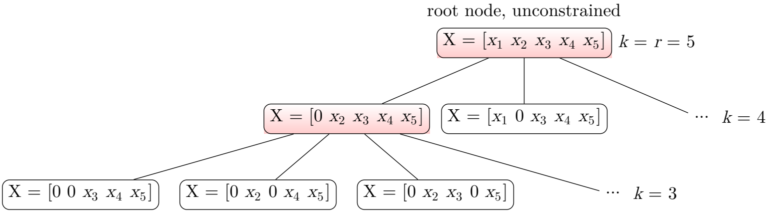


Illustration of arborescent, $r = 5$ and $k = 2$

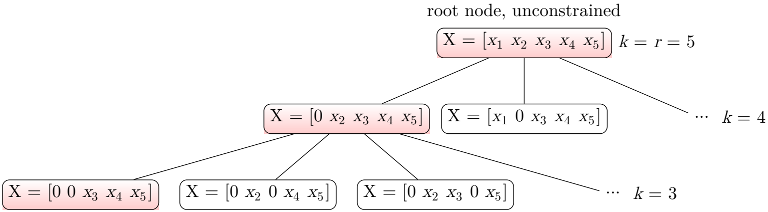


Illustration of arborescent, $r = 5$ and $k = 2$

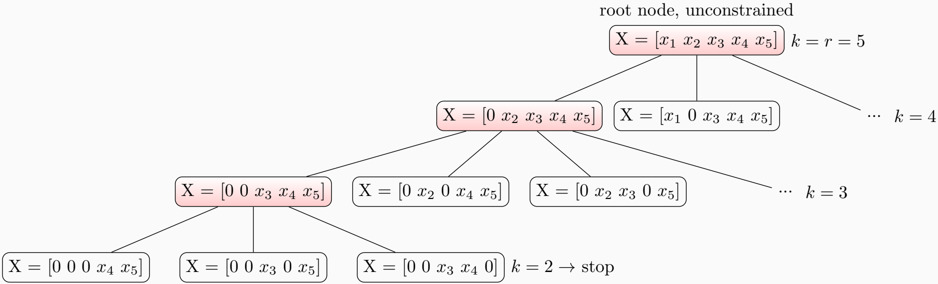


Illustration of arborescent, $r = 5$ and $k = 2$

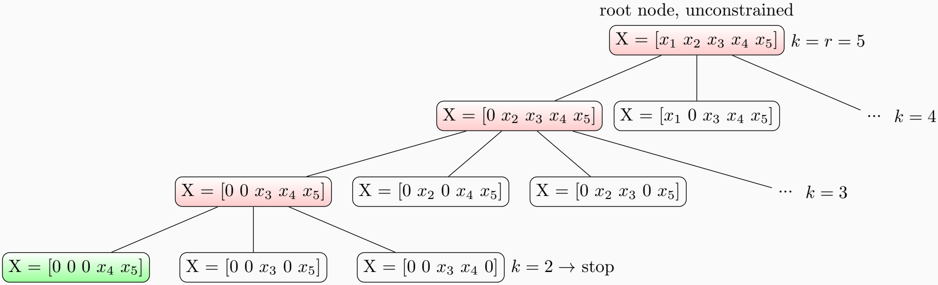


Illustration of arborescent, $r = 5$ and $k = 2$

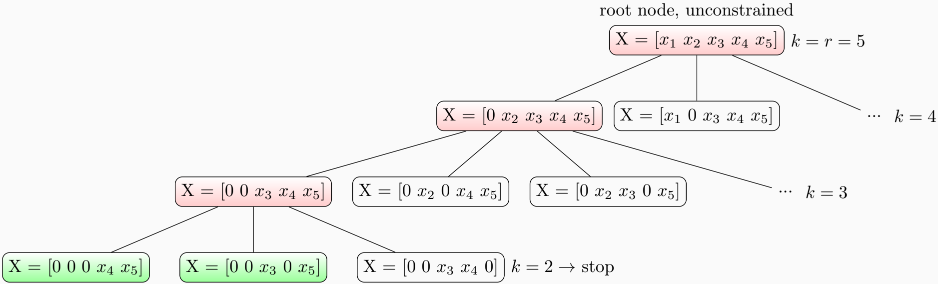


Illustration of arborescent, $r = 5$ and $k = 2$

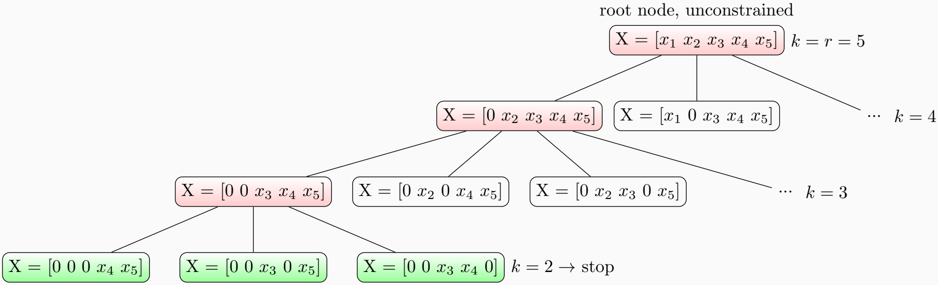


Illustration of arborescent, $r = 5$ and $k = 2$

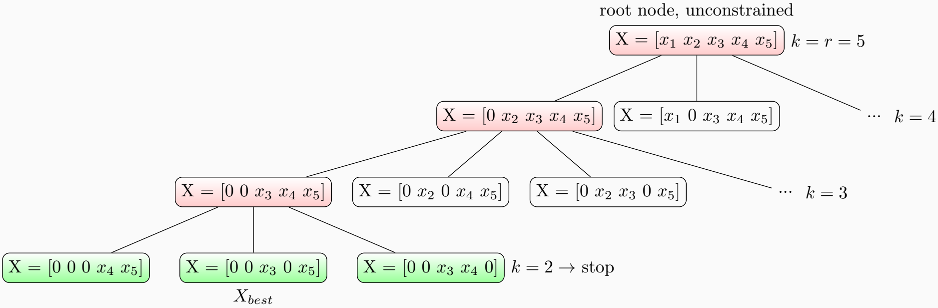


Illustration of arborescent, $r = 5$ and $k = 2$

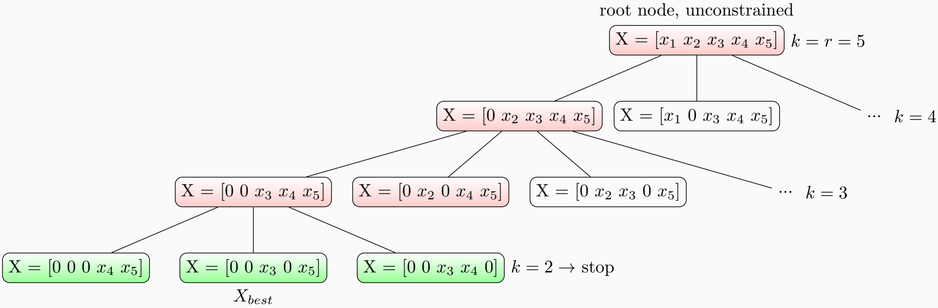


Illustration of arborescent, $r = 5$ and $k = 2$

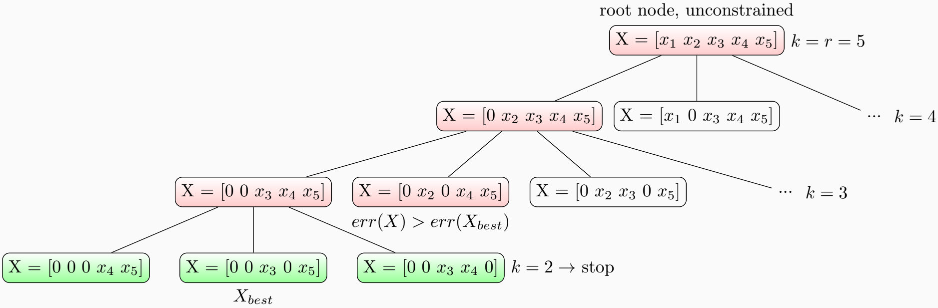
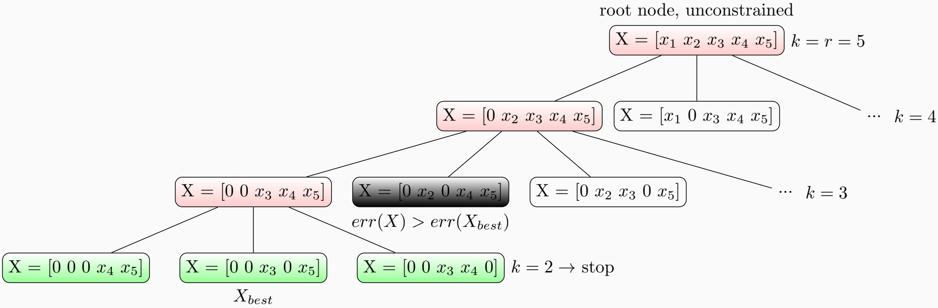
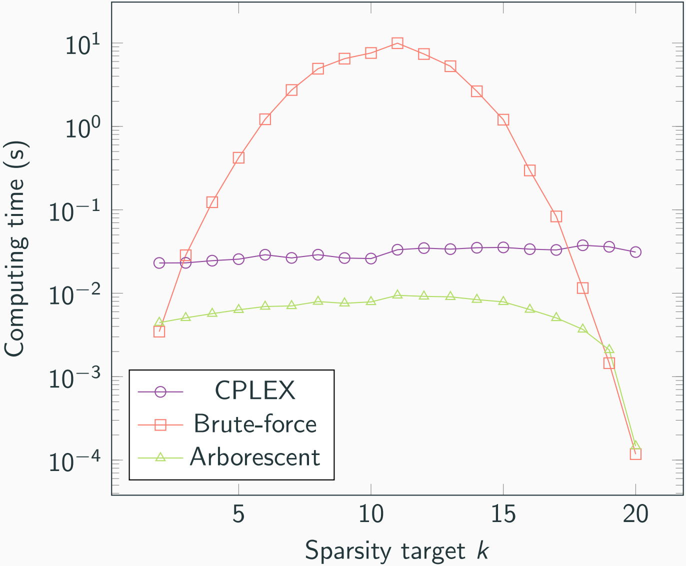


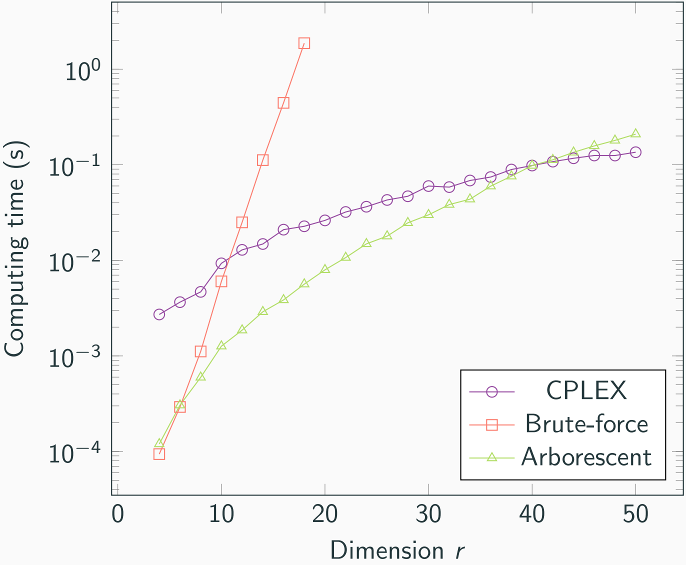
Illustration of arborescent, $r = 5$ and $k = 2$



Comparison with brute force and generic MIP solvers



Comparison with brute force and generic MIP solvers



Second contribution: biobjective extension

Why? Constrained formulation is not always practical

- k can be difficult to estimate
- In a multicolumn problem, k can vary between columns

What? Biobjective extension of arborescent

Biobjective k-sparse NNLS:

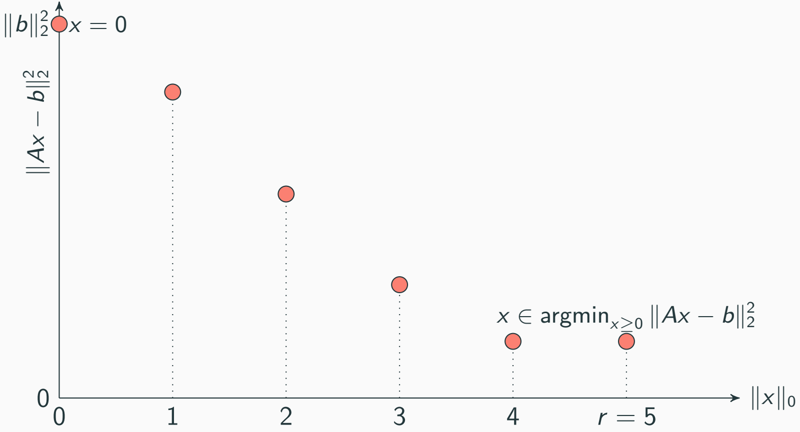
$$\min_{x \geq 0} \{\|Ax - b\|_2^2, \|x\|_0\}$$

Bi-objective formulation

$$\min_{x \geq 0} \begin{cases} \|Ax - b\|_2^2 \\ \|x\|_0 \end{cases}$$

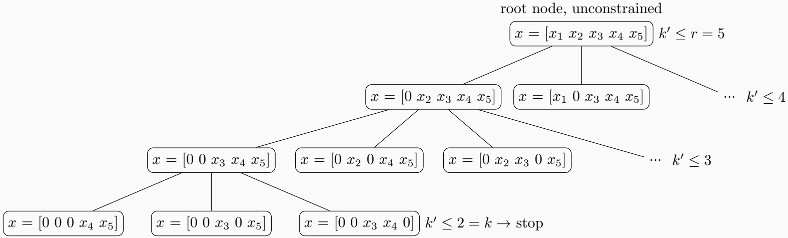
Equivalent to $\min_{x \geq 0} \|b - Ax\|_2^2$ s.t. $\|x\|_0 \leq k$ for all $k \in \{0, \ldots, r\}$

Example for $r = 5$



How to solve the biobjective problem?

An extension of the existing branch-and-bound algorithm for k-sparse NNLS

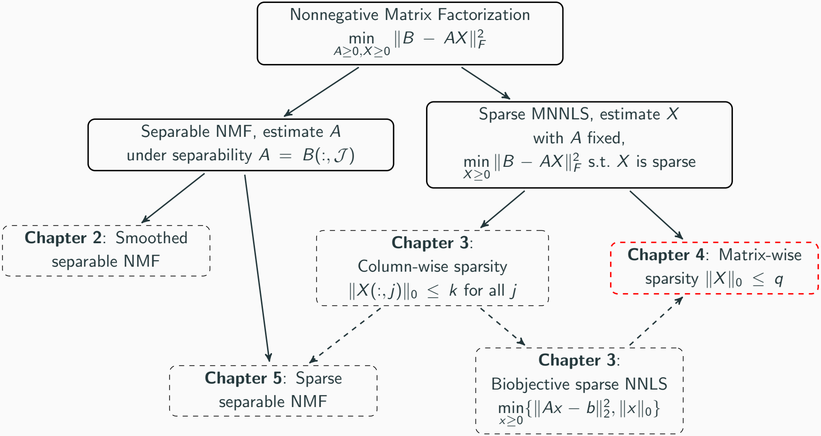


Sum-up

- We proposed arborescent, a branch-and-bound algorithm to solve exactly the k-sparse NNLS problem.
- Faster than brute force and generic solver
- Biobjective extension
  - Useful when k is hard to set
  - Can be used as a subroutine in a larger framework (next chapter...)

Matrix-wise ℓ_0-constrained nonnegative least squares

Overview


Nonnegative Matrix Factorization
$\min_{A \geq 0, X \geq 0} \|B - AX\|_F^2$
Separable NMF, estimate A
under separability $A = B(:, \mathcal{J})$
Sparse MNNLS, estimate X
with A fixed,
$\min_{X \geq 0} \|B - AX\|_F^2$ s.t. X is sparse
Chapter 2: Smoothed
separable NMF
Chapter 3:
Column-wise sparsity
$\|X(:, j)\|_0 \leq k$ for all j
Chapter 4: Matrix-wise
sparsity $\|X\|_0 \leq q$
Chapter 5: Sparse
separable NMF
Chapter 3:
Biobjective sparse NNLS
$\min_{x \geq 0} \{\|Ax - b\|_2^2, \|x\|_0\}$

Matrix-wise ℓ_0-constrained NNLS

Chapter 4 of the thesis. Presented in the article:

NN, Jeremy E. Cohen, Arnaud Vandaele, and Nicolas Gillis (2022). "Matrix-wise L0-constrained sparse nonnegative least squares". In: *preprint arXiv:2011.11066*.

Why? Column-wise sparsity is sometimes not practical, few works handle matrix-wise sparsity (mostly heuristics, e.g. ℓ_1-relaxation)

What? Algorithmic framework with optimality guarantees under conditions

A matrix-wise constraint

Matrix-wise q-sparse MNNLS

$$\min_{X \geq 0} \|B - AX\|_2^2 \text{ s.t. } \|X\|_0 \leq q$$

- Can be seen as a global sparsity budget
- If $q = k \times n$, this enforces an average k-sparsity on the columns of X

A matrix-wise constraint

Matrix-wise q-sparse MNNLS

$$\min_{X \geq 0} \|B - AX\|_2^2 \text{ s.t. } \|X\|_0 \leq q$$

- Can be seen as a global sparsity budget
- If $q = k \times n$, this enforces an average k-sparsity on the columns of X

How to solve it?

- With a k-sparse NNLS methods, by vectorizing the problem
  $\Rightarrow$ leads to a huge NNLS problem, too expensive to solve
- Our contribution: dedicated algorithm

Our contribution: a two-step algorithm

Algorithm Salmon[2]:

1. Generate a set of solutions for every column of X, with different tradeoffs between reconstruction error and sparsity
   - Divide the sparse MNNLS problem into n biobjective sparse NNLS subproblems

$$\min_{X(:,j)\geq 0} \left\{ \quad \|B(:,j) - AX(:,j)\|_2^2 \quad , \quad \|X(:,j)\|_o \quad \right\}$$

   - Solve with arborescent, or heuristic (homotopy, greedy algo)
   - Build a cost matrix C

[2] Salmon Applies ℓ_0-constraints Matrix-wise On NNLS problems

Our contribution: a two-step algorithm

Algorithm Salmon[2]:

1. Generate a set of solutions for every column of X, with different tradeoffs between reconstruction error and sparsity
   - Divide the sparse MNNLS problem into n biobjective sparse NNLS subproblems

$$\min_{X(:,j)\geq 0}\left\{ \quad \|B(:,j) - AX(:,j)\|_2^2 \quad , \quad \|X(:,j)\|_o \quad \right\}$$

   - Solve with arborescent, or heuristic (homotopy, greedy algo)
   - Build a cost matrix C
2. Select one solution per column such that in total X has q nonzero entries and the error is minimized $\Rightarrow$ assignment-like problem
   - Dedicated greedy algorithm proved near-optimal

[2] Salmon Applies ℓ_0-constraints Matrix-wise On NNLS problems

Cost matrix C

- Each row = one sparsity level
- Each column = one column of the MNNLS problem

$$\begin{pmatrix} C_{0,1} & C_{0,2} & \cdots & C_{0,n} \\ C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r,1} & C_{r,2} & \cdots & C_{r,n} \end{pmatrix}$$

Illustration of Salmon — Step 1

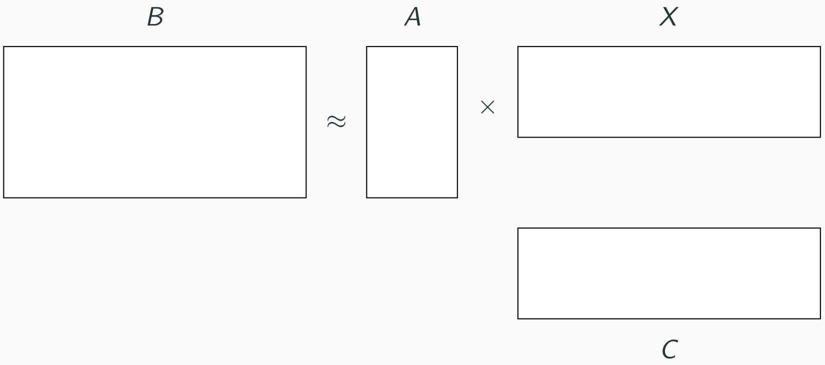


Illustration of Salmon — Step 1

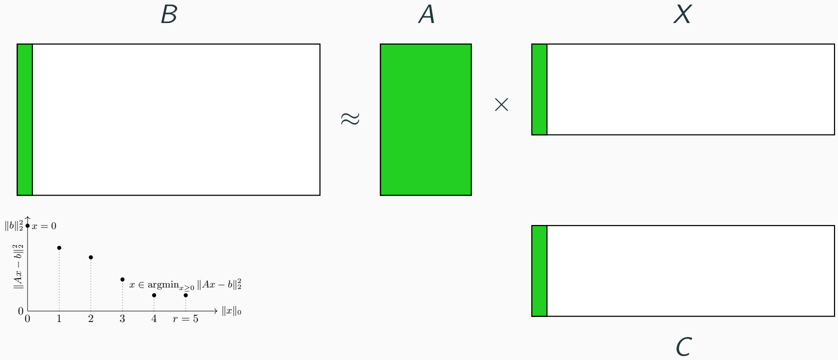


Illustration of Salmon — Step 1

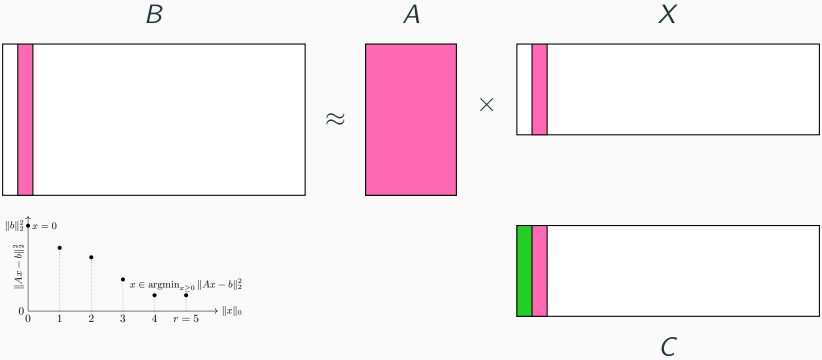


Illustration of Salmon — Step 1

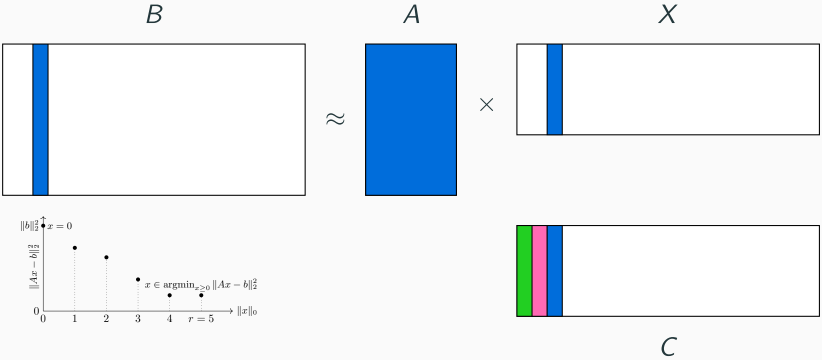


Illustration of Salmon — Step 1

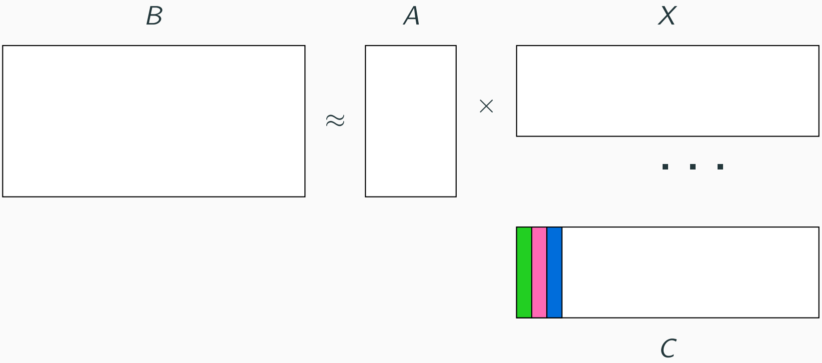
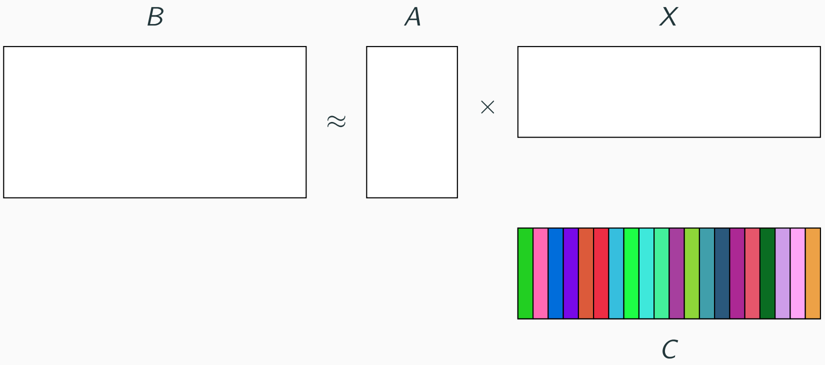


Illustration of Salmon — Step 1



Salmon — Step 2

Similar to an assignment problem

$$\begin{pmatrix} C_{0,1} & C_{0,2} & \cdots & C_{0,n} \\ C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r,1} & C_{r,2} & \cdots & C_{r,n} \end{pmatrix}$$

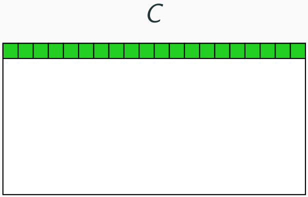
Given $z_{i,j} \in \{0,1\}$ such that $z_{i,j} = 1$ if and only if the jth column of X is i-sparse,

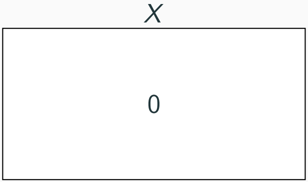
$$\min_{z \in \{0,1\}^{r \times n}} \sum_{i,j} z_{i,j} C(i,j)$$

$$\text{such that } \sum_i z_{i,j} = 1 \text{ for all } j, \quad \text{and} \quad \sum_{i,j} i\, z_{i,j} \leq q.$$

Solved with a dedicated greedy algorithm, fast but proved near-optimal

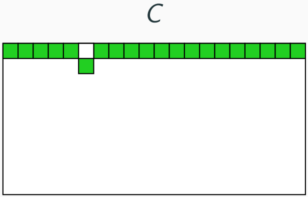
Salmon — Step 2

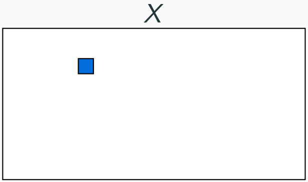




$\|X\|_0 = 0$

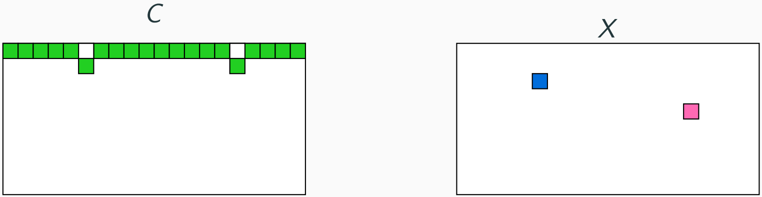
Salmon — Step 2





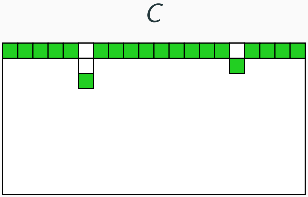
$\|X\|_0 = 1$

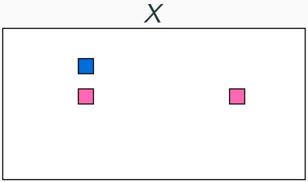
Salmon — Step 2



$\|X\|_0 = 2$

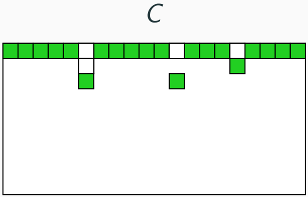
Salmon — Step 2

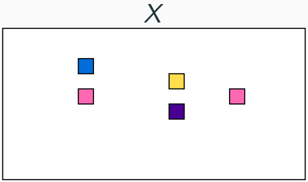




$\|X\|_0 = 3$

Salmon — Step 2

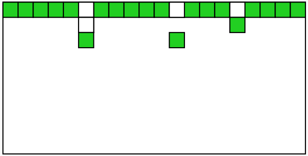




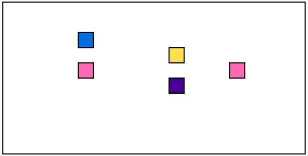
$\|X\|_0 = 5$

Salmon — Step 2

C



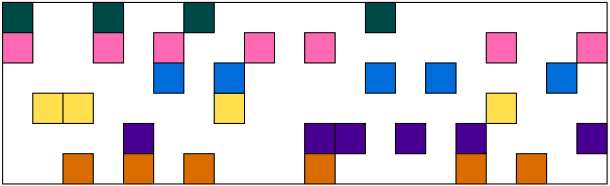
X



$\|X\|_0 = 5$

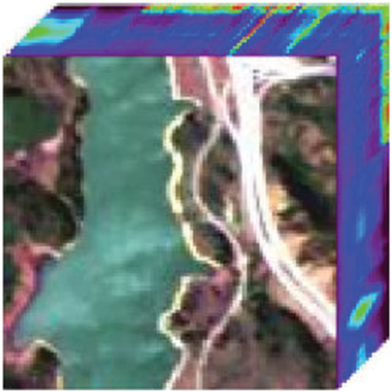
Iterate while $\|X\|_0 < q$

Salmon — Step 2



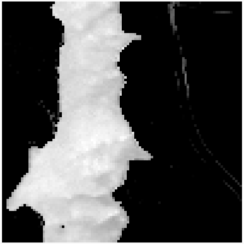
Final solution X, q-sparse matrix

Experiment: unmixing of hyperspectral image Jasper

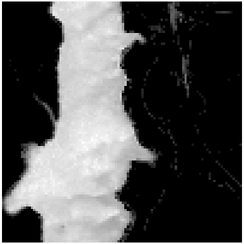


Experiment: unmixing of hyperspectral image Jasper

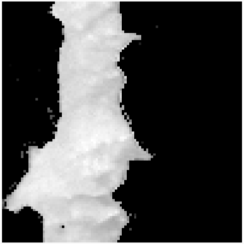
NNLS (no sparse)



Col-wise, $k = 2$



Salmon, $q/n = 2$



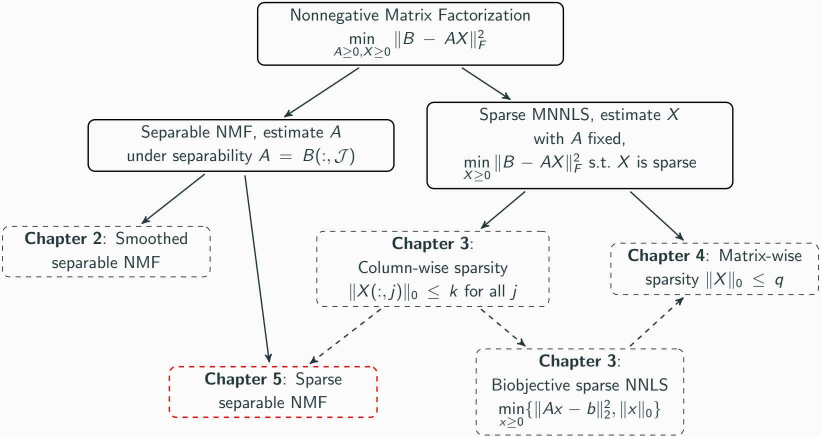
Salmon, $q/n = 1.8$

Conclusion

- We introduced a sparse MNNLS model with matrix-wise ℓ_0-sparsity constraint
- We developed a 2-step algorithm to tackle it
- Makes tractable some problems that are too big for standard NNLS solvers
- Improves results, allows a finer parameter tuning
- Interesting where sparsity varies between columns

Sparse separable nonnegative matrix factorization

Overview



Nonnegative Matrix Factorization
$\min_{A\geq 0, X\geq 0} \|B - AX\|_F^2$
Separable NMF, estimate A
under separability $A = B(:, \mathcal{J})$
Sparse MNNLS, estimate X
with A fixed,
$\min_{X\geq 0} \|B - AX\|_F^2$ s.t. X is sparse
Chapter 2: Smoothed
separable NMF
Chapter 3:
Column-wise sparsity
$\|X(:, j)\|_0 \leq k$ for all j
Chapter 4: Matrix-wise
sparsity $\|X\|_0 \leq q$
Chapter 5: Sparse
separable NMF
Chapter 3:
Biobjective sparse NNLS
$\min_{x\geq 0}\{\|Ax - b\|_2^2, \|x\|_0\}$

Sparse separable nonnegative matrix factorization

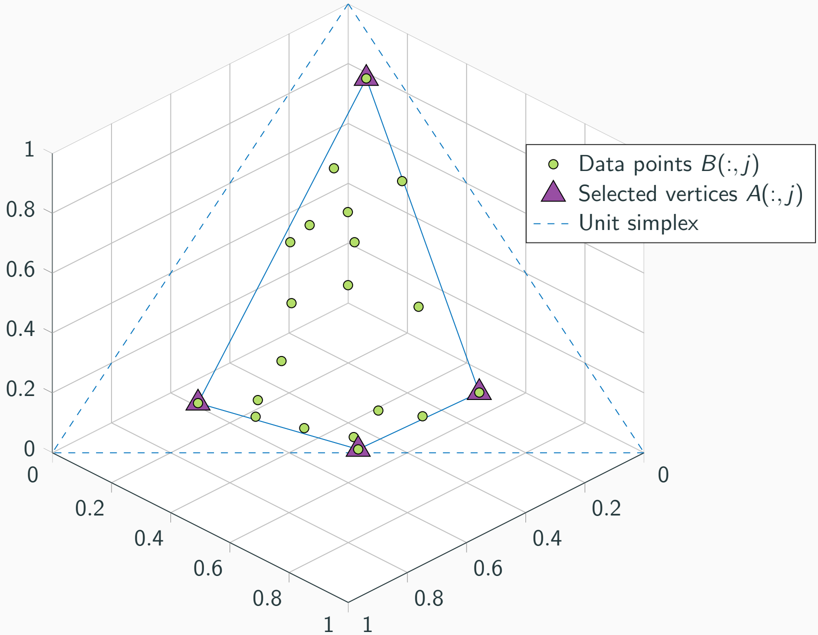
Chapter 5 of the thesis. Presented in the article:

NN, Arnaud Vandaele, Jeremy E Cohen, and Nicolas Gillis (2020). "Sparse separable nonnegative matrix factorization". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECMLPKDD)*, pp. 335–350.

Why? No work handles the underdetermined case with interior vertices, nor leverages sparsity

What? New model and exact algorithm for separable NMF with sparsity constraints, identifiability and complexity proofs

Starting point — Separable NMF



A limitation of Separable NMF

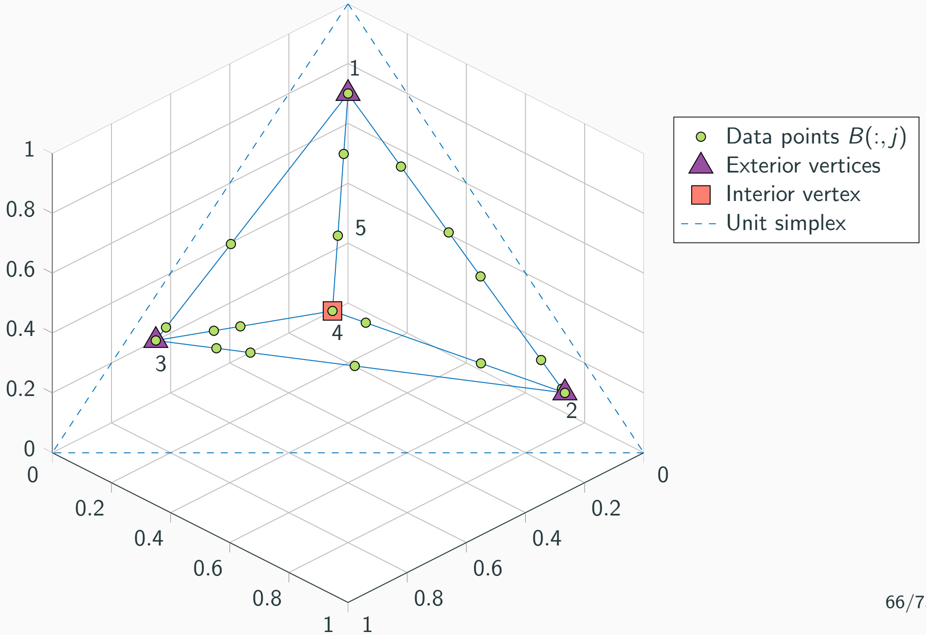
What if one column of A is a combination of others columns of A?

Ex: multispectral unmixing with $m < r$

$\rightarrow$ Interior vertex

Not identifiable by separable NMF, because it belongs to the convex hull of the other vertices.

A limitation of Separable NMF



Combining separability and k-sparsity

Sparse separable NMF

$B = B(:,\mathcal{J})X$ s.t. for all i, $\|X(:,i)\|_0 \leq k$

Given B, find $\mathcal{J}$ and X.

Our approach for SSNMF

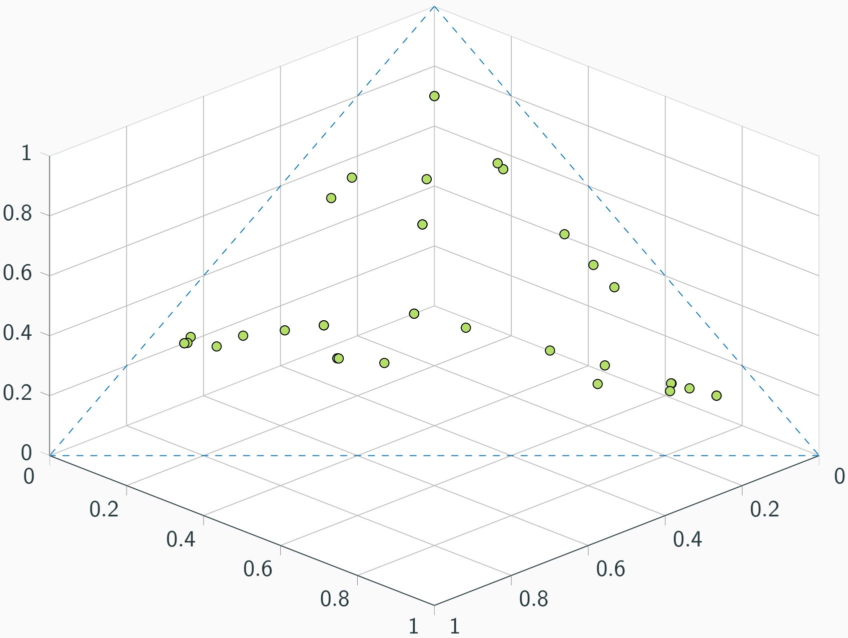
In a nutshell, 3 steps:

1. Identify exterior vertices with Separable NMF algorithm (SNPA)
2. Identify candidate interior vertices with k-sparse SNPA
3. Discard bad candidates, those that are k-sparse combinations of other selected points (they cannot be vertices)

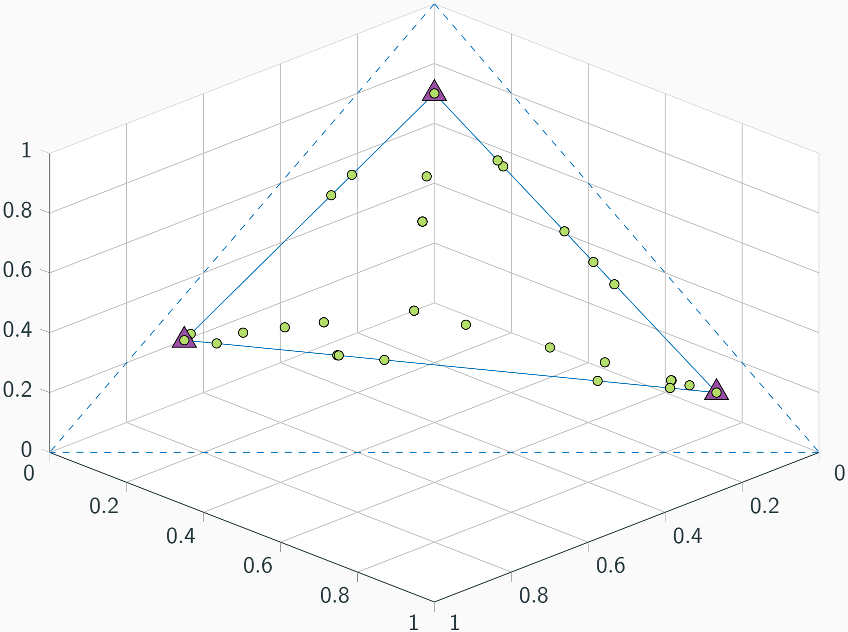
Our algorithm: Brassens[3]

[3] Brassens Relies on Assumptions of Separability and Sparsity for Elegant NMF Solving

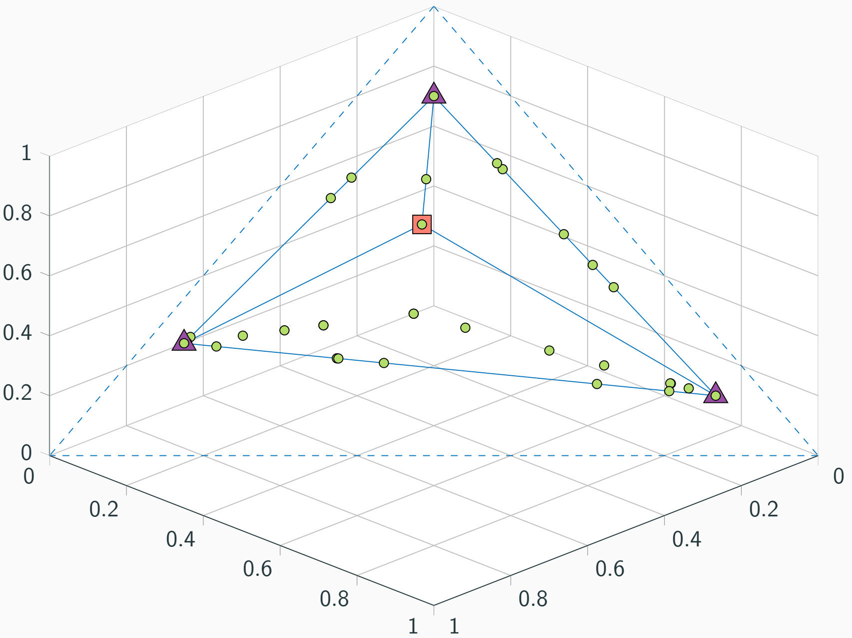
Brassens with sparsity $k = 2$



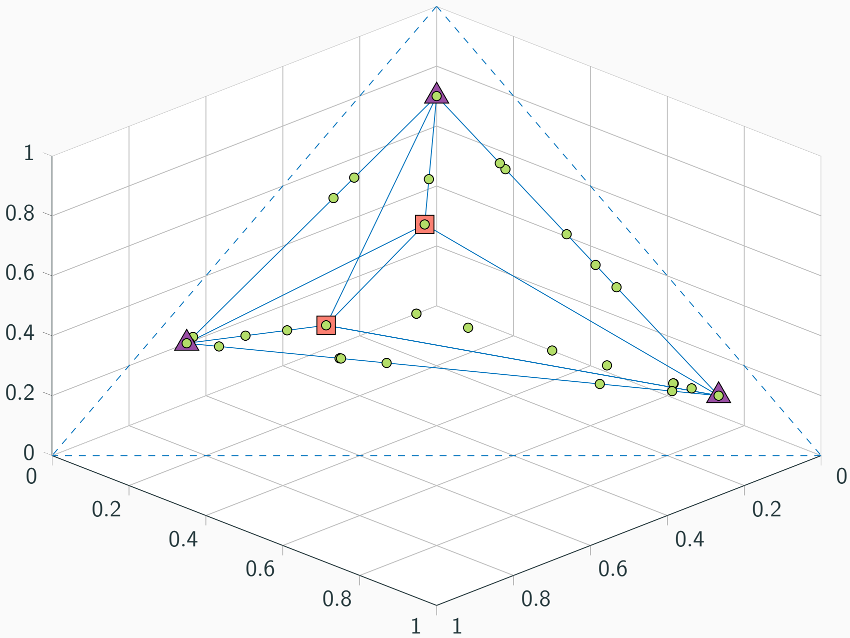
Brassens with sparsity $k = 2$



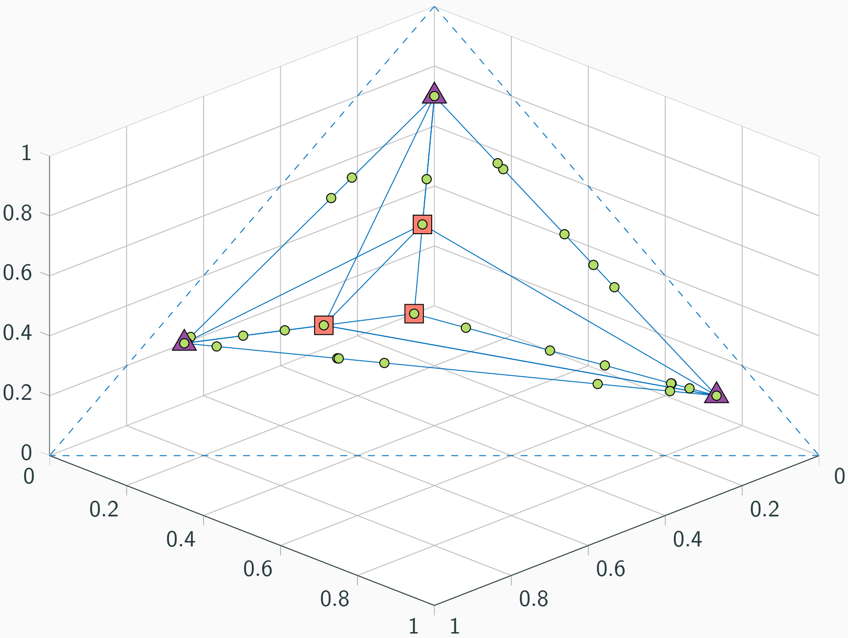
Brassens with sparsity $k = 2$



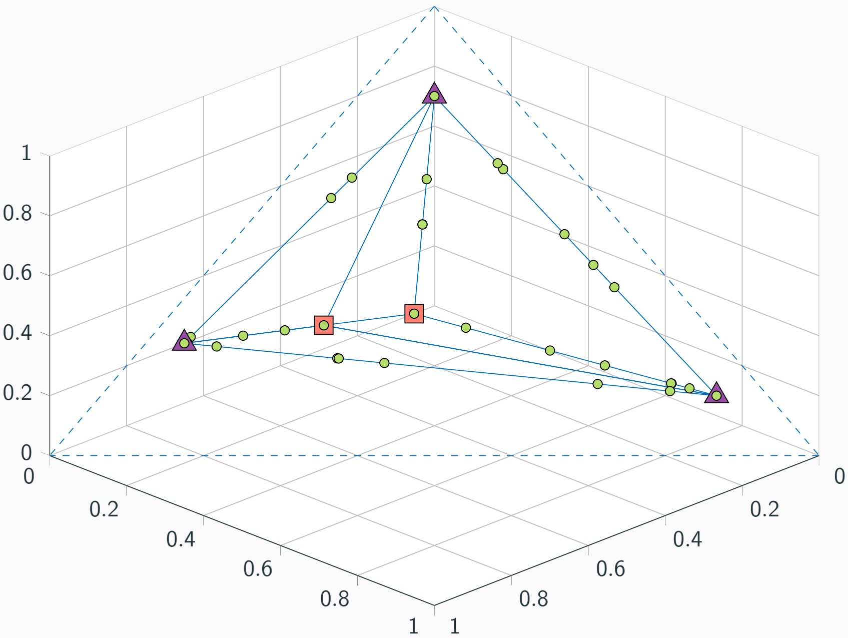
Brassens with sparsity $k = 2$



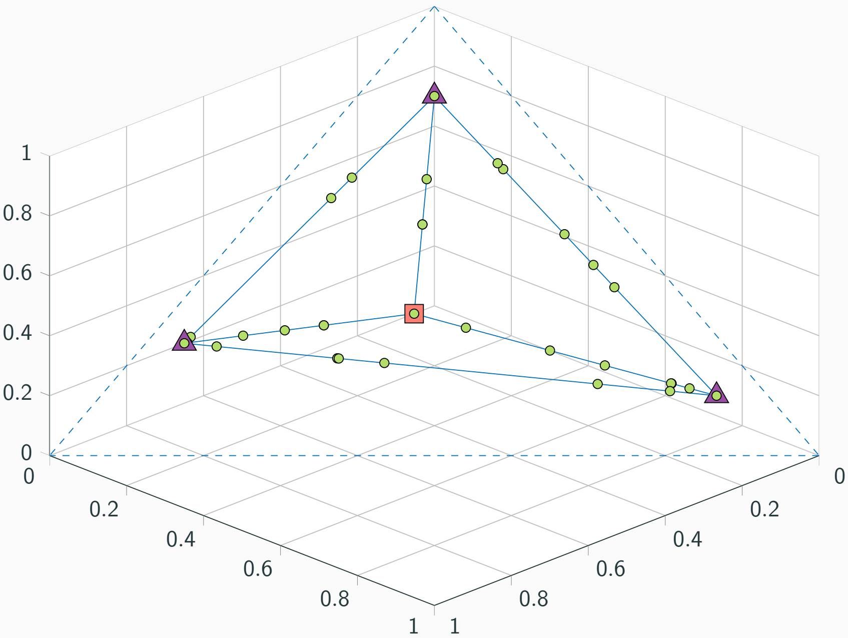
Brassens with sparsity $k = 2$



Brassens with sparsity $k = 2$



Brassens with sparsity $k = 2$



Sum-up

Sparse Separable NMF, a new model that combine constraints of separability and k-sparsity:

- Can handle some cases that Separable NMF cannot handle, such as interior vertices
- We proved it is NP-hard (unlike Sep NMF), but actually "not so hard" for small r
- It is provably solved by our algorithm Brassens under mild assumptions

Limitations:

- Brassens does not scale well
- Theoretical results limited to the noiseless case

Conclusion

Conclusion

Our contributions:

- Leverage more a priori knowledge to improve models
- Focus on ℓ_0-"norm" constraints: more intuitive formulations for sparse models
- Provide exact algorithms: guaranteed results but with higher computing cost

Future lines of research

- A whole new class of smoothed separable NMF algorithms
- Better branch-and-bound algorithms
- Generalize our algorithms to other sparse optimization problems (e.g. simultaneous sparse optimization)
- Enforce other discrete contraints (binary, integer, . . .) using combinatorial techniques, such as branch-and-bound
- Study the sparsity assumption in other kinds of data and applications: audio processing, text mining, chemometrics, . . .

Thanks!

Contact: `nicolas.nadisic@umons.ac.be`

Thesis, paper and code:

`http://nicolasnadisic.xyz`