Sparsity and Nonnegativity in Least Squares Problems and Matrix Factorizations

PhD Defense

Nicolas Nadisic

22 March 2022

Université de Mons, Belgium

Outline

1. Introduction
2. Smoothed separable nonnegative matrix factorization
3. Exact sparse nonnegative least squares
4. Matrix-wise ℓ_0-constrained nonnegative least squares
5. Sparse separable nonnegative matrix factorization
6. Conclusion

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: Nonnegative matrix factorization

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

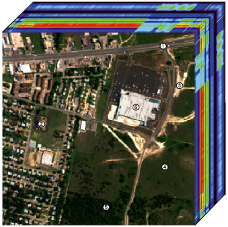
$$\min_{A\geq 0, X\geq 0} \|B - AX\|_F^2$$

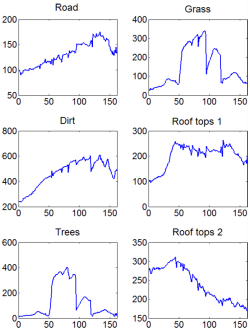
Why nonnegativity?

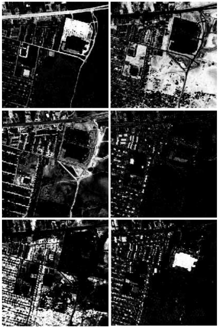
- More interpretable factors (part-based representation)
- Naturally favors sparsity
- Is natural in many applications (image processing, hyperspectral unmixing, text mining, . . .)

One application — Hyperspectral unmixing

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

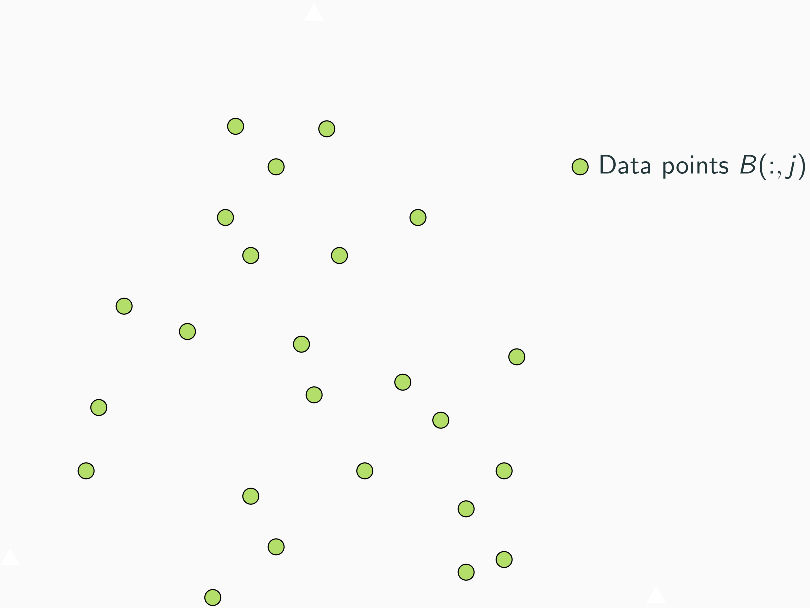




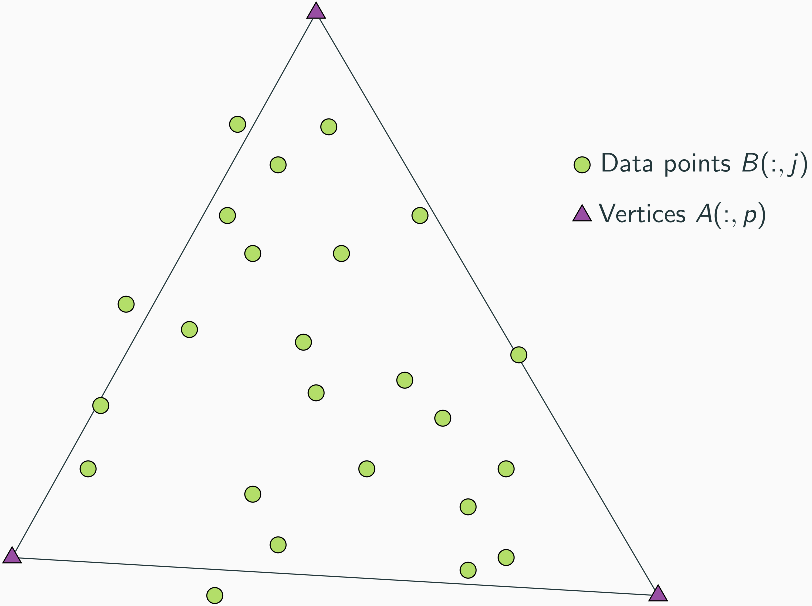


Images from Bioucas Dias and Nicolas Gillis.

NMF Geometry ($B \approx AX$)



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



Nonnegative least squares

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

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

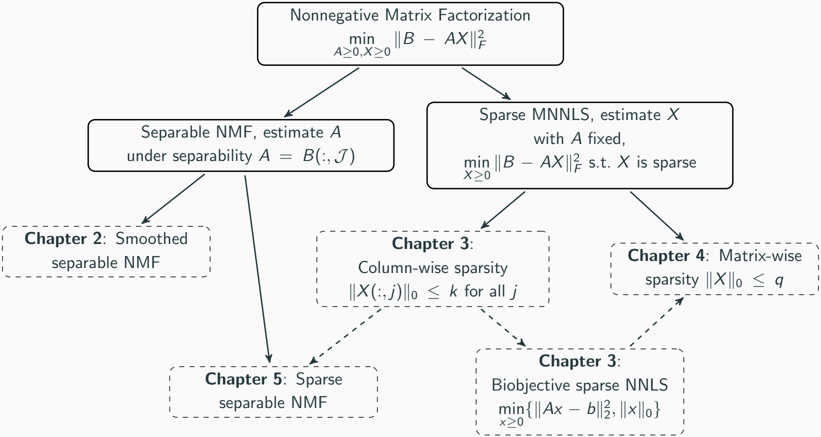
$$\min_{X \geq 0} \|B - AX\|_F^2,$$

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

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

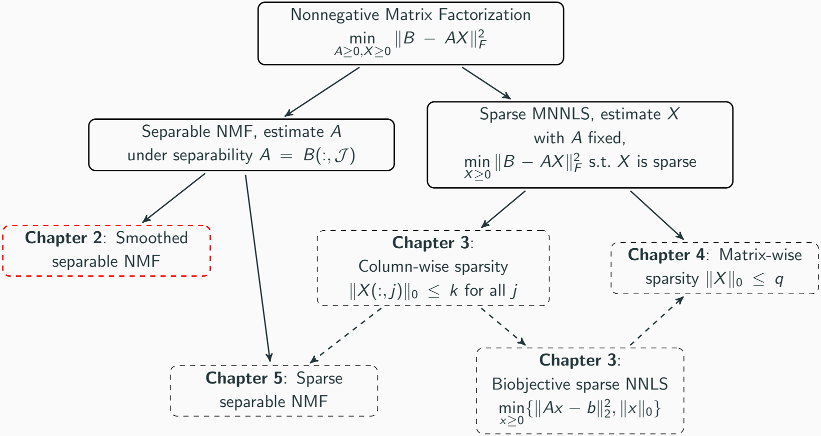
where $X(:,j)$, A, and $B(:,j)$ correspond respectively to x, A, and b.

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.

Separability assumption

There exists an index set $\mathcal{J}$ with $|\mathcal{J}| = r$ such that

$$B = B(:, \mathcal{J})X + N$$

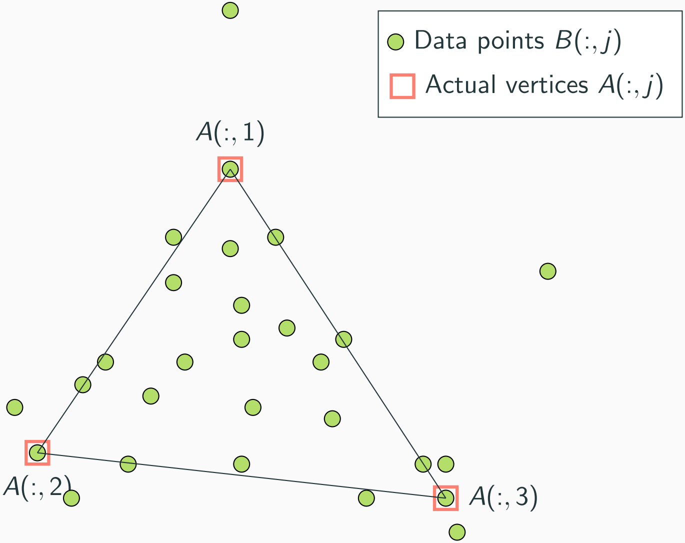
(where N is bounded noise)

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

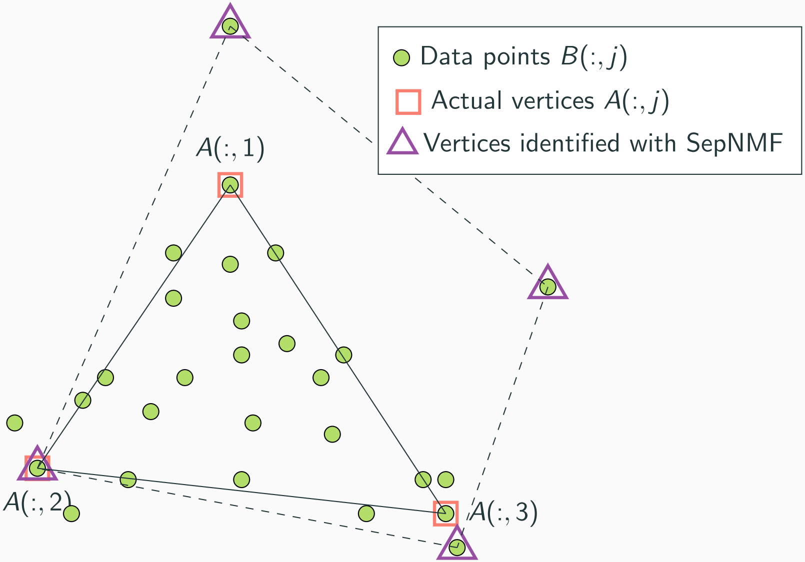
Algorithms: we focus on two greedy algorithms

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

Issues of Separable NMF: outliers, extreme points



Issues of Separable NMF: outliers, extreme points



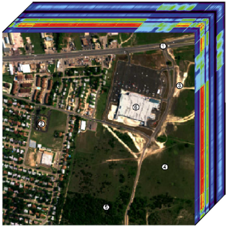
Model 2: Proximal latent points (Bhattacharyya et al. 2020)

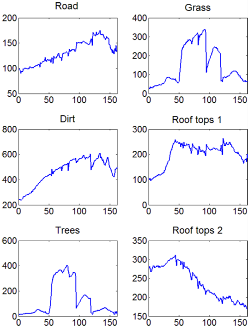
Interpretation: Each vertex has at least p data points close to it.

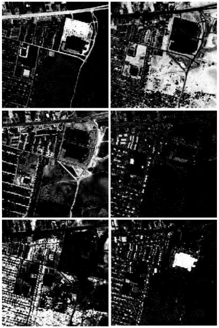
- Assumption is stronger than separability, but it allows more noise, and is realistic in practice.
- The proposed Algorithm to Learn a Latent Simplex (ALLS) has practical issues.

Hyperspectral unmixing

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





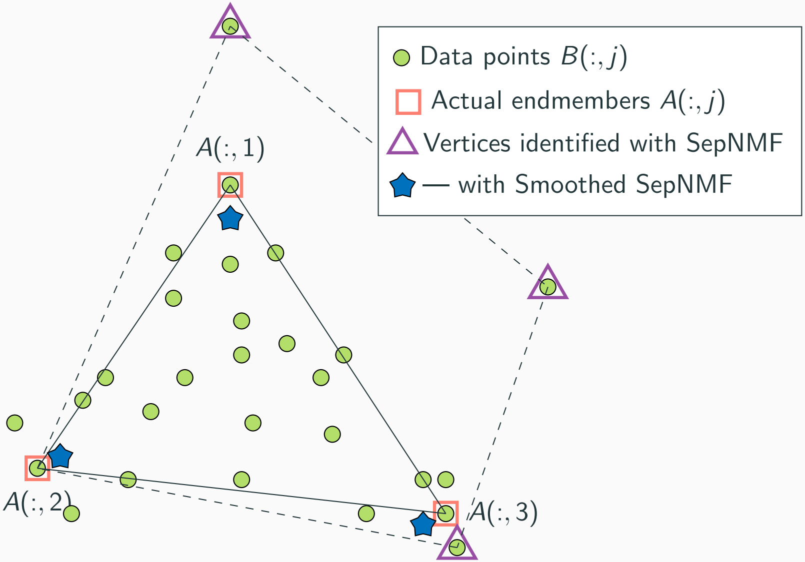


Images from Bioucas Dias and Nicolas Gillis.

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
- Empirically better than VCA, SPA, and ALLS

With smoothed separable NMF



Exact sparse nonnegative least squares

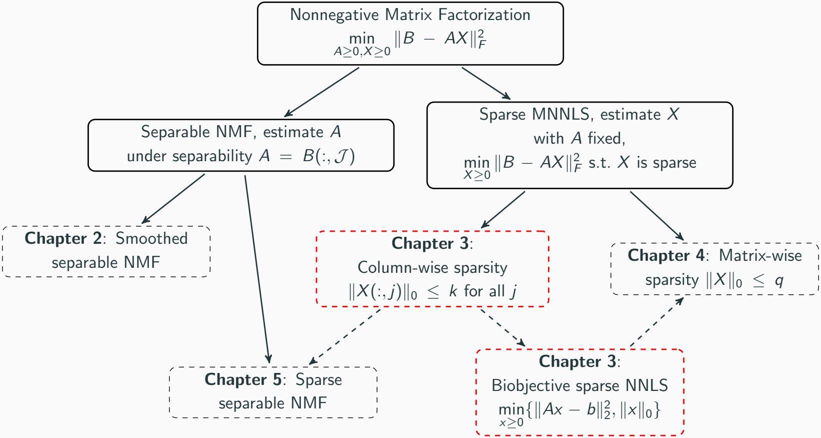
Exact sparse NNLS

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

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

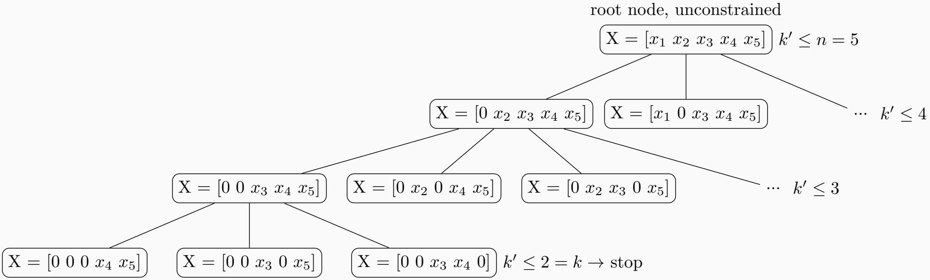
Can we do better than brute-force?

Pruning the search space

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

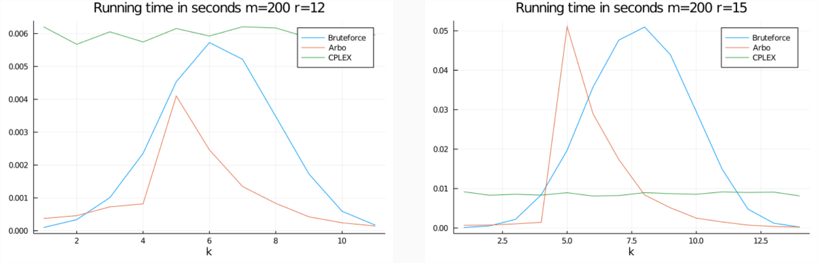
- Branch-and-bound
- Idea: when adding constraints to a problem, the optimal solution can only worsen (or stay the same)
- Our algorithm: arborescent

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



Comparison with generic MIP solvers

(I forgot to include in the thesis: this will be fixed)



Running time on synthetic data sets when k varies

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:

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

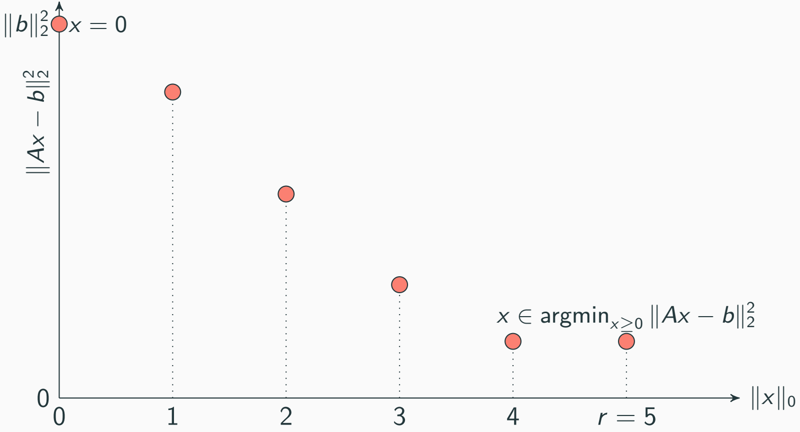
Bi-objective formulation

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

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

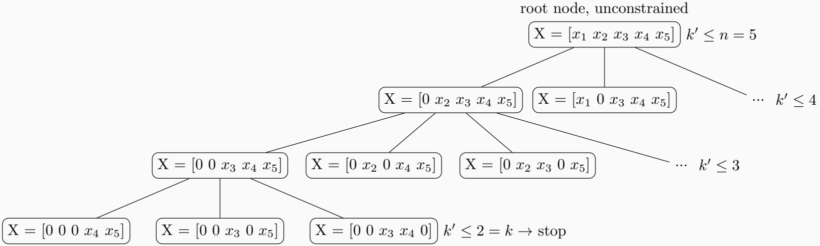
Pareto front

Example for $r = 5$



How to solve the biobjective problem?

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

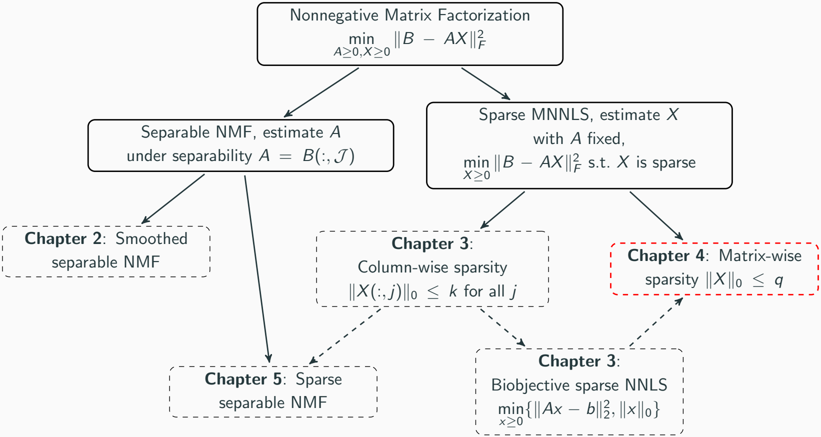


Sum-up

- We proposed arborescent, a branch-and-bound algorithm to solve exactly the k-sparse NNLS problem.
- It works in very general settings (ill-conditioned or noisy data), when traditional approaches fail.
  - At the cost of higher computation time
- 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

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

- 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

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

- 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:

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

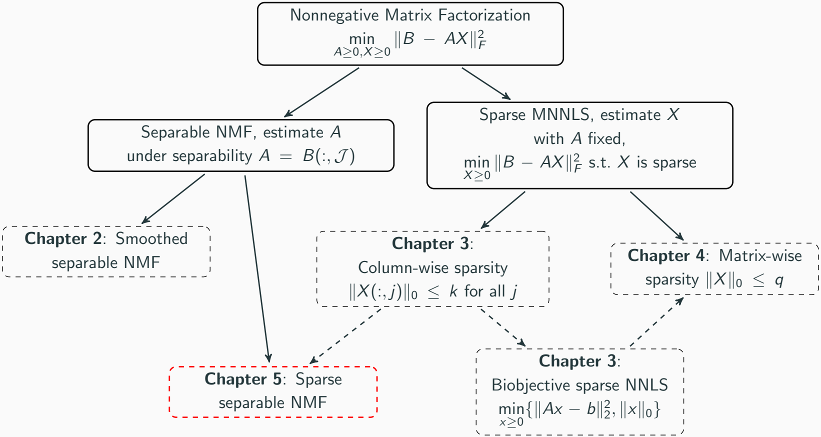
$$\min_{X(:,j)\geq 0} \{ \quad \|B(:,j) - AX(:,j)\|_2^2 \quad , \quad \|X(:,j)\|_o \quad \}$$

   - Solve with arborescent, or heuristic (homotopy, greedy algo)
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

- 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

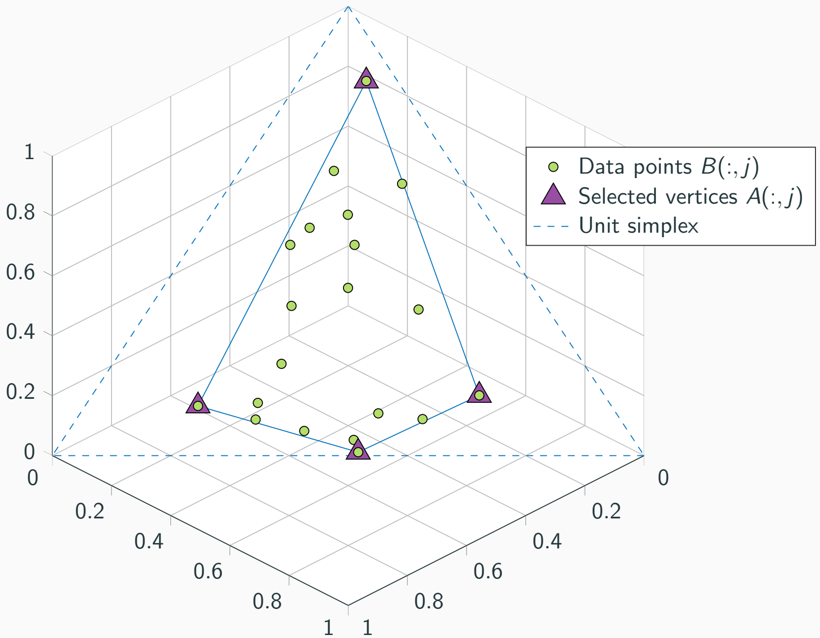
Chapter 5 of the thesis. Presented in the article:

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

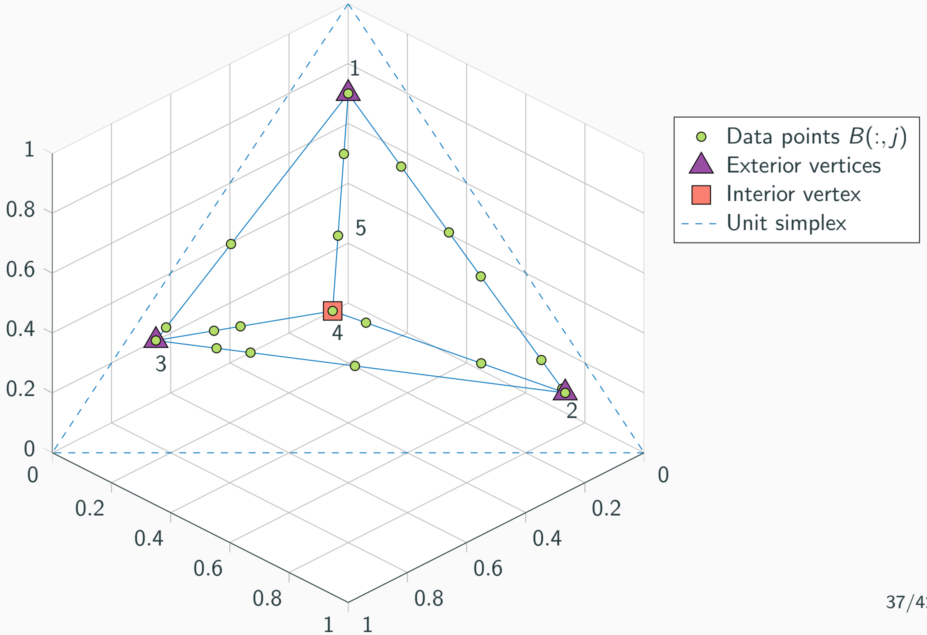
What if one column of A is a combination of others columns of A?

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.

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 guaranteed algorithms: better results at the cost of 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`

Website: `http://nicolasnadisic.xyz`