Sparse Separable Nonnegative Matrix Factorization

Extending Separable NMF with ℓ_0 sparsity constraints

Nicolas Nadisic, Arnaud Vandaele, Jeremy Cohen, Nicolas Gillis

3 December 2020 — iTWIST'2020

Université de Mons, Belgium

Nonnegative Matrix Factorization

Given a data matrix $M \in \mathbb{R}_+^{m \times n}$ and a rank $r \ll \min(m, n)$, find $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{r \times n}$ such that $M \approx WH$.

In optimization terms, standard NMF is equivalent to:

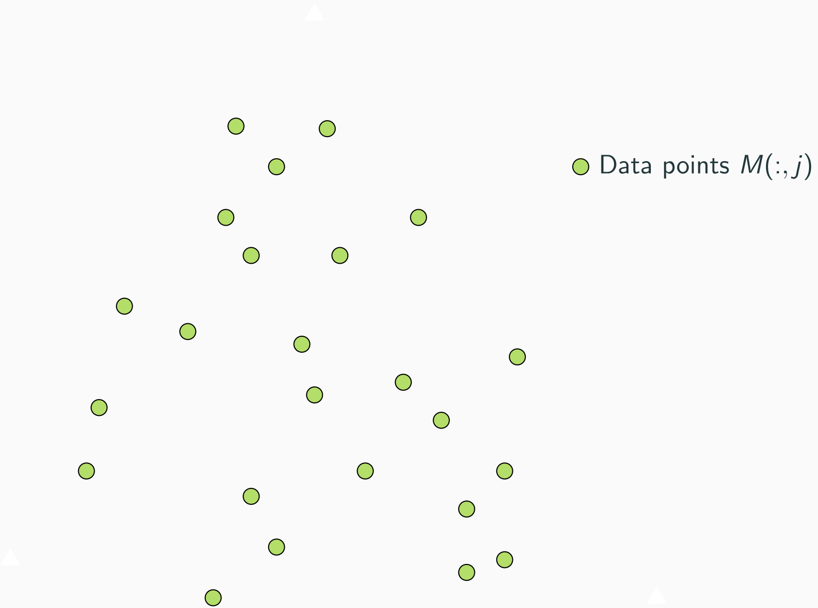
$$\min_{W \geq 0, H \geq 0} \|M - WH\|_F^2$$

Nonnegative Matrix Factorization

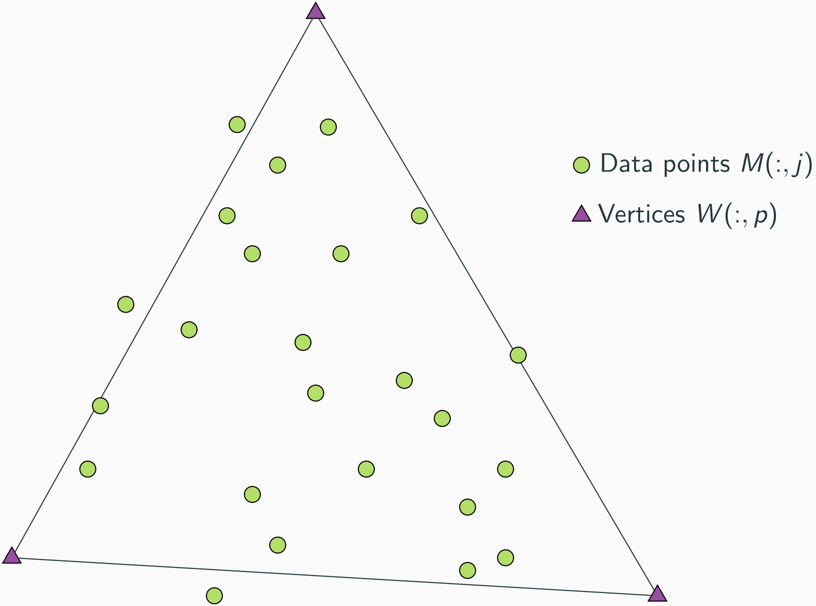
Why nonnegativity?

- More interpretable factors (part-based representation)
- Naturally favors sparsity (solution with few nonzeros)
- Makes sense in many applications (image processing, hyperspectral unmixing, text mining, ...)

NMF Geometry ($M \approx WH$)

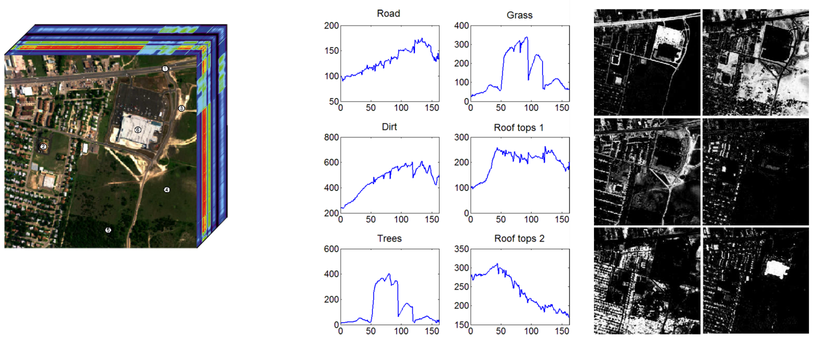


NMF Geometry ($M \approx WH$)



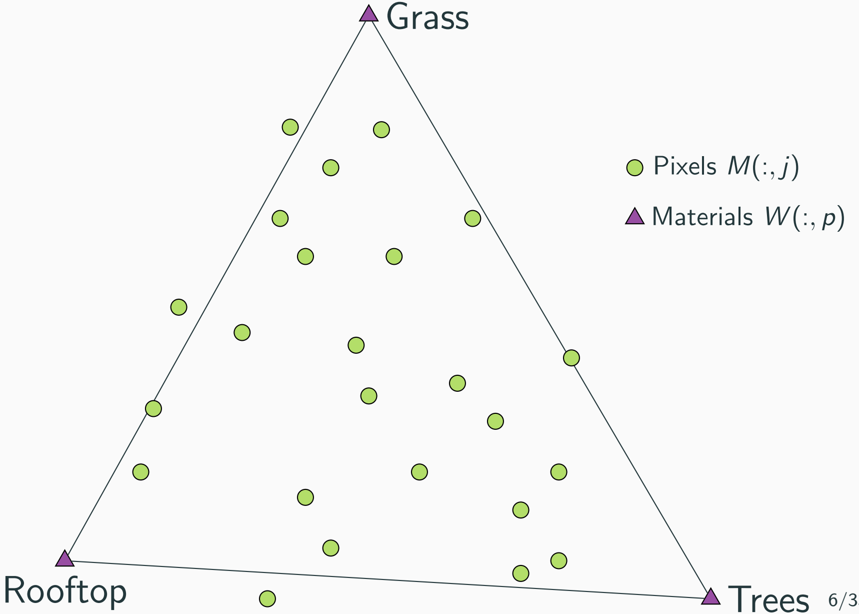
Application – hyperspectral unmixing

$$\underbrace{M(:,j)}_{\text{spectral signature of j-th pixel}} \approx \sum_p \underbrace{W(:,p)}_{\text{spectral signature of p-th material}} \underbrace{H(p,j)}_{\text{abundance of p-th material in j-th pixel}}$$



Images from Bioucas Dias and Nicolas Gillis.

Application – hyperspectral unmixing



Starting point 1/2 – Separable NMF

- NMF is NP-hard [Vavasis, 2010].
- Under the separability assumption, it's solvable in polynomial time [Arora et al., 2012].

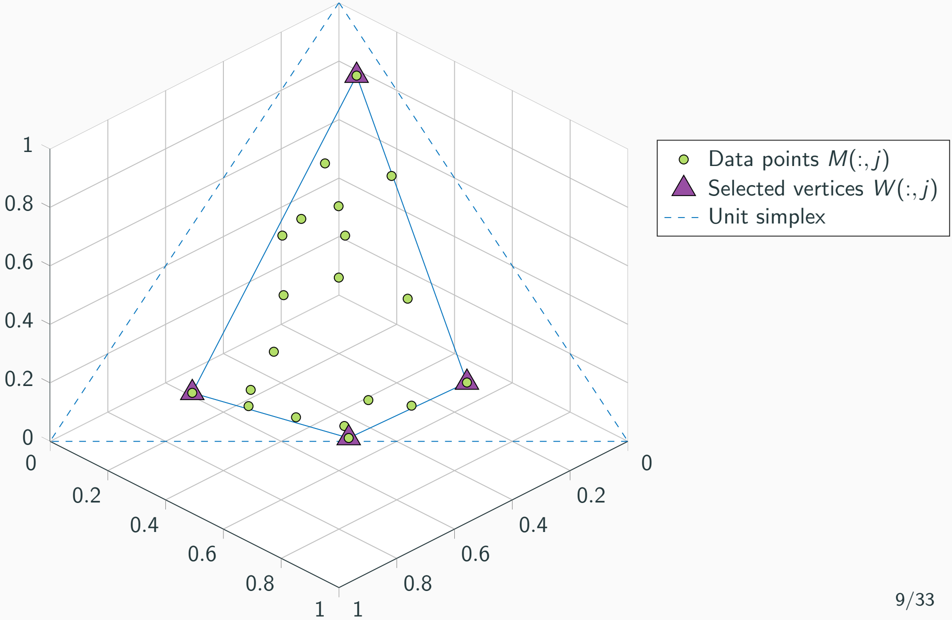
Starting point 1/2 – Separable NMF

Separability:

- The vertices are selected among the data points
- In hyperspectral unmixing, equivalent to Pure-pixel assumption

Standard NMF model	$M = WH$
Separable NMF	$M = M(:, \mathcal{J})H$

Separable NMF – Geometry



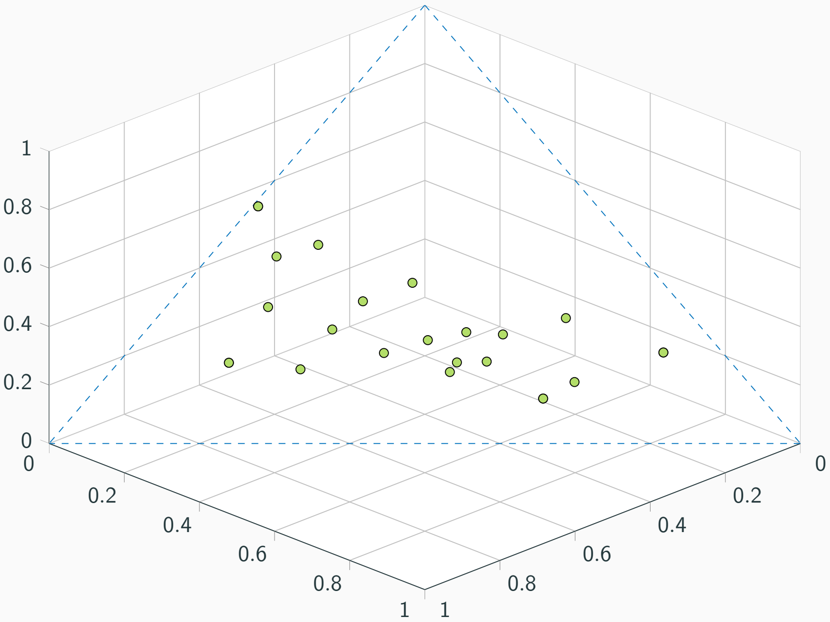
Algorithm for Separable NMF – SNPA

SNPA = Successive Nonnegative Projection Algorithm [Gillis, 2014]

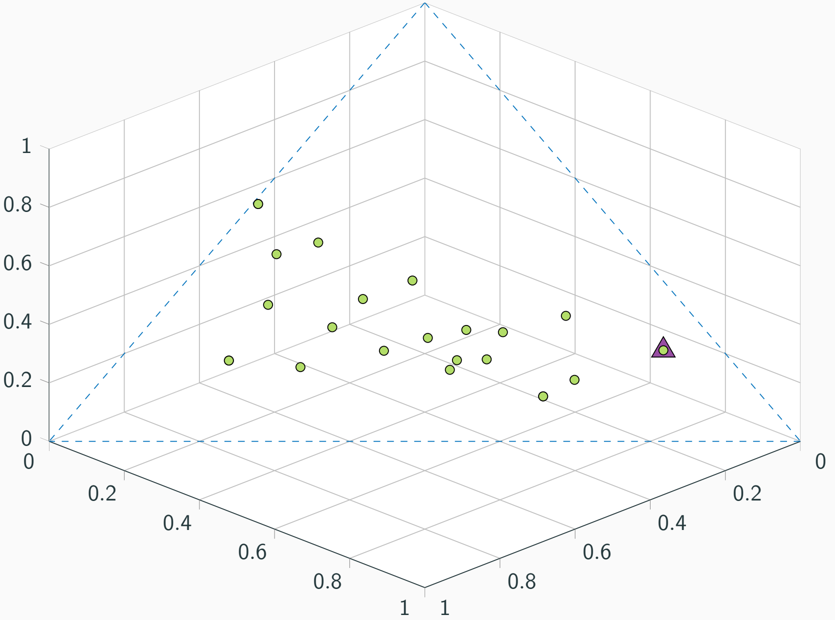
- Start with empty W, and residual $R = M$
- Alternate between
  - Greedy selection of one column of R to be added to W
  - Projection of R on the convex hull of the origin and columns of W
- Stop when reconstruction error $= 0$ (or $< \epsilon$)

(Condition: columns of M have unit ℓ_1-norm)

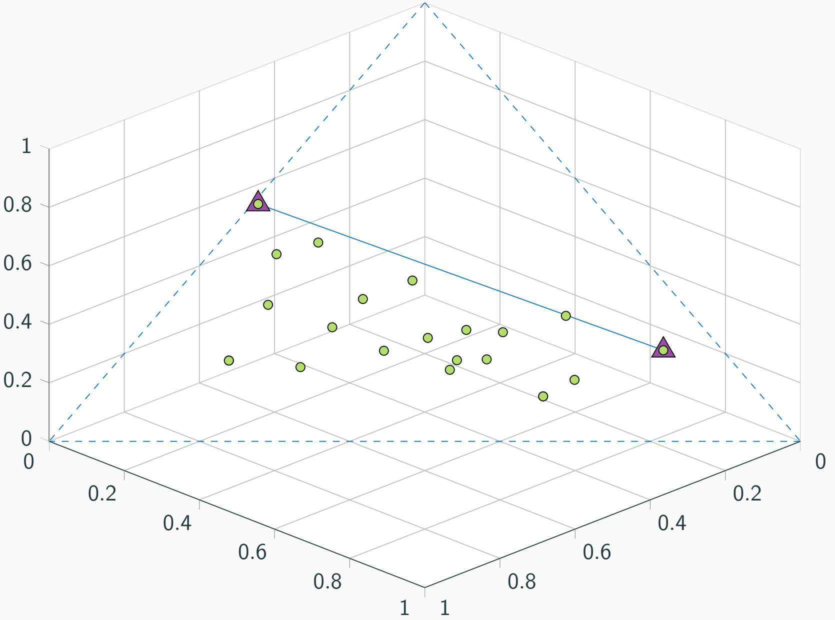
SNPA



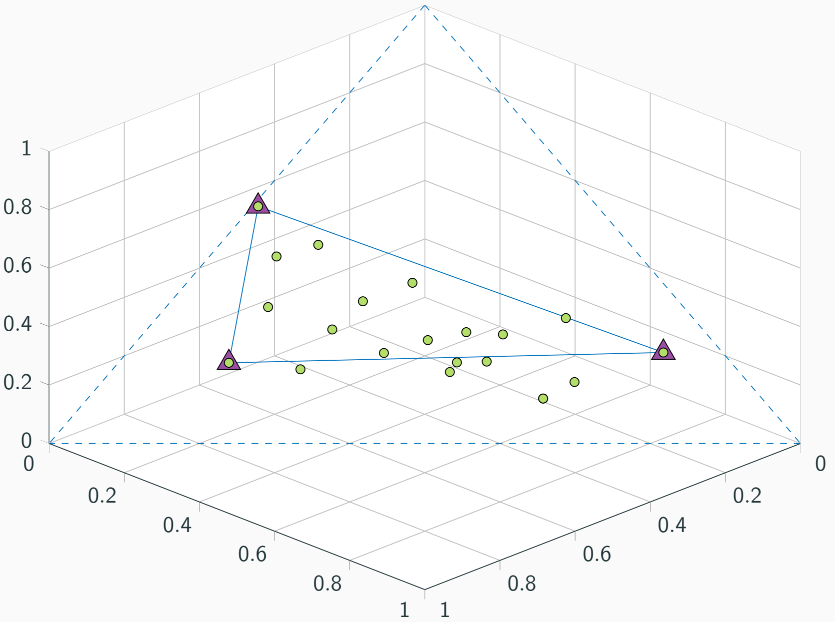
SNPA



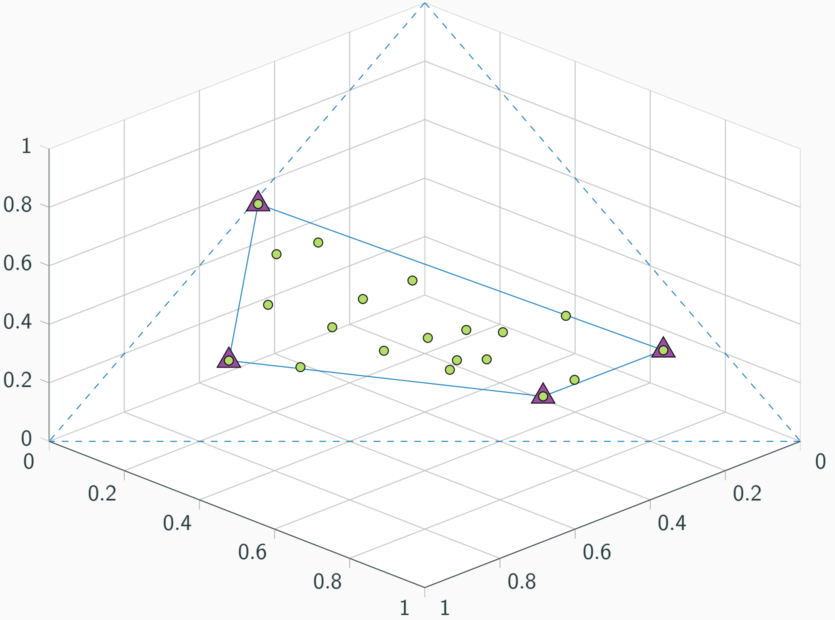
SNPA



SNPA



SNPA



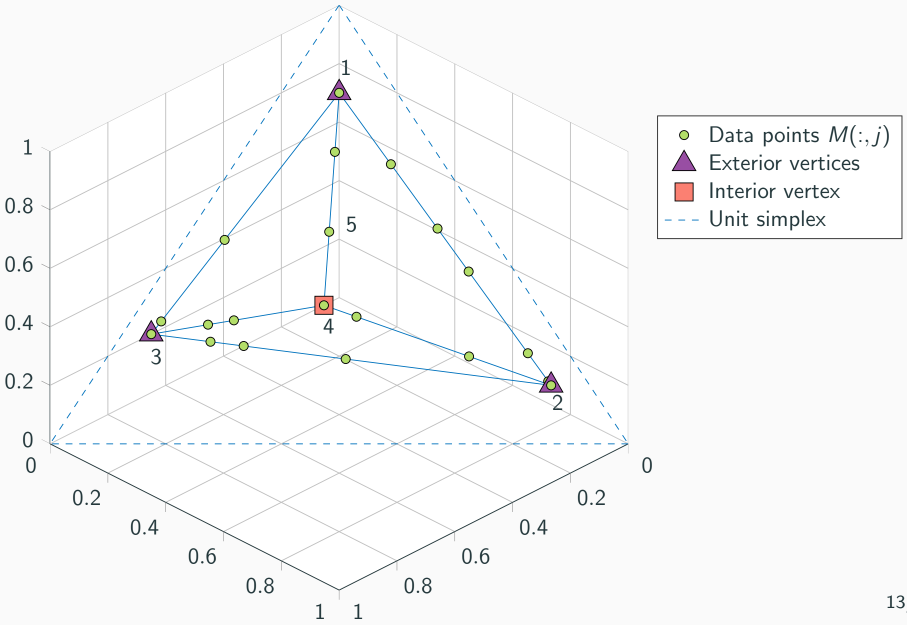
Limitations of Separable NMF

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

$\rightarrow$ Interior vertex

SNPA cannot identify it, because it belongs to the convex hull of the other vertices.

Limitations of Separable NMF



Limitations of Separable NMF

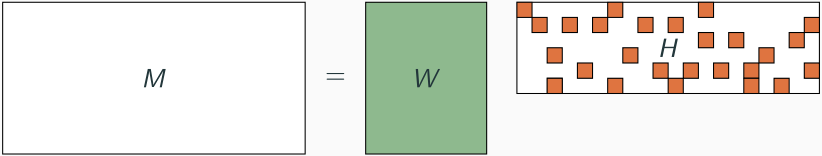
SNPA is unable to handle this case, the interior vertex is not identifiable.

However, if columns of H are sparse (a data point is a combination of only $k < r$ vertices), this interior vertex may be identifiable.

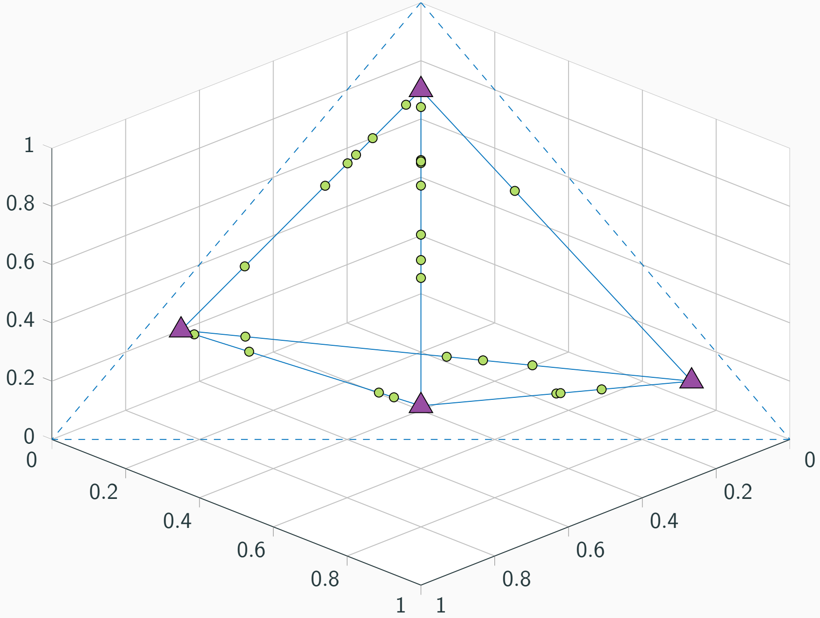
Starting point 2/2 — k-Sparse NMF

$M \approx WH$ s.t. H is column-wise k-sparse (for all i, $\|H(:,i)\|_0 \leq k$)

- Motivation $\rightarrow$ better interpretability
- Motivation $\rightarrow$ improve results using prior sparsity knowledge
- Ex: a pixel expressed as a combination of at most k materials



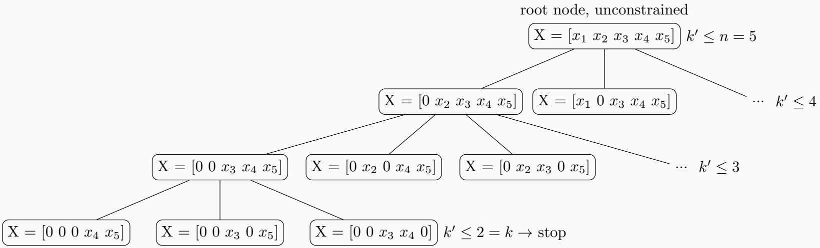
k-Sparse NMF – Geometry



k-Sparse NMF

k-Sparse NMF is combinatorial, with $\binom{r}{k}$ possible combinations per column of H.

Previous work: a branch-and-bound algorithm for Exact k-Sparse NNLS [Nadisic et al., 2020].



Sparse Separable NMF

Standard NMF model	$M = WH$
Separable NMF	$M = M(:, \mathcal{J})H$
SSNMF	$M = M(:, \mathcal{J})H$ s.t. for all i, $\|H(:, i)\|_0 \leq k$

Our approach for SSNMF

Replace the projection step of SNPA, from projection on convex hull to projection on k-sparse hull, done with our BnB solver $\Rightarrow$ kSSNPA.

kSSNPA

- Identifies all interior vertices (non-selected points are never vertices)
- May also identify wrong vertices (explanation to come!)

$\Rightarrow$ kSSNPA can be seen as a screening technique to reduce the number of points to check.

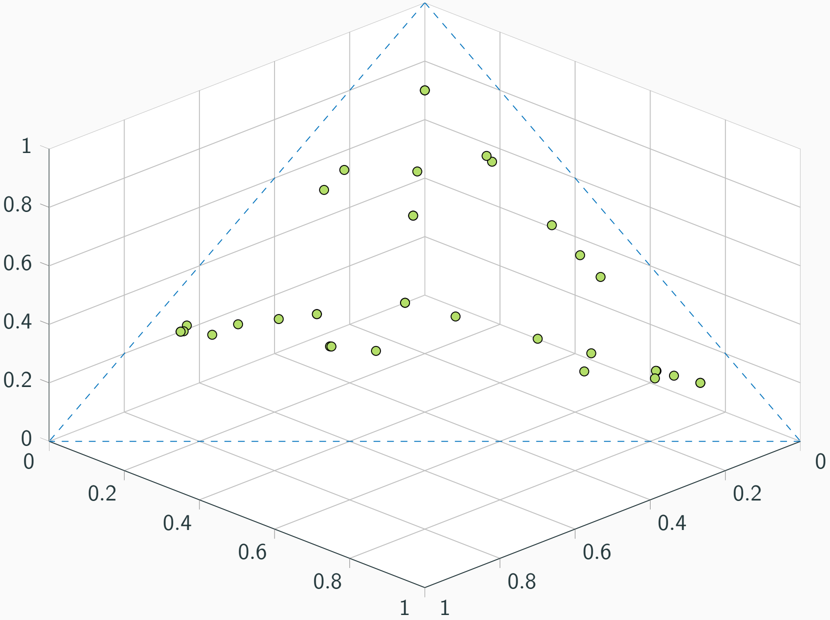
Our approach for SSNMF

In a nutshell, 3 steps:

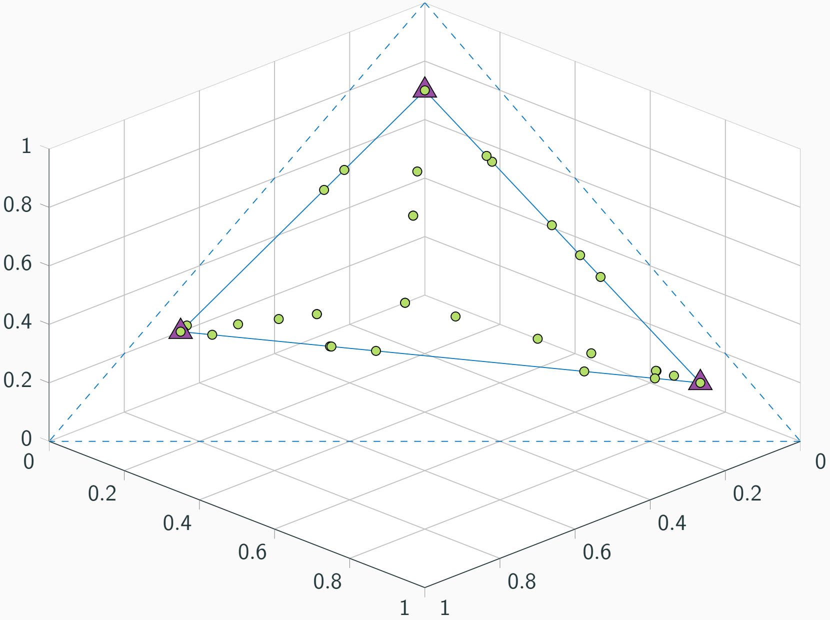
1. Identify exterior vertices with SNPA
2. Identify candidate interior vertices with kSSNPA
3. Discard bad candidates, those that are k-sparse combinations of other selected points (they cannot be vertices)

Our algorithm: BRASSENS Relies on Assumptions of Sparsity and Separability for Elegant NMF Solving.

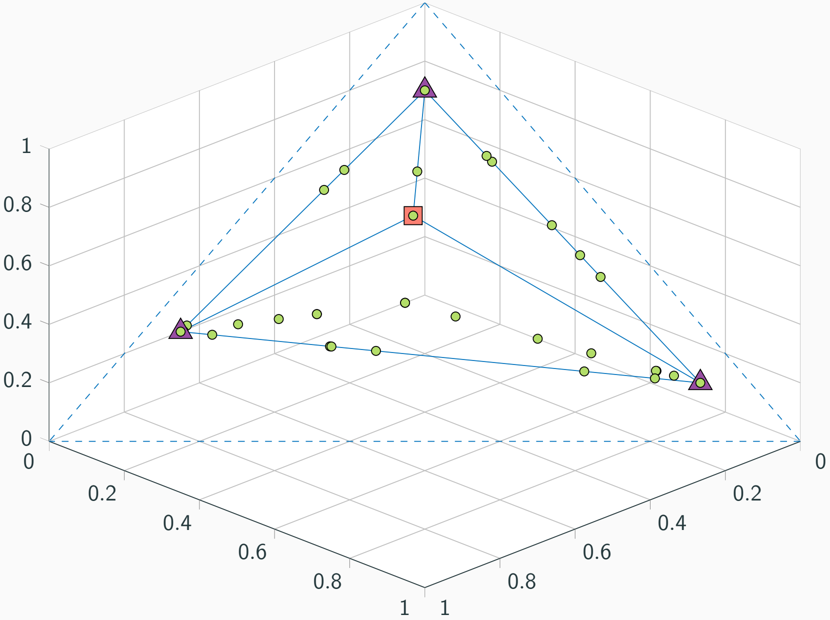
BRASSENS with sparsity $k = 2$



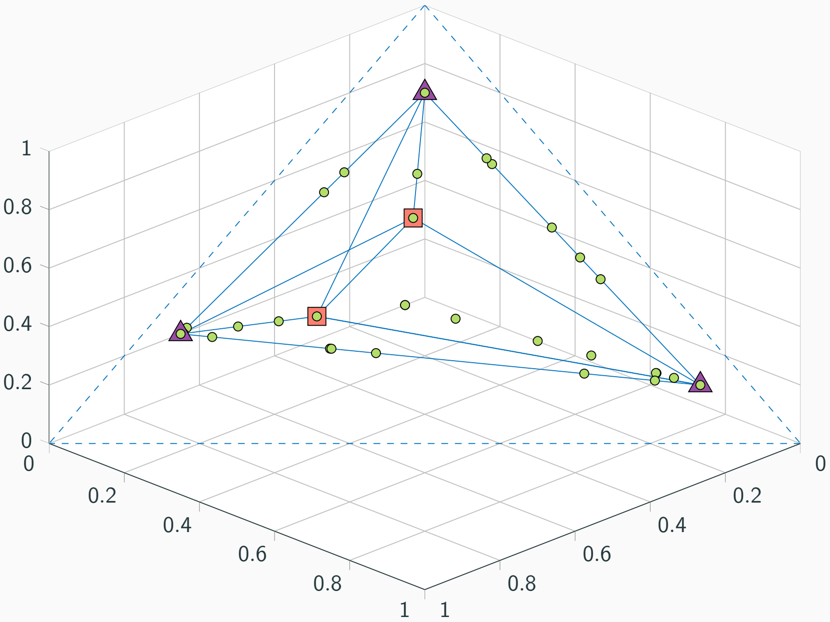
BRASSENS with sparsity $k = 2$



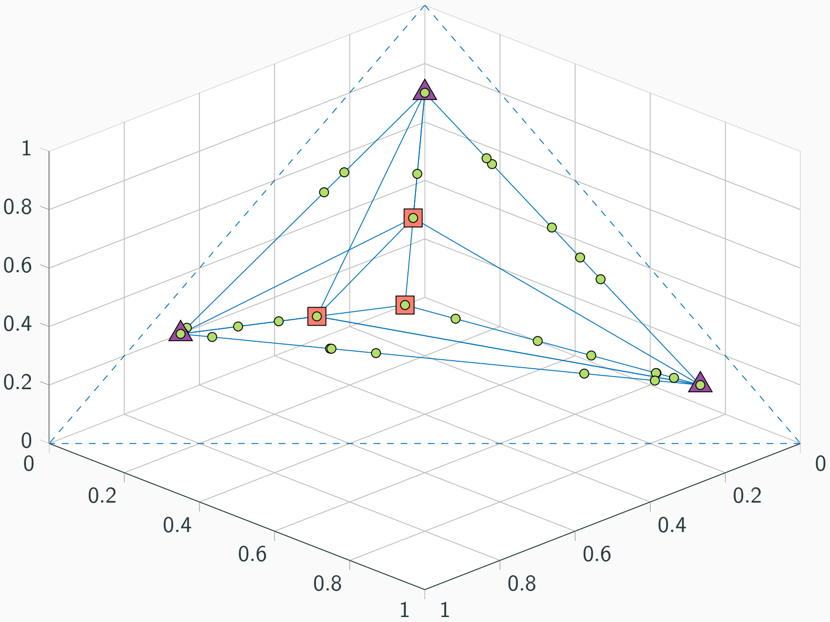
BRASSENS with sparsity $k = 2$



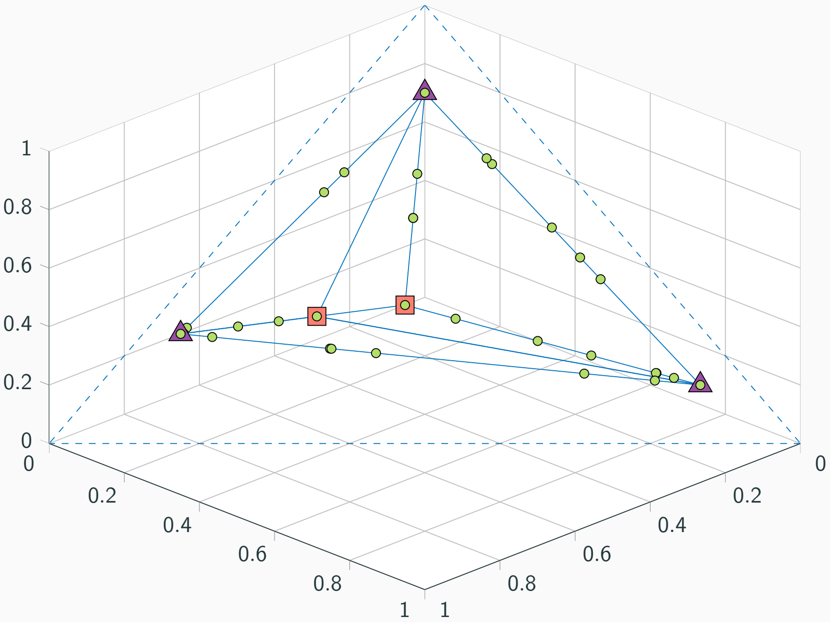
BRASSENS with sparsity $k = 2$



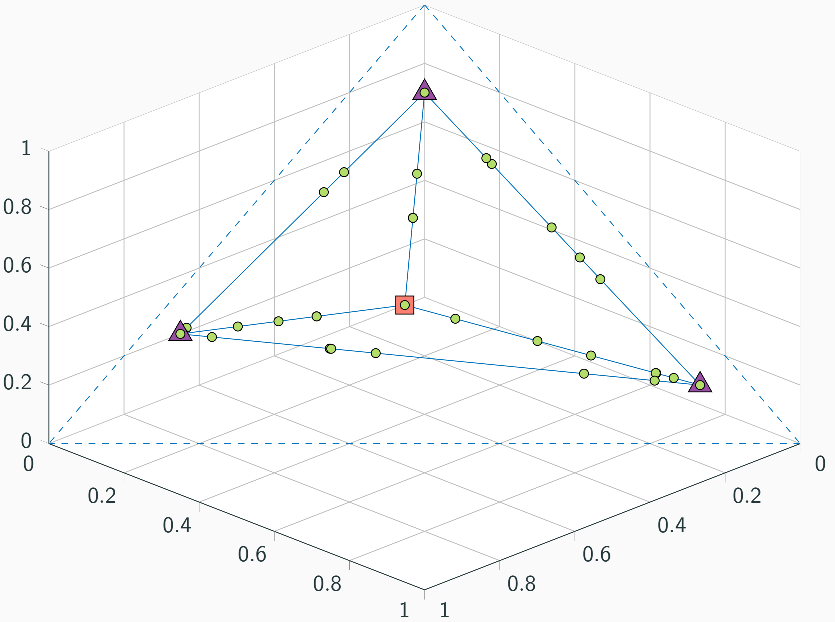
BRASSENS with sparsity $k = 2$



BRASSENS with sparsity $k = 2$



BRASSENS with sparsity $k = 2$



Complexity

- As opposed to Sep NMF, SSNMF is NP-hard (Arnaud proved it, see the paper)
- Hardness comes from the k-sparse projection
- Not too bad when r is small, with our BnB solver

Correctness

Assumption 1 No column of W is a nonnegative linear combination of k other columns of W.
⇒ necessary condition for recovery by BRASSENS

Assumption 2 No column of W is a nonnegative linear combination of k other columns of M.
⇒ sufficient condition for recovery by BRASSENS

If data points are *k*-sparse and generated at random, **Assumption 2** is true with probability one.

Related work

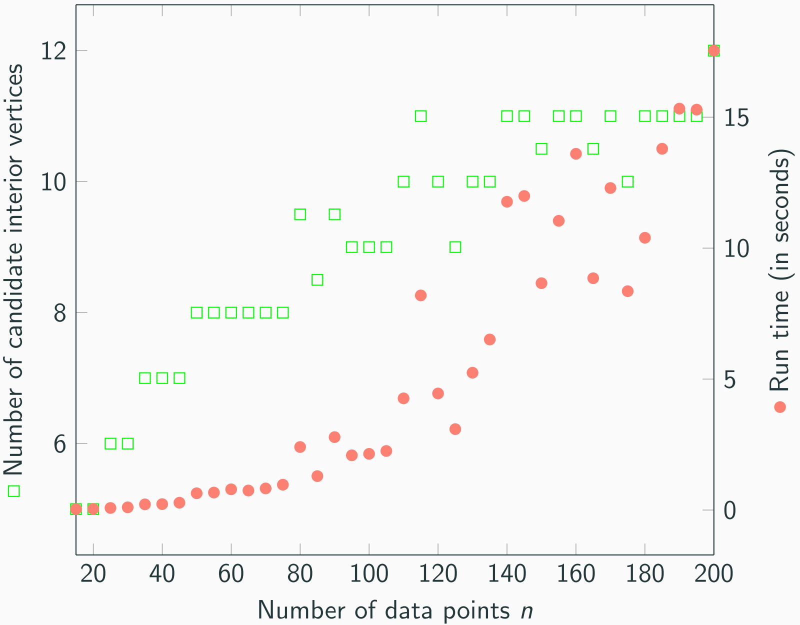
Only one similar work: [Sun and Xin, 2011]

- Handles only one interior vertex
- Non-optimal bruteforce-like method

Experiments

- Experiments on synthetic datasets with interior vertices
- Experiment on underdetermined multispectral unmixing (Urban image, 309×309 pixels, limited to $m = 3$ spectral bands, and we search for $r = 5$ materials)
- No other algorithm can tackle SSNMF, so comparisons are limited

XP Synthetic: 3 exterior and 2 interior vertices, n grows



XP Synthetic 2: dimensions grow

m	n	r	k	Number of candidates	Run time in seconds
3	25	5	2	5.5	0.26
4	30	6	3	8.5	3.30
5	35	7	4	9.5	38.71
6	40	8	5	13	395.88

Conclusion from experiments:

- kSSNPA is efficient to select few candidates
- Still, BRASSENS does not scale well :(

XP on 3-bands Urban dataset with $r = 5$

SNPA

Grass+Trees +Rooftops	Rooftops 1	Dirt+Road +Rooftops	Dirt+Grass	Rooftops 1 +Dirt+Road

BRASSENS (finds 1 interior point)

Grass+Trees	Rooftops 1	Road	Rooftops+Road	Dirt+Grass

Future work

- Theoretical analysis of robustness to noise
- New real-life applications

Take-home messages

Sparse Separable NMF:

- Combine constraints of separability and k-sparsity
- A new way to regularize NMF
- Can handle some cases that Separable NMF cannot
  - Underdetermined case
  - Interior vertices
- Is NP-hard (unlike Sep NMF), but actually “not so hard” for small r
- Is provably solved by our approach
- Does not scale well

References i

Arora, S., Ge, R., Kannan, R., and Moitra, A. (2012).
Computing a nonnegative matrix factorization – provably.
STOC '12.

Gillis, N. (2014).
Successive Nonnegative Projection Algorithm for Robust Nonnegative Blind Source Separation.
SIAM Journal on Imaging Sciences, 7(2):1420–1450.

Nadisic, N., Vandaele, A., Gillis, N., and Cohen, J. E. (2020).
Exact Sparse Nonnegative Least Squares.
In *ICASSP 2020*, pages 5395 – 5399.

Sun, Y. and Xin, J. (2011).
Underdetermined Sparse Blind Source Separation of Nonnegative and Partially Overlapped Data.
SIAM Journal on Scientific Computing, 33(4):2063–2094.

Vavasis, S. A. (2010).
On the Complexity of Nonnegative Matrix Factorization.
SIAM Journal on Optimization.

Contact: nicolas.nadisic@umons.ac.be

Code and exp.: https://gitlab.com/nnadisic/ssnmf

Slides and paper: http://nicolasnadisic.xyz