Matrix-wise ℓ_0-constrained Sparse Nonnegative Least Squares

Nicolas Nadisic[1,(2)], Arnaud Vandaele[2], Jeremy E. Cohen[3], Nicolas Gillis[2]

[1]Ghent University, Belgium [2]University of Mons, Belgium [3]CNRS, Université de Lyon, France

Starting point: nonnegative low-rank linear model

$$B \approx AX, \qquad \text{where}$$

- $B \in \mathbb{R}_+^{m\times n}$, $A \in \mathbb{R}_+^{m\times r}$, $X \in \mathbb{R}_+^{r\times n}$
- $r \ll \min(m,n)$
- each data point $B(:,j)$ is an additive linear combination of r features $A(:,p)$

Given A and B, finding X is a nonnegative least squares problem with multiple right-hand sides (MNNLS),

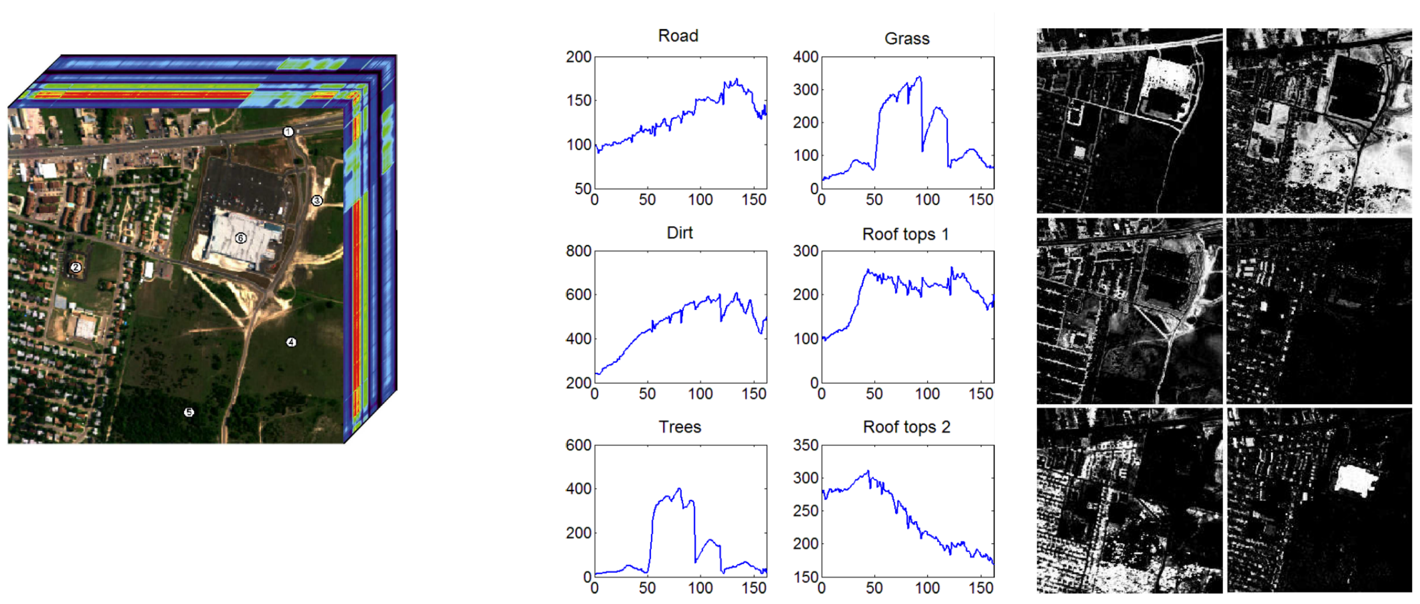
$$\min_{X\geq 0} \|B - AX\|_F^2$$

Can be divided in n independent NNLS subproblems,

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

One application: hyperspectral unmixing

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



Sparse nonnegative least squares

Sparsity of X means each data point is a combination of a few features
$\Rightarrow$ Improve interpretability, regularize the problem
How to enforce it?
ℓ_1-norm penalty

$$\min_{X\geq 0} \|B - AX\|_F^2 + \lambda \|X\|_1$$

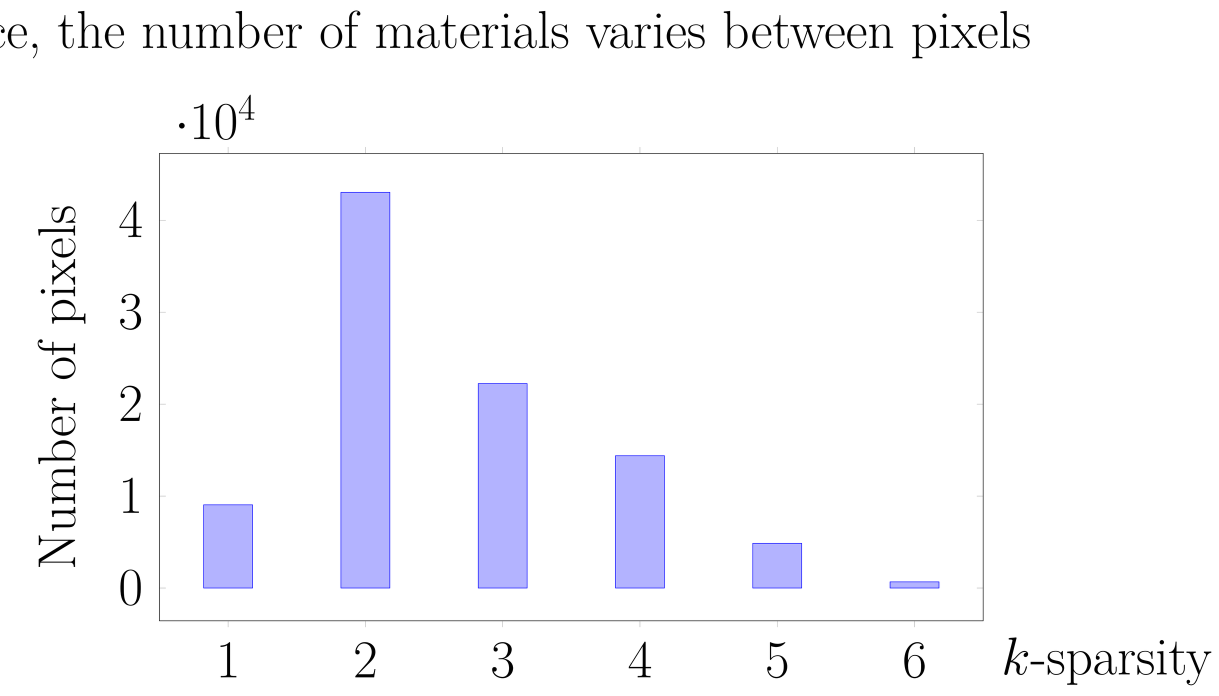
Column-wise ℓ_0 constraint, $\|x\|_0 = |\{i : x(i) \neq 0\}|$

$$\min_{X\geq 0} \|B - AX\|_2^2 \text{ s.t. } \|X(:,j)\|_0 \leq k \text{ for all } j$$

solved locally with greedy algo, globally with branch-and-bound

Issue of the column-wise sparsity constraint

- What if the relevant k varies between columns?
- For instance, the number of materials varies between pixels



Sparsity of the pixels of the ground truth X of the HSI Urban, $n = 94249$, $r = 6$

Our contribution: matrix-wise constrained NNLS

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

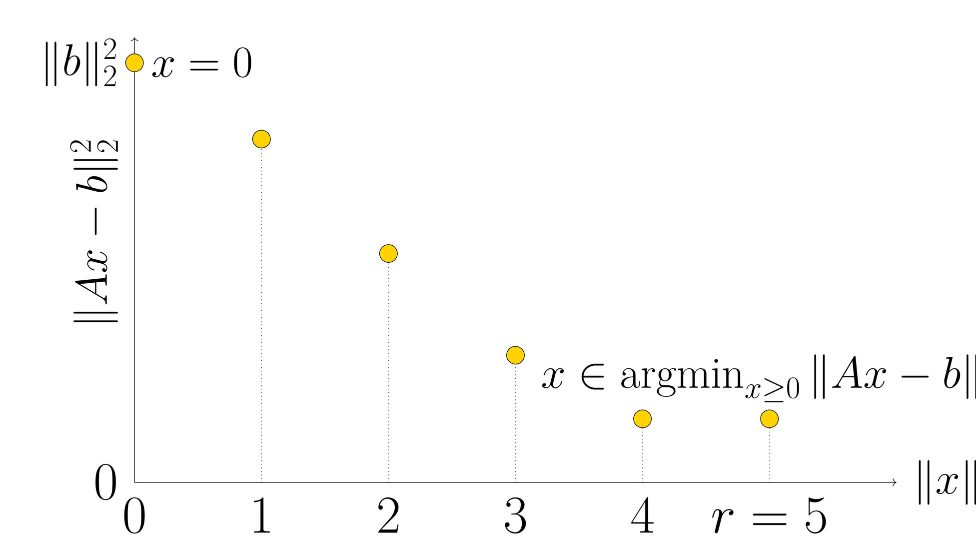
- $\|X\|_0$ is the number of non-zero entries in X
- if $q = k \times n$, equivalent to an average column-wise k-sparsity constraint

Problem too big to be tackled directly $\Rightarrow$ divide and conquer

Column-wise bi-objective problem, Pareto fronts

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

- Equivalent to
  $\min_{x\geq 0} \|b - Ax\|_2^2$ s.t. $\|x\|_0 \leq k$
  for all $k \in \{0,\dots,r\}$
- Generate a Pareto front for each subproblem

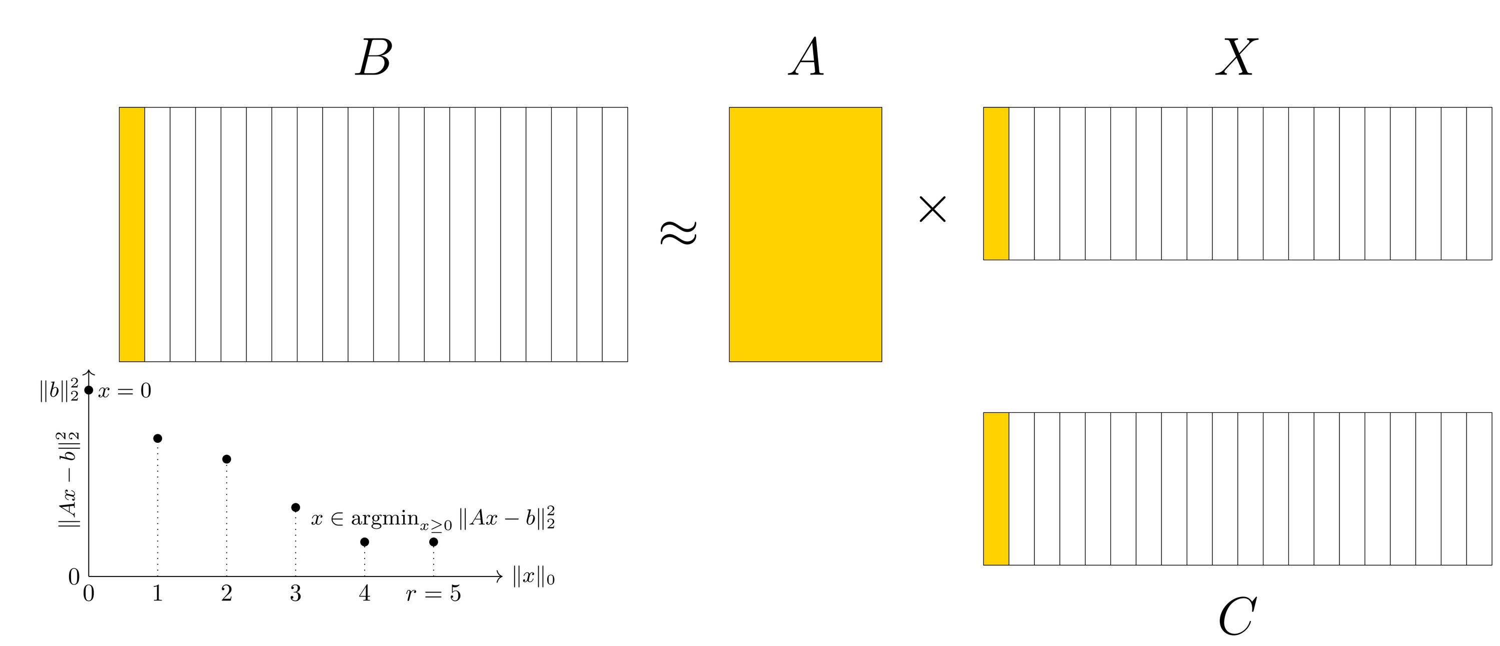


Our algorithm: SALMON

SALMON Applies ℓ_0-constraints Matrix-wise On NNLS problems

Step 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
  $$\min_{X(:,j)\geq 0} \{ \quad \|B(:,j) - AX(:,j)\|_2^2 \quad , \quad \|X(:,j)\|_0 \quad \}$$
- Solve with branch-and-bound, or heuristic (homotopy, greedy algo)
- Build a cost matrix C



Step 2: Select one solution per column such that in total X has q nonzero entries and the error is minimized
$\Rightarrow$ assignment-like problem

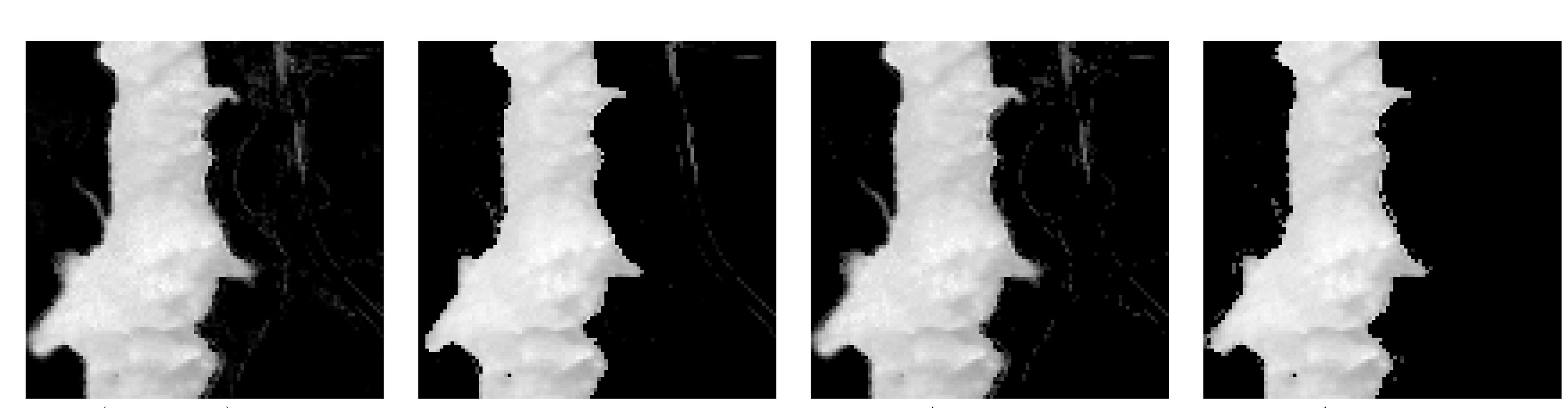
$$\begin{pmatrix} C_{0,1} & C_{0,2} & \cdots & C_{0,n} \\ C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r,1} & C_{r,2} & \cdots & C_{r,n} \end{pmatrix}$$

- Dedicated greedy algorithm proved near-optimal

$$C(i,j) \approx \min_{x\geq 0} \|B(:,j) - Ax\|_2^2 \text{ s.t. } \|x\|_0 \leq i$$

Results

Using homotopy in step 1. Unmixing of the HSI Jasper Ridge.
We show only one material (row of X reshaped), it corresponds to water.



NNLS (no sparse)
Error = 5.71%
Actual sparsity = 2.27

Col-wise, $k = 2$
Error = 6.99%
Actual sparsity = 1.78

Salmon, $q/n = 2$
Error = 5.72%

Salmon, $q/n = 1.8$
Error = 5.95%

		AS	ℓ_1-CD	Hcw	**H+S**	OGcw	OGg	**OG+S**	BaBcw	**BaB+S**
Samson	Time	0.22	0.24	0.2	0.26	0.31	3.67	0.57	0.52	0.8
$r = 3$	Error	3.3	**3.3**	3.34	**3.3***	6.76	3.32	**3.3***	3.4	**3.3***
$k = 2$	Sparsity	2.2	2	1.85	2	1.6	1.99	1.99	1.83	2
Urban	Time	5.08	4.31	4.86	7.79	3.38	958	16.4	33.5	73.1
$r = 6$	Error	7.67	8.13	8.62	7.83*	8.97	8.07	**7.76***	8.27	**7.71***
$k = 2$	Sparsity	2.63	2	1.9	2	1.7	2	2	1.83	2
Cuprite	Time	5.19	3.32	7.86	10.1	5.06	620	31.5	784	4829
$r = 12$	Error	1.74	3.17	2.37	2.01	2.32	1.97	**1.89***	1.93	**1.83***
$k = 4$	Sparsity	6.61	4	3.92	4	3.53	4	4	3.81	4

Code and contact

- http://nicolasnadisic.xyz
- nicolas.nadisic@ugent.be

UMONS University of Mons
GHENT UNIVERSITY