Exact and Heuristic Methods for Simultaneous Sparse Coding

Application to dictionary-based NMF

Alexandra Dache[1], Nicolas Nadisic[2], Arnaud Vandaele[1], Nicolas Gillis[1]

EUSIPCO 2023

[1]University of Mons, Belgium
[2]Ghent University, Belgium

Our motivations

- Approximate data points as linear combinations of a few features selected from an overcomplete dictionary.
- In the standard (N)MF setting with

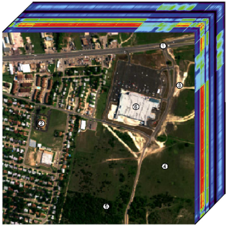
$$X \approx W\hat{H},$$

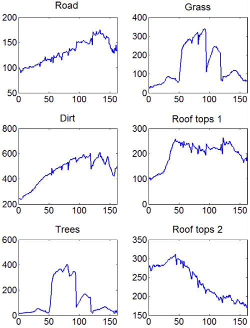
it means the columns of W are a subset of some fixed dictionary,

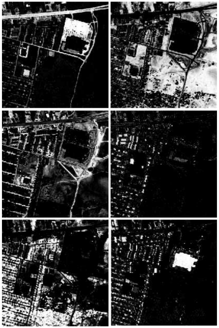
$$W = D(:, \mathcal{J}).$$

Example: hyperspectral unmixing

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







- Select endmembers spectral signatures in a overcomplete dictionary.
- Dictionary can be a library of spectra (eg USGS).
- Using the input X as a self-dictionary reduces to pure-pixel search or separable NMF.

Simultaneous Sparse Coding (SSC)

Given

- an input matrix $X \in \mathbb{R}^{m \times n}$,
- an overcomplete dictionary $D \in \mathbb{R}^{m \times s}$,
- and a sparsity target $r \in \mathbb{N}$,

find $H \in \mathbb{R}^{s \times n}$,

Simultaneous Sparse Coding (SSC)

$$\min_{H} \|X - DH\|_F^2 \quad \text{s.t.} \quad \|H\|_{row-0} \leq r,$$

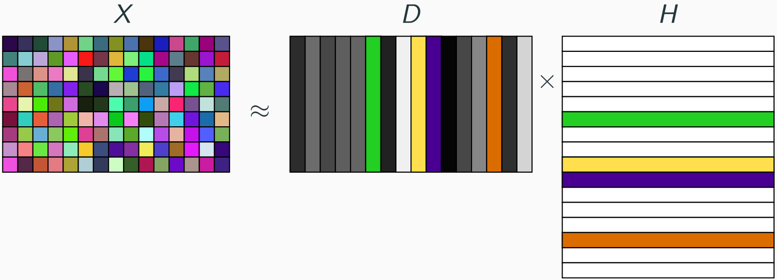
where $\|H\|_{row-0} = |\{i | H(i,:) \neq 0\}|$ is the number of non-zero rows of H.

Simultaneous Sparse Coding (SSC)

$$\text{(SSC)} \qquad \min_{H} \|X - DH\|_F^2 \quad \text{s.t.} \quad \|H\|_{row-0} \leq r.$$

SSC $\Leftrightarrow$ Find H with at most r non-zero rows.

$\Leftrightarrow$ Select the best r columns of D to reconstruct X.



Existing works and limitations

SSC is a combinatorial problem, hard to solve up to global optimality.

Existing works and limitations

SSC is a combinatorial problem, hard to solve up to global optimality.

Existing methods rely on fast but approximate heuristics:

- Convex relaxation, eg group LASSO with $\ell_{2,1}$ penalty, $\sum_i \|H(i,:)\|_2$,
- Greedy algorithms, mostly variants of orthogonal matching pursuit (OMP).

Existing works and limitations

SSC is a combinatorial problem, hard to solve up to global optimality.

Existing methods rely on fast but approximate heuristics:

- Convex relaxation, eg group LASSO with $\ell_{2,1}$ penalty, $\sum_i \|H(i,:)\|_2$,
- Greedy algorithms, mostly variants of orthogonal matching pursuit (OMP).

Issue: condition for exact recovery are restrictive and virtually never met in practice.

Our proposal: global optimization

Motivations:

- In somes critical cases, we need guarantees on the solution.
- The acquisition of hyperspectral images is long and expensive, why not spend more time/energy on their analysis?

Our proposal: global optimization

Motivations:

- In somes critical cases, we need guarantees on the solution.
- The acquisition of hyperspectral images is long and expensive, why not spend more time/energy on their analysis?

The sparsity constraint is actually a classical cardinality constraint, standard in combinatorial optimization. Can we tackle SSC with a generic solver?

Our proposal: global optimization

Motivations:

- In somes critical cases, we need guarantees on the solution.
- The acquisition of hyperspectral images is long and expensive, why not spend more time/energy on their analysis?

The sparsity constraint is actually a classical cardinality constraint, standard in combinatorial optimization. Can we tackle SSC with a generic solver?

We first need to reformulate SSC in a standard problem form.

Reformulation as a Mixed-Integer Program (MIP)

Let the binary vector $y \in \{0,1\}^s$ represents the row-sparsity of H,

$$y_i = 0 \Leftrightarrow H(i,:) = 0 \text{ for all } i. \quad (1)$$

Reformulation as a Mixed-Integer Program (MIP)

Let the binary vector $y \in \{0,1\}^s$ represents the row-sparsity of H,

$$y_i = 0 \Leftrightarrow H(i,:) = 0 \text{ for all } i. \quad (1)$$

Then the row-0 "norm" can be written as a sum,

$$\|H\|_{row-0} = \sum_i y_i,$$

Reformulation as a Mixed-Integer Program (MIP)

Let the binary vector $y \in \{0,1\}^s$ represents the row-sparsity of H,

$$y_i = 0 \Leftrightarrow H(i,:) = 0 \text{ for all } i. \tag{1}$$

Then the row-0 “norm” can be written as a sum,

$$\|H\|_{row-0} = \sum_i y_i,$$

and (1) can be rewritten as a linear constraint using a big-M constant,

$$-My_i \leq H(i,j) \leq My_i \text{ for all } i, j.$$

Reformulation as a Mixed-Integer Program (MIP)

The SSC problem can be reformulated with continuous and binary variables, a quadratic objective, and linear constraints ⇒ MIQP:

$$\min_{H} \sum_{j} H(:,j)^T D^T D H(:,j) - 2X(:,j)^T D H(:,j)$$

$$\text{s.t.} \begin{cases} -My_i \leq H(i,j) \leq My_i \text{ for all } i,j, \\ \sum_i y_i \leq r. \end{cases}$$

Reformulation as a Mixed-Integer Program (MIP)

The SSC problem can be reformulated with continuous and binary variables, a quadratic objective, and linear constraints $\Rightarrow$ MIQP:

$$\min_{H} \sum_{j} H(:,j)^T D^T D H(:,j) - 2X(:,j)^T D H(:,j)$$

$$\text{s.t.} \begin{cases} -My_i \leq H(i,j) \leq My_i \text{ for all } i,j, \\ \sum_i y_i \leq r. \end{cases}$$

Adding nonnegativity is easy!

Reformulation as a Mixed-Integer Program (MIP)

The SSC problem can be reformulated with continuous and binary variables, a quadratic objective, and linear constraints ⇒ MIQP:

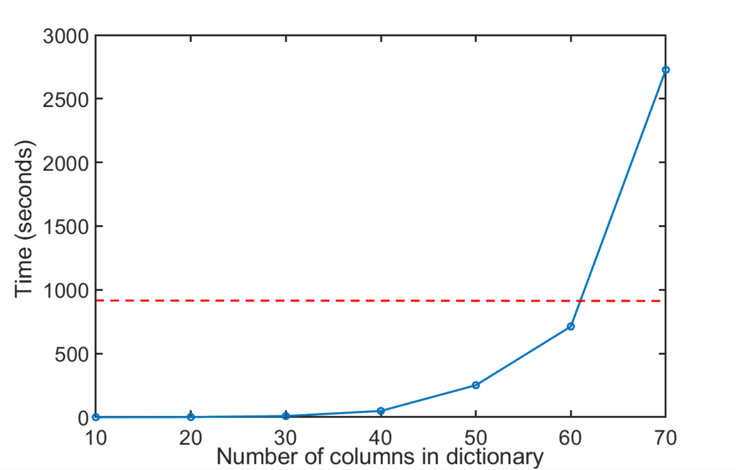
$$\min_{H} \sum_{j} H(:,j)^T D^T D H(:,j) - 2X(:,j)^T D H(:,j)$$

$$\text{s.t. } \begin{cases} 0 \leq H(i,j) \leq M y_i \text{ for all } i,j, \\ \sum_i y_i \leq r. \end{cases}$$

Adding nonnegativity is easy!

Solving this MIQP in Gurobi: experiment

Number of columns in dictionary s varies, $r = 4$, $n = 2s$



How to scale up?

Reduce the size of the dataset by removing the least useful columns using heuristics for SSC.

Intuition: if we want the 10 best columns, running a heuristic to extract 30 columns may return the 10 best ones + 20 other.

How to scale up?

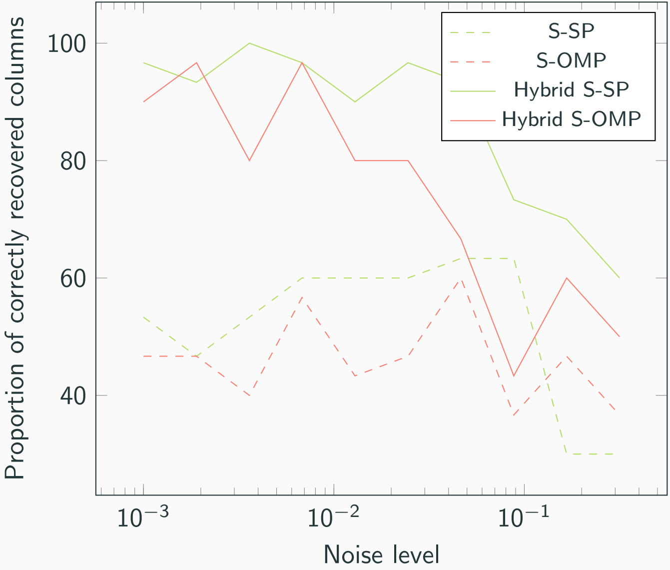
Reduce the size of the dataset by removing the least useful columns using heuristics for SSC.

Intuition: if we want the 10 best columns, running a heuristic to extract 30 columns may return the 10 best ones + 20 other.

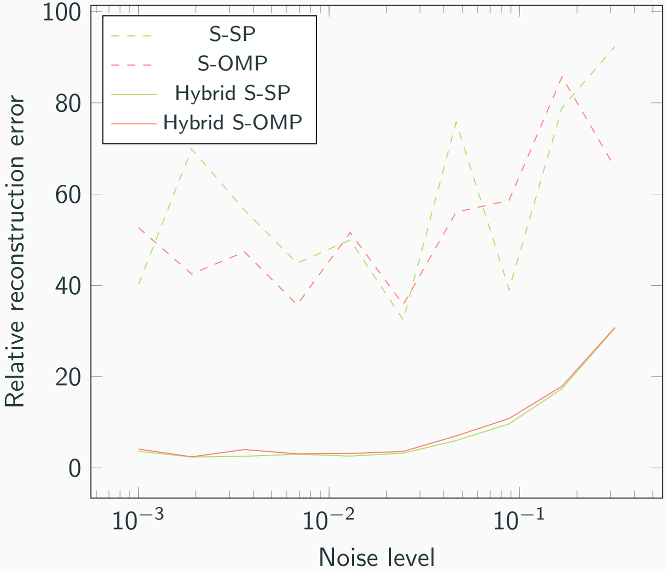
Hybrid method for SSC:

1. $H' \leftarrow \text{heuristic}(D, X, r')$
2. $J' \leftarrow \{i | H'(i, :) \neq 0\}$
3. $H^* \leftarrow \text{argmin}_{H, \|H\|_{row-0} \leq r} \|X - D(:, J')H\|_F^2$

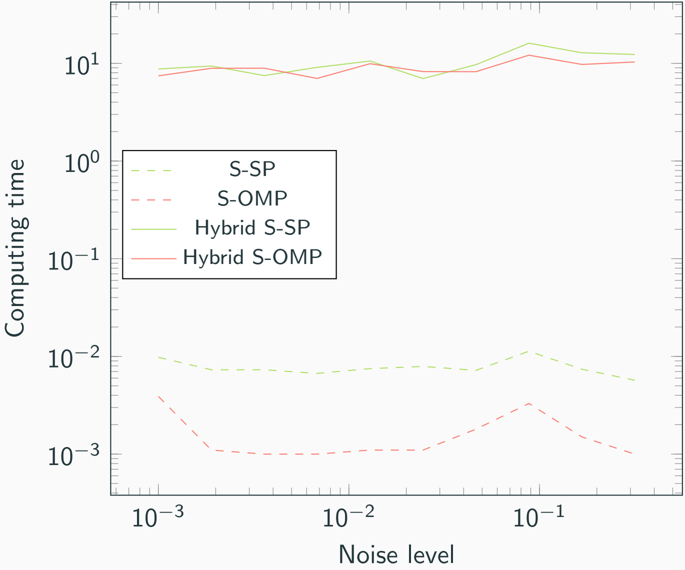
Experiments with hybrid method — recovery of columns



Experiments with hybrid method — reconstruction error



Experiments with hybrid method — computing time



Experiments with hybrid method — real-world hyperspectral unmixing

We perform Non-Negative SSC with self-dictionary ($D = X$) using our hybrid method on real-world hyperspectral images.

Experiments with hybrid method — real-world hyperspectral unmixing

We perform Non-Negative SSC with self-dictionary $(D = X)$ using our hybrid method on real-world hyperspectral images.

As a pre-processing heuristic, we use here the clustering-based algorithm H2NMF
We compare our results with two methods:

- FGNSR , an algorithm based on convex relaxation for NSSC
- NMFdico , a greedy algorithm for NSSC.

Experiments with hybrid method — real-world hyperspectral unmixing

After selecting the subset $X(:, J)$ of r columns of X, the computation of matrix $\hat{H}$ is a standard nonnegative least squares (NNLS) problem and we solve it using a block coordinate descent.

Experiments with hybrid method — real-world hyperspectral unmixing

After selecting the subset $X(:, J)$ of r columns of X, the computation of matrix $\hat{H}$ is a standard nonnegative least squares (NNLS) problem and we solve it using a block coordinate descent.

To evaluate the computed solutions without knowing the ground-truth solution, we measure the relative reconstruction error $\frac{\|X - X(:,J)\hat{H}\|_F}{\|X\|_F}$.

Experiments with hybrid method — real-world hyperspectral unmixing

Results of the unmixing of real-world hyperspectral images. Time is in seconds, error is the relative reconstruction error in percents. Bold numbers correspond to our hybrid method using H2NMF as a pre-processing heuristic.

	Data	San Diego	Urban	Terrain	Samson
	r	8	6	5	3
	r'	80	60	50	30
Time	FGNSR	0.04	0.03	0.03	0.01
	NMFdico	0.01	0.01	0.02	0.01
	Ours (Alg. 1)	**83.6**	**7.76**	**1.21**	**0.64**
Error	FGNSR	9.21	6.03	3.73	3.48
	NMFdico	9.05	6.03	3.52	3.2
	Ours (Alg. 1)	**8.35**	**4.27**	**3.32**	**3.06**

Conclusion

SSC

$$\min_{H} \|X - DH\|_F^2 \quad \text{s.t.} \quad \|H\|_{row-0} \leq r.$$

- We reformulated SSC as a MIQP, solvable globally by generic solvers.
- Doable for medium-size data, but does not scale well.
- Handle larger data sets with a hybrid method, using heuristics as a preprocess to pre-select columns and reduce the size of the data set.

Future work?

- Leverage the strong structure of the cardinality-constrained problem, to which MIP solvers are mostly indifferent.

Future work?

- Leverage the strong structure of the cardinality-constrained problem, to which MIP solvers are mostly indifferent.
- A branch-and-bound algorithm to solve SSC globally?

Future work?

- Leverage the strong structure of the cardinality-constrained problem, to which MIP solvers are mostly indifferent.
- A branch-and-bound algorithm to solve SSC globally?
- Key challenges:
  - Find a relaxation that is cheap to compute and allows for efficient pruning;
  - Find a way to solve efficiently the subproblems.

Thanks!

Contact: alexandra.dache@student.umons.ac.be

Code: https://gitlab.com/Alexia1305/SSC

Presentation based on the article :
Alexandra Dache, Nicolas Nadisic, Arnaud Vandaele, Nicolas Gillis (2023). Exact and Heuristic Methods for Simultaneous Sparse Coding. EUSIPCO 2023.