FACULTY OF ENGINEERING
AND ARCHITECTURE

16 September 2024 - Institut Henri Poincaré, Paris

Global Optimization for Simultaneous Sparse Coding

(a work in progress)

Nicolas Nadisic, Sébastien Bourguignon

GHENT UNIVERSITY

IRPA KIK Royal Institute for Cultural Heritage

Simultaneous sparse coding (SSC)

Given

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

the simultaneous sparse coding problem consists in finding $X \in \mathbb{R}^{s \times n}$ with at most r non-zero rows such that $Y \approx DX$.

$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|X\|_{row-0} \leq r. \tag{SSC}$$

where $\|X\|_{row-0} = \text{Card}(\{i | X(i,:) \neq 0\})$.

Simultaneous sparse coding (SSC)

$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|X\|_{row-0} \leq r. \tag{SSC}$$

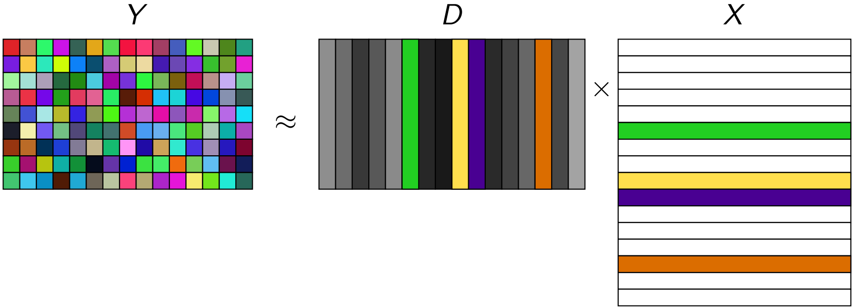
(SSC) is equivalent to finding a subset J of columns of D such that $|J| \leq r$ and $Y \approx D(:,J)\hat{X}$ for some matrix $\hat{X}$.

Also called:

- Simultaneous sparse approximation/recovery
- Joint sparse approximation
- Multiple measurement vectors
- Dictionary-based nonnegative matrix factorization

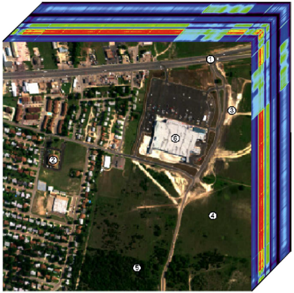
Simultaneous sparse coding — Illustration

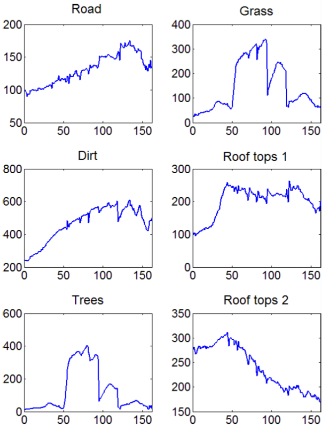
$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|X\|_{row-0} \leq r. \qquad \text{(SSC)}$$

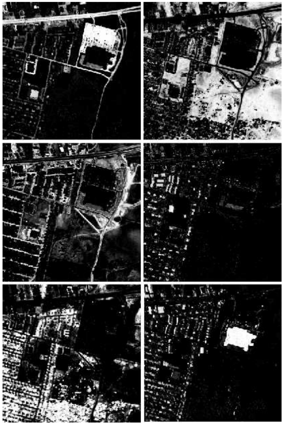


Application — Hyperspectral unmixing with known spectra library

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

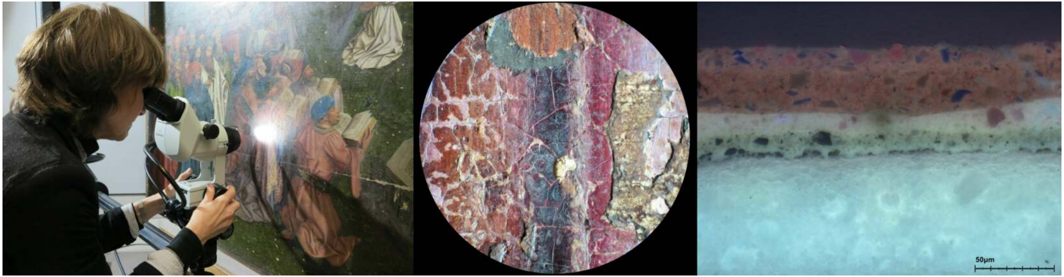






Images from J. Bioucas Dias and Nicolas Gillis.

Application — Fourier-transform infrared spectroscopy for polychromes



Images from KIK-IRPA.

Non-destructive analysis of the color/varnish layers.

Current state of the art for SSC

- Greedy algorithms: simultaneous variants of OMP, OLS, etc [Tropp, Gilbert, and Strauss 2006; Rakotomamonjy 2011; Kim and Haldar 2016; Belmerhnia et al. 2021]
- Convex relaxations: group LASSO, penalty or constraint on a mixed $\ell_{p,q}$ norm, etc [Tropp 2006; Gillis and Glineur 2012]
- Both have at best restrictive conditions for exact support recovery, or no guarantee at all.

A combinatorial problem

- (SSC) can be reduced to finding the row-support of X (set of nonzero rows).
- It is a combinatorial problem, with $\binom{s}{r}$ solutions.
- Bruteforce search is not realistic for large problems.

MIP reformulation

Let a binary decision variable $b \in \{0,1\}^s$, with $\begin{cases} b_i = 0 \text{ if } X(i,:) = 0, \\ b_i = 1 \text{ if } X(i,:) \neq 0. \end{cases}$

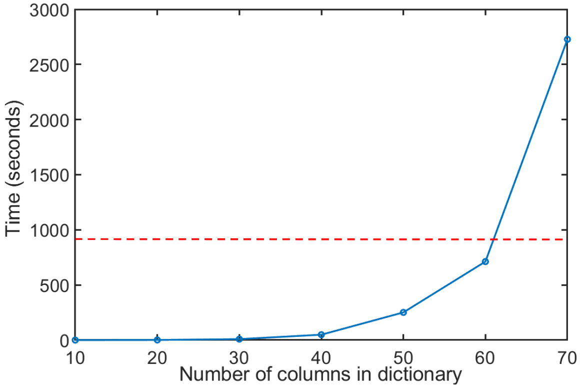
Problem (SSC) becomes:

$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \begin{cases} \sum_{i \in \{1,\dots,s\}} b_i \leq r, \\ -b_i M \leq X(i,j) \leq b_i M \text{ for all } i,j. \end{cases} \qquad \text{(SSC-MIP)}$$

Where M is a constant.

Solving the MIP directly

We solved (SSC-MIP) in [Dache et al. 2023], using Gurobi.

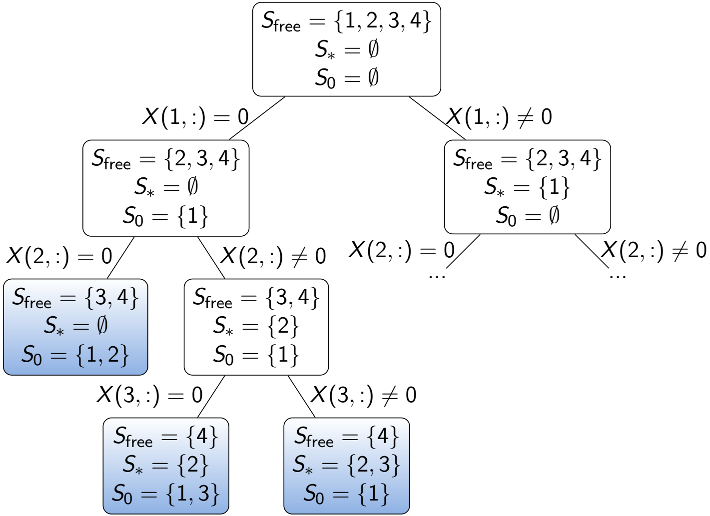


$r = 4$, $n = 2s$
The red dotted line represents 15 minutes.

Branch and Bound (BaB)

- Idea of branch and bound: explore the search space in a smart way, pruning useless parts to avoid computations but still finding the global optimum.

Portion of a BaB search tree for $s = 4$ and $r = 2$.



Pruning the search space

We can prune a node and leave the descending nodes unxplored when we are certain that the optimal solution does not descend from this node.
In practice, we compute bounds on the error of the best solution:

- Global upper bound: given by an admissible solution.
- Lower bound: given by a relaxation of the problem.
- If lower bound $>$ upper bound, we can prune.

Finding a relaxation

The simplest one, but not tight at all: remove the constraint.
Already helps to prune the search space and is cheap to compute.

Finding a relaxation

A tight convex relaxation of the row-0 "norm" is the $\ell_{1,\infty}$-norm [Tropp 2006],

$$\|X\|_{1,\infty} = \sum_{i=1}^{s} \max_{j \in \{1,\dots,n\}} |X(i,j)|.$$

So we can relax (SSC) into

$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \begin{cases} \|X\|_{1,\infty} \leq rM, \\ -M \leq X(i,j) \leq M \text{ for all } i,j. \end{cases} \tag{SSC-Rel}$$

Rewriting the relaxation to solve it more easily

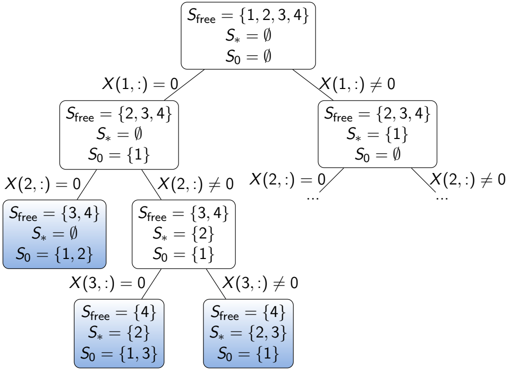
$$\min_{X \in \mathbb{R}^{s \times n}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \begin{cases} \|X\|_{1,\infty} \leq rM, \\ -M \leq X(i,j) \leq M \text{ for all } i,j. \end{cases} \tag{SSC-Rel}$$

We introduce an auxiliary variable $w \in \mathbb{R}^s$ to rewrite the constraint $\|X\|_{1,\infty} \leq rM$ linearly. The idea is to have for all i, w_i "replace" $\max_{j \in \{1,\ldots,n\}} X(i,j)$.

$$\min_{X \in \mathbb{R}^{s \times n}, w \in \mathbb{R}^s} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \begin{cases} \sum_{i=1}^{s} |w_i| \leq rM, \\ X(i,j) \leq w_i \text{ for all } j, \\ -M \leq X(i,j) \leq M \text{ for all } i,j. \end{cases} \tag{SSC-Rel-Lin}$$

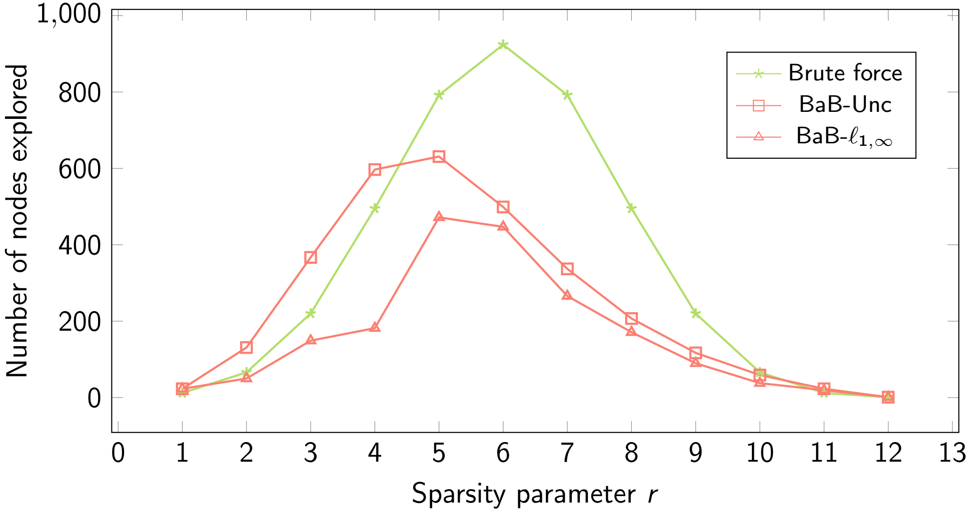
This is a problem with quadratic objective and linear constraints, easy to tackle with a generic solver.

Order of exploration of the tree

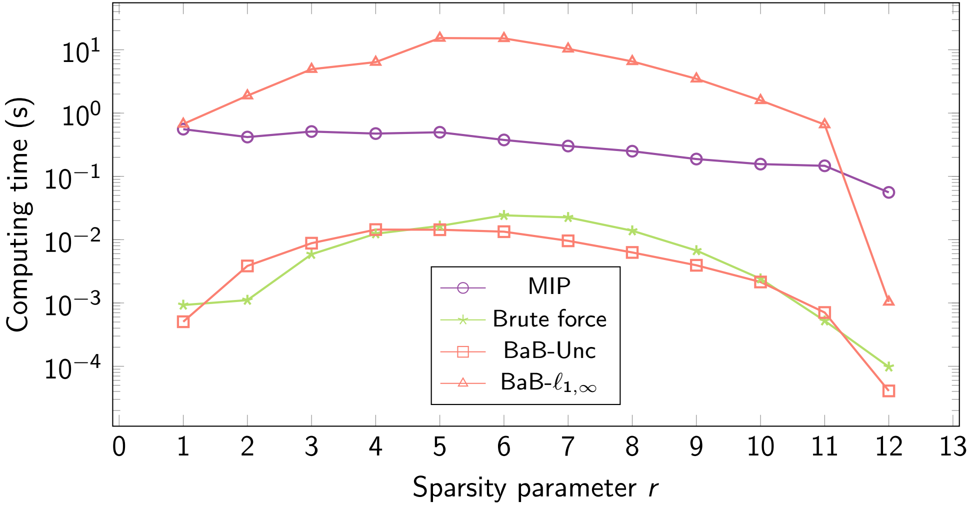


Depth-first search $\Rightarrow$ get an admissible solution faster $\Rightarrow$ early pruning

Numerical results on synthetic noisy data — $Y \in \mathbb{R}^{20\times 100}, s = 12$



Numerical results on synthetic noisy data — $Y \in \mathbb{R}^{20\times 100}$, $s = 12$



Conclusion and future work

The good news:

- Branch-and-bound reduces significantly the number of nodes to explore
- The relaxation $\ell_{1,\infty}$ is efficient in pruning nodes

The things to improve:

- An efficient way to compute the relaxed $\ell_{1,\infty}$-constrained problem?
- A way to compute a lower bound to this relaxed problem (eg using the dual)?
- Consider other relaxations, maybe less tight but easier to compute (eg $\ell_{1,2}$, $\ell_{1,1}$)?
- Adapt the branching order to sparsity parameter r?

References I

- Leila Belmerhnia et al. "Simultaneous variable selection for the classification of near infrared spectra". In: Chemometr. Intell. Lab. Syst. 211 (2021), p. 104268.
- Alexandra Dache et al. "Exact and Heuristic Methods for Simultaneous Sparse Coding". In: 31st European Signal Processing Conference (EUSIPCO). IEEE. 2023, pp. 1753–1757.
- Nicolas Gillis and François Glineur. "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization". In: Neural computation 24.4 (2012), pp. 1085–1105.
- Daeun Kim and Justin P Haldar. "Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery". In: Signal processing 125 (2016), pp. 274–289.
- Alain Rakotomamonjy. "Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms". In: Signal processing 91.7 (2011), pp. 1505–1526.

References II

Joel A Tropp. "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation". In: Signal Processing 86.3 (2006), pp. 589–602.

Joel A Tropp, Anna C Gilbert, and Martin J Strauss. "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit". In: Signal processing 86.3 (2006), pp. 572–588.

Thanks!

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