Node-screening tests for the L0-penalized least-squares problem

- **Theo Guyard** (Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000 Rennes, France)
- Cedric Herzet (INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35000 Rennes, France)
- Clement Elvira (SCEE/IETR UMR CNRS 6164, CentraleSupélec, 35510 Cesson Sévigné, France)

Keywords: Sparse approximation, Mixed-integer problems, Branch and bound, Safe screening.

Abstract: Finding a sparse representation is a fundamental problem in the field of statistics, machine learning and inverse problems. It consists in decomposing some input vector $\mathbf{y} \in \mathbf{R}^m$ as a linear combination of a few columns of a dictionary $\mathbf{A} \in \mathbf{R}^{m \times n}$. This task can be addressed by solving

$$\min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 \tag{1}$$

where $\|\mathbf{x}\|_0$ counts the number of nonzero entries in $\mathbf{x}$ and $\lambda > 0$ is a tuning parameter. There has recently been a surge of interest for methods solving (1) based on its Mixed-Integer Program reformulation, most of them leveraging a tailored Branch-and-Bound (BnB) algorithm [1, 2].

In this context, Atamtürk *et al.* extended the notion of safe screening introduced in [3] from sparsity-promoting convex problems to some non-convex ℓ_0-penalized problems [4]. In particular, they introduced a new methodology allowing to detect some of the positions of zero and non-zero entries in the minimizers of a particular ℓ_0-penalized problem. This methodology can be used as a preprocessing step of any algorithmic procedure and allows to reduce the problem dimension. We go one step further in the development of numerical methods addressing large-scale ℓ_0-penalized problems by proposing *node-screening* rules that allow to prune *nodes* within the BnB search tree. In contrast to [4], we emphasize the existence of a *nesting property* between screening tests at different nodes. This enables to potentially fix *multiple* entries to either zero or non-zero at *any step* of the optimization process with a marginal cost.

Our method leverages the connection between the objectives of the Fenchel dual of the relaxed problems solved at each node of the BnB algorithm. More precisely, the objective of this dual problem at a given node is composed of a term common to all nodes and terms corresponding to the current branching constraints. Hence, the dual objective only differs by one term between two consecutive nodes in the BnB tree. Using this relation, we construct two *node-screening* tests allowing to identify branching constraints that cannot lead to a global solution to (1). If one of our test is passed at a given node, we can avoid processing half of its children. This results in a reduction in the number of nodes processed by the BnB algorithm.

In Table 1, we compare CPLEX to a BnB algorithm tailored to (1) with or without our node-screening methodology. We use two different setups where $\mathbf{A} \in \mathbf{R}^{m \times n}$ is either a random matrix with normally-generated entries with $(m, n) = (500, 1000)$ or inheriting from a Tœplitz structure with $(m, n) = (500, 300)$. For a given sparsity level k, we generate the input vector as $\mathbf{y} = \mathbf{A}\mathbf{x}^\star + \boldsymbol{\epsilon}$ where $\mathbf{x}^\star$ is a k-sparse vector and $\boldsymbol{\epsilon}$ is a noise with 10dB Signal-to-Noise ratio. Our results emphasize the acceleration brought by our methodology, which is directly induced by the reduction in the number of nodes explored.

		CPLEX			BnB			BnB+screening		
	k	Nds	T	F	Nds	T	F	Nds	T	F
Gauss.	5	0.1	26	0	0.1	2	0	**0.0**	**1**	**0**
	7	0.3	61	0	0.2	5	0	**0.1**	**3**	**0**
	9	0.8	103	10	0.5	16	0	**0.4**	**10**	**0**
Tœpl.	5	1.4	10	0	1.0	6	0	**0.7**	**4**	**0**
	7	17.7	107	0	10.5	79	0	**7.9**	**52**	**0**
	9	80.7	353	50	47.8	346	48	**41.2**	**267**	**40**

Table 1: Number of nodes explored $\times 10^3$ (Nds), optimization time in seconds (T) and percentage of instances not solved within 1,000 seconds (F).

References:

[1] R. Ben Mhenni, S. Bourguignon and J. Ninin. Global optimization for sparse solution of least squares problems. *Optimization Methods and Software*, 2021, p. 1-30.

[2] D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *The annals of statistics*, 2016, vol. 44, no 2, p. 813-852.

[3] L. El Ghaoui, V. Viallon, and T. Rabbani. Safe feature elimination for the lasso and sparse supervised learning problems. *Pacific Journal of Optimization*, 2021, vol. 8, no 4, p. 667-698.

[4] A. Atamturk and A. Gomez. Safe screening rules for l0-regression from perspective relaxations. *International conference on machine learning*, PMLR, 2020. p. 421-430.