SCREEN & RELAX: ACCELERATING THE RESOLUTION OF ELASTIC-NET BY SAFE IDENTIFICATION OF THE SOLUTION SUPPORT

Théo Guyard[⋆] *Cédric Herzet*[†] *Clément Elvira*[‡]

[⋆] Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000, France
[†] INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35000 Rennes, France
[‡] IETR UMR CNRS 6164, CentraleSupelec Rennes Campus, 35576 Cesson Sévigné, France
firstname.lastname@{insa-rennes,inria,centralesupelec}.fr

ABSTRACT

In this paper, we propose a procedure to accelerate the resolution of the well-known "Elastic-Net" problem. Our procedure is based on the (partial) identification of the solution support and the reformulation of the original problem into a problem of reduced dimension. The identification of the support leverages the novel concept of *"safe relaxing"* where one aims to identify *non-zero* coefficients of the solution. It can be viewed as a dual approach to *"safe screening"* introduced in the last decade and allowing to reduce the problem dimension using the identification of *zero* coefficients of the solution. We show numerically that combining both methodologies in a *"Screen & Relax"* strategy enables to significantly improve the trade-off between complexity and accuracy achievable by standard resolution techniques.

***Index Terms*—** Convex optimization, Sparsity, Safe screening, Acceleration techniques, Constraint relaxation.

1. INTRODUCTION

Sparse decomposition aims at finding some approximation of a vector $\mathbf{y} \in \mathbb{R}^m$ as the linear combination of a few columns (dubbed *atoms*) of a dictionary $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$. Unfortunately, identifying the sparsest decomposition of a vector according to some accuracy criterion often turns out to be a combinatorial problem [1, Sec. 2.3]. A standard strategy to circumvent this issue consists in approximating this ideal decomposition as the solution of a problem of the form

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x} \in \mathbb{R}^n} \; \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \Omega(\mathbf{x}) \tag{1}$$

where $\Omega \colon \mathbb{R}^n \to \mathbb{R}_+$ is some *sparsity-inducing* convex regularizer. The common choice $\Omega(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ for some $\lambda > 0$ leads to the well-known *"Lasso"* problem and has been extensively studied in the literature [2, 3]. Another standard choice is $\Omega(\mathbf{x}) = \lambda\|\mathbf{x}\|_1 + \frac{\varepsilon}{2}\|\mathbf{x}\|_2^2$ for some parameters $\lambda > 0, \varepsilon > 0$. In this case, problem (1) is known as *"Elastic-Net"* and is popular in many applicative domains because its solution enjoys desirable statistical properties [4].

Because of its clear practical interest, many contributions of the literature have proposed efficient solving procedures for (1), see *e.g.*, [5–8]. Of particular interest in this paper is the "safe screening" acceleration technique proposed by El Ghaoui *et al.* in [9]. Safe screening consists in performing simple tests to identify the *zero* elements of the minimizers of an optimization problem. This knowledge can then be exploited to reduce the dimensionality of the problem by discarding the atoms of the dictionary weighted by zero safely, *i.e.*, without changing the solution set. Over the past decade, many authors have identified safe screening as a simple procedure to significantly speed up the resolution of many optimization problems, see *e.g.*, [10–15].

In this paper, we introduce a dual approach to safe screening, dubbed *"safe relaxing"*. We focus on a specific instance of problem (1), namely the non-negative version of Elastic-Net. Our method aims at identifying the position of the *non-zero* coefficients of the minimizer of this problem. We show that, similarly to screening, this knowledge can be exploited to safely reduce the dimensionality of the target problem and accelerate its resolution. We use the terminology "relaxing" as the reduction of the problem dimensionality results from the relaxation of some constraints.

The rest of the paper is organized as follows. The target problem is defined in Sec. 2. The concepts of "safe screening" and "safe relaxing" are presented in Sec. 3 and 4. In Sec. 5, we combine screening and relaxing methodologies in a *"Screen & Relax"* strategy. A numerical evaluation of the proposed method is finally carried out in Sec. 6.

Notations. Boldface uppercase (*e.g.*, $\mathbf{A}$) and lowercase (*e.g.*, $\mathbf{x}$) letters respectively represent matrices and vectors. $\mathbf{0}_n$ and $\mathbf{1}_n$ stand for the n-dimensional all-zeros and all-ones vectors. $\mathbf{I}$ represents the identity matrix whose dimension is usually clear from the context. The ith component of $\mathbf{x}$ is

The research presented in this paper is reproducible. Code and data are available at `https://gitlab.insa-rennes.fr/Theo.Guyard/screen-and-relax`

978-1-6654-0540-9/22/$31.00 ©2022 IEEE

denoted $\mathbf{x}(i)$. Calligraphic letters (*e.g.*, $\mathcal{I}$) are used to denote sets and the notation $\overline{\mathcal{I}}$ refers to the complementary set of $\mathcal{I}$. We denote by $\mathbf{x}_{\mathcal{I}}$ the restriction of $\mathbf{x}$ to its elements indexed by $\mathcal{I}$ and $\mathbf{A}_{\mathcal{I}}$ corresponds to the restriction of $\mathbf{A}$ to its columns indexed by $\mathcal{I}$. Finally, for any real symmetric positive definite matrix $\mathbf{M}$, we let $\|\mathbf{x}\|_{\mathbf{M}}^2 \triangleq \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}$. Throughout this paper, we assume without loss of generality that the columns of $\mathbf{A}$ are normalized to one.

2. TARGET PROBLEM

We focus on the non-negative version of Elastic-Net:

$$\min_{\mathbf{x}\geq \mathbf{0}_n} P(\mathbf{x}) \triangleq \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x} + \tfrac{\varepsilon}{2}\|\mathbf{x}\|_2^2 \qquad (2\text{-}\mathcal{P})$$

where $\boldsymbol{\lambda} \in \mathbb{R}_+^n$ and $\varepsilon > 0$. We note that the standard formulation of Elastic-Net can be seen as a particular case of (2-$\mathcal{P}$) (see *e.g.*, [16, Sec. 2]). Since $P(\cdot)$ is continuous, coercive and strongly convex, (2-$\mathcal{P}$) admits a unique minimizer $\mathbf{x}^\star$. The goal of this paper is to accelerate the resolution of (2-$\mathcal{P}$) by identifying the position of the zero and non-zero coefficients of $\mathbf{x}^\star$. Our strategy leverages the primal-dual optimality conditions described below.

The dual problem associated to (2-$\mathcal{P}$) reads

$$\max_{\mathbf{u}\in\mathbb{R}^m} \tfrac{1}{2}\|\mathbf{y}\|_2^2 - \tfrac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2 - \tfrac{1}{2\varepsilon}\|[\mathbf{A}^{\mathrm{T}}\mathbf{u}-\boldsymbol{\lambda}]_+\|_2^2 \qquad (3\text{-}\mathcal{D})$$

where $[\mathbf{x}]_+ \triangleq \max(\mathbf{0}_n, \mathbf{x})$ and with the maximum taken component-wise [17, Sec. 5.2]. Similarly to (2-$\mathcal{P}$), the cost function in (3-$\mathcal{D}$) is continuous, coercive and strongly concave. Problem (3-$\mathcal{D}$) thus admits a unique maximizer $\mathbf{u}^\star$. By Slater's constraint qualification, strong duality holds between (2-$\mathcal{P}$) and (3-$\mathcal{D}$). As a consequence, a couple $(\mathbf{x}^\star, \mathbf{u}^\star)$ is a primal-dual solution of (2-$\mathcal{P}$)-(3-$\mathcal{D}$) if and only if

$$\mathbf{u}^\star = \mathbf{y} - \mathbf{A}\mathbf{x}^\star \qquad (4)$$

$$\mathbf{x}^\star = \varepsilon^{-1}[\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star - \boldsymbol{\lambda}]_+. \qquad (5)$$

See [18, Prop. 5.1.5 and 5.3.1] for technical details. In particular, letting $\mathcal{J}^\star \triangleq \{\ell\colon \mathbf{x}^\star(\ell) > 0\}$, we also easily obtain from (4)-(5) that

$$\mathbf{x}^\star_{\mathcal{J}^\star} = (\mathbf{A}_{\mathcal{J}^\star}^{\mathrm{T}}\mathbf{A}_{\mathcal{J}^\star} + \varepsilon\mathbf{I})^{-1}(\mathbf{A}_{\mathcal{J}^\star}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\lambda}_{\mathcal{J}^\star}). \qquad (6)$$

3. SAFE SCREENING

The goal of safe screening is to identify the zero components of $\mathbf{x}^\star$ in order to transform (2-$\mathcal{P}$) into a problem of reduced dimension and speed-up its resolution. More precisely, let

$$\mathcal{I} \subseteq \{\ell\colon \mathbf{x}^\star(\ell) = 0\} \qquad (7)$$

denote a subset of the zero components of $\mathbf{x}^\star$. Then, (2-$\mathcal{P}$) is equivalent to

$$\mathbf{x}^\star = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) \;\text{ s.t. } \begin{cases} \mathbf{x}_{\overline{\mathcal{I}}} & \geq \mathbf{0}_{n_r} \\ \mathbf{x}_{\mathcal{I}} & = \mathbf{0}_{n-n_r} \end{cases} \qquad (8)$$

where $n_r \triangleq \operatorname{card}(\overline{\mathcal{I}})$. This problem can also be rewritten more explicitly as

$$\mathbf{x}^\star_{\overline{\mathcal{I}}} = \operatorname*{arg\,min}_{\mathbf{x}_r \geq \mathbf{0}_{n_r}} \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}_r\mathbf{x}_r\|_2^2 + \boldsymbol{\lambda}_r^{\mathrm{T}}\mathbf{x}_r + \tfrac{\varepsilon}{2}\|\mathbf{x}_r\|_2^2 \qquad (9a)$$

$$\mathbf{x}^\star_{\mathcal{I}} = \mathbf{0}_{n-n_r} \qquad (9b)$$

where $\mathbf{A}_r \triangleq \mathbf{A}_{\overline{\mathcal{I}}} \in \mathbb{R}^{m\times n_r}$ and $\boldsymbol{\lambda}_r \triangleq \boldsymbol{\lambda}_{\overline{\mathcal{I}}} \in \mathbb{R}^{n_r}$. In the above formulation, we note that (9a) has the same structure as (2-$\mathcal{P}$) but with a reduced optimization domain of dimension n_r instead of n. Hence, if $n_r \ll n$, huge computational savings can potentially be achieved by considering the reduced formulation (9a) instead of (2-$\mathcal{P}$).

Safe screening tests aim to identify some subset $\mathcal{I} \subseteq \{1,\ldots,n\}$ verifying (7). The design of such tests usually leverages the optimality conditions of the problem at stake. As far as (2-$\mathcal{P}$) is concerned, we have from (5) that

$$\forall \ell \in \{1,\ldots,n\}:\; \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}^\star \leq \boldsymbol{\lambda}(\ell) \iff \mathbf{x}^\star(\ell) = 0. \qquad (10)$$

The left-hand side of the equivalence is thus a sufficient condition for $\mathbf{x}^\star(\ell)$ to be equal to zero. Unfortunately, computing $\mathbf{u}^\star$ is usually as difficult as solving primal problem (2-$\mathcal{P}$) and (10) is therefore of poor practical interest.

This difficulty can be circumvented by using *"safe regions"*, that is subsets of the dual domain that are guaranteed to contain $\mathbf{u}^\star$. For example, assuming that $\mathbf{u}^\star$ belongs to a safe spherical regions, that is

$$\mathbf{u}^\star \in \mathcal{S}(\mathbf{c}, r) \triangleq \{\mathbf{u} \in \mathbb{R}^m\colon \|\mathbf{u}-\mathbf{c}\|_2 \leq r\}, \qquad (11)$$

test (10) can be relaxed as

$$\max_{\mathbf{u}\in\mathcal{S}(\mathbf{c},r)} \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u} = \mathbf{a}_\ell^{\mathrm{T}}\mathbf{c} + r \leq \boldsymbol{\lambda}(\ell) \implies \mathbf{x}^\star(\ell) = 0. \qquad (12)$$

Methods to construct safe spheres have been extensively studied in the literature over the past decade, see *e.g.*, [9–12, 14, 15, 19, 20].

4. SAFE RELAXING

In this section, we expose our *"safe relaxing"* methodology. In contrast to safe screening, our goal is to identify the positions of the *non-zero* coefficients of $\mathbf{x}^\star$. We show that the identification of these components can also lead to an equivalent problem of reduced dimension. More precisely, let

$$\mathcal{J} \subseteq \{\ell\colon \mathbf{x}^\star(\ell) > 0\} \qquad (13)$$

denote a subset of non-zero components of $\mathbf{x}^\star$. Problem (2-$\mathcal{P}$) can then be equivalently expressed as

$$\mathbf{x}^\star = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) \;\text{ s.t. } \begin{cases} \mathbf{x}_{\overline{\mathcal{J}}} & \geq \mathbf{0}_{n_r} \\ \mathbf{x}_{\mathcal{J}} & \in \mathbb{R}^{n-n_r} \end{cases} \qquad (14)$$

where $n_r \triangleq \operatorname{card}(\overline{\mathcal{J}})$. We note that the constraints on the elements in $\mathcal{J}$ have been totally removed in (14). This is in contrast with screening where the elements $\mathbf{x}^\star_{\mathcal{I}}$ were set to zero. Similarly to screening, this relaxation allows to express (2-$\mathcal{P}$) as a problem of reduced dimension.

Let us first notice that the restriction of (14) to $\mathbf{x}^\star_{\overline{\mathcal{J}}}$ can be written as :

$$\mathbf{x}^\star_{\overline{\mathcal{J}}} = \operatorname*{arg\,min}_{\mathbf{x}_{\overline{\mathcal{J}}}\geq \mathbf{0}_{n_r}} \left(\min_{\mathbf{x}_{\mathcal{J}}\in\mathbb{R}^{n-n_r}} P(\mathbf{x}) \right). \tag{15}$$

Since the inner minimization in (15) is a strongly-convex quadratic problem, it admits the unique optimizer

$$\mathbf{x}_{\mathcal{J}} = \mathbf{B}\mathbf{x}_{\overline{\mathcal{J}}} + \mathbf{b} \tag{16}$$

where

$$\mathbf{B} \triangleq -(\mathbf{A}_{\mathcal{J}}^{\mathrm{T}}\mathbf{A}_{\mathcal{J}} + \varepsilon\mathbf{I})^{-1}\mathbf{A}_{\mathcal{J}}^{\mathrm{T}}\mathbf{A}_{\overline{\mathcal{J}}} \tag{17a}$$

$$\mathbf{b} \triangleq -(\mathbf{A}_{\mathcal{J}}^{\mathrm{T}}\mathbf{A}_{\mathcal{J}} + \varepsilon\mathbf{I})^{-1}\big(\mathbf{A}_{\mathcal{J}}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\lambda}_{\mathcal{J}}\big). \tag{17b}$$

Plugging (16) into the cost function $P(\mathbf{x})$ then leads to the following equivalent formulation of (2-$\mathcal{P}$):

$$\mathbf{x}^\star_{\overline{\mathcal{J}}} = \operatorname*{arg\,min}_{\mathbf{x}_r\geq\mathbf{0}_{n_r}} \tfrac{1}{2}\|\mathbf{y}_r - \mathbf{A}_r\mathbf{x}_r\|_2^2 + \boldsymbol{\lambda}_r^{\mathrm{T}}\mathbf{x}_r + \tfrac{\varepsilon}{2}\|\mathbf{x}_r\|_{\mathbf{M}}^2 \tag{18a}$$

$$\mathbf{x}^\star_{\mathcal{J}} = \mathbf{B}\mathbf{x}^\star_{\overline{\mathcal{J}}} + \mathbf{b} \tag{18b}$$

where

$$\mathbf{A}_r \triangleq \mathbf{A}_{\overline{\mathcal{J}}} + \mathbf{A}_{\mathcal{J}}\mathbf{B} \tag{19a}$$

$$\boldsymbol{\lambda}_r \triangleq \boldsymbol{\lambda}_{\overline{\mathcal{J}}} + \mathbf{B}^{\mathrm{T}}(\boldsymbol{\lambda}_{\mathcal{J}} + \varepsilon\mathbf{b}) \tag{19b}$$

$$\mathbf{y}_r \triangleq \mathbf{y} - \mathbf{A}_{\mathcal{J}}\mathbf{b} \tag{19c}$$

$$\mathbf{M} \triangleq \mathbf{I} + \mathbf{B}^{\mathrm{T}}\mathbf{B}. \tag{19d}$$

Similarly to screening, the reduced problem (18a) has the same mathematical structure as (2-$\mathcal{P}$). The definition of the parameters $(\mathbf{A}_r, \boldsymbol{\lambda}_r, \mathbf{y}_r, \mathbf{M})$ in (18a) differs however quite significantly from those in (9a). In particular, whereas the construction of $\mathbf{A}_r$ only requires to remove some columns from $\mathbf{A}$ in (9a), it involves a matrix inversion in (18a). This operation introduces some complexity overhead and must therefore be performed with care as discussed in Sec. 6.

Optimality condition (5) can be exploited to identify some subset $\mathcal{J}$ verifying (13). In particular, we have

$$\forall \ell \in \{1,\dots,n\}: \ \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}^\star > \boldsymbol{\lambda}(\ell) \iff \mathbf{x}^\star(\ell) > 0. \tag{20}$$

Similarly to screening, we can resort to a safe sphere (11) to obtain a weaker, yet practical, version of (20). This leads to the following *relaxing test*:

$$\min_{\mathbf{u}\in\mathcal{S}(\mathbf{c},r)} \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u} = \mathbf{a}_\ell^{\mathrm{T}}\mathbf{c} - r > \boldsymbol{\lambda}(\ell) \implies \mathbf{x}^\star(\ell) > 0. \tag{21}$$

Algorithm 1: "Screen & Relax" solving procedure

```
   Input: $\mathbf{x}^{(0)}$, $\mathbf{A}$, $\mathbf{y}$, $\boldsymbol{\lambda}$, $\varepsilon$
1  $t \leftarrow 1$
2  $(\mathcal{I}, \mathcal{J}, \mathcal{K}) \leftarrow (\emptyset, \emptyset, \emptyset)$
3
4  $(\mathbf{A}_r, \boldsymbol{\lambda}_r, \mathbf{y}_r, \mathbf{M}) \leftarrow (\mathbf{A}, \mathbf{y}, \boldsymbol{\lambda}, \mathbf{I})$
5  while convergence criterion is not met do
6      $\mathbf{x}^{(t)}_{\overline{\mathcal{K}}} \leftarrow$ DescentStep($\mathbf{x}^{(t-1)}_{\overline{\mathcal{K}}}, \mathbf{A}_r, \mathbf{y}_r, \boldsymbol{\lambda}_r, \mathbf{M}, \varepsilon$)
7      Compute a new safe sphere $\mathcal{S}(\mathbf{c}^{(t)}, r^{(t)})$
8      Update $\mathcal{I}$ with test (12)          // Screening test
9      Update $\mathcal{J}$ with test (21)          // Relaxing test
10     $\mathcal{K} \leftarrow \mathcal{I} \cup \mathcal{J}$
11     Update $\mathbf{A}_r, \mathbf{y}_r, \boldsymbol{\lambda}_r, \mathbf{M}$ with (19a)-(19d)
12     $t \leftarrow t + 1$
13 end
```

5. SCREEN AND RELAX

The "screening" and "relaxing" procedures described in Sec. 3 and 4 can obviously be combined in a *"Screen & Relax"* strategy to benefit from the identification of *both* zero and non-zero components of $\mathbf{x}^\star$. More precisely, let $\mathcal{I}$ and $\mathcal{J}$ be subsets respectively verifying (7) and (13) and let $\mathcal{K} \triangleq (\mathcal{I} \cup \mathcal{J})$ be the set of components of $\mathbf{x}^\star$ already identified as zero or non-zero. Applying the same reasoning as in Sec. 3 and 4, we then obtain that (2-$\mathcal{P}$) is equivalent to

$$\mathbf{x}^\star_{\overline{\mathcal{K}}} = \operatorname*{arg\,min}_{\mathbf{x}_r\geq\mathbf{0}_{n_r}} \tfrac{1}{2}\|\mathbf{y}_r - \mathbf{A}_r\mathbf{x}_r\|_2^2 + \boldsymbol{\lambda}_r^{\mathrm{T}}\mathbf{x}_r + \tfrac{\varepsilon}{2}\|\mathbf{x}_r\|_{\mathbf{M}}^2 \tag{22a}$$

$$\mathbf{x}^\star_{\mathcal{J}} = \mathbf{B}\mathbf{x}^\star_{\overline{\mathcal{K}}} + \mathbf{b} \tag{22b}$$

$$\mathbf{x}^\star_{\mathcal{I}} = \mathbf{0}_{\mathrm{card}(\mathcal{I})}, \tag{22c}$$

where the parameters $(\mathbf{A}_r, \boldsymbol{\lambda}_r, \mathbf{y}_r, \mathbf{M})$ are defined as in (17a)-(19d) by using $\overline{\mathcal{K}}$ instead of $\overline{\mathcal{J}}$. The dimension of reduced problem (22a) is equal to $n_r = \mathrm{card}(\overline{\mathcal{K}})$ and thus benefits from the identification of both the zero and non-zero components of $\mathbf{x}^\star$ in its dimensionality reduction.

Quite interestingly, when equality holds in (7) and (13), relations (22b)-(22c) entirely define the solution of (2-$\mathcal{P}$). In this case, (22b) reduces to (6). The solution of (2-$\mathcal{P}$) can therefore be computed to *machine-precision* via simple linear-algebra operations when all components of $\mathbf{x}^\star$ have either passed a screening or a relaxing test.

6. NUMERICAL RESULTS

In this section, we evaluate the computational gain induced by the proposed safe relaxing strategy. We focus on the resolution of (2-$\mathcal{P}$) with $\boldsymbol{\lambda} = \lambda\mathbf{1}_n$ for some $0 < \lambda < \lambda_{\max} \triangleq \max(\mathbf{A}^{\mathrm{T}}\mathbf{y})$. We mention that $\mathbf{x}^\star = \mathbf{0}_n$ as soon as $\lambda \geq \lambda_{\max}$.

We consider the "Screen & Relax" (S&R) procedure described in Alg. 1. The function "DescentStep" in line 6 corresponds to one iteration of an accelerated proximal gradient

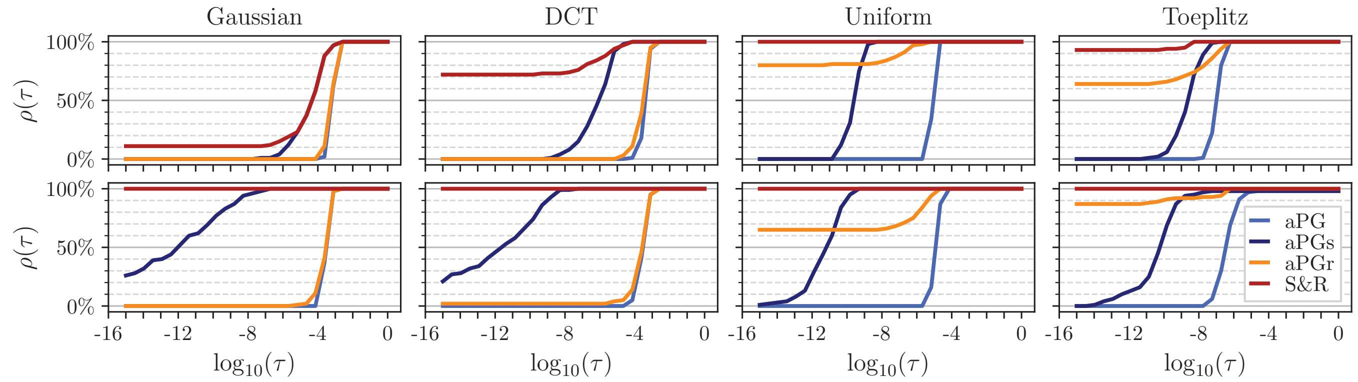

Fig. 1. Dolan-Moré performance profiles for $(\lambda, \varepsilon) = (0.2, 0.5)\lambda_{\max}$ (top) and $(\lambda, \varepsilon) = (0.5, 0.2)\lambda_{\max}$ (bottom).

algorithm [6, Sec. 4.3] applied to problem (22a). Recall that at the beginning of the solving procedure, $\mathcal{I} = \mathcal{J} = \emptyset$. The evaluation of the safe sphere parameters $\mathbf{c}^{(t)}$ and $r^{(t)}$ in line 7 follows the "GAP" methodology presented in [21, Th. 6]. At each iteration, problem (22a) is updated in line 11 upon the identification of additional zero or non-zero components. We note that $\mathcal{J}$ typically only varies by (at most) a few elements at each iteration of Alg. 1. This behavior can be exploited to efficiently compute the inverse in (17a)-(17b) by using rank-one update rules [22].

We compare the performance of the S&R procedure with three restricted versions of Alg. 1 : *i)* no screening and no relaxing is performed (*i.e.*, lines 8-9 are skipped); *ii)* only screening is performed (*i.e.*, line 9 is skipped); *iii)* only relaxing is performed (*i.e.*, line 8 is skipped). These variants will respectively be denoted "aPG", "aPGs" and "aPGr" in the sequel. Both aPG and aPGs correspond to standard methodologies of the literature while aPGr and S&R are contributions of the present paper.

We use "Dolan-Moré" performance profiles [23] to assess the performance of these four methods. Our results are gathered in Fig. 1. To generate each curve, we run a solving method with a *given* computational budget on 100 different instances of problem (2-$\mathcal{P}$). The curve corresponds to the percentage $\rho(\tau)$ of problem instances for which the solving strategy achieves a duality gap [17] lower than τ.

To generate problem data, we consider the four following setups: the elements of $\mathbf{A}$ are i.i.d. realizations of *i)* a standard normal distribution or *ii)* a uniform law on $[0, 1]$; *iii)* the rows of $\mathbf{A}$ are randomly-sampled from a DCT matrix [24]; *iv)* $\mathbf{A}$ has a Toeplitz structure [25] with shifted versions of a Gaussian curve. In all setups, the columns of $\mathbf{A}$ are normalized to one. The observation $\mathbf{y}$ is drawn according to a uniform distribution on the m-dimensional sphere for "Gaussian" and "DCT" dictionaries and is restricted to the positive orthant for "Uniform" and "Toeplitz" dictionaries. We set $m = 100$, $n = 300$ and consider the following choices for the regularization parameters : $(\lambda, \varepsilon) = (0.2, 0.5) \times \lambda_{\max}$ or $(\lambda, \varepsilon) = (0.5, 0.2) \times \lambda_{\max}$. Each problem instance is solved with a budget of 2×10^6 FLOPs (the number of floating-point operations) for "Gaussian" and "DCT" dictionaries, and 2×10^7 FLOPs for "Uniform" and "Toeplitz" dictionaries. The difference in the FLOPs budgets stems from the bad conditioning of the "Uniform" and "Toeplitz" dictionaries which leads to slower convergence of standard numerical solvers.

As far as our simulation setups are concerned, we notice that safe relaxing enables us to significantly improve the performance. Safe relaxing alone (aPGr) proves to be of particular interest for dictionaries with highly-correlated atoms (*e.g.*, "Uniform" or "Toeplitz"). A careful study of our simulation results led us to the conclusion that this behavior is due to an improvement of the problem conditioning when moving from problem (2-$\mathcal{P}$) to (22a) and therefore of the convergence rate of the proximal gradient algorithm. The combination of screening and relaxing significantly outperforms all the other methods. We notice that S&R attains machine precision ($\tau = 10^{-16}$) for a large proportion of problem instances in most setups. This can be explained by the behavior emphasized in Sec. 5: when all the zero and non-zero elements of $\mathbf{x}^\star$ are identified, the minimizer can be explicitly computed from (22b)-(22c) with simple linear operations. Now, perfect identification of zero and non-zero elements of $\mathbf{x}^\star$ always occurs after a finite number of iterations when the GAP methodology is used to construct the safe sphere in tests (12) and (21) since $\mathbf{c}^{(t)} \rightarrow \mathbf{u}^\star$ and $r^{(t)} \rightarrow 0$ as $t \rightarrow \infty$.

7. CONCLUSION

In this paper, we proposed a new safe relaxing methodology to detect the position of non-zero components in the solution of the Elastic-Net problem. We showed how to leverage this knowledge to reduce the dimension of the optimization problem, enabling potential computational gains in the resolution. Numerical simulations show the interest of the method, especially when combined with safe screening.

References

[1] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer New York, 2013.

[2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996.

[3] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by Basis Pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.

[4] Hui Zou and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[5] Mário A. T. Figueiredo, Robert D. Nowak, and Stephen J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[6] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[7] Bangti Jin, Dirk A. Lorenz, and Stefan Schiffler, "Elastic-net regularization: error estimates and active set methods," vol. 25, no. 11, pp. 115022, 2009.

[8] Stephen Boyd, Neal Parikh, and Eric Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011.

[9] Laurent El Ghaoui, Vivian Viallon, and Tarek Rabbani, "Safe feature elimination for the lasso and sparse supervised learning problems," *Pacific Journal of Optimization*, vol. 8, no. 4, pp. 667–698, 2010.

[10] Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon, "Mind the duality gap: safer rules for the lasso," in *International Conference on Machine Learning*. PMLR, 2015, pp. 333–342.

[11] Zhen J. Xiang, Yun Wang, and Peter J. Ramadge, "Screening tests for lasso problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1008–1027, 2017.

[12] Jun Liu, Zheng Zhao, Jie Wang, and Jieping Ye, "Safe screening with variational inequalities and its application to lasso," in *ICML-14*. 2014, pp. 289–297, JMLR Workshop and Conference Proceedings.

[13] Jie Wang, Jiayu Zhou, Peter Wonka, and Jieping Ye, "Lasso screening rules via dual polytope projection," in *Advances in Neural Information Processing Systems*. 2013, vol. 26, Curran Associates, Inc.

[14] Cédric Herzet and Abed Malti, "Safe screening tests for LASSO based on firmly non-expansiveness," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4732–4736.

[15] Cédric Herzet, Clément Dorffer, and Angélique Drémeau, "Gather and conquer: Region-based strategies to accelerate safe screening tests," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3300–3315, 2019.

[16] Zhen James Xiang, Yun Wang, and Peter J Ramadge, "Screening tests for lasso problems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 1008–1027, 2016.

[17] Celestine Dünner, Simone Forte, Martin Takác, and Martin Jaggi, "Primal-dual rates and certificates," in *International Conference on Machine Learning*. PMLR, 2016, pp. 783–792.

[18] Dimitri P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[19] Liang Dai and Kristiaan Pelckmans, "An ellipsoid based, two-stage screening test for bpdn," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. Aug. 2012, pp. 654–658, IEEE.

[20] Jie Wang, Peter Wonka, and Jieping Ye, "Lasso screening rules via dual polytope projection," *Journal of Machine Learning Research*, 2015.

[21] Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon, "Gap safe screening rules for sparse-group lasso," in *Advances in neural information processing systems*, 2016, pp. 388–396.

[22] William W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.

[23] Elizabeth D. Dolan and Jorge J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical programming*, vol. 91, no. 2, pp. 201–213, 2002.

[24] Nasir U. Ahmed, Raj Natarajan, and Kamisetty R. Rao, "Discrete cosine transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[25] Robert M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2005.