Safe screening tests for lasso based on firmly non-expansiveness

Cédric Herzet, Abed Malti

▶ To cite this version:

HAL Id: hal-01260129
https://hal.inria.fr/hal-01260129

Submitted on 21 Jan 2016

SAFE SCREENING TESTS FOR LASSO BASED ON FIRMLY NON-EXPANSIVENESS

C. Herzet[*]

INRIA Rennes
Campus de Beaulieu
35000 Rennes, France

A. Malti

University of Tlemcen
Electrical Engineering Department
13000 Tlemcen, Algeria

ABSTRACT

This technical report focusses on safe screening techniques for the LASSO problem. We derive a new sphere test, coined RFNE, exploiting the firmly non-expansiveness of projection operators. Our test generalizes some methods of the literature but, unlike the latter, exploits approximated primal-dual solutions of the LASSO problem while remaining safe and effective. Our simulation results show that the proposed RFNE test outperforms the best methodology of the state of the art, namely the GAP test derived by Fercoq *et al.*

Index Terms— ℓ_1-norm minimization, LASSO, screening techniques.

1. INTRODUCTION

In this document, we focus on the (nonnegative) LASSO problem[1]:

$$\mathbf{x}_\lambda^\star \in \arg\min_{\mathbf{x} \geq \mathbf{0}} P_\lambda(\mathbf{y}, \mathbf{x}), \tag{1}$$

where $P_\lambda(\mathbf{y}, \mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$, and $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ with $\|\mathbf{a}_i\|_2 = 1$ for all i. During the last decade, ℓ_1-norm penalization has revealed to be a proper regularization scheme for many inverse problems (see [2]). Therefore, the study of computationally-efficient procedures solving (1) has become an active field of research [3–6].

Recently, an important step in that direction has been carried out in [7] by El Ghaoui *et al.*: the authors introduced a "screening" technique, allowing for a *safe*[2] identification of some of the zeros in $\mathbf{x}_\lambda^\star$; the latter components can thus be ignored in the LASSO computations leading to (potentially) huge memory and computational savings. Since the cornerstone paper by El Ghaoui *et al.*, many contributions addressing the screening problem have appeared in the literature, see [8–15].

---

[1]The standard LASSO problem can be reduced to a particular instance of (1), see [1, Section 2].

[2]The term "safe" refers to the fact that the elements identified by the screening method *always* correspond to zeros of $\mathbf{x}_\lambda^\star$.

In this paper, we propose a new screening method belonging to the family of "safe sphere tests". Our approach relies on properties of projection onto convex sets and generalizes some methods of the state of the art [13, 14]. In particular, unlike previous works, the proposed screening technique is ensured being safe even when the primal-dual solution exploited in the test is corrupted by some error[3]. We show that our methodology leads to superior performance than the best sphere test of the current state of the art (namely GAP [14]), at both a theoretical (in some specific setups, see Lemma 1) and an empirical level.

2. SOME CONVEX CONSIDERATIONS

In this section, we remind some elements of convex analysis which will serve as building blocks in the next sections.

Problem (1) is convex and has always (at least) one solution $\mathbf{x}_\lambda^\star$ [16]. The dual problem associated to (1) can be written as (see for example [12]):

$$\boldsymbol{\theta}_\lambda^\star = \arg\min_{\boldsymbol{\theta} \in \mathcal{D}} D_\lambda(\mathbf{y}, \boldsymbol{\theta}), \tag{2}$$

where

$$D_\lambda(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{2}\|\mathbf{y}\|_2^2 - \frac{1}{2}\|\mathbf{y} - \lambda\boldsymbol{\theta}\|_2^2,$$
$$\mathcal{D} = \left\{\boldsymbol{\theta} \in \mathbb{R}^m : \mathbf{a}_i^{\mathrm{T}}\boldsymbol{\theta} \leq 1, 1 \leq i \leq n\right\}.$$

Since $D_\lambda(\mathbf{y}, \boldsymbol{\theta})$ is a strictly convex function and $\mathcal{D}$ a closed set, there exists one unique $\boldsymbol{\theta}_\lambda^\star$ satisfying (2). We can also notice that (2) can be recast as a projection problem, $\boldsymbol{\theta}_\lambda^\star$ being the projection of $\mathbf{y}/\lambda$ onto $\mathcal{D}$. We will exploit this property in section 4 to state Theorem 1.

From the KKT conditions, it can be shown that any primal-dual optimal couple $(\mathbf{x}_\lambda^\star, \boldsymbol{\theta}_\lambda^\star)$ must satisfy the following optimality conditions [17, Prop. 5.1.5]:

$$\mathbf{x}_\lambda^\star \geq \mathbf{0}, \; \mathbf{a}_i^{\mathrm{T}}\boldsymbol{\theta}_\lambda^\star \leq 1 \text{ for all } i, \tag{3}$$
$$\left(\mathbf{a}_i^{\mathrm{T}}\boldsymbol{\theta}_\lambda^\star - 1\right)\mathbf{x}_\lambda^\star(i) = 0 \text{ for all } i, \tag{4}$$
$$\mathbf{y} = \lambda\boldsymbol{\theta}_\lambda^\star + \mathbf{A}\mathbf{x}_\lambda^\star, \tag{5}$$

---

[3]This error is related for instance to the finite numerical precision of numerical solvers.

where $\mathbf{x}_\lambda^\star(i)$ denotes the ith component of $\mathbf{x}_\lambda^\star$. From these conditions, it can be seen that if $\lambda \geq \lambda_{max} \triangleq \left\| \mathbf{A}^T \mathbf{y} \right\|_\infty$, then the only possible solution is $\mathbf{x}_\lambda^\star = \mathbf{0}$. In the sequel we will therefore restrict our attention to the case $\lambda < \lambda_{max}$.

3. SCREENING TESTS

It is well-known that the solutions of (1) are typically sparse. Optimality conditions (3)-(5) provide some information about the position of zeros in $\mathbf{x}_\lambda^\star$. Indeed, if $\langle \mathbf{a}_i, \boldsymbol{\theta}_\lambda^\star \rangle < 1$ then necessarily $\mathbf{x}_\lambda^\star(i) = 0$ from (4). Defining the *support* of $\mathbf{x} \geq \mathbf{0}$ and the *active set* associated to $\boldsymbol{\theta} \in \mathcal{D}$ as

$$\mathrm{supp}(\mathbf{x}) = \{i \in \{1, \dots, n\} : \mathbf{x}(i) > 0\},$$
$$\mathrm{acts}(\boldsymbol{\theta}) = \{i \in \{1, \dots, n\} : \mathbf{a}_i^T \boldsymbol{\theta} = 1\},$$

this can also be rewritten as

$$\mathrm{supp}(\mathbf{x}_\lambda^\star) \subseteq \mathrm{acts}(\boldsymbol{\theta}_\lambda^\star). \qquad (6)$$

Knowing $\boldsymbol{\theta}_\lambda^\star$ thus gives a valuable information about the support of $\mathbf{x}_\lambda^\star$. Of course, in practice $\boldsymbol{\theta}_\lambda^\star$ is unknown and solving (2) is unfortunately as difficult as solving (1).

Nevertheless, as first noted in [7], (4) can be of interest to identify the position of *some* of the zeros in $\mathbf{x}_\lambda^\star$: assume one can identify a region $\mathcal{R} \subset \mathbb{R}^m$ such that $\boldsymbol{\theta}_\lambda^\star \in \mathcal{R}$; then the following inequality trivially holds

$$\mathbf{a}_i^T \boldsymbol{\theta}_\lambda^\star \leq \max_{\boldsymbol{\theta} \in \mathcal{R}} \mathbf{a}_i^T \boldsymbol{\theta},$$

and from (4), we thus have

$$\max_{\boldsymbol{\theta} \in \mathcal{R}} \mathbf{a}_i^T \boldsymbol{\theta} < 1 \Rightarrow \mathbf{x}_\lambda^\star(i) = 0. \qquad (7)$$

This observation was the cornerstone of the "safe" screening test proposed in [7]. Since this seminal paper, many different tests, corresponding to different choices of $\mathcal{R}$ (sphere, dome, etc), have been proposed in the literature, see [9–15]. In this paper, we focus on the family of *sphere* tests because: *i)* they are conceptually very simple; *ii)* they have been shown to achieve state-of-the-art performance [15]. In such a case, $\mathcal{R} = S(\mathbf{q}, r) \triangleq \{\boldsymbol{\theta} \in \mathbb{R}^m : \left\| \boldsymbol{\theta} - \mathbf{q} \right\|_2 \leq r\}$ and the screening test (7) takes a very simple form:

$$\mathbf{a}_i^T \mathbf{q} + r < 1 \Rightarrow \mathbf{x}_\lambda^\star(i) = 0. \qquad (8)$$

The sphere tests proposed in the literature differ in the definition of the center $\mathbf{q}$ and the radius r, see [7,9,13–15]. We review the most competitive sphere tests of the literature in sections 4 and 5. We present some methodologies deriving from the properties of projection onto convex sets [13,14] in section 4. In section 5, we recall the main equations of the GAP test [15], which is a safe screening method currently achieving the best performance of the state of the art.

4. TESTS BASED ON PROJECTION PROPERTIES

A recent line of research has concentrated on the design of screening methods exploiting the knowledge of a dual optimal point $\boldsymbol{\theta}_{\lambda'}^\star$ with $\lambda' \neq \lambda$. Among the most effective screening methods relying on this approach, let us mention the sphere test based on "variational inequalities" presented in [13] and the methods exploiting the non-expansiveness property of projection operators onto convex sets in [14]. The inequalities exploited in these papers are summarized and extended in the following result:

Theorem 1. *Let* $\boldsymbol{\theta} \in \mathcal{D}$. *Then,*

$$\left\| \boldsymbol{\theta}_\lambda^\star - \left(\boldsymbol{\theta} + \frac{1}{2}\left(\frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right) \right) \right\|_2 \leq \frac{1}{2} \left\| \frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right\|_2, \quad (9)$$

for any $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta})$, *where*

$$\mathcal{Y}(\boldsymbol{\theta}) = \left\{ \mathbf{y}' : \mathbf{y}' = \lambda' \boldsymbol{\theta} + \sum_{i \in \mathrm{acts}(\boldsymbol{\theta})} \mathbf{a}_i \mathbf{x}(i), \ \mathbf{x} \geq \mathbf{0} \right\}. \quad (10)$$

This result is a direct consequence of the firmly non-expansiveness (FNE) property of projection operators (see [14]) and of the optimality conditions (3)-(5). A proof can be found in section 9. In the sequel we will refer to the screening methods deriving from Theorem 1 as *FNE sphere tests*.

As a consequence of Theorem 1, safe screening tests can be derived from any sphere $S(\mathbf{q}_{\mathrm{fne}}, r_{\mathrm{fne}})$ satisfying

$$\begin{aligned} \mathbf{q}_{\mathrm{fne}} &= \boldsymbol{\theta} + \tfrac{1}{2}\left(\tfrac{\mathbf{y}}{\lambda} - \tfrac{\mathbf{y}'}{\lambda'} \right), \\ r_{\mathrm{fne}} &= \tfrac{1}{2}\left\| \tfrac{\mathbf{y}}{\lambda} - \tfrac{\mathbf{y}'}{\lambda'} \right\|_2, \end{aligned} \qquad (11)$$

where $\boldsymbol{\theta} \in \mathcal{D}$ and $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta})$. In particular, the spheres proposed in [13, 14] are obtained as particular cases of (11): if $\boldsymbol{\theta} = \boldsymbol{\theta}_{\lambda'}^\star$ and $\mathbf{y}' = \lambda' \boldsymbol{\theta}_{\lambda'}^\star$ for some $\lambda' \neq \lambda$, (11) corresponds to the "variational" sphere proposed in [13, Section 2]; similarly, if $\boldsymbol{\theta} = \boldsymbol{\theta}_{\lambda'}^\star$ and $\mathbf{x} = \beta \mathbf{x}_{\lambda'}^\star$, for some $\beta > 0$, one recovers the EDDP sphere proposed in [14, Th. 15].

It can be seen from Theorem 1 that the choice of $\mathbf{y}'$ constitutes a degree of freedom in the construction of an FNE sphere test. In practice, a good rule of thumb consists in choosing $\mathbf{y}'$ close to $\mathbf{y}$: as seen from (11), the radius of the FNE sphere can be decreased by a proper choice of $\mathbf{y}'$. Since $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta})$ and $\mathbf{y} \in \mathcal{Y}(\boldsymbol{\theta}_\lambda^\star)$, one should therefore choose $\boldsymbol{\theta}$ such that $\mathcal{Y}(\boldsymbol{\theta})$ is close to $\mathcal{Y}(\boldsymbol{\theta}_\lambda^\star)$. In particular, if some $\boldsymbol{\theta}_{\lambda'}^\star$ with $\lambda \simeq \lambda'$ is available (as it occurs in sequential implementations of LASSO), setting $\boldsymbol{\theta} = \boldsymbol{\theta}_{\lambda'}^\star$ is certainly a sensible choice.

However, in practice, this approach faces an important bottleneck: only *finite*-precision approximations of $\boldsymbol{\theta}_{\lambda'}^\star$, say[4] $\hat{\boldsymbol{\theta}}_{\lambda'} \in \mathcal{D}$, are available in numerical solvers. Unfortunately, simply using $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\lambda'}$ in Theorem 1 (as implicitly done in

---

[4]Without loss of generality we assume that $\hat{\boldsymbol{\theta}}_{\lambda'} \in \mathcal{D}$ because this configuration can always be achieved by dual scaling, see [7, Section 3.4].

[13]) typically leads to poor results for the following reason: in most cases the cardinality of $\mathrm{acts}(\hat{\boldsymbol{\theta}}_{\lambda'})$ is equal to 1 and $\mathcal{Y}(\hat{\boldsymbol{\theta}}_{\lambda'})$ thus corresponds to a half-line in $\mathbb{R}^m$; on the other hand, the cardinality of $\mathrm{acts}(\boldsymbol{\theta}_\lambda^\star)$ is often much larger[5] than 1 and $\mathcal{Y}(\boldsymbol{\theta}_\lambda^\star)$ is a high-dimensional cone in $\mathbb{R}^m$; hence, we are likely to fail finding $\mathbf{y}' \in \mathcal{Y}(\hat{\boldsymbol{\theta}}_{\lambda'})$ close to $\mathbf{y} \in \mathcal{Y}(\boldsymbol{\theta}_\lambda^\star)$.

To deal with this issue, an approach (implicitly) considered in [14] consists in approximating $\mathrm{acts}(\hat{\boldsymbol{\theta}}_{\lambda'})$ by $\mathrm{supp}(\hat{\mathbf{x}}_{\lambda'})$ in the definition of $\mathcal{Y}(\hat{\boldsymbol{\theta}}_{\lambda'})$ (see (10)), where $\hat{\mathbf{x}}_{\lambda'}$ is a finite-precision approximation of $\mathbf{x}_{\lambda'}^\star$. Unfortunately, such approximation may lead to a violation of the hypotheses of Theorem 1 and this approach is then no longer ensured to lead to a *safe* screening test.

We propose a solution to this issue in section 6 by deriving a new safe sphere test relaxing the constraint of a perfect knowledge of $\boldsymbol{\theta}_{\lambda'}^\star$ while allowing for a fine tuning of $\mathbf{y}'$. In the next section, we recall the expressions of the GAP sphere test which constitutes one of the building blocks of the result presented in Theorem 2.

5. THE GAP SPHERE TEST

In [15], the authors used another route to derive a safe sphere test: they exploited the dual gap between the primal and the dual functions of the LASSO problem. More specifically, they showed that the following inequality holds for *any* primal-dual feasible points $(\mathbf{x}, \boldsymbol{\theta})$:

$$\|\boldsymbol{\theta}_\lambda^\star(\mathbf{y}) - \boldsymbol{\theta}\|_2 \leq \sqrt{2\lambda^{-2}(P_\lambda(\mathbf{y}, \mathbf{x}) - D_\lambda(\mathbf{y}, \boldsymbol{\theta}))}. \quad (12)$$

It immediately follows that the sphere $S(\mathbf{q}_\mathrm{gap}, r_\mathrm{gap})$ with

$$\begin{aligned} \mathbf{q}_\mathrm{gap} &= \boldsymbol{\theta}, \\ r_\mathrm{gap} &= \sqrt{2\lambda^{-2}(P_\lambda(\mathbf{y}, \mathbf{x}) - D_\lambda(\mathbf{y}, \boldsymbol{\theta}))}, \end{aligned} \quad (13)$$

and $\mathbf{x} \geq \mathbf{0}$, $\boldsymbol{\theta} \in \mathcal{D}$, defines a safe sphere test. The GAP sphere test has been shown to lead to very impressive results. In particular, the radius of the sphere r_gap tends to zero as $(\mathbf{x}, \boldsymbol{\theta})$ converges to $(\mathbf{x}_\lambda^\star, \boldsymbol{\theta}_\lambda^\star)$.

Because the GAP sphere (13) is safe for any primal-dual feasible point, it might be interesting to compare it with the FNE sphere test defined in (11) for $(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$, $\lambda' \neq \lambda$. The next lemma, whose proof is available in section 9, provides an answer to this question:

Lemma 1. *Assume one wants to solve* (1) *for a given* $\lambda > 0$ *and a primal-dual solution* $(\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$, $\lambda \neq \lambda' < \lambda_{max}$, *is available. Then, there exists* $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$ *such that the FNE sphere defined in* (11) *and the GAP sphere defined in* (13) *with* $(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$ *verify*

$$S(\mathbf{q}_\mathrm{fne}, r_\mathrm{fne}) \subset S(\mathbf{q}_\mathrm{gap}, r_\mathrm{gap}), \quad (14)$$

that is the FNE sphere is strictly included in the GAP sphere. In particular, (14) *holds when* $\mathbf{y}'$ *is a minimizer of* r_fne *over* $\mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$.

---

[5]From (6), we have that $\mathrm{card}\{\mathrm{acts}(\boldsymbol{\theta}_\lambda^\star)\} \geq \mathrm{card}\{\mathrm{supp}(\mathbf{x}_\lambda^\star)\}$.

The inclusion (14) implies that any zero identified by the GAP test will also be identified by some FNE test. Hence, as far as a primal-dual optimal point $(\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$ is available, Lemma 1 suggests that using FNE rather than GAP screening tests may be rewarding. We discuss the implementation of FNE tests based on the approximate knowledge of $(\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$ in the next section.

6. RELAXED FNE SPHERE TESTS

Motivated by Lemma 1, it may be tempting to derive FNE sphere tests by setting $\boldsymbol{\theta} = \boldsymbol{\theta}_{\lambda'}^\star$ with $\lambda \simeq \lambda'$. Unfortunately, as discussed at the end of section 4, one has to face to the problem of the finite precision of $\boldsymbol{\theta}_{\lambda'}^\star$ in practical numerical solvers.

In the next theorem, we provide a solution to this issue by deriving a relaxed version of the FNE sphere test:

Theorem 2. *Let* $\mathbf{x}$ *and* $\boldsymbol{\theta}$ *be primal and dual feasible points and let*

$$\mathbf{y}' = \lambda' \boldsymbol{\theta} + \mathbf{A}\mathbf{x}. \quad (15)$$

Then,

$$\|\boldsymbol{\theta}_\lambda^\star - \mathbf{q}_\mathrm{fne}\|_2 \leq r_\mathrm{fne} + r_\mathrm{slack}, \quad (16)$$

where $\mathbf{q}_\mathrm{fne}$ *and* r_fne *are defined in* (11) *and*

$$r_\mathrm{slack} = (\lambda')^{-1/2} \sqrt{\sum_i \mathbf{x}(i)(1 - \mathbf{a}_i^\mathsf{T}\boldsymbol{\theta})}$$

A proof of this result can be found in section 9. It essentially relies on the arguments developed in Theorem 1 and the use of the GAP inequality (12).

There are two main differences between Theorems 1 and 2. First, the definition of $\mathbf{y}'$ in Theorem 2 is a relaxed version of the one in Theorem 1. More specifically, unlike in Theorem 1, Theorem 2 does not require that the positions of the nonzero elements in $\mathbf{x}$ correspond to indices in $\mathrm{acts}(\boldsymbol{\theta})$. Second, the radius of the sphere defined in (16) is larger than the one defined in (9): the extra term in (16), r_slack, penalizes the violation of the complementary slackness condition (4). This term disappear as soon as $\mathrm{supp}(\mathbf{x}) \subseteq \mathrm{acts}(\boldsymbol{\theta})$. In particular, if $(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$, we have $r_\mathrm{slack} = 0$ and (16) corresponds to the standard FNE test defined in Theorem 1. In this sequel, we will refer to the screening methods deriving from Theorem 2 as *"robust FNE"* (RFNE) tests.

We next consider the RFNE sphere test obtained by particularizing Theorem 2 to $\mathbf{y}'$ defined as follows:

$$\mathbf{y}' = \underset{\tilde{\mathbf{y}} \in \hat{\mathcal{Y}}(\hat{\boldsymbol{\theta}}_\lambda)}{\arg\min} \, r_\mathrm{fne}, \quad (17)$$

with

$$\hat{\mathcal{Y}}(\hat{\boldsymbol{\theta}}_\lambda) = \left\{ \tilde{\mathbf{y}} \in \mathbb{R}^m : \lambda' \hat{\boldsymbol{\theta}}_\lambda + \sum_{i \in \mathrm{supp}(\hat{\mathbf{x}}_\lambda)} \mathbf{a}_i \mathbf{x}(i), \, \mathbf{x} \geq \mathbf{0} \right\},$$

and $(\hat{\mathbf{x}}_\lambda, \hat{\boldsymbol{\theta}}_\lambda)$ are some finite-precision approximation of $(\mathbf{x}_\lambda^\star, \boldsymbol{\theta}_\lambda^\star)$. In practice, $(\hat{\mathbf{x}}_\lambda, \hat{\boldsymbol{\theta}}_\lambda)$ can for example correspond to the output of the numerical solver at a given iteration or, in a sequential implementation, to (an approximation of) $(\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$ with $\lambda' > \lambda$. We assess the performance of such a sphere test in the next section.

7. SIMULATION RESULTS

In this section, we assess the performance of the screening test presented in section 6. We compare our approach to GAP sphere test which is currently the most competitive safe method of the state of the art.

Our simulation setup is as follows. The entries of $\mathbf{A} \in \mathbb{R}^{128 \times 512}$ are drawn from a uniform distribution on $[0, 1]$. The columns of $\mathbf{A}$ are then normalized such that $\|\mathbf{a}_i\|_2 = 1$ for all i. $\mathbf{y} = \mathbf{A}\tilde{\mathbf{x}}$ where $\tilde{\mathbf{x}}$ is a vector with 50 nonzero elements whose amplitude is drawn from a uniform distribution on $[0, 1]$. We consider a sequential implementation of LASSO where the solution of (1) has to be computed for $\lambda_t/\lambda_{max} = 0.95 \times 10^{-2t/[T-1]}$, $t \in \{1, \ldots, T-1\}$ and $T = 30$. For each λ_t, we search for a solution of (1) with a FISTA procedure [5] initialized with the solution obtained for λ_{t-1} (warm start). The solver is run until a duality gap $P(\mathbf{y}, \hat{\mathbf{x}}_{\lambda_t}) - D(\mathbf{y}, \hat{\boldsymbol{\theta}}_{\lambda_t})$ smaller than $\epsilon = 10^{-5}$ is reached. For each value of λ_t, we apply a dynamic screening strategy as suggested in [18, 19]: a screening test is carried out every 10 iterations by exploiting the output $(\hat{\mathbf{x}}_{\lambda_t}, \hat{\boldsymbol{\theta}}_{\lambda_t})$ of FISTA; the zeros which have been identified are removed from the problem.

Fig. 1 illustrates the proportion of zeros (with respect to the total number of zeros in $\mathbf{x}_{\lambda_t}^\star$) identified by the GAP and the proposed RFNE sphere test. The x-axis (resp. y-axis) corresponds to the logarithm of the ratio $(\lambda_t/\lambda_{max})^{-1}$ (resp. the number of iterations). One can notice that the proposed RFNE test allows for the identification of many more zeros of $\mathbf{x}_\lambda^\star$ than the GAP method at the very first iteration of the numerical solver. This can be understood as follows. The first screening test performed when considering problem (1) with λ_t is based on $(\hat{\mathbf{x}}_{\lambda_t}, \hat{\boldsymbol{\theta}}_{\lambda_t}) = (\hat{\mathbf{x}}_{\lambda_{t-1}}, \hat{\boldsymbol{\theta}}_{\lambda_{t-1}})$ where $(\hat{\mathbf{x}}_{\lambda_{t-1}}, \hat{\boldsymbol{\theta}}_{\lambda_{t-1}})$ is the final approximation of $(\mathbf{x}_{\lambda_{t-1}}^\star, \boldsymbol{\theta}_{\lambda_{t-1}}^\star)$ computed by FISTA; now, provided the FISTA solver has reached a fair convergence, the gap in the complementary slackness condition (and thus r_{slack} in (16)) must be very small. In this case, the RFNE test presented in section 6 is close to the FNE test in (11). We can thus expect the result in Lemma 1 to apply, that is the (R)FNE sphere is strictly included in the GAP sphere. This is indeed what we observe in terms of screening performance in Fig. 1. On the other hand, during the subsequent iterations, the two screening methods have similar performance. We have run simulations in different setups and with different dictionaries (not presented here for conciseness) and always observed this same behavior.
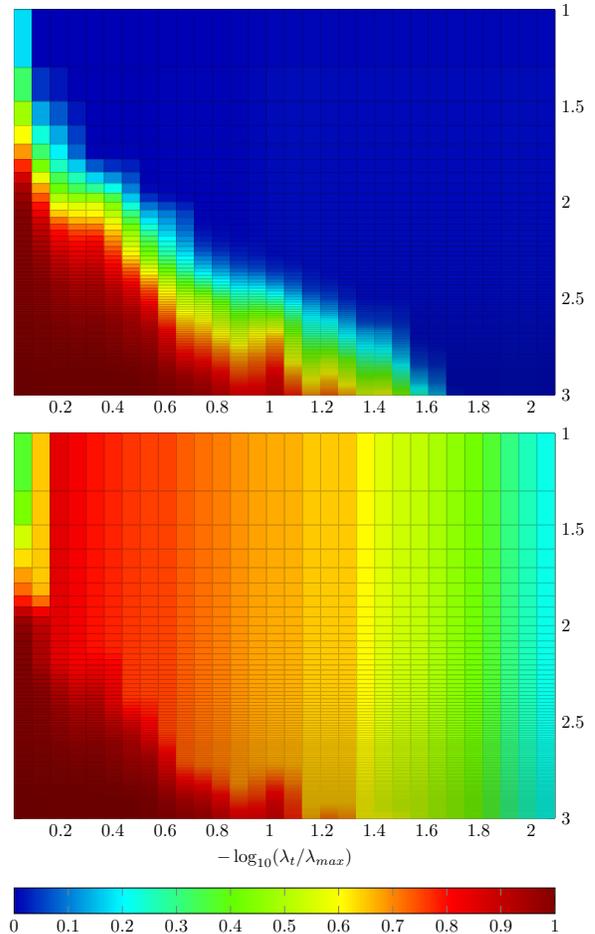

Fig. 1. Proportion of zeros identified by the screening procedures as a function of $-\log_{10}(\lambda_t/\lambda_{max})$ (x-axis) and the ($\log_{10}$ of the) number of iterations (y-axis): GAP (top) and safe FNE (bottom).

8. CONCLUSIONS

In this report, we proposed a new safe sphere screening test, namely RFNE. Our method generalizes state of the art methods and is ensured to be safe. It relies on estimated primal-dual solutions rather than exact ones which are unreachable in practice. We compared the proposed method to the best safe sphere test of the state of the art, namely GAP. We showed that RFNE performs at least as well as GAP in terms of zeros identification. Moreover, RFNE is able to identify far many more zeros at early iterations of the resolution process.

9. APPENDIX

Proof of Theorem 1: Let $\boldsymbol{\theta} \in \mathcal{D}$ and $\mathbf{y}'$ be defined as in (10) for some $\mathbf{x}$ such that

$$\begin{cases} \mathbf{x}(i) \geq 0 & i \in \text{acts}(\boldsymbol{\theta}), \\ \mathbf{x}(i) = 0 & \text{otherwise.} \end{cases} \tag{18}$$

Clearly, $\boldsymbol{\theta}$, $\mathbf{y}'$ and $\mathbf{x}$ satisfy the optimality conditions (3)-(5). As a consequence, $\boldsymbol{\theta}$ is a solution of the dual problem (2) with $\mathbf{y}'$ and λ' as input: $\boldsymbol{\theta}$ can thus also be seen as the projection of $\mathbf{y}'/\lambda'$ onto $\mathcal{D}$ (see section 2). The same remark applies for $\boldsymbol{\theta}_\lambda^\star$ and $\mathbf{y}/\lambda$. Hence, $\boldsymbol{\theta}_\lambda^\star$, $\mathbf{y}/\lambda$, $\boldsymbol{\theta}$ and $\mathbf{y}'/\lambda'$ must satisfy the firmly non-expansiveness property of projections onto convex sets, see [14, Th. 12]. More specifically, we have

$$\|\boldsymbol{\theta}_\lambda^\star - \boldsymbol{\theta}\|_2^2 + \left\| \left(\frac{\mathbf{y}}{\lambda} - \boldsymbol{\theta}_\lambda^\star \right) - \left(\frac{\mathbf{y}'}{\lambda'} - \boldsymbol{\theta} \right) \right\|_2^2 \leq \left\| \frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right\|_2^2. \tag{19}$$

Inequality (9) follows by rearranging the terms. $\qquad\square$

Proof of Lemma 1: Assume we search for the solution of the LASSO problem (1) for some $0 < \lambda < \lambda_{max}$. Suppose moreover that a primal-dual solution $(\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$ is known for some $\lambda' \neq \lambda$, $\lambda' < \lambda_{max}$. Consider the sphere $S(\mathbf{q}_{\mathrm{fne}}, r_{\mathrm{fne}})$ built from (11) by setting $\boldsymbol{\theta} = \boldsymbol{\theta}_{\lambda'}^\star$ and $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$. From inequality (19), it is easy to see that any $\boldsymbol{\theta}_\lambda^\star \in S(\mathbf{q}_{\mathrm{fne}}, r_{\mathrm{fne}})$ must also verify $\boldsymbol{\theta}_\lambda^\star \in S(\mathbf{q}_{\mathrm{ne}}, r_{\mathrm{ne}})$ where

$$\begin{aligned} \mathbf{q}_{\mathrm{ne}} &= \boldsymbol{\theta}_{\lambda'}^\star, \\ r_{\mathrm{ne}} &= \left\| \frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right\|_2. \end{aligned} \tag{20}$$

Hence, $S(\mathbf{q}_{\mathrm{fne}}, r_{\mathrm{fne}}) \subseteq S(\mathbf{q}_{\mathrm{ne}}, r_{\mathrm{ne}})$. To prove the statement of Lemma 1, it is therefore sufficient to show that

$$S(\mathbf{q}_{\mathrm{ne}}, r_{\mathrm{ne}}) \subset S(\mathbf{q}_{\mathrm{gap}}, r_{\mathrm{gap}}), \tag{21}$$

for some $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$. Since $\mathbf{q}_{\mathrm{ne}} = \mathbf{q}_{\mathrm{gap}} = \boldsymbol{\theta}_{\lambda'}^\star$, this is equivalent to proving that

$$r_{\mathrm{ne}} < r_{\mathrm{gap}} \text{ for some } \mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star). \tag{22}$$

First, particularizing the definition of r_{gap} in (13) to $(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{x}_{\lambda'}^\star, \boldsymbol{\theta}_{\lambda'}^\star)$, and taking optimality conditions (3)-(5) into account, we find:

$$r_{\mathrm{gap}} = \left| 1 - \frac{\lambda'}{\lambda} \right| \|\boldsymbol{\theta}_{\lambda'}^\star\|_2. \tag{23}$$

Moreover, evaluating r_{ne} in (20) for $\mathbf{y}' = \lambda' \boldsymbol{\theta}_{\lambda'}^\star + \frac{\lambda'}{\lambda} \mathbf{A} \mathbf{x}_{\lambda'}^\star \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$, we obtain:

$$r_{\mathrm{ne}} = \left| 1 - \frac{\lambda'}{\lambda} \right| \|\boldsymbol{\theta}_{\lambda'}^\star\|_2. \tag{24}$$

Hence,

$$\min_{\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)} r_{\mathrm{ne}} \leq r_{\mathrm{gap}}. \tag{25}$$

The above inequality is actually strict because of the following arguments. Let us first exploit the definition of $\mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$ to rewrite r_{ne} as a function of $\mathbf{x}$:

$$r_{\mathrm{ne}} = \left\| \frac{\mathbf{y}}{\lambda} - \lambda' \boldsymbol{\theta}_{\lambda'}^\star - \lambda^{-1} \sum_{i \in \mathrm{acts}(\boldsymbol{\theta}_{\lambda'}^\star)} \mathbf{a}_i \mathbf{x}(i) \right\|_2. \tag{26}$$

Clearly, optimizing r_{ne} over $\mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$ in (25) is equivalent to optimizing the right-hand side of (26) over $\mathbf{x} \geq \mathbf{0}$. The particular value of r_{ne} obtained in (24) corresponds to the choice $\mathbf{x} = \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star$. We next show that $\mathbf{x} = \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star$ is not a minimizer of r_{ne} because optimality conditions cannot be satisfied.

Let $\mathbf{x}_{\mathrm{ne}}^\star$ be a minimizer of r_{ne}; then it must satisfy [17, Prop. 2.1.2]:

$$\sum_i \left(\frac{\partial r_{\mathrm{ne}}}{\partial \mathbf{x}(i)} \right)_{\mathbf{x} = \mathbf{x}_{\mathrm{ne}}^\star} (\mathbf{x}(i) - \mathbf{x}_{\mathrm{ne}}^\star(i)) \geq 0 \text{ for all } \mathbf{x} \geq \mathbf{0} \tag{27}$$

First note that $\mathrm{supp}(\mathbf{x}_{\lambda'}^\star) \neq \emptyset$ (and thus $\mathrm{acts}(\boldsymbol{\theta}_{\lambda'}^\star) \neq \emptyset$) because we assume $\lambda' < \lambda_{max}$. Now we have for any $i \in \mathrm{acts}(\boldsymbol{\theta}_{\lambda'}^\star)$:

$$\begin{aligned} \left(\frac{\partial r_{\mathrm{ne}}}{\partial \mathbf{x}(i)} \right)_{\mathbf{x} = \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star} &= -2 \mathbf{a}_i^{\mathrm{T}} \left(\frac{\mathbf{y}}{\lambda} - \boldsymbol{\theta}_{\lambda'}^\star - \frac{\mathbf{A} \mathbf{x}_{\lambda'}^\star}{\lambda} \right), \\ &= 2 \left(1 - \frac{\lambda'}{\lambda} \right) \mathbf{a}_i^{\mathrm{T}} \boldsymbol{\theta}_{\lambda'}^\star, \\ &= 2 \left(1 - \frac{\lambda'}{\lambda} \right), \end{aligned} \tag{28}$$

where the second equality follows from the fact that $\mathbf{y} = \lambda' \boldsymbol{\theta}_{\lambda'}^\star + \mathbf{A} \mathbf{x}_{\lambda'}^\star$ and the last one from $\mathbf{a}_i^{\mathrm{T}} \boldsymbol{\theta}_{\lambda'}^\star = 1$ for all $i \in \mathrm{acts}(\boldsymbol{\theta}_{\lambda'}^\star)$. If $\lambda' < \lambda$, then $(1 - \lambda'/\lambda) > 0$ and (27) cannot be satisfied with $\mathbf{x}_{\mathrm{ne}}^\star = \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star$ because setting $\mathbf{x}$ as

$$\begin{aligned} \mathbf{x}(i) &< \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star(i), \quad i \in \mathrm{supp}(\mathbf{x}_{\lambda'}^\star), \\ \mathbf{x}(i) &= 0, \qquad\qquad \text{otherwise.} \end{aligned} \tag{29}$$

leads to a violation of the inequality in (27). Similarly, if $\lambda' > \lambda$, $(1 - \lambda'/\lambda) < 0$ and a violation of (27) with $\mathbf{x}_{\mathrm{ne}}^\star = \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star$ occurs by setting

$$\begin{aligned} \mathbf{x}(i) &> \frac{\lambda'}{\lambda} \mathbf{x}_{\lambda'}^\star(i), \quad i \in \mathrm{supp}(\mathbf{x}_{\lambda'}^\star), \\ \mathbf{x}(i) &= 0, \qquad\qquad \text{otherwise.} \end{aligned} \tag{30}$$

This proves that $r_{\mathrm{ne}} < r_{\mathrm{gap}}$ for some $\mathbf{y}' \in \mathcal{Y}(\boldsymbol{\theta}_{\lambda'}^\star)$ and thus $S(\mathbf{q}_{\mathrm{fne}}, \mathbf{r}_{\mathrm{fne}}) \subset S(\mathbf{q}_{\mathrm{gap}}, \mathbf{r}_{\mathrm{gap}})$.

Finally, the last assertion of Lemma 1 directly follows from the fact that, by definition, $r_{\mathrm{fne}} = r_{\mathrm{ne}}/2$.

$\qquad\square$

Proof of Theorem 2: Let $\mathbf{y}'$ be defined as in (15) and let $\boldsymbol{\theta}_{\lambda'}^\star(\mathbf{y}')$ denotes the projection of $\mathbf{y}'/\lambda'$ onto $\mathcal{D}$. Applying Theorem 1, we have that (9) holds (with the substitution $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_{\lambda'}^\star(\mathbf{y}')$). Using a triangle inequality leads to

$$\begin{aligned} & \left\| \boldsymbol{\theta}_\lambda^\star - \left(\boldsymbol{\theta} + \frac{1}{2} \left(\frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right) \right) \right\|_2 \\ & \qquad \leq \frac{1}{2} \left\| \frac{\mathbf{y}}{\lambda} - \frac{\mathbf{y}'}{\lambda'} \right\|_2 + \|\boldsymbol{\theta}_{\lambda'}^\star(\mathbf{y}') - \boldsymbol{\theta}\|_2. \end{aligned}$$

Since $\boldsymbol{\theta}_{\lambda'}^\star(\mathbf{y}') \in \mathcal{D}$, we can use Ferocq's inequality (12) to bound the second term in the right-hand of this inequality.

More specifically, let $\mathbf{x} \geq \mathbf{0}$, $\boldsymbol{\theta} \in \mathcal{D}$ and $\mathbf{y}' = \lambda'\boldsymbol{\theta} + \mathbf{A}\mathbf{x}$ as in (15), we find after some algebraic manipulations that

$$\|\boldsymbol{\theta}^\star_{\lambda'}(\mathbf{y}') - \boldsymbol{\theta}\|_2 \leq (\lambda')^{-1/2} \sqrt{\sum_i \mathbf{x}(i)\big(1 - \mathbf{a}_i^{\mathrm{T}}\boldsymbol{\theta}\big)}. \quad (31)$$

$\square$

10. REFERENCES

[1] Mario A. T. Figueiredo, "Teaching a new trick to an old dog: Revisiting the quadratic programming formulation of sparse recovery using admm," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. May 2014, pp. 1512–1516, IEEE.

[2] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing.*, Applied and Numerical Harmonic Analysis. Birkhaüser, 2013.

[3] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA Journal of Numerical Analysis*, vol. 20, no. 3, pp. 389–403, July 2000.

[4] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–451, 2004.

[5] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[6] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2345–2356, Sept. 2010.

[7] Laurent El Ghaoui, Vivian Viallon, and Tarek Rabbani, "Safe feature elimination in sparse supervised learning," Tech. Rep. UC/EECS-2010-126, EECS Dept., University of California at Berkeley, Sept. 2010.

[8] Yun Wang, Zhen J. Xiang, and Peter J. Ramadge, "Lasso screening with a small regularization parameter," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. May 2013, pp. 3342–3346, IEEE.

[9] Zhen J. Xiang, Hao Xu, and Peter J. Ramadge, "Learning sparse representations of high dimensional data on large scale dictionaries," in *Advances in Neural Information Processing Systems 24*, J. Shawe-taylor, Zemel, P. Bartlett, Pereira, and Weinberger, Eds., pp. 900–908. 2011.

[10] Zhen J. Xiang and Peter J. Ramadge, "Fast lasso screening tests based on correlations," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. Mar. 2012, pp. 2137–2140, IEEE.

[11] Hao Wu and Peter J. Ramadge, "The 2-codeword screening test for lasso problems," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. May 2013, pp. 3307–3311, IEEE.

[12] Zhen J. Xiang, Yun Wang, and Peter J. Ramadge, "Screening tests for lasso problems," arXiv:1405.4897v1, May 2014.

[13] Jun Liu, Zheng Zhao, Jie Wang, and Jieping Ye, "Safe screening with variational inequalities and its application to lasso," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, Tony Jebara and Eric P. Xing, Eds. 2014, pp. 289–297, JMLR Workshop and Conference Proceedings.

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

[15] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap: safer rules for the lasso," in *ICML*, 2015.

[16] Ryan J. Tibshirani and Others, "The lasso problem and uniqueness," *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.

[17] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, USA, 2003.

[18] Antoine Bonnefoy, Valentin Emiya, Liva Ralaivola, and Remi Gribonval, "A dynamic screening principle for the lasso," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. 2014, pp. 6–10, IEEE.

[19] Antoine Bonnefoy, Valentin Emiya, Liva Ralaivola, and Remi Gribonval, "Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso," *Signal Processing, IEEE Transactions on*, vol. 63, no. 19, pp. 5121–5132, Oct. 2015.