Region-free Safe Screening Tests for ℓ_1-penalized Convex Problems

Cédric Herzet
SIMSMART team
INRIA Rennes - Bretagne Atlantique
Rennes, France
cedric.herzet@inria.fr

Clément Elvira
IETR UMR CNRS 6164,
CentraleSupelec Rennes Campus
35576 Cesson Sévigné, France
clement.elvira@centralesupelec.fr

Hong-Phuong Dang
LaTIM, INSERM-UMR1101
Univ. de Bretagne Occidentale
Brest, France
dang@univ-brest.fr

***Abstract*—We address the problem of safe screening for ℓ_1-penalized convex regression/classification problems, *i.e.*, the identification of zero coordinates of the solutions. Unlike previous contributions of the literature, we propose a screening methodology which does not require the knowledge of a so-called "safe region". Our approach does not rely on any other assumption than convexity (in particular, no strong-convexity hypothesis is needed) and therefore applies to a wide family of convex problems. When the Fenchel conjugate of the data-fidelity term is strongly convex, we show that the popular "GAP sphere test" proposed by Fercoq *et al.* can be recovered as a particular case of our methodology (up to a minor modification). We illustrate numerically the performance of our procedure on the "sparse support vector machine classification" problem.**

***Index Terms*—sparsity, convex problem, safe screening.**

I. Introduction

In the last decades, solving optimization problems promoting sparsity has become a standard task in signal processing, machine learning or statistics. A common formulation of these problems reads as follows

$$(\mathbf{x}^\star, x^\star) \in \operatorname*{arg\,min}_{(\mathbf{x},x)\in\mathbb{R}^{n+1}} f(\mathbf{A}\mathbf{x} + \mathbf{b}x + \mathbf{c}) + \lambda\|\mathbf{x}\|_1 \tag{1}$$

where $f\colon \mathbb{R}^m \to (-\infty, +\infty]$ is a closed, convex, proper function and $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^m$, $\lambda > 0$ are some problem-specific parameters. Particular instances of (1) include (among many others) ℓ_2 [1] or Kullback-Leibler [2] sparse regression, logistic [3] and sparse support vector machine classification [4], etc.

The practical relevance of (1) has given birth to many numerical procedures to efficiently solve it, see *e.g.*, [5, 6]. Of particular interest in this paper is an acceleration method first proposed by El Ghaoui *et al.* in [7] and known as *"safe screening"*. The objective of safe screening is the identification of zero coordinates of $\mathbf{x}^\star$ via simple tests; the variables corresponding to zeros can then be removed from the optimization problem and thus potentially lead to huge computational and memory savings. Since the seminal work by El Ghaoui *et al.*, the effectiveness of safe screening has been acknowledged by many authors in different setups, see *e.g.*, [8–12].

In the current state of the art, all the contributions dealing with safe screening for convex problems leverage the concept of *"safe region"*, *i.e.*, a region of the dual space containing the solution of the dual problem of (1). It is well-known that the effectiveness of these methods improves as the size of the safe region decreases. Loosely speaking, "smaller" regions lead to screening procedures able to identify more zeros. Many authors have therefore addressed the problem of designing "good" safe regions, *i.e.*, small-volume regions computable at low cost, see *e.g.*, [13–18].

The current state-of-the-art methodology in this respect is the *"GAP sphere"* first proposed by Fercoq *et al.* [19] for LASSO and later on generalized by Ndiaye *et al.* in [20]. The construction of a GAP sphere relies on the identification of some primal-dual feasible couple $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}}) \in \mathbb{R}^{n+1} \times \mathbb{R}^m$ –$\bar{\mathbf{u}}$ denotes the dual feasible point– and owns its popularity to the fact that the radius of the sphere is proportional to the squared-root of the duality gap evaluated at $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$. In particular, when the iterates of a solving procedure for (1) are used to construct the GAP sphere, its radius converges to zero: all the zero components of $\mathbf{x}^\star$ can then be asymptotically identified by the screening test (under mild regularity conditions).

Despite this desirable feature, the construction of GAP spheres as suggested by Ndiaye *et al.* only applies to problems where the Fenchel conjugate of f is strongly convex, see [20, Theorem 6]. This precludes the use of GAP safe regions in popular problems where f corresponds for example to a Kullback-Leiber divergence or a hinge function. This restriction has been partially relaxed in [21, 22] where the construction of the GAP sphere is shown to be possible when the Fenchel conjugate of f is strongly-convex over (sufficiently simple) subsets of the dual space.

In this paper, we introduce a new methodology to build screening tests for convex problem (1). Our method is applicable to *any convex* function f (in particular no strong-convexity assumption is needed) and does not require the knowledge of any safe region. In the case where the Fenchel conjugate of f is strongly convex, we show that a quadratic relaxation of our procedure leads to the same test (up to a minor variation) as the standard GAP sphere region. We illustrate numerically the relevance of our proposed approach in the case where f is a hinge function and the Fenchel conjugate of f is therefore simply convex but not even strictly so.

II. Problem and optimality conditions

Without loss of generality[1] we consider the nonnegative version of (1), that is

$$p^\star = \min_{(\mathbf{x},x)\in\mathbb{R}^n_+\times\mathbb{R}} f(\mathbf{A}\mathbf{x}+\mathbf{b}x+\mathbf{c})+\lambda\mathbf{1}_n^{\mathrm{T}}\mathbf{x}. \tag{2}$$

Since f is assumed to be convex, closed and proper, the dual problem of (2) can be written as [7, Section 3.1]:

$$d^\star = \max_{\mathbf{u}\in\mathbb{R}^m} -\mathbf{c}^{\mathrm{T}}\mathbf{u} - f^*(-\mathbf{u}) \quad \text{s.t.} \begin{cases} \mathbf{a}_i^{\mathrm{T}}\mathbf{u} \le \lambda \ \forall i \\ \mathbf{b}^{\mathrm{T}}\mathbf{u} = 0 \end{cases} \tag{3}$$

where f^* denotes the Fenchel conjugate of f and $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. Problems (2)-(3) can also be expressed more compactly as

$$p^\star = \min_{(\mathbf{x},x)\in\mathbb{R}^{n+1}} p(\mathbf{x},x), \quad d^\star = \max_{\mathbf{u}\in\mathbb{R}^m} d(\mathbf{u}), \tag{4}$$

by using the following definitions:

$$p(\mathbf{x},x) \triangleq f(\mathbf{A}\mathbf{x}+\mathbf{b}x+\mathbf{c}) + \lambda\mathbf{1}_n^{\mathrm{T}}\mathbf{x} + \mathbb{I}\{\mathbf{x}\ge\mathbf{0}_n\}$$
$$d(\mathbf{u}) \triangleq -\mathbf{c}^{\mathrm{T}}\mathbf{u} - f^*(-\mathbf{u}) - \mathbb{I}\{\mathbf{A}^{\mathrm{T}}\mathbf{u}\le\lambda\} - \mathbb{I}\{\mathbf{b}^{\mathrm{T}}\mathbf{u}=0\}$$

where $\mathbb{I}\{\cdot\}$ denotes the "indicator" function which is equal to 0 when the statement in the braces is true and to $+\infty$ otherwise.

In the rest of this paper, we suppose that (2) (resp. (3)) admits a nonempty set of minimizers $\mathcal{X}^\star$ (resp. maximizers $\mathcal{U}^\star$). We assume moreover that strong duality holds, *i.e.*,

$$\mathrm{gap}(\mathbf{x},x,\mathbf{u}) \triangleq p(\mathbf{x},x) - d(\mathbf{u}) \ge 0 \tag{5}$$

with equality if and only if $(\mathbf{x},x)\in\mathcal{X}^\star$ and $\mathbf{u}\in\mathcal{U}^\star$.

Under strong duality assumption, standard primal-dual optimality conditions write as follows [22, Theorem 1]: $(\mathbf{x}^\star,x^\star)\in\mathcal{X}^\star$ and $\mathbf{u}^\star\in\mathcal{U}^\star$ if and only if

$$\mathbf{b}^{\mathrm{T}}\mathbf{u}^\star = 0,\ \mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star\le\lambda,\ x_i^\star\ge 0\ \forall i \tag{6}$$
$$(\mathbf{a}_i^{\mathrm{T}}\mathbf{u}^\star-\lambda)\,x_i^\star = 0 \quad \forall i \tag{7}$$
$$\mathbf{u}^\star \in -\partial f(\mathbf{A}\mathbf{x}^\star+\mathbf{b}x^\star+\mathbf{c}) \tag{8}$$

where ∂ denotes the subdifferential operator [23, Chapter 3], and $x_i^\star$ the ith component of $\mathbf{x}^\star$.

Safe screening procedures essentially leverage optimality condition (7). More specifically, if $(\mathbf{x}^\star,x^\star,\mathbf{u}^\star)$ is a primal-dual optimal couple we have $\forall\ell\in\{1,\dots,n\}$:

$$\mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}^\star < \lambda \implies x_\ell^\star = 0. \tag{9}$$

If f is moreover differentiable (that is $\partial f = \{\nabla f\}$), we have from (8) that the latter condition can be written as

$$-\mathbf{a}_\ell^{\mathrm{T}}\nabla f(\mathbf{A}\mathbf{x}^\star+\mathbf{b}x^\star+\mathbf{c}) < \lambda \implies x_\ell^\star = 0. \tag{10}$$

We will see hereafter that the so-called "GAP sphere test" and a quadratic relaxation of the proposed approach can be seen as generalizations of (9)-(10), valid for any primal-dual feasible couple $(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}})$.

[1] Problem (2) can be seen as a generalization of (1) since the latter can always be rewritten as a particular case of the former [9, Section 2].

III. Region-based screening

Standard screening tests, as originally proposed in [7], rely on the concept of *"safe region"*. Let $(\mathbf{x}^\star,x^\star,\mathbf{u}^\star)$ be a primal-dual optimal solution. We say that a region $\mathcal{R}\subset\mathbb{R}^m$ is safe for $\mathbf{u}^\star$ if $\mathbf{u}^\star\in\mathcal{R}$. In this case, optimality condition (9) can be relaxed as

$$\max_{\mathbf{u}\in\mathcal{R}} \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u} < \lambda \implies x_\ell^\star = 0. \tag{11}$$

The left-hand side of (11) provides a practical way to identify zero coordinates of $\mathbf{x}^\star$ provided that the evaluation of the maximum over $\mathcal{R}$ is tractable. This requirement is usually achieved by constructing safe regions with favorable geometries (*e.g.*, spheres or domes), see [9].

When sphere regions are considered and f^* is μ-strongly convex, one of the most effective screening procedure of the literature is the so-called "GAP sphere test" [19, 20]. Letting $(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}})$ denote a primal-dual feasible couple, this test reads as follows:

$$\mathbf{a}_\ell^{\mathrm{T}}\bar{\mathbf{u}} < \lambda - \sqrt{\tfrac{2}{\mu}\mathrm{gap}(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}})} \implies x_\ell^\star = 0. \tag{12}$$

It can be regarded as a generalization of (9) which holds for any primal-dual feasible couple $(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}})$. In particular, if $(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}})$ is primal-dual optimal then $\mathrm{gap}(\bar{\mathbf{x}},\bar{x},\bar{\mathbf{u}}) = 0$ and one recovers optimality condition (9).

IV. Region-free screening

In this section, we present our "region-free" safe screening method. The rationale and the main expressions of the proposed approach are introduced in Section IV-A. We emphasize that no strong-convexity assumption is made in our derivations. In the case where f^* is μ-strongly convex, we show in Section IV-B that (a slight variant of) GAP test (12) can be recovered as a particular case of our method.

A. Rationale and main expressions

Consider the following problem:

$$d_\ell^\star = \max_{\mathbf{u}\in\mathbb{R}^m} -\mathbf{c}^{\mathrm{T}}\mathbf{u} - f^*(-\mathbf{u}) \quad \text{s.t.} \begin{cases} \mathbf{a}_i^{\mathrm{T}}\mathbf{u}\le\lambda \ \ i\ne\ell \\ \mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}=\lambda \\ \mathbf{b}^{\mathrm{T}}\mathbf{u}=0. \end{cases} \tag{13}$$

We note that (13) corresponds to dual problem (3) where the constraint "$\mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}\le\lambda$" has been strengthened to "$\mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}=\lambda$". Since the feasible set of (13) is contained in the feasible set of (3), we necessarily have:

$$d^\star \ge d_\ell^\star \tag{14}$$

with equality if and only if there exists $\mathbf{u}^\star\in\mathcal{U}^\star$ verifying $\mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}^\star = \lambda$. As a consequence, if $d^\star > d_\ell^\star$ there is no maximizer of (3) such that $\mathbf{a}_\ell^{\mathrm{T}}\mathbf{u}^\star=\lambda$. Using (9), this in turn implies that $x_\ell^\star = 0$ for all minimizers of (2). Formally, this writes:

$$d^\star > d_\ell^\star \implies \forall(\mathbf{x}^\star,x^\star)\in\mathcal{X}^\star:\ x_\ell^\star = 0. \tag{15}$$

This implication is the basis of our region-free screening test, which is encapsulated in the following theorem:

Theorem 1. *If* $\bar{\mathbf{u}} \in \mathbb{R}^m$ *is dual feasible then*

$$d(\bar{\mathbf{u}}) > p_\ell(\mathbf{x}, x) \implies \forall (\mathbf{x}^\star, x^\star) \in \mathcal{X}^\star : x_\ell^\star = 0 \tag{16}$$

where

$$p_\ell(\mathbf{x}, x) \triangleq f(\mathbf{A}\mathbf{x} + \mathbf{b}x + \mathbf{c}) + \lambda \mathbf{1}_n^{\mathrm{T}} \mathbf{x} + \sum_{i \neq \ell} \mathbb{I}\{x_i \geq 0\}.$$

Proof: In view of (15) it is sufficient to show that $d(\bar{\mathbf{u}})$ is a lower bound on $d^\star$ and $p_\ell(\mathbf{x}, x)$ an upper bound on $d_\ell^\star$. Obviously, we have $d^\star \geq d(\bar{\mathbf{u}})$ since $\bar{\mathbf{u}}$ is dual feasible. Moreover, straightforward calculations show that (13) is the dual problem of "$\min_{(\mathbf{x},x)\in\mathbb{R}^{n+1}} p_\ell(\mathbf{x}, x)$". We thus have $p_\ell(\mathbf{x}, x) \geq d_\ell^\star$ by weak duality. □

We note that the implementation of test (16) does not require the knowledge of any safe region but simply involves the comparison between the values of a relaxed primal function $p_\ell(\mathbf{x}, x)$ and the dual function at some feasible point $\bar{\mathbf{u}}$. The choice of the primal-dual couple $(\mathbf{x}, x, \bar{\mathbf{u}})$ is left as a degree of freedom to the practitioner. Obviously, one should try to optimize the effectiveness of the test by minimizing p_ℓ over $(\mathbf{x}, x)$ (resp. maximizing d over $\bar{\mathbf{u}}$) while keeping the computational cost of the test reasonable. In this paper, we consider the following option: given a primal-dual *feasible* couple $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$, we optimize the value of p_ℓ with respect to its ℓth component. Test (16) then particularizes as

$$d(\bar{\mathbf{u}}) > \min_{x_\ell} \bar{p}_\ell(x_\ell) \implies \forall (\mathbf{x}^\star, x^\star) \in \mathcal{X}^\star : x_\ell^\star = 0, \tag{17}$$

where $\bar{p}_\ell$ is a function defined for all scalars x_ℓ as

$$\bar{p}_\ell(x_\ell) = p_\ell(\bar{\mathbf{x}} + (x_\ell - \bar{x}_\ell)\mathbf{e}_\ell, \bar{x}) \tag{18}$$

and $\mathbf{e}_\ell$ refers to the ℓth vector of the canonical basis of $\mathbb{R}^n$. In the next subsection, we draw a connection between (17) and standard GAP sphere test (12). More specifically, we show that the latter can be obtained from a quadratic relaxation of the former when f^* is strongly convex.

B. Connection with GAP

Suppose that $\bar{p}_\ell(x_\ell)$ admits the following upper bound:

$$\forall x_\ell : \ \bar{p}_\ell(x_\ell) \leq \bar{p}_\ell(\bar{x}_\ell) + \alpha(x_\ell - \bar{x}_\ell) + \tfrac{\beta}{2}(x_\ell - \bar{x}_\ell)^2 \tag{19}$$

for some parameters $\alpha \in \mathbb{R}$, $\beta > 0$. Minimizing both sides of this inequality with respect to x_ℓ leads to

$$\min_{x_\ell} \bar{p}_\ell(x_\ell) \leq p(\bar{\mathbf{x}}, \bar{x}) - \frac{\alpha^2}{2\beta}, \tag{20}$$

where we used the fact that $\bar{p}_\ell(\bar{x}_\ell) = p_\ell(\bar{\mathbf{x}}, \bar{x}) = p(\bar{\mathbf{x}}, \bar{x})$ since $(\bar{\mathbf{x}}, \bar{x})$ is primal feasible. Plugging (20) into (17), we obtain after a few algebraic manipulations:

$$\sqrt{2\beta \, \mathrm{gap}(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})} < |\alpha| \Rightarrow \forall (\mathbf{x}^\star, x^\star) \in \mathcal{X}^\star : x_\ell^\star = 0. \tag{21}$$

Let us now particularize (21) to the case where f^* is μ-strongly convex. The function $f : \mathbb{R}^m \to (-\infty, \infty]$ is then necessarily μ^{-1}-smooth differentiable [23, Theorem 5.26] and admits a quadratic upper bound at any point $\bar{\mathbf{z}} \in \mathbb{R}^m$ [23, Lemma 5.7]:

$$f(\mathbf{z}) \leq f(\bar{\mathbf{z}}) + \nabla^{\mathrm{T}} f(\bar{\mathbf{z}})(\mathbf{z} - \bar{\mathbf{z}}) + \tfrac{1}{2\mu}\|\mathbf{z} - \bar{\mathbf{z}}\|_2^2. \tag{22}$$

Using (22) with $\bar{\mathbf{z}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b}\bar{x} + \mathbf{c}$ and $\mathbf{z} = \bar{\mathbf{z}} + (x_\ell - \bar{x}_\ell)\mathbf{a}_\ell$, it is easy to see that (19) holds with

$$\alpha = \mathbf{a}_\ell^{\mathrm{T}} \nabla f(\mathbf{A}\bar{\mathbf{x}} + \mathbf{b}\bar{x} + \mathbf{c}) + \lambda, \quad \beta = \mu^{-1}. \tag{23}$$

Letting $\mathbf{u}_{\bar{\mathbf{x}}} \triangleq -\nabla f(\mathbf{A}\bar{\mathbf{x}} + \mathbf{b}\bar{x} + \mathbf{c})$, the left-hand side of (21) then takes the following form:

$$\sqrt{\tfrac{2}{\mu} \mathrm{gap}(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})} < |\lambda - \mathbf{a}_\ell^{\mathrm{T}} \mathbf{u}_{\bar{\mathbf{x}}}|. \tag{24}$$

We note that this test is very similar to (12). More specifically, if $\mathbf{u}_{\bar{\mathbf{x}}}$ is dual feasible (thus $\mathbf{a}_\ell^{\mathrm{T}} \mathbf{u}_{\bar{\mathbf{x}}} \leq \lambda$), (24) leads to

$$\mathbf{a}_\ell^{\mathrm{T}} \mathbf{u}_{\bar{\mathbf{x}}} < \lambda - \sqrt{\tfrac{2}{\mu} \mathrm{gap}(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})} \implies \forall (\mathbf{x}^\star, x^\star) \in \mathcal{X}^\star : x_\ell^\star = 0. \tag{25}$$

Quite interestingly, this test is structurally equivalent to (12). The two tests differ only in that the two dual points involved in (25) (namely $\mathbf{u}_{\bar{\mathbf{x}}}$ and $\bar{\mathbf{u}}$) need not be linked. Moreover, (25) can be regarded as a generalization of (10) which holds for any primal-dual feasible couple $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$. In particular, if $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$ is primal-dual optimal then $\mathrm{gap}(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}}) = 0$ and (25) reduces to optimality condition (10).

V. Example: Screening for Sparse SVM

We consider the following classification problem to illustrate the performance of our screening procedure:

$$\min_{(\mathbf{x},x)\in\mathbb{R}_+^n \times \mathbb{R}} \mathbf{1}_m^{\mathrm{T}} [\mathbf{1}_m - \mathrm{diag}(\mathbf{y})\mathbf{P}\mathbf{x} - \mathbf{y}x]_+ + \lambda \mathbf{1}_n^{\mathrm{T}} \mathbf{x} \tag{26}$$

where $[\cdot]_+ \triangleq \max(0, \cdot)$ and is applied component-wise to vector inputs. The rows of $\mathbf{P} \in \mathbb{R}^{m \times n}$ represent data points and the elements of $\mathbf{y} \in \{-1, +1\}^m$ the corresponding labels. Problem (26) is often referred to as "sparse support vector machine classification" and has been considered in many contributions of the literature, see *e.g.*, [4, 24].

Problem (26) can be written as a particular instance of (2) with the following identifications:

$$\begin{array}{ll} \mathbf{A} = -\mathrm{diag}(\mathbf{y})\mathbf{P} & \mathbf{c} = \mathbf{1}_m \\ \mathbf{b} = -\mathbf{y} & f = \mathbf{1}_m^{\mathrm{T}} [\cdot]_+. \end{array} \tag{27}$$

The Fenchel conjugate of f is equal to

$$f^*(\mathbf{u}) = \mathbb{I}\{\mathbf{0}_m \leq \mathbf{u} \leq \mathbf{1}_m\}, \tag{28}$$

see *e.g.*, [23, Section 4.4.3], and the dual problem of (26) thus takes the form:

$$d^\star = \max_{\mathbf{u}\in\mathbb{R}^m} -\mathbf{1}_m^{\mathrm{T}} \mathbf{u} \quad \text{s.t.} \begin{cases} -\mathbf{1}_m \leq \mathbf{u} \leq \mathbf{0}_m \\ -\mathbf{P}^{\mathrm{T}} \mathrm{diag}(\mathbf{y})\mathbf{u} \leq \lambda \mathbf{1}_n \\ \mathbf{y}^{\mathrm{T}} \mathbf{u} = 0. \end{cases} \tag{29}$$

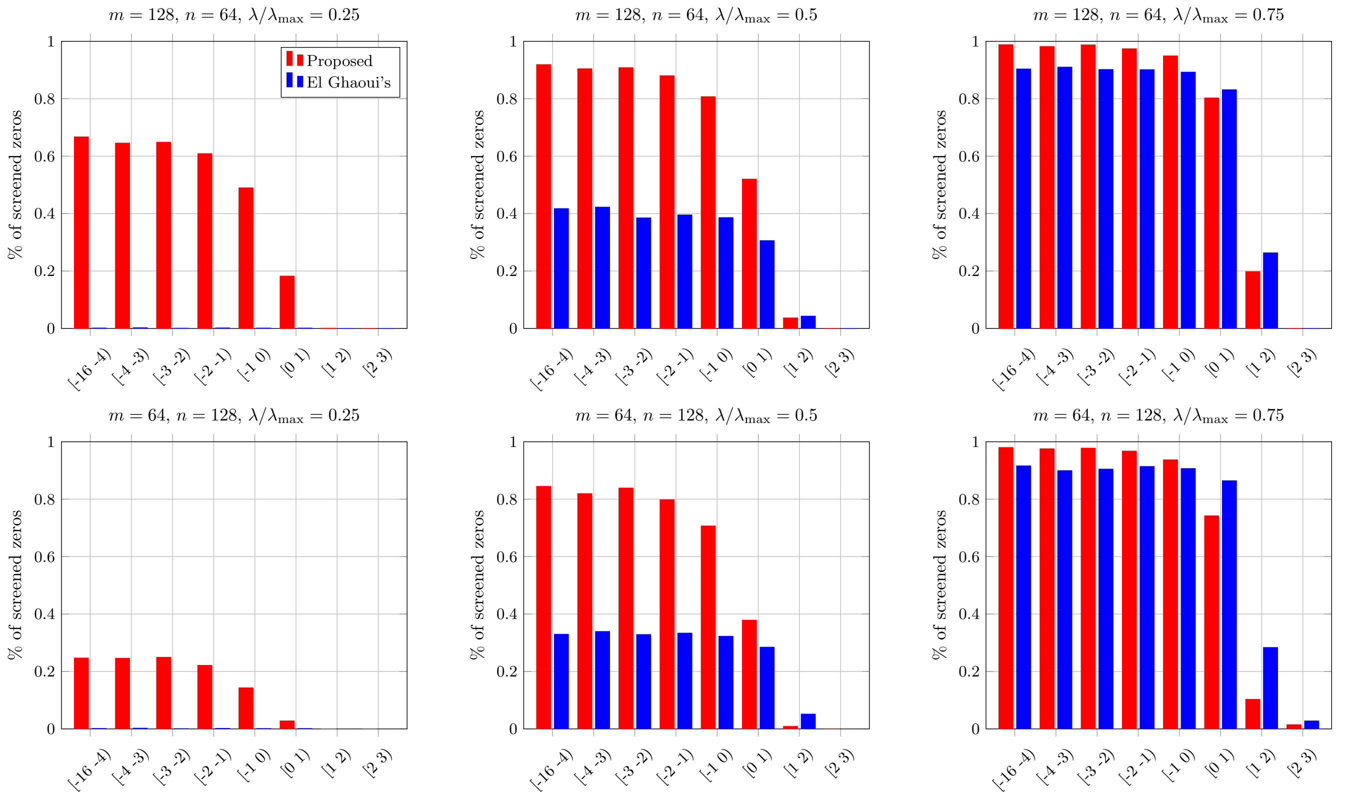

Figure 1. Percentage of zeros in the minimizers of (26) identified by the screening tests as a function of the duality gap attained by the couple $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$ used in the tests. The results are averaged over 100 problem instances. Blue: El Ghaoui *et al.*'s test [7]; Red: our region-free test (17).

Hereafter, we assess numerically the effectiveness of our region-free screening procedure (17) for problem (26).[2] The minimization over x_ℓ in (17) particularizes as

$$\min_{x_\ell \in \mathbb{R}} \ \mathbf{1}_m^{\mathrm{T}}[\bar{\mathbf{r}} - \mathrm{diag}(\mathbf{y})\mathbf{p}_\ell x_\ell]_+ + \lambda(x_\ell + \bar{x}_\ell)$$

where $\bar{\mathbf{r}} \triangleq \mathbf{1}_m - \mathbf{P}\mathrm{diag}(\mathbf{y})\bar{\mathbf{x}} - \mathbf{y}\bar{x}$. This problem is tantamount to finding the minimum of a one-dimensional piece-wise linear function and can be solved in $\mathcal{O}(m \log m)$ [7, Appendix B.1].

We compare our screening method with the test proposed by El Ghaoui *et al.* in [7, Equation (5)]. This test has the same order of complexity as our proposed method and is (to the best of our knowledge) the only safe screening procedure dealing with problem (26). In particular, since f^* is convex but even not strictly so, the GAP screening procedure does not apply here. Moreover, other methods of the literature dealing with safe screening for sparse support vector machine classification either consider screening for data points (rather than features), see *e.g.*, [25–27], or add an extra quadratic regularization term to enforce strong convexity of the cost function, see *e.g.*, [28].

Figure 1 represents the percentage of zeros of the minimizers of (26) identified by the two screening tests. The results are plotted as a function of the range of duality gap (in logarithm) attained by the couples $(\bar{\mathbf{x}}, \bar{x}, \bar{\mathbf{u}})$ used in the test. The entries of $\mathbf{P}$ are generated as $i.i.d$ realizations of a zero-mean Gaussian distribution and its columns are normalized to one. The results are averaged over 100 problem instances. The top and bottom rows respectively correspond to $(m, n) = (128, 64)$ and $(m, n) = (64, 128)$.

For the sake of simplicity, we focus on the case where $\mathbf{1}_m^{\mathrm{T}}\mathbf{y} = 0$ (that is there are as many labels "+1" and "−1"). Under this hypothesis, we have from (6)-(8) that $(\mathbf{x}^\star, x^\star) = (\mathbf{0}_n, 0)$ is the unique solution of (26) if and only if

$$\lambda_{\max} \triangleq \|[\mathbf{P}^{\mathrm{T}}\mathrm{diag}(\mathbf{y})\mathbf{1}_m]_+\|_\infty \leq \lambda.$$

In our simulation, we thus consider values of λ which are fractions of $\lambda_{\max}$: $\lambda/\lambda_{\max} \in \{0.25, 0.5, 0.75\}$.

The primal and dual feasible points used in the tests are generated as follows. For each problem instance, we solve (26) to machine-precision with the interior-point method of Matlab function `linprog`. The iterates of this method serve to define the primal points $(\bar{\mathbf{x}}, \bar{x})$ used in our screening test. For each $(\bar{\mathbf{x}}, \bar{x})$, we build a dual feasible point $\bar{\mathbf{u}}$ by "dual scaling" as suggested in [7, Section 3.3].

We see from Figure 1 that the proposed test outperforms El Ghaoui's strategy in almost all the considered operating regimes. As already noticed for other problems in the literature, El Ghaoui's method performs correctly for high values of $\lambda/\lambda_{\max}$ but its performance rapidly degrades when this ratio decreases. For example, when $\lambda/\lambda_{\max} = 0.25$ El

[2]The research presented in this paper is reproducible. Code is available at https://gitlab.inria.fr/cherzet/region-free-screening.

Ghaoui's test was not able to detect any zero of the solutions in the two setups considered in our simulations. Although the performance of our method also deteriorates when $\lambda/\lambda_{\max}$ becomes smaller, our test can still safely identify a significant proportion of zeros in most operating regimes.

VI. Conclusions

In this paper, we presented a new framework for safe screening. Unlike other methods of the literature, the construction of our test does not require the identification of a safe region of the dual space. If some strong-convexity hypothesis on the conjugate of the primal function holds, we emphasized that the state-of-the-art "GAP sphere test", proposed by Fercoq *et al.*, can be recovered as a particular case of our method. Our framework does however not require any strong-convexity assumption and therefore applies to more general families of problems. As an illustrative example, we applied our methodology to design an effective screening test for the well-known "sparse support vector machine classification" problem. As far as our simulation setup is concerned, we showed that the proposed screening procedure clearly outperforms the screening method originally proposed by El Ghaoui *et al.*

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[2] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse poisson intensity reconstruction algorithms—theory and practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[3] K. Koh, S.-J. Kim, and S. Boyd, "An interior-point method for large-scale ℓ_1-regularized logistic regression," *Journal of Machine Learning Research*, vol. 8, no. 54, pp. 1519–1555, 2007.

[4] J. Bi, K. Bennett, M. Embrechts, C. Breneman, and M. Song, "Dimensionality reduction via sparse support vector machines," *Journal of Machine Learning Research*, vol. 3, pp. 1229–1243, 2003.

[5] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] L. El Ghaoui, V. Viallon, and T. Rabbani, "Safe feature elimination for the lasso and sparse supervised learning problems.," *Pacific Journal of Optimization*, vol. 8, no. 4, pp. 667—698, 2012.

[8] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5121–5132, 2015.

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

[10] C. Herzet, C. Dorffer, and A. 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.

[11] C. Elvira and C. Herzet, "Safe squeezing for antisparse coding," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3252–3265, 2020.

[12] C. Elvira and C. Herzet, "Safe rules for the identification of zeros in the solutions of the SLOPE problem," arXiv 2110.11784, 2021.

[13] Z. J. Xiang, H. Xu, and P. J. Ramadge, "Learning sparse representations of high dimensional data on large scale dictionaries," in *Proceedings of Advances in neural information processing systems*, pp. 900–908. 2011.

[14] L. Dai and K. Pelckmans, "An ellipsoid based, two-stage screening test for BPDN," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2012, pp. 654–658.

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

[16] J. Liu, Z. Zhao, J. Wang, and J. Ye, "Safe screening with variational inequalities and its application to Lasso," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2014, pp. 289–297.

[17] J. Wang, P. Wonka, and J. Ye, "Lasso screening rules via dual polytope projection," *Journal of Machine Learning Research*, 2015.

[18] C. Herzet and A. Malti, "Safe screening tests for LASSO based on firmly non-expansiveness," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 4732–4736.

[19] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap: safer rules for the Lasso," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2015, pp. 333–342.

[20] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, "Gap safe screening rules for sparsity enforcing penalties," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 4671–4703, 2017.

[21] C. F. Dantas, E. Soubies, and C. Févotte, "Safe screening for sparse regression with the Kullback-Leibler divergence," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2021, pp. 5544–5548.

[22] C. Dantas, E. Soubies, and C. Févotte, "Expanding boundaries of Gap safe screening," *Journal of Machine Learning Research*, vol. 22, no. 236, pp. 1–57, 2021.

[23] A. Beck, *First-Order Methods in Optimization*, SIAM-Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2017.

[24] J. Zhu, S. Rosset, R. Tibshirani, and T. Hastie, "1-norm support vector machines," in *Proceedings of Advances in neural information processing systems*, 2004, vol. 16.

[25] K. Ogawa, Y. Suzuki, and I. Takeuchi, "Safe screening of non-support vectors in pathwise SVM computation," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2013, vol. 28, pp. 1382–1390.

[26] J. Wang, P. Wonka, and J. Ye, "Scaling SVM and least absolute deviations via exact data reduction," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2014, vol. 32, pp. 523–531.

[27] J. Zimmert, C. Schröder, G. Kerg, and M. Kloft, "Safe screening for support vector machines," in *Proceedings of Advances in neural information processing systems*, 2015.

[28] W. Zhang, B. Hong, W. Liu, J. Ye, D. Cai, X. He, and J. Wang, "Scaling up sparse support vector machines by simultaneous feature and sample reduction," in *Proceedings of the International Conference on Machine Learning (ICML)*, Aug. 2017, vol. 70, pp. 4016–4025.