# Beyond GAP screening for Lasso by exploiting new dual cutting half-spaces with supplementary material

Thu-Le Tran*, Clément Elvira†, Hong-Phuong Dang‡, and Cédric Herzet§

*Univ Rennes, IRMAR - UMR 6625, F-35000 Rennes, France
†IETR UMR CNRS 6164, CentraleSupelec Rennes Campus, 35576 Cesson Sévigné, France
‡LaTIM, INSERM-UMR1101, *Univ. de Bretagne Occidentale*, Brest, France
§Inria Rennes - Bretagne Atlantique, Rennes, France
emails: thu-le.tran@univ-rennes1.fr, clement.elvira@centralesupelec.fr, dang@univ-brest.fr, cedric.herzet@inria.fr

***Abstract*—In this paper, we propose a novel safe screening test for Lasso. Our procedure is based on a safe region with a dome geometry and exploits a canonical representation of the set of half-spaces (referred to as "dual cutting half-spaces" in this paper) containing the dual feasible set. The proposed safe region is shown to be always included in the state-of-the-art "GAP Sphere" and "GAP Dome" proposed by Fercoq *et al.* (and strictly so under very mild conditions) while involving the same computational burden. Numerical experiments confirm that our new dome enables to devise more powerful screening tests than GAP regions and lead to significant acceleration to solve Lasso.**

***Index Terms*—Lasso, convex optimization, safe screening.**

## I. Introduction

Finding sparse representations is a fundamental problem in signal processing and machine learning. It consists in decomposing some vector $\mathbf{y} \in \mathbb{R}^m$ as a linear combination of a few columns (referred to as *atoms*) of a matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ called *dictionary*. A popular strategy for obtaining sparse representations is to solve the so-called Lasso problem:

$$\mathbf{x}^\star \in \operatorname*{arg\,min}_{\mathbf{x} \in \mathbb{R}^n} \; P(\mathbf{x}) \triangleq \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \qquad (1)$$

where $\lambda > 0$, see [1]. Inasmuch as solving (1) may require a heavy computational burden as $n$ becomes large, the design of efficient optimization techniques tackling this problem has become an active field of research. Among the most popular approaches addressing (1), one can mention [2–4].

A noteworthy approach in this field is the acceleration method proposed by El Ghaoui *et al.* in [5] and known as *safe screening*, which aims to design simple tests to identify zeros entries in the minimizers of (1). This knowledge can then be exploited to (potentially significantly) reduce the dimensionality of the problem by discarding the atoms of the dictionary weighted by zero. Over the past few years, safe screening has sparked a surge of interest in the literature, see *e.g.*, [6–13] and beyond $\ell_1$-regularization [14–17].

Standard screening methodologies leverage the concept of "*safe region*", a set provably containing the optimal solution of the dual problem of (1), see *e.g.*, [18, Section 4]. The choice of the safe region reveals to be crucial to the final effectiveness and efficiency of the screening tests. On the one hand, loosely speaking, "smaller" regions lead to more effective tests, see [18, Lemma 1]. On the other hand, the complexity of the tests is closely related to the geometry of the safe region. As a consequence, safe regions with "simple" geometries such as spheres [5–9] or domes [12, 13] are commonly considered in the literature.

One state-of-the-art methodology to find a good compromise between these two requirements was proposed in [8]: the authors introduced two new safe regions (referred to as *"GAP sphere"* and *"GAP dome"*) whose radii are proportional to the duality gap attained by the primal-dual feasible couple used to design the region. The radii of GAP regions have therefore the desirable feature to converge to zero when the primal-dual feasible couple tends to a primal-dual solution, and thus lead to extremely effective screening tests.

In this paper, we propose a new safe dome which is provably contained in the GAP regions. The definition of our dome is based on a fine characterization of the set of half-spaces (referred to as "*dual cutting half-spaces*" hereafter) containing the whole dual feasible set. Its construction has the same complexity as GAP dome and also relies on the identification of some primal-dual feasible couple.

The paper is organized as follows. In the next section, we define the notations used throughout the paper. In Section III, some background on Lasso and safe screening is provided. In Section IV, we derive a fine characterization of the set of dual cutting half-spaces and present our new safe dome. The relevance of our proposed approach is finally illustrated in Section V via numerical simulations.

## II. NOTATIONS

We use the following notational conventions throughout the paper. Boldface uppercase (*e.g.*, $\mathbf{A}$) and lowercase (*e.g.*, $\mathbf{x}$) letters respectively represent matrices and vectors. $\mathbf{0}_n$ denotes the all-zeros vector of $\mathbb{R}^n$. $\langle\cdot,\cdot\rangle$ stands for the canonical inner product. The $i$th entry of a vector $\mathbf{x}$ is denoted $\mathbf{x}(i)$. Calligraphic letters (*e.g.*, $\mathcal{S}$) are used for sets.

## III. BACKGROUND

In this section, we provide some elements of convex analysis for problem (1) and recall the main ingredients underlying the concept of safe screening for Lasso.

### A. Dual problem and optimality conditions

We first note that (1) admits at least one minimizer since $P(\cdot)$ is continuous, proper and coercive [19, Theorem 2.14]. The dual problem of (1) writes as

$$\mathbf{u}^\star = \operatorname*{arg\,max}_{\mathbf{u}\in\mathcal{U}} \; D(\mathbf{u}) \triangleq \tfrac{1}{2}\|\mathbf{y}\|_2^2 - \tfrac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2 \tag{2}$$

where $\mathcal{U} \triangleq \{\mathbf{u}\in\mathbb{R}^m : \|\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_\infty \le \lambda\}$ is the so-called *dual feasible set*, see [20, Appendix A]. Since $\mathcal{U}$ is closed and $D(\cdot)$ is strictly concave, problem (2) admits a unique maximizer $\mathbf{u}^\star$.

It is known that strong duality holds between problems (1) and (2), that is

$$\forall\mathbf{x}\in\mathbb{R}^n, \forall\mathbf{u}\in\mathcal{U}: \;\; \mathrm{gap}(\mathbf{x},\mathbf{u}) \triangleq P(\mathbf{x}) - D(\mathbf{u}) \ge 0 \tag{3}$$

with equality if and only if $(\mathbf{x},\mathbf{u})$ is primal-dual optimal, see *e.g.*, [19, Theorem A.2]. Moreover, any primal-dual optimal couple $(\mathbf{x}^\star,\mathbf{u}^\star)$ must verify the following optimality conditions [20, Section 2]:

$$\mathbf{u}^\star = \mathbf{y} - \mathbf{A}\mathbf{x}^\star \tag{4}$$

and,

$$\langle\mathbf{a}_i,\mathbf{u}^\star\rangle = \begin{cases}\lambda\,\mathrm{sign}(\mathbf{x}^\star(i)) & \text{if } \mathbf{x}^\star(i)\neq 0,\\ s\in[-\lambda,\lambda] & \text{otherwise.}\end{cases} \tag{5}$$

for all $i=1,\dots,n$. It is easy to see from these conditions that $\mathbf{x}^\star = \mathbf{0}_n$ is the unique solution of (1) if and only if

$$\lambda \ge \lambda_{\max} \triangleq \|\mathbf{A}^{\mathrm{T}}\mathbf{y}\|_\infty. \tag{6}$$

### B. Safe screening

Safe screening tests leverage the following consequence of (5):

$$|\langle\mathbf{a}_i,\mathbf{u}^\star\rangle| < \lambda \implies \mathbf{x}^\star(i) = 0. \tag{7}$$

If the inequality in the left-hand side of (7) is verified for some index $i$, the corresponding column of $\mathbf{A}$ can therefore be safely removed without changing the minimum value of (1). Although computing $\mathbf{u}^\star$ is (generally) as difficult as solving (1), a weaker version of (7) can be obtained if a region $\mathcal{R}\subseteq\mathbb{R}^n$ containing $\mathbf{u}^\star$ (often called "safe region") is known. (7) can then be relaxed to:

$$\max_{\mathbf{u}\in\mathcal{R}} |\langle\mathbf{a}_i,\mathbf{u}\rangle| < \lambda \implies \mathbf{x}^\star(i) = 0. \tag{8}$$

From an effectiveness point of view, $\mathcal{R}$ should be chosen as small as possible. In particular, if $\mathcal{R}\subseteq\mathcal{R}'$ then obviously

$$\max_{\mathbf{u}\in\mathcal{R}'} |\langle\mathbf{a}_i,\mathbf{u}\rangle| < \lambda \implies \max_{\mathbf{u}\in\mathcal{R}} |\langle\mathbf{a}_i,\mathbf{u}\rangle| < \lambda, \tag{9}$$

that is the screening test built from $\mathcal{R}$ will always detect at least as many zeros as that constructed with $\mathcal{R}'$.

From a complexity point of view, the computational cost of (8) mostly depends on the evaluation of $\max_{\mathbf{u}\in\mathcal{R}}|\langle\mathbf{a}_i,\mathbf{u}\rangle|$. A simple strategy to lower the cost of this operation consists in building safe regions with "appropriate" geometries. Two standard choices are spheres and domes.

Sphere regions are defined by their center $\mathbf{c}\in\mathbb{R}^m$ and radius $R\ge 0$:

$$\mathcal{R} = \mathcal{B}(\mathbf{c},R) \triangleq \{\mathbf{u}\in\mathbb{R}^m : \|\mathbf{u}-\mathbf{c}\|_2 \le R\}. \tag{10}$$

Particularizing the left-hand side of (8) to the case where $\mathcal{R}$ is a sphere, we obtain:

$$\max_{\mathbf{u}\in\mathcal{B}(\mathbf{c},R)} |\langle\mathbf{a}_i,\mathbf{u}\rangle| = |\langle\mathbf{a}_i,\mathbf{c}\rangle| + R\|\mathbf{a}_i\|_2. \tag{11}$$

We see that the solution of the maximization problem then admits a simple closed form. Its computation only requires the evaluation of one inner product between $\mathbf{a}_i$ and $\mathbf{c}$.

Another popular choice of geometry is dome region. A dome is defined as the intersection of a sphere and an half-space, that is

$$\mathcal{R} = \mathcal{D}(\mathbf{c},R,\mathbf{g},\delta) \triangleq \mathcal{B}(\mathbf{c},R)\cap\mathcal{H}(\mathbf{g},\delta) \tag{12}$$

where $\mathbf{g}\in\mathbb{R}^m$, $\delta\in\mathbb{R}$ and[1]

$$\mathcal{H}(\mathbf{g},\delta) \triangleq \{\mathbf{u}\in\mathbb{R}^m : \langle\mathbf{g},\mathbf{u}\rangle \le \delta\}. \tag{13}$$

In this case, the left-hand side of (8) also admits a closed-form solution, see [20, Lemma. 3]. More precisely, we have

$$\max_{\mathbf{u}\in\mathcal{D}} |\langle\mathbf{a}_i,\mathbf{u}\rangle| = \max\left(\max_{\mathbf{u}\in\mathcal{D}} \langle\mathbf{a}_i,\mathbf{u}\rangle, \max_{\mathbf{u}\in\mathcal{D}} \langle-\mathbf{a}_i,\mathbf{u}\rangle\right) \tag{14}$$

and $\forall\mathbf{a}\in\mathbb{R}^m\setminus\{\mathbf{0}_m\}$, $\mathcal{D}\neq\emptyset$:

$$\max_{\mathbf{u}\in\mathcal{D}} \langle\mathbf{a},\mathbf{u}\rangle = \langle\mathbf{a},\mathbf{c}\rangle + R\|\mathbf{a}\|_2 f(\psi_1,\psi_2), \tag{15}$$

where

$$f(\psi_1,\psi_2) = \begin{cases}1 & \text{if } \psi_1 \le \psi_2,\\ \psi_1\psi_2 + \sqrt{1-\psi_1^2}\sqrt{1-\psi_2^2} & \text{otherwise}\end{cases}$$
$$\psi_1 = \frac{\langle\mathbf{a},\mathbf{g}\rangle}{\|\mathbf{a}\|_2\|\mathbf{g}\|_2}, \quad \psi_2 = \min\left(\frac{\delta-\langle\mathbf{g},\mathbf{c}\rangle}{R\|\mathbf{g}\|_2}, 1\right).$$

The cost associated to (15) is mostly dominated by the computation of $\langle\mathbf{g},\mathbf{c}\rangle$, $\langle\mathbf{a}_i,\mathbf{c}\rangle$ and $\langle\mathbf{a}_i,\mathbf{g}\rangle$. Dome tests are therefore usually slightly more complex to implement than sphere tests. However, since dome geometry enables smaller safe regions, they potentially lead to much more effective tests.

[1] Note that when $\mathbf{g}=\mathbf{0}_m$, our definition implies that $\mathcal{H}(\mathbf{g},\delta)$ either reduces to $\mathbb{R}^m$ if $\delta\ge 0$ or is empty if $\delta<0$.

### C. GAP regions

To conclude this section, we provide the expressions of two well-known state-of-the-art safe regions proposed in [8]. The first region takes the form of a sphere and is known as the *"GAP sphere"* in the literature. It is defined by the following choice of parameters:

$$\mathbf{c} = \mathbf{u} \tag{16}$$
$$R = \sqrt{2\text{gap}(\mathbf{x}, \mathbf{u})}) \tag{17}$$

where $(\mathbf{x}, \mathbf{u})$ can be any primal-dual feasible couple. The second has the geometry of a dome and is usually dubbed *"GAP dome"*. It is defined by the following set of parameters:

$$\mathbf{c} = \tfrac{1}{2}(\mathbf{y} + \mathbf{u}) \tag{18}$$
$$R = \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2 \tag{19}$$
$$\mathbf{g} = \mathbf{y} - \mathbf{c} \tag{20}$$
$$\delta = \langle \mathbf{g}, \mathbf{c} \rangle + \text{gap}(\mathbf{x}, \mathbf{u}) - R^2 \tag{21}$$

where $(\mathbf{x}, \mathbf{u})$ can be any primal-dual feasible couple. In the sequel, with a slight abuse of notation, we will respectively denote the GAP sphere and the GAP dome as $\mathcal{B}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ and $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ to explicitly emphasize their dependence on the choice of the primal-dual feasible couple $(\mathbf{x}, \mathbf{u})$.

Although not proved explicitly in [8], it can be easily shown (see Appendix B) that

$$\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}) \subseteq \mathcal{B}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}). \tag{22}$$

These two regions have the desirable feature of having a radius (see (32)) that decreases to zero when the couple $(\mathbf{x}, \mathbf{u})$ tends to a primal-dual solution. In particular, if $(\mathbf{x}, \mathbf{u}) = (\mathbf{x}^\star, \mathbf{u}^\star)$, $\{\mathbf{u}^\star\} = \mathcal{B}_{\texttt{gap}}(\mathbf{x}^\star, \mathbf{u}^\star)$ and the screening tests based on GAP regions reduce to (7).

## IV. Safe dome with general dual cutting half-spaces

In this section, we present and motivate our new safe dome region. The construction of our dome is based on the set of dual cutting half-spaces $\mathcal{H}(\mathbf{g}, \delta)$:

$$\mathcal{G} \triangleq \{(\mathbf{g}, \delta) \in \mathbb{R}^m \times \mathbb{R} \colon \langle \mathbf{g}, \mathbf{u} \rangle \leq \delta \ \forall \mathbf{u} \in \mathcal{U}\}. \tag{23}$$

The next lemma provides a canonical characterization of $\mathcal{G}$. A proof can be found in Appendix A-A.

**Lemma 1.**

$$\mathcal{G} = \{(\mathbf{A}\mathbf{x}, \delta) \colon \mathbf{x} \in \mathbb{R}^n, \delta \geq \lambda \|\mathbf{x}\|_1\}. \tag{24}$$

We now expose our proposed new dome region and state a result showing that it is guaranteed to perform at least as well as GAP safe regions presented in Section III-C.

The definition of the proposed dome is encapsulated in the following theorem:

**Theorem 1.** *Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathcal{U}$ and*

$$\mathbf{c} = \tfrac{1}{2}(\mathbf{y} + \mathbf{u}) \tag{25}$$
$$R = \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2 \tag{26}$$
$$\mathbf{g} = \mathbf{A}\mathbf{x} \tag{27}$$
$$\delta = \lambda \|\mathbf{x}\|_1 \tag{28}$$

*Then*

$$\mathbf{u}^\star \in \mathcal{D}(\mathbf{c}, R, \mathbf{g}, \delta). \tag{29}$$

A proof of this result can be found in Appendix A-B. The safeness of $\mathcal{H}(\mathbf{A}\mathbf{x}, \lambda\|\mathbf{x}\|_1)$ can in fact be seen as a simple consequence of the Hölder inequality:

$$\langle \mathbf{A}\mathbf{x}, \mathbf{u}^\star \rangle = \langle \mathbf{x}, \mathbf{A}^{\mathrm{T}}\mathbf{u}^\star \rangle \leq \|\mathbf{x}\|_1 \|\mathbf{A}^{\mathrm{T}}\mathbf{u}^\star\|_\infty \leq \lambda \|\mathbf{x}\|_1,$$

where the last inequality follows from dual feasibility of $\mathbf{u}^\star$. In the sequel we will therefore refer to the dome defined in Theorem 1 as "*Hölder dome*". We note that, similarly to the GAP regions, the Hölder dome is completely specified by the choice of a primal-dual feasible couple $(\mathbf{x}, \mathbf{u})$. Hereafter, we will use the notation $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$ to emphasize this fact.

We note that the definition of $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$ only differs from that of $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ in the choice of the half-space $\mathcal{H}(\mathbf{g}, \delta)$, where the canonical characterization of $\mathcal{G}$ in Lemma 1 is directly exploited in the former. Our next result shows that this choice is beneficial regarding the size of the safe region:

**Theorem 2.** *For $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathcal{U}$:*

$$\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}) \subseteq \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}). \tag{30}$$

*Moreover, if $P(\mathbf{x}) < P(\mathbf{0}_n)$ and $(\mathbf{x}, \mathbf{u})$ is not primal-dual optimal, then the inclusion is strict.*

A proof of this result can be found in Appendix A-C. Theorem 2 shows that the Hölder dome is guaranteed to be a subset of the GAP dome and sphere. From (9), this suggests that screening tests based on our proposed dome are ensured to perform at least as well as the two other safe regions. We also note that the condition "$P(\mathbf{x}) < P(\mathbf{0}_n)$", ensuring $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}) \subset \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$, is very mild and verified in many practical setups. It is for example the case when the iterates $\{\mathbf{x}^{(t)}\}_{t=1}^{\infty}$ of an optimization procedure, initialized at $\mathbf{x}^{(0)} = \mathbf{0}_n$ and monotically decreasing the cost function, are used to define the primal point $\mathbf{x}$ used in the construction of the dome.

## V. Numerical Experiment

This section reports an empirical study of the relevance of the Hölder dome presented in Section IV.[2] For the two experiments, our simulation setup is as follows. We set $(m, n) = (100, 500)$. For each trial, new realizations of $\mathbf{A}$ and $\mathbf{y}$ are generated. The observation $\mathbf{y}$ is drawn according to a uniform distribution on the $m$-dimensional unit sphere while $\mathbf{A}$ either satisfies *i)* the entries are i.i.d. realizations of a normal distribution or *ii)* $\mathbf{A}$ has a Toeplitz structure, *i.e.*, columns are shifted versions of a Gaussian curve. The columns of $\mathbf{A}$ are then normalized such that $\|\mathbf{a}_i\|_2 = 1$ for all $i$.

[2] The code is available at https://gitlab.inria.fr/cherzet/holder_dome_lasso.

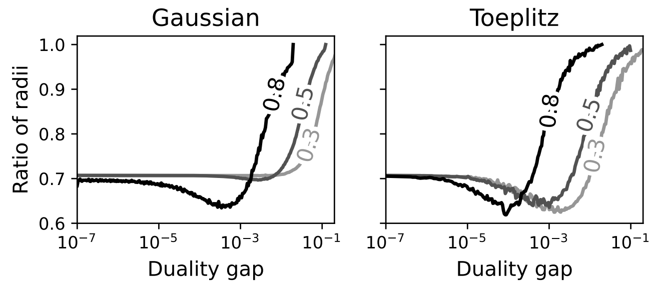


Fig. 1. Expected value of the ratio (31) as a function of the duality gap achieved by $(\mathbf{x}, \mathbf{u})$ for the two dictionaries. Each curve corresponds to a value of the ratio $\lambda/\lambda_{\max}$ (the value is indicated along the line).

*a) Radius of safe regions:* In this first experiment, we investigate the size difference between the Hölder and GAP domes. More precisely, we evaluate the ratio

$$\frac{\texttt{Rad}(\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}))}{\texttt{Rad}(\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}))} \tag{31}$$

for different choices of $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{u} \in \mathcal{U}$, where $\texttt{Rad}(\mathcal{S})$ denotes the radius of the (closed bounded) set $\mathcal{S}$:

$$\texttt{Rad}(\mathcal{S}) \triangleq \max_{\mathbf{u}, \mathbf{u}' \in \mathcal{S}} \ \tfrac{1}{2}\|\mathbf{u}' - \mathbf{u}\|_2. \tag{32}$$

Figure 1 shows the average value of the ratio (31) as a function of the duality gap achieved by $(\mathbf{x}, \mathbf{u})$ for the two considered dictionaries and three values of the ratio $\lambda/\lambda_{\max}$. Results have been obtained by averaging over 50 trials. As expected (cf. Theorem 2), the ratios are always lower than 1. One also observes that the radius of the proposed Hölder dome is up to $0.6$ smaller than the GAP dome. As far as our simulation is concerned, all the curves seem to converge to a ratio close to $0.7$ as the dual gap tends to zero.

*b) Benchmarks:* In this second experiment, we assess the computational gain obtained with the Hölder dome and GAP-based regions. To do so, we compare three variants of FISTA –a standard method to address (1), see [3]– where the iterations are interleaved with screening tests that leverage *i)* the GAP sphere, *ii)* the GAP dome, *iii)* the Hölder dome. More precisely, at each iteration $t$, a screening test is carried out with the corresponding safe region obtained with parameters $(\mathbf{x}^{(t)}, \mathbf{u}^{(t)})$ where $\mathbf{x}^{(t)}$ refers to the current iterate and $\mathbf{u}^{(t)}$ is obtained by dual scaling of $\mathbf{y} - \mathbf{A}\mathbf{x}^{(t)}$ (see [5, Section 3.3]). We use the "Dolan-Moré" performance profiles [21] to assess the performance of the three methods.

We run each method with a prescribed computational budget (the number of floating point operations) on 200 instances of problem (1). We then evaluate the (empirical) probability $\rho(\tau)$ that a solver achieves a duality gap lower that $\tau$ upon completion. For each setup, the budget is adjusted so that $\rho(10^{-7}) = 50\%$ for the solver using the Hölder dome.

Figure 2 shows the performance profiles for the two considered dictionaries and different values of the ratio $\lambda/\lambda_{\max}$. One sees that, as far as our simulation setup is concerned, implementing the screening test (8) with the Hölder dome improves quite significantly the average accuracy achieved in all but one setup. These findings support our claim that the Hölder dome leads to more effective test than the one using GAP based-region. The case where the three safe regions lead to comparable results (Gaussian dictionary and $\lambda/\lambda_{\max} = 0.3$) has to be understood as follows: even though the Hölder and GAP domes are expected to perform better than the GAP sphere, the profiles result from a compromise between the effectiveness of the test and its complexity. In particular, a study of our simulation results shows that even though the tests are less effective, more iterations are carried (in average) with the GAP sphere in that specific setup, thus leading to a potentially more accurate solution.

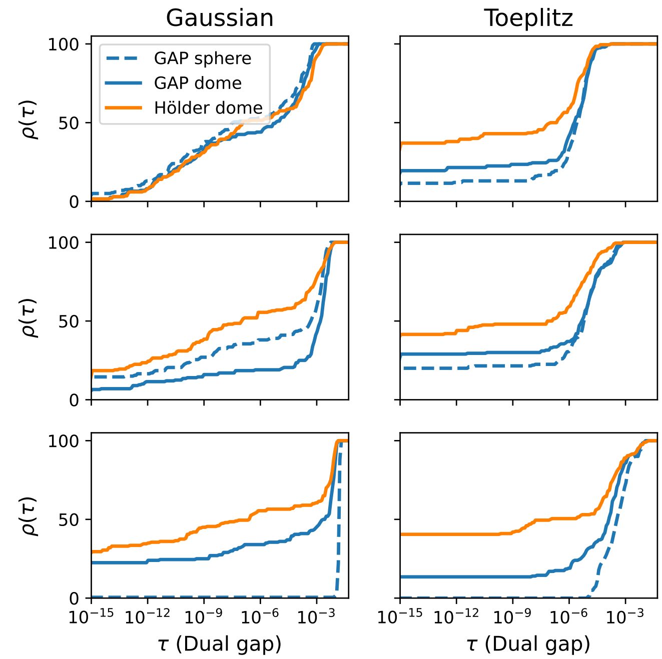


Fig. 2. Performance profiles of screening methods with different safe regions. Each row corresponds to a value of $\lambda/\lambda_{\max}$. From top to bottom: .3, .5, .8.

## VI. Conclusion

In this paper, we introduced a novel safe dome region for Lasso that can be used to design safe screening tests. We showed that our proposed dome region is always a (potentially strict) subset of the GAP sphere and dome, two ubiquitous safe regions in the literature. The proposed methodology is shown to allow significant computational gains when solving Lasso with a prescribed computational budget.

## References

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

[2] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," vol. 1, no. 4, pp. 586–597, Dec. 2007.

[3] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” vol. 2, no. 1, pp. 183–202, 2009.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” vol. 3, no. 1, pp. 1–122, 2011.
[5] L. El Ghaoui, V. Viallon, and T. Rabbani, “Safe feature elimination for the Lasso and sparse supervised learning problems,” vol. 8, no. 4, pp. 667—698, Jan. 2012.
[6] Z. Xiang, H. Xu, and P. J. Ramadge, “Learning sparse representations of high dimensional data on large scale dictionaries,” vol. 24, 2011.
[7] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, “A dynamic screening principle for the Lasso,” 2014, pp. 6–10.
[8] O. Fercoq, A. Gramfort, and J. Salmon, “Mind the duality gap: safer rules for the Lasso,” 2015, pp. 333–342.
[9] C. Herzet and A. Malti, “Safe screening tests for Lasso based on firmly non-expansiveness,” Mar. 2016, pp. 4732–4736.
[10] C. Herzet, C. Dorffer, and A. Drémeau, “Gather and conquer: Region-based strategies to accelerate safe screening tests,” vol. 67, no. 12, pp. 3300–3315, 2019.
[11] L. Dai and K. Pelckmans, “An ellipsoid based, two-stage screening test for BPDN,” 2012, pp. 654–658.
[12] Z. J. Xiang and P. J. Ramadge, “Fast Lasso screening tests based on correlations.” IEEE, 2012, pp. 2137–2140.
[13] “Safe screening with variational inequalities and its application to Lasso,” 2014, pp. 289–297.
[14] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, “Gap safe screening rules for sparsity enforcing penalties,” vol. 18, no. 1, pp. 4671–4703, 2017.
[15] C. Elvira and C. Herzet, “Safe squeezing for antisparse coding,” vol. 68, pp. 3252–3265, 2020.
[16] ——, “Safe rules for the identification of zeros in the solutions of the SLOPE problem,” arXiv:2110.11784, 2021.
[17] T. Guyard, C. Herzet, and C. Elvira, “Screen & relax: accelerating the resolution of Elastic-Net by safe identification of the solution support,” arXiv:2110.07281, 2021.
[18] Z. J. Xiang, Y. Wang, and P. J. Ramadge, “Screening tests for Lasso problems,” vol. 39, no. 5, pp. 1008–1027, 2017.
[19] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017.
[20] Z. J. Xiang, Y. Wang, and P. J. Ramadge, “Screening tests for Lasso problems,” vol. 39, no. 5, pp. 1008–1027, 2016.
[21] E. D. Dolan and J. J. Moré, “Benchmarking optimization software with performance profiles,” vol. 91, no. 2, pp. 201–213, Jan. 2002.

# APPENDIX A
# PROOFS

## A. Proof of Lemma 1

Let $\mathbf{g} \in \mathbb{R}^m$ and consider the following optimization problem

$$\delta^\star = \sup_{\mathbf{u}\in\mathcal{U}} \langle \mathbf{g}, \mathbf{u}\rangle. \tag{33}$$

By strong duality [19, Equation (12.4) combined with Theorem A.1], we have

$$\delta^\star = \inf_{\mathbf{x}\in\mathbb{R}^n} \lambda\|\mathbf{x}\|_1 + \eta\{\mathbf{g} = \mathbf{A}\mathbf{x}\}, \tag{34}$$

where $\eta\{\cdot\}$ denotes the indicator function, which is equal to 0 if the statement in the braces is true and $+\infty$ otherwise.

Hence, if $\mathbf{g} = \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, we have from (34) that $\delta^\star \leq \lambda\|\mathbf{x}\|_1$ and therefore $(\mathbf{A}\mathbf{x}, \delta) \in \mathcal{G}$ for all $\delta \geq \lambda\|\mathbf{x}\|_1$. Conversely, if $(\mathbf{g}, \delta) \in \mathcal{G}$ then $\delta^\star \leq \delta < \infty$. From (34), we thus have that there exists some $\mathbf{x}$ such that $\mathbf{g} = \mathbf{A}\mathbf{x}$ and $\delta \geq \lambda\|\mathbf{x}\|_1$.

## B. Proof of Theorem 1

First remember that a dome is defined as the intersection of the ball $\mathcal{B}(\mathbf{c}, R)$ and the half-space $\mathcal{H}(\mathbf{g}, \mathcal{H})$, see (12). It is then sufficient to show that both $\mathcal{B}(\mathbf{c}, R)$ and $\mathcal{H}(\mathbf{g}, \delta)$ are safe. Finally, the safeness of $\mathcal{B}(\mathbf{c}, R)$ and $\mathcal{H}(\mathbf{g}, \mathcal{H})$ respectively follows from [13, Section 2.2] and Lemma 1.

## C. Proof of Theorem 2

Let $\mathcal{H}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$ and $\mathcal{H}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ respectively denote the half-space defining $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$ and $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$. Consider

$$\mathbf{u}' \in \mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}) = \mathcal{B}(\mathbf{c}, R) \cap \mathcal{H}_{\texttt{new}}(\mathbf{x}, \mathbf{u}), \tag{35}$$

where $\mathbf{c}, R$ are the parameters defined in (18) and (19), respectively. Since $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$ and $\mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ only differ in the definition of their half-space, it is sufficient to show that $\mathbf{u}' \in \mathcal{H}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$, *i.e.*,

$$\langle \mathbf{y} - \mathbf{c}, \mathbf{u}'\rangle \leq \langle \mathbf{y} - \mathbf{c}, \mathbf{c}\rangle + \mathrm{gap}(\mathbf{x}, \mathbf{u}) - R^2. \tag{36}$$

We have

$$\begin{aligned} 2\langle \mathbf{y} - \mathbf{c}, \mathbf{u}' - \mathbf{c}\rangle &= \|\mathbf{y} - \mathbf{c}\|_2^2 + \|\mathbf{u}' - \mathbf{c}\|_2^2 - \|\mathbf{y} - \mathbf{u}'\|_2^2 \\ &\leq 2\Big(R^2 - \tfrac{1}{2}\|\mathbf{y} - \mathbf{u}'\|_2^2\Big) \\ &= 2\big(D(\mathbf{u}') - D(\mathbf{u}) - R^2\big) \end{aligned} \tag{37}$$

where the inequality follows from the fact that both $\mathbf{y}$ and $\mathbf{u}'$ belong to $\mathcal{B}(\mathbf{c}, R)$. Moreover, one can write

$$\begin{aligned} 2D(\mathbf{u}') &= \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{y} - \mathbf{u}' - \mathbf{A}\mathbf{x}\|_2^2 + 2\langle \mathbf{A}\mathbf{x}, \mathbf{u}'\rangle \\ &\leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + 2\langle \mathbf{A}\mathbf{x}, \mathbf{u}'\rangle \\ &\leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + 2\lambda\|\mathbf{x}\|_1 = 2P(\mathbf{x}) \end{aligned} \tag{38}$$

where the inequalities result from the non-negativity of $\|\mathbf{y} - \mathbf{u}' - \mathbf{A}\mathbf{x}\|_2^2$ and the hypothesis that $\mathbf{u}' \in \mathcal{H}_{\texttt{new}}(\mathbf{x}, \mathbf{u})$. Combining (37) and (38), we finally obtain (36), that is $\mathbf{u}' \in \mathcal{H}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$.

Assume now that $P(\mathbf{x}) < P(\mathbf{0}_n)$ and $(\mathbf{x}, \mathbf{u})$ is not primal-dual optimal. We next show that the strict inclusion $\mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}) \subset \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ holds in that case. To this end, it is sufficient to identify some $\mathbf{u}_0 \in \mathbb{R}^m$ satisfying

$$\mathbf{u}_0 \in \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u}) \setminus \mathcal{D}_{\texttt{new}}(\mathbf{x}, \mathbf{u}). \tag{39}$$

Let $\mathbf{u}_0 \triangleq \mathbf{c} + \frac{1}{R^2}(P(\mathbf{x}) - D(\mathbf{u}) - R^2)(\mathbf{y} - \mathbf{c})$. We note that $\mathbf{u}_0$ is well defined since $R \neq 0$ by construction.

We can show that $\mathbf{u}_0 \in \mathcal{D}_{\texttt{gap}}(\mathbf{x}, \mathbf{u})$ as follows. We first easily verify that

$$\langle \mathbf{y} - \mathbf{c}, \mathbf{u}_0 - \mathbf{c}\rangle = P(\mathbf{x}) - D(\mathbf{u}) - R^2. \tag{40}$$

The result is then proved if $\mathbf{u}_0 \in \mathcal{B}(\mathbf{c}, R)$, *i.e.*,

$$\left|P(\mathbf{x}) - D(\mathbf{u}) - R^2\right| < R^2. \tag{41}$$

We can show that (41) holds by distinguishing between two cases. First, if $P(\mathbf{x}) - D(\mathbf{u}) \geq R^2$, the absolute value in (41) can be removed and the result follows from $P(\mathbf{x}) < P(\mathbf{0}_n) = D(\mathbf{u}) + 2R^2$. Second, if $P(\mathbf{x}) - D(\mathbf{u}) < R^2$, the same conclusion holds by using the fact $P(\mathbf{x}) - D(\mathbf{u}) > 0$ since $(\mathbf{x}, \mathbf{u})$ is not primal-dual optimal by hypothesis.

Let us finally show that $\mathbf{u}_0 \notin \mathcal{D}_{\mathtt{new}}(\mathbf{x}, \mathbf{u})$. We have

$$\begin{aligned}
2D(\mathbf{u}_0) &= \|\mathbf{y}\|_2^2 - \|\mathbf{y} - \mathbf{u}_0\|_2^2 \\
&= \|\mathbf{y}\|_2^2 - \|\mathbf{y} - \mathbf{c}\|_2^2 + 2\langle \mathbf{y} - \mathbf{c}, \mathbf{u}_0 - \mathbf{c}\rangle - \|\mathbf{u}_0 - \mathbf{c}\|_2^2 \\
&> \|\mathbf{y}\|_2^2 - R^2 + 2(P(\mathbf{x}) - D(\mathbf{u}) - R^2) - R^2 \\
&= 2P(\mathbf{x})
\end{aligned}$$

where the inequality follows from (40) and (41). Hence $\mathbf{u}_0 \notin \mathcal{D}_{\mathtt{new}}(\mathbf{x}, \mathbf{u})$ since we have from (38) that any $\mathbf{u}' \in \mathcal{D}_{\mathtt{new}}(\mathbf{x}, \mathbf{u})$ must verify $D(\mathbf{u}') \leq P(\mathbf{x})$.

## Appendix B
## Proof of (22)

Let $\mathbf{u}' \in \mathcal{D}_{\mathtt{gap}}(\mathbf{x}, \mathbf{u})$. By definition, we thus have

$$\|\mathbf{u}' - \mathbf{c}\|_2 \leq R \tag{42}$$

$$\langle \mathbf{g}, \mathbf{u}' - \mathbf{c}\rangle \leq \mathrm{gap}(\mathbf{x}, \mathbf{u}) - R^2 \tag{43}$$

where $\mathbf{c}$, $R$ and $\mathbf{g}$ are the parameters of the GAP dome defined in (19)-(20). We want to show that

$$\|\mathbf{u}' - \mathbf{u}\|_2 \leq \sqrt{2\mathrm{gap}(\mathbf{x}, \mathbf{u})}. \tag{44}$$

This follows from

$$\begin{aligned}
\|\mathbf{u}' - \mathbf{u}\|_2^2 &= \|\mathbf{u}' - \mathbf{c}\|_2^2 + 2\langle \mathbf{c} - \mathbf{u}, \mathbf{u}' - \mathbf{c}\rangle + \|\mathbf{c} - \mathbf{u}\|_2^2 \\
&\leq R^2 + 2\left(\mathrm{gap}(\mathbf{x}, \mathbf{u}) - R^2\right) + R^2 \\
&= 2\,\mathrm{gap}(\mathbf{x}, \mathbf{u})
\end{aligned} \tag{45}$$

where the inequality follows from (42), (43) and the fact that $\mathbf{c} - \mathbf{u} = \mathbf{y} - \mathbf{c}$.