On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising

Patrick J C Tardivel, Malgorzata Bogdan

On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising

Patrick J.C. Tardivel[a*] and Małgorzata Bogdan[a,b],

[a] Institute of Mathematics, University of Wrocław, Wrocław, Poland

[b] Department of Statistics, Lund University, Lund, Sweden

Abstract

In the high-dimensional regression model $Y = X\beta + \varepsilon$, we provide new theoretical results on the probability to recover the sign of β by the Least Absolute Selection and Shrinkage Operator (LASSO) and by the thresholded LASSO.

It is well known that "irrepresentability" is a necessary condition for LASSO to recover the sign of β with a large probability. In this article we extend this result by providing a tight upper bound for the probability of LASSO sign recovery. This upper bound is smaller than 1/2 when the irrepresentable condition does not hold and thus generalizes Theorem 2 of Wainwright [27]. The bound is attained when non-null components of β tend to infinity and its value, which depends on the tuning parameter λ. The bound is reached and its value, which depends on the tuning parameter λ, is the limit of the probability that every null component of β is correctly estimated at 0 when the non-null components of β tend to infinity. The bound can be subsequently used to select λ for LASSO, so as to control the probability that the set of non-null components of LASSO estimator is contained in the set of non-null components of β.

While "irrepresentability" is a stringent condition to recover the sign of β by LASSO, it can be substantially relaxed when LASSO estimates are additionally filtered out with an appropriately selected threshold. Indeed, it is well known that LASSO estimates are consistent under much weaker assumptions than the irrepresentability condition. In this article we provide new theoretical results on LASSO and thresholded LASSO in the asymptotic setup under which X is fixed and non-null components of β tend to infinity. Apart from LASSO, our results cover also Basis Pursuit DeNoising (BPDN). Compared to the classical asymptotics, where X is a $n \times p$ matrix and both n and p tend to $+\infty$, our approach allows for reduction of the technical burden. In the result our main Theorem takes a simple form:

When non-null components of β are sufficiently large, appropriately thresholded LASSO or thresholded BPDN can recover the sign of β if and only if β is *identifiable* with respect to the

*Corresponding author: tardivel@math.uni.wroc.pl

L_1 **norm, i.e.**

$$\textbf{If} \quad X\gamma = X\beta \ \textbf{and} \ \gamma \neq \beta \ \ \textbf{then} \ \|\gamma\|_1 > \|\beta\|_1.$$

To illustrate our results we present examples of *irrepresentability* and *identifiability* curves for some selected design matrices X. These curves provide the proportion of k sparse vectors β for which the *irrepresentability* and *identifiability* conditions hold. Our examples illustrate that "irrepresentability" is a much stronger condition than "identifiability", especially when the entries in each row of X are strongly correlated. Finally, we illustrate how the knockoff methodology [1, 8] allows to select an appropriate threshold and that thresholded BPDN and LASSO can recover the sign of β with a larger probability than adaptive LASSO [31].

Keywords: Multiple regression, Basis Pursuit, LASSO, Sparsity, Active set estimation, Sign estimation, Identifiability condition, Irrepresentability condition

1 Introduction

Let us consider the high-dimensional linear model

$$Y = X\beta + \varepsilon, \tag{1}$$

where $X = (X_1|\dots|X_p)$ is a $n \times p$ design matrix, with $n \leq p$, ε is a random vector in $\mathbb{R}^n$, and $\beta \in \mathbb{R}^p$ is an unknown vector of regression coefficients. The sign vector of β is $S(\beta) = (S(\beta_1), \dots S(\beta_p)) \in \{-1, 0, 1\}^p$, where for $x \in \mathbb{R}$, $S(x) = \mathbf{1}_{x>0} - \mathbf{1}_{x<0}$. Our main purpose is to recover $S(\beta)$. This objective is slightly more general than the aim of recovering the active set, $\text{supp}(\beta) := \{i \in \{1, \dots, p\} \mid \beta_i \neq 0\}$. The sign of β can be estimated by the sign of the well known LASSO estimator [25]:

$$\widehat{\beta}^{\text{L}} := \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \ \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1. \tag{2}$$

When $\text{rank}(X) = n$, an alternative formulation of LASSO is provided by the Basis Pursuit DeNoising (BPDN) estimator [9]:

$$\widehat{\beta}^{\text{BPDN}} := \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \ \|b\|_1 \text{ subject to } \|Y - Xb\|_2^2 \leq R. \tag{3}$$

Given a particular vector $Y \in \mathbb{R}^n$, there is a one to one correspondance between the tuning parameter $\lambda > 0$ and the regularization parameter $R > 0$, under which LASSO and BPDN estimates take the same value (see *e.g* page 64 of [17] or the chapter 5.3 of [3]). For example, when $\lambda = \|X'Y\|_\infty$ and when $R = \|Y\|_2^2$ then both LASSO and BPDN estimators are equal to $\mathbf{0}$. However, the relationship between λ and R depends on the specific realization of Y and, in broad generality, given a fixed $\lambda > 0$ for LASSO, we cannot pick a fixed $R > 0$

for BPDN under which these both estimators equal. Thus, BPDN and LASSO are not equivalent estimators. The Basis Pursuit (BP) estimator, solution of (3) when $R = 0$, is a particular case of BPDN. As discussed e.g. in [11, 15], BP can be thought of as the limit of LASSO when the tuning parameter λ tends to 0.

1.1 Sign recovery by LASSO

Properties of the LASSO sign estimator $S(\widehat{\beta}^{\mathrm{L}}(\lambda)) := \left(S(\widehat{\beta}_1^{\mathrm{L}}(\lambda)), \dots, S(\widehat{\beta}_p^{\mathrm{L}}(\lambda))\right)$ (or properties of the active set estimator $\mathrm{supp}(\widehat{\beta}^{\mathrm{L}}(\lambda)) := \{i \in \{1, \dots, p\} \mid \widehat{\beta}_i(\lambda) \neq 0\}$) have been intensively studied [16, 20, 27, 30, 31]. Specifically, Zhao and Yu [30] and Zou [31] consider the asymptotic setup under which n tends to $+\infty$ and p is fixed and observe that LASSO can recover $S(\beta)$ only if the restrictive irrepresentable condition is fulfilled. These results were further extended to the case of the fixed design matrix X, where the irrepresentable condition is formulated as follows:

Definition 1 (Irrepresentability condition) *Let $b \in \mathbb{R}^p$, $I := \{i \in \{1, \dots, p\} \mid b_i \neq 0\}$, and $X_I, X_{\overline{I}}$ be the matrices whose columns are respectively $(X_i)_{i \in I}$ and $(X_i)_{i \notin I}$. The vector b satisfies the irrepresentable condition if $\ker(X_I) = \mathbf{0}$ and $\|X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(b_I)\|_\infty \leq 1$.*

According to the Theorem 2 of Wainwright [27], the irrepresentability condition is necessary to recover $S(\beta)$ with high probability. Indeed, when $\ker(X_I) = \mathbf{0}$, $\|X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I)\|_\infty > 1$ and both ε and $-\varepsilon$ have the same distribution, then for any selection of the tuning parameter $\lambda > 0$, $\mathbb{P}(S(\widehat{\beta}^{\mathrm{L}}(\lambda)) = S(\beta)) \leq 1/2$. This result holds also in the noiseless case when $\varepsilon = \mathbf{0}$, where the probability to recover $S(\beta)$ is equal to zero. Moreover, Bühlmann and van de Geer [5] (page 192-194) showed that, in the noiseless case, when the irrepresentability strictly holds (*i.e* when $\|X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I)\|_\infty < 1$) then the non-random set $\mathrm{supp}(\beta^L(\lambda))$ recovers $\mathrm{supp}(\beta)$ as soon as non-null components of β are sufficiently large. The proof provided in [5] can be easily adapted for the sign recovery.

In this article we provide a new theoretical result on the sign recovery by LASSO. Specifically, Theorem 1 in Section 2 provides an upper bound for the probability to recover the sign of β which depends from $X, S(\beta), \lambda$ and the distribution of ε. The formula for the bound is not analytic but its value can be well approximated by simple Monte Carlo simulations. We also show that the bound is attained when non-null components of β tend to infinity and its value is the limit of the probability that $\mathrm{supp}(\hat{\beta}^L) \subset \mathrm{supp}(\beta)$. Therefore, the bound can be used to select λ in order to control the probability that at least one null component of β is not estimated at 0 (the Family Wise Error Rate, FWER), when non-zero elements of β are sufficiently large. Moreover, as shown in our simulation study, in many examples FWER increases with the magnitude of non-zero elements of β. Consequently, in such cases our selected λ allows to control FWER independently of the magnitude of the non-null components of β.

1.2 Sign recovery by thresholded LASSO

It is clear that in the noiseless case, the following identifiability condition is necessary and sufficient to recover $S(\beta)$ by the non-random basis pursuit.

Definition 2 (Identifiability condition) *The vector $b \in \mathbb{R}^p$ is identifiable with respect to the design matrix X and the L_1 norm (or just identifiable with respect to the L_1 norm) if the following implication holds*

$$X\gamma = Xb \text{ and } \gamma \neq b \Rightarrow \|\gamma\|_1 > \|b\|_1. \tag{4}$$

Under the identifiability assumption, β is sparse. Indeed, Lemma 3 in Tardivel et al. [24] shows that $k = \text{card}\{i \in \{1, \ldots, p\} \mid \beta_i \neq 0\} \leq n$, i.e. β has at least $p - n$ zeros. On the other hand some assumptions on the sparsity of β assure that β is identifiable with respect to the L_1 norm. For example when $\|X_1\|_2 = \cdots = \|X_p\|_2 = 1$ and the number of nonzero elements of β satisfies the following inequality (called mutual coherence condition)

$$k = \text{card}\{i \in \{1, \ldots, p\} \mid \beta_i \neq 0\} \leq \frac{1}{2}\left(1 + \frac{1}{M}\right), \text{ where } M := \max_{i \neq j} |\langle X_i, X_j \rangle| \ , \tag{5}$$

then β is identifiable with respect to the L_1 norm [13, 17, 19]. When entries of X are i.i.d $\mathcal{N}(0, 1)$ and n, p are both very large, the phase transition curve of Donoho and Tanner [14] provides, with respect to the ratio n/p, an upper bound on k/n so that β having a sparsity k is identifiable with respect to the L_1 norm.

According to the Theorem 2, reported in Section 3, for any value of the tuning parameter λ or the regularization parameter R, the identifiability condition is sufficient and necessary so that LASSO or BPDN estimator allow to separate asymptotically negative, null, and positive components of β. This means that, when non null components of β are sufficiently large, appropriately thresholded LASSO or BPDN can properly identify the sign of β if and only if the identifiability condition holds for β.

1.3 Graphical illustrations of main results

By definition, the irrepresentability condition depends only on $S(\beta)$ and not on how large are the non-null components of β. Moreover, as claimed in the Proposition 2, the identifiability condition also depends only on $S(\beta)$. Thus, the comparison of these two conditions can be performed by considering vectors of parameters such that $\beta = S(\beta)$. In Figure 1, we provide the irrepresentability and the identifiability curves for a selected matrix X of dimension 100×300, whose elements were independently drawn from the normal $\mathcal{N}(0, 1)$ distribution. These curves provide the proportion of the sign vectors with k nonzero elements which satisfy the identifiability condition or irrepresentability condition. Figure 1 illustrates that the identifiability curve is above the irrepresentability curve. This observation is not surprising since, according to the Proposition 1, the irrepresentability condition implies the identifiability condition. Based on these curves, when non null components of β are suffi-

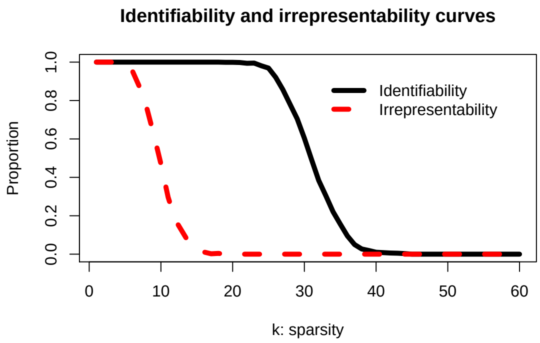

Figure 1: This figure provides the identifiability and irrepresentability curves for the design matrix X of dimension 100×300, whose entries were independently generated from $\mathcal{N}(0,1)$ distribution. The x-axis represents the sparsity k and the y-axis represents the proportion of sign vectors satisfying the identifiability condition (resp. irrepresentability condition)

ciently large, one can expect that LASSO allows to recover the sign of β when the sparsity k is smaller than 5, while thresholded LASSO allows to recover the sign of β when $k \leq 25$.

Figure 2 illustrates Theorem 1, which provides an upper bound for the probability of LASSO sign recovery. In this Figure we use the same design matrix X as in Figure 1 and the noise ε is a standard Gaussian vector. According to the irrepresentability curve provided in Figure 1, the irrepresentability condition holds when $k \leq 5$. Thus, when $k \leq 5$, one can select the tuning parameter λ in order to obtain any fixed value for this bound. In our experiment the value of λ was selected so that the average value of the bound over 1000 randomly sampled vectors β with $k = 5$ non-zero elements is equal to 0.95. The y axis in Figure 2 represents the probability of recovering $S(\beta)$ calculated based on 1000 randomly sampled vectors β having $k = 5$ non-null components, which are all equal to $t > 0$. Figure 2 shows that indeed, the upper bound for LASSO sign recovery is reached when non null components of β tend to $+\infty$ and that the selected λ allows to control the FWER below 0.05 for the whole range of the magnitudes of β.

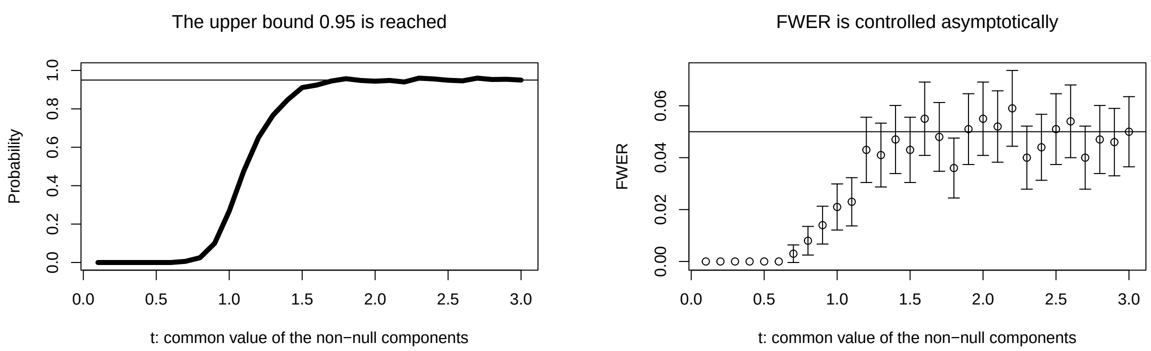

Figure 2: When $k = 5$, the left figure provides the probability to recover $S(\beta)$ with LASSO sign estimator and the right figure provides the FWER: the probability that at least one null component of β is selected by LASSO estimator. The x-axis represents $t > 0$ which measure how large are the non-null components of β (in both figures), the y-axis represents the probability of the sign recovery (left figure) and the FWER (right figure). The horizontal lines correspond to $y = 0.95$ (left figure) and $y = 0.05$ (right figure).

Finally, Figure 3 illustrates Theorem 2 which states that when non null components of β are large enough, "identifiability" is a sufficient condition under which appropriately thresholded BPDN and thresholded LASSO can recover $S(\beta)$.

In this figure we present results for $k = 20$, for which the irrepresentability condition does not hold (see Figure 1). Consequently, the probability to recover $S(\beta)$ is theoretically smaller than 1/2 and more precisely, the empirical value of this probability is almost 0. On the other hand, due to a fact that for $k = 20$ the identifiability condition holds, we expect that thresholded LASSO and thresholded BP can recover $S(\beta)$ when the magnitude of β is large enough. In Figure 3 the y axis represents the probability of recovering $S(\beta)$ by thresholded LASSO and thresholded BP, calculated based on 1000 randomly sampled vectors β having $k = 20$ non-null components, which are all equal to $t > 0$. The threshold was selected so as to control the FWER at the level 0.05. For both of these procedures, in order to pick a threshold, we approximate the distribution of LASSO (resp. BP) estimators associated to null components of β using control variables created according to the knockoff methodology [1] (see Section 5 for details). In our case, a control variable is just a column added to the design matrix and generated according to a standard Gaussian $\mathcal{N}(0, Id_n)$ distribution. Figure 3 shows that, indeed, both thresholded BP and thresholded LASSO can recover $S(\beta)$ with probability converging to 0.95 when non null components of β increase in magnitude.

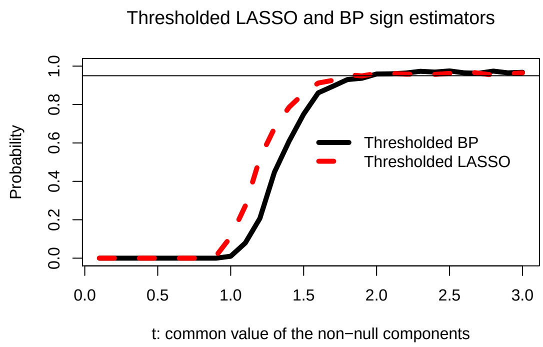

Figure 3: When $k = 20$, this figure provides the probability to recover $S(\beta)$ with thresholded LASSO and thresholded BP. The x-axis represents $t > 0$ which measure how large are the non-null components of β and the y-axis represents the sign recovery probability.

1.4 Organization of the article

In Section 2 we formulate and discuss Theorem 1, which provides a tight upper bound for LASSO sign estimator to recover $S(\beta)$. This sections contains also Proposition 1, which shows that the irrepresentability condition is stronger than the identifiability condition. In Section 3, Theorem 2 shows that identifiability is a necessary and suficient condition for LASSO to separate the non-null components of β from the noise and to recover asymptotically $S(\beta)$ with thresholded LASSO and thresholded BPDN. In Section 4, Proposition 2 shows that identifiability condition depends only on $S(\beta)$ and not on the magnitude of the non-null components of β. Here we also introduce the irrepresentability and identifiability curves which provide respectively the proportion of sign vectors satisfying the irrepresentability condition and identifiability condition. Section 5 is devoted to numerical experiments which illustrate that sign estimators derived from the thresholded LASSO and thresholded BPDN can be better than sign estimators derived from LASSO and adaptive LASSO and that knockoff methodology allows for the appropriate selection of the threshold for both of these methods.

1.5 Notations and assumptions

In this article we always assume that the design matrix X is in general condition (see *e.g* [26] or the supplementary material for this manuscript). This assumption assures that the minimizer of (2) (resp. minimizer of (3)) is unique and thus that the LASSO estimator (resp. BPDN estimator) is well defined. This assumption is very weak and generically holds. Indeed, when X is a random matrix such that the entries $(X_{11}, X_{12}, \ldots, X_{np})$ have a density on $\mathbb{R}^{np}$ then, almost surely, X is in general position [26].

The main notation used in the subsequent sections is as follows:

- Let I be the subset of $\{1, \ldots, p\}$. We denote by $\overline{I}$ the complement of I, namely $\overline{I} := \{1, \ldots, p\} \setminus I$.
- The notation X_I represents a matrix whose columns are indexed by the elements of I: $(X_i)_{i\in I}$.
- For $b \in \mathbb{R}^p$, b_I denotes the sub-vector containing elements of b with indices in I.
- Symbols $\text{supp}(b), \text{supp}^+(b)$ and $\text{supp}^-(b)$ denote respectively the sets $\{i \in \{1, \ldots, p\} \mid b_i \neq 0\}, \{i \in \{1, \ldots, p\} \mid b_i > 0\}$ and $\{i \in \{1, \ldots, p\} \mid b_i < 0\}$.
- LASSO and BPDN estimators depend on X, β, r, ε and on the tuning parameter $\lambda > 0$ or the regularization parameter $R \geq 0$. When it is useful, we use the parentheses to recall these dependencies. The estimator $\widehat{\beta}$ represents indistinctly the LASSO estimator or the BPDN estimator.

To formulate our asymptotic results we will often consider a sequence of regression parameters $\beta^{(r)}_{r\in\mathbb{N}}$ for which non-null components tend to infinity in the following way.

Assumption 1

1) *The sign of $\beta^{(r)}_{r\in\mathbb{N}}$ is invariant namely, there exists a sign vector $s^0 \in \{-1, 0, 1\}^p$ such that whatever $r \in \mathbb{N}, S(\beta^{(r)}) = s^0$.*

2) *The following limit holds* $\lim_{r\to+\infty} \min\{|\beta^{(r)}_i|, i \in \text{supp}(s^0)\} = +\infty$

3) *There exists $q > 0$ such that*

$$\forall r \in \mathbb{N}, \frac{\min\{|\beta^{(r)}_i|, i \in \text{supp}(s^0)\}}{\|\beta^{(r)}\|_\infty} \geq q.$$

According to the Proposition 2 the identifiability condition just depends from the sign, that is the reason why in **1)** the sign of $\beta^{(r)}$ is fixed. Consequently, whatever $r \geq 0$, $\beta^{(r)}$ is identifiable with respect to the L_1 norm if and only if s^0 is identifiable with respect to the L_1 norm. The condition **2)** means that non-null components of $\beta^{(r)}$ tend to $+\infty$. The condition **3)** exclude sequences $(\beta^{(r)})_{r\in\mathbb{N}}$ for which the smallest non-null component of $\beta^{(r)}$ is asymptotically infinitely smaller than the largest component of $\beta^{(r)}$.

2 Sign recovery with LASSO sign estimator

In this section we formulate Theorem 1, which provides an upper bound for the probability to recover $S(\beta)$ with LASSO estimator. When β is identifiable with respect to the L_1 norm, this upper bound is reached asymptotically when $\min\{|\beta_i|, i \in \mathrm{supp}(\beta)\}$ tends to $+\infty$.

Theorem 1 *Let $I := \mathrm{supp}(\beta)$ and let $X_I, X_{\overline{I}}$ be matrices whose columns are $(X_i)_{i\in I}$ and $(X_i)_{i\notin I}$, respectively. Additionally, let $\zeta_{X,\lambda,S(\beta)}$ be the random vector $\zeta_{X,\lambda,S(\beta)} = X'_{\overline{I}}X_I(X'_IX_I)^{-1}S(\beta_I)+\frac{1}{\lambda}X'_{\overline{I}}\left(Id - X_I(X'_IX_I)^{-1}X'_I\right)\varepsilon$ and let us assume that $\ker(X_I) = \mathbf{0}$.*

Upper bound: *The following upper bound for the sign recovery holds.*

$$\mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda)) = S(\beta)\right) \leq \mathbb{P}\left(\left\|\zeta_{X,\lambda,S(\beta)}\right\|_\infty \leq 1\right) = \bar{\gamma}.$$

Now, let $(\beta^{(r)})_{r\in\mathbb{N}}$ be a sequence of $\mathbb{R}^p$ satisfying the assumption 1 and let us remind that the sign $\beta^{(r)}$ does not depend from r and is equal to s^0 and let $I = \mathrm{supp}(s^0)$. If s^0 is identifiable with respect to the L_1 norm then the following asymptotic results hold.

Sharpness of the upper bound: *Asymptotically, the upper bound is reached.*

$$\limsup_{r\to+\infty}\mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda,r)) = s^0\right) \leq \bar{\gamma},$$
$$\liminf_{r\to+\infty}\mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda,r)) = s^0\right) \geq \mathbb{P}\left(\left\|\zeta_{X,\lambda,S(\beta)}\right\|_\infty < 1\right) = \gamma.$$

Asymptotic Full power and asymptotic control of the FWER: *Asymptotically, the power is equal to 1 and the FWER is controlled at the level $1-\gamma$.*

$$\lim_{r\to+\infty}\mathbb{P}\left(\forall i \in I, S(\widehat{\beta}^{\mathrm{L}}_i(\lambda,r)) = s^0_i\right) = 1,$$
$$\limsup_{r\to+\infty}\mathbb{P}\left(\exists i \notin I, \widehat{\beta}^{\mathrm{L}}_i(\lambda,r) \neq 0\right) \leq 1-\gamma,$$
$$\liminf_{r\to+\infty}\mathbb{P}\left(\exists i \notin I, \widehat{\beta}^{\mathrm{L}}_i(\lambda,r) \neq 0\right) \geq 1-\bar{\gamma}.$$

Remark 1 *Results given in Theorem 1 are quite straightforward when X is orthogonal (i.e. when $X'X = Id_p$). Indeed, in this case the upper bound $\bar{\gamma}$ is just the probability that null components of β are simultaneously estimated at* 0 *namely* $\bar{\gamma} = \mathbb{P}(\forall i \notin \text{supp}(\beta), \hat{\beta}_i^{\text{L}}(\lambda) = 0)$.

Remark 2 *Theorem 1 immediately implies Theorem 2 of Wainwright [27] which claims that when ε and $-\varepsilon$ have the same distribution and when $\|X'_{\bar{I}}X_I(X'_IX_I)^{-1}S(\beta_I)\|_\infty > 1$ then, for any $\lambda > 0$, the probability of the sign recovery by LASSO is always smaller than* $1/2$.

The bound for the probability of recovering $S(\beta)$ provided in Theorem 1 is not analytic but can be empirically computed using Monte Carlo simulations (e.g. see Figure 1). When the irrepresentable condition strictly holds for s^0, namely when $\|X'_{\bar{I}}X_I(X'_IX_I)^{-1}s^0_I\|_\infty < 1$, the tuning parameter λ can be selected to fix γ at an arbitrary level in $(0,1)$ (e.g. see Figure 1). Because the irrepresentable condition implies the irrepresentable condition (as claimed in Proposition 1) such a tuning parameter allows to control asymptotically the FWER at level $1-\gamma$ when non-null components tends to $+\infty$. To our knowledge, Theorem 1 is the first theoretical result providing a guide on how to select the tuning parameter in order to control a type I error at a specified level for a given design matrix X.

Proposition 1 *Let X be a $n \times p$ matrix with $n \leq p$ in general position, let $\beta \in \mathbb{R}^p$, let $I := \text{supp}(\beta)$ and let us assume* $\ker(X_I) = \mathbf{0}$. *If $\|X'_{\bar{I}}X_I(X'_IX_I)^{-1}S(\beta^0_I)\|_\infty \leq 1$, then the parameter β is identifiable with respect to the L_1 norm.*

Let us notice that when the inequality in the irrepresentable condition is strict instead of large, the Theorem 1 remains true without assuming that X is in general position. The proof of the proposition 1 given in this article is the one reported in the PhD manuscript of Tardivel [23].

3 Identifiability is a necessary and sufficient condition for sign recovery

When β does not satisfy the irrepresentable condition then, whatever $\lambda > 0$ and even if non-null components of β are infinitely large, the LASSO sign estimator $S(\widehat{\beta}^{\text{L}}(\lambda))$ fails to recover $S(\beta)$. However, the irrepresentable condition is not an unsurpassable limitation to recover $S(\beta)$. Actually the Theorem 2 shows that an appropriately thresholded LASSO (resp. thresholded BPDN) recover asymptotically $S(\beta)$ under the identifiability condition on β (which is, by the proposition 1, weaker than the irrepresentability condition). We remind that, $\widehat{\beta}$ represents indistinctly the LASSO or BPDN estimator with a fixed tuning parameter $\lambda > 0$ or with a fixed regularization parameter $R \geq 0$.

Theorem 2 *Let X be a $n \times p$ matrix in general position such that* $\mathrm{rank}(X) = n$*, let $\beta^{(r)}$ be a sequence of parameter satisfying assumption 1. If s^0 is identifiable with respect to the L_1 norm then for any fixed $\varepsilon \in \mathbb{R}^n$ and sufficiently large $r > r_0(\varepsilon)$ the following inequality holds*

i)

$$\mathrm{supp}^-(s^0) \subset \mathrm{supp}^-(\widehat{\beta}_i(\varepsilon, r)) \text{ and } \mathrm{supp}^+(s^0) \subset \mathrm{supp}^+(\widehat{\beta}_i(\varepsilon, r)).$$

ii)

$$\max_{i \in \mathrm{supp}^-(s^0)} \left\{\widehat{\beta}_i(\varepsilon, r)\right\} < \min_{i \notin \mathrm{supp}(s^0)} \left\{\widehat{\beta}_i(\varepsilon, r)\right\} \leq \max_{i \notin \mathrm{supp}(s^0)} \left\{\widehat{\beta}_i(\varepsilon, r)\right\} < \min_{i \in \mathrm{supp}^+(s^0)} \left\{\widehat{\beta}_i(\varepsilon, r)\right\}.$$

This inequalities mean that when s^0 is identifiable with respect to the L_1 norm and when r is large, the estimator $\widehat{\beta}(\varepsilon, r)$ separates negative components of s^0 (i.e $i \in \mathrm{supp}^-(s^0)$), null components of s^0 (i.e $i \notin \mathrm{supp}(s^0)$) and positive components of s^0 (i.e $i \in \mathrm{supp}^+(s^0)$).

If s^0 is not identifiable with respect to the L_1 norm then whatever $r \geq 0$ the property i) does not hold namely

$$\forall r \in \mathbb{N}, \mathrm{supp}^-(s^0) \not\subset \mathrm{supp}^-(\widehat{\beta}_i(\varepsilon, r)) \text{ or } \mathrm{supp}^+(s^0) \not\subset \mathrm{supp}^+(\widehat{\beta}_i(\varepsilon, r).$$

Consequently, when s^0 is not identifiable with respect to the L_1 norm, then sign estimators derived from thresholded LASSO or thresholded BPDN cannot recover s^0.

Let us notice that the assumptions on X are very weak and generically hold when $n \leq p$. The assumption $\mathrm{rank}(X) = n$ assures that, whatever $R \geq 0$, the BPDN estimator is well defined. The general position condition assures the uniqueness of both LASSO and BPDN estimators (see *e.g* the proposition 1 given in supplementary material for a proof).

Because the almost sure convergence (and thus the convergence for every fixed ε) implies the convergence in probability when s^0 is identifiable with respect to the L_1 norm, according to the Theorem 2, the following convergences in probability hold

$$\begin{aligned}
1 &= \lim_{r \to +\infty} \mathbb{P}\left(\mathrm{supp}^-(s^0) \subset \mathrm{supp}^-(\widehat{\beta}_i(r)) \text{ and } \mathrm{supp}^+(s^0) \subset \mathrm{supp}^+(\widehat{\beta}_i(r))\right), \\
1 &= \lim_{r \to +\infty} \mathbb{P}\left(\max_{i \in \mathrm{supp}^-(s^0)} \left\{\widehat{\beta}_i(r)\right\} < \min_{i \notin \mathrm{supp}(s^0)} \left\{\widehat{\beta}_i(r)\right\} \leq \max_{i \notin \mathrm{supp}(s^0)} \left\{\widehat{\beta}_i(r)\right\} < \max_{i \in \mathrm{supp}^+(s^0)} \left\{\widehat{\beta}_i(r)\right\}\right).
\end{aligned}$$

The Theorem 2 stress that one cannot recover $S(\beta)$ with a sign estimator derived from LASSO or BPDN when β is not identifiable with respect to the L_1 norm. When β is identifiable with respect to the L_1 norm, the Theorem 2 suggest to recover $S(\beta)$ by deriving sign estimators from the thresholded LASSO or thresholded BPDN. Expressions of these thresholded estimators are reported in (6) given below. By the the separation property, one knows that it remains to select a good threshold τ to construct a consistent sign estimator (with

τ depending from r for the consistency).

The Theorem 2 confirms recent results given by Bogdan et al. [4]. Indeed, if X has i.i.d $\mathcal{N}(0,1)$ entries, $n/p \to \delta \in (0,1)$ and if asymptotically the point $(\text{card}(\text{supp}(\beta))/n, n/p)$ is below the asymptotic phase transition curve [12] (*i.e.* if β is asymptotically identifiable with respect to the L_1 norm) then the thresholded LASSO almost surely recovers $S(\beta)$ (as soon as non-null components of β are large enough).

In the following section, we are going to give some properties about identifiability condition. In particular, we show that identifiability (as irrepresentability) just depends from $S(\beta)$ and not on how large are the non null components of β.

4 Identifiability and irrepresentability sign applications

By definition the irrepresentable condition just depends from the sign of β. Given a particular design matrix X, the irrepresentability sign application is defined hereafter.

Irrepresentabity sign application:

$$\Phi_{\text{IC}}^X : s \in \{-1,0,1\}^p \mapsto \begin{cases} 1 \text{ if } s = (0,\dots,0) \\ 1 \text{ if } \ker(X_I) = \mathbf{0} \text{ and } \|X_{\overline{I}}' X_I (X_I' X_I)^{-1} s_I\|_\infty \leq 1 \text{ where } I := \text{supp}(s) \\ 0 \text{ otherwise} \end{cases}.$$

Such a sign application provides the limitation of the LASSO sign estimator to recover $S(\beta)$. Indeed, if $\phi_{\text{IC}}^X(S(\beta)) = 0$ then $S(\beta)$ cannot be recovered with the LASSO sign estimator even if non-null components of β are extremely large. The proposition 2 shows that the identifiability condition just depends from $S(\beta)$ and not how large are the non-null components of β.

Proposition 2 *Let $b \in \mathbb{R}^p$ be identifiable with respect to X and the L_1 norm and let $\tilde{b} \in \mathbb{R}^p$ such that $S(b) = S(\tilde{b})$ then $\tilde{b}$ is identifiable with respect X and to the L_1 norm.*

Given a particular design matrix X, the identifiability sign application is defined hereafter.

Identifiability sign application:

$$\Phi_{\text{Idtf}}^X : s \in \{-1,0,1\}^p \mapsto \begin{cases} 0 \text{ if } s \neq \underset{b \in \mathbb{R}^p}{\text{argmin}} \|b\|_1 \text{ subject to } Xb = Xs \\ 1 \text{ otherwise} \end{cases}.$$

Such a sign application for the identifiability condition provides the limitation of sign estimators derived from thresholded LASSO and thresholded BPDN to recover $S(\beta)$. Indeed, if $\phi_{\text{Idtf}}^X(S(\beta)) = 0$ then thresholded LASSO (resp. thresholded BPDN) sign estimator cannot recovered $S(\beta)$ even if non-null components of β are extremely large.

According to the proposition 2 given supplementary material when $(X_i)_{i\in\text{supp}(\beta)}$ is not linearly independent then β does not satisfy the identifiability condition. Consequently, when $\text{card}(\text{supp}(\beta)) > n$ then $\phi^X_{\text{IC}}(S(\beta)) = \phi^X_{\text{Idtf}}(S(\beta)) = 0$. Let us provides some basic properties and comments about these sign applications.

1. These two sign applications are even.
2. Due to the proposition 1, whatever $s \in \{-1, 0, 1\}^p$, $\Phi^X_{\text{IC}}(s) \leq \Phi^X_{\text{Idtf}}(s)$.
3. The computation of Φ^X_{IC} is a straightforward matricial computation; the computation of Φ^X_{Idtf} is no more difficult and need to solve a basis pursuit problem.

The last remark shows that given a parameter $\beta \in \mathbb{R}^p$, it is easy to check weather or not β is identifiable with respect to the L_1 norm.

Given a sparsity k, the identifiability (resp. irrepresentability) curve provides the proportion of sign vectors satisfying the identifiability condition (resp. irrepresentability condition). These curves illustrate that the identifiability condition is much weaker than the irrepresentability condition and thus emphasize the theoretical result given in the proposition 1.

4.1 Illustrations of identifiability and irrepresentability curves

The number of sign vectors is very huge (3^p), that is why we are not going to provide explicitly Φ^X_{Idtf} and Φ^X_{IC} for each sign vector. Instead, for each sparsity $k \in \{1, \dots, n\}$, we are going to compute empirically $p^X_{\text{Idtf}}(k) := \mathbb{E}_U(\Phi^X_{\text{Idtf}}(U))$ and $p^X_{\text{IC}}(k) := \mathbb{E}_U(\Phi^X_{IC}(U))$ where U is a uniformly distributed on $\{u \in \{-1, 0, 1\}^p \mid \text{card}(\text{supp}(u)) = k\}$. The identifiability and irrepresentability curves represents respectively the curves of the functions $k \in \{1, \dots, n\} \mapsto p^X_{\text{Idtf}}(k)$ and $k \in \{1, \dots, n\} \mapsto p^X_{\text{IC}}(k)$. In the numerical experiments given in the figure 4, X is a Gaussian matrix described hereafter.

Setting 1: The matrix X is a $n \times p$ matrix with $n = 100$, $p = 300$ and $(X_{ij})_{1\leq i\leq n, 1\leq j\leq p}$ are i.i.d $\mathcal{N}(0, 1)$.

Setting 2: The matrix X is a $n \times p$ matrix with $n = 100$, $p = 300$ and the vectors $(X_{ij})_{1\leq j\leq p}$ where $i \in \{1, \dots, n\}$ is a family of i.i.d Gaussian vector $\mathcal{N}(\mathbf{0}, \Gamma)$. In this setting Γ is a $p \times p$ matrix where $\Gamma_{ii} = 1$ with $i \in \{1, \dots, p\}$ and $\Gamma_{ij} = 0.9$ when $i \neq j$.

From now on, X is a particular observation of Gaussian matrix as described in setting 1 and setting 2 (by using the R command set.seed(123)).

Surprisingly the two identifiability curves given in the setting 1 and 2 are very similar. *A priori*, we expected to recover a curve in the setting 2 much below than the one given in the setting 1. Indeed, classical conditions implying the identifiability of β with respect to the L_1 norm are the mutual coherence condition (5) and the restricted isometry property [6, 7]. These conditions are quite weak when the family $(X_i)_{1\leq i\leq p}$ is almost

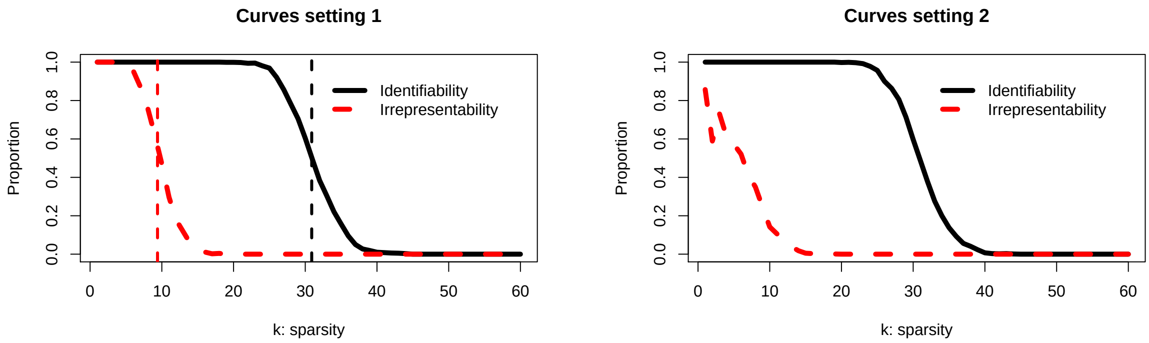

Figure 4: This figure gives the curves of the functions $k \mapsto p^X_{\mathrm{Idtf}}(k)$ and $k \mapsto p^X_{\mathrm{IC}}(k)$ when X is a Gaussian matrix given in the setting 1 (left panel) and setting 2 (right panel). Due to the proposition 1, whatever the sparsity k, $p^X_{\mathrm{Idtf}}(k) \geq p^X_{\mathrm{IC}}(k)$ thus this figure just emphasizes that the identifiability condition is a much weaker assumption than the irrepresentability condition. The vertical lines in the left panel provides, in the setting 1, an asymptotic approximation of the identifiability and irrepresentability curves. Indeed by the Theorem 1 in [14] and the Theorem 1 in [27], when p is very large and $n/p = 1/3$ then the identifiability and irrepresentability conditions hold respectively when $k \leq 0.31n$ and $k \leq 0.09n$. To plot these these curves, for a sparsity k the quantities $p^X_{\mathrm{Idtf}}(k)$ and $p^X_{\mathrm{IC}}(k)$ have been computed by simulating 1000 observations of the random vector U.

orthogonal (as in the setting 1 since $\mathbb{E}(X'X) = nId_n$) but are very strong when $(X_i)_{1\leq i\leq p}$ is far from an orthogonal family (as in the setting 2 since $\mathbb{E}(X'X) = n\Gamma$).

The asymptotic phase transition curve given in Donoho and Tanner [14] provides an approximation of the identifiability curve in the setting 1. Such an approximation is useful when n and p are too much large so that the identifiability curve is too much time expensive to obtain. Unfortunately, to our knowledge, there is not such asymptotic phase transition curve for Gaussian matrices with correlated entries as in the setting 2 (see *e.g.* [21] for more details about asymptotic phase transition curve).

One notices that in the setting 2, the irrepresentability curve is not monotonic in the neighbourhood of 0; it is not a numerical problem. Actually when k is very small, components of U are all positive or all negative with a quite large probability. Furthermore the figure 5 illustrates that, in the setting 2, when the sign vector s is positive componentwise (resp. negative componentwise), the irrepresentable condition becomes a very strong condition. These both remarks, aim at explaining why, in the setting 2, the irrepresentabilty curve is not monotonic. Hereafter, without any loss of generality, we focus on the particular case in which sign vector is positive componentwise. The figure 5 provides the positive irrepresentability and identifiability curves, which are respectively the curves of the functions $k \mapsto p^X_{\mathrm{Idtf}+}(k) := \mathbb{E}_U(\Phi^X_{\mathrm{Idtf}}(U))$ and $k \mapsto p^X_{\mathrm{IC}+}(k) := \mathbb{E}_U(\Phi^X_{IC}(U))$ where U has uniform distribution over the set $\{u \in \{0,1\}^p \mid \mathrm{card}(\mathrm{supp}(u)) = k\}$.

Performance of the sign estimators derived from LASSO, thresholded LASSO and thresholded BPDN depends from the tuning parameter λ, regularization parameter R and threshold τ. In the following section, we are going to prescribe values for these parameters.

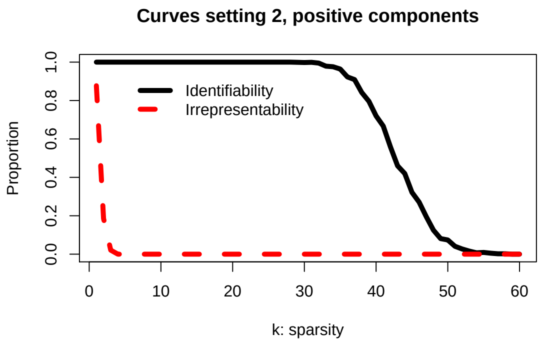

Figure 5: This figure gives the curves of the functions $k \mapsto p^X_{\text{Idtf}+}(k)$ and $k \mapsto p^X_{\text{IC}+}(k)$ when X is a Gaussian matrix given in setting 2. One notices that, with respect to the curves given in the figure 4, the gap between the irrepresentable condition and the identifiability condition becomes larger. When k is small $p^X_{\text{IC}+}(k) \approx p^X_{\text{IC}}(k)$ (more precisely, $p^X_{\text{IC}+}(k) = p^X_{\text{IC}}(k)$ when $k = 1$) and when k is large enough $p^X_{\text{IC}}(k)$ weakly depends from the correlation. This remark aim to explain why, in the setting 2, the function $k \mapsto p^X_{\text{IC}}(k)$ is not monotonic in the neighbourhood of 0. To plot these these curves, for a sparsity k, the quantities $p^X_{\text{Idtf}+}(k)$ and $p^X_{\text{IC}+}(k)$ have been computed by simulating 1000 observations of U.

5 Numerical comparisons of sign estimators

Theorem 2 states that the sign estimators provided by thresholded LASSO or thresholded BPDN allow to recover $\text{sign}(\beta)$ as long as the identifiability condition is satisfied. Another way to recover $\text{sign}(\beta)$ is to use a sign estimator derived from adaptive LASSO. Indeed, as claimed by Theorem 2 of Zou [31], by deriving weights for adaptive LASSO from a consistent estimator of β and by selecting properly the tuning parameter λ, one obtains a sign estimator derived from adaptive LASSO which is consistent for $\text{sign}(\beta)$. Weights for adaptive LASSO can be appropriately derived from LASSO. Indeed, according to Lemma 2, under the identifiability assumption LASSO estimator converges to β. The purpose of this section is to provide a numerical comparison of sign estimators derived from LASSO, thresholded LASSO, thresholded BPDN and adaptive LASSO.

As explained hereafter, there are some recommendations on how to select the tuning parameter $\lambda > 0$ for the LASSO estimator whereas, to our knowledge, there are not recommendation on how to select $R \geq 0$ for BPDN estimator. That is the reason why, for BPDN estimator, we arbitrary set $R = 0$ and thus we only consider BP estimator.

5.1 Selection of the tuning parameter

As explained in [4, 29], a value of the optimal tuning parameter for the sign recovery by thresholded LASSO is substantially smaller than the optimal value of the tuning parameter for LASSO sign estimator. Specifically:

- For LASSO sign estimator, the tuning parameter has to be large enough so that it prevents including false discoveries.
- For thresholded LASSO sign estimator the tuning parameter needs to be selected so as to minimize the

mean square error of the estimation of β. This tuning parameter does not need to be large, since the threshold will allow to correctly estimate at 0 null components of β.

5.1.1 Tuning parameter for LASSO sign estimator

When β has a known sign so that $S(\beta)$ satisfies the irrepresentable condition, by Proposition 1, one may pick a tuning parameter λ_L so that $\mathbb{P}(S(\hat{\beta}(\lambda_L)) = S(\beta))$ is smaller than a given value (say 0.95). According to the irrepresentability curve associated to the matrix X described in the setting 1 of the subsection 4.1, the irrepresentability condition is satisfied with probability close to 1 when β contains $k = 5$ nonzero elements. Thus in this setting, we can chose λ_L such that the average value of the upper-bound given in Proposition 1 is equal to 0.95. In other words, λ_L is chosen so that $\mathbb{E}_S(\zeta_{X,\lambda_L,S}) = 0.95$, where S is a random sign vector having a uniform distribution over the set $\{s \in \{-1,0,1\}^p \mid \mathrm{card}(\mathrm{supp}(s)) = 5\}$. The computation of this value gives $\lambda_L = 81.18$. Since under the remaining scenarios of our simulation study the irrepresentability condition is typically not satisfied and thus the FWER can not be controlled at a low level, we decided to use the same value $\lambda_L = 81.18$ for all our simulations.

5.1.2 Tuning parameter for thresholded LASSO sign estimator

When X is the gaussian matrix with independent entries the tuning parameter was selected with the help of the asymptotic theory of Approximate Message Passing (AMP) algorithm for LASSO, provided e.g. in [2, 4, 22]. In the set-up of this theory the design matrix is Gaussian with i.i.d $\mathcal{N}(0, 1/\sqrt{n})$ and components of β are i.i.d random variables having $\Pi = (1-\gamma)\delta_0 + \gamma\Pi^\star$ mixture distribution, where δ_0 and $\Pi^\star$ are point mass at 0 distribution and an arbitrary distribution. The number of observation n, the number of explanatory variables p becomes infinity large and $n/p \to \delta > 0$. The tuning parameter λ_{AMP}, depending on $\delta, \gamma, \Pi^\star$, is selected so as to minimize the asymptotic mean square error according to the prescription provided in [2, 4, 29]. As discussed in [4, 29], for any fixed type I error, such a tuning parameter allows to maximize the power. In practice, to compute λ_{AMP}, we replaced the asymptotic parameters of the AMP theory with their finite sample counterparts. Namely, $\delta = n/p = 100/300$, $\gamma = k/p = k/300$ and $\Pi^\star = 1/2\delta_t + 1/2\delta_{-t}$, where δ_t is a point mass distribution at t. Given these parameters, the formula to evaluate λ_{AMP} is provided e.g. in [2, 4, 22]. In case of strongly correlated design we additionally use $\lambda_s = 0.5\lambda_{AMP}$.

5.2 Selection of the threshold

We aim to construct a sign estimator derived from the thresholded LASSO estimator (resp. thresholded BP estimator) as defined hereafter

$$\forall i \in \{1,\dots,p\}, \widehat{\beta}_i^\tau := \widehat{\beta}_i \mathbf{1}_{\{|\widehat{\beta}_i| > \tau\}} \tag{6}$$

By taking $\tau_{1-\alpha}$ as the $1-\alpha$ quantile of $\max\left\{\left|\widehat{\beta}_i\right|, i \notin \text{supp}(\beta)\right\}$ then the probability to estimate simultaneously every null components of β at zero is $1-\alpha$. Consequently, using the threshold $\tau_{1-\alpha}$ and when non-null components of β are very large then thresholded LASSO sign estimator (resp. thresholded BP estimator) recovers $S(\beta)$ with a probability arbitrarily close to $1-\alpha$. Obviously $\tau_{1-\alpha}$ cannot be obtained by a straightforward computation since β is not known.

Given a threshold $\tau > 0$, let us set the FWER as follows

$$\text{FWER} := \mathbb{P}\left(\exists i \notin \text{supp}(\beta), \left|\widehat{\beta}_i^{\tau}\right| \neq 0\right).$$

In order to provide a threshold larger than $\tau_{1-\alpha}$ (and thus to control the FWER at level α), it could seem appealing to look at the distribution of supremum norm of the LASSO estimator (resp. BP estimator) in the full null model when $\beta = \mathbf{0}$ [18]. For the BP estimator, Descloux and Sardy [11] suggest the threshold $\tau_{1-\alpha}^{\text{fn}}$ defined as the $1-\alpha$ quantile of $\max\left\{\left|\widehat{\beta}_1^{\text{fn}}\right|, \ldots, \left|\widehat{\beta}_p^{\text{fn}}\right|\right\}$ where $\widehat{\beta}^{\text{fn}}$ is the following estimator

$$\widehat{\beta}^{\text{fn}} := \text{ argmin } \|\beta\|_1 \text{ subject to } X\beta = \varepsilon.$$

Unfortunately, in the high-dimensional linear model, this intuitive method provides a threshold $\tau_{1-\alpha}^{\text{fn}}$ which is smaller than $\tau_{1-\alpha}$ and thus does not assure that FWER $\leq \alpha$ (see also Su et al. [22] for additional explanations).

Recently developed knockoff methodology [1, 8], allows to approximate the distribution of $\widehat{\beta}(\lambda)$ associated to null-components of β by creating controlled variables. Consequently, the knockoff methodology is useful to compute a threshold. For this numerical study, we use model free knockoffs proposed in [8] to recover a threshold which heuristically control the FWER at a given level. The approach developed hereafter is available when X is a Gaussian matrix having a distribution invariant by columns' permutation. In this setting, the size of the knockoff matrix can be as small as possible (see Weinstein et al. [28] for a similar approach). Because adding some controlled variables can change some relevant properties (such as the identifiability condition for β), ideally the knockoff matrix should have just one column. Specifically, at the first step we use model free knockoffs [8] to generate $30 = p/10$ of controlled variables. Then Lasso or BP is run on the matrix supplemented with these additional columns and the maximum of the absolute values of regression coefficients over 30 controlled variables is saved. This step is repeated 10 times and the overall maximum of the $p = 300$ absolute values of regression coefficients over controlled variables is calculated. The whole procedure is repeated many times (here 1000) and 0.95 quantile of the obtained maxima is used as the threshold to identify null-components of β.

To confirm with the set-up of simulations used to derive the irrepresentability and identifiability curves, in all replicates of our simulation study we used the same fixed design matrix X described in settings 1 and 2 of the subsection 4.1 In our numerical experiments we randomly sampled the location of the true signals and we randomly generated the error term.

5.2.1 LASSO and Adaptive LASSO

In our numerical experiments we selected the following values of the tuning parameters for LASSO and adaptive LASSO:

- For LASSO we selected $\lambda_L = 81.18$.
- For the adaptive LASSO the weights are derived using initial estimates $\widehat{\beta}^L(\lambda_{AMP})$, where the tuning parameter is selected according to AMP theory, described above. For $i \in \{1, \dots, p\}$, weights $w(\beta_i)$ are defined as $w(\beta_i) := 1/(\widehat{\beta}_i^L(\lambda_{AMP}) + 10^{-7})$. Using these weights and the tuning parameter λ_L described above, the adaptive LASSO has the following expression

$$\widehat{\beta}^{\text{adapt}} := \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda_L \sum_{i=1}^{p} w(\beta_i)|\beta_i|. \tag{7}$$

In all our simulations LASSO is calculated with *glmnet*.

5.3 Numerical comparisons

The rows of the design matrix X are sampled as the independent vectors from the multivariate Gaussian distribution, as in setting 1 and 2. All numerical experiments are performed with a particular observation of X (the same as the one used in the previous section). We set $\beta \in \mathbb{R}^p$ such that $k := \text{card}(\text{supp}(\beta))$ where $k = \{5, 20\}$ and $\text{supp}(\beta)$ is a k sample without replacement of $\{1, \dots, p\}$. The non null components of β have a uniform distribution $\{-t, t\}$ where $t > 0$. Additionally, in the setting 2, we consider the set-up where all non-zero coefficients are equal to t. In all simulations the error term is generated as $\varepsilon \sim \mathcal{N}(0, Id_n)$.

Figures 4-6 provide the comparison between the following sign estimators.

- The sign estimator **L** is derived from LASSO with $\lambda = \lambda_L$.
- The sign estimator **aL** is derived from the adaptive LASSO estimator, described in (7).
- The sign estimator **BP** is derived from the thresholded BP, with threshold selected as in [11].
- The sign estimator **BPk** is derived from the thresholded BP, with a threshold given by the "knockoff" methodology described above.
- The sign estimator **Lk** is derived from the thresholded LASSO with $\lambda = \lambda_{AMP}$ and with a threshold given by the "knockoff" methodology described above.
- The sign estimator **Lks** is derived from the thresholded LASSO with $\lambda = 0.5\lambda_{AMP}$ and with a threshold given by the "knockoff" methodology described above.

In order to recover the sign of β, null components of β have to be estimated simultaneously at zero. This naive remark motivate us to report the curves illustrating the following statistical properties as the function of $t > 0$:

- **FWER** is the proportion of 1000 replicates that at least one null components of β is not estimated at zero.

We report the curve illustrating the probability to recover the sign as the function of $t > 0$:

- **Probability** is the proportion of 1000 replicates for which the sign is recovered.

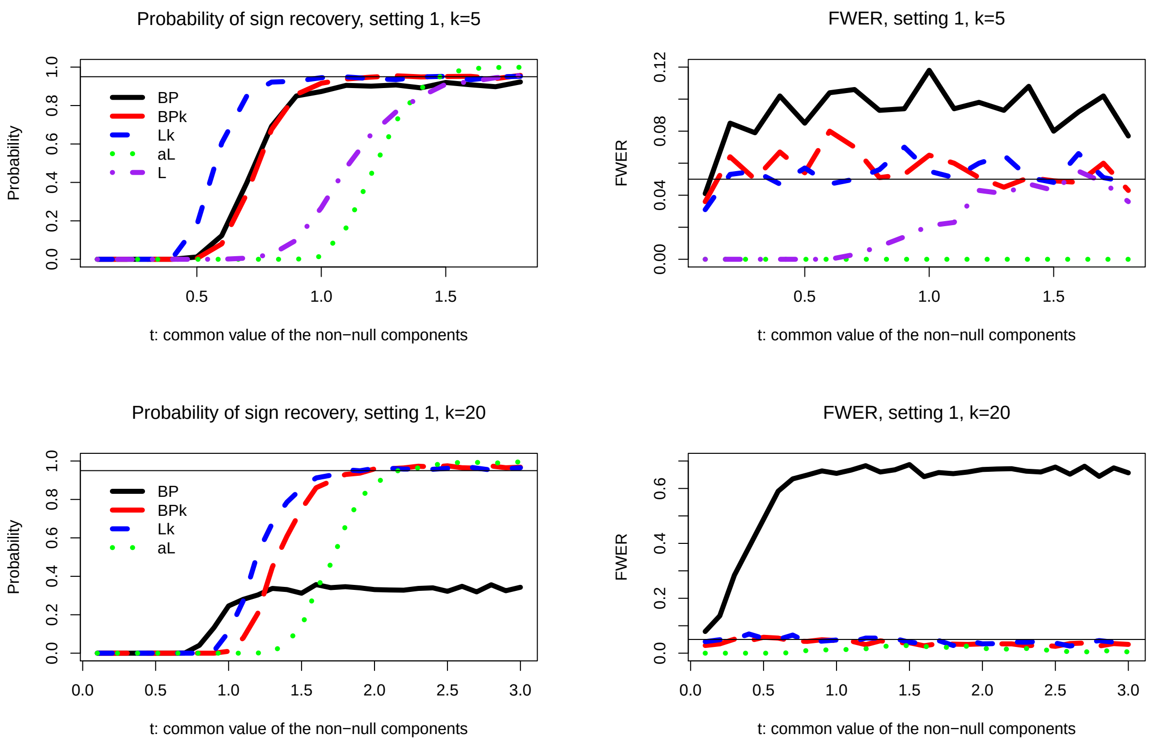

Figure 6: This figure provides the FWER and the probability to recover $S(\beta)$ for each sign estimators and when X is the design matrix given in setting 1. Graphics on the left provide the probability to recover $S(\beta)$ (on the y-axis) as a function of t, where t measures how large are the non-null components of β. Graphics on the right provide the FWER (on the y-axis) as a function of t (on the x-axis). Among these sign estimators, one notices that the thresholded LASSO sign estimator is the one which recovers $S(\beta)$ with the largest probability. These sign estimators recover approximately $S(\beta)$ with a probability close to 0.95 when t is large.

Figure 6-8 illustrate that the upper bound for the probability of sign recovery by LASSO is reached and the FWER is controlled when non null component of β are large ($i.e$ when t is large). On the other hand, thresholded LASSO and thresholded BP can appropriately identify $S(\beta)$ when the identifiability condition holds. Indeed, when $k \in \{5, 20\}$ as illustrated in figures 4 and 5, the identifiability condition occurs and thus sign estimators derived from thresholded LASSO and thresholded BP recover $S(\beta)$ as soon as the threshold is well calibrated and the non null components are large enough. In our simulated setup, thresholded BP performs pretty well but is never optimal. Indeed using an appropriate tuning parameter λ, the probability to recover $S(\beta)$ is larger

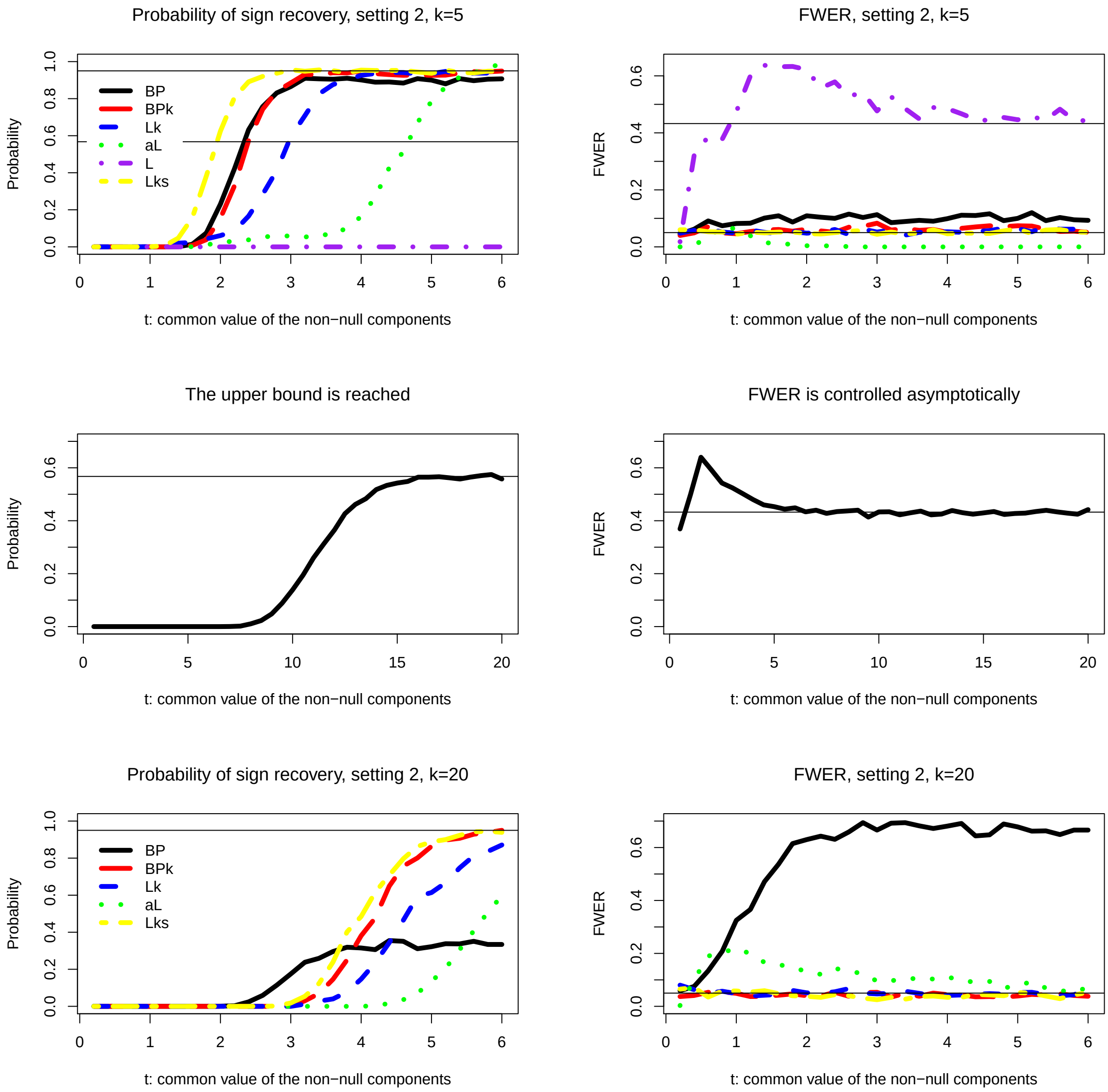

Figure 7: This figure provides the FWER and the probability to recover $S(\beta)$ for each sign estimators and when X is the design matrix given in setting 2. Graphics on the left provide the probability to recover $S(\beta)$ (on the y-axis) as a function of t (on the x-axis), where t measures how large are the non-null components of β. Graphics on the right provide the FWER (on the y-axis) as a function of t. The horizontal lines $y = 0.55$ and $y = 0.45$ represent respectively the average values of the upper bound for the probability of sign recovery and FWER associated to LASSO (see the Theorem 1). One notices that the upper-bound is approximately reached and the FWER is approximately controlled when t is very large as illustrated by graphics in the middle. Sign estimators (except LASSO sign estimator) recover approximately $S(\beta)$ with a probability close to 0.95 when t is large.

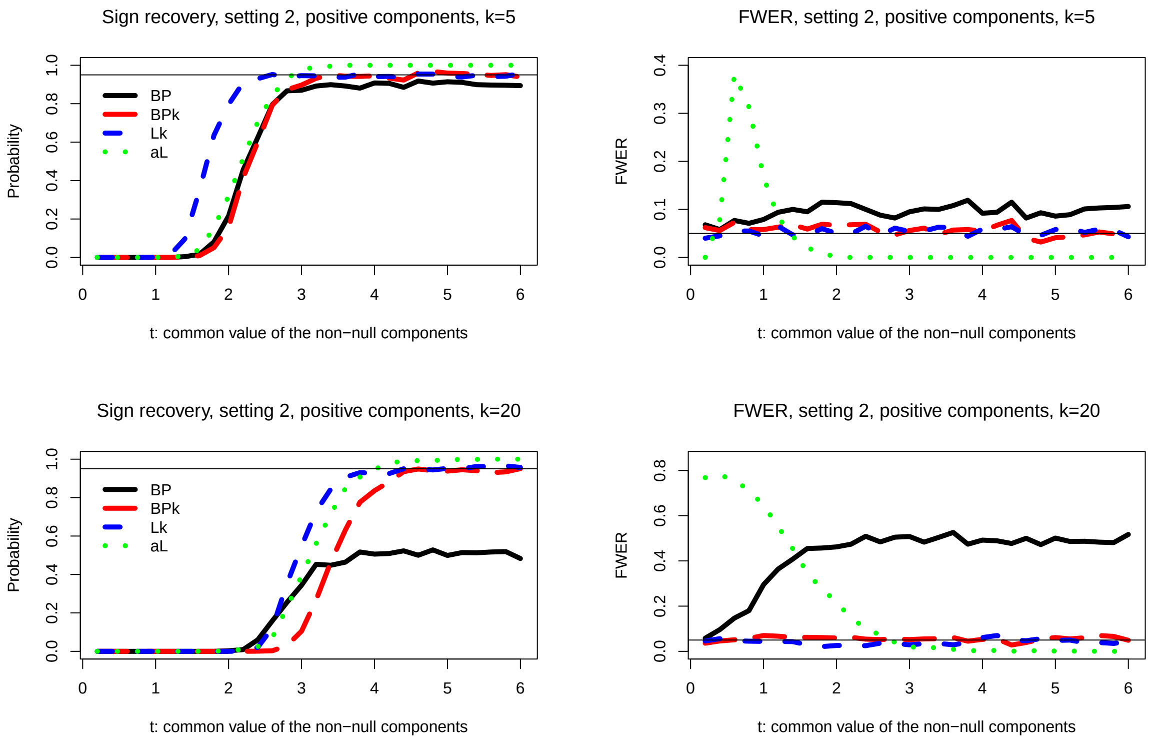

Figure 8: This figure provides the FWER and the probability to recover $S(\beta)$ for each sign estimators when X is the design matrix given in setting 2 and non-null components of β are positive. Graphics on the left provide the probability to recover $S(\beta)$ (on the y-axis) as a function of t (on the x-axis), where t measures how large are the non-null components of β. Graphics on the right provide the FWER (on the y-axis) as a function of t. These sign estimators recover approximately $S(\beta)$ with a probability close to 0.95 when t is large.

with thresholded LASSO than with thresholded BP. When entries of X are i.i.d $\mathcal{N}(0,1)$, the optimal value of λ selected by AMP theory provides a thresholded LASSO for which the derived sign estimator is the best one to recover $S(\beta)$. One notices that the threshold selection provided in Descloux and Sardy [11] does not allow to recover $S(\beta)$ with a large probability when β has lot of large components (intuitively when β is far from $\mathbf{0}$). Instead, our heuristic application of the knockoff methodology allows for almost perfect control of FWER at level 0.05. Consequently, when non null components of β are large enough and when the threshold is given by knockoff methodology, sign estimator derived from thresholded LASSO (resp. thresholded BP) recovers $S(\beta)$ with a probability close to 0.95.

6 Conclusion

This article main focus on theoretical properties of sign estimators derived from LASSO, thresholded LASSO and thresholded BPDN. We provided an upper bound for LASSO sign recovery which is reached when non-null components of β are infinitely large and the identifiability condition holds. In addition, when the irrepresentable condition occurs (implying that the identifiability condition occurs), we have shown that λ can be selected appropriately in order to control asymptotically the FWER at an arbitrary level.

When $S(\beta)$ is identifiable with respect to the L_1 norm and when non-null components of β are infinitely large, we have shown that sign estimators derived from thresholded LASSO and thresholded BPDN recover $S(\beta)$. On the other hand, if $S(\beta)$ is not identifiable with respect to the L_1 norm, sign estimators derived from thresholded LASSO and thresholded cannot recover $S(\beta)$.

We have introduced identifiability curve (resp. irrepresentability curve) which is useful to know for which sparsity β is identifiable with respect to the L_1 norm (resp. for which sparsity β the irrepresentable condition holds).

The performances of sign estimators derived from LASSO, thresholded LASSO and thresholded BPDN depend obviously from the tuning parameter, the regularization parameter and the threshold. We have illustrated that AMP theory and knockoff methodology are useful to select these parameters. Our simulations show that thresholded LASSO and thresholded BPDN sign estimators outperform adaptive LASSO and LASSO sign estimators.

Acknowledgments

We would like to thank Emmanuel J. Candès for helpful comments. The research of Małgorzata Bogdan was funded by the NCN grant 2016/23/B/ST1/00454.

7 appendix

7.1 Proof of the Theorem 1

First, let us provide lemmas which are useful to prove both Theorems 1 and 2. Lemma 2 partially prove the Theorem 1. Indeed, according to this Lemma, when $(\beta^{(r)})_{r\in\mathbb{N}}$ is a sequence of $\mathbb{R}^p$ satisfying assumptions 1 then the following asymptotic result holds

$$\lim_{r\to+\infty}\mathbb{P}\left(\forall i\in\text{supp}(s^0), S(\widehat{\beta}_i^{\text{L}}(\lambda,r))=s_i^0\right)=1.$$

Lemma 1 *Let $(\beta^{(r)})_{r\in\mathbb{N}}$ be a sequence of $\mathbb{R}^p$ satisfying the conditions* **1)** *and* **2)** *of the assumption 1, let us assume that s^0 is identifiable with respect to the L_1 norm and let us set $u_r=\|\beta^{(r)}\|_1$ then*

$$\lim_{r\to+\infty}\frac{\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)}}{u_r}=0.$$

Proof: Because $\widehat{\beta}^{\text{L}}(\varepsilon,r)$ is the LASSO estimator as defined in (2) then the following inequality occurs

$$\frac{1}{2}\|Y-X\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_2^2+\lambda\|\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_1\leq\frac{1}{2}\|Y-X\beta^{(r)}\|_2^2+\lambda\|\beta^{(r)}\|_1.$$

Since $Y-X\beta^{(r)}=\varepsilon$ one may deduce the following inequalities

$$\begin{aligned}&\lambda\|\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_1\leq\frac{1}{2}\|\varepsilon\|_2^2+\lambda\|\beta^{(r)}\|_1,\\ \Rightarrow\quad&\|\widehat{\beta}^{\text{L}}(\varepsilon,r)/u_r\|_1\leq\frac{\|\varepsilon\|_2^2}{2\lambda u_r}+1.\end{aligned}\tag{8}$$

In addition, Cauchy-Schwarz inequality gives the following implications

$$\begin{aligned}&\frac{1}{2}\|\varepsilon+X\beta^{(r)}-X\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_2^2+\lambda\|\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_1\leq\frac{1}{2}\|\varepsilon\|_2^2+\lambda\|\beta^{(r)}\|_1,\\ \Rightarrow\quad&-\|\varepsilon\|_2\|X\beta^{(r)}-X\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_2+\frac{1}{2}\|X\beta^{(r)}-X\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_2^2+\lambda\|\widehat{\beta}^{\text{L}}(\varepsilon,r)\|_1\leq\lambda\|\beta^{(r)}\|_1,\\ \Rightarrow\quad&-\frac{\|\varepsilon\|_2}{u_r}\left\|X\left(\frac{\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)}}{u_r}\right)\right\|_2+\frac{1}{2}\left\|X\left(\frac{\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)}}{u_r}\right)\right\|_2^2+\frac{\lambda}{u_r}\left\|\frac{\widehat{\beta}^{\text{L}}(\varepsilon,r)}{u_r}\right\|_1\leq\frac{\lambda}{u_r}.\end{aligned}\tag{9}$$

Because u_r tends to $+\infty$ then, according to (8), the sequence $((\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)})/u_r)_{r\in\mathbb{N}^*}$ is bounded since the following superior limit is finite

$$\limsup_{r\to+\infty}\left\|\frac{\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)}}{u_r}\right\|_1\leq 2.$$

Consequently, to prove that $\lim_{r\to+\infty}(\widehat{\beta}^{\text{L}}(\varepsilon,r)-\beta^{(r)})/u_r=\mathbf{0}$ it is sufficient to show that $\mathbf{0}$ is the unique limit point of this sequence. Let $((\widehat{\beta}^{\text{L}}(\varepsilon,\phi(r))-\beta^{(\phi(r))})/u_{\phi(r)})_{r\in\mathbb{N}^*}$ be a converging subsequence to l (with

$\phi : \mathbb{N}^* \to \mathbb{N}^*$ strictly increasing) and without loss of generality, let us assume $\lim_{r\to+\infty} \widehat{\beta}^{\mathrm{L}}(\varepsilon, \phi(r))/u_{\phi(r)} = v$ and $\lim_{r\to+\infty} \beta^{(\phi(r))}/u_{\phi(r)} = v'$ so that $l = v - v'$. By (8) and (9) one deduces that

$$Xv = Xv' \text{ and } \|v\|_1 \leq 1.$$

Since, whatever $r \geq 0$, we have $S(\beta^{(\phi(r))}/u_{\phi(r)}) = s^0$ where s^0 is identifiable with respect to the L_1 norm then, according to the Proposition 2, one may deduce that $\beta^{(\phi(r))}/u_{\phi(r)}$ is an unitary vector satisfying the identifiability condition. Consequently, $\|v'\|_1 = 1$ and v' is identifiable with respect to the L_1 norm. Consequently, $v = v'$ and thus $l = \mathbf{0}$ is the unique limit point, which implies that

$$\lim_{r\to+\infty} \frac{\widehat{\beta}^{\mathrm{L}}(\varepsilon, r) - \beta^{(r)}}{u_r} = \mathbf{0}.$$

□

For the proof of the Lemma 1, we do not have used the third condition of the assumption 1. This condition, under which the smallest non-null component of $\beta^{(r)}$ is not asymptotically infinitely smaller than $\|\beta^{(r)}\|_\infty$, is useful to prove the Lemma 2.

Lemma 2 *Let $(\beta^{(r)})_{r\in\mathbb{N}}$ be a sequence of $\mathbb{R}^p$ satisfying the assumption 1 then*

$$\lim_{r\to+\infty} \mathbb{P}(\forall i \in \operatorname{supp}(s^0), S(\widehat{\beta}_i^{\mathrm{L}}(\lambda, r)) = s_i^0) = 1.$$

Proof: Let ε be a fixed vector in $\mathbb{R}^p$. According to the third condition of the assumption 1 we have $\min\{|\beta_i^{(r)}|, i \in \operatorname{supp}(s^0)\}/\|\beta^{(r)}\|_\infty \geq q > 0$, consequently the following inequalities occur

$$\forall i \in \operatorname{supp}(s^0), s_i^0 \frac{\widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r) - \beta_i^{(r)}}{\|\beta^{(r)}\|_\infty} = \frac{s_i^0 \widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r)}{\|\beta^{(r)}\|_\infty} - \frac{|\beta_i^{(r)}|}{\|\beta^{(r)}\|_\infty} \leq \frac{s_i^0 \widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r)}{\|\beta^{(r)}\|_\infty} - q.$$

According to the Lemma 1, the following inequality occurs

$$0 = \liminf_{r\to+\infty} s_i^0 \frac{\widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r) - \beta_i^{(r)}}{\|\beta^{(r)}\|_\infty} \leq \liminf_{r\to+\infty} \frac{s_i^0 \widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r)}{\|\beta^{(r)}\|_\infty} - q.$$

Which implies that for r large enough $s_i^0 \widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r) > 0$ and thus $S(\widehat{\beta}_i^{\mathrm{L}}(\varepsilon, \lambda, r)) = s_i^0$. When ε is no longer fixed then, for $i \in \operatorname{supp}(s^0)$, almost surely $S(\widehat{\beta}_i^{\mathrm{L}}(r))$ converges to s_i^0 and consequently

$$\lim_{r\to+\infty} \mathbb{P}\left(\forall i \in \operatorname{supp}(s^0), S(\widehat{\beta}_i^{\mathrm{L}}(\lambda, r)) = s_i^0\right) = 1.$$

$\square$

Proof of the Theorem 1: Let A be the set $A := \mathrm{supp}(\widehat{\beta}^{\mathrm{L}}(\lambda))$.

Upper bound) Let us give two expressions met by the LASSO estimator as defined in (2). The vector $\widehat{\beta}^{\mathrm{L}}(\lambda)$ is the LASSO estimator if and only if the following two inequalities occur simultaneously.

$$\begin{aligned} X_A'(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda)) &= \lambda S(\widehat{\beta}_A^{\mathrm{L}}(\lambda)), && (10)\\ \|X_{\overline{A}}'(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda))\|_\infty &\le \lambda. && (11) \end{aligned}$$

These two expressions are given in Bühlmann and van de Geer [5] page 15 or in the proof of the Theorem 1 of Zou [31]. Using the equality (10) and the inequality (11), we are going to show that if $S(\widehat{\beta}^{\mathrm{L}}(\lambda)) = S(\beta)$ then the following event holds

$$\left\| X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I) + \frac{1}{\lambda} X_{\overline{I}}' \left(Id - X_I (X_I' X_I)^{-1} X_I' \right) \varepsilon \right\|_\infty \le 1.$$

Let us assume that $S(\widehat{\beta}^{\mathrm{L}}(\lambda)) = S(\beta)$ thus $A = I$ (where $I = \mathrm{supp}(\beta)$). Since $Y = X\beta + \varepsilon = X_I\beta_I + \varepsilon$ and $X\widehat{\beta}^{\mathrm{L}}(\lambda) = X_I\widehat{\beta}_I^{\mathrm{L}}(\lambda)$ then the equality (10) and the inequality (11) lead to the following expressions

$$\begin{aligned} X_I' \left(\varepsilon + X_I(\beta_I - \widehat{\beta}_I^{\mathrm{L}}(\lambda)) \right) &= \lambda S(\beta_I), && (12)\\ \left\| X_{\overline{I}}' \left(\varepsilon + X_I(\beta_I - \widehat{\beta}_I^{\mathrm{L}}(\lambda)) \right) \right\|_\infty &\le \lambda. && (13) \end{aligned}$$

The equality (12) assures that

$$\beta_I - \widehat{\beta}_I^{\mathrm{L}}(\lambda) = (X_I'X_I)^{-1} \left(\lambda S(\beta_I) - X_I'\varepsilon \right).$$

Let us notice that since $\ker(X_I) = \mathbf{0}$ then the Gram matrix $X_I'X_I$ is invertible. Using the previous expression in the inequality (13) gives

$$\begin{aligned} \left\| X_{\overline{I}}' X_I (X_I'X_I)^{-1} (\lambda S(\beta_I) - X_I'\varepsilon) + X_{\overline{I}}'\varepsilon \right\|_\infty &\le \lambda,\\ \left\| X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I) + \frac{1}{\lambda} X_{\overline{I}}' \left(Id - X_I (X_I' X_I)^{-1} X_I' \right) \varepsilon \right\|_\infty &\le 1. \end{aligned}$$

Consequently, one deduces the folowing inequality

$$\mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda)) = S(\beta) \right) \le \underbrace{\mathbb{P}\left(\left\| X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I) + \frac{1}{\lambda} X_{\overline{I}}' \left(Id - X_I (X_I' X_I)^{-1} X_I' \right) \varepsilon \right\|_\infty \le 1 \right)}_{=\mathbb{P}\left(\left\| \zeta_{X,\lambda,S(\beta)} \right\|_\infty \le 1 \right) = \bar{\gamma}}.$$

Sharpness of the upper bound) Since the upper bound just depends from s^0 and not on how large are the non-null components $\beta^{(r)}$ then

$$\limsup_{r\to+\infty} \mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = s^0\right) \le \bar{\gamma}.$$

So it remains to prove that $\liminf_{r\to+\infty} \mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = s^0\right) \ge \mathbb{P}\left(\left\|\zeta_{X,\lambda,S(\beta)}\right\|_\infty < 1\right) = \gamma$. Let us remind that $I = \mathrm{supp}(s^0)$ and let us assume that the following events hold simultaneously

$$X_I'(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda)) = \lambda s_I^0 \text{ and } \underbrace{\left\|X_{\overline{I}}'X_I(X_I'X_I)^{-1}\lambda s_I^0 + X_{\overline{I}}'\left(Id - X_I(X_I'X_I)^{-1}X_I'\right)\varepsilon\right\|_\infty < \lambda}_{=\left\|\zeta_{X,\lambda,s^0}\right\|_\infty<1}. \tag{14}$$

We aim to show that the inequalities given above imply that $\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) = \mathbf{0}$. For convenience, let us set H be the projection matrix $H := X_I(X_I'X_I)^{-1}X_I'$. When (14) occurs then the following inequalities holds

$$\begin{aligned}
\left\|X_{\overline{I}}'H(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda)) + X_{\overline{I}}'(Id - H)\,\varepsilon\right\|_\infty &< \lambda,\\
\left\|X_{\overline{I}}'\left(H(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda)) + (Id - H)\varepsilon\right)\right\|_\infty &< \lambda,\\
\left\|X_{\overline{I}}'\left(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda) + X_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) - HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\right)\right\|_\infty &< \lambda.
\end{aligned} \tag{15}$$

The inequality (15) comes from the following two identities

$$\begin{aligned}
HY &= H(X\beta^{(r)}) + H\varepsilon = H(X_I\beta_I^{(r)}) + H\varepsilon = X_I\beta_i^{(r)} + H\varepsilon = X(\beta^{(r)}) + H\varepsilon \text{ and,}\\
HX\widehat{\beta}^{\mathrm{L}}(\lambda) &= HX_I\widehat{\beta}^{\mathrm{L}}_I(\lambda) + HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) = X_I\widehat{\beta}^{\mathrm{L}}_I(\lambda) + HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) = X\widehat{\beta}^{\mathrm{L}}(\lambda) - X_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) + HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda).
\end{aligned}$$

Let v be the vector $v := X_{\overline{I}}'\left(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda) + X_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) + HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\right)$, we are going to see that the inequality (15) implies that $\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) = \mathbf{0}$. Let us assume that $\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) \neq \mathbf{0}$ then, on the first hand, the following inequality occurs

$$\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)'v \le \|\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\|_1\|v\|_\infty < \lambda\|\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\|_1. \tag{16}$$

According to (10) the identity $\widehat{\beta}^{\mathrm{L}}_i(\lambda)X_i'(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda)) = \lambda|\widehat{\beta}^{\mathrm{L}}_i(\lambda)|$ occurs. Consequently, on the other hand, the following inequalities hold

$$\begin{aligned}
\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)'v &= \widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)'X_{\overline{I}}'\left(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda) + X_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) - HX_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\right),\\
&= \lambda\|\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)\|_1 + \widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda)'X_{\overline{I}}'(Id - H)X_{\overline{I}}\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda),\\
&\ge \lambda\|\widehat{\beta}^{L}_{\overline{I}}(\lambda)\|_1.
\end{aligned} \tag{17}$$

The last inequality occurs because the projection matrix $Id - H$ is positive semi-definite. Inequalities (16) and

(17) provide a contradiction which implies that $\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda) = \mathbf{0}$.

According to (10), the following implication holds

$$S(\widehat{\beta}^{\mathrm{L}}_I(\lambda, r)) = s^0_I \Rightarrow X'_I(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = \lambda s^0_I.$$

Because s^0 is identifiable with respect to the L_1 norm then, according to the Lemma 2, the following convergence in probability occurs

$$\lim_{r\to+\infty} \mathbb{P}(S(\widehat{\beta}^{\mathrm{L}}_I(\lambda, r)) = s^0_I) = \lim_{r\to+\infty} \mathbb{P}(X'_I(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = \lambda s^0_I) = 1. \tag{18}$$

Using this asymptotic result and since when (14) occurs then $\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda, r) = \mathbf{0}$, one may deduce the following inequalities

$$\begin{aligned}
\liminf_{r\to+\infty} \mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = s^0\right) &= \liminf_{r\to+\infty} \mathbb{P}\left(S(\widehat{\beta}^{\mathrm{L}}_I(\lambda, r)) = s^0_I \text{ and } \widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda, r) = \mathbf{0}\right), \\
&= \liminf_{r\to+\infty} \mathbb{P}(\widehat{\beta}^{\mathrm{L}}_{\overline{I}}(\lambda, r) = \mathbf{0}), \\
&\geq \liminf_{r\to+\infty} \mathbb{P}\left(X'_I(Y - X\widehat{\beta}^{\mathrm{L}}(\lambda, r)) = s^0_I \text{ and } \left\|\zeta_{X,\lambda,s^0}\right\|_\infty < 1\right), \\
&\geq \liminf_{r\to+\infty} \mathbb{P}\left(\left\|\zeta_{X,\lambda,s^0}\right\|_\infty < 1\right) = \gamma.
\end{aligned}$$

Asymptotic full power and asymptotic control of the FWER) According to (18), asymptotically the power is equal to 1, namely $\lim_{r\to+\infty} \mathbb{P}(\forall i \in I, S(\widehat{\beta}^{\mathrm{L}}_i(\lambda, r)) = s^0_i) = 1$. Now let us prove that the FWER is controlled asymptotically. Using asymptotic results given above one may deduce the following inequalities.

$$\begin{aligned}
\bar{\gamma} &\geq \limsup_{r\to+\infty} \mathbb{P}(S(\widehat{\beta}^L(\lambda, r)) = s^0), \\
&\geq \limsup_{r\to+\infty} \mathbb{P}\left(\forall i \in I, S(\widehat{\beta}^L_i(\lambda, r)) = s^0_i \text{ and } \forall i \notin I, \widehat{\beta}^L_i(\lambda, r) = 0\right), \\
&\geq \limsup_{r\to+\infty} \mathbb{P}(\forall i \notin I, \widehat{\beta}^L_i(\lambda, r) = 0).
\end{aligned} \tag{19}$$

The last inequality come from (18). Similarly, we have

$$\gamma \leq \liminf_{r\to+\infty} \mathbb{P}(\forall i \notin I, \widehat{\beta}^L_i(\lambda, r) = 0). \tag{20}$$

Consequently, by taking the complement to 1 of the inequalities given in (19) and (20), one may deduce that

$$\liminf_{r\to+\infty} \mathbb{P}(\exists i \notin I, \widehat{\beta}^L_i(\lambda, r) \neq 0) \geq 1 - \bar{\gamma} \text{ and } \limsup_{r\to+\infty} \mathbb{P}(\exists i \notin I, \widehat{\beta}^L_i(\lambda, r) \neq 0) \leq 1 - \gamma.$$

□

Proof of the Theorem 2

The Lemma 3 provides the same result for BPDN than the Lemma 1 for LASSO. These both lemmas are the keystone to prove the Theorem 2.

Lemma 3 *Let $(\beta^{(r)})_{r\in\mathbb{N}}$ be a sequence of $\mathbb{R}^p$ satisfying the conditions* **1)** *and* **2)** *of the assumption 1, let us assume that s^0 is identifiable with respect to the L_1 norm and let set $u_r = \|\beta^{(r)}\|_1$ then*

$$\lim_{r\to+\infty} \frac{\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)-\beta^{(r)}}{u_r} = 0.$$

Proof: Let us define $u(\varepsilon)\in\mathbb{R}^p$ as follows

$$u(\varepsilon) := \operatorname*{argmin}_{b\in\mathbb{R}^p} \|b\|_1 \text{ subject to } Xb=\varepsilon.$$

Because $X(u(\varepsilon))=\varepsilon$, we have $Y(\varepsilon)=X(\beta^{(r)}+u(\varepsilon))$ and because $\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)$ is an admissible point of (3) one deduces the following inequality

$$\left\|\frac{1}{u_r}X\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)-\frac{1}{u_r}X\beta^{(r)}\right\|_2 \le \left\|\frac{1}{u_r}X\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)-\frac{1}{u_r}Y\right\|_2+\left\|\frac{1}{u_r}Y-\frac{1}{u_r}X\beta^{(r)}\right\|_2 \le \frac{\sqrt{R}}{u_r}+\frac{\|Xu(\varepsilon)\|_2}{u_r}. \quad (21)$$

Because $\beta^{(r)}+u(\varepsilon)$ is an admissible point of the problem (3) and because $\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)$ is the minimizer of (3), one deduces the following inequalities hold

$$\frac{1}{u_r}\|\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)\|_1 \le \frac{1}{u_r}\|\beta^{(r)}+u(\varepsilon)\|_1 \le 1+\frac{\|u(\varepsilon)\|_1}{u_r}. \quad (22)$$

Because u_r tends to $+\infty$ then, according to (22), the sequence $((\widehat{\beta}^{\mathrm{L}}(\varepsilon,r)-\beta^{(r)})/u_r)_{r\in\mathbb{N}^*}$ is bounded since the following superior limit is finite

$$\limsup_{r\to+\infty}\left\|\frac{\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)-\beta^{(r)}}{u_r}\right\|_1 \le 2.$$

Consequently, to prove that $\lim_{r\to+\infty}(\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,r)-\beta^{(r)})/u_r=\mathbf{0}$ it is sufficient to show that $\mathbf{0}$ is the unique limit point of this sequence. Let $((\widehat{\beta}^{\mathrm{L}}(\varepsilon,\phi(r))-\beta^{(\phi(r))})/u_{\phi(r)})_{r\in\mathbb{N}^*}$ be a converging subsequence to l (with $\phi:\mathbb{N}^*\to\mathbb{N}^*$ strictly increasing) and without loss of generality, let us assume $\lim_{r\to+\infty}\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon,\phi(r))/u_{\phi(r)}=v$ and $\lim_{r\to+\infty} b^{(\phi(r))}/u_{\phi(r)}=v'$ so that $l=v-v'$. By (21) and (22) one deduces that

$$Xv = Xv' \text{ and } \|v\|_1\le 1.$$

Since, whatever $r\ge 0$, we have $S(\beta^{(\phi(r))}/u_{\phi(r)})=s^0$ where s^0 is identifiable with respect to the L_1 norm then, according to the Proposition 2, one may deduce that $\beta^{(\phi(r))}/u_{\phi(r)}$ is an unitary vector satisfying the identifiability condition. Consequently, $\|v'\|_1=1$ and v' is identifiable with respect to the L_1 norm. Consequently, $v=v'$

and thus $l = \mathbf{0}$ is the unique limit point, which implies that

$$\lim_{r \to +\infty} \frac{\widehat{\beta}^{\mathrm{BPDN}}(\varepsilon, r) - \beta^{(r)}}{u_r} = \mathbf{0}.$$

$\square$

Lemma 4 *Let X be a matrix in general position then, the random vector $\widehat{\beta}$ is identifiable with respect to X and the L_1 norm.*

Proof: Let us remind that when X is in general position then the minimizer $\widehat{\beta}$ is unique. Let us assume that $\widehat{\beta}$ is not identifiable with respect to X and the L_1 norm then there exists $b \in \mathbb{R}^p$ such that $Xb = X\widehat{\beta}$ and $\|b\|_1 \leq \|\widehat{\beta}\|_1$. Consequently, for LASSO, one may deduce that

$$\|Y - Xb\|^2 + \lambda \|b\|_1 \leq \|Y - X\widehat{\beta}^{\mathrm{L}}\|^2 + \lambda \|\widehat{\beta}^{\mathrm{L}}\|_1.$$

This inequality contradicts that $\widehat{\beta}^{\mathrm{L}}$ is the unique minimizer of (2). Similarly when $\widehat{\beta}^{\mathrm{BPDN}}$ is not identifiable with respect to the L_1 norm then $\widehat{\beta}^{\mathrm{BPDN}}$ is not the unique minimizer of (3) which provides a contradiction. $\square$

For the proofs of the Theorem 2 and the proof of the Proposition 2, we need to introduce the following inequality which characterizes the identifiability condition [10]. A vector $b \in \mathbb{R}^p$ is identifiable with respect to X and the L_1 norm if and only if the following inequality holds

$$\forall h \in \ker(X) \setminus \{\mathbf{0}\}, \left| \sum_{i \in \mathrm{supp}(b)} S(b) h_i \right| < \sum_{i \notin \mathrm{supp}(b)} |h_i|. \tag{23}$$

Proof of the Theorem 2: Let us remind that according to the condition **3)** of the assumption 1 the following inequality holds

$$\forall r \in \mathbb{N}, \frac{\min\{|\beta_i^{(r)}|, i \in \mathrm{supp}(s^0)\}}{\|\beta^{(r)}\|_\infty} \geq q > 0.$$

According to the lemmas 1 and 3, when s^0 is identifiable with respect to the L_1 norm then

$$\lim_{r \to +\infty} \frac{\widehat{\beta}(\varepsilon, r) - \beta^{(r)}}{\|\beta^{(r)}\|_\infty} = 0.$$

Therefore, there exists $r_0(\varepsilon) \geq 0$ such that

$$\forall r \geq r_0(\varepsilon), \left\| \frac{\widehat{\beta}(\varepsilon, r) - \beta^{(r)}}{\|\beta^{(r)}\|_\infty} \right\|_\infty < q/2 \Leftrightarrow \forall i \in \{1, \ldots, p\}, \forall r \geq r_0(\varepsilon), \left| \frac{\widehat{\beta}_i(\varepsilon, r) - \beta_i^{(r)}}{\|\beta^{(r)}\|_\infty} \right| < q/2.$$

Consequently, when $r \geq r_0(\varepsilon)$, whatever $i \notin \mathrm{supp}(s^0)$ (thus when $\beta_i^{(r)} = 0$) the following inequalities hold

$$
\begin{aligned}
& \forall i \notin \mathrm{supp}(s^0), \left| \frac{\widehat{\beta}_i(\varepsilon, r)}{\|\beta^{(r)}\|_\infty} \right| < q/2, \\
\Rightarrow \quad & -\|\beta^{(r)}\|_\infty q/2 < \min_{i \notin \mathrm{supp}(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} \leq \max_{i \notin \mathrm{supp}(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} < \|\beta^{(r)}\|_\infty q/2.
\end{aligned}
$$

Whatever $i \in \mathrm{supp}^+(s^0)$ (thus when $\beta_i^{(r)} > 0$) the following inequalities hold

$$
\begin{aligned}
& \forall i \in \mathrm{supp}^+(s^0), \frac{\widehat{\beta}_i(\varepsilon, r)}{\|\beta^{(r)}\|_\infty} \geq - \left| \frac{\widehat{\beta}_i(\varepsilon, r) - \beta_i^{(r)}}{\|\beta^{(r)}\|_\infty} \right| + \frac{\beta_i^{(r)}}{\|\beta^{(r)}\|_\infty}, \\
\Rightarrow \quad & \min_{i \in \mathrm{supp}^+(s^0)} \left\{ \frac{\widehat{\beta}_i(\varepsilon, r)}{\|\beta^{(r)}\|_\infty} \right\} > -q/2 + q = q/2, \\
\Rightarrow \quad & \min_{i \in \mathrm{supp}^+(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} > \|\beta^{(r)}\|_\infty q/2.
\end{aligned}
$$

Whatever $i \in \mathrm{supp}^-(s^0)$ (thus when $\beta_i^{(r)} < 0$) the following inequalities hold

$$
\begin{aligned}
& \forall i \in \mathrm{supp}^+(s^0), \frac{\widehat{\beta}_i(\varepsilon, r)}{\|\beta^{(r)}\|_\infty} \leq \left| \frac{\widehat{\beta}_i(\varepsilon, r) - \beta_i^{(r)}}{\|\beta^{(r)}\|_\infty} \right| + \frac{\beta_i^{(r)}}{\|\beta^{(r)}\|_\infty}, \\
\Rightarrow \quad & \max_{i \in \mathrm{supp}^-(s^0)} \left\{ \frac{\widehat{\beta}_i(\varepsilon, r)}{\|\beta^{(r)}\|_\infty} \right\} < q/2 - q = -q/2, \\
\Rightarrow \quad & \max_{i \in \mathrm{supp}^-(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} < -\|\beta^{(r)}\|_\infty q/2.
\end{aligned}
$$

Finally, when $r \geq r_0(\varepsilon)$ then

i)

$$
\mathrm{supp}^-(s^0) \subset \mathrm{supp}^-(\widehat{\beta}_i(\varepsilon, r)) \text{ and } \mathrm{supp}^+(s^0) \subset \mathrm{supp}^+(\widehat{\beta}_i(\varepsilon, r)).
$$

ii)

$$
\max_{i \in \mathrm{supp}^-(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} < \min_{i \notin \mathrm{supp}(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} \leq \max_{i \notin \mathrm{supp}(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\} < \min_{i \in \mathrm{supp}^+(s^0)} \left\{ \widehat{\beta}_i(\varepsilon, r) \right\}.
$$

Which achieve the first part of the proof. Now, let us assume that s^0 is not identifiable with respect to the L_1 norm. Let us show that when the following events hold

$$
\mathrm{supp}^-(s^0) \subset \mathrm{supp}^-(\widehat{\beta}) \text{ and } \mathrm{supp}^+(s^0) \subset \mathrm{supp}^+(\widehat{\beta}), \tag{24}
$$

then the inequality (23) occurs which contradicts that s^0 is not identifiable. Let $h \in \ker(X) \setminus \{\mathbf{0}\}$ on the first hand, when (24) occurs, we have

$$
\left| \sum_{i \in \mathrm{supp}(s^0)} s_i^0 h_i \right| = \left| - \sum_{\mathrm{supp}^-(s^0)} h_i + \sum_{\mathrm{supp}^+(s^0)} h_i \right| \leq \left| - \sum_{i \in \mathrm{supp}^-(\widehat{\beta})} h_i + \sum_{i \in \mathrm{supp}^+(\widehat{\beta})} h_i \right| + \sum_{i \in \mathrm{supp}(\widehat{\beta}) \setminus \mathrm{supp}(s)} |h_i|.
$$

On the other hand, according to the Lemma 4, $\widehat{\beta}$ is identifiable with respect to the L_1 norm then (23) occurs implying thus the following inequality

$$\left| - \sum_{i \in \mathrm{supp}^-(\widehat{\beta})} h_i + \sum_{i \in \mathrm{supp}^+(\widehat{\beta})} h_i \right| + \sum_{i \in \mathrm{supp}(\widehat{\beta}) \setminus \mathrm{supp}(s)} |h_i| < \sum_{i \notin \mathrm{supp}(\widehat{\beta})} |h_i| + \sum_{i \in \mathrm{supp}(\widehat{\beta}) \setminus \mathrm{supp}(s)} |h_i| = \sum_{i \notin \mathrm{supp}(s)} |h_i|.$$

Consequently the following inequality holds

$$\forall h \in \ker(X) \setminus \{\mathbf{0}\}, \left| \sum_{i \in \mathrm{supp}(s^0)} s_i^0 h_i \right| < \sum_{i \notin \mathrm{supp}(s)} |h_i|,$$

which, according to (23), contradicts that s^0 is not identifiable with respect to the L_1 norm. $\square$

Proof of propositions

Proof of the Proposition 1: From Daubechies et al. [10], β is a parameter having a minimal L_1 norm, namely $X\beta = X\gamma \Rightarrow \|\gamma\|_1 \geq \|\beta\|_1$ holds, if and only if the following inequality occurs

$$\forall h \in \ker(X), \left| \sum_{i \in I} S(\beta_i) h_i \right| \leq \sum_{i \notin I} |h_i|. \tag{25}$$

We are going to show that when the irrepresentable condition holds for β then the inequality (23) holds.

Let $h \in \ker(X)$ and let us remind that h_I and $h_{\overline{I}}$ denote respectively vectors $(h_i)_{i \in I}$ and $(h_i)_{i \notin I}$ then, the following equality holds

$$\sum_{i \in I} S(\beta_i) h_i = h_I' S(\beta_I) = h_I' X_I' X_I (X_I' X_I)^{-1} S(\beta_I).$$

Because $\mathbf{0} = Xh = X_I h_I + X_{\overline{I}} h_{\overline{I}}$, one deduces the following inequalities

$$\begin{aligned} |h_I' S(\beta_I)| &= \left| h_{\overline{I}}' X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I) \right|, \\ &\leq \|h_{\overline{I}}\|_1 \|X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(\beta_I)\|_\infty. \end{aligned} \tag{26}$$

Consequently, when the irrepresentable condition holds for β namely, when $\|X_{\overline{I}}' X_I (X_I' X_I)^{-1} S(b_I^*)\|_\infty \leq 1$ then, the inequality (26) gives $|h_I' S(\beta_I)| \leq \|h_{\overline{I}}\|_1$. Thus, by the equivalence given in (25), β is a solution of the following basis pursuit problem

$$\text{minimize } \|\gamma\|_1 \text{ subject to } X\gamma = X\beta$$

Because X is in general position the previous optimisation problem has a unique solution (see *e.g.* the proposition 1 in appendix) thus $X\beta = X\gamma$ and $\gamma \neq \beta$ implies that $\|\gamma\|_1 > \|\beta\|_1$ namely β is identifiable with respect to the L_1 norm. $\square$

Proof of the Proposition 2: Because b is identifiable with respect to the L_1 norm and because $S(\tilde{b}) = S(b)$ implies $\text{supp}(\tilde{b}) = \text{supp}(b)$, then the following inequality holds

$$\forall h \in \ker(X) \setminus \{\mathbf{0}\}, \left| \sum_{i \in \text{supp}(\tilde{b})} S(\tilde{b}_i) h_i \right| < \sum_{i \notin \text{supp}(\tilde{b})} |h_i|.$$

Consequently, according to (23), the parameter $\tilde{b}$ is identifiable with respect to the L_1 norm. $\square$

Supplementary material

We already said that when X is in general position the minimizer of problem (2) (resp. problem (3)) is unique. Concerning LASSO, a sketch of proof given in Tibshirani [26] shows the uniqueness of the LASSO estimator when X is in general position. In order to provide a self content article, we show that when X is in general position the minimizer of the problem (3) is unique when $R = 0$ as well as when $R > 0$. We already stressed that when β is identifiable with respect to the L_1 norm then β is sparse. We are going to show that when the identifiability holds for β then the family $(X_i)_{i \in \text{supp}(\beta)}$ is linearly independent and thus the number of components of β equal to 0 is larger than $p - n$.

References

[1] R. F. Barber and E. J. Candès. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085, 2015.

[2] Mohsen Bayati and Andrea Montanari. The LASSO risk for Gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, 2012.

[3] Dimitri P Bertsekas. *Nonlinear programming.* Athena scientific Belmont, 1999.

[4] M. Bogdan, E. J. Candès, W. Su, and A. Weinstein. Off the beaten path: ranking variables with cross-validated lasso. *Technical Report, University of Wroclaw*, 2018.

[5] Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Springer, 2011.

[6] T Tony Cai and Anru Zhang. Sharp rip bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1):74–93, 2013.

[7] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[8] Emmanuel J. Candès, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: Model-free knockoffs for high-dimensional controlled variable selection. *arXiv preprint arXiv:1610.02351*, 2016. To appear in Journal of the Royal Statistical Society Series B.

[9] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[10] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on pure and applied mathematics*, 63(1):1–38, 2010.

[11] Pascaline Descloux and Sylvain Sardy. Model selection with lasso-zero: adding straw to the haystack to better find needles. *arXiv preprint arXiv:1805.05133*, 2018.

[12] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Trans. R. Soc. A*, 367(1906):4273–4293, 2009.

[13] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[14] David L Donoho and Jared Tanner. Precise undersampling theorems. *Proceedings of the IEEE*, 98(6):913–924, 2010.

[15] Charles Dossal. A necessary and sufficient condition for exact sparse recovery by 1 minimization. *Comptes Rendus Mathematique*, 350(1):117–120, 2012.

[16] Charles Dossal, Marie-Line Chabanol, Gabriel Peyré, and Jalal Fadili. Sharp support recovery from noisy random measurements by 1-minimization. *Applied and Computational Harmonic Analysis*, 33(1):24–43, 2012.

[17] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Springer, 2013.

[18] Caroline Giacobino, Sylvain Sardy, Jairo Diaz-Rodriguez, Nick Hengartner, et al. Quantile universal threshold. *Electronic Journal of Statistics*, 11(2):4701–4722, 2017.

[19] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.

[20] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

[21] Hatef Monajemi, Sina Jafarpour, Matan Gavish, David L Donoho, Sivaram Ambikasaran, Sergio Bacallado, Dinesh Bharadia, Yuxin Chen, Young Choi, Mainak Chowdhury, et al. Deterministic matrices matching the compressed sensing phase transitions of gaussian random matrices. *Proceedings of the National Academy of Sciences*, 110(4):1181–1186, 2013.

[22] Weijie J Su, Małgorzata Bogdan, and Emmanuel J. Candès. False discoveries occur early on the lasso path. *The Annals of Statistics*, 45(5):2133–2150, 2017.

[23] Patrick Tardivel. *Représentation parcimonieuse et procédures de tests multiples: application à la métabolomique*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2017.

[24] Patrick JC Tardivel, Rémi Servien, and Didier Concordet. Sparsest representations and approximations of an underdetermined linear system. *Inverse Problems*, 34(5):055002, 2018.

[25] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[26] Ryan J Tibshirani et al. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

[27] Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55(5):2183–2202, 2009.

[28] Asaf Weinstein, Rina Barber, and Emmanuel J. Candès. A power and prediction analysis for knockoffs with lasso statistics. *arXiv preprint arXiv:1712.06465*, 2017.

[29] Shuaiwen Wang Haolei Weng and Arian Maleki. Which bridge estimator is the best for variable selection? *arxiv*, 2018.

[30] Peng Zhao and Bin Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.

[31] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.