Submitted to the Annals of Statistics

PATTERN RECOVERY BY SLOPE

BY MAŁGORZATA BOGDAN[1], XAVIER DUPUIS[2], PIOTR GRACZYK[3], BARTOSZ KOŁODZIEJEK[4], TOMASZ SKALSKI[5], PATRICK TARDIVEL[6], AND MACIEJ WILCZYŃSKI[7*]

[1] University of Wroclaw, Poland & Lund University, Sweden, malgorzata.bogdan@uwr.edu.pl

[2] University of Burgundy, France, xavier.dupuis@u-bourgogne.fr

[3] Université d'Angers, France, piotr.graczyk@univ-angers.fr

[4] Warsaw University of Technology, Poland, b.kolodziejek@mini.pw.edu.pl

[5] Wrocław University of Science and Technology, Poland & Université d'Angers, France, tomasz.skalski@pwr.edu.pl

[6] University of Burgundy, France, patrick.tardivel@u-bourgogne.fr

[7] Wroclaw University of Science and Technology, Poland, maciej.wilczynski@pwr.edu.pl

LASSO and SLOPE are two popular methods for dimensionality reduction in the high-dimensional regression. LASSO can eliminate redundant predictors by setting the corresponding regression coefficients to zero, while SLOPE can additionally identify clusters of variables with the same absolute values of regression coefficients. It is well known that LASSO Irrepresentability Condition is sufficient and necessary for the proper estimation of the sign of sufficiently large regression coefficients. In this article we formulate an analogous Irrepresentability Condition for SLOPE, which is sufficient and necessary for the proper identification of the SLOPE pattern, i.e. of the proper sign as well as of the proper ranking of the absolute values of individual regression coefficients, while proper ranking guarantees a proper clustering. We also provide asymptotic results on the strong consistency of pattern recovery by SLOPE when the number of columns in the design matrix is fixed while the sample size diverges to infinity.

1. Introduction. High-dimensional data is currently ubiquitous in many areas of science and industry. Efficient extraction of information from such data sets often requires dimensionality reduction based on identifying the low-dimensional structure behind the data generation process. One of the simplest forms of dimensionality reduction is the elimination of irrelevant parameters of the statistical models describing the data. In case of the high-dimensional multiple regression model

$$Y = X\beta + \varepsilon, \tag{1.1}$$

with the design matrix $X \in \mathbb{R}^{n\times p}$, an unknown parameter vector $\beta \in \mathbb{R}^p$ and a random noise ε, elimination of irrelevant parameters in the vector β corresponds to elimination of irrelevant predictors. One of the most popular methods for performing this task is the Least Absolute Shrinkage and Selection Operator (LASSO, [10, 20]), which estimates β by solving the following convex optimization problem

$$\text{minimize } \left[\frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1\right] \text{ over } b \in \mathbb{R}^p, \tag{1.2}$$

*The order of authors is alphabetical.

MSC2020 subject classifications: Primary 62J05; secondary 62J07.

Keywords and phrases: linear regression, SLOPE, pattern recovery, irrepresentability condition.

where $\|b\|_1 = \sum_{i=1}^p |\beta_i|$ is the ℓ_1 norm of β and $\lambda > 0$ is a tuning parameter. The well-known selection properties of the ℓ_1 norm are related to the fact that it is not differentiable when at least one coordinate of β is equal to zero. The sparsity of the LASSO solution is governed by the tuning parameter λ and it is guaranteed that at least one of the solutions of the optimization problem (1.2) has at most n nonzero coordinates.
In this article we discuss the Sorted L-One Penalized Estimator (SLOPE) [4, 5, 23], defined as

$$\text{minimize } \left[\frac{1}{2}\|Y - Xb\|_2^2 + J_\Lambda(b)\right] \text{ over } b \in \mathbb{R}^p, \tag{1.3}$$

with the penalty

$$J_\Lambda(b) = \sum_{i=1}^p \lambda_i |b|_{(i)} \ ,$$

where the sequence $(|b|_{(1)}, \ldots, |b|_{(p)})$ with $|b|_{(1)} \geq \cdots \geq |b|_{(p)}$ contains absolute values of coordinates of b sorted in the nonincreasing order and the vector of hyperparameters $\Lambda = (\lambda_1, \ldots, \lambda_p)'$ satisfies $\lambda_1 > 0$ and $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$. SLOPE is an extension of OSCAR [6] (where the hyper-parameter Λ has arithmetically decreasing components) and an extension of LASSO [10, 20]. Indeed, if for all i we have $\lambda_i = \lambda > 0$, then the SLOPE penalty equals $\lambda\|b\|_1$ and coincides with the ℓ_1 penalty for the LASSO.
When the sequence Λ is strictly decreasing, SLOPE has substantially different properties than LASSO. Specifically, when the design matrix is orthogonal or the covariates are independent, SLOPE adapts to unknown signal sparsity and, contrarily to LASSO, obtains sharp minimax rates of the estimation and prediction errors [17] (see also [1, 2]). Moreover, in this case SLOPE allows for exact or asymptotic control of the False Discovery Rate (see, e.g., [5, 7, 8, 13]).
Another interesting feature of SLOPE is that it can group variables together so that the estimates of the regression coefficients in each group have exactly the same magnitude. As a result, SLOPE can yield more than n nonzero regression coefficients as long as the number of groups created is less than the rank of X [14] and can achieve full selection power in regimes where the lasso power is constrained [9]. In Figueiredo and Nowak [11], Bondell and Reich [6] it is observed that SLOPE clusters groups of correlated predictors, while in Kremer et al. [14] it is shown that the driving force behind the clustering phenomenon is the similarity of the influence of the predictors on the value of the response variable. Thus, clustering can also occur because of the similarity of the values of the regression coefficients. In particular, it can also occur when the design matrix is orthogonal (see Bogdan et al. [5], Skalski et al. [16]).
In this article we take a novel perspective on the low-dimensional pattern recovery by SLOPE. In a recent article [15], the authors argue that the low-dimensional patterns, which can be recovered by the penalized regression are characterized by properties of the penalty subdifferential.
For example, LASSO pattern corresponds to the sign vector of β. Therefore, under certain circumstances, LASSO can recover the sign of β with large probability. As explained in Section 1.2 of our paper, if the SLOPE sequence Λ is strictly decreasing, then the SLOPE pattern is substantially more refined and contains also the information on the ranking of the absolute values of the coordinates of β. Thus, under certain circumstances, SLOPE can recover both the proper sign and the proper ranking of the absolute values of β with a large probability. This clearly implies that SLOPE can identify clusters of the same absolute values in coordinates of β. Hence, from this new perspective, the clustering properties of SLOPE are just a specific realization of a much more important property, which is recovering the proper ranking of the explanatory variables. Consequently, it is natural to treat the triple: ranking,

clusters and the sign of β as a SLOPE pattern and to expect that, under certain circumstances, SLOPE will recover this pattern.

The main purpose of this article is to identify the circumstances under which the pattern recovery by SLOPE holds with a large probability. In case of LASSO, the capacity of the sign recovery is characterized by the well known Irrepresentability Condition, which takes the following form:

$$\ker(X_S) = \{0\} \quad \text{and} \quad \|X'_{\overline{S}} X_S (X'_S X_S)^{-1} \mathrm{sign}(\beta_S)\|_\infty \leq 1 \ , \tag{1.4}$$

where $S \subset \{1, \ldots, p\}$ is the set of columns for which the true regression coefficients are different from zero, $\overline{S} = \{1, \ldots, p\} \setminus S$ and $\mathrm{sign}(\beta_S)$ is the sign vector of these coefficients. This condition is necessary for the sign discovery by LASSO. It is also sufficient for discovering the sign of the sufficiently strong signal [24].

In this article we derive the analogous irrepresentability condition for SLOPE (see (4.7) and (4.8)) and prove that it is necessary for discovering the sign and the ranking of the coordinates of β, and it is sufficient to do so when the nonzero elements of β are large enough. We also present asymptotic results which illustrate the consistency and strong consistency of the pattern recovery when p is fixed and n diverges to infinity.

2. Preliminaries and basic notions.

Throughout this paper we will always suppose that

$$\lambda_1 > \cdots > \lambda_p > 0,$$

even though some results do not require this condition to be held.

For the design matrix $X \in \mathbb{R}^{n\times p}$, vector $Y \in \mathbb{R}^n$ and tuning parameter Λ, we define the set of SLOPE minimizers as

$$S_{X,\Lambda}(Y) = \underset{b\in\mathbb{R}^p}{\arg\min} \left[\frac{1}{2}\|Y - Xb\|_2^2 + J_\Lambda(b)\right].$$

In the multiple linear regression model (1.1) the SLOPE estimator of β is defined as a vector $\hat{\beta}^{\mathrm{SLOPE}}$ such that

$$\hat{\beta}^{\mathrm{SLOPE}} \in S_{X,\Lambda}(X\beta + \varepsilon).$$

Note that there may exists many SLOPE estimators of β. However, if $\ker(X) = \{0\}$, then $S_{X,\Lambda}(Y)$ consists of one element. As observed in [21] in the LASSO case, if there are two minimizers $\hat{\beta}_1 \neq \hat{\beta}_2$, then the fitted values must agree, i.e. $X\hat{\beta}_1 = X\hat{\beta}_2$ and so $\hat{\beta}_1 = \hat{\beta}_2$. Otherwise, by the strict convexity of the function $\|Y - \cdot\|_2^2$ and from the triangle inequality for the norm J_Λ, the value of $\frac{1}{2}\|Y - X\hat{\beta}_0\|_2^2 + J_\Lambda(\hat{\beta}_0)$, for $\hat{\beta}_0 = (\hat{\beta}_1 + \hat{\beta}_2)/2$, is smaller than the value for $\hat{\beta}_1$ and $\hat{\beta}_2$, contradicting their minimality. For a complete characterization of the unicity of SLOPE minimizer, see [15].

The column (resp. row) space of a matrix A is denoted by $\mathrm{col}(A)$ (resp. $\mathrm{row}(A)$). I_n denotes the identity matrix of rank n.

2.1. *SLOPE Pattern and related objects.*

The following definition of the pattern of a vector on $\mathbb{R}^p$ is equivalent to the definition of the SLOPE model introduced in [15].

DEFINITION 1. The SLOPE pattern is a function $\mathrm{patt}\colon \mathbb{R}^p \to \mathbb{Z}^p$ defined by

$$\mathrm{patt}(b)_i = \mathrm{sign}(b_i)\mathrm{rank}(|b_i|), \qquad i = 1, \ldots, p,$$

where $\mathrm{rank}(|b_i|) \in \{1, 2, \ldots, k\}$ is the rank of $|b_i|$ in a set of nonzero distinct values of $\{|b_1|, \ldots, |b_p|\}$. We adopt the convention that $\mathrm{sign}(0) = 0$.

Abusing terminology we call the elements $M \in \mathrm{patt}(\mathbb{R}^p)$ patterns as well. We denote the set of patterns in $\mathbb{Z}^p$ by $\mathcal{M}_p$.

Clearly, we have $b = 0 \in \mathbb{R}^p$ if and only if $\mathrm{patt}(b) = 0 \in \mathbb{Z}^p$.

REMARK 2. For $b \in \mathbb{R}^p$, we have
The SLOPE pattern defined above has the following properties, used in the definition of the SLOPE model in [15]. For any $b \in \mathbb{R}^p$, we have

(i) sign preservation: $\mathrm{sign}(\mathrm{patt}(b)) = \mathrm{sign}(b)$,
(ii) cluster preservation: $|b_i| = |b_j| \implies |\mathrm{patt}(b)_i| = |\mathrm{patt}(b)_j|$,
(iii) hierarchy preservation: $|b_i| > |b_j| \implies |\mathrm{patt}(b)_i| > |\mathrm{patt}(b)_j|$.

The sets of indices $i \in \{1, \dots, p\}$ for which the absolute values $|M_i|$ are equal are called clusters. For $M \in \mathcal{M}_p$, $\|M\|_\infty$ is the number of nonzero clusters of M.

DEFINITION 3. Let $M \neq 0$ be a pattern in $\mathbb{R}^p$ with $k = \|M\|_\infty$ nonzero clusters. The model matrix $U_M \in \mathbb{R}^{p \times k}$ is defined as follows

$$(U_M)_{ij} = \mathrm{sign}(m_i)\mathbf{1}_{(|m_i| = k+1-j)}, \qquad i \in \{1, \dots, p\}, \, j \in \{1, \dots, k\}.$$

If $M = 0$, we set $U_M = 0 \in \mathbb{R}^p$.

EXAMPLE 4. If $\beta = (-4, 3, 0, -3, 4)$, then $\mathrm{patt}(\beta) = (-2, 1, 0, -1, 2)$ and

$$U_M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

For $k \in \mathbb{N}$ we denote

$$\mathbb{R}^{k+} = \{\kappa \in \mathbb{R}^k \colon \kappa_1 > \ldots > \kappa_k > 0\}.$$

Definition 3 implies that for $0 \neq b \in \mathbb{R}^p$ and $k = \max\{\|M\|_\infty, 1\}$,

$$\mathrm{patt}(b) = M \iff \text{there exists } \kappa \in \mathbb{R}^{k+} \text{ such that } b = U_M \kappa. \tag{2.1}$$

If $b = U_M \kappa$, then κ_i equals the absolute value of terms in the i-th cluster of b, where the clusters are ordered decreasingly by their absolute value.
We are about to define the clustered matrix and the clustered parameter. For the purpose of the latter definition we denote element-wise absolute value $|M| = (|m| \colon m \in M)$ of $M \in \mathcal{M}_p$. Let $|M|_\downarrow$ be the vector $|M|$ ordered non-increasingly. Both of these objects play an essential role in the characterization of the pattern recovery by SLOPE, see Theorem 5.

DEFINITION 5. Let $M \neq 0$ be a pattern in $\mathbb{R}^p$ and $k = \max\{\|M\|_\infty, 1\}$.
For $X \in \mathbb{R}^{n \times p}$ we define the clustered matrix X by $\tilde{X}_M = X U_M \in \mathbb{R}^{n \times k}$.
For $\Lambda \in \mathbb{R}^{p+}$, the clustered parameter Λ is defined by $\tilde{\Lambda}_M = U'_{|M|_\downarrow} \Lambda \in \mathbb{R}^k$.

If $M = \mathrm{patt}(\beta)$ for $\beta \in \mathbb{R}^p$ satisfies $\|M\|_\infty < p$, the pattern $M = (m_1, \dots, m_p)'$ leads naturally to reduce the dimension of the design matrix X in the regression problem, by replacing X by $\tilde{X}_M$. Actually, if $\mathrm{patt}(\beta) = M$, then $X\beta = X U_M \kappa = \tilde{X}_M \kappa$ for $\kappa \in \mathbb{R}^{k+}$. In particular, in order to construct the clustered matrix $\tilde{X}_M$, by Definitions 3 and 5,

(i) null components $m_i = 0$ lead to discard the column X_i from the design matrix X,
(ii) a cluster $K \subset \{1, \dots, p\}$ of M (component of M equal in absolute value) leads to replace the columns $(X_i)_{i \in K}$ by one column equal to the signed sum: $\sum_{i \in K} \mathrm{sign}(m_i) X_i$.

2.2. *Pattern and the subdifferential of the SLOPE norm.* Recall that for a function f on $\mathbb{R}^p$, the subdifferential at $b \in \mathbb{R}^p$ is defined by:

$$\partial_f(b) = \{v \in \mathbb{R}^p \colon f(z) \geq f(b) + v'(z-b)\ \forall z \in \mathbb{R}^p\}.$$

PROPOSITION 1. *The SLOPE minimizer is characterized by the following conditions*

$$\begin{aligned}\hat\beta \in S_{X,\Lambda}(Y) &\overset{(1)}{\Longleftrightarrow} \pi^* := X'(Y - X\hat\beta) \in \partial_{J_\Lambda}(\hat\beta)\\ &\overset{(2)}{\Longleftrightarrow} \left[J^*_\Lambda(\pi^*) \leq 1 \text{ and } \hat\beta'\pi^* = J_\Lambda(\hat\beta)\right]\\ &\overset{(3)}{\Longleftrightarrow} \left[J^*_\Lambda(\pi^*) \leq 1 \text{ and } U'_M\pi^* = \tilde\Lambda_M\right],\end{aligned}$$

where $M = \mathrm{patt}(\hat\beta)$ *and* J^*_Λ *is the dual norm on* $\mathbb{R}^p$ *given by*

$$J^*_\Lambda(x) := \max\left\{\frac{|x|_{(1)}}{\lambda_1}, \ldots, \frac{\sum_{i=1}^p |x|_{(i)}}{\sum_{i=1}^p \lambda_i}\right\}.$$

PROOF. The equivalence (1) follows immediately from the general fact that a function f attains its minimum at a point b if and only if $0 \in \partial_f(b)$. For (2), see [12, Example VI (3.1)]. Equivalence (3) follows directly from Proposition 3.

□

When it comes to pattern of the SLOPE minimizer, the second condition in (3) is a substantial simplification compared to the second condition in (2). If the pattern is known, it gives an explicit system of linear equations for π^* and so for $\hat\beta$ as well. Moreover, (2) should be seen as an optimization problem, see (2.3).
The following Theorem 2 states that patterns $\mathcal{M}_p$ are in bijective correspondence with possible subdifferentials of the SLOPE norm J_Λ and it is proven in [15]. For the sake of completeness, we provide in the Appendix a short proof of Theorem 2 as a by-product of our explicit description of SLOPE norm subdifferentials in Proposition 3.

THEOREM 2. *Let* a *and* b *be any vectors in* $\mathbb{R}^p$*. Then,*

$$\mathrm{patt}(a) = \mathrm{patt}(b) \text{ if and only if } \partial_{J_\Lambda}(a) = \partial_{J_\Lambda}(b).$$

We denote

$$D_b := \partial_{J_\Lambda}(b).$$

Let $C_\Lambda = \{x \in \mathbb{R}^p \colon J^*_\Lambda(x) \leq 1\}$ be the closed unit ball in the dual SLOPE norm. It follows from the definition of J^*_Λ that $x \in C_\Lambda$ if and only if the following p inequalities hold

$$|x|_{(1)} + \ldots |x|_{(j)} \leq \lambda_1 + \ldots + \lambda_j, \qquad j = 1, \ldots, p. \tag{2.2}$$

It follows from a definition of dual norm $\|\cdot\|^*$ and the general fact that $\|\cdot\| = (\|\cdot\|^*)^*$,

$$J_\Lambda(\beta) = \max_{\pi \in C_\Lambda} \beta'\pi. \tag{2.3}$$

For each pattern $M \in \mathcal{M}_p$ we define the pattern affine space A_M by

$$A_M = \{x \in \mathbb{R}^p \colon U'_M x = \tilde\Lambda_M\}. \tag{2.4}$$

Note that if $M=0$, then $A_M=\mathbb{R}^p$. When $\|M\|_\infty=k>0$, then the set A_M is an affine subspace of $\mathbb{R}^p$ of dimension $p-k$.
Let $M=(m_1,\dots,m_p)'$ be a pattern with $k\geq 1$ clusters and let $p_j=|\{i:|m_i|=k+1-j\}|$ be the number of elements of the j-th biggest cluster. Denote $P_j=\sum_{i\leq j}p_i$, $j=1,\dots,k$. Then,

$$\tilde{\Lambda}_M=\begin{pmatrix}\lambda_1+\ldots+\lambda_{P_1}\\ \vdots\\ \lambda_{P_{k-1}+1}+\ldots+\lambda_{P_k}\end{pmatrix}. \tag{2.5}$$

Moreover, if $x\in D_M$, then (see the proof of Proposition 3 (i))

$$U_M'x=\begin{pmatrix}|x|_{(1)}+\ldots+|x|_{(P_1)}\\ \vdots\\ |x|_{(P_{k-1}+1)}+\ldots+|x|_{(P_k)}\end{pmatrix}.$$

Thus, the system of equations $U_M'x=\tilde{\Lambda}_M$ is equivalent to equalities in (2.2) for $j=P_i$, $i=1,\dots,k$.

PROPOSITION 3. *Let $b\in\mathbb{R}^p$ and $M=\mathrm{patt}(b)$.*

(i) *We have*

$$\begin{aligned}D_b&=C_\Lambda\cap A_M\\&=\left\{x\in\mathbb{R}^p:\begin{array}{l}(2.2)\text{ holds for all } j=1,2,\dots,p\\ \text{equalities hold in (2.2) for } j=p_1,p_1+p_2,\dots,p_1+\ldots+p_k\end{array}\right\},\end{aligned}$$

where p_j is the number of elements of the j-th cluster.

(ii) *The subdifferential depends on b only through the pattern of b, i.e.*

$$D_b=D_{\mathrm{patt}(b)}.$$

(iii) *A_M is the smallest affine space containing D_M, i.e.*

$$A_M=\mathrm{aff}(D_M). \tag{2.6}$$

As we shall see in Theorem 11 and Theorem 19, the relative interior of the pattern subdifferential D_M plays a preponderant role. We give an explicit form of $\mathrm{ri}(D_M)$ it in the following proposition, having the advantage of being computationally efficient.

PROPOSITION 4. *Let M be a pattern with $k\geq 1$ clusters and let p_j be the number of elements of the j-th cluster, $j=1,\dots,k$. Then,*

$$\mathrm{ri}(D_M)=\left\{x\in\mathbb{R}^p:\begin{array}{l}\text{equalities hold in (2.2) for } j=p_1,p_1+p_2,\dots,p_1+\ldots+p_k\\ \text{strict inequalities hold in (2.2) for } j\neq p_1,\dots,p_1+\ldots+p_k\end{array}\right\}.$$

PROOF. The set $\mathrm{ri}(D_M)$ is the interior of D_M within the affine space containing D_M, i.e. within A_M. Thus, the result follows by Proposition 3. □

3. Characterization of pattern recovery by SLOPE. We derive a novel characterization of the pattern of the SLOPE minimizer. Theorem below is the main tool of the proofs of the results on model recovery and is crucial for the paper. For the definition and the basic properties of the Moore-Penrose pseudo-inverse A^+ of a matrix A, see [3]. Recall the definitions of $\tilde{X}_M$ and $\tilde{\Lambda}_M$ from Definition 5.
In the following result we do not require the unicity of the SLOPE minimizer. Recall that $\tilde{P}_M=(\tilde{X}_M')^+\tilde{X}_M'=\tilde{X}_M\tilde{X}_M^+$ is the orthogonal projection onto $\mathrm{col}(\tilde{X}_M)$.

THEOREM 5. *Let $X \in \mathbb{R}^{n\times p}$, $Y \in \mathbb{R}^n$ and $\Lambda \in \mathbb{R}^{p+}$.*
Let $M \in \mathcal{M}_p$ be a SLOPE pattern and denote $k = \max\{\|M\|_\infty, 1\}$. Define

$$\pi := X'(\tilde{X}'_M)^+\tilde{\Lambda}_M + X'(I_n - \tilde{P}_M)Y. \tag{3.1}$$

There exists $\hat{\beta} \in S_{X,\Lambda}(Y)$ with $\mathrm{patt}(\hat{\beta}) = M$ if and only if the two conditions below hold true:

$$\begin{cases} \text{there exists } s \in \mathbb{R}^{k+} \text{ such that } \tilde{X}'_M Y - \tilde{\Lambda}_M = \tilde{X}'_M \tilde{X}_M s, & \text{(positivity condition)} \\ J^*_\Lambda(\pi) \le 1. & \text{(dual norm condition)} \end{cases}$$

If the positivity and dual norm conditions are satisfied, then $\hat{\beta} = U_M s$ and $\pi = X'(Y - X\hat{\beta}) \in D_M$.

PROOF OF THEOREM 5. *Necessity.* Let us assume that there exists $\hat{\beta} \in S_{X,\Lambda}(Y)$ with $\mathrm{patt}(\hat{\beta}) = M$. Consequently, $\hat{\beta} = U_M s$ for some $s \in \mathbb{R}^{k+}$.
By Proposition 1, $X'(Y - X\hat{\beta}) \in \partial_{J_\Lambda}(\hat{b}) = D_M$. Multiplying this inclusion by U'_M, due to (2.6), we get $\tilde{X}'_M(Y - X\hat{\beta}) = \tilde{\Lambda}_M$ and so

$$\tilde{X}'_M Y - \tilde{\Lambda}_M = \tilde{X}'_M X\hat{\beta} = \tilde{X}'_M \tilde{X}_M s. \tag{3.2}$$

The positivity condition is proven.
We apply $(\tilde{X}'_M)^+$ from the left to (3.2) and use the fact that $\tilde{P}_M = (\tilde{X}'_M)^+\tilde{X}'_M$ is the projection onto $\mathrm{col}(\tilde{X}_M)$: since $X\hat{\beta} \in \mathrm{col}(\tilde{X}_M)$, we have $\tilde{P}_M X\hat{\beta} = X\hat{\beta}$. Thus,

$$\tilde{P}_M Y - (\tilde{X}'_M)^+\tilde{\Lambda}_M = X\hat{\beta}.$$

The dual norm condition follows from Proposition 1, since

$$X'(Y - X\hat{\beta}) = X'(Y - (\tilde{P}_M Y - (\tilde{X}'_M)^+\tilde{\Lambda}_M)) = \pi.$$

Sufficiency. Assume that the positivity condition and the dual norm conditions hold true. Then, by the positivity condition, one may pick $s \in \mathbb{R}^{k+}$ for which

$$\tilde{X}'_M Y - \tilde{\Lambda}_M = \tilde{X}'_M \tilde{X}_M s.$$

We will show that $U_M s \in S_{X,\Lambda}(Y)$. Multiplying on the left by $(\tilde{X}'_M)^+$, the equality above gives

$$\tilde{P}_M Y - (\tilde{X}'_M)^+\tilde{\Lambda}_M = \tilde{P}_M \tilde{X}_M s = \tilde{X}_M s = XU_M s.$$

Moreover, by definition of U_M, we have $\mathrm{patt}(U_M s) = M$. Clearly, the following equalities occur

$$X'(Y - XU_M s) = X'(Y - (\tilde{P}_M Y - (\tilde{X}'_M)^+\tilde{\Lambda}_M)) = \pi. \tag{3.3}$$

Thus, by the dual norm condition we have $J^*_\Lambda(X'(Y - XU_M s) \le 1$.
Since $I_n - \tilde{P}_M$ is the projection onto $\mathrm{col}(\tilde{X}_M)^\perp$, we have $(I_n - \tilde{P}_M)\tilde{X}_M = 0$ and hence $U'_M X'(I_n - \tilde{P}_M)Y = \tilde{X}'_M(I_n - \tilde{P}_M)Y = 0$. Thus, by definition of π and equation (3), we get

$$U'_M\pi = \tilde{X}'_M(\tilde{X}'_M)^+\tilde{\Lambda}_M = \tilde{X}'_M(\tilde{X}'_M)^+(\tilde{X}'_M Y - \tilde{X}'_M\tilde{X}_M s) = \tilde{X}'_M(Y - \tilde{X}_M s) = \tilde{\Lambda}_M.$$

In view of (2.6), this implies that $\pi \in D_M$. Finally, by (3.3) and Proposition 1 we have $U_M s \in S_{X,\Lambda}(Y)$. □

REMARK 6. The assertion of Theorem 5 cannot be strengthened in the sense that if $S_{X,\Lambda}(Y)$ contains more than one element, then these different minimizers in general have different SLOPE patterns.

COROLLARY 6.

(i) *The following condition*

$$\tilde{\Lambda}_M \in \mathrm{col}(\tilde{X}_M') = \mathrm{row}(\tilde{X}_M) \tag{3.4}$$

is necessary for the positivity condition and thus also for the pattern recovery.

(ii) *The condition* (3.4) *is equivalent to*

$$X'(\tilde{X}_M')^+\tilde{\Lambda}_M \in A_M.$$

(iii) *The condition* (3.4) *is equivalent to* $\pi \in A_M$, *where* π *is defined in* (3.1).

PROOF. (i) is immediate. For (ii) observe that $\tilde{X}_M'(\tilde{X}_M')^+$ is the orthogonal projection onto $\mathrm{col}(\tilde{X}_M')$. Thus, (3.4) is equivalent to $\tilde{X}_M'(\tilde{X}_M')^+\tilde{\Lambda}_M = \tilde{\Lambda}_M$, which, by (2.6), is equivalent to $X'(\tilde{X}_M')^+\tilde{\Lambda}_M \in A_M$. Since $I_n - \tilde{P}_M$ is the orthogonal projection onto $\mathrm{col}(\tilde{X}_M)^\perp$, we have

$$U_M\left(\pi - X'(\tilde{X}_M')^+\tilde{\Lambda}_M\right) = X_M'(I_n - \tilde{P}_M)Y = 0.$$

Hence, (iii) follows. □

REMARK 7. Similar approach as in Theorem 5 for $\lambda_1 = \ldots = \lambda_p$ concerns the LASSO estimator. In particular, it allows new simple proofs of results on the sign recovery by LASSO. For LASSO we have the LASSO pattern being a sign vector, cf. [19]. For $S \in \{-1,0,1\}^p$, $\|S\|_1$ denotes the number of nonzero coordinates. If $\|S\|_1 = k \geq 1$, then we define the corresponding pattern matrix $U_S \in \mathbb{R}^{p\times k}$ by

$$U_S = \mathrm{diag}(S)_{\mathrm{supp}(S)},$$

where for $A \in \mathbb{R}^{p\times p}$ and $B \subset \{1,\ldots,p\}$, A_B denotes the submatrix of A obtained by keeping columns corresponding to indices in B. Observe that for any $0 \neq \beta \in \mathbb{R}^p$ there exist unique $S \in \{-1,0,1\}^p$ and $\kappa \in \mathbb{R}_+^k$ such that $\beta = U_S\kappa$. Define the reduced matrix and reduced parameter by

$$\tilde{X}_S = XU_S \qquad \text{and} \qquad \tilde{\lambda}_S = \lambda 1_k,$$

where $1_k = (1,\ldots,1)' \in \mathbb{R}^k$. It can be shown that the necessary and sufficient conditions for the LASSO pattern recovery are the following

$$\begin{cases} \text{there exists } \kappa \in \mathbb{R}_+^k \text{ such that } \tilde{X}_S'Y - \tilde{\lambda}_S = \tilde{X}_S'\tilde{X}_S\kappa, \\ \|X'(\tilde{X}_S')^+1_k + \frac{1}{\lambda}X'(I_n - \tilde{X}_S\tilde{X}_S^+)Y\|_\infty \leq 1. \end{cases}$$

4. Pattern recovery for fixed n.

In the regression problem (1.1) with $M = \mathrm{patt}(\beta) \neq 0$, by Theorem 5 we obtain

COROLLARY 7.

$$\begin{aligned} &\mathbb{P}\left(\exists \hat{\beta} \in S_{X,\Lambda}(Y) \textit{ such that } \mathrm{patt}(\hat{\beta}) = M\right) \\ &\quad = \mathbb{P}\left(J_\Lambda^*(\pi) \leq 1, \exists\, s \in \mathbb{R}^{k+}\colon\ \tilde{X}_M'Y - \tilde{\Lambda}_M = \tilde{X}_M'\tilde{X}_M s\right) \leq \mathbb{P}\left(J_\Lambda^*(\pi) \leq 1\right), \end{aligned} \tag{4.1}$$

where π *is a random vector defined in* (3.1).

Clearly, if the necessary condition for the positivity condition $\tilde{\Lambda}_M \notin \mathrm{row}(\tilde{X}_M)$ is not satisfied, then the probability of model recovery is 0. A similar upper bound for the sign recovery by LASSO is given in [18].
As we will argue, under natural assumptions (like in the noisy case with the large signal to noise ratio or in the informative noiseless case) we have

$$\mathbb{P}\left(\exists\hat{\beta} \in S_{X,\Lambda}(Y) \text{ such that } \mathrm{patt}(\hat{\beta}) = M\right) \approx \mathbb{P}\left(J^*_{\Lambda}(\pi) \leq 1\right),$$

meaning that the positivity condition is not restrictive and that the inequality bound in (4.1) is sharp.
Below we formulate the main result of this section, which states that the upper bound in (4.1) is attained when the signal magnitude diverges to infinity. For this aim, we consider a sequence of signals with the same pattern $M \in \mathcal{M}_p$, namely

$$(\beta^{(r)})_{r\geq 1}, \qquad \beta^{(r)} = U_M s^{(r)}, \quad s^{(r)} \in \mathbb{R}^{k+}, \tag{4.2}$$

whose strength is increasing in the following sense:

$$\Delta_r = \min_{1\leq i<k}\left(s_i^{(r)} - s_{i+1}^{(r)}\right) \overset{r\to\infty}{\longrightarrow} \infty, \text{ with the convention } s_{k+1}^{(r)} = 0. \tag{4.3}$$

With the above sequence, we consider the following sequence of linear regression models

$$Y^{(r)} = X\beta^{(r)} + \varepsilon, \qquad r = 1, 2, \ldots, \tag{4.4}$$

where $X \in \mathbb{R}^{n\times p}$ and $\varepsilon \in \mathbb{R}^n$ do not change with r.
We will assume that the vector of tuning parameters varies with r. More precisely, let

$$\Lambda^{(r)} = \alpha_r \Lambda, \tag{4.5}$$

where $\Lambda \in \mathbb{R}^p$ is fixed and $\alpha_r > 0$ for each r.

4.1. *Positivity condition.* We start with a lemma giving the sufficient assumptions on sequences $(\beta^{(r)})_{r\geq 1}$ and $(\Lambda^{(r)})_{r\geq 1}$ for the positivity condition to hold for large r.

LEMMA 8. *Let $\beta^{(r)}$ have the same pattern M for every r. Assume that $\beta^{(r)}$ satisfies (4.2) and (4.3). Then, if $\frac{\alpha_r}{\Delta_r} \to 0$, then the positivity condition holds for large r.*

PROOF. Recall that $\beta^{(r)} = U_M s^{(r)}$ for some $s^{(r)} \in \mathbb{R}^{k+}$. Recall that for any matrix A the matrix AA^+ is an orthogonal projector onto $\mathrm{col}(A)$. As $\tilde{\Lambda}_M \in \mathrm{col}(\tilde{X}'_M)$, we have $\tilde{\Lambda}_M = \tilde{X}'_M(\tilde{X}'_M)^+\tilde{\Lambda}_M$. Therefore, as $(\tilde{X}'_M)^+ = \tilde{X}'_M(\tilde{X}'_M\tilde{X}_M)^+$ and $Y^{(r)} = \tilde{X}_M s^{(r)} + \varepsilon$, we obtain

$$\begin{aligned}
\tilde{X}'_M Y^{(r)} - \tilde{\Lambda}_M^{(r)} &= \tilde{X}'_M\tilde{X}_M s^{(r)} - \tilde{\Lambda}_M^{(r)} + \tilde{X}'_M\varepsilon \\
&= \tilde{X}'_M\tilde{X}_M s^{(r)} - \tilde{X}'_M\tilde{X}_M(\tilde{X}'_M\tilde{X}_M)^+\tilde{\Lambda}_M^{(r)} + \tilde{X}'_M\tilde{X}_M(\tilde{X}'_M\tilde{X}_M)^+\tilde{X}'_M\varepsilon \\
&= \tilde{X}'_M\tilde{X}_M\Delta_r\left(\frac{s^{(r)}}{\Delta_r} - \frac{(\tilde{X}'_M\tilde{X}_M)^+\tilde{\Lambda}_M}{\Delta_r/\alpha_r} + \frac{(\tilde{X}'_M\tilde{X}_M)^+\tilde{X}'_M\varepsilon}{\Delta_r}\right).
\end{aligned}$$

$S^{(r)} := \frac{s^{(r)}}{\Delta_r}$ is a signal with strength not smaller than 1, i.e. the differences between neighbour terms $S_{m+1}^{(r)} - S_m^{(r)} \geq 1$ and $S_k^{(r)} \geq 1$. Then, since the next terms converge pointwise to 0, for r large enough the above sum is of the form $\tilde{X}'_M\tilde{X}_M w^{(r)}$ with $w^{(r)} \in \mathbb{R}^{k+}$, which finishes the proof.

□

REMARK 8. Lemma 8 evidently holds in the special case of Λ independent of r, i.e. when $\alpha_r = 1$.

4.2. *Dual norm condition.* Observe that the vectors π_r defined as in (3.1) can be represented in a following way:

$$\pi_r = \pi_r^{(1)} + \pi_r^{(2)},$$
$$\pi_r^{(1)} = X'(\tilde{X}'_M)^+ \tilde{\Lambda}_M^{(r)}, \qquad \pi_r^{(2)} = X'(I_n - \tilde{P}_M)Y^{(r)} = X'(I_n - \tilde{P}_M)\varepsilon$$

Note that in the noiseless case ($\varepsilon = 0$) we have $\pi_r^{(2)} = 0$.
Then the dual norm condition $J^*_{\Lambda_r}(\pi_r) \leq 1$ is equivalent to

$$1 \geq J^*_\Lambda\left(\frac{\pi_r}{\alpha_r}\right) = J^*_\Lambda\left(\frac{\pi_r^{(1)}}{\alpha_r} + \frac{\pi_r^{(2)}}{\alpha_r}\right) = J^*_\Lambda\left(X'(\tilde{X}'_M)^+\tilde{\Lambda}_M + \frac{\pi_r^{(2)}}{\alpha_r}\right). \tag{4.6}$$

Thus, if $\varepsilon \neq 0$, then the properties of pattern recovery require that $\alpha_r \overset{r\to\infty}{\longrightarrow} \infty$.

4.3. *Consistency of pattern of the SLOPE estimator.*

DEFINITION 9. We say that X and Λ satisfy the SLOPE irrepresentability (SLOPE IR) condition

$$J^*_\Lambda\left(X'(\tilde{X}'_M)^+\tilde{\Lambda}_M\right) \leq 1 \text{ and } \tilde{\Lambda}_M \in \mathrm{row}(\tilde{X}_M), \tag{4.7}$$

where $M = \mathrm{patt}(\beta)$.

REMARK 10. Recall that the condition $\tilde{\Lambda}_M \in \mathrm{row}(\tilde{X}_M)$ is necessary for the pattern recovery (Corollary 6 (i)). When $\ker(\tilde{X}_M) = \{0\}$, SLOPE IR condition takes a form

$$J^*_\Lambda\left(X'\tilde{X}_M(\tilde{X}'_M\tilde{X}_M)^{-1}\tilde{\Lambda}_M\right) \leq 1. \tag{4.8}$$

REMARK 11. By Corollary 6 (ii), the SLOPE irrepresentability condition (4.7) is equivalent to condition

$$X'(\tilde{X}'_M)^+\tilde{\Lambda}_M \in D_M. \tag{4.9}$$

PROPOSITION 9. *Suppose that there is no error, i.e.* $\varepsilon = 0$ *and that* $\Delta_r \to \infty$.
For any fixed $\Lambda \in \mathbb{R}^{k+}$*, the pattern recovery is equivalent to SLOPE IR condition* (4.7).
Moreover, if the SLOPE IR condition does not hold, then for any $\hat{\beta} \in S_{X,J_\Lambda}(X\beta)$ *we have* $\mathrm{patt}(\hat{\beta}) \neq \mathrm{patt}(\beta)$.

PROOF. We have $\pi_r^{(2)} = 0$. The theorem follows immediately from Theorem 5, formula (4.6) and Lemma 8. □

LEMMA 10. *Let* $\varepsilon \in \mathbb{R}^n$ *be a random error. Let* $Y^{(r)} = X\beta^{(r)} + \varepsilon$ *and assume that* $\tilde{\Lambda}_M \in \mathrm{row}(\tilde{X}_M)$. *Then the upper bound* (4.1) *is asymptotically reached when* r *tends to* ∞*:*

$$\lim_{r\to\infty} \mathbb{P}\left(\exists\hat{\beta} \in S_{X,\Lambda}(Y^{(r)}) \text{ such that } \mathrm{patt}(\hat{\beta}) = M\right) = \mathbb{P}\left(J^*_\Lambda(\pi) \leq 1\right),$$

where π *is defined in* (3.1). *Moreover, if* ε *is a Gaussian vector* $N(0, \sigma^2 I_n)$*, then*

$$\pi \sim \mathrm{N}\left(X'(\tilde{X}'_M)^+\tilde{\Lambda}_M, \sigma^2 X'(I_n - \tilde{P}_M)X\right).$$

PROOF. It is sufficient to show that the probability of the positivity condition tends to 1. This is an instant conclusion of Lemma 8 and Remark 8.
The distribution of π may be deduced directly from $\pi^{(1)}$ being deterministic and $\pi^{(2)}$ being a Gaussian vector multiplied by a matrix. □

When there is a noise $\varepsilon \neq 0$, the sufficient condition for the discovery of strong signals takes a slightly stronger form, called Open SLOPE IR condition. If the vector $X'(\tilde{X}'_M)^+\tilde{\Lambda}_M$ is in the boundary of D_M, adding an error term to it may provoke getting out of the dual ball C_Λ and violating the dual norm condition.

DEFINITION 12. We say that X and Λ satisfy the Open SLOPE irrepresentability (Open SLOPE IR) Condition if

$$X'(\tilde{X}'_M)^+\tilde{\Lambda}_M \in \mathrm{ri}(D_M), \tag{4.10}$$

where $\mathrm{ri}(D_M)$ is the relative interior of the subdifferential D_M of the SLOPE norm at the pattern vector.

We will use the sharpness of the upper bound (4.1) to prove the sufficiency of the Open SLOPE IR condition for the SLOPE pattern recovery when the signal to noise ratio diverges to infinity.
Theorem 11 formally states the sufficiency of the Open SLOPE IR condition (4.10).

THEOREM 11. *(Sufficiency) Let $Y^{(r)} = X\beta^{(r)} + \varepsilon$, where $\beta^{(r)}$ is the sequence of regression coefficient vectors defined in (4.2) and $\varepsilon \in \mathbb{R}^n$ is a random error.*
Assume (4.3) and that a sequence (α_r) is such that $\alpha_r \to \infty$ and $\alpha_r/\Delta_r \to 0$ as $r \to \infty$. Let $\Lambda_r = \alpha_r\Lambda$ for $\Lambda \in \mathbb{R}^{p+}$. If the Open SLOPE IR condition is satisfied, then

$$\lim_{r\to\infty} \mathbb{P}\left(\exists\hat{\beta} \in S_{X,\Lambda_r}(Y^{(r)}) \text{ such that } \mathrm{patt}(\hat{\beta}) = M\right) = 1. \tag{4.11}$$

PROOF. We prove that for all realizations of $\varepsilon(\omega)$ there exists such r_0 that if $r \geq r_0$, then there exists

$$\hat{\beta} \in S_{X,\Lambda_r}(Y^{(r)}) \text{ such that } \mathrm{patt}(\hat{\beta}) = M.$$

We apply Theorem 5. By the Open SLOPE IR condition, $\frac{\pi_r^{(1)}}{\alpha_r} \in ri(D_M)$. Then the above statement is a quick consequence of a formula (4.6), Lemma 8 and the fact that $\frac{\pi_r^{(2)}}{\alpha_r} \in A_M$. The convergence for all realizations of $\varepsilon(\omega)$ implies the convergence in probability, so that (4.11) holds.

□

COROLLARY 12. *(Necessity) Assume that the error ε has a symmetric distribution. If the SLOPE IR condition does not hold, then the probability of the pattern recovery is smaller than $1/2$.*

PROOF. The result easily follows from the separation theorem, convexity of C_Λ and the argument used in [22], see also [18]. □

For LASSO, a similar result on the probability of sign recovery is given in [22]. Note that another proof of Corollary 12 may be given using Theorem 2 in [19] and our Proposition 9.

5. Asymptotic Pattern recovery. In this section we discuss asymptotic properties of the SLOPE estimator in the low-dimensional regression model in which p is fixed and the sample size n tends to infinity.
For each $n \geq p$ we consider a linear regression problem

$$Y_n = X_n\beta + \varepsilon_n, \tag{5.1}$$

where $Y_n \in \mathbb{R}^n$ is a vector of observations, $X_n \in \mathbb{R}^{n\times p}$ is a random design matrix, $\beta \in \mathbb{R}^p$ is a vector of unknown regression coefficients and $\varepsilon_n \in \mathbb{R}^n$ is a noise term with the normal distribution $\mathrm{N}(0, \sigma^2 I_n)$. We assume that

$$X_n \text{ and } \varepsilon_n \text{ are independent for each } n, \tag{5.2}$$

but we do not make any assumptions on the dependence structure of ε_n and ε_m for $n \neq m$. In particular, our setting covers the case of deterministic X_n.
From now on we suppose that the random matrices X_n, $n \geq 1$, have full column rank almost surely, i.e.

$$\ker(X_n) = \{0\} \quad \text{a.s.} \tag{5.3}$$

When defining the sequence $(\hat\beta_n^{\mathrm{SLOPE}})_n$ of SLOPE estimators, we will assume that the vector of tuning parameters varies with n. More precisely, we will assume that

$$\Lambda_n = \alpha_n\Lambda, \tag{5.4}$$

where $\Lambda \in \mathbb{R}^{p+}$ is fixed and $(\alpha_n)_n$ is a sequence of positive numbers. Under (5.3) the set of SLOPE minimizers $S_{X_n,\Lambda_n}(Y_n)$ consists of one element. For each n, we denote this unique element by $\hat\beta_n^{\mathrm{SLOPE}}$.
We further assume that the design matrices $X_1, X_2, \ldots$ satisfy the condition

$$\frac{1}{n}X_n'X_n \xrightarrow{a.s.} C, \tag{5.5}$$

where C is a deterministic positive definite symmetric $p \times p$ matrix.

REMARK 13. Condition (5.5) is satisfied in the natural setting when the rows of $X_n \in \mathbb{R}^{n\times p}$ are i.i.d. random vectors. Let $\xi = (\xi_1, \ldots, \xi_p)$ be a random vector and assume that each (independent) row of X_n has the same law as ξ. In such case, the assumption (5.5) is implied by $\mathbb{E}[\xi_i^2] < \infty$ for all $i = 1, \ldots, p$. Then, the strong law of large numbers ensures that (5.5) holds with $C = (C_{ij})_{ij}$, where $C_{ij} = \mathbb{E}[\xi_i\xi_j]$ for $i, j = 1, \ldots, p$. Moreover, C is positive definite if and only if the random variables $(\xi_1, \ldots, \xi_p)$ are linearly independent a.s. Indeed, for $t \in \mathbb{R}^p$ we have

$$t'Ct = \mathbb{E}\left[\left(\sum_{i=1}^p t_i\xi_i\right)^2\right] > 0$$

if and only if $\sum_{i=1}^p t_i\xi_i \neq 0$ a.s. for all $t \in \mathbb{R}^p$.

Under (5.5), the strong consistency of $\hat\beta_n^{\mathrm{SLOPE}}$ can be characterized in terms of behaviour of the tuning parameter.

THEOREM 13. *Consider a linear regression* (5.1) *and assume* (5.2), (5.3) *and* (5.5). *Let* $\Lambda_n = (\lambda_1^{(n)}, \ldots, \lambda_p^{(n)})'$. *If* $\beta \neq 0$, *then* $\hat\beta_n^{\mathrm{SLOPE}} \xrightarrow{a.s.} \beta$ *if and only if*

$$\lim_{n\to\infty} \frac{\lambda_1^{(n)}}{n} = 0. \tag{5.6}$$

If $\beta = 0$ *and* (5.6) *holds true, then* $\hat\beta_n^{\mathrm{SLOPE}} \xrightarrow{a.s.} 0$.

As before, $M = \mathrm{patt}(\beta)$, $k = \|M\|_\infty$ and U_M is the corresponding model matrix. Recall from Definition 5 that $\tilde{\Lambda} = U'_{|M|\downarrow}\Lambda$, $\tilde{\Lambda}_n = \alpha_n\tilde{\Lambda}$ and $\tilde{X}_n = X_n U_M$. To ease the notation, we write the clustered matrices and clustered parameters without the subscript indicating the model M. Let

$$s_n := (\tilde{X}'_n\tilde{X}_n)^{-1}[\tilde{X}'_n Y_n - \tilde{\Lambda}_n],$$

$$\pi_n := X'_n(\tilde{X}'_n)^+\tilde{\Lambda}_n + X'_n(I_n - \tilde{P}_n)Y_n,$$

where $\tilde{P}_n = (\tilde{X}'_n)^+\tilde{X}'_n$. We note that if $\ker(X_n) = \{0\}$, then $(\tilde{X}'_n)^+ = \tilde{X}_n(\tilde{X}'_n\tilde{X}_n)^{-1}$ and so $\tilde{P}_n = \tilde{X}_n(\tilde{X}'_n\tilde{X}_n)^{-1}\tilde{X}'_n$.
As the positivity condition and the dual norm condition from Theorem 5 have different nature, we consider them separately in the two following subsections. It turns out that for the positivity condition (α_n) cannot be too large, while for the dual norm condition it cannot be too small. In this way we determine the assumptions for the sequence $(\alpha_n)_n$ for which both positivity and dual norm conditions hold true and for which the pattern is recovered.

5.1. *Positivity condition.* If $M = 0$, then the positivity condition is trivially satisfied. Thus, we consider $M \neq 0$. Since $\tilde{X}'_n\tilde{X}_n$ is invertible, the positivity condition is equivalent to

$$s_n := (\tilde{X}'_n\tilde{X}_n)^{-1}[\tilde{X}'_n Y_n - \tilde{\Lambda}_n] \in \mathbb{R}^{k+}.$$

Let $s_0 \in \mathbb{R}^{k+}$ be defined through $\beta = U_M s_0$, where $k = \|M\|_\infty$.

LEMMA 14. *Let $M = \mathrm{patt}(\beta)$. Assume (5.2) and (5.5).*
If $\alpha_n/n \to 0$, then the positivity condition is almost surely satisfied for large n.

PROOF. We show in the Appendix that if $\alpha_n/n \to 0$, then

$$s_n \xrightarrow{a.s.} s_0. \tag{5.7}$$

Since $s_0 \in \mathbb{R}^{k+}$ and $\mathbb{R}^{k+}$ is an open set, there exists $n_0(\omega)$ such $s_n(\omega) \in \mathbb{R}^{k+}$ for all $n \geq n_0(\omega)$, for $\mathbb{P}$ almost all ω. □

5.2. *Dual norm condition.* For $M \neq 0$ we denote

$$\pi_n^{(1)} = X'_n(\tilde{X}'_n)^+\tilde{\Lambda}_n, \qquad \pi_n^{(2)} = X'_n(I_n - \tilde{P}_n)Y_n,$$

$$\pi_n = \pi_n^{(1)} + \pi_n^{(2)},$$

which simplifies in the $M = 0$ case to $\pi_n = \pi_n^{(2)} = X'_n Y_n$.
For any $\alpha > 0$ we have $C_{\alpha\Lambda} = \alpha C_\Lambda$. Recall that the dual norm condition is $J^*_{\Lambda_n}(\pi_n) \leq 1$, which, under (5.4), is equivalent to

$$1 \geq J^*_\Lambda\left(\alpha_n^{-1}\pi_n\right) = J^*_\Lambda\left(\alpha_n^{-1}\pi_n^{(1)} + \frac{\sqrt{n}}{\alpha_n}n^{-1/2}\pi_n^{(2)}\right).$$

In view of results shown below, $\alpha_n^{-1}\pi_n^{(1)}$ converges almost surely, while $n^{-1/2}\pi_n^{(2)}$ converges in distribution to a Gaussian vector. Thus, the pattern recovery properties of SLOPE estimator strongly depend on the behavior of the sequence $(\alpha_n/\sqrt{n})_n$.

LEMMA 15. *Assume $M \neq 0$, (5.4) and (5.5). Then,*

$$\frac{1}{\alpha_n}\pi_n^{(1)} \xrightarrow{a.s.} CU_M(U'_M C U_M)^{-1}\tilde{\Lambda}.$$

PROOF. We note that (5.5) implies that almost surely

$$\lim_{n\to\infty} X_n'\tilde{X}_n(\tilde{X}_n'\tilde{X}_n)^{-1} = \lim_{n\to\infty}\frac{1}{n}X_n'X_nU_M(U_M'n^{-1}X_n'X_nU_M)^{-1} = CU_M(U_M'CU_M)^{-1},$$

what ends the proof. □

LEMMA 16. *Consider a linear regression* (5.1) *and assume* (5.2), (5.3), (5.4) *and* (5.5). *Then, the sequence* $\left(n^{-1/2}\pi_n^{(2)}\right)_n$ *converges in distribution to a Gaussian vector* Z *on* $\mathrm{aff}(D_M)$ *with*

$$Z \sim \mathrm{N}\left(0, \sigma^2\left[C - CU_M(U_M'CU_M)^{-1}U_M'C\right]\right).$$

PROOF. We note that when $\beta = U_M s_0$, then the linear regression model (5.1) takes the form $Y_n = \tilde{X}_n s_0 + \varepsilon_n$. Since $\tilde{P}_n$ is the projection matrix onto the vector subspace $\mathrm{col}(\tilde{X}_n)$, we have $(I_n - \tilde{P}_n)\tilde{X}_n = 0$. Thus,

$$\alpha_n^{-1}X_n'(I_n - \tilde{P}_n)Y_n = \alpha_n^{-1}X_n'(I_n - \tilde{P}_n)(\tilde{X}_n s_0 + \varepsilon_n) = \alpha_n^{-1}X_n'(I_n - \tilde{P}_n)\varepsilon_n.$$

We consider the sequence of random matrices $(Q_n)_n$ defined by

$$Q_n' = X_n'(I_n - \tilde{P}_n).$$

We have

$$n^{-1}Q_n'Q_n = n^{-1}X_n'(I_n - \tilde{P}_n)X_n.$$

Since $n(\tilde{X}_n'\tilde{X}_n)^{-1} \xrightarrow{a.s.} (U_M'CU_M)^{-1}$, it follows that (A.6) holds with

$$C_0 = C - CU_M(U_M'CU_M)^{-1}U_M'C.$$

Lemma 22 completes the proof. □

PROPOSITION 17. *Under the assumptions of Lemma 16, if*

$$\lim_{n\to\infty}\frac{\alpha_n}{\sqrt{n\log n}} = \infty,$$

then $\alpha_n^{-1}\pi_n^{(2)} \xrightarrow{a.s.} 0$.

PROOF. Follows directly from Lemma 16 and Lemma 21 (i) with $f(n) = \alpha_n/\sqrt{n\log(n)}$. □

5.3. *Consistency of pattern of the SLOPE estimator.* The symbols D_M, $\mathrm{ri}(D_M)$ and A_M denote the subdifferential, its relative interior and the affine space corresponding to fixed Λ.

DEFINITION 14. Let $M = \mathrm{patt}(\beta)$. We say that the matrix C satisfy the SLOPE Irrepresentability (SLOPE IR) condition if

$$J_\Lambda^*\left(CU_M(U_M'CU_M)^{-1}\tilde{\Lambda}\right) \leq 1. \tag{5.8}$$

We say that the matrix C satisfies the Open SLOPE Irrepresentability (Open SLOPE IR) condition if

$$CU_M(U_M'CU_M)^{-1}\tilde{\Lambda} \in \mathrm{ri}(D_M). \tag{5.9}$$

First of our result concerns with the consistency of the pattern recovery by the SLOPE estimator.

THEOREM 18. *Under the assumptions of Lemma 16, let* $\hat{\beta}_n^{\mathrm{SLOPE}}$ *be the unique SLOPE estimator of* $\beta \in \mathbb{R}^p$.

(i) *If* $\alpha_n = \sqrt{n}$, *then*

$$\lim_{n\to\infty} \mathbb{P}\left(\mathrm{patt}(\hat{\beta}_n^{\mathrm{SLOPE}}) = \mathrm{patt}(\beta)\right) = \mathbb{P}\left(J_\Lambda^*(Z) \le 1\right),$$

where $Z \sim N(CU_M(U_M'CU_M)^{-1}\tilde{\Lambda}, \sigma^2[C - CU_M(U_M'CU_M)^{-1}U_M'C])$.

(ii) *Assume* (5.8). *The pattern of SLOPE estimator is consistent, i.e.*

$$\mathrm{patt}(\hat{\beta}_n^{\mathrm{SLOPE}}) \xrightarrow{\mathbb{P}} \mathrm{patt}(\beta),$$

if and only if

$$\lim_{n\to\infty} \frac{\alpha_n}{n} = 0 \qquad \textit{and} \qquad \lim_{n\to\infty} \frac{\alpha_n}{\sqrt{n}} = \infty.$$

PROOF. (i) is a direct consequence of Lemmas 14, 15 and 16. It is clear that the sequence $(\alpha_n^{-1}\pi_n)_n$ converges in probability to 0 if and only if $\alpha_n/\sqrt{n} \to \infty$. □

The above conditions on $(\alpha_n)_n$ in general do not suffice for the almost sure convergence of the SLOPE pattern - see Remark 15. The following asymptotic theorem concerns the strong consistency.

THEOREM 19. *Under the assumptions of Lemma 16, let* $\hat{\beta}_n^{\mathrm{SLOPE}}$ *be the unique SLOPE estimator of* $\beta \in \mathbb{R}^p$.
Assume that a sequence $(\alpha_n)_n$ *satisfies*

$$\lim_{n\to\infty} \frac{\alpha_n}{n} = 0 \qquad \textit{and} \qquad \lim_{n\to\infty} \frac{\alpha_n}{\sqrt{n\log n}} = \infty. \tag{5.10}$$

(i) *If the Open SLOPE IR* (5.9) *condition is satisfied, then the sequence* $(\hat{\beta}_n^{\mathrm{SLOPE}})_n$ *recovers almost surely the pattern of* β *asymptotically, i.e.*

$$\mathrm{patt}(\hat{\beta}_n^{\mathrm{SLOPE}}) \xrightarrow{a.s.} \mathrm{patt}(\beta). \tag{5.11}$$

(ii) *The SLOPE IR* (5.8) *condition is necessary for the asymptotic pattern recovery* (5.11).

PROOF. (i) By Lemma 14, the positivity condition is satisfied for large n almost surely. By Lemma 15 and Proposition 17, we have

$$a_n := \frac{1}{\alpha_n}\pi_n \xrightarrow{a.s.} CU_M(U_M'CU_M)^{-1}\tilde{\Lambda} =: a_0.$$

By Corollary 6 (iii), we have $\pi_n \in \alpha_n A_M$, thus $a_n \in A_M$. By the Open SLOPE IR condition $a_0 \in \mathrm{ri}(D_M)$, and thus it follows that $a_n \in D_M \subset C_\Lambda$ almost surely for sufficiently large n. Therefore $J_{\Lambda_n}(\pi_n) \le 1$ for large n almost surely.
(ii) By Lemma 15 and Proposition 17 we have $\mathbb{P}(A) = 1$, where

$$A = \left\{\omega \colon \lim_{n\to\infty} \frac{\pi_n(\omega)}{\alpha_n} = CU_M(U_M'CU_M)^{-1}\tilde{\Lambda}\right\}.$$

Suppose that $\mathrm{patt}(\hat{\beta}_n^{\mathrm{SLOPE}}(\omega)) \to \mathrm{patt}(\beta)$ for $\omega \in A$. Then it follows from Theorem 5 that for sufficiently large n we have $J_{\Lambda_n}^*(\pi_n(\omega)) \le 1$ or equivalently, $a_n(\omega) \in C_\Lambda$. Since $(a_n(\omega))_n$ is convergent and the set C_Λ is closed, it follows that (5.8) is satisfied. □

REMARK 15. Assume that the Open SLOPE IR condition is satisfied. Assume (5.4) with $\alpha_n = c\sqrt{n\log n}$ for $c > 0$. Then (5.10) is not satisfied. If random vectors $(\varepsilon_n)_n$ are independent and $(X_n)_n$ are deterministic matrices, then with positive probability, the true SLOPE pattern is not recovered.

APPENDIX A: PROOFS

A.1. Proofs from Section 2.

COROLLARY 20. *Let $M = (m_1, \dots, m_p)'$ be a pattern with k clusters. Assume that $x \in D_M$. Then,*

(i) $\{|x_i|: |m_i| = k\} > \{|x_i|: |m_i| = k-1\} > \dots > \{|x_i|: m_i = 0\}$*, where for $A, B \subset \mathbb{R}$, we write $A > B$ if $a > b$ for all $a \in A$ and $b \in B$.*
(ii) $x_i m_i \geq 0$*, $i = 1, \dots, p$. If $m_i \neq 0$, then $x_i m_i > 0$.*

PROOF. By Proposition 3 (i) we have

$$D_M = \left\{ x \in \mathbb{R}^p : \begin{array}{l} \text{equalities hold in (2.2) for } j = p_1, p_1+p_2, \dots, p_1 + \dots + p_k \\ \text{(2.2) holds for } j \neq p_1, p_1+p_2, \dots, p_1 + \dots + p_k \end{array} \right\},$$

where p_j is the number of elements of the j-th biggest cluster. The strict inequalities in (i) need to be justified. The inequality (2.2) for $j = p_i - 1$ and the equality (2.2) for $j = p_i$ imply that $|x|_{(p_i)} \geq \lambda_{p_i}$. The same equality and the inequality (2.2) for $j = p_i + 1$ imply that $|x|_{(p_i+1)} \leq \lambda_{p_i+1}$. We have $\lambda_{p_i} > \lambda_{p_i+1}$, so that $|x|_{(p_i)} > |x|_{(p_i+1)}$.
We have $x \in D_M$ if and only if $x \in C_\Lambda$ and $M'x = J_\Lambda(M) = \max_{\pi \in C_\Lambda} M'\pi$. Since for any $\varepsilon \in \{-1, 1\}^p$ the vector $(\varepsilon_1 x_1, \dots, \varepsilon_p x_p)$ belongs to C_Λ, we have

$$M'x \geq \sum_{i=1}^{p} \varepsilon_i m_i x_i,$$

which ensures (ii). ☐

PROOF OF PROPOSITION 3. (i) Let $\beta \in \mathbb{R}^p$ and $M = \text{patt}(\beta) \neq 0$. By Proposition 1 (2), we know that $\pi^* \in C_\Lambda$ belongs to $D_\beta = D_M$ if and only if $\beta'\pi^* = J_\Lambda(\beta)$. We will show that if $\pi^* \in C_\Lambda$, then

$$\beta'\pi^* = J_\Lambda(\beta) \quad \text{if and only if} \quad U_M'\pi^* = \tilde{\Lambda}_M. \tag{A.1}$$

First we will need a precise description of $\tilde{\Lambda}_M$, the clustered parameter.
Let $M = (m_1, \dots, m_p)'$ be a pattern with k clusters and let $p_j = |\{i : |m_i| = k + 1 - j\}|$ be the number of elements of the j-th biggest cluster. Denote $P_j = \sum_{i \leq j} p_i$ with $P_0 = 0$. Subsequent columns of the model matrix U_M contain $p_1, \dots, p_k$ nonzero elements 1 or -1. Then, by (2.5), we have

$$\tilde{\Lambda}_M = U_{|M|\downarrow}'\Lambda = \begin{pmatrix} \lambda_1 + \dots + \lambda_{P_1} \\ \vdots \\ \lambda_{P_{k-1}+1} + \dots + \lambda_{P_k} \end{pmatrix}. \tag{A.2}$$

Clearly, if $\beta = (\beta_1, \dots, \beta_p)'$ is any vector with $\text{patt}(\beta) = M$, then there exists a vector $s = (s_1, \dots, s_k)' \in \mathbb{R}^{k+}$ such that $\beta = U_M s$ and hence p_1 of the absolute values $|\beta_1|, \dots, |\beta_p|$ are equal to s_1, p_2 of them are equal to s_2, …, p_k of them are equal to s_k. In particular, we have

$$J_\Lambda(\beta) = \sum_{i=1}^{p} \lambda_i |U_M s|_{(i)} = \sum_{i=1}^{k} s_i \sum_{j=P_{i-1}+1}^{P_i} \lambda_j.$$

Sufficiency in (A.1). Suppose that $\pi^* = (\pi_1^*, \dots, \pi_p^*)' \in C_\Lambda$ satisfies the condition $U_M'\pi^* = \tilde{\Lambda}_M$. Then, $\pi^* \in D_M$, because $J_\Lambda(\beta) = s_1(\lambda_1 + \dots + \lambda_{P_1}) + s_2(\lambda_{P_1+1} + \dots + \lambda_{P_2}) + \dots + {}$ $+ s_k(\lambda_{P_{k-1}+1} + \dots + \lambda_{P_k}) = s'\tilde{\Lambda}_M = s'U_M'\pi^* = (U_M s)'\pi^* = \beta'\pi^*$.

Necessity in (A.1). Suppose that $\pi \in C_\Lambda$ satisfies $\beta'\pi^* = J_\Lambda(\beta)$. Recall by (2.3) that $J_\Lambda(\beta) = \max_{\pi \in C_\Lambda} \beta'\pi$. By Corollary 20 (ii), it follows that $\text{sign}(\beta_i) \cdot \text{sign}(\pi_i^*) \geq 0$, $i = 1, \ldots, p$. Moreover, the rearrangement inequality implies the following: if $|\beta_i| = s_1$, then $|\pi_i^*| \in \{|\pi^*|_{(1)}, \ldots, |\pi^*|_{(P_1)}\}$, if $|\beta_i| = s_2$, then $|\pi_i^*| \in \{|\pi^*|_{(P_1+1)}, \ldots, |\pi^*|_{(P_2)}\}$ and so on. These two facts imply that

$$U_M'\pi^* = \begin{pmatrix} |\pi^*|_{(1)} + \ldots + |\pi^*|_{(P_1)} \\ \vdots \\ |\pi^*|_{(P_{k-1}+1)} + \ldots + |\pi^*|_{(P_k)} \end{pmatrix}.$$

To prove that $U_M'\pi^* = \tilde{\Lambda}_M$, we use the fact that $J_\Lambda^*(\pi^*) \leq 1$. By the inequalities (2.2) there exist nonnegative $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$, such that $\sum_{j=1}^{P_i} |\pi^*|_{(j)} = \sum_{j=1}^{P_i} \lambda_j - \varepsilon_i, i = 1, 2, \ldots, k$. Hence we have $\sum_{j=P_{i-1}+1}^{P_i} |\pi^*|_{(j)} = \sum_{j=P_{i-1}+1}^{P_i} \lambda_j + \varepsilon_{i-1} - \varepsilon_i$, for $i = 1, 2, \ldots, k$, where $\varepsilon_0 = 0$. This implies that

$$\begin{aligned}
\beta'\pi^* &= \sum_{i=1}^{p} \beta_i \pi_i^* = \sum_{i=1}^{p} |\beta_i||\pi_i^*| = \sum_{i=1}^{k} s_i \sum_{j=P_{i-1}+1}^{P_i} |\pi^*|_{(j)} \\
&= \sum_{i=1}^{k} s_i \sum_{j=P_{i-1}+1}^{P_i} \lambda_j + \sum_{i=1}^{k} s_i(\varepsilon_{i-1} - \varepsilon_i) \\
&= J_\Lambda(\beta) + \varepsilon_0 s_1 + \sum_{i=1}^{k-1} \varepsilon_i (s_{i+1} - s_i) - \varepsilon_k s_k.
\end{aligned}$$

Since ε_i's are nonnegative with $\varepsilon_0 = 0$ and since $s_{i+1} - s_i < 0$, $i = 0, \ldots, k-1$ and $s_k > 0$, it follows that $(\beta)'\pi^* = J_\Lambda(\beta) \iff \varepsilon_1 = \varepsilon_2 = \ldots = \varepsilon_k = 0$.
(ii) and (iii) follow directly from (i). ∎

PROOF OF THEOREM 2. Let $M = (m_1, \ldots, m_p)'$ and $\widetilde{M} = (\widetilde{m}_1, \ldots, \widetilde{m}_p)'$ be patterns corresponding to the vectors b and $\widetilde{b}$, respectively. By Proposition 3, if $M = \widetilde{M}$, then $D_M = C_\Lambda \cap A_M = C_\Lambda \cap A_{\widetilde{M}} = D_{\widetilde{M}}$.
It remains to prove that if $D_M = D_{\widetilde{M}}$, then $M = \widetilde{M}$. Suppose that $D_M = D_{\widetilde{M}}$. Then, the patterns $(v_1 m_{\sigma(1)}, \ldots, v_p m_{\sigma(p)})'$ and $(v_1 \widetilde{m}_{\sigma(1)}, \ldots, v_p \widetilde{m}_{\sigma(p)})'$ both have the same subdifferential for any permutation σ of $(1, 2, \ldots, p)$ and any sequence $(v_1, \ldots, v_p) \in \{-1, 1\}^p$. Indeed, if $\pi^* = (\pi_1^*, \ldots, \pi_p^*)' \in C_\Lambda$ maximizes $(m_1, \ldots, m_p)'(\pi_1, \ldots, \pi_p)$, then $(v_1 \pi_{\sigma(1)}^* \ldots, v_p \pi_{\sigma(p)}^*)$ maximizes $(v_1 m_{\sigma(1)} \ldots, v_p m_{\sigma(p)})'(v_1 \pi_{\sigma(1)} \ldots, v_p \pi_{\sigma(p)})$.
Hence we assume without loss of generality that $m_1 \geq m_2 \geq \ldots \geq m_p \geq 0$, which implies $\Lambda = (\lambda_1, \ldots, \lambda_p)' \in D_M$. Since $D_M = D_{\widetilde{M}}$, it follows from Corollary 20 that $\widetilde{m}_1 \geq \ldots \geq \widetilde{m}_p \geq 0$, because $\Lambda \in D_{\widetilde{M}}$. The condition $D_M = D_{\widetilde{M}}$ also implies that $\|M\|_\infty = \|\widetilde{M}\|_\infty$ and that the kth cluster of M has the same size as the kth cluster of $\widetilde{M}$, $k = 1, \ldots, \|M\|_\infty$. To see this, suppose that there is i such that $m_{i+1} = m_i$ and $\widetilde{m}_{i+1} < \widetilde{m}_i$. Then, $\Lambda^{(i)} = (\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \lambda_i, \lambda_{i+2}, \ldots, \lambda_p)' \in D_M$ and $\Lambda^{(i)} \notin D_{\widetilde{M}}$, which is in contradiction to the fact $D_M = D_{\widetilde{M}}$ (the same argument applies to the case $m_{i+1} < m_i$ and $\widetilde{m}_{i+1} = \widetilde{m}_i$). To complete the proof, suppose that $m_p \neq \widetilde{m}_p$, say $m_p = 0$ and $\widetilde{m}_p = 1$. Then, $(\lambda_1, \lambda_2, \ldots, \lambda_{p-1}, 0)'$ is in D_M but not in $D_{\widetilde{M}}$, which is impossible. ∎

A.2. Proofs from Section 5.

PROOF OF THEOREM 13. By Proposition 1, $\hat{\beta}_n^{\mathrm{SLOPE}}$ is the SLOPE estimator of β in a linear regression model (5.1) if and only if

$$\pi_n^* = X_n'(Y_n - X_n\hat{\beta}_n^{\mathrm{SLOPE}}) \in C_{\Lambda_n} \tag{A.3}$$

and

$$U_{M_n}'\pi_n^* = \tilde{\Lambda}_n, \tag{A.4}$$

where $M_n = \mathrm{patt}(\hat{\beta}_n^{\mathrm{SLOPE}})$ and $\tilde{\Lambda}_n = U_{|M_n|\downarrow}'\Lambda_n$. By the definition of π_n^* we have

$$\hat{\beta}_n^{\mathrm{SLOPE}} = (X_n'X_n)^{-1}X_n'Y_n - (X_n'X_n)^{-1}\pi^* = \hat{\beta}_n^{\mathrm{OLS}} - \left(\frac{1}{n}X_n'X_n\right)^{-1}\left(\frac{1}{n}\pi_n^*\right).$$

Since in our setting $\hat{\beta}_n^{\mathrm{OLS}}$ is strongly consistent, $\hat{\beta}_n^{SLOPE} \xrightarrow{a.s.} \beta$ if and only if $(n^{-1}X_n'X_n)^{-1}\left(n^{-1}\pi_n^*\right) \xrightarrow{a.s.} 0$. In view of (5.5), we have $(n^{-1}X_n'X_n)^{-1}\left(n^{-1}\pi_n^*\right) \xrightarrow{a.s.} 0$ if and only if $n^{-1}\pi_n^* \xrightarrow{a.s.} 0$.
Assume $n^{-1}\lambda_1^{(n)} \to 0$. By (A.3) we have $\pi_n^* \in C_{\Lambda_n}$ and thus $\|\pi_n^*\|_\infty \le \lambda_1^{(n)}$, which gives

$$\left\|\frac{\pi_n^*}{n}\right\|_\infty \le \frac{\lambda_1^{(n)}}{n} \to 0$$

Therefore, (5.6) implies that $\hat{\beta}_n^{SLOPE} \xrightarrow{a.s.} \beta$.
Now assume that $\beta \neq 0$ and $\hat{\beta}^{\mathrm{SLOPE}}$ is strongly consistent, i.e. $n^{-1}\pi_n^* \xrightarrow{a.s.} 0$. Then, (A.4) gives

$$p\|\pi_n^*\|_\infty \ge \|U_{M_n}'\pi_n^*\|_\infty = \|\tilde{\Lambda}_n\|_\infty \ge \lambda_1^{(n)} \tag{A.5}$$

provided $M_n \neq 0$. Applying (A.3) for $\hat{\beta}_n^{SLOPE} = 0$, we note that $M_n(\omega) = 0$ if and only if

$$n^{-1}X_n(\omega)'Y_n(\omega) \in C_{n^{-1}\Lambda_n}.$$

It can be easily verified that $n^{-1}X_n'Y_n \xrightarrow{a.s.} C\beta$.
Since

$$\left\|\frac{1}{n}\pi_n^*\right\|_\infty \quad \ge \quad \left\|\frac{1}{n}\pi_n^*\right\|_\infty \mathbf{1}_{(M_n=0)} \quad = \quad \left\|\frac{1}{n}X_n'Y_n\right\|_\infty \mathbf{1}_{(M_n=0)},$$

we see that for $\beta \neq 0$, we have $M_n \neq 0$ for large n almost surely. Thus, for $\beta \neq 0$ we eventually obtain for large n

$$\frac{\lambda_1^{(n)}}{n} \le p\left\|\frac{\pi_n^*}{n}\right\|_\infty \quad \text{a.s.}$$

□

LEMMA 21. *Assume that $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of random variables, defined on the same probability space, which converges in distribution to $\mathrm{N}(0,\sigma^2)$ for some $\sigma \in (0,\infty)$.*

(i) *For any real function with $f(n) \to \infty$,*

$$\frac{\xi_n}{f(n)\sqrt{\log n}} \xrightarrow{a.s.} 0.$$

(ii) *If ξ_n are independent, then the sequence $(\frac{\xi_n}{\sqrt{\log n}})_n$ does not converge a.s. to 0.*

PROOF. Convergence in distribution to a continuous law implies the uniform convergence of cumulative distribution functions. Thus, for any $c > 1$ and N large enough we have

$$\sum_{n=N}^{\infty} \mathbb{P}\left(\frac{|\xi_n|}{f(n)\sqrt{\log n}} > \varepsilon\right) = \sum_{n=N}^{\infty}\left(F_{\xi_n}\left(-\varepsilon f(n)\sqrt{\log n}\right) + 1 - F_{\xi_n}\left(\varepsilon f(n)\sqrt{\log n}\right)\right)$$

$$\leq 2c\sum_{n=N}^{\infty}\left(1 - \Phi\left(\varepsilon f(n)\sqrt{\log n}/\sigma\right)\right),$$

where Φ is the cumulative distribution function of $\mathrm{N}(0,1)$.

The assertion (i) quickly follows from the Borel-Cantelli Lemma and the well-known tail inequality: $1-\Phi(t) \leq t^{-1}e^{-t^2/2}/\sqrt{2\pi}$, $t > 0$. Indeed, we have

$$1 - \Phi\left(\varepsilon f(n)\sqrt{\log n}/\sigma\right) \leq \frac{1}{\sqrt{2\pi}}\frac{\sigma}{\varepsilon f(n)\sqrt{\log n}}\left(\frac{1}{n}\right)^{\frac{\varepsilon^2 f^2(n)}{2\sigma^2}} \leq \frac{c}{n^2}$$

for a positive constant c and n large enough.

For (ii) we observe that the events $A_n = \{\frac{|\xi_n|}{\sqrt{\log n}} > \varepsilon\}$, $n = 1, 2, \ldots$, are independent and moreover

$$\sum_{n=1}^{\infty}\left(1 - \Phi\left(\varepsilon\sqrt{\log n}\right)\right) = \infty$$

for any $\varepsilon > 0$. Thus, the result follows again from the uniform convergence of cumulative distribution function. □

LEMMA 22. *Assume that $\varepsilon_n \sim \mathrm{N}(0, \sigma^2 I_n)$ and that the sequence $(Q_n)_n$ of random matrices $Q_n \in \mathbb{R}^{n\times p}$ satisfy*

$$\frac{1}{n}Q_n'Q_n \xrightarrow{a.s.} C_0, \tag{A.6}$$

where C_0 is a deterministic $p \times p$ matrix. If ε_n and Q_n are independent for each n, then the sequence $(\frac{1}{\sqrt{n}}Q_n'\varepsilon_n)_n$ of random vectors converges in distribution to $\mathrm{N}(0, \sigma^2 C_0)$.

PROOF. By the independence of Q_n and ε_n, we have for $t \in \mathbb{R}^p$,

$$\mathbb{E}\left[\exp\left(it'\frac{1}{\sqrt{n}}Q_n'\varepsilon_n\right)\right] = \mathbb{E}\left[\exp\left(-\frac{1}{2}t'\left(\sigma^2\frac{1}{n}Q_n'Q_n\right)t\right)\right] \to \exp\left(-\frac{1}{2}t'\sigma^2 C_0 t\right),$$

where the convergence follows from the Lebesgue dominated convergence theorem. Indeed, since $Q_n'Q_n$ is nonnegative definite, we have $-t'Q_n'Q_n t \leq 0$ for any $t \in \mathbb{R}^p$. □

PROOF OF (5.7). First we rewrite s_n as

$$s_n = (\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'Y_n - \alpha_n(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{\Lambda}.$$

Since $\beta = U_M s_0$, we conclude $X_n\beta = X_n U_M s_0 = \tilde{X}_n s_0$, so the linear regression model (5.1) takes the form $Y_n = \tilde{X}_n s_0 + \varepsilon_n$. Thus, $(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'Y_n$ is the OLS estimator of s_0. Note that, with $Q_n' = n(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n' \in \mathbb{R}^{n\times k}$, we have

$$\frac{1}{n}Q_n'Q_n = n(\tilde{X}_n'\tilde{X}_n)^{-1} \xrightarrow{a.s.} (U_M'CU_M)^{-1} =: C_0.$$

A simple argument using Lemma 22 and Lemma 21 (i) shows that

$$(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'Y_n = s_0 + \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}Q_n'\varepsilon_n\right) \xrightarrow{a.s.} s_0.$$

To complete the proof, we note that

$$\alpha_n (\tilde{X}_n' \tilde{X}_n)^{-1} \tilde{\Lambda} = \frac{\alpha_n}{n} \left[n (\tilde{X}_n' \tilde{X}_n)^{-1} \tilde{\Lambda} \right] \xrightarrow{a.s.} 0 \left[(U_M' C U_M)^{-1} \tilde{\Lambda} \right] = 0.$$

□

Funding. T. Skalski was supported by a French Government Scholarship.

REFERENCES

[1] ABRAMOVICH, F. and GRINSHTEIN, V. (2019). High-dimensional classification by sparse logistic regression. *IEEE Trans. Inform. Theory* **65** 3068–3079. MR3951383

[2] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2018). Slope meets Lasso: improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663

[3] BEN-ISRAEL, A. and GREVILLE, T. N. E. (2003). *Generalized inverses*, second ed. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* **15**. Springer-Verlag, New York Theory and applications. MR1987382

[4] BOGDAN, M., VAN DEN BERG, E., SU, W. and CANDÈS, E. J. (2013). Statistical estimation and testing via the sorted L1 norm. *arXiv preprint arXiv:1310.1969*.

[5] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717

[6] BONDELL, H. D. and REICH, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64** 115–123, 322–323. MR2422825

[7] BRZYSKI, D., PETERSON, C. B., SOBCZYK, P., CANDÈS, E. J., BOGDAN, M. and SABATTI, C. (2017). Controlling the rate of GWAS false discoveries. *Genetics* **205** 61–75.

[8] BRZYSKI, D., GOSSMANN, A., SU, W. and BOGDAN, M. (2019). Group SLOPE—adaptive selection of groups of predictors. *J. Amer. Statist. Assoc.* **114** 419–433. MR3941265

[9] BU, Z., KLUSOWSKI, J. M., RUSH, C. and SU, W. J. (2021). Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. *IEEE Trans. Inform. Theory* **67** 506–537. MR4231969

[10] CHEN, S. and DONOHO, D. (1994). Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers* **1** 41–44. IEEE.

[11] FIGUEIREDO, M. and NOWAK, R. (2016). Ordered Weighted L1 Regularized Regression with Strongly Correlated Covariates: Theoretical Aspects. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* (A. GRETTON and C. C. ROBERT, eds.). *Proceedings of Machine Learning Research* **51** 930–938. PMLR, Cadiz, Spain.

[12] HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993). *Convex analysis and minimization algorithms. I. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **305**. Springer-Verlag, Berlin Fundamentals. MR1261420

[13] KOS, M. and BOGDAN, M. (2020). On the asymptotic properties of SLOPE. *Sankhya A* **82** 499–532. MR4140775

[14] KREMER, P. J., BRZYSKI, D., BOGDAN, M. and PATERLINI, S. (2022). Sparse index clones via the sorted ℓ_1-Norm. *Quant. Finance* **22** 349–366. MR4390846

[15] SCHNEIDER, U. and TARDIVEL, P. (2020). The Geometry of Uniqueness, Sparsity and Clustering in Penalized Estimation. *arXiv preprint arXiv:2004.09106*.

[16] SKALSKI, T., GRACZYK, P., KOŁODZIEJEK, B. and WILCZYŃSKI, M. (2022). Pattern recovery and signal denoising by SLOPE when the design matrix is orthogonal. *arXiv preprint arXiv:2202.08573*.

[17] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953

[18] TARDIVEL, P. J. C. and BOGDAN, M. On the sign recovery by least absolute shrinkage and selection operator, thresholded least absolute shrinkage and selection operator, and thresholded basis pursuit denoising. *Scandinavian Journal of Statistics* 1-33.

[19] TARDIVEL, P., SKALSKI, T., GRACZYK, P. and SCHNEIDER, U. (2021). The Geometry of Model Recovery by Penalized and Thresholded Estimators. *HAL preprint hal-03262087*.

[20] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

[21] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. MR3066375

[22] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873

[23] ZENG, X. and FIGUEIREDO, M. A. (2014). Decreasing Weighted Sorted l_1 Regularization. *IEEE Signal Processing Letters* **21** 1240–1244.

[24] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449