arXiv: 1906.12072v2

Multiple Testing and Variable Selection along Least Angle Regression's path

Jean-Marc Azaïs• **and Yohann De Castro**⋆

•*Institut de Mathématiques de Toulouse*
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France

⋆*Institut Camille Jordan & École Centrale de Lyon, France*

Abstract: In this article we investigate Least Angle Regression (LAR) algorithm in high dimensions under the Gaussian noise assumption. For the first time, we give the exact joint law of sequence of knots conditional on the sequence of variables entering the model. Numerical experiments are provided to demonstrate the perfect fit of our finding. Based on this result, we prove an exact control of the existence of false negatives in the general design case and an exact control of the False Discovery Rate (FDR) in the orthogonal design case.

Our contribution is two fold. First, we build testing procedures on variables entering the model along the LAR's path and we introduce a new exact testing procedure on the existence of false negatives in the general design case when the noise level can be unknown. This testing procedures are referred to as the Generalized t-Spacing Test (GtST). Second, we give an exact control of the FDR in the orthogonal design case. Monte-Carlo simulations and a real data experiment are provided to illustrate our results.

MSC 2010 subject classifications: Primary 62E15, 62F03, 60G15, 62H10, 62H15; secondary 60E05, 60G10, 62J05, 94A08.

Keywords and phrases: Multiple Testing, False Discovery Rate, High-Dimensional Statistics, Post-Selection Inference.

Preprint version of May 22, 2020

1. Introduction

Parsimonious models have become ubiquitous tools to tackle high-dimensional representations with a small budget of observations. Successful applications may be found in signal processing (see for instance the pioneering works [13, 12] and references therein), biology (see for instance [3] or [11, Chapter 1.4] and references therein). These applications have shown that there exists interesting *almost sparse representations* in some well chosen basis. Nowadays, in many practical situations, this sparsity assumption is reasonable.

These important successes have put focus on High-Dimensional Statistics in the past decades and they may be due to the deployment of tractable algorithms with strong theoretical guarantees. Among the large panoply of methods, one may have seen emerged ℓ_1-regularization which may have found a fine balance between tractability and performances. Nowadays, sparse regression techniques based on ℓ_1-regularization are a common and powerful tool in high-dimensional settings. Popular estimators, among which one may point LASSO [27] or SLOPE [10], are known to achieve minimax rate of prediction and to satisfy sharp oracle inequalities under conditions on the design such as Restricted Eigenvalue [7, 4] or Compatibility [11, 30].

Recent avances have focused on a deeper understanding of these techniques looking at confidence intervals and testing procedures (see [30, Chapter 6] and references therein) or false discovery rate control (*e.g.*, [3]) for instance. These new results aim at describing the (asymptotic or non-asymptotic) law of the outcomes of ℓ_1-minimization regression. This line of works adresses important issues encountered in practice. Assessing the uncertainty of popular estimators give strong guarantees on the estimate produced, *e.g.*, the false discovery rate is controlled or a confidence interval on linear statistics of the estimator can be given.

Algorithm 1: LAR algorithm ("recursive" formulation)

Data: Correlations vector $\overline{Z}$ and variance-covariance matrix $\overline{R}$.

`/* Given a response variable` Y `and a design` X`, we have` $\overline{Z} = X^\top Y$ `and` $\overline{R} = X^\top X$ `*/`

Result: Sequence $((\lambda_k, \bar{\imath}_k, \varepsilon_k))_{k\geq 1}$ where $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ are the knots, and $\bar{\imath}_1, \bar{\imath}_2, \ldots$ are the variables that enter the model with signs $\varepsilon_1, \varepsilon_2, \ldots$ ($\varepsilon_k = \pm 1$).

`/* Define the recursive function Rec() that would be applied repeatedly. The inputs of Rec() are` Z `a vector,` R `a SDP matrix and` T `a vector. */`

Function `Rec`(R, Z, T):

Compute

$$\lambda = \max_{\{j\colon T_j<1\}} \left\{ \frac{Z_j}{1-T_j} \right\} \text{ and } \mathbf{i} = \arg \max_{\{j\colon T_j<1\}} \left\{ \frac{Z_j}{1-T_j} \right\}.$$

`/* The following recursions are given by (55), (56) and (57). */`

Update

$$\begin{aligned}
\mathbf{x} &= R_{\mathbf{i}}/R_{\mathbf{ii}} \\
R &\leftarrow R - \mathbf{x}R_{\mathbf{i}}^\top \\
Z &\leftarrow Z - \mathbf{x}Z_{\mathbf{i}} \\
T &\leftarrow T + \mathbf{x}(1-T_{\mathbf{i}})
\end{aligned}$$

return (R, Z, T, λ, $\mathbf{i}$)

1 Set $k = 0$, $T = 0$, $Z = (\overline{Z}, -\overline{Z})$ and $R = \begin{bmatrix} \overline{R} & -\overline{R} \\ -\overline{R} & \overline{R} \end{bmatrix}$.

`/* Use the following recursion function to compute the LAR path. */`

2 Update $k \leftarrow k+1$ and compute

$$(R, Z, T, \lambda_k, \widehat{\imath}_k) = \texttt{Rec}(R,\ Z,\ T)$$

Set $\bar{\imath}_k = \widehat{\imath}_k \mod p$ and $\varepsilon_k = 1 - 2(\widehat{\imath}_k - \bar{\imath}_k)/p \in \pm 1$.

`/*` $\widehat{\imath}_k$ `is referred to as the "signed" variable, see (3) */`

1.1. *Least Angle Regression algorithm, Support Selection and FDR*

Least Angle Regression (LAR) algorithm has been introduced in the seminal article [14]. This forward procedure produces a sequence of knots $\lambda_1, \lambda_2, \ldots$ based on a control of the residuals in ℓ_∞-norm. This sequence of knots is closely related to the sequence of knots of LASSO [27], as they differ by only one rule: "*Only in the LASSO case, if a nonzero coefficient crosses zero before the next variable enters, drop it from the active set and recompute the current joint least-squares direction*", as mentioned in [28, Page 120] or [14, Theorem 1] for instance. We focus on the LAR algorithm and:

- We present three equivalent formulations of *LAR algorithm*, see Section A.1;
- As far as we know, Algorithm 1, based on a *recursive* function, is new.

One specific task is to estimate the support of the target sparse vector, namely identify *true positives* in a context of multiple testing procedure. In particular, one may take the support of LASSO (or SLOPE) solution as an estimate of the support solution. This strategy has been intensively studied in the literature, one may consider [32, 10, 30, 4] and references therein. Support selection has been studied under the so-called "*Irrepresentable Condition*" (IC), as presented for instance in the books [30, Page 53] and [11, Sec. 7.5.1] and also referred to as the "*Mutual Incoherence Condition*" [32]. Under the so-called "*Beta-Min Condition*", one may prove [11, 30] that the LASSO asymptotically returns the true support. In this article, we investigate the existence of false non-negatives and we present exact non-asymptotic testing procedures, see Section 3.5.

Another recent issue is controling *False Discovery Rate* (FDR) in high-dimensional setting, as for instance in [3] and references therein; or the *Joint family-wise Error Rate* as in [8] and references therein. In this paper, we investigate the consecutive *spacings* of knots of the LAR as testing statistics and we prove an exact FDR control using a Benjamini–Hochberg procedure [5] in the orthogonal design case, see Section 3.6. Our proof (see Appendix C.7) is based on the *Weak*

Positive Regression Dependency (WPRDS), the reader may consult [9] or the survey [23], and *Knothe-Rosemblatt transport*, see for instance [24, Sec.2.3, P.67] or [31, P.20], which is based on conditional quantile transforms.

1.2. Post-Selection Inference in High-Dimensions with LAR

Consider the simplest linear model where one observes the target $\beta^0 \in \mathbb{R}^p$, namely there is no noise and the design is the identity. In this case, LASSO and LAR give the same knots $\lambda_1, \lambda_2, \ldots$ and LASSO reduces the proximal operator of the ℓ_1-norm at point $\beta^0 \in \mathbb{R}^p$, see for instance [28, Chapter 2]. In this simple case, we deduce that the knots are such that

$$\lambda_k = \beta^0_{(k)}\,, \tag{1}$$

where we have considered the reordering $\beta^0_{(1)} \geq \beta^0_{(2)} \geq \ldots$ on the entries of the target. Obviously, this fact is no longer true for general designs in high-dimensions with noise but one may ask

[Q1] *What is the* **joint law** *of the LAR's knots $\lambda_1, \lambda_2, \ldots$ and how do they relate to the target β^0?*

We will answer this question in high-dimensions under Gaussian noise assumption in Section 3.1.3 and Section 3.2. Under the so-called "*Irrepresentable Check*" Condition ($\mathcal{A}_{\mathrm{Irr.}}$), the simple fact (1) can be extended to general design in high-dimensions as we will show that the joint distribution of the LAR knots is a *mixture of Gaussian order statistics*, see Theorem 5.

Next, this paper introduces a class of exact tests built from ℓ_1-minimization regression in high-dimensions. Following the original idea of [20], we study tests based on the knots of the LAR's path. Note that, conditional on the sequence of indexes selected by LAR, the law of three **consecutive** knots has been studied by [20] referred to as the *Spacing test* (ST) [29], and

[Q2] *Can we provide exact testing procedures based on knots that are* **not** *consecutive?*

A positive answer is given in Section 3.2 and we referred to these new tests are *Generalized Spacing Tests* (GST), see Theorem 8 and Remark 11; and, when the noise variance is unknown, *Generalized t-Spacing Tests* (GtST), see Theorem 14 and Remark 13

The article [1] proved that *Spacing test* is unbiased and introduce a *Studentized* version of this test. In the same direction, inference after model selection has been studied in several papers, as [15, 25] and respectively [26] for *selective inference* and respectively a joint estimate of the noise level. It raises the following questions.

[Q3] *What is the power of the Generalized Spacing Tests?*

[Q4] *Can we provide* **exact** *testing procedures when the noise level is not known, and how do we estimate the noise level without bias in this case?*

We will answer these questions in Section 3.3 and respectively Section 3.5.

The above line of works studies a single test on a linear statistics while one may ask

[Q5] *Can we provide an exact control of multiple Spacing Tests?*

[Q6] *Can we provide an exact false negative testing procedure after model selection?*

To the best of our knowledge, this paper is the first to study the joint law and an exact control of multiple spacing tests of LAR's knots in a non-asymptotic frame, see Sections 3.2 and 3.6. The control of the false negatives after model selection is given in Section 3.4 and the procedure is introduced in Algorithm 2.

One may point others approaches for building confidence intervals and/or testing procedures in high-dimensional settings as follows. Simultaneous controls of confidence intervals independently of the selection procedure have been studied under the concept of *post-selection constants* as introduced in [6] and studied for instance in [2]. Asymptotic confidence intervals can be build using the *de-sparsified LASSO*, the reader may refer to [30, Chapter 5] and references therein. We also point a recent study [19] of the problem of FDR control as the sample size tends to infinity using *de-biased LASSO*.

1.3. Outline of the paper

Section 2 introduces notation (see a summary in Appendix E), assumptions and variance estimates. The main assumption is based on the condition of "*Irrepresentable Check*" ($\mathcal{A}_{\text{Irr.}}$) that can be checked in practice, see Section 2.2. The variance estimates are a key step in our testing procedures: we introduce new variance estimates with useful property to derive exact and non-asymptotic post-selection laws, see Section 2.3.

Section 3 gives the main results: Section 3.2 describes the joint distribution of the LAR knots as a *mixture of Gaussian order statistics* and the main test, the orthogonal case being considered in Section 3.3. Control of false negative in a post selection inference with estimation of the variance is presented in Section 3.5. A procedure of control of the false discovery rate in the orthogonal case is presented in Section 3.6.

An illustration of our method, both on simulated data and on real data, is presented in Section 4.

2. Assumptions, Variance Estimations and Admissible Procedures

2.1. Notation and Least Angle Regression (LAR)

2.1.1. Linear model in high-dimensions

Consider the linear model in high-dimensions where the number of observations n may be less than the number of predictors p. We denote by $Y \in \mathbb{R}^n$ the response variable and we assume that

$$Y = X\beta^0 + \eta \sim \mathcal{N}_n(X\beta^0, \sigma^2 \mathrm{Id}_n)\,, \tag{2}$$

where $\eta \sim \mathcal{N}_n(0, \sigma^2 \mathrm{Id}_n)$ is some Gaussian noise, the noise level $\sigma > 0$ may be known or that has to be estimated depending on the context, and $X \in \mathbb{R}^{n\times p}$ has rank r. Consider the Least Angle Regression (LAR) algorithm where we denote by $(\lambda_k)_{k\geq 1}$ the sequence of knots and by $(\bar{\imath}_k, \varepsilon_k)_{k\geq 1}$ the sequence of variables $\bar{\imath}_k \in [p]$ and signs $\varepsilon_k \in \{\pm 1\}$ that enter the model along the LAR path, see for instance [28, Chapter 5.6] or [14] for standard description of this algorithm. We recall this algorithm in Algorithm 3 and we present equivalent formulations in Algorithm 4 (using orthogonal projections) and Algorithm 1 (using a recursion). The interested reader may find their analysis in Appendices A and B. In particular, we present here Algorithm 1 that consists in three lines, applying the same function recursively.

2.1.2. Signed variables of LAR

We give some notation that we will be useful. We denote by $(\widehat{\imath}_1, \ldots, \widehat{\imath}_k) \in [2p]^k$ the "*signed*" variables that enter the model along the LAR path with the convention that

$$\widehat{\imath}_k := \bar{\imath}_k + p\left(\frac{1-\varepsilon_k}{2}\right), \tag{3}$$

so that $\widehat{\imath}_k \in [2p]$ is a useful way of encoding both the variable $\bar{\imath}_k \in [p]$ and its sign $\varepsilon_k = \pm 1$ as used in Algorithm 1. We denote by $\overline{Z} := X^\top Y$ the vector such that $\overline{Z}_k$ is the scalar product between the k-th predictor and the response variable, and we denote by $\sigma^2\overline{R}$ its *variance-covariance* matrix. For sake of presentation, we may consider the $2p$-vector

$$Z := (\overline{Z}, -\overline{Z}) = (X^\top Y, -X^\top Y)\,, \tag{4}$$

whose mean is given by

$$\mu^0 := (\overline{R}\beta^0, -\overline{R}\beta^0) = (X^\top X\beta^0, -X^\top X\beta^0) = (\overline{\mu}^0, -\overline{\mu}^0)\,, \tag{5}$$

and its variance-covariance matrix is $\sigma^2 R$ with

$$R = \begin{bmatrix} \overline{R} & -\overline{R} \\ -\overline{R} & \overline{R} \end{bmatrix} = \begin{bmatrix} X^\top X & -X^\top X \\ -X^\top X & X^\top X \end{bmatrix}. \tag{6}$$

We also denote by

- $\widehat{\imath}_1, \ldots, \widehat{\imath}_k$, the first k signed variables entering the LAR,
- $i_1, \ldots, i_k$, a generic value of the sequence above,

- $\bar{\imath}_1, \ldots, \bar{\imath}_k$, the first k variables entering the LAR,
- $j_1, \ldots, j_k$, a generic value of the sequence above,

- $\varepsilon_1, \ldots, \varepsilon_k$, the first k signs of the coefficients of the variables entering in the LAR,
- $s_1, \ldots, s_k$, a generic value of the sequence above.

The quantities above are related by (3) and

$$i_k := j_k + p\left(\frac{1-s_k}{2}\right). \tag{7}$$

2.1.3. Models and the K notation

We are interested in selecting the true support S^0 of β^0, where the support is defined by

$$S^0 := \left\{k \in [p] \;:\; \beta^0_k \neq 0\right\}.$$

To estimate this support, we will consider the models that appear along the LAR's path: the selected model $\widehat{S}$ would be chosen among the family of nested models

$$\underbrace{\{\bar{\imath}_1\}}_{\overline{S}^1} \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2\}}_{\overline{S}^2} \subset \cdots \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2, \ldots, \bar{\imath}_k\}}_{\overline{S}^k} \subset \cdots \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2, \ldots, \bar{\imath}_K\}}_{\overline{S}^K}, \tag{8}$$

where K denotes the maximal model size. Respectively, denote

$$\{0\} =: H_0 \subset H_1 \subset \cdots \subset H_k := \mathrm{Span}(X_{\bar{\imath}_1}, \ldots, X_{\bar{\imath}_k}) \subset \cdots \subset H_K, \tag{9}$$

the corresponding family of nested subspaces of $\mathbb{R}^n$. Once the model has been selected, we will construct tests based on the $K+1$ first knots of the LAR.

Throughout this paper, we assume that

$$K \text{ is fixed and such that } 1 \leq K < \min(n-3, r) \text{ where } r = \mathrm{rank}(X). \tag{10}$$

In practice, K can be considerably much smaller than n. Our analysis is conditional on $(\widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1})$ and in this spirit it can be referred to as a "*Post-Section*" procedure, see *e.g.* [25, 29, 28].

2.2. Irrepresentable Check on the Active sets

We define the set of *Active Sets* $\mathcal{A}_K$ as all the sequences $i_1, \ldots, i_K$ of signed variables such that $j_1, \ldots, j_K$ are pairwise different, where the j's are defined by (7), namely

$$\mathcal{A}_K := \left\{(i_1, \ldots, i_K) \in [2p]^K \;:\; j_1, \ldots, j_K \text{ are pairwise different}\right\}.$$

Sometimes it would be useful to consider $\mathcal{A}_{K+1}$, the set of active sets of size $K+1$. We introduce *Irrepresentable Check* which is the only assumption on the design and the selected active set in most of our results.

Definition 1 (Irrepresentable Check)**.** *An active set $(i_1, \ldots, i_K) \in \mathcal{A}_K$ said to satisfy the Irrepresentable Check condition if*

$$\forall k \in [K]\,,\ \forall j \notin T^k := \{j_1, \ldots, j_k\}, \quad X_j^\top X_{T^k}\left(X_{T^k}^\top X_{T^k}\right)^{-1} s_k < 1\,, \tag{$\mathcal{A}_{\mathrm{Irr.}}$}$$

where j_k and s_k are defined from i_k using (7). By a small abuse of notation we will denote by ($\mathcal{A}_{\mathrm{Irr.}}$) the set of sequences $(i_1, \ldots, i_K)$ that satisfy this property.

In our procedures and theoretical results, we will limit our attention to sequences of chosen variables $\widehat{\imath}_1, \ldots, \widehat{\imath}_K$ by LAR that satisfy ($\mathcal{A}_{\mathrm{Irr.}}$). A particular case is when the property is true for all possible active sets. This is equivalent to the *Irrepresentable Condition* that we recall here.

Definition 2 (Irrepresentable Condition of order K)**.** *The design matrix X satisfies the Irrepresentable Condition of order K if and only if*

$$\forall S \subset [p] \text{ s.t. } \#S \leq K, \quad \max_{j \in [p]\setminus S} \max_{||v||_\infty \leq 1} X_j^\top X_S \big(X_S^\top X_S\big)^{-1} v < 1\,, \tag{Irrep.}$$

where X_j denotes the j^{th} column of X and X_S the sub-matrix of X obtained by keeping the columns indexed by S.

Remark 1. *Note that the "Irrepresentable Condition" is a standard condition, as presented for instance in the books [30, Page 53] and [11, Sec. 7.5.1] and also referred to as the "Mutual Incoherence Condition" [32].*

Remark 2. *This condition has been intensively studied in the literature and it is now well established that some random matrix models satisfies it with high probability. For instance, one may refer to the article [32] where it is shown that a design matrix $X \in \mathbb{R}^{n\times p}$ whose rows are drawn independently with respect to a centered Gaussian distribution with variance-covariance matrix satisfying* (Irrep.) *(for instance the Identity matrix) satisfies* (Irrep.) *with high probability when $n \gtrsim K \log(p - K)$, where $\gtrsim$ denotes an inequality up to some multiplicative constant.*

In practice, the Irrepresentable Condition (Irrep.) is a strong requirement on the design X and, additionally, this condition cannot be checked in polynomial time. One important feature of our results is that we do not require *Irrepresentable Condition* (Irrep.) but the weaker requirement *Irrepresentable Check* ($\mathcal{A}_{\mathrm{Irr.}}$) on the selected active set. Namely, we would assume that

$$\text{For } K \text{ defined by (10)}, \ (\widehat{\imath}_1, \ldots, \widehat{\imath}_K) \text{ satifies } (\mathcal{A}_{\mathrm{Irr.}})\,. \tag{Assumption}$$

Given $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$, note that this condition can be checked in polynomial time.

Example 3. *Taking the* (*signed*) *variables entering the model with LAR on a iid Gaussian design and response variable a centered Gaussian vector with iid entries on* $10{,}000$ *Monte-Carlo repetitions, we draw the law of the maximal order $K_{\max}$ for which Irrepresentable Check holds in Figure 1. For example, we found that for $p = 1{,}000$ and $n = 100$* (*with ratio $n/p = 0.1$*) *resp. $n = 500$* (*with ratio $n/p = 0.5$*)*, Irrepresentable Check* ($\mathcal{A}_{\mathrm{Irr.}}$) *of order $K_{\max}$ holds when $K_{\max}$ is about $K_{\max} \simeq 0.16 \times n = 16$ resp. $K_{\max} \simeq 0.12 \times n = 60$, see Figure 1.*

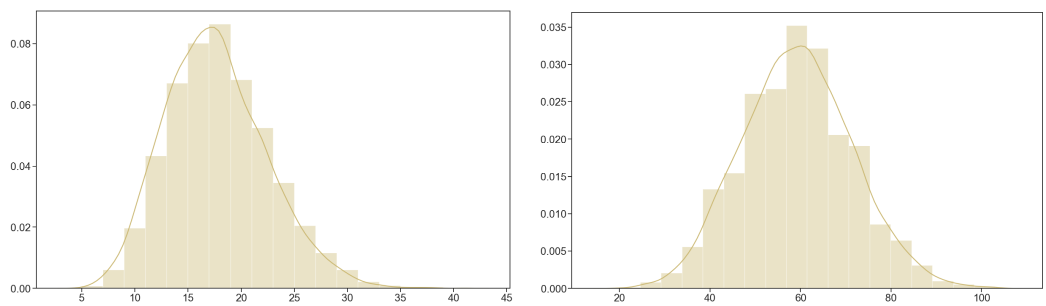

Figure 1: Taking the (*signed*) variables entering the model with LAR on a iid Gaussian design and response variable a centered Gaussian vector with iid entries on $10{,}000$ Monte-Carlo repetitions, we draw the law of the maximal order K for which *Irrepresentable Check* holds. In these experiments, we have $p = 1{,}000$ predictors and $n = 100$ (left) $n = 500$ (right) observations, and we find that $K_{\max} \in [10, 27]$ (left) and $K_{\max} \in [39, 81]$ (right) for 95% of the values.

Now, we can define

$$\forall (i_1, \ldots, i_k) \in [2p]^k, \quad \theta_j(i_1, \ldots, i_k) := \big(R_{j,i_1} \cdots R_{j,i_k}\big) M^{-1}_{i_1,\ldots,i_k}(1, \ldots, 1)\,, \tag{11}$$

where $(1, \ldots, 1)$ is the column vector of size k whose entries are equal to one; $\sigma^2 M_{i_1,\ldots,i_k}$ is the variance-covariance matrix of the vector $(Z_{i_1}, \cdots, Z_{i_k})$ and $(R_{j,i_1} \cdots R_{j,i_k})$ is a row vector of size k. Note that $M_{i_1,\ldots,i_k}$ is the submatrix of R obtained by keeping the columns and the rows indexed by $\{i_1, \ldots, i_k\}$, namely

$$M_{i_1,\ldots,i_k} := (R_{i,j})_{i,j=i_1,\ldots,i_k}\,. \tag{12}$$

Remark that

$$\theta_j(i_1, \ldots, i_k) = \mathbb{E}\big[Z_j \mid Z_{i_1} = 1, \ldots, Z_{i_k} = 1\big]\,,$$

when $\mathbb{E}Z = 0$. Then Proposition 1 shows that the *Irrepresentable Condition* (Irrep.) of order K is equivalently given by

$$\forall k \leq K,\ \forall (i_1, \ldots, i_k) \in [2p]^k,\ \forall j \notin \{i_1, \ldots, i_k\}\,, \quad \theta_j(i_1, \ldots, i_k) < 1\,, \tag{13}$$

where $\theta_j(i_1, \ldots, i_k)$ is given by (11).

Proposition 1. *Let X and R be defined by (6) then the following assumptions are equivalent:*

- *the design matrix X satisfies (Irrep.) of order K,*
- *the variance-covariance matrix R satisfies (13) of order K.*

Furthermore, they imply that for all $(i_1, \ldots, i_K) \in \mathcal{A}_K$ it holds

$$\max\Big[\max_{j \neq \hat{\imath}_1} \theta_j(\imath_1), \ \ldots\ , \max_{j \neq i_1,\ldots,i_K} \theta_j(i_1, \ldots, i_K)\Big] < 1$$

which is an equivalent formulation of $(i_1, \ldots, i_K)$ satisfying ($\mathcal{A}_{\text{Irr.}}$).

Proof. Let $S = \{j_1, \ldots, j_k\} \subset [p]$ and $j \in [2p] \setminus S$. Let $\overline{v} = (\overline{v}_1, \ldots, \overline{v}_k) \in \{-1, 1\}^k$ and define $i_\ell = j_\ell + p(1 - \overline{v}_\ell)/2$ for $\ell \in [k]$. Note that

$$\begin{aligned}
\theta_j(i_1, \ldots, i_k) &= \big(R_{j,i_1} \cdots R_{j,i_k}\big) M^{-1}_{i_1,\ldots,i_k}(1, \ldots, 1) \\
&= \Big[X_j^\top X_S \mathrm{Diag}(\overline{v})\Big] M^{-1}_{i_1,\ldots,i_k}(1, \ldots, 1) \\
&= \Big[X_j^\top X_S \mathrm{Diag}(\overline{v})\Big] M^{-1}_{i_1,\ldots,i_k}\Big[\mathrm{Diag}(\overline{v})\overline{v}\Big] \\
&= X_j^\top X_S \Big[\mathrm{Diag}(\overline{v}) M^{-1}_{i_1,\ldots,i_k} \mathrm{Diag}(\overline{v})\Big]\overline{v} \\
&= X_j^\top X_S \big(X_S^\top X_S\big)^{-1}\overline{v}\,.
\end{aligned}$$

Now, observe that

$$\max_{||v||_\infty \leq 1} v^\top \big(X_S^\top X_S\big)^{-1} X_S^\top X_j = \max_{\overline{v} \in \{-1,1\}^k} X_j^\top X_S \big(X_S^\top X_S\big)^{-1}\overline{v}\,,$$

showing the equivalence between the two assumptions. $\square$

Remark 4. *One may require that the design is "normalized" so that $R_{i,i} = 1$, namely its columns have unit Euclidean norm. Under this normalization, one can check that R satisfies (Irrep.) of order $K = 1$. Hence, up to some normalization, one can always assume (Irrep.) of order $K = 1$.*

Remark 5. *When computing the LAR path, one has to compute the values $X_j^\top X_{\overline{S}^k}\big(X_{\overline{S}^k}^\top X_{\overline{S}^k}\big)^{-1}\varepsilon^k$, see for instance Algorithm 3 or Algorithm 4, where these values are given by θ as shown by Proposition 1. It implies that, in practice, along the LAR's path, one witnesses the maximal order K for which Irrepresentable Check holds.*

2.3. Variance Estimates

In our analysis, we introduce two independent estimates of the variance, namely $\widehat{\sigma}^2_{\mathrm{select}}$ to select the model and $\widehat{\sigma}^2_{\mathrm{test}}$ to perform post-selection test. The *budget* devoted to estimation of the variance is $n-(K+1)$ and we can divide it equitably into n_1 and n_2, namely

$$n = n_1 + n_2 + (K+1) \quad \text{with} \quad |n_1 - n_2| \geq 1\,.$$

Let us fix, for the moment $j_1, \ldots, j_{K+1}$ the indexes that are putative indices for the selected variables. Let $P^{\perp}_{K+1}$ be the orthogonal projector on the orthogonal to $X_{j_1}, \ldots, X_{j_{K+1}}$. For ℓ positive define $V_\ell := \mathrm{Span}(e_1, \ldots, e_\ell)$ where e_i is the *ith* vector of the canonical basis. It is easy to check that we can find some ℓ such that

$$E_1 := P^{\perp}_{K+1}(V_\ell) \text{ has dimension } n_1.$$

We set also

$$E_2 := E_1^{\perp} \cap (\mathrm{Span}(X_{j_1}, \ldots, X_{j_{K+1}}))^{\perp} \text{ that has dimension } n_2,$$

and

$$\widehat{\sigma}^{j_1,\ldots,j_{K+1}}_{\mathrm{select}} := \frac{||P_{E_1}Y||_2}{\sqrt{n_1}} \quad , \quad \widehat{\sigma}^{j_1,\ldots,j_{K+1}}_{\mathrm{test}} := \frac{||P_{E_2}Y||_2}{\sqrt{n_2}}. \tag{14}$$

By a small abuse of notation, we can also index the estimator above by *signed* indexes $i_1, \ldots, i_{K+1}$. Eventually we set

$$\widehat{\sigma}_{\mathrm{select}} := \widehat{\sigma}^{\bar{\imath}_1,\ldots,\bar{\imath}_{K+1}}_{\mathrm{select}} \quad \text{and} \quad \widehat{\sigma}_{\mathrm{test}} := \widehat{\sigma}^{\bar{\imath}_1,\ldots,\bar{\imath}_{K+1}}_{\mathrm{test}}\,,$$

the estimates of the standard deviation σ.

2.4. Admissible Selection Procedures

Note that choosing a model $\widehat{S}$ is equivalent to choosing a model size $\widehat{m}$ so that

$$\widehat{S} = \{\bar{\imath}_1, \bar{\imath}_2, \ldots, \bar{\imath}_{\widehat{m}}\}\,. \tag{15}$$

Our procedure is flexible on this point and it allows any choice of $\widehat{m}$ as long as the following property ($\mathcal{A}_{\mathrm{Stop}}$) is satisfied

Stopping Rule: *The estimated model size $\widehat{m}$ is a "stopping time": $\widehat{m} \in [K-1]$ and*

$$\mathbb{1}_{\{\widehat{m} \leq a\}} \textit{ is a measurable function of } (\lambda_1, \ldots, \lambda_a, \widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1}, \widehat{\sigma}_{\mathrm{select}})\,, \tag{$\mathcal{A}_{\mathrm{Stop}}$}$$

for all $a \in [K-1]$.

In other words, the decision to select a model of size $\{\widehat{m} = a\}$ depends only on the a first variables entering LAR and on $\widehat{\sigma}_{\mathrm{select}}$. Of course, when σ^2 is known, $\widehat{\sigma}_{\mathrm{select}}$ can be removed from ($\mathcal{A}_{\mathrm{Stop}}$).

Remark 6. *Let us give an example to show that ($\mathcal{A}_{\mathrm{Stop}}$) implies some restriction. Suppose for example that we want to decide wether the target β^0 is two sparse or one sparse. A natural rule is to look at the second knot λ_2, if "$\lambda_2 >$ (some threshold)" chose $m = 2$ otherwise $m = 1$. This rule does not meet ($\mathcal{A}_{\mathrm{Stop}}$), since looking at λ_2 we can choose only sizes m greater or equal to 2.*

In the sequel, we will present some examples of such "*stopping time*" procedures, see Section 4.1.

Denote by $P_k(Y)$ (resp. $P_k^{\perp}(Y)$) the orthogonal projection of the observation Y onto H_k (resp. the orthogonal of H_k) for all $k \geq 1$ where H_k are defined by (9). A class of selection procedures satisfying ($\mathcal{A}_{\mathrm{Stop}}$) is given by

$$\mathbb{1}_{\{\widehat{m} \leq a\}} = h(P_a(Y), \widehat{\sigma}_{\mathrm{select}})\,,$$

where h is any measurable function. These procedures decide to stop at $\{\widehat{m} = a\}$ based on the information given by $P_a(Y)$.

Once one has selected a model of size $\widehat{m}$, one may be willing to test if $\widehat{S}$ contains the true support S^0 by considering the null hypothesis

$$\mathbb{H}_0 \,:\, \text{“}S^0 \subseteq \widehat{S}\text{”}\,,$$

namely there is no *false negatives*. Equivalently, one aim at testing the null hypothesis

$$\mathbb{H}_0 \,:\, \text{“}X\beta^0 \in H_{\widehat{m}}\text{”}\,, \tag{16}$$

at an exact significance level $\alpha \in (0,1)$, where $(H_a)_{a=0}^{K-1}$ is defined by (9).

Remark 7. *This hypothesis is not standard since $(H_a)_{a=1}^{K-1}$ are random subspaces. This hypothesis has to be understood in the framework of* selective testing*, namely the hypothesis is considered conditional on the selection event $\{\widehat{m} = a, \widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_K = i_K\}$, for some fixed $a \in [K-1]$. Conditional on the event, note that H_a is constant. By convention, we may consider the case $a = 0$, that is testing the* global nul hypothesis*.*

3. Exact Controls using Least Angle Regression: Main Results

3.1. Key notion: the “frozen” knots and means

We introduce some useful notation referred to as the *frozen* random variables. These random variables are defined for a *generic* sequence $i_1, \ldots, i_{K+1}$ of *signed* variables.

3.1.1. Frozen knots

An interesting feature of the LAR's knots is that they have a simple expression onto the partitioning given by the identity

$$\sum_{(i_1,\ldots,i_K)\in\mathcal{A}_K} \mathbb{1}_{\{\widehat{\imath}_1=i_1,\ldots,\widehat{\imath}_K=i_K\}} = 1 \quad \text{almost surely}\,.$$

For instance, as we will see in (21), it holds that

$$\forall k \in [K]\,,\quad \lambda_k = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_k} \mathbb{1}_{\{\widehat{\imath}_1=i_1,\ldots,\widehat{\imath}_k=i_k\}} \underbrace{Z_{i_k}^{i_1,\ldots i_{k-1}}}_{=:\lambda_k^f}\,,$$

giving the definition of the *frozen* knots λ_k^f below. In the same spirit, we will define the frozen means m_k^f by (23). These means are the building blocks of any hypotheses tested along the LAR as we will see in Proposition 2 and Theorem 5.

Given K as defined in (10) and fixed $i_1, \ldots, i_{K+1} \in [2p]$, one may define

$$\forall j \text{ s.t. } \theta_j(i_1,\ldots,i_k) \neq 1\,,\quad Z_j^{(i_1,\ldots,i_k)} := \frac{Z_j - \Pi_{i_1,\ldots,i_k}(Z_j)}{1-\theta_j(i_1,\ldots,i_k)}\,, \tag{17}$$

where

$$\Pi_{i_1,\ldots,i_k}(Z_j) := \big(R_{j,i_1}\cdots R_{j,i_k}\big)M^{-1}_{i_1,\ldots,i_k}(Z_{i_1},\ldots,Z_{i_k}) \tag{18}$$

and θ given by (11). When $\mathbb{E}Z = 0$, one may remark that $\Pi_{i_1,\ldots,i_k}(Z_j)$ is the *regression* of Z_j on the vector $(Z_{i_1},\cdots,Z_{i_k})$ whose variance-covariance matrix is $\sigma^2 M_{i_1,\ldots,i_k}$, namely

$$\text{When } \mathbb{E}Z = 0\,,\quad \Pi_{i_1,\ldots,i_k}(Z_j) = \big(R_{j,i_1}\cdots R_{j,i_k}\big)M^{-1}_{i_1,\ldots,i_k}(Z_{i_1},\ldots,Z_{i_k}) = \mathbb{E}\big[Z_j|Z_{i_1},\cdots,Z_{i_k}\big]\,.$$

From this point, we can introduce

$$\forall k \geq 0,\quad \lambda_{k+1}^{(i_1,\ldots,i_k)} := \max_{j:\theta_j(i_1,\ldots,i_k)<1} Z_j^{(i_1,\ldots,i_k)}\,, \tag{19}$$

and we emphasize that

$$\forall k \geq 0, \quad \lambda_{k+1}\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}} = \lambda_{k+1}^{(i_1,\dots,i_k)}\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}}\,, \tag{20}$$

as proven in Appendix A.3 (Eq. (52)) and Proposition 3. We are now able to define the "*frozen*" values of the knots by

$$\lambda_1^f := Z_{i_1}, \dots, \lambda_{K+1}^f := Z_{i_{K+1}}^{i_1,\dots i_K}\,. \tag{21}$$

They are the Gaussian variables that coincide with $\lambda_1, \lambda_2, \cdots, \lambda_{K+1}$ when the random indices $\widehat{\imath}_1, \widehat{\imath}_2, \dots, \widehat{\imath}_{K+1}$ take the particular values $i_1, i_2, \dots, i_{K+1}$.

3.1.2. Mean and centering of the frozen knots

Now, denote

$$\forall y \in \mathbb{R}^n\,, \quad P^{(i_1,\dots,i_k)}(y) = \big(X_{j_1}\cdots X_{j_k}\big) M^{-1}_{j_1,\dots,j_k}(X_{j_1}^\top, \dots, X_{j_k}^\top)\, y \tag{22}$$

the orthogonal projection onto $\mathrm{Span}(X_{j_1}, \dots, X_{j_k})$. Note that

$$\begin{aligned} P_k\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}} &= P^{(i_1,\dots,i_k)}\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}}\,, \\ P_k^\perp\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}} &= \big(\mathrm{Id}_n - P^{(i_1,\dots,i_k)}\big)\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}}\,, \\ \text{and for all } i \in [2p]\,, \quad \Pi_{i_1,\dots,i_k}(Z_i) &= s\langle X_j, P^{(i_1,\dots,i_k)}(Y)\rangle\,, \end{aligned}$$

where $i = j + p(1-s)/2$. The mean m_k^f and standard deviation $\sigma\,\rho_k^f$ of λ_k^f are important values defined for all $k \in [K]$,

$$\begin{aligned} m_k\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}} &= m_k^f\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}}\,, \\ \rho_k\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}} &= \rho_k^f\,\mathbb{1}_{\{\widehat{\imath}_1=i_1,\dots,\widehat{\imath}_k=i_k\}}\,, \end{aligned}$$

with

$$m_k^f = \frac{s_k\langle X_{j_k}, \big(\mathrm{Id}_n - P^{(i_1,\dots,i_k)}\big)X\beta^0\rangle}{1-\theta_{i_k}(i_1,\dots,i_{k-1})}\,, \tag{23}$$

$$\rho_k^f = \frac{\sqrt{\langle X_{j_k}, \big(\mathrm{Id}_n - P^{(i_1,\dots,i_k)}\big)X_{j_k}\rangle}}{1-\theta_{i_k}(i_1,\dots,i_{k-1})}\,, \tag{24}$$

and this definition is equivalent to (28) and (29), see Section 3.2. Recall that $\overline{\mu}^0 = \overline{R}\beta^0$ as defined in (5) and note that

$$m_k^f = 0 \quad \Leftrightarrow \quad \overline{\mu}^0_{i_k} - \Pi_{i_1,\dots,i_{k-1}}(\overline{\mu}^0_{i_k}) = 0 \quad \Leftrightarrow \quad \langle X_{j_k}, \big(\mathrm{Id}_n - P^{(i_1,\dots,i_k)}\big)X\beta^0\rangle = 0\,, \tag{25}$$

which is implied when *the true support S^0 of β^0 is included in $\overline{S}^k$*, defined by (8). It shows the next proposition.

Proposition 2. *For fixed $0 \leq a \leq K-1$, conditional on the selection event $\{\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_K = i_K\}$, the hypothesis*

$$\mathbb{H}_0\,:\ \text{“}X\beta^0 \in H_a\text{”}$$

implies that $m_k^f = 0$ for all $a < k \leq K$, namely $(Z_{i_{a+1}}^{(i_1,\dots,i_a)}, \dots, Z_{i_K}^{(i_1,\dots,i_{K-1})})$ is centered.

This proposition is important to define the hypothesis under consideration, see also Remark 7.

3.1.3. One key result: characterization of the selection event

Regarding the joint law of *frozen* knots, one has the following important proposition whose proof can be found in Section C.1.

Proposition 3. *Let $(i_1, \dots, i_K, i_{K+1}) \in \mathcal{A}_{K+1}$, namely a fixed active set of size $K+1$.*

- *It holds that*

$$(Z_j^{(i_1,\dots,i_{k+1})})_{j \neq i_1,\dots,i_{k+1}} \perp\!\!\!\perp Z_{i_{k+1}}^{(i_1,\dots,i_k)} \perp\!\!\!\perp Z_{i_k}^{(i_1,\dots,i_{k-1})} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Z_{i_2}^{(i_1)} \perp\!\!\!\perp Z_{i_1}$$

  are mutually independent, for any $k \in [K]$. And it holds that

$$\widehat{\sigma}_{\text{select}}^{i_1,\dots,i_{K+1}} \perp\!\!\!\perp \widehat{\sigma}_{\text{test}}^{i_1,\dots,i_{K+1}} \perp\!\!\!\perp Z_{i_{K+1}}^{(i_1,\dots,i_K)} \perp\!\!\!\perp Z_{i_K}^{(i_1,\dots,i_{K-1})} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Z_{i_2}^{(i_1)} \perp\!\!\!\perp Z_{i_1}\,. \tag{26}$$

- *If $(i_1, \dots, i_K)$ satisfies $(\mathcal{A}_{\text{Irr.}})$ then*

$$\begin{aligned}
&\big\{\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_{k+1} = i_{k+1}\big\} \\
&= \big\{\lambda_{k+1}^{(i_1,\dots,i_k)} = Z_{i_{k+1}}^{(i_1,\cdots,i_k)} \leq Z_{i_k}^{(i_1,\cdots,i_{k-1})} \leq \cdots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\big\} \\
&= \big\{\lambda_{k+1}^{(i_1,\dots,i_k)} = Z_{i_{k+1}}^{(i_1,\cdots,i_k)} \leq Z_{i_k}^{(i_1,\cdots,i_{k-1})} \leq \cdots \leq Z_{i_{a+1}}^{(i_1,\cdots,i_a)} \leq Z_{i_a}^{(i_1,\cdots,i_{a-1})} = \lambda_a^{(i_1,\dots,i_{a-1})}\big\} \\
&\quad \bigcap \big\{\lambda_{k+1} = Z_{i_{k+1}}^{(i_1,\cdots,i_k)}, \dots, \lambda_a = Z_{i_a}^{(i_1,\cdots,i_{a-1})}\big\} \\
&\quad \bigcap \big\{\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_a = i_a\big\}\,,
\end{aligned}$$

  for any $0 \leq a < k \leq K$ with the convention $\lambda_0 = \infty$.

From Proposition 3 we know that *frozen* values of the knots $Z_{i_1}, Z_{i_2}^{(i_1)}, \cdots, Z_{i_K}^{(i_1,\dots,i_{K-1})}, Z_{i_{K+1}}^{(i_1,\dots,i_K)}$ are independent with Gaussian distribution. The properties above are the basis of our reasoning. Remark that if we condition by $\lambda_{K+1}^f = \ell_{K+1}$, because of the independence above, the distribution of the other *frozen* variables $Z_{i_1}, Z_{i_2}^{(i_1)}, \cdots, Z_{i_K}^{(i_1,\dots,i_{K-1})}$ and $\widehat{\sigma}_{\text{select}}^{i_1,\dots,i_{K+1}}, \widehat{\sigma}_{\text{test}}^{i_1,\dots,i_{K+1}}$ remain unchanged.

From Proposition 3, we may understand how the LAR selects a support.

Proposition 4. *Assume that the design X is such that Irrepresentable Condition (Irrep.) of order K holds. Almost surely, one has*

- *Among all possible sets $(i_1, \dots, i_K) \in \mathcal{A}_K$, there is one and only one such that*

$$\max_{i_{K+1} \neq i_1,\dots,i_K} Z_{i_{K+1}}^{(i_1,\cdots,i_K)} \leq Z_{i_K}^{(i_1,\cdots,i_{K-1})} \leq \cdots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\,; \tag{27}$$

- *this set is the selected set by LAR, namely $\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_K = i_K$;*
- *and it holds for all $(i_1, \dots, i_K) \in \mathcal{A}_K$,*

$$\mathbb{P}\big(\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_K = i_K\big) = \mathbb{P}\Big(\max_{i_{K+1} \neq i_1,\dots,i_K} Z_{i_{K+1}}^{(i_1,\cdots,i_K)} \leq Z_{i_K}^{(i_1,\cdots,i_{K-1})} \leq \cdots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\Big)\,.$$

Proof. Note that (Irrep.) implies $(\mathcal{A}_{\text{Irr.}})$ by Proposition 1. Then apply the second point of Proposition 3 to conclude. □

Finding the set $\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_K = i_K$ selected by LAR may be related to a combinatorial search testing (27) among all possible candidates $(i_1, \dots, i_K) \in \mathcal{A}_K$. Under *Irrepresentable Condition*, this support selected by LAR is given by (27) which can be seen as the extension of (1) introducing **[Q1]**.

3.2. Main results: joint law and post-selection tests construction

We assume that K is defined as in (10). Except in Section 3.5, σ^2 is assumed to be *known*. Let $(\widehat{\imath}_1,\ldots,\widehat{\imath}_K)$ be the first *signed* variables entering along the LAR path. In this section, we are interested in the joint law of the LAR's knots $(\lambda_1,\ldots,\lambda_K)$ *conditional on* λ_{K+1} *and* $(\widehat{\imath}_1,\ldots,\widehat{\imath}_K)$. To define this joint law, we need to the centering parameters m_k by (see also (23))

$$m_k := \frac{\mu^0_{\widehat{\imath}_k} - \big(R_{\widehat{\imath}_k,\widehat{\imath}_1}\cdots R_{\widehat{\imath}_k,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}\big(\mu^0_{\widehat{\imath}_1},\cdots,\mu^0_{\widehat{\imath}_{k-1}}\big)}{1-\theta^{k-1}_{\widehat{\imath}_k}}\,, \tag{28}$$

the first standard deviation $\sigma\rho_1$ with $\rho_1 := \sqrt{R_{\widehat{\imath}_1,\widehat{\imath}_1}}$, and the others $\sigma\rho_k$ by (see also (24))

$$\rho_\ell := \frac{\sqrt{R_{\widehat{\imath}_\ell,\widehat{\imath}_\ell} - \big(R_{\widehat{\imath}_\ell,\widehat{\imath}_1}\cdots R_{\widehat{\imath}_\ell,\widehat{\imath}_{\ell-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{\ell-1}}\big(R_{\widehat{\imath}_\ell,\widehat{\imath}_1},\cdots,R_{\widehat{\imath}_\ell,\widehat{\imath}_{\ell-1}}\big)}}{1-\theta^{\ell-1}_{\widehat{\imath}_\ell}} \quad \text{for } 2\leq \ell\leq K+1\,, \tag{29}$$

where

$$\theta^{\ell-1} := \theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{\ell-1})\,,\quad \text{for } 2\leq \ell\leq K+1\,,$$

is defined by (11) and $M_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{\ell-1}}$ is defined by (12).

Theorem 5 (Conditional Joint Law of the LAR Knots)**.** *Let $(\lambda_1,\ldots,\lambda_K,\lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1,\ldots,\widehat{\imath}_K,\widehat{\imath}_{K+1})$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1,\ldots,\widehat{\imath}_K)$ satisfies ($\mathcal{A}_{\mathrm{Irr.}}$) then, conditional on $\{\widehat{\imath}_1,\ldots,\widehat{\imath}_K,\lambda_{K+1}\}$, the vector $(\lambda_1,\ldots,\lambda_K)$ has law with density (w.r.t. the Lebesgue measure)*

$$\mathrm{Q}^{-1}_{(\widehat{\imath}_1,\ldots,\widehat{\imath}_K,\lambda_{K+1})}\Big(\prod_{k=1}^K \varphi_{m_k,v_k^2}(\ell_k)\Big)\mathbb{1}_{\{\ell_1\geq\ell_2\geq\cdots\geq\ell_K\geq\lambda_{K+1}\}} \text{ at point } (\ell_1,\ell_2,\ldots,\ell_K)\,,$$

where $\mathrm{Q}_{(\widehat{\imath}_1,\ldots,\widehat{\imath}_K,\lambda_{K+1})}$ is a normalizing constant, φ_{m_k,v_k^2} is the standard Gaussian density with mean m_k and variance $v_k^2 := \sigma^2\rho_k^2$ as in (28) and (29).

Proof. From the definition of the Gaussian random variable $Z^{(i_1,\ldots,i_{k-1})}_{i_k}$ in (17) one can deduce that its mean m_k is given by (28) and its standard deviation v_k by (29), considering putative indices for the selected variables. Furthermore by the first point of Proposition 3 we know that these variables are independent. We deduce that their joint density $(Z_{i_1},Z^{(i_1)}_{i_2},\ldots,Z^{(i_1,\ldots,i_{K-1})}_{i_K})$ is

$$\prod_{k=1}^K \varphi_{m_k,v_k^2}(\ell_k)\,,$$

with respect to the Lebesgue measure. Now, we are conditional on

$$\mathcal{E} := \{\widehat{\imath}_1 = i_1,\ldots,\widehat{\imath}_K = i_K,\widehat{\imath}_{K+1}=i_{K+1},\lambda_{K+1}\}\,,$$

and $(i_1,\ldots,i_K)$ satisfies ($\mathcal{A}_{\mathrm{Irr.}}$), Proposition 3 implies $\mathcal{E} = \big\{\lambda_{K+1}\leq Z^{(i_1,\ldots,i_{K-1})}_{i_K}\leq\cdots\leq Z_{i_1}\big\}$, and *on this event* $\mathcal{E}$ it holds

$$(Z_{i_1},Z^{(i_1)}_{i_2},\ldots,Z^{(i_1,\ldots,i_{K-1})}_{i_K}) = (\lambda_1,\lambda_2,\ldots,\lambda_K)\,. \tag{30}$$

Conditional on $\mathcal{E}$, the joint density of $(Z_{i_1},Z^{(i_1)}_{i_2},\ldots,Z^{(i_1,\ldots,i_{K-1})}_{i_K})$ is proportional to

$$\Big(\prod_{k=1}^K \varphi_{m_k,v_k^2}(\ell_k)\Big)\mathbb{1}_{\{\ell_1\geq\ell_2\geq\cdots\geq\ell_K\geq\lambda_{K+1}\}}\,, \tag{31}$$

with respect to the Lebesgue measure and by (30) it is the conditional density of the knots. Since this conditional density does not depend on i_{K+1}, one can de-condition on $\widehat{\imath}_{K+1}$. $\square$

A useful consequence of this theorem is that one can explicitly describe the joint law of the LAR's knots after having selected a support $\widehat{S}$ of size $\widehat{m}$ with any procedure satisfying ($\mathcal{A}_{\text{Stop}}$).

In the sequel, we denote

$$F_i := \Phi_i(\lambda_i) := \Phi\Big(\frac{\cdot}{\sigma\rho_i}\Big) \quad \text{and} \quad \mathcal{P}_{i,j} := \Phi_i \circ \Phi_j^{-1}\,, \quad \text{for } i,j \in [K+1]\,, \tag{32}$$

where $\Phi_k(\cdot)$ is the CDF of the centered Gaussian law with variance $\sigma^2\rho_k^2$ for $k \geq 1$, $\lambda_0 = \infty$ and $F_0 = 1$ by convention.

Proposition 6. *Let $a \in \mathbb{N}$ be such that $0 \leq a \leq K-1$. Let $\widehat{m}$ be a selection procedure satisfying ($\mathcal{A}_{\text{Stop}}$). Under the conditions of Theorem 5, under the null hypothesis*

$$\mathbb{H}_0 \;:\; \text{“}X\beta^0 \in H_a\text{”}, \tag{33}$$

and conditional on the selection event $\{\widehat{m} = a, F_a, F_{K+1}, \widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1}\}$, $(F_{a+1}, \ldots, F_K)$ is uniformly distributed on

$$\begin{aligned}\mathcal{D}_{a+1,K} :=\big\{&(f_{a+1}, \ldots, f_K) \in \mathbb{R}^{K-a} : \\ &\mathcal{P}_{a+1,a}(F_a) \geq f_{a+1} \geq \mathcal{P}_{a+1,a+2}(f_{a+2}) \geq \cdots \geq \mathcal{P}_{a+1,K}(f_K) \geq \mathcal{P}_{a+1,K+1}(F_{K+1})\big\}\,,\end{aligned}$$

where $\mathcal{P}_{i,j}$ are described in (32).

A proof of this corollary can be found in Appendix C.2.

Remark 8. *The previous statement is consistent with the case $a = 0$ corresponding to the* global null *hypothesis $\mathbb{H}_0$: “$X\beta^0 = 0$” (or equivalently $\mathbb{E}Z = 0$).*

Therefore, if Z is centered then, conditional on F_{K+1}, $(F_1, \ldots, F_K)$ is uniformly distributed on

$$\mathcal{D}_{1,K} := \big\{(f_1, \ldots, f_K) \in \mathbb{R}^K \;:\; 1 \geq f_1 \geq \mathcal{P}_{1,2}(f_2) \geq \cdots \geq \mathcal{P}_{1,K}(f_K) \geq \mathcal{P}_{1,K+1}(F_{K+1}))\big\}\,.$$

Remark 9. *In the orthogonal case where $\overline{R} = \mathrm{Id}$, note that $\theta_j(i_1, \ldots, i_\ell) = 0$ for all $\ell \geq 1$ and all $i_1, \ldots, i_\ell \neq j$, $\rho_j = 1$ and $\mathcal{P}_{i,j}(f) = f$. We recover that $\mathcal{D}_{1,K}$ is the set of order statistics*

$$1 \geq f_1 \geq f_2 \geq \ldots \geq f_K \geq \Phi(\lambda_{K+1}/\sigma)\,.$$

In this case, knots λ_i are Gaussian order statistics $\lambda_1 = Z_{\widehat{\imath}_1} \geq \lambda_2 = Z_{\widehat{\imath}_2} \geq \ldots \geq \lambda_K = Z_{\widehat{\imath}_K} \geq \lambda_{K+1}$ of the vector Z.

From Theorem 5, we deduce several testing statistics. To this end, we introduce some notation. Define

$$\mathcal{I}_{ab}(s,t) := \int_{\mathcal{P}_{(a+1),b}(t)}^{\mathcal{P}_{(a+1),a}(s)} \mathrm{d}f_{a+1} \int_{\mathcal{P}_{(a+2),b}(t)}^{\mathcal{P}_{(a+2),(a+1)}(f_{a+1})} \mathrm{d}f_{a+2} \int_{\mathcal{P}_{(a+3),b}(t)}^{\mathcal{P}_{(a+3),(a+2)}(f_{a+2})} \mathrm{d}f_{a+3} \cdots \int_{\mathcal{P}_{(b-1),b}(t)}^{\mathcal{P}_{(b-1),(b-2)}(f_{b-2})} \mathrm{d}f_{b-1} \tag{34}$$

$$\text{for } 0 \leq a < b \text{ and } s,t \in \mathbb{R}\,, \text{ with the convention that } \mathcal{I}_{ab} = 1 \text{ when } b = a+1\,,$$

and also

$$\mathbb{F}_{abc}(t) := \mathbb{1}_{\{\lambda_c \leq t \leq \lambda_a\}} \int_{\Phi_b(\lambda_c)}^{\Phi_b(t)} \mathcal{I}_{ab}(F_a, f_b)\mathcal{I}_{bc}(f_b, F_c)\,\mathrm{d}f_b \tag{35}$$

$$\text{for } 0 \leq a < b < c \leq K+1\,,\; t \in \mathbb{R} \text{ where } F_a = \Phi_a(\lambda_a) \text{ and } F_c = \Phi_c(\lambda_c).$$

On the numerical side, note that this quantity can be computed using *Quasi Monte Carlo* (QMC) methods as in [17, Chapter 5.1] or Appendix D. The function $\mathbb{F}_{abc}$ gives the CDF of λ_b conditional on λ_a, λ_c and on some selection event, as shown in the next proposition.

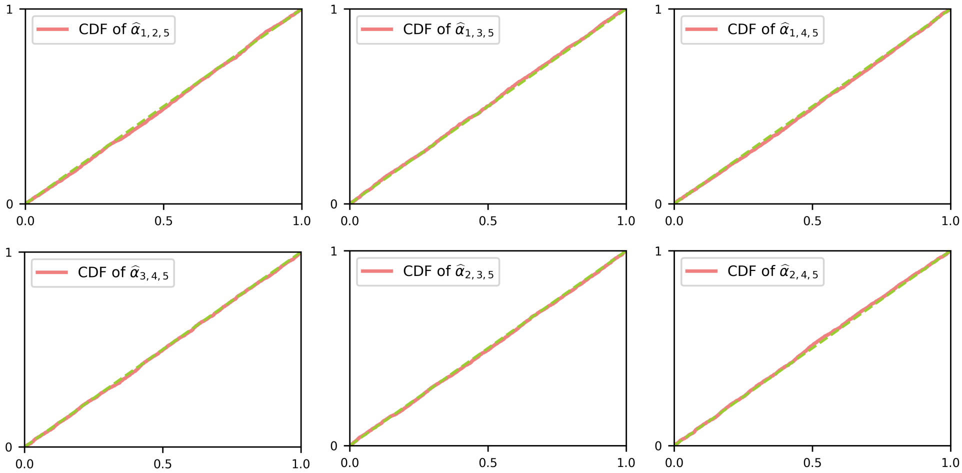

Figure 2: Observed empirical law of $\widehat{\alpha}_{abc}$ over 5,000 Monte-Carlo repetitions with $n = 200$ and $p = 300$. We considered a design $X \in \mathbb{R}^{n\times p}$ with independent column vectors uniformly distributed on the sphere and an independent $y \in \mathbb{R}^n$ with i.i.d. standard Gaussian entries, and we computed the indexes $(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)$ and the knots $(\lambda_1, \ldots, \lambda_n)$ entering the model with LAR. The displays are the empirical CDF of the $\widehat{\alpha}_{abc}$. We observe a perfect fit with the uniform distribution: the conditional law of the LAR's knots obtained theoretically is numerically validated.

Proposition 7. *Let a, b, c be such that $0 \le a < b < c \le K + 1$. Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies $(\mathcal{A}_{\mathrm{Irr.}})$ and $\widehat{m}$ is chosen according to a procedure satisfying $(\mathcal{A}_{\mathrm{Stop}})$ then under the null hypothesis*

$$\mathbb{H}_0 \; : \; \text{“}X\beta^0 \in H_a\text{”},$$

it holds that

$$\mathbb{P}\big[\lambda_b \le t \mid \widehat{m} = a, \lambda_a, \lambda_c, \widehat{\imath}_{a+1}, \ldots, \widehat{\imath}_{c-1}\big] = \frac{\mathbb{F}_{abc}(t)}{\mathbb{F}_{abc}(\lambda_a)} \,. \tag{36}$$

A proof of this proposition can be found in Appendix C.3.

Remark 10. *Note that for a fixed $a \in [K-1]$, the deterministic $\widehat{m}_{\mathrm{deter.}} = a$ is a procedure satisfying $(\mathcal{A}_{\mathrm{Stop}})$ and Proposition 7 holds true with that selection procedure $\widehat{m}_{\mathrm{deter.}}$.*

It shows that if $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies $(\mathcal{A}_{\mathrm{Irr.}})$ then, under the null hypothesis $\mathbb{H}_0$: “$X\beta^0 \in H_a$”, it holds that

$$\mathbb{P}\big[\lambda_b \le t \mid \lambda_a, \lambda_c, \widehat{\imath}_{a+1}, \ldots, \widehat{\imath}_{c-1}\big] = \frac{\mathbb{F}_{abc}(t)}{\mathbb{F}_{abc}(\lambda_a)} \,, \tag{37}$$

for any $0 \le a < b < c \le K + 1$.

From this point, we derive an exact p-value $\widehat{\alpha}_{abc}$ in the next theorem.

Theorem 8. *Let a, b, c be such that $0 \le a < b < c \le K + 1$. Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies $(\mathcal{A}_{\mathrm{Irr.}})$ and $\widehat{m}$ is chosen according to a procedure satisfying $(\mathcal{A}_{\mathrm{Stop}})$ then under the null hypothesis*

$$\mathbb{H}_0 \; : \; \text{“}X\beta^0 \in H_a\text{”},$$

and conditional on the selection event $\{\widehat{m} = a\}$ it holds that

$$\widehat{\alpha}_{abc} = \widehat{\alpha}_{abc}(\lambda_a, \lambda_b, \lambda_c, \widehat{\imath}_{a+1}, \ldots, \widehat{\imath}_{c-1}) := 1 - \frac{\mathbb{F}_{abc}(\lambda_b)}{\mathbb{F}_{abc}(\lambda_a)} \sim \mathcal{U}(0,1) \,, \tag{38}$$

namely, it is uniformly distributed over $(0, 1)$.

Proof. From Proposition 7 we know that under $\mathbb{H}_0$ and conditional on the selection event $\{\widehat{m} = a\}$, Eq. (36) gives the conditional cumulative distribution function of λ_b. As a consequence and under the same conditioning, one has

$$\frac{\mathbb{F}_{abc}(\lambda_b)}{\mathbb{F}_{abc}(\lambda_a)} \sim \mathcal{U}(0,1).$$

Finally considerations of distribution under the alternative show that to obtain a p-value we must consider the complement to 1 of the quantity above. □

This theoretical result can be numerically illustrated in Figure 2. Note that we have a perfect fit with the uniform law: the conditional law of the LAR's knots obtained theoretically is numerically validated[1].

This testing statistic generalizes previous testing statistics that appeared in "Spacing Tests", as presented in [29, Chapter 5] for instance, and will be referred to as the *Generalized Spacing test*.

Remark 11. *If one considers $a = 0$, $b = 1$ and $c = 2$ then one gets*

$$\widehat{\alpha}_{012} = 1 - \frac{\Phi_1(\lambda_1) - \Phi_1(\lambda_2)}{\Phi_1(\lambda_0) - \Phi_1(\lambda_2)} = \frac{1 - \Phi_1(\lambda_1)}{1 - \Phi_1(\lambda_2)}\,.$$

Similarly, taking $b = a + 1$ and $c = a + 2$ one gets

$$\widehat{\alpha}_{a(a+1)(a+2)} = \frac{\Phi_{a+1}(\lambda_{a+1}) - \Phi_{a+1}(\lambda_a)}{\Phi_{a+1}(\lambda_{a+2}) - \Phi_{a+1}(\lambda_a)}\,.$$

which is the spacing test as presented in [29, Chapter 5].

One can consider the following testing procedures

$$\mathcal{S}_{abc} := \mathbb{1}_{\{\widehat{\alpha}_{abc} \leq \alpha\}}\,, \tag{39}$$

that rejects if the p-value $\widehat{\alpha}_{abc}$ is less than the level α of the test. One may remark that

the p-value $\widehat{\alpha}_{abc}$ detects abnormally large values of λ_b conditional on (λ_a, λ_c).

3.3. Power studies

One may investigate the power of these tests detecting false negatives, namely alternatives given by: there exists $k \in S^0$ such that $k \notin \{\bar{\imath}_1, \ldots, \bar{\imath}_a\}$. In particular, *what is the most powerful test among these latter (39) testing procedures?*

3.3.1. Theoretical result on power for the orthogonal design case

A comprehensive study in the orthogonal case is given by the following theorem.

Theorem 9. *Assume that the design is orthogonal, namely $\overline{R} = \mathrm{Id}_p$. Let a_0 be s.t. $0 \leq a_0 \leq K-1$. If $\widehat{m}$ is chosen according to a procedure satisfying ($\mathcal{A}_{\text{Stop}}$) then under the null hypothesis*

$$\mathbb{H}_0 \;:\; \text{“}X\beta^0 \in H_{a_0}\text{”},$$

and conditional on the selection event $\{\widehat{m} = a_0\}$, it holds the test $\mathcal{S}_{a_0,a_0+1,K+1}$ is uniformly most powerful than any of the tests $\mathcal{S}_{a,b,c}$ for $a_0 \leq a < b < c \leq K+1$.

The proof of this result is given in Appendix C.4. It shows that the best choice among the tests $(\mathcal{S}_{a,b,c})_{a_0 \leq a < b < c \leq K+1}$ is

the test $\mathcal{S}_{a_0,a_0+1,K+1}$ with the smallest a and the largest c.

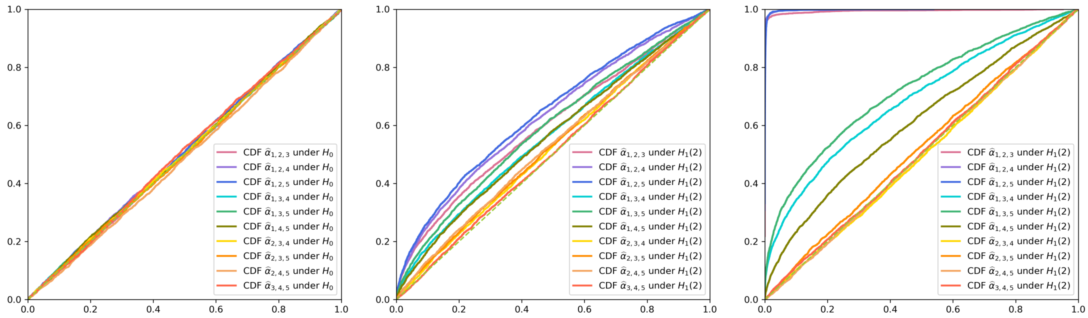

Figure 3: CDF of p-values $\widehat{\alpha}_{abc}$ over $3,000$ Monte-Carlo iterations and a random design X given by $p = 300$ independent column vectors uniformly distributed on the Euclidean sphere $\mathbb{S}^{199}$ ($n = 200$). Central panel represents alternative composed by 2-sparse vector, right panel alternative composed by 2-sparse vector 5 times larger while left panel corresponds to the null.

3.3.2. Numerical studies on power for the general design case

In the orthogonal design case, Theorem 9 shows that the test based on $\widehat{\alpha}_{a,a+1,K+1}$ is uniformly more powerful than tests based on $\widehat{\alpha}_{x,y,z}$ with $a \leq x < y < z \leq K+1$. Numerical experiments on the power of these tests are presented in Figure 3 and they witness the same phenomenon for Gaussian designs. It presents the CDF of the p-value $\widehat{\alpha}_{abc}$ under the null and under two 2-sparse alternatives, one with low signal and one with 5 times more signal. Numerical results show that all the tests are exact (leftmost panel) and the test $\mathcal{S}_{125}$ is the most powerful. More precisely, it holds, as proved in the orthogonal case by Theorem 9 (and its proof), that

- $\widehat{\alpha}_{125} \preccurlyeq \widehat{\alpha}_{124} \preccurlyeq \widehat{\alpha}_{123}$;
- $\widehat{\alpha}_{125} \preccurlyeq \widehat{\alpha}_{135} \preccurlyeq \widehat{\alpha}_{235} \preccurlyeq \widehat{\alpha}_{234}$;
- $\widehat{\alpha}_{125} \preccurlyeq \widehat{\alpha}_{135} \preccurlyeq \widehat{\alpha}_{145} \preccurlyeq \widehat{\alpha}_{245}$.

where $\preccurlyeq$ denotes stochastic ordering. In the proof of Theorem 9 it is shown that

$$\widehat{\alpha}_{ab(c+1)} \preccurlyeq \widehat{\alpha}_{abc} \text{ and } \widehat{\alpha}_{a(b-1)c} \preccurlyeq \widehat{\alpha}_{abc} \text{ and } \widehat{\alpha}_{(a-1)bc} \preccurlyeq \widehat{\alpha}_{abc}\,,$$

for orthogonal designs.

3.4. Exact false negative testing after model selection

We get back to the *general design* case. Given $\alpha \in (0,1)$ and using Theorem 8, one can consider the following *exact* testing procedure at level α on false negatives, see the pseudo-code in Algorithm 2. The theoretical guarantee of this algorithm is given by the next proposition. It shows that conditional on the event that *there is no false negatives*, namely "$X\beta^0 \in H_{\widehat{m}}$", the observed significance $\widehat{\alpha}$ has uniform law and hence $\mathbb{1}_{\{\widehat{\alpha} \leq \alpha\}}$ is a testing procedure with level exactly α.

Proposition 10. *Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies $(\mathcal{A}_{\mathrm{Irr.}})$ and $\widehat{m}$ is chosen according to a procedure satisfying $(\mathcal{A}_{\mathrm{Stop}})$ then conditional on the null hypothesis*

$$\mathbb{H}_0 \,:\, \text{“}X\beta^0 \in H_{\widehat{m}}\text{”},$$

it holds that

$$\widehat{\alpha}_{\widehat{m}(\widehat{m}+1)(K+1)} := 1 - \frac{\mathbb{F}_{\widehat{m}(\widehat{m}+1)(K+1)}(\lambda_{\widehat{m}+1})}{\mathbb{F}_{\widehat{m}(\widehat{m}+1)(K+1)}(\lambda_{\widehat{m}})} \sim \mathcal{U}(0,1)\,,$$

namely, it is uniformly distributed over $(0,1)$.

[1] A reproducible experience given in a Python notebook is available at `https://github.com/ydecastro/lar_testing/blob/master/Law_LAR.ipynb`

Algorithm 2: Exact false negative testing after model selection

Data: K satisfying (10), selection procedure $\widehat{m}$ satisfying ($\mathcal{A}_{\text{Stop}}$), couple (X, Y) giving design and response.

Result: p-value $\widehat{\alpha}$ on the existence of false negative.

/* $\mathbb{1}_{\{\widehat{\alpha} \leq \alpha\}}$ is a testing procedure with level exactly α */

1 Compute the LAR's path from (X, Y).

2 Check that $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies ($\mathcal{A}_{\text{Irr.}}$). If not **Stop**.

3 Compute $\widehat{m}$, the size of the selected model.

4 **Return** $\widehat{\alpha} = \widehat{\alpha}_{\widehat{m}(\widehat{m}+1)(K+1)}$, see (38). /* When variance is unknown, $\widehat{\alpha} = \widehat{\beta}_{\widehat{m},\widehat{m}+1,K+1}$, see (42). */

Proof. By Theorem 8, the conditional law of $\widehat{\alpha}_{a(a+1)(K+1)}$ when $\{\widehat{m} = a\}$. Remark the conditional law (38) does not depend on $a, b = a + 1, c = K + 1$, hence this law is unconditional on $\widehat{m}$. □

When the variance σ^2 is unknown, one can "Studentized" this test as presented in the next section. The reader may consult Section 3.5 for a definition and check that the quantities $\widehat{\beta}_{abc}, \widetilde{\mathbb{F}}, \Lambda_k$ do not require σ to be computed.

Proposition 11. *Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies ($\mathcal{A}_{\text{Irr.}}$) and $\widehat{m}$ is chosen according to a procedure satisfying ($\mathcal{A}_{\text{Stop}}$) then conditional on the null hypothesis*

$$\mathbb{H}_0 \; : \; \text{“} X\beta^0 \in H_{\widehat{m}} \text{”},$$

it holds that

$$\widehat{\beta}_{\widehat{m}(\widehat{m}+1)(K+1)} := 1 - \frac{\widetilde{\mathbb{F}}_{\widehat{m}(\widehat{m}+1)(K+1)}(\Lambda_{\widehat{m}+1})}{\widetilde{\mathbb{F}}_{\widehat{m}(\widehat{m}+1)(K+1)}(\Lambda_{\widehat{m}})} \sim \mathcal{U}(0,1)\,,$$

namely, it is uniformly distributed over $(0, 1)$.

Proof. By Theorem 14, the conditional law of $\widehat{\beta}_{a(a+1)(K+1)}$ when $\{\widehat{m} = a\}$. Remark the conditional law (42) does not depend on $a, b = a + 1, c = K + 1$, hence this law is unconditional on $\widehat{m}$. □

3.5. Exact Testing Procedure on False Negatives with Variance Estimation

From the result of Section 3.2, one can present a method to select a model and propose an exact test on false negatives in the *general design case* when the variance is *unknown*. We introduce a new *exact* testing procedure that can be deployed when ($\mathcal{A}_{\text{Stop}}$) holds, namely an "*admissible*" selection procedure is used to build $\widehat{S}$. We start by a preliminary result.

Proposition 12. *Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1})$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies Condition ($\mathcal{A}_{\text{Irr.}}$) then conditional on $\{\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1}, \lambda_{K+1}\}$*

- *the random variables $\big((\lambda_1, \ldots, \lambda_K), \widehat{\sigma}_{\text{select}}, \widehat{\sigma}_{\text{test}}\big)$ are independent;*
- *and, under the null hypothesis $\mathbb{H}_0$: "$X\beta^0 \in H_K$", the distribution of $(\lambda_1, \ldots, \lambda_K)$ is given by Theorem 5, the distribution of $\widehat{\sigma}_{\text{select}}$ and of $\widehat{\sigma}_{\text{test}}$ are $\sigma\chi(n_1)/\sqrt{n_1}$ and $\sigma\chi(n_2)/\sqrt{n_2}$ respectively, where the $\chi(d)$ distribution is defined as the square root of a $\chi^2(d)$ distribution.*

Remark 12. *Note that under the null hypothesis $\mathbb{H}_0$: "$X\beta^0 \in H_K$", the Gaussian vectors $P_{E_1}(Y)$ and $P_{E_2}(Y)$ (see (14)), defining the variance estimates $\widehat{\sigma}^2_{\text{select}}$ and of $\widehat{\sigma}^2_{\text{test}}$, are centered.*
This null hypothesis means that the true support is included in the set of the K first indices chosen by LAR. One may chose K large enough to guarantee this null hypothesis.

Proof of Proposition 12. Let us fix some values $i_1, \ldots, i_{K+1}$, we recall that the *frozen* values of the knots

$$\lambda_1^f := Z_{i_1}, \ldots, \lambda_K^f := Z_{i_K}^{(i_1, \ldots i_{K-1})}, \lambda_{K+1}^f := Z_{i_{K+1}}^{(i_1, \ldots i_K)},$$

are Gaussian independent and

$$\widehat{\sigma}_{\text{select}}^{i_1, \ldots, i_{K+1}} \perp\!\!\!\perp \widehat{\sigma}_{\text{test}}^{i_1, \ldots, i_{K+1}} \perp\!\!\!\perp Z_{i_{K+1}}^{(i_1, \ldots, i_K)} \perp\!\!\!\perp Z_{i_K}^{(i_1, \ldots, i_{K-1})} \perp\!\!\!\perp \cdots \perp\!\!\!\perp Z_{i_2}^{(i_1)} \perp\!\!\!\perp Z_{i_1},$$

see Proposition 3 and (26). Let us condition by $\{\widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_K = i_K, \lambda_{K+1}^f = \ell_{K+1}\}$ because of the independence above, the distribution of $\widehat{\sigma}_{\text{select}}$ and $\widehat{\sigma}_{\text{test}}$ remain unchanged and independent of the other variables. By Remark 12, these standard deviation estimates are defined from centered Gaussian vectors and hence follow χ distributions. Furthermore on this event, $\lambda_h^f = \lambda_h$, $h \in [K+1]$ and by Proposition 3 this event is equivalent to

$$\{Z_{i_1} > \cdots > Z_{i_K}^{(i_1, \ldots i_{K-1})} > \lambda_{K+1} = \ell_{K+1}\}.$$

This implies that the conditional distribution is the one claimed. □

We recall that, up to some numerical constant, the probability density function of the multivariate t-distribution with ν degrees of freedom, mean $m = (m_1, \ldots, m_K)$ and variance-covariance matrix $\mathrm{Diag}(\rho_1, \ldots, \rho_K)$ is given by

$$\widetilde{\varphi}(t_1, \ldots, t_K) := \left[1 + \frac{1}{\nu} \sum_{k=1}^{K} \left(\frac{t_k - m_k}{\rho_k}\right)^2\right]^{-\frac{\nu+K}{2}}.$$

We have an analogue to Theorem 5 giving the joint law of

$$\Lambda_k := \frac{\lambda_k}{\widehat{\sigma}_{\text{test}}} \text{ for } k = 1, \ldots, K+1, \tag{40}$$

where $\widehat{\sigma}_{\text{test}}$ is given by (14) has n_2 degrees of freedom, see Proposition 12. In the sequel, we will consider that $\nu = n_2$ in the expression of $\widetilde{\varphi}$.

Theorem 13 (Conditional Joint Law of the Studentized LAR Knots)**.** *Let* $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ *be the first knots and let* $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1})$ *be the first variables entering along the LAR path.*

If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ *satisfies* $(\mathcal{A}_{\text{Irr.}})$ *then, under the null hypothesis* $\mathbb{H}_0$: *"*$X\beta^0 \in H_K$*" and conditional on* $\{\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \Lambda_{K+1}\}$*, the vector* $(\Lambda_1, \ldots, \Lambda_K)$ *has law with density (w.r.t. the Lebesgue measure)*

$$\mathrm{P}^{-1}_{(\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \Lambda_{K+1})} \widetilde{\varphi}(t_1, \ldots, t_K) \mathbb{1}_{\{t_1 \geq t_2 \geq \cdots \geq t_K \geq \Lambda_{K+1}\}},$$

at point $(t_1, t_2, \ldots, t_K)$*, where* $\mathrm{P}_{(\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \Lambda_{K+1})}$ *is a normalizing constant,* m_k *and* ρ_k *are as in (28) and (29).*

Proof of Theorem 13. Let us fix some values $i_1, \ldots, i_{K+1}$. From the definition of the Gaussian random variable $Z_{i_k}^{(i_1, \ldots, i_{k-1})}$ in (17) one can deduce that its mean m_k is given by (28) and its standard deviation $v_k := \sigma \rho_k$ by (29), considering putative indices for the selected variables. By the proof of Proposition 12, we know that these variables are independent of $\widehat{\sigma}_{\text{test}}^{i_1, \ldots, i_{K+1}}$, which has $\sigma \frac{\chi(n_2)}{\sqrt{n_2}}$ distribution. We deduce that the vector $(Z_{i_1}, Z_{i_2}^{(i_1)}, \ldots, Z_{i_K}^{(i_1, \ldots, i_{K-1})}) / \widehat{\sigma}_{\text{test}}^{i_1, \ldots, i_{K+1}}$ has density a multivariate t-distribution with ν degrees of freedom, mean $m = (m_1, \ldots, m_K)$ and variance-covariance matrix $\mathrm{Diag}(\rho_1, \ldots, \rho_K)$. Now recall that, conditional on

$$\mathcal{E} := \{\widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_K = i_K, \widehat{\imath}_{K+1} = i_{K+1}, \lambda_{K+1}\},$$

and $(i_1, \ldots, i_K)$ satisfies $(\mathcal{A}_{\text{Irr.}})$, Proposition 3 implies that $\mathcal{E} = \{\lambda_{K+1} \leq Z_{i_K}^{(i_1, \ldots, i_{K-1})} \leq \cdots \leq Z_{i_1}\}$ or equivalently

$$\mathcal{E} = \left\{\Lambda_{K+1} \leq \frac{Z_{i_K}^{(i_1, \ldots, i_{K-1})}}{\widehat{\sigma}_{\text{test}}^{i_1, \ldots, i_{K+1}}} \leq \cdots \leq \frac{Z_{i_1}}{\widehat{\sigma}_{\text{test}}^{i_1, \ldots, i_{K+1}}}\right\}$$

and *on this event* $\mathcal{E}$ it holds

$$(Z_{i_1}, Z_{i_2}^{(i_1)}, \ldots, Z_{i_{K+1}}^{(i_1,\ldots,i_K)}, \widehat{\sigma}_{\text{test}}^{i_1,\ldots,i_{K+1}}) = (\lambda_1, \lambda_2, \ldots, \lambda_{K+1}, \widehat{\sigma}_{\text{test}})\,.$$

Since this conditional density does not depend on i_{K+1}, one can de-condition on $\widehat{\imath}_{K+1}$. $\square$

For $0 \le a < b \le K+1$, we introduce

$$\widetilde{\varphi}_{ab}(t_{a+1}, \ldots, t_{b-1}) := \Big[1 + \frac{1}{\nu}\sum_{k=a+1}^{b-1}\Big(\frac{t_k}{\rho_k}\Big)^2\Big]^{-\frac{\nu+b-a}{2}}$$
$$\widetilde{\mathcal{I}}_{ab}(s,t) := \int_{\{s\ge t_{a+1}\ge\cdots\ge t_{b-1}\ge t\}} \widetilde{\varphi}_{ab}$$

with the convention $\widetilde{\mathcal{I}}_{ab}(s,t) = 1$ when $b = a+1$: and also

$$\widetilde{\mathbb{F}}_{abc}(t) := \mathbb{1}_{\{\Lambda_c \le t \le \Lambda_a\}} \int_{\Lambda_c}^{t} \widetilde{\mathcal{I}}_{ab}(\Lambda_a, \ell_b)\, \widetilde{\mathcal{I}}_{bc}(\ell_b, \Lambda_c) \Big[1 + \frac{1}{\nu}\Big(\frac{\ell_b}{\rho_b}\Big)^2\Big]^{-\frac{\nu+1}{2}} \mathrm{d}\ell_b \tag{41}$$
$$\text{for } 0 \le a < b < c \le K+1\,,\ t \in \mathbb{R}\,.$$

When $m_{a+1} = \cdots = m_{c-1} = 0$, the function $\widetilde{\mathbb{F}}_{abc}$ gives the CDF of Λ_b conditional on Λ_a, Λ_c and on some selection event, as shown below in Theorem 14 and (43).

For $0 \le a < b < c \le K+1$, we introduce the p-value

$$\widehat{\beta}_{abc} = \widehat{\beta}_{abc}(\lambda_a, \lambda_b, \lambda_c, \widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1}, \widehat{\sigma}_{\text{test}}) = 1 - \frac{\widetilde{\mathbb{F}}_{abc}(\Lambda_b)}{\widetilde{\mathbb{F}}_{abc}(\Lambda_a)} \tag{42}$$

On the numerical side, note that this quantity can be computed using *Quasi Monte Carlo* (QMC) methods as in [17, Chapter 5.1]. We have the following result which is the analogue to Theorem 8.

Theorem 14. *Let a, b, c be such that $0 \le a < b < c \le K+1$. Let $(\lambda_1, \ldots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1})$ be the first variables entering along the LAR path. If $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ satisfies ($\mathcal{A}_{\text{Irr.}}$) and $\widehat{m}$ is chosen according to a procedure satisfying ($\mathcal{A}_{\text{Stop}}$) then under the null hypothesis*

$$\mathbb{H}_0\ :\ \text{“}X\beta^0 \in H_a\text{”},$$

and conditional on the selection event $\{\widehat{m} = a\}$ it holds that

$$\widehat{\beta}_{abc} \sim \mathcal{U}(0,1)\,,$$

namely, it is uniformly distributed over $(0,1)$.

Proof. Fix a such that $0 \le a \le K-1$ and consider any selection procedure $\widehat{m}$ satisfying ($\mathcal{A}_{\text{Stop}}$). From Proposition 2, conditional on

$$\mathcal{F} := \{\widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_K = i_K, \widehat{\imath}_{K+1} = i_{K+1}, \Lambda_a, \Lambda_{K+1}\}$$

and under the null hypothesis $\mathbb{H}_0$: “$X\beta^0 \in H_a$”, we know that $m_{a+1} = \ldots = m_K = 0$. From Theorem 13 we know that the density of $(\Lambda_{a+1}, \Lambda_{a+2}, \ldots, \Lambda_K)$ conditional on $\mathcal{F}$ is given by

$$(const)\Big[1 + \frac{1}{\nu}\sum_{k=a+1}^{K}\Big(\frac{t_k}{\rho_k}\Big)^2\Big]^{-\frac{\nu+K-a}{2}} \mathbb{1}_{\Lambda_a \le \ell_{a+1} \le \cdots \le \ell_K \le \Lambda_{K+1}}\,.$$

From definition ($\mathcal{A}_{\text{Stop}}$) and on the event $\mathcal{F}$, we know that $\mathbb{1}_{\{\widehat{m}=a\}}$ is a measurable function of $\lambda_1, \ldots, \lambda_{a-1}, \widehat{\sigma}_{\text{select}}$ which are respectively equal to $Z_{i_1}, \ldots, Z_{i_{a-1}}^{(i_1,\ldots,i_{a-2})}, \widehat{\sigma}_{\text{select}}^{i_1,\ldots,i_{K+1}}$ on $\mathcal{F}$ by (21) (as proven in Appendix A.3 and Eq. (52)). By Proposition 3 (more precisely (26)), we also know that this function is independent of $(\lambda_{a+1}, \lambda_{a+2}, \ldots, \lambda_K)$ conditional on $\mathcal{F}$. Remark that its is also independent of $\widehat{\sigma}_{\text{test}}^{i_1,\ldots,i_{K+1}}$ conditional on $\mathcal{F}$ for the same reason. We deduce that the conditional density above is also the conditional density on the event

$$\mathcal{G} := \left\{\widehat{m} = a, \widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_K = i_K, \widehat{\imath}_{K+1} = i_{K+1}, \Lambda_a, \Lambda_{K+1}\right\}.$$

Now, a simple integration shows that

$$\mathbb{P}\big[\Lambda_b \leq t \mid \widehat{m} = a, \Lambda_a, \Lambda_c, \widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1}\big] = \frac{\widetilde{\mathbb{F}}_{abc}(t)}{\widetilde{\mathbb{F}}_{abc}(\Lambda_a)}. \tag{43}$$

As a consequence and under the same conditioning, one has

$$\frac{\widetilde{\mathbb{F}}_{abc}(\Lambda_b)}{\widetilde{\mathbb{F}}_{abc}(\Lambda_a)} \sim \mathcal{U}(0,1).$$

Finally considerations of distribution under the alternative show that to obtain a p-value we must consider the complement to 1 of the quantity above. □

One can consider the following testing procedures

$$\mathcal{T}_{abc} := \mathbb{1}_{\{\widehat{\beta}_{abc} \leq \alpha\}}, \tag{44}$$

that rejects if the p-value $\widehat{\beta}_{abc}$ is less than the level α of the test. This testing statistic generalizes previous testing statistics that appeared in t-Spacing Tests, as presented in [1] for instance, and will be referred to as the *Generalized t-Spacing Test* (GTST).

Remark 13. *If one considers $a = 0$, $b = 1$ and $c = 2$ then one gets*

$$\widehat{\beta}_{012} = 1 - \frac{\boldsymbol{T}_1(\Lambda_1) - \boldsymbol{T}_1(\Lambda_2)}{\boldsymbol{T}_1(\Lambda_0) - \boldsymbol{T}_1(\Lambda_2)} = \frac{1 - \boldsymbol{T}_1(\Lambda_1)}{1 - \boldsymbol{T}_1(\Lambda_2)}.$$

Similarly, taking $b = a+1$ and $c = a+2$ one gets

$$\widehat{\beta}_{a(a+1)(a+2)} = \frac{\boldsymbol{T}_{a+1}(\Lambda_{a+1}) - \boldsymbol{T}_{a+1}(\Lambda_a)}{\boldsymbol{T}_{a+1}(\Lambda_{a+2}) - \boldsymbol{T}_{a+1}(\Lambda_a)}.$$

which is the t-spacing test as presented in [1], where

$$\boldsymbol{T}_k(\ell) := \int_{-\infty}^{\ell} \left[1 + \frac{1}{\nu}\Big(\frac{\ell}{\rho_k}\Big)^2\right]^{-\frac{\nu+1}{2}} \mathrm{d}\ell \tag{45}$$

is, up to some numerical constant, the CDF of centered t-Student distribution with variance ρ_k^2 and $\nu = n_2$ degrees of freedom.

3.6. Control of False Discovery Rate in the Orthogonal Design case

3.6.1. Presentation in the general case

For sake of readability, we will assume, for the moment, that σ is known. We understand that the law of test statistics are parametrized by the hypotheses $(m_k)_{k\in[K]}$, where m_k is given by (28).

We recall that we denote $\overline{\mu}^0 = X^\top X\beta^0$ and $\overline{\mu}_i^0$ it ith coordinate. Assuming that predictors are normalised, in the general case, this quantity is the sum of β_i^0 and a linear combination of the β_j^0's whose predictors X_j are highly correlated with the predictor X_i. Now, given $\bar{\imath}_1,\ldots,\bar{\imath}_k \in [p]$ and signs $\varepsilon_1,\ldots,\varepsilon_k \in \{\pm 1\}^k$, we denote by $(\Pi^{\perp}_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}(\overline{\mu}^0))_{\bar{\imath}_k}$ the orthogonal projection given by

$$(\Pi^{\perp}_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}(\overline{\mu}^0))_{\bar{\imath}_k} := \varepsilon_k X^{\top}_{\bar{\imath}_k}\Big[\mathrm{Id}_n - X_{\overline{S}^{k-1}}\big(X^{\top}_{\overline{S}^{k-1}}X_{\overline{S}^{k-1}}\big)^{-1}X^{\top}_{\overline{S}^{k-1}}\Big]X\beta^0. \tag{46}$$

The tested hypotheses are conditional on some sub-sequence of variables $(\bar{\imath}_1,\ldots,\bar{\imath}_{K+1}) \in [p]^{K+1}$ and signs $\varepsilon_1,\ldots,\varepsilon_{K+1} \in \{\pm 1\}^{K+1}$ entering the model. The p-values under consideration are

$$\begin{aligned}
&\circ\ \widehat{p}_1 := \widehat{\alpha}_{0,1,2} \text{ is the } p\text{-value testing } \mathbb{H}_{0,1} : \text{“}m_1 = 0\text{” namely } \overline{\mu}^0_{\bar{\imath}_1} = 0;\\
&\circ\ \widehat{p}_2 := \widehat{\alpha}_{1,2,3} \text{ is the } p\text{-value testing } \mathbb{H}_{0,2} : \text{“}m_2 = 0\text{” namely } (\Pi^{\perp}_1(\overline{\mu}^0))_{\bar{\imath}_2} = 0;\\
&\circ\ \widehat{p}_3 := \widehat{\alpha}_{2,3,4} \text{ is the } p\text{-value testing } \mathbb{H}_{0,3} : \text{“}m_3 = 0\text{” namely } (\Pi^{\perp}_2(\overline{\mu}^0))_{\bar{\imath}_3} = 0;\\
&\circ\ \text{and so on...}
\end{aligned} \tag{47}$$

We write I_0 of the set

$$I_0 = \big\{k \in [K] \ :\ \mathbb{H}_{0,k} \text{ is true}\big\}\,,$$

Given a subset $\widehat{R} \subseteq [K]$ of hypotheses that we consider as rejected, we call *false positive* (FP) and *true positive* (TP) the quantities $\mathrm{FP} = \mathrm{card}(\widehat{R} \cap I_0)$ and $\mathrm{TP} = \mathrm{card}(\widehat{R} \setminus I_0)$.

Denote by $\widehat{p}_{(1)} \le \ldots \le \widehat{p}_{(K)}$ the p-values ranked in a nondecreasing order. Let $\alpha \in (0,1)$ and consider the Benjamini-Hochberg procedure, see for instance [5], defined by a rejection set $\widehat{R} \subseteq [K]$ such that $\widehat{R} = \emptyset$ when $\{k \in [K] \ :\ \widehat{p}_{(k)} \le \alpha k/K\} = \emptyset$ and

$$\widehat{R} = \{k \in [K] \ :\ \widehat{p}_k \le \alpha\widehat{k}/K\} \quad \text{where} \quad \widehat{k} = \max\big\{k \in [K] \ :\ \widehat{p}_{(k)} \le \alpha k/K\big\}\,. \tag{48}$$

Recall the definition of FDR as the mean of False Discovery Proportion (FDP), namely

$$\mathrm{FDR} := \mathbb{E}\Big[\underbrace{\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}}\mathbb{1}_{\mathrm{FP}+\mathrm{TP}\ge 1}}_{\mathrm{FDP}}\Big]\,,$$

where the expectation is unconditional on the sequence of variables entering the model, while the hypotheses that are being tested are conditional on the sequence of variables entering the model. This FDR can be understood invoking the following decomposition

$$\mathrm{FDR} = \sum_{(\imath_1,\ldots,\imath_K)\in[p]^K} \overline{\pi}_{(\imath_1,\ldots,\imath_K)}\,\mathbb{E}\big[\mathrm{FDP}|\bar{\imath}_1 = i_1,\ldots,\bar{\imath}_K = i_K\big]\,,$$

where $\overline{\pi}_{(\imath_1,\ldots,\imath_K)} = \mathbb{P}\big\{\bar{\imath}_1 = \imath_1,\ldots,\bar{\imath}_K = \imath_K\big\}$.

3.6.2. FDR control of Benjamini–Hochberg procedure in the orthogonal design case

We now consider the orthogonal design case where $X^{\top}X = \mathrm{Id}_p$ and the set of p-values given by (47). Note that I_0 is simply the set of null coordinates of β. Remark also that, Irrepresentable Condition (Irrep.) of order p holds and so does Empirical Irrepresentable Check ($\mathcal{A}_{\mathrm{Irr.}}$), see Proposition 1.

Theorem 15. *Assume that the design is orthogonal, namely it holds $X^{\top}X = \mathrm{Id}_p$, and let $K \in [p]$. Let $(\bar{\imath}_1,\ldots,\bar{\imath}_K)$ be the first variables entering along the LAR's path. Consider the p-values given by (47) and the set $\widehat{R}$ given by (48). Then*

$$\mathbb{E}\big[\mathrm{FDP}|\bar{\imath}_1 = i_1,\ldots,\bar{\imath}_K = i_K\big] \le \alpha\,,$$

and so FDR *is upper bounded by α.*

The proof of this result is given in Appendix C.7.

One interpretation *of post-selection type* may be given as follows: if one looks at all the experiments giving the same sequence of variables entering the model $\{\bar{\imath}_1 = i_1,\ldots,\bar{\imath}_K = i_K\}$ and if one considers the Benjamini–Hochberg procedure for the hypotheses described in Section 3.6.1, then the FDR is exactly controlled by α.

4. Numerical experiments

In some cases some numerical limitations appear due to multivariate integration. All our testing statistics can be efficiently computed using Quasi Monte Carlo methods (QMC) for Multi-Variate Normal (MVN) and t (MVT) distributions, see the book [17] for a comprehensive treatment of this issue or Appendix D for a short overview of the method we used in our Python codes that can be found at https://github.com/ydecastro/lar_testing.

4.1. Examples of Admissible procedures

We are now able to present here an admissible procedure to build an estimate $\widehat{S}$ of the support that satisfies ($\mathcal{A}_{\rm Stop}$). We chose a level α' and we define a light modification of (44)

$$\mathcal{T}_{abc} = \mathbb{1}_{\{\widehat{\beta}_{abc} \leq \alpha'\}}\,, \tag{49}$$

by replacing $\widehat{\sigma}^2_{\rm test}$ by $\widehat{\sigma}^2_{\rm select}$. The number of degrees of freedom of the χ^2 distribution is now $\nu = n_1$. By a small abuse of notation, we still set

$$\Lambda_i := \frac{\lambda_i}{\widehat{\sigma}_{\rm select}}.$$

We limit our attention to consecutive a, b, c. By Remark 13, one has

$$\widehat{\beta}_{a(a+1)(a+2)} = \frac{\boldsymbol{T}_{a+1}(\Lambda_{a+1}) - \boldsymbol{T}_{a+1}(\Lambda_a)}{\boldsymbol{T}_{a+1}(\Lambda_{a+2}) - \boldsymbol{T}_{a+1}(\Lambda_a)}\,.$$

which is the *t-spacing test* as presented in [1], where $\boldsymbol{T}_{a+1}$ is defined by (45) with $\nu = n_1$ degrees of freedom, and so the quantity above is easy to compute. We are now in condition to present our algorithm.

- Begin with $a = 0$,
- at each step, perform the test $\mathcal{T}_{a,a+1,a+2}$ at the level α',
- if the test is significant, we set $a = a + 1$ and keep on going,
- if it is non-significant, we stop and set $\widehat{m} = a + 2$.

Recall that possible selected supports along the LAR's path are nested models of the form (8). Denote $k^0 \geq 1$ the smallest integer k such that the true support S^0 is contained in $\overline{S}^k$ namely

$$S^0 \subseteq \overline{S}^{k^0} \text{ and } S^0 \nsubseteq \overline{S}^{(k^0-1)}.$$

We understand that admissible procedures depend on the sequence of false positives appearing along the LAR's path.

If the experimenter believes that there is no more than γ_{FP} **consecutive** false positives in S^0 a less conservative admissible procedure would be the following.

- Begin with $a = 0$,
- at each step, perform the test $T_{a,a+1,a+2}$ at level α',
- if the test is significant, we set $a = a + 1$ and keep on going,
- if the γ_{FP} consecutive tests $T_{a,a+1,a+2}, T_{a+1,a+2,a+3}, \ldots, T_{a+\gamma_{FP}-1,a+\gamma_{FP},a+\gamma_{FP}+1}$ are all non-significants, we stop and set $\widehat{m} = a + \gamma_{FP} + 1$.

This method has been deployed on real data in Section 4.3 with $\gamma_{FP} = 3$.

4.2. Experiments

To study the relative power in a *non-orthogonal* design case we have build a Monte-Carlo experiment with $3,000$ repetitions. We have considered a model with $n = 200$, $p = 300$ and a random design matrix X given by 300 independent column vectors uniformly distributed on the Euclidean sphere $\mathbb{S}^{199}$. The computation of the function $\mathbb{F}_{abc}$ given by (35) is an important issue that demands multivariate integration tools, see Appendix D for a solution using cubature of integral by lattice rule. This has lead to some limitations namely we can compute spacings of length 4 that implies $c \leq 5$ when $a = 1$ in our experimental framework.

A python notebook and codes are given at `https://github.com/ydecastro/lar_testing`. The base function is

```
observed_significance_CBC(lars, sigma, start, end, middle)
```

in the file `multiple_spacing_tests.py`. It gives the p-value $\widehat{\alpha}_{(\mathtt{start})(\mathtt{middle})(\mathtt{end})}$ of knots and indexes given by `lars` and an estimate of (or the true) standard deviation given by `sigma`. We have run $3,000$ repetitions of this function to get the laws displayed in Figure 3. It presents the CDF of the p-value $\widehat{\alpha}_{abc}$ under the null and under two 2-sparse alternatives, one with low signal and one with 5 times more signal. Results show, in our particular case, that all the tests are exact and the test $\mathcal{S}_{125}$ is the most powerful, see Section 3.3.2 for further details.

4.3. Real data

A detailed presentation in a Python notebook is available at `https://github.com/ydecastro/lar_testing/blob/master/multiple_spacing_tests.ipynb`.

We consider a data set about HIV drug resistance extracted from [3] and [22]. The experiment consists in identifying mutations on the genes of the HIV-virus that are involved with drug resistance. The data set contains about $p = 200$ and $n = 700$ observations. Since some protocol is used to remove some gene or some individuals, the exact numbers depend on the considered drug.

We used a procedure referred to as "*spacing-BH*" procedure which is a Benjamini–Hochberg procedure based on the sequence of spacing tests

$$\widehat{\beta}_{012}, \widehat{\beta}_{123}, \ldots, \widehat{\beta}_{a(a+1)(a+2)}, \ldots$$

as described in Section 3.5 with $\alpha = 0.2$. The results for Knockoff of [3] and of Benjamini–Hochberg procedure on the coefficients of linear regression (BHq) are for the R-vignette `knockoff` of the dedicated web page of Stanford. All results are evaluated using the TSM data base that gives, in some sense, the list of true positives. A comparison of our results with those of [3] is displayed in Figure 4. Our procedure is a bit more conservative but, in most of the case, gives a better control of the FDP.

In addition we have performed on the same dataset a false negative detection as in Section 3.5, and Section 4.1 with $\gamma_{FP} = 3$. We refer to the aforementioned Python notebook for further details.

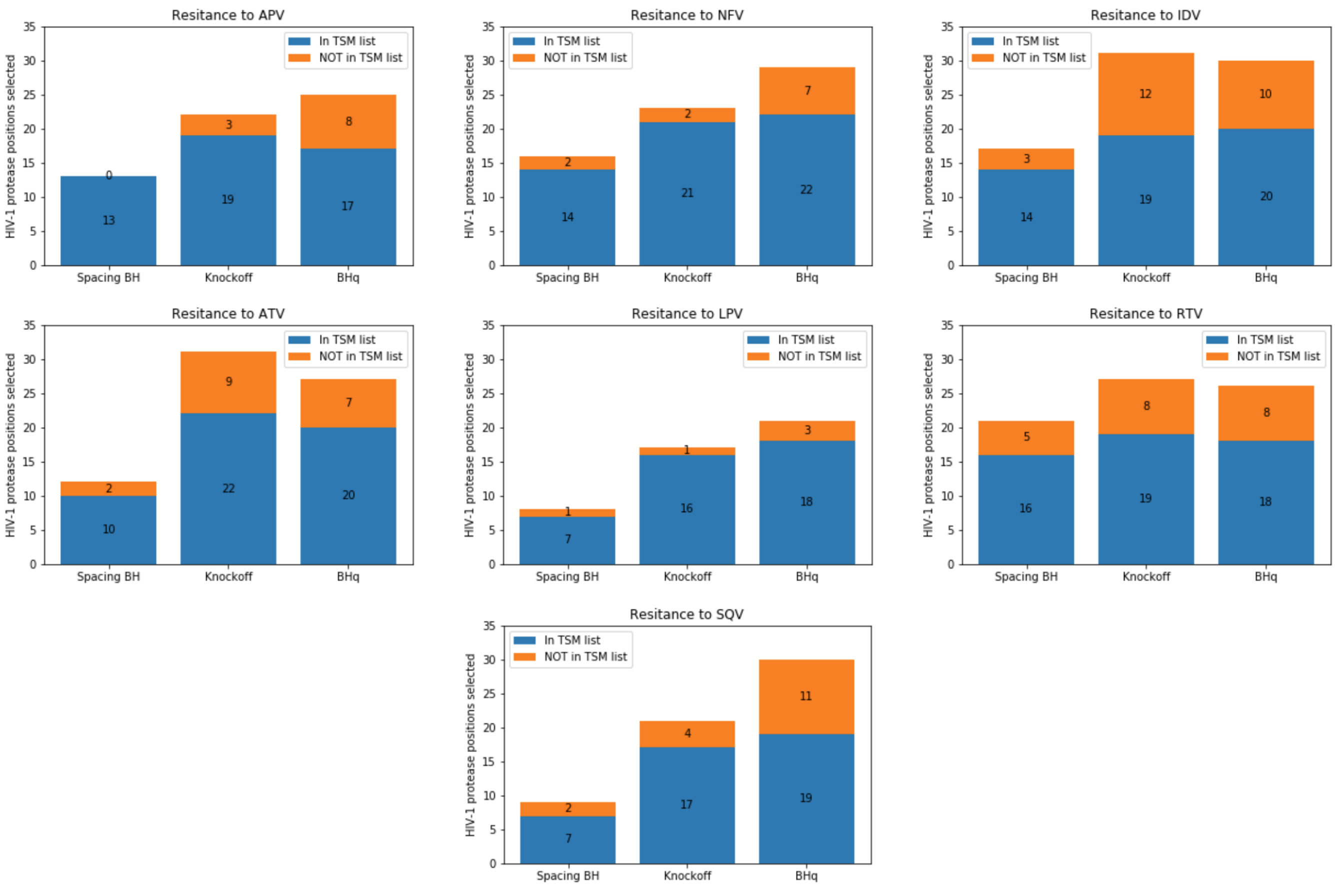

Figure 4: Comparison of numbers true and false positives for three procedures: Spacing-BH; Knockoff [3] and Benjamini–Hochberg procedure on the coefficient of linear regression BHq. For each drug we indicate the number of true positives in blue and false positives in orange. In the three procedures the aimed FDR is $\alpha = 20\%$.

References

[1] J.-M. Azaïs, Y. De Castro, and S. Mourareau. Power of the spacing test for least-angle regression. *Bernoulli*, 24(1):465–492, 2018.

[2] F. Bachoc, G. Blanchard, P. Neuvial, et al. On the post selection inference constant under restricted isometry properties. *Electronic Journal of Statistics*, 12(2):3736–3757, 2018.

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

[4] P. C. Bellec, G. Lecué, A. B. Tsybakov, et al. Slope meets lasso: improved oracle bounds and optimality. *The Annals of Statistics*, 46(6B):3603–3642, 2018.

[5] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[6] R. Berk, L. Brown, A. Buja, K. Zhang, L. Zhao, et al. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

[7] P. J. Bickel, Y. Ritov, A. B. Tsybakov, et al. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

[8] G. Blanchard, P. Neuvial, and E. Roquain. Post hoc inference via joint family-wise error rate control. *arXiv preprint arXiv:1703.02307*, 2017.

[9] G. Blanchard, E. Roquain, et al. Two simple sufficient conditions for fdr control. *Electronic journal of Statistics*, 2:963–992, 2008.

[10] M. Bogdan, E. Van Den Berg, C. Sabatti, W. Su, and E. J. Candès. Slope—adaptive variable selection via convex optimization. *The annals of applied statistics*, 9(3):1103, 2015.

[11] P. Bühlmann and S. van de Geer. *Statistics for high-dimensional data*. Springer Series in Statistics. Springer, Heidelberg, 2011. Methods, theory and applications.

[12] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61 (electronic), 1998.

[14] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[15] W. Fithian, D. Sun, and J. Taylor. Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*, 2014.

[16] A. Genz. Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1(2):141–149, 1992.

[17] A. Genz and F. Bretz. *Computation of multivariate normal and t probabilities*, volume 195. Springer Science & Business Media, 2009.

[18] C. Giraud. *Introduction to high-dimensional statistics*. Chapman and Hall/CRC, 2014.

[19] A. Javanmard, H. Javadi, et al. False discovery rate control via debiased lasso. *Electronic Journal of Statistics*, 13(1):1212–1253, 2019.

[20] R. Lockhart, J. Taylor, R. J. Tibshirani, and R. Tibshirani. A significance test for the lasso. *Annals of statistics*, 42(2):413, 2014.

[21] D. Nuyens and R. Cools. Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel hilbert spaces. *Mathematics of Computation*, 75(254):903–920, 2006.

[22] S.-Y. Rhee, J. Taylor, G. Wadhera, A. Ben-Hur, D. L. Brutlag, and R. W. Shafer. Genotypic predictors of human immunodeficiency virus type 1 drug resistance. *Proceedings of the National Academy of Sciences*, 103(46):17355–17360, 2006.

[23] E. Roquain. Type i error rate control for testing many hypotheses: a survey with proofs. *Journal de la Société Française de Statistique*, 152(2):3–38, 2011.

[24] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55:58–63, 2015.

[25] J. Taylor and R. J. Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25):7629–7634, 2015.

[26] X. Tian, J. R. Loftus, and J. E. Taylor. Selective inference with unknown variance via the square-root lasso. *Biometrika*, 105(4):755–768, 2018.

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

[28] R. Tibshirani, M. Wainwright, and T. Hastie. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Monographs on Statistics & Applied Probability. Chapman and Hall/CRC press, 2015.

[29] R. J. Tibshirani, J. Taylor, R. Lockhart, and R. Tibshirani. Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association*, 111(514):600–620, 2016.

[30] S. van de Geer. Estimation and testing under sparsity. *Lecture Notes in Mathematics*, 2159, 2016.

[31] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[32] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55(5):2183–2202, 2009.

Appendix A: Representing the LAR Knots

A.1. The equivalent formulations of the LAR algorithm

We present here three equivalent formulations of the LAR that are a consequence of the analysis provided in Appendices A and B. One formulation is given by Algorithm 1.

Algorithm 3: LAR algorithm (standard formulation)

Data: Correlations vector $\overline{Z}$ and variance-covariance matrix $\overline{R}$.
Result: Sequence $((\lambda_k, \bar{\imath}_k, \varepsilon_k))_{k\geq 1}$ where $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ are the knots, and $\bar{\imath}_1, \bar{\imath}_2, \ldots$ are the variables that enter the model with signs $\varepsilon_1, \varepsilon_2, \ldots$ ($\varepsilon_k = \pm 1$).

`/* Initialize computing` $(\lambda_1, \bar{\imath}_1, \varepsilon_1)$ `and defining a "residual"` $\overline{N}^{(1)}$. `*/`

1 Set $k = 1$, $\lambda_1 := \max |\overline{Z}|$, $\bar{\imath}_1 := \arg\max |\overline{Z}|$ and $\varepsilon_1 = \overline{Z}_{\bar{\imath}_1}/\lambda_1 \in \pm 1$, and $\overline{N}^{(1)} := \overline{Z}$.

`/* Note that` $((\lambda_\ell, \bar{\imath}_\ell, \varepsilon_\ell))_{1\leq \ell\leq k-1}$ `and` $\overline{N}^{(k-1)}$ `have been defined at the previous step. */`

2 Set $k \leftarrow k+1$ and compute the least-squares fit

$$\bar{\theta}_j := \big(\overline{R}_{j,\bar{\imath}_1} \cdots \overline{R}_{j,\bar{\imath}_{k-1}}\big) M^{-1}_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}} (\varepsilon_1, \ldots, \varepsilon_{k-1}) \,, \quad j = 1, \ldots, p \,,$$

where $M_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}$ is the sub-matrix of $\overline{R}$ keeping the columns and the rows indexed by $\{\bar{\imath}_1, \ldots, \bar{\imath}_{k-1}\}$.

3 For $0 < \lambda \leq \lambda_{k-1}$ compute the "residuals" $\overline{N}^{(k)}(\lambda) = (\overline{N}^{(k)}_1(\lambda), \ldots, \overline{N}^{(k)}_p(\lambda))$ given by

$$\overline{N}^{(k)}_j(\lambda) := \overline{N}^{(k-1)}_j - (\lambda_{k-1} - \lambda)\bar{\theta}_j \,, \quad j = 1, \ldots, p \,,$$

and pick

$$\lambda_k := \max\big\{\beta > 0 \,;\ \exists j \notin \{\bar{\imath}_1, \ldots, \bar{\imath}_{k-1}\},\ \text{s.t. } |\overline{N}^{(k)}_j(\beta)| = \beta\big\} \text{ and } \bar{\imath}_k := \underset{j\notin\{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}\}}{\arg\max} |\overline{N}^{(k)}_j(\lambda_k)| \,,$$

$$\varepsilon_k := \overline{N}^{(k)}_{\bar{\imath}_k}(\lambda_k)/\lambda_k \in \pm 1 \text{ and } \overline{N}^{(k)} := \overline{N}^{(k)}(\lambda_k) \,.$$

Then, iterate from **2**.

Algorithm 4: LAR algorithm ("projected" formulation)

Data: Correlations vector $\overline{Z}$ and variance-covariance matrix $\overline{R}$.
Result: Sequence $((\lambda_k, \bar{\imath}_k, \varepsilon_k))_{k\geq 1}$ where $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ are the knots, and $\bar{\imath}_1, \bar{\imath}_2, \ldots$ are the variables that enter the model with signs $\varepsilon_1, \varepsilon_2, \ldots$ ($\varepsilon_k = \pm 1$).

`/* Initialize computing` $(\lambda_1, \bar{\imath}_1, \varepsilon_1)$. `*/`

1 Define $Z = (\overline{Z}, -\overline{Z})$ and R as in (6), and set $k = 1$, $\lambda_1 := \max Z$, $\widehat{\imath}_1 := \arg\max Z$, $\bar{\imath}_1 = \widehat{\imath}_1 \mod p$ and $\varepsilon_1 = 1 - 2(\widehat{\imath}_1 - \bar{\imath}_1)/p \in \pm 1$.

`/* Note that` $((\lambda_\ell, \widehat{\imath}_\ell))_{1\leq \ell\leq k-1}$ `have been defined at the previous step/loop. */`

2 Set $k \leftarrow k+1$ and compute

$$\lambda_k = \max_{\{j\colon \theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1})<1\}} \left\{\frac{Z_j - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j)}{1 - \theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})}\right\} \quad \text{and} \quad \widehat{\imath}_k = \underset{\{j\colon \theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1})<1\}}{\arg\max} \left\{\frac{Z_j - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j)}{1 - \theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})}\right\},$$

where

$$\Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j) := \big(R_{j,\widehat{\imath}_1} \cdots R_{j,\widehat{\imath}_{k-1}}\big) M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}} (Z_{\widehat{\imath}_1}, \ldots, Z_{\widehat{\imath}_{k-1}})$$
$$\theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}) := \big(R_{j,\widehat{\imath}_1} \cdots R_{j,\widehat{\imath}_{k-1}}\big) M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}} (1, \ldots, 1)$$

and set $\bar{\imath}_k = \widehat{\imath}_k \mod p$ and $\varepsilon_k = 1 - 2(\widehat{\imath}_k - \bar{\imath}_k)/p \in \pm 1$. Then, iterate from **2**.

A.2. Initialization: First Knot

The first step of the LAR algorithm (Step **1** in Algorithm 3) seeks the most correlated predictor with the observation. In our formulation, introduce the first residual $N^{(1)} := Z$ and observe that $N^{(1)} := (\overline{N}^{(1)}, -\overline{N}^{(1)})$. We define the first knot $\lambda_1 > 0$ as

$$\lambda_1 = \max Z \quad \text{and} \quad \widehat{\imath}_1 = \arg\max Z\,.$$

One may see that this definition is consistent with λ_1 in Algorithm 3 and note that $\widehat{\imath}_1$ and $(\bar{\imath}_1, \varepsilon_1)$ are related as in (3).

The LAR algorithm is a forward algorithm that selects a new variable and maintains a residual at each step. We also define

$$N^{(2)}(\lambda) = N^{(1)} - (\lambda_1 - \lambda)\theta(\widehat{\imath}_1)\,, \quad 0 < \lambda \leq \lambda_1\,, \tag{50}$$

and one can check that $N^{(2)}(\lambda) = (\overline{N}^{(2)}(\lambda), -\overline{N}^{(2)}(\lambda))$ where $\overline{N}(\lambda)$ is defined in Algorithm 3. It is clear that the coordinate $\widehat{\imath}_1$ of $N^{(2)}(\lambda)$ is equal to λ. On the other hand $N^{(1)} = Z$ attains its maximum at the single point $\widehat{\imath}_1$. By continuity this last property is kept for λ in a left neighborhood of λ_1. We search for the first value of λ such that this property is not met, *i.e.* the largest value of λ such that

$$\exists j \neq \widehat{\imath}_1 \text{ such that } N^{(2)}(\lambda) = \lambda\,,$$

as in Step **3** of Algorithm 3. We call this value λ_2 and one may check that this definition is consistent with λ_2 in Algorithm 3.

Now, we can be more explicit about the expression of λ_2. Indeed, we make the following discussion on the values of $\theta_j(\widehat{\imath}_1)$.

- If $\theta_j(\widehat{\imath}_1) \geq 1$, since $N_j^{(1)} < N_{\widehat{\imath}_1}^{(1)}$ for $j \neq \widehat{\imath}_1$ there is no hope to achieve the equality between $N_j^{(2)}(\lambda)$ and $N_{\widehat{\imath}_1}^{(2)}(\lambda) = \lambda$ for $0 < \lambda \leq \lambda_1$ in view of (50).
- Thus we limit our attention to the j's such that $\theta_j(\widehat{\imath}_1) < 1$. We have equality $N_j^{(2)}(\lambda) = \lambda$ when
$$\lambda = \frac{N_j^{(1)} - \lambda_1\theta_j(\widehat{\imath}_1)}{1 - \theta_j(\widehat{\imath}_1)}.$$

So we can also define the second knot λ_2 of the LAR as

$$\lambda_2 = \max_{j : \theta_j(\widehat{\imath}_1) < 1} \left\{ \frac{Z_j - \Pi_{\widehat{\imath}_1}(Z_j)}{1 - \theta_j(\widehat{\imath}_1)} \right\}.$$

where $\Pi_{i_1}(Z_j) := Z_{i_1}\theta_j(i_1)$. Remark that $\Pi_{i_1}(Z_j) = \mathbb{E}(Z_j \mid Z_{i_1})$ is the regression of Z_j on Z_{i_1} when $\mathbb{E}Z = 0$.

A.3. Recursion: Next Knots

The loop ($\mathbf{2} \rightleftarrows \mathbf{3}$) in Algorithm 3 builds iteratively the knots $\lambda_1, \lambda_2 \ldots$ of the LAR algorithm and some "residuals" $\overline{N}^{(1)}, \overline{N}^{(2)}, \ldots$ defined in Step **3**. We will present here an equivalent formulation of these knots.

Assume that $k \geq 2$ and we have build $\lambda_1, \ldots, \lambda_{k-1}$ and selected the "signed" variables $\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}$. Introduce $N^{(k-1)} := (\overline{N}^{(k-1)}, -\overline{N}^{(k-1)})$ and define

$$N^{(k)}(\lambda) = N^{(k-1)} - (\lambda_{k-1} - \lambda)\theta(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})\,, \quad 0 < \lambda \leq \lambda_{k-1}\,.$$

Check that $\theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}) = (\overline{\theta}_j, -\overline{\theta}_j)$ where we recall that we define

$$\overline{\theta}_j := (\overline{R}_{j,\bar{\imath}_1} \cdots \overline{R}_{j,\bar{\imath}_{k-1}}) M^{-1}_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}(\varepsilon_1, \ldots, \varepsilon_{k-1})\,, \quad j = 1, \ldots, p\,,$$

at Step **2** and it holds that $\widehat{\imath}_\ell$ and $(\bar{\imath}_\ell, \varepsilon_\ell)$ are related as in (3). From this equality, we deduce that it holds $N^{(k)}(\lambda) = (\overline{N}^{(k)}(\lambda), -\overline{N}^{(k)}(\lambda))$. One may also check that the coordinates $\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}$ of $N^{(k)}(\lambda)$ are equal to λ.

Again if we want to solve $N_j^{(k)}(\lambda) = \lambda$ for some j, we have to limit our attention to j's such that $\theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}) < 1$. Solving this latter equality yields to

$$\lambda_k = \max_{j:\theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1})<1} \left\{ \frac{N_j^{(k-1)} - \lambda_{k-1}\theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})}{1 - \theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})} \right\}.$$

This expression is consistent with λ_k in Algorithm 3.

Now, we can give an other expression of λ_k that will be useful in the proofs of our main theorems. Note that the residuals satisfy the relation

$$N^{(k)} = N^{(k-1)} - (\lambda_{k-1} - \lambda_k)\theta(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}), \tag{51}$$

and that $N_j^{(k-1)} = \lambda_{k-1}$ for $j = \widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}$. The following lemma permits a drastic simplification of the expression of the knots. Its proof is given in Appendix C.5.

Lemma 16. *It holds*

$$N^{(k-1)} - \lambda_{k-1}\theta(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}) = Z - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z)$$

where we denote $\Pi_{i_1,\ldots,i_{k-1}}(Z) = (\Pi_{i_1,\ldots,i_{k-1}}(Z_1), \ldots, \Pi_{i_1,\ldots,i_{k-1}}(Z_{2p}))$ *and for all* $j \in [2p]$ *one has* $\Pi_{i_1,\ldots,i_{k-1}}(Z_j) = \big(R_{j,i_1} \cdots R_{j,i_{k-1}}\big) M^{-1}_{i_1,\ldots,i_{k-1}}(Z_{i_1}, \ldots, Z_{i_{k-1}})$.

Using Lemma 16 we deduce that λ_k in Algorithm 3 is consistent with

$$\lambda_k = \max_{j:\theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1})<1} \left\{ \frac{Z_j - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j)}{1 - \theta_j(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1})} \right\}. \tag{52}$$

where $\Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j) = \big(R_{j,\widehat{\imath}_1} \cdots R_{j,\widehat{\imath}_{k-1}}\big) M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_{\widehat{\imath}_1}, \ldots, Z_{\widehat{\imath}_{k-1}})$. When $\mathbb{E}Z = 0$, one may remark that $\Pi_{i_1,\ldots,i_{k-1}}(Z_j)$ is the regression of Z_j on the vector $(Z_{i_1}, \cdots, Z_{i_{k-1}})$ whose variance-covariance matrix is $M_{i_1,\ldots,i_{k-1}}$. This analysis leads to an equivalent formulation of the LAR algorithm (Algorithm 3). We present this formulation in Algorithm 4.

Remark 14. *Note that Algorithm 3 implies that* $\widehat{\imath}_1, \ldots, \widehat{\imath}_k$ *are pairwise different, but also that they differ modulo* p.

Appendix B: First Steps to Derive the Joint Law of the LAR Knots

B.1. Law of the First Knot

One has the following lemma governing the law of λ_1.

Lemma 17. *It holds that*

- Z_{i_1} *is independent of* $(Z_j^{(i_1)})_{j \neq i_1}$,
- *If* $\theta_j(i_1) < 1$ *for all* $j \neq i_1$ *then*

$$\big\{\widehat{\imath}_1 = i_1\big\} = \big\{\lambda_2^{(i_1)} \leq Z_{i_1}\big\},$$

- *If* $\theta_j(i_1) < 1$ *for all* $j \neq i_1$ *then, conditional on* $\{\widehat{\imath}_1 = i_1\}$ *and* λ_2, λ_1 *is a truncated Gaussian random variable with mean* $\mathbb{E}(Z_{i_1})$ *and variance* $\rho_1^2 := R_{\widehat{\imath}_1,\widehat{\imath}_1}$ *subject to be greater than* λ_2.

Proof. The first point is a consequence or the properties of Gaussian regression. Now, observe that

$$
\begin{aligned}
\big\{\lambda_2^{(i_1)} \le Z_{i_1}\big\} &\Leftrightarrow \Big\{\forall j \neq i_1\,,\ \frac{Z_j - Z_{i_1}\theta_j(i_1)}{1-\theta_j(i_1)} \le Z_{i_1}\Big\} \\
&\Leftrightarrow \big\{\forall j \neq i_1\,,\ Z_j - Z_{i_1}\theta_j(i_1) \le Z_{i_1} - Z_{i_1}\theta_j(i_1)\big\} \\
&\Leftrightarrow \big\{\forall j \neq i_1\,,\ Z_j \le Z_{i_1}\big\} \\
&\Leftrightarrow \big\{\widehat{\imath}_1 = i_1\big\}\,,
\end{aligned}
$$

as claimed. The last statement is a consequence of the two previous points. □

B.2. Recursive Formulation of the LAR

One has the following proposition whose proof can be found in Section C.6. As we will see in this section, this intermediate result as a deep consequence, the LAR algorithm can be stated in a recursive way applying the same function repeatedly, as presented in Algorithm 1.

Proposition 18. *Set*

$$
\tau_{j,i_k} := \frac{R_{j,i_k} - \big(R_{j,i_1} \cdots R_{j,i_{k-1}}\big) M^{-1}_{i_1,\ldots,i_{k-1}} \big(R_{i_k,i_1},\cdots,R_{i_k,i_{k-1}}\big)}{(1-\theta_j(i_1,\ldots,i_{k-1}))(1-\theta_{i_k}(i_1,\ldots,i_{k-1}))}\,,
$$

and observe that τ_{j,i_k} is the covariance between $Z_j^{(i_1,\ldots,i_{k-1})}$ and $Z_{i_k}^{(i_1,\ldots,i_{k-1})}$. Furthermore, it holds

$$
\frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} = 1 - \frac{1-\theta_j(i_1,\ldots,i_k)}{1-\theta_j(i_1,\ldots,i_{k-1})} \tag{53}
$$

and

$$
\forall j \neq i_1,\ldots,i_k\,,\quad Z_j^{(i_1,\ldots,i_k)} = \frac{Z_j^{(i_1,\ldots,i_{k-1})} - Z_{i_k}^{(i_1,\ldots,i_{k-1})}\tau_{j,i_k}/\tau_{i_k,i_k}}{1-\tau_{j,i_k}/\tau_{i_k,i_k}}\,. \tag{54}
$$

Now, we present Algorithm 1. Define $R(0) := R$, $Z(0) = Z$ and $T(0) = 0$. For $k \ge 1$ and fixed $i_1,\ldots,i_k \in [2p]$, introduce

$$
\begin{aligned}
R(k) &:= \Big(R_{j,\ell} - \big(R_{j,i_1} \cdots R_{j,i_k}\big) M^{-1}_{i_1,\ldots,i_k} \big(R_{\ell,i_1},\cdots,R_{\ell,i_k}\big)\Big)_{j,\ell} \\
Z(k) &:= Z - \Pi_{i_1,\ldots,i_k}(Z) \\
T(k) &:= (\theta_j(i_1,\ldots,i_k))_j\,,
\end{aligned}
$$

and note that $R(k)$ is the variance-covariance matrix of the Gaussian vector $Z(k)$. The key property is following. Let $v_1,\ldots,v_k$, be k linearly independent vectors of an Euclidean space and let u be any vector of the space. Set

$$
v := P^{\perp}_{(v_1,\ldots,v_{k-1})} v_k,
$$

the projection of v_k orthogonally to $v_1,\ldots,v_{k-1}$.Then

$$
P^{\perp}_{(v_1,\ldots,v_k)} u = P^{\perp}_{v} P_{(v_1,\ldots,v_{k-1})} u\,.
$$

Using this result we deduce that

$$
\begin{aligned}
Z(k) &= \Pi^{\perp}_{i_1,\ldots,i_k}(Z) \\
&= \Pi^{\perp}_{i_k}(\Pi^{\perp}_{i_1,\ldots,i_{k-1}}(Z)) \\
&= \Pi^{\perp}_{i_k}(Z(k-1)) \\
&= Z(k-1) - \Pi_{i_k}(Z(k-1)) \\
&= Z(k-1) - \mathbf{x}(k) Z(k-1)\,,
\end{aligned} \tag{55}
$$

where $\mathbf{x}(k) = R_{i_k}(k-1)/R_{i_k,i_k}(k-1)$. It yields that

$$R(k) = R(k-1) - \mathbf{x}(k) R_{i_k}(k-1)^\top \,. \tag{56}$$

Using (53) (or (66)), remark that

$$T(k) = T(k-1) - \mathbf{x}(k)(1 - T_{i_k}(k-1)) \,. \tag{57}$$

These relations give a recursive formulation of the LAR as presented in Algorithm 1.

Appendix C: Proofs

C.1. Proof of Proposition 3

The proof of the first point can be lead by induction. The initialization of the proof is given by the first point of Lemma 17. Now, observe that $Z_{i_k}^{(i_1,\ldots,i_{k-1})}, \cdots, Z_{i_2}^{(i_1)}, Z_{i_1}$ are measurable functions of $(Z_{i_1}, \ldots, Z_{i_k})$ and one may check that the vector $(Z_{i_1}, \ldots, Z_{i_k})$ is independent of the vector $(Z_{i_{k+1}}^{(i_1,\ldots,i_k)}, (Z_j^{(i_1,\ldots,i_{k+1})})_{j\neq i_1,\ldots,i_{k+1}})$. We deduce that $Z_{i_k}^{(i_1,\ldots,i_{k-1})}, \cdots, Z_{i_2}^{(i_1)}, Z_{i_1}$ are independent of $(Z_{i_{k+1}}^{(i_1,\ldots,i_k)}, (Z_j^{(i_1,\ldots,i_{k+1})})_{j\neq i_1,\ldots,i_{k+1}})$. One can also check that $Z_{i_{k+1}}^{(i_1,\ldots,i_k)}$ is independent of $(Z_j^{(i_1,\ldots,i_{k+1})})_{j\neq i_1,\ldots,i_{k+1}}$. We know also that they are independent of $P_{K+1}^{\perp} Y$ leading to their independence of the estimates of the standard error given by (14). These two estimates are independent since they are built on orthogonal Gaussian vectors.

The second point also works by induction. The initialization of the proof is given by the second point of Lemma 17. We will use Proposition 18 to prove the second point.
Now, we have

$$\begin{aligned}
&\lambda_{k+1}^{(i_1,\ldots,i_k)} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} \\
&\Leftrightarrow \forall j \neq i_1,\ldots,i_k\,,\ Z_j^{(i_1,\ldots,i_k)} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} \\
&\Leftrightarrow \forall j \neq i_1,\ldots,i_k\,,\ Z_j^{(i_1,\ldots,i_{k-1})} - Z_{i_k}^{(i_1,\ldots,i_{k-1})} \frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} - Z_{i_k}^{(i_1,\ldots,i_{k-1})} \frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} \\
&\Leftrightarrow \forall j \neq i_1,\ldots,i_k\,,\ Z_j^{(i_1,\ldots,i_{k-1})} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} \\
&\Leftrightarrow \lambda_k^{(i_1,\ldots,i_{k-1})} = Z_{i_k}^{(i_1,\ldots,i_{k-1})} \,.
\end{aligned} \tag{58}$$

using that (54) and that $1 - \tau_{j,i_k}/\tau_{i_k,i_k} > 0$ (which is a consequence of (53) and ($\mathcal{A}_{\mathrm{Irr.}}$)) in (58). By induction and using (58), it holds that

$$\begin{aligned}
&\{\lambda_{k+1}^{(i_1,\ldots,i_k)} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} \leq Z_{i_{k-1}}^{(i_1,\ldots,i_{k-2})} \leq \ldots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\} \\
&\quad \Leftrightarrow \{\forall j \neq i_1,\ldots,i_k\,,\ Z_j^{(i_1,\ldots,i_{k-1})} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})} \leq Z_{i_{k-1}}^{(i_1,\ldots,i_{k-2})} \leq \ldots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\} \\
&\quad \Leftrightarrow \{\forall j \neq i_1,\ldots,i_{k-1}\,,\ Z_j^{(i_1,\ldots,i_{k-1})} \leq Z_{i_{k-1}}^{(i_1,\ldots,i_{k-2})} \leq \ldots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1} \\
&\qquad\quad \text{and } \forall j \neq i_1,\ldots,i_k\,,\ Z_j^{(i_1,\ldots,i_{k-1})} \leq Z_{i_k}^{(i_1,\ldots,i_{k-1})}\} \\
&\quad \Leftrightarrow \{\lambda_k^{(i_1,\ldots,i_{k-1})} \leq Z_{i_{k-1}}^{(i_1,\ldots,i_{k-2})} \leq \ldots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1} \text{ and } \lambda_k^{(i_1,\ldots,i_{k-1})} = Z_{i_k}^{(i_1,\ldots,i_{k-1})}\} \\
&\quad \vdots \\
&\quad \Leftrightarrow \{\lambda_a^{(i_1,\ldots,i_{a-1})} \leq Z_{i_{a-1}}^{(i_1,\ldots,i_{a-2})} \leq \ldots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1} \\
&\qquad\quad \text{and } \lambda_k^{(i_1,\ldots,i_{k-1})} = Z_{i_k}^{(i_1,\ldots,i_{k-1})} \leq \ldots \leq \lambda_a^{(i_1,\ldots,i_{a-1})} = Z_{i_a}^{(i_1,\ldots,i_{a-1})}\} \qquad (s_a) \\
&\quad \vdots \\
&\quad \Leftrightarrow \{\widehat{\imath}_1 = i_1, \ldots, \widehat{\imath}_k = i_k\} \,.
\end{aligned}$$

Now, observe that $\widehat{i}_{k+1}$ is the (unique) arg max of $\lambda_{k+1}^{(i_1,\ldots,i_k)}$ on the event $\{\widehat{\imath}_1 = i_1,\ldots,\widehat{\imath}_k = i_k\}$. It yields that

$$\{\widehat{\imath}_1 = i_1,\ldots,\widehat{\imath}_{k+1} = i_{k+1}\} \quad = \{\lambda_{k+1}^{(i_1,\ldots,i_k)} = Z_{i_{k+1}}^{(i_1,\cdots,i_k)} \leq Z_{i_k}^{(i_1,\cdots,i_{k-1})} \leq \cdots \leq Z_{i_2}^{(i_1)} \leq Z_{i_1}\},$$

as claimed. Stopping at a as in (s_a) gives the second part of the statement.

C.2. Proof of Proposition 6

Fix a such that $0 \leq a \leq K-1$ and consider any selection procedure $\widehat{m}$ satisfying ($\mathcal{A}_{\text{Stop}}$). From Theorem 5 (more precisely (31)) we know that the density of $(\lambda_{a+1},\lambda_{a+2},\ldots,\lambda_K)$ conditional on

$$\mathcal{F} := \{\widehat{\imath}_1 = i_1,\ldots,\widehat{\imath}_K = i_K, \widehat{\imath}_{K+1} = i_{K+1}, \lambda_a, \lambda_{K+1}\}$$

is given by

$$(const)\Big(\prod_{k=a+1}^{K} \varphi_{m_k,v_k^2}(\ell_k)\Big)\,\mathbb{1}_{\lambda_a \leq \ell_{a+1} \leq \cdots \leq \ell_K \leq \lambda_{K+1}}\,.$$

From Proposition 2, conditional on $\mathcal{F}$ and under the null hypothesis of Proposition 6, we know that $m_{a+1} = \ldots = m_K = 0$. It implies that Φ_k is the CDF of λ_k for $a < k \leq K$.

From definition ($\mathcal{A}_{\text{Stop}}$) and on the event $\mathcal{F}$, we know that $\mathbb{1}_{\{\widehat{m}=a\}}$ is a measurable function of $\lambda_1,\ldots,\lambda_{a-1},\widehat{\sigma}_{\text{select}}$ which are respectively equal to $Z_{i_1},\ldots,Z_{i_{a-1}}^{(i_1,\ldots,i_{a-2})},\widehat{\sigma}_{\text{select}}^{i_1,\ldots,i_{K+1}}$ on $\mathcal{F}$ by (21) (as proven in Appendix A.3 and Eq. (52)). By Proposition 3 (more precisely (26)), we also know that this function is independent of $(\lambda_{a+1},\lambda_{a+2},\ldots,\lambda_K)$ conditional on $\mathcal{F}$. Remark that its is also independent of $\widehat{\sigma}_{\text{test}}^{i_1,\ldots,i_{K+1}}$ conditional of $\mathcal{F}$ for the same reason (it would be useful later, when we will build testing procedures when the variance is unknown). We deduce that the conditional density above is also the conditional density on the event

$$\mathcal{G} := \{\widehat{m} = a, \widehat{\imath}_1 = i_1,\ldots,\widehat{\imath}_K = i_K, \widehat{\imath}_{K+1} = i_{K+1}, \lambda_a, \lambda_{K+1}\}\,.$$

From $F_k = \Phi_k(\lambda_k)$ (*i.e.*, applying the CDF) we deduce by a change of variables that conditional on the selection event $\mathcal{G}$, the vector $(F_{a+1},\ldots,F_K)$ is uniformly distributed on

$$\begin{aligned}\mathcal{D}_{a+1,K} :=\{&(f_{a+1},\ldots,f_K) \in \mathbb{R}^{K-a} :\\ &\mathcal{P}_{a+1,a}(F_a) \geq f_{a+1} \geq \mathcal{P}_{a+1,a+2}(f_{a+2}) \geq \cdots \geq \mathcal{P}_{a+1,K}(f_K) \geq \mathcal{P}_{a+1,K+1}(F_{K+1})\}\,,\end{aligned}$$

where $\mathcal{P}_{i,j}$ are described in (32).

C.3. Proof of Proposition 7

By Proposition 6, a simple integration shows that

$$\mathbb{P}\big[\lambda_b \leq t \mid \widehat{m} = a, \lambda_a, \lambda_c, \widehat{\imath}_1,\ldots,\widehat{\imath}_a,\widehat{\imath}_{a+1},\ldots,\widehat{\imath}_{c-1},\widehat{\imath}_c,\ldots,\widehat{\imath}_{K+1}\big] = \frac{\mathbb{F}_{abc}(t)}{\mathbb{F}_{abc}(\lambda_a)}\,,$$

under the null hypothesis of Proposition 7 (which implies that $m_{a+1} = \ldots = m_K = 0$). Then note that the function $\mathbb{F}_{abc}$ is defined by $\sigma,\lambda_a,\lambda_c,\widehat{\imath}_{a+1},\ldots,\widehat{\imath}_{c-1}$ only. We deduce that we can de-condition on $\widehat{m} = a, \widehat{\imath}_1,\ldots,\widehat{\imath}_a,\widehat{\imath}_c,\ldots,\widehat{\imath}_{K+1}$, which gives the result.

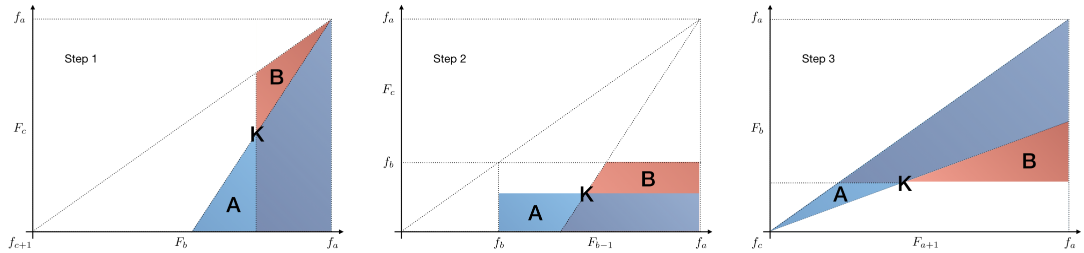

Figure 5: Rejection domains associated to the different comparison sets appearing in steps of the proof of Theorem 9.

C.4. Orthogonal Case: Proof of Theorem 9

Let $\mathcal{I}$ the set of admissible indexes

$$\mathcal{I} := \{a, b, c : a_0 \leq a < b < c \leq K + 1\}.$$

◦ **Step 1:** We prove that, when the considered indexes such that $c + 1 \leq K + 1$ belong to $\mathcal{I}$, $\mathcal{S}_{a,b,c+1}$ is more powerful than $\mathcal{S}_{a,b,c}$. Our proof is conditional to $F_a = f_a, F_{c+1} = f_{c+1}$. Note that $(F_{a+1}, \dots, F_c)$ has for distribution the uniform distribution on the simplex

$$\mathcal{S} := \{f_a > F_{a+1} > \dots > F_c > f_{c+1}\}.$$

This implies by direct calculations that

$$\mathcal{I}_{ab}(s,t) = \frac{(s-t)^{b-a-1}}{(b-a-1)!}$$

and that

$$\frac{\mathbb{F}_{abc}(\lambda_b)}{\mathbb{F}_{abc}(\lambda_a)} = \boldsymbol{F}_{\beta((b-a),(c-b))}\left(\frac{F_b - F_c}{f_a - F_c}\right). \tag{59}$$

where $\boldsymbol{F}_\beta$ is the cumulative distribution of the Beta distribution in reference. Using monotony of this function the $\mathcal{S}_{abc}$ test has for rejection region

$$(F_b - F_c) \geq z_1(f_a - F_c) \Leftrightarrow F_b \geq z_1 f_a + (1 - z_1)F_c, \tag{60}$$

where z_1 is some threshold, depending on α, that belongs to $(0, 1)$.

Similarly $\mathcal{S}_{ab(c+1)}$ has for rejection region

$$F_b \geq z_2(f_a - f_{c+1}) + f_{c+1}, \tag{61}$$

where z_2 is some other threshold belonging to $(0, 1)$. We use the following lemmas.

Lemma 19. *Let $c \leq K$. The density h_μ of $f_1, \dots, f_c$, conditional on F_{c+1} with respect of the Lebesgue measure under the alternative is coordinate-wise non-decreasing and given by* (62).

Proof. Observe that it suffices to prove the result when $\sigma = 1$. Note that

$$\lambda_{c+1}^{i_1,\dots,i_c} = \max_{j\in[p]\,,\ j\neq\bar{\imath}_1,\dots,j\neq\bar{\imath}_c} |Z_j|.$$

Thus its density $p_{\mu^0,i_1,\dots,i_c}$ does not depend on $\mu^0_{\bar{\imath}_1}, \dots, \mu^0_{\bar{\imath}_c}$. As a consequence the following variables have the same distribution ; $\lambda_{c+1}^{i_1+\epsilon_1 p,\dots,i_c+\epsilon_c p}$, where $\epsilon_1, \dots, \epsilon_c$ take the value 0 or 1 and indices are taken modulo p.

Because of independence of the different variables, the joint density, under the alternative, of $\lambda_1,\dots,\lambda_{c+1}$ taken at $\ell_1,\dots,\ell_{c+1}$, on the domain $\{\lambda_1>\dots>\lambda_{c+1}\}$ takes the value

$$(Const)\sum{}^{'}\big(\varphi(\ell_1-\mu^0_{j_1})+\varphi(\ell_1+\mu^0_{j_1})\big),\dots,\big(\varphi(\ell_c-\mu^0_{j_c})+\varphi(\ell_c+\mu^0_{j_c})\big)p_{\mu^0,j_1,\dots,j_K}(\ell_{k+1}).$$

Here the sum $\sum{}^{'}$ is taken over all different $j_1,\dots,j_c$ belonging to $[\![1,p]\!]$.

Then the density, conditional on $F_{c+1}=f_{c+1}$, of $F_1,\dots,F_c$ at $f_1,\dots,f_c$ takes the value

$$(\text{const})\sum{}^{'}\cosh(\mu_{j_1}f_1)\dots\cosh(\mu_{j_c}f_c)\mathbb{1}_{f_1>\dots>f_c>F_{c+1}},\tag{62}$$

implying that this density is coordinate-wise non-decreasing. □

Lemma 20. *Let ν_0 the image on the plane (F_b,F_c) on the uniform probability on $\mathcal{S}$: it is the distribution under the null of (F_b,F_c). The two rejection regions : $\mathcal{R}_1$ associated to (60) and $\mathcal{R}_2$ associated to (61) have of course the same probability α under ν_0. Let η_{μ^0} the density w.r.t. ν_0 of the distribution of (F_b,F_c) under the alternative.*

Then η_{μ^0} is non decreasing coordinate-wise.

Proof. Integration yields that density of ν_0 w.r.t. the Lebesgue measure taken at point (f_b,f_c) is

$$\frac{(f_a-f_b)^{b-a-1}(f_b-f_c)^{c-b-1}}{(b-a-1)!(c-b-1)!}.$$

The density of ν_{μ^0} w.r.t. Lebesgue measure is

$$\int_{f_b}^{f_a}df_{a+1}\dots\int_{f_b}^{f_b-2}df_{a+1}\int_{f_b}^{f_b-2}df_{b-1}\int_{f_c}^{f_b}df_{b+1}\dots\int_{f_c}^{f_{c-2}}df_{c-1}h_{\mu^0}(f_a,\dots,f_c).\tag{63}$$

Thus η_{μ^0} which is the quotient of these two quantities is just a mean value of h_{μ^0} on the domain of integration $\mathcal{D}_{f_b,f_c}$ in (63).

Suppose that f_b and f_c increase, then all the borns of the domain $\mathcal{D}_{f_b,f_c}$ increase also. By Lemma 19 the mean value increases. □

We finish now the proof of Step 1: For a given level α let us consider the two rejection regions $R_{a,b,c}$ and $R_{a,b,(c+1)}$ of the two considered tests in the plane F_b,F_c and set

$$A:=R_{a,b,c}\setminus R_{a,b,(c+1)}\text{ and }B:=R_{a,b,(c+1)}\setminus R_{a,b,c}\,,$$

see Figure 5. These two regions have the same ν_0 measure. By elementary geometry there exist a point $K=(K_b,K_c)$ in the plane such that

- For every point of A, $F_b\le K_b$, $F_c\le K_c$,
- For every point of B, $F_b\ge K_b$, $F_c\ge K_c$,

By transport of measure there exists a transport function $\mathcal{T}$ that preserve the measure ν_0 and that is one-to one $A\to B$. As a consequence the transport by $\mathcal{T}$ improve the probability under the alternative: the power of $\mathcal{S}_{a,b,c+1}$ is larger than that of $\mathcal{S}_{a,b,c}$.

∘ **Step 2:** We prove that, when the considered indexes belong to $\mathcal{I}$ such that $a<b-1$, $\mathcal{S}_{a,(b-1),c}$ is more powerful than $\mathcal{S}_{a,b,c}$. Our proof is conditional on $F_a=f_a,F_b=f_b$ and is located in the plane (F_{b-1},F_c).

The rejection region $R_{a,b,c}$ takes the form $F_c\le\frac{1}{1-z_1}f_b-\frac{z_1}{1-z_1}f_a$ for some threshold z_1 belonging to $(0,1)$.

The rejection region $R_{a,(b-1),c}$ takes the form $F_c\le\frac{1}{1-z_2}F_{b-1}-\frac{z_2}{1-z_2}f_a$ for some other threshold z_2

belonging to $(0, 1)$.
These regions as well as the regions A and B and the point K are indicated in Figure 5.

Transport of measure and the convenient modification of Lemma 20 imply that the power of $\mathcal{S}_{a,(b-1),c}$ is greater of equal that that of $\mathcal{S}_{a,b,c}$.

∘ **Step 3:** We prove that, when the considered indexes belong to $\mathcal{I}$ such that $a+1<b$, $\mathcal{S}_{a,b,c}$ is more powerful than $\mathcal{S}_{(a+1),b,c}$. Our proof is conditional on $F_a = f_a, F_c = f_c$ and is located in the plane F_{a+1}, F_b.
The rejection region $R_{a,b,c}$ takes the form $F_b \geq z_1 f_a + (1-z_1) f_c$ for some threshold $z_1$belonging to $(0, 1)$.
The rejection region $R_{a+1,b,c}$ takes the form $F_b \geq z_2 F_{a+1} + (1-z_2) f_c$ for some other threshold z_2 belonging to $(0, 1)$.
These regions as well as the regions A and B and the point K are indicated in Figure 5.

Transport of measure and the convenient modification of Lemma 20 imply that the power of $\mathcal{S}_{a,b,c}$ is greater of equal that that of $\mathcal{S}_{(a+1),b,c}$.

Considering the three cases above, we get the desired result.

C.5. *Proof of Lemma 16*

The proof works by induction. Let us check the relation for $k = 2$, namely

$$N^{(1)} - \lambda_1 \theta(\widehat{\imath}_1) = Z - Z_{\widehat{\imath}_1}\theta(\widehat{\imath}_1) = Z - \Pi_{\widehat{\imath}_1}(Z)\,.$$

Now, let $k \geq 3$. First, the three perpendicular theorem implies that for every $j, i_1, \ldots, i_{k-1}$,

$$\begin{aligned}\theta_j(i_1,\ldots,i_{k-2}) &= \big(R_{j,i_1}\cdots R_{j,i_{k-1}}\big)M^{-1}_{i_1,\ldots,i_{k-1}}(\theta_{i_1}(i_1,\ldots,i_{k-2}),\ldots,\theta_{i_{k-1}}(i_1,\ldots,i_{k-2}))\,,\\ \text{and } \Pi_{i_1,\ldots,i_{k-2}}(Z_j) &= \big(R_{j,i_1}\cdots R_{j,i_{k-1}}\big)M^{-1}_{i_1,\ldots,i_{k-1}}(\Pi_{i_1,\ldots,i_{k-2}}(Z_{i_1}),\ldots,\Pi_{i_1,\ldots,i_{k-2}}(Z_{i_{k-1}}))\,.\end{aligned}$$

By induction, using (51), we get that

$$\begin{aligned}N^{(k-1)} &= N^{(k-2)} - (\lambda_{k-2}-\lambda_{k-1})\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})\,,\\ &= (N^{(k-2)} - \lambda_{k-2}\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})) + \lambda_{k-1}\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})\,,\\ &= Z - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}}(Z) + \lambda_{k-1}\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})\,.\end{aligned} \tag{64}$$

Then, recall that $N_j^{(k-1)} = \lambda_{k-1}$ for $j = \widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}$ and remark that

$$\lambda_{k-1}\theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}) = \big(R_{j,\widehat{\imath}_1}\cdots R_{j,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(N^{(k-1)}_{\widehat{\imath}_1},\ldots,N^{(k-1)}_{\widehat{\imath}_{k-1}})\,.$$

Using (64) at indices $j = \widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}$, we deduce that

$$\begin{aligned}&\lambda_{k-1}\theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1})\\ &\quad= \big(R_{j,\widehat{\imath}_1}\cdots R_{j,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(N^{(k-1)}_{\widehat{\imath}_1},\ldots,N^{(k-1)}_{\widehat{\imath}_{k-1}})\\ &\quad= \big(R_{j,\widehat{\imath}_1}\cdots R_{j,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_{\widehat{\imath}_1},\ldots,Z_{\widehat{\imath}_{k-1}})\\ &\qquad- \big(R_{j,\widehat{\imath}_1}\cdots R_{j,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(\Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}}(Z_{\widehat{\imath}_1}),\ldots,\Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}}(Z_{\widehat{\imath}_{k-1}}))\\ &\qquad+ \lambda_{k-1}\big(R_{j,\widehat{\imath}_1}\cdots R_{j,\widehat{\imath}_{k-1}}\big)M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(\theta_{\widehat{\imath}_1}(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}),\ldots,\theta_{\widehat{\imath}_{k-1}}(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}))\\ &\quad= \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j) - \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}}(Z_j) + \lambda_{k-1}\theta_j(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})\,,\end{aligned}$$

Namely

$$\Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z) - \lambda_{k-1}\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}) = \Pi_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2}}(Z) - \lambda_{k-1}\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-2})\,.$$

Using again (64) we get that

$$\begin{aligned} N^{(k-1)} &= Z - \Pi_{\widehat{\imath}_1,\dots,\widehat{\imath}_{k-2}}(Z) + \lambda_{k-1}\theta(\widehat{\imath}_1,\dots,\widehat{\imath}_{k-2})\,,\\ &= Z - \Pi_{\widehat{\imath}_1,\dots,\widehat{\imath}_{k-1}}(Z) + \lambda_{k-1}\theta(\widehat{\imath}_1,\dots,\widehat{\imath}_{k-1})\,, \end{aligned}$$

as claimed.

C.6. *Proof of Proposition 18*

We denote

$$\begin{aligned} R_j &:= \big(R_{j,i_1},\dots,R_{j,i_{k-1}}\big)\,,\\ R_{i_k} &:= \big(R_{i_k,i_1},\dots,R_{i_k,i_{k-1}}\big)\,,\\ M &:= M_{i_1,\dots,i_{k-1}}\,,\\ \overline{M} &:= M_{i_1,\dots,i_k} = \begin{bmatrix} M & R_{i_k} \\ R_{i_k}^\top & R_{i_k,i_k} \end{bmatrix},\\ \overline{R} &:= \big(R_{j,i_1},\dots,R_{j,i_k}\big)\,,\\ x &:= \frac{1-\theta_j(i_1,\dots,i_{k-1})}{1-\theta_{i_k}(i_1,\dots,i_{k-1})}\,\frac{\tau_{j,i_k}}{\tau_{i_k,i_k}}\,, \end{aligned}$$

and observe that

$$\begin{aligned} x &= \frac{R_{j,i_k} - R_j^\top M^{-1} R_{i_k}}{R_{i_k,i_k} - R_{i_k}^\top M^{-1} R_{i_k}}\,,\\ \overline{M}^{-1} &= \begin{bmatrix} \mathrm{Id}_{k-1} & -M^{-1}R_{i_k} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M^{-1} & 0 \\ 0 & \big(R_{i_k,i_k} - R_{i_k}^\top M^{-1} R_{i_k}\big)^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{Id}_{k-1} & 0 \\ -R_{i_k}^\top M^{-1} & 1 \end{bmatrix},\\ \overline{M}^{-1}\overline{R} &= \begin{bmatrix} M^{-1}\big(R_j - xR_{i_k}\big) \\ x \end{bmatrix}, \end{aligned} \tag{65}$$

using Schur complement of block M of the matrix $\overline{M}$ and a LU decomposition. Note also that

$$\frac{Z_j^{(i_1,\dots,i_{k-1})} - Z_{i_k}^{(i_1,\dots,i_{k-1})}\tau_{j,i_k}/\tau_{i_k,i_k}}{1-\tau_{j,i_k}/\tau_{i_k,i_k}} = \frac{Z_j - \Pi_{i_1,\dots,i_{k-1}}(Z_j) - x\,(Z_{i_k} - \Pi_{i_1,\dots,i_{k-1}}(Z_{i_k}))}{1-\theta_j(i_1,\dots,i_{k-1}) - x(1-\theta_{i_k}(i_1,\dots,i_{k-1}))}\,.$$

To prove (54), it suffices to show that the R.H.S term above is equal to the following R.H.S term

$$Z_j^{(i_1,\dots,i_k)} = \frac{Z_j - \Pi_{i_1,\dots,i_k}(Z_j)}{1-\theta_j(i_1,\dots,i_k)}\,.$$

We will prove that numerators are equal and that denominators are equal. For denominators,

$$\begin{aligned} &1-\theta_j(i_1,\dots,i_{k-1}) - x(1-\theta_{i_k}(i_1,\dots,i_{k-1}))\\ &= 1-\theta_j(i_1,\dots,i_{k-1}) - x + x\,\theta_{i_k}(i_1,\dots,i_{k-1})\\ &= 1 - (\underbrace{1\cdots 1}_{k\ \text{times}}) \begin{bmatrix} M^{-1}\big(R_j - xR_{i_k}\big) \\ x \end{bmatrix}\\ &= 1-\theta_j(i_1,\dots,i_k)\,, \end{aligned} \tag{66}$$

using (65). Furthermore, it proves (53). For the numerators, we use that

$$\begin{aligned} &Z_j - \Pi_{i_1,\dots,i_{k-1}}(Z_j) - x\,(Z_{i_k} - \Pi_{i_1,\dots,i_{k-1}}(Z_{i_k}))\\ &= Z_j - \Pi_{i_1,\dots,i_{k-1}}(Z_j) - xZ_{i_k} + x\Pi_{i_1,\dots,i_{k-1}}(Z_{i_k})\\ &= Z_j - (Z_{i_1}\cdots Z_{i_k}) \begin{bmatrix} M^{-1}\big(R_j - xR_{i_k}\big) \\ x \end{bmatrix}\\ &= Z_j - \Pi_{i_1,\dots,i_k}(Z_j)\,. \end{aligned}$$

using (65).

C.7. Proof of Theorem 15

We rely on the **Weak Positive Regression Dependency (WPRDS) property** to prove the result, one may consult [18, Page 173] for instance. We say that a function $g : [0,1]^K \to \mathbb{R}^+$ is *nondecreasing* if for any $p,q \in [0,1]^K$ such that $p_k \geq q_k$ for every $k = 1,\dots,K$, we have $g(p) \geq g(q)$. We say that a Borel set $\Gamma \in [0,1]^K$ is *nondecreasing* if $g = \mathbb{1}_\Gamma$ is nondecreasing. In other words if $y \in \gamma$ and if $z \geq 0$, then $y+z \in \gamma$. We say that the p-values $(\,\widehat{p}_1 = \widehat{\alpha}_{0,1,K+1},\dots,\widehat{p}_K = \widehat{\alpha}_{K-1,K,K+1})$ satisfy the WPRDS property if for any nondecreasing set Γ and for all $k^0 \in I_0$, the function

$$u \mapsto \mathbb{P}_{\mu^0}\big[(\,\widehat{p}_1,\dots,\widehat{p}_K) \in \Gamma \big| \widehat{p}_{k^0} \leq u\big] \text{ is nondecreasing}$$

where $\mu^0 = \beta^0$ in our orthogonal design case, and we recall that

$$I_0 = \big\{k \in [K] \ : \ \mathbb{H}_{0,k} \text{ is true}\big\}\,.$$

To prove Theorem 15, note that it is sufficient [18, Chapter 8] to prove that

$$u \mapsto \overline{\mathbb{P}}\big[(\,\widehat{p}_1,\dots,\widehat{p}_K) \in \Gamma \big| \widehat{p}_{k^0} \leq u\big] \text{ is nondecreasing} \tag{67}$$

where $\overline{\mathbb{E}}, \overline{\mathbb{P}}$ will denote that expectations and probabilities are conditional on $\{\bar{\imath}_1,\dots,\bar{\imath}_K,\lambda_{K+1}\}$ and under the hypothesis that $\mu^0 = X^\top X \beta^0$. Note that one can integrate in λ_{K+1} to get the statement of Theorem 15.

$\circ$ **Step 1:** We start by giving the joint law of the LAR's knots under the alternative in the orthogonal design case. Lemma 19 and (62) show that, conditional on $\{\bar{\imath}_1,\dots,\bar{\imath}_K,\lambda_{K+1}\}$, $(\lambda_1,\dots,\lambda_K)$ is distributed on the set $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K \geq \lambda_{K+1}$ and it has a coordinate-wise nondecreasing density. Now we can assume without loss of generality that $\sigma^2 = 1$, in addition because of orthogonality $\rho_k^2 = 1$ implying that $F_k = \Phi(\lambda_k)$ $\mathcal{P}_{i,j} = \Phi_i \circ \Phi_j^{-1} = \mathrm{Id}$. We deduce that, conditional on $\{\bar{\imath}_1,\dots,\bar{\imath}_K,F_{K+1}\}$, $(F_1,\dots,F_K)$ is distributed on the set

$$\big\{(f_1,\dots,f_K) \in \mathbb{R}^K \ : \ 1 \geq f_1 \geq f_2 \geq \dots \geq f_K \geq F_{K+1})\big\}\,,$$

it has an **explicit density** given by (62), and we denote it by h_{μ^0}. By the change of variables $G_k := \frac{F_k - F_{K+1}}{F_{k-1} - F_{K+1}}$ one obtains that the distribution of $(G_1,\dots,G_K)$ is supported on $[0,1]^K$. More precisely, define

$$\begin{aligned}
\psi(f_1,\dots,f_K) &:= (g_1,\dots,g_K) := (\frac{f_1 - F_{K+1}}{1 - F_{K+1}},\dots,\frac{f_K - F_{K+1}}{f_{K-1} - F_{K+1}}) \\
\psi^{-1}(g_1,\dots,g_K) &:= \Big((1-F_{K+1})g_1 + F_{K+1},\dots,(1-F_{K+1})g_1 g_2 \dots g_K + F_{K+1}\Big),
\end{aligned}$$

whose inverse Jacobian determinant is

$$\det\Big[\frac{\partial\psi}{\partial f_1} \cdots \frac{\partial\psi}{\partial f_K}\Big]^{-1} = \prod_{k=1}^K (f_{k-1} - F_{K+1}) = (1-F_{K+1})^K \prod_{k=1}^K g_k^{K-k}\,.$$

We deduce that the density of $(G_1,\dots,G_K)|\{\bar{\imath}_1,\dots,\bar{\imath}_K,F_{K+1}\}$ at point g with respect to Lebesgue measure is

$$\mathbf{p}(g) := (\text{const})\mathbb{1}_{g\in(0,1)^K} \prod_{k=1}^K g_k^{K-k} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\prod_{\ell=1}^k g_\ell + F_{K+1})\big]\,, \tag{68}$$

where we have used (62). From (38) and (59), one has

$$\widehat{p}_k = 1 - \boldsymbol{F}_{\beta(1,K-k+1)}\bigg(\frac{F_k - F_{K+1}}{F_{k-1} - F_{K+1}}\bigg) = 1 - \boldsymbol{F}_{\beta(1,K-k+1)}(G_k) \tag{69}$$

where $\boldsymbol{F}_\beta$ is the cumulative distribution of the Beta distribution in reference. We deduce that for any $v \in (0,1)$ and for any $\ell \in [K]$,

$$\widehat{p}_k = v \Leftrightarrow (G_1, \ldots, G_K) \in [0,1]^K \cap \{\boldsymbol{F}^{-1}_{\beta(1,K-k+1)}(1-v) = G_k\}\,,$$

so that

$$\overline{\mathbb{P}}\big[(\widehat{p}_1, \ldots, \widehat{p}_K) \in \Gamma \big| \widehat{p}_{k^0} \leq u\big] = \overline{\mathbb{P}}\big[(G_1, \ldots, G_K) \in \overline{\Gamma} \big| G_{k^0} \geq \boldsymbol{F}^{-1}_{\beta(1,K-\ell+1)}(1-u)\big]\,, \tag{70}$$

where $\overline{\Gamma}$ can be proved to be a **nonincreasing Borel set** from (69).

◦ **Step 2:** Let $0 < x < y < 1$ and denote by μ_x the following conditional law

$$\mu_x := \mathrm{law}\big[(G_1, \ldots, G_K) | \{\bar{\imath}_1, \ldots, \bar{\imath}_K, F_{K+1}, G_{k^0} \geq x\}\big]\,.$$

Remark that if there exists a measurable $T : [0,1]^K \mapsto [0,1]^K$ such that

- T is nondecreasing, meaning that for any $g \in [0,1]^K$, $T(g) \geq g$;
- T is such that push-forward of μ_x by T gives μ_y, namely $T_\#\mu_x = \mu_y$;

then it holds

- $\mathbb{1}_{\{T(g) \in \overline{\Gamma}\}} \leq \mathbb{1}_{\{g \in \overline{\Gamma}\}}$;
- $\mathrm{law}\big[T(G) | \{\bar{\imath}_1, \ldots, \bar{\imath}_K, F_{K+1}, G_{k^0} \geq x\}\big] = \mathrm{law}\big[G | \{\bar{\imath}_1, \ldots, \bar{\imath}_K, F_{K+1}, G_{k^0} \geq y\}\big]$ where $G = (G_1, \ldots, G_K)$.

In this case, we deduce that

$$\overline{\mathbb{P}}\big[G \in \overline{\Gamma} \big| G_{k^0} \geq x\big] \geq \overline{\mathbb{P}}\big[T(G) \in \overline{\Gamma} \big| G_{k^0} \geq x\big] = \overline{\mathbb{P}}\big[G \in \overline{\Gamma} \big| G_{k^0} \geq y\big]\,.$$

If one can prove that such function T exists for any $0 < x < y < 1$, it proves that

$$x \mapsto \overline{\mathbb{P}}\big[G \in \overline{\Gamma} \big| G_{k^0} \geq x\big] \text{ is nonincreasing}\,,$$

and, in view of (70), it proves (67). Proving that such function T exists is done in the next step.

◦ **Step 3:** Let $0 < x < y < 1$. Consider the **Knothe-Rosenblatt transport map** T of μ_x toward μ_y following the order

$$k^0 \to k^0 + 1 \to \cdots \to K \to k^0 - 1 \to k^0 - 2 \to \cdots \to 1\,.$$

It is based on a sequence of conditional quantile transforms defined following the ordering above. Its construction is presented for instance in [24, Sec.2.3, P.67] or [31, P.20]. The transport T is defined as follows. Given $z, z' \in [0,1]^K$ such that $z' = T(z)$ it holds

$$\begin{aligned}
z'_{k^0} &= T^{(k^0)}(z_{k^0});\\
z'_{k^0+1} &= T^{(k^0+1)}(z_{k^0+1}, z'_{k^0});\\
&\vdots\\
z'_K &= T^{(K)}(z_K, z'_{K-1}, \ldots, z'_{k^0});\\
z'_{k^0-1} &= T^{(k^0-1)}(z_{k^0-1}, z'_K, \ldots, z'_{k^0});\\
&\vdots\\
z'_1 &= T^{(1)}(z_1, z'_2, \ldots, z'_{k^0-1}, z'_K, \ldots, z'_{k^0});
\end{aligned}$$

where $T^{(k^0)}, T^{(k^0+1)}, \ldots, T^{(K)}, T^{(k^0-1)} \ldots, T^{(1)}$ will be build in the sequel, in which we will drop their dependencies in the z'_k's to ease notations. It remains to prove that

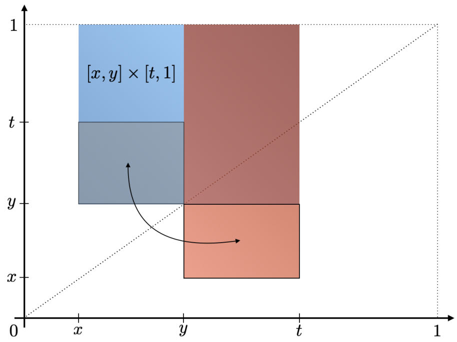

Figure 6: Note that, by symmetry the two boxed regions have same $\mathbf{p}\otimes\mathbf{p}$ measure. The blue region is $[x,t]\times[y,1]$, its measure is the measure of the red region (namely $cD_y(t)\times\mathcal{D}_x(1)$) more the bluest upper left corner (namely $[x,y]\times[t,1]$).

- T is nondecreasing, meaning that for any $g\in[0,1]^K$, $T(g)\geq g$;
- T is such that push-forward of μ_x by T gives μ_y, namely $T_\#\mu_x=\mu_y$;

to conclude. The last point is a property of the Knothe-Rosenblatt transport map. Proving the first point will be done in the rest of the proof.

○ *Step 3.1:* We start by the first transport map $T^{(k^0)}:[0,1]\mapsto[0,1]$. Denote $\mu_x^{(k^0)}$ the following conditional law

$$\mu_x^{(k^0)}:=\mathrm{law}\big[G_{k^0}|\{\bar\imath_1,\ldots,\bar\imath_K,F_{K+1},G_{k^0}\geq x\}\big]\,,$$

and $\mathbb{F}_x^{(k^0)}$ its cdf. Note that the Knothe-Rosenblatt construction gives $T^{(k^0)}=(\mathbb{F}_y^{(k^0)})^{-1}\circ\mathbb{F}_x^{(k^0)}$. We would like to prove that $T^{(k^0)}(t)\geq t$ for all $z\in(0,1)$. This is equivalent to prove that it holds $\mathbb{F}_x^{(k^0)}\geq\mathbb{F}_y^{(k^0)}$. For $t\leq y$, $\mathbb{F}_y^{(k^0)}(t)=0$ and it implies that $\mathbb{F}_x^{(k^0)}(t)\geq\mathbb{F}_y^{(k^0)}(t)$. Let $t>y$, using the conditional density $\mathbf{p}$ defined in (68), note that

$$\begin{aligned}\mathbb{F}_x^{(k^0)}(t)\geq\mathbb{F}_y^{(k^0)}(t)\quad&\Leftrightarrow\quad\frac{\int_x^t\mathbf{p}}{\int_x^1\mathbf{p}}\geq\frac{\int_y^t\mathbf{p}}{\int_y^1\mathbf{p}}\\&\Leftrightarrow\quad\int_x^t\int_y^1\mathbf{p}\otimes\mathbf{p}\geq\int_y^t\int_x^1\mathbf{p}\otimes\mathbf{p},\end{aligned}$$

where, for example

$$\int_x^t\text{ means the integral over the hyper rectangle }[x,t]:=\Big\{(g_1,\ldots,g_K)\in[0,1]^K\ :\ x\leq g_{k^0}\leq t\Big\}.$$

A simple calculation (see also Figure 6) gives that

$$\int_x^t\int_y^1\mathbf{p}\otimes\mathbf{p}=\int_y^t\int_x^1\mathbf{p}\otimes\mathbf{p}+\int_{[x,y]\times[t,1]}\mathbf{p}\otimes\mathbf{p}\,,$$

and it proves that $\mathbb{F}_x^{(k^0)}\geq\mathbb{F}_y^{(k^0)}$.

○ *Step 3.2:* We continue with the second transport map in Knothe-Rosenblatt construction. Let $z_{k^0}\in(x,1)$ and denote $\mu_{z_{k^0}}^{(k^0+1)}$ the following conditional law

$$\mu_{z_{k^0}}^{(k^0+1)}:=\mathrm{law}\big[G_{k^0+1}|\{\bar\imath_1,\ldots,\bar\imath_K,F_{K+1},G_{k^0}=z_{k^0}\}\big]\,,$$

and $\mathbb{F}^{(k^0+1)}_{z_{k^0}}$ its cdf. Let $z'_{k^0} := T^{(k^0)}(z'_{k^0})$ and denote $\mu^{(k^0+1)}_{z'_{k^0}}$ the following conditional law

$$\mu^{(k^0+1)}_{z'_{k^0}} := \mathrm{law}\big[G_{k^0+1}|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}=z'_{k^0}\}\big]\,,$$

and $\mathbb{F}^{(k^0+1)}_{z'_{k^0}}$ its cdf. Note that $x<z_{k^0}\leq z'_{k^0}=T^{(k^0)}(z_{k^0})\leq 1$. Again, we would like to prove that $\mathbb{F}^{(k^0+1)}_{z_{k^0}}\geq \mathbb{F}^{(k^0+1)}_{z'_{k^0}}$ which implies that the transport map $T^{(k^0+1)} := \big(\mathbb{F}^{(k^0+1)}_{z'_{k^0}}\big)^{-1}\circ\mathbb{F}^{(k^0+1)}_{z_{k^0}}$ satisfies $T^{(k^0+1)}(u)\geq u$ for all $u\in(0,1)$.

Recall that the conditional density $\mathbf{p}$ of $G|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1}\}$ is given by (68) and recall that $k^0\in I_0$. Observe that $\mu^0_{k^0}=0$, so that the conditional density of $G|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}=\mathbf{z}\}$ is

$$\begin{aligned}(\mathrm{const})\,\mathbb{1}_{g\in(0,1)^K}\mathbb{1}_{g_{k^0}=\mathbf{z}}\prod_{k<k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\prod_{\ell=1}^k g_\ell+F_{K+1}\big)\Big]&\\ \times\prod_{k>k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\,\mathbf{z}\prod_{1\leq\ell\neq k^0\leq k}g_\ell+F_{K+1}\big)\Big]&\,.\end{aligned}\tag{71}$$

Set $\tau := z'_{k^0}/z_{k^0}\geq 1$ and $G'_{k^0+1}=\tau G_{k^0+1}$ so that

$$z_{k^0}G_{k^0+1}=z'_{k^0}G'_{k^0+1}\,.$$

Denote $G' := (G_1,\dots,G_{k^0-1},G'_{k^0+1},G_{k^0+2},\dots,G_K)\in(0,1)^{k^0-1}\times(0,\tau)\times(0,1)^{K-k^0-1}$ and note that the conditional density of $G'|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}=\tau\mathbf{z}\}$ is

$$\begin{aligned}(\mathrm{const})\,\mathbb{1}_{g\in(0,1)^{k^0-1}\times(0,\tau)\times(0,1)^{K-k^0-1}}\prod_{k<k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\prod_{\ell=1}^k g_\ell+F_{K+1}\big)\Big]&\\ \times\prod_{k>k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\,\mathbf{z}\prod_{1\leq\ell\neq k^0\leq k}g_\ell+F_{K+1}\big)\Big]&\,,\end{aligned}$$

which, up to some normalising constant, is the same as (71) up to the following change of support

$$\mathbb{1}_{g\in(0,1)^K}\leftrightarrow\mathbb{1}_{g'\in(0,1)^{k^0-1}\times(0,\tau)\times(0,1)^{K-k^0-1}}\,.$$

By an abuse of notation, we denote by $\mathbf{p}$ this function, namely

$$\begin{aligned}\mathbf{p}(g)=\prod_{k<k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\prod_{\ell=1}^k g_\ell+F_{K+1}\big)\Big]&\\ \times\prod_{k>k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\,\mathbf{z}\prod_{1\leq\ell\neq k^0\leq k}g_\ell+F_{K+1}\big)\Big]&\,.\end{aligned}$$

We deduce that

$$\begin{aligned}\mathbb{F}^{(k^0+1)}_{z_{k^0}}(t)\geq\mathbb{F}^{(k^0+1)}_{z'_{k^0}}(t)\quad&\Leftrightarrow\quad\overline{\mathbb{P}}(G_{k^0+1}\leq t|G_{k^0}=z_{k^0})\geq\overline{\mathbb{P}}(G_{k^0+1}\leq t|G_{k^0}=z'_{k^0})\\ &\Leftrightarrow\quad\overline{\mathbb{P}}(G_{k^0+1}\leq t|G_{k^0}=z_{k^0})\geq\overline{\mathbb{P}}(G'_{k^0+1}\leq\tau t|G_{k^0}=\tau z_{k^0})\\ &\Leftrightarrow\quad\frac{\int_{\mathcal{D}(t)}\mathbf{p}}{\int_{\mathcal{D}(1)}\mathbf{p}}\geq\frac{\int_{\mathcal{D}(\tau t)}\mathbf{p}}{\int_{\mathcal{D}(\tau)}\mathbf{p}}\\ &\Leftrightarrow\quad\int_{\mathcal{D}(t)\times\mathcal{D}(\tau)}\mathbf{p}\otimes\mathbf{p}\geq\int_{\mathcal{D}(\tau t)\times\mathcal{D}(1)}\mathbf{p}\otimes\mathbf{p}\,,\end{aligned}\tag{72}$$

where

$$\mathcal{D}(s) := \Big\{(g_1,\dots,g_{k^0-1},g_{k^0+1}\dots,g_K)\in(0,1)^{K-1}\ :\ 0<g_{k^0+1}\leq s\Big\}\,.$$

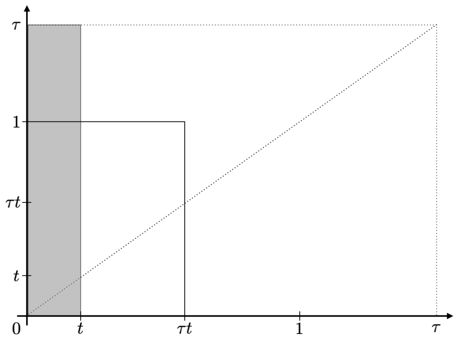

Figure 7: The two boxed rectangles have Lebesgue measure, namely τt. The $\mathbf{p} \otimes \mathbf{p}$ measure of the grey box is greater than the $\mathbf{p} \otimes \mathbf{p}$ measure of the white box.

We now present an inequality on the to conclude. Observe that we are integrating on domains depicted in Figure 7. The two boxes have same area for the uniform measure and we would like to compare their respective measure for the $\mathbf{p} \otimes \mathbf{p}$ measure. We start by the next lemma whose proof is omitted.

Lemma 21. *Let $a, b \geq 0$. The function*

$$z \mapsto \cosh(a\, z + b) \times \cosh(a/z + b)$$

is non-decreasing on the domain $[1, \infty)$.

Now, let $(g_1, \ldots, g_{k^0-1}, g_{k^0+2} \ldots, g_K) \in (0,1)^{K-1}$ be fixed in the integrals (72). We are the looking at the weights of the domains $(h_1, h_2) \in (0,t) \times (0,\tau)$ and $(h_3, h_4) \in (0,\tau t) \times (0,1)$ for the weight function w given by

$$\begin{aligned}
w(h_1,h_2) =& \mathrm{C}_1 h_1^{K-k^0-1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell<k^0} g_\ell \times h_1 + F_{K+1})\big] \\
&\times \prod_{k>k^0+1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell\neq k^0, k^0+1\leq k} g_\ell \times h_1 + F_{K+1})\big] \\
&\times h_2^{K-k^0-1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell<k^0} g_\ell \times h_2 + F_{K+1})\big] \\
&\times \prod_{k>k^0+1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell\neq k^0, k^0+1\leq k} g_\ell \times h_2 + F_{K+1})\big]\,,
\end{aligned}$$

where the constant C_1 depends on $(g_1, \ldots, g_{k^0-1}, g_{k^0+2} \ldots, g_K) \in (0,1)^{K-1}$. By the change of variables $h'_1 = h_3/t$ and $h'_2 = t h_4$, the right hand term of (72) is given by the integration on the domain $(h'_1, h'_2) \in (0,t) \times (0,\tau)$ of the weight function w' given by

$$\begin{aligned}
w'(h'_1,h'_2) =& \mathrm{C}_1 {h'_1}^{K-k^0-1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell<k^0} g_\ell \times t \times h'_1 + F_{K+1})\big] \\
&\times \prod_{k>k^0+1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell\neq k^0, k^0+1\leq k} g_\ell \times t \times h'_1 + F_{K+1})\big] \\
&\times {h'_2}^{K-k^0-1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell<k^0} g_\ell \times h'_2/t + F_{K+1})\big] \\
&\times \prod_{k>k^0+1} \cosh\big[\mu^0_{\bar{\imath}_k}((1-F_{K+1})\ z_{k^0} \prod_{1\leq \ell\neq k^0, k^0+1\leq k} g_\ell \times h'_2/t + F_{K+1})\big]\,.
\end{aligned}$$

Now, invoke Lemma 21 with

$$\begin{aligned}
a &= \mu^0_{\bar\imath_k}(1-F_{K+1})\; z_{k^0} \prod_{1\le \ell<k^0} g_\ell \times h \\
b &= \mu^0_{\bar\imath_k} F_{K+1} \\
z &= t \ge 1\,,
\end{aligned}$$

where $h = h_1$ or h_2, to get that $w' \ge w$ and so

$$\int_{\mathcal{D}(t)\times\mathcal{D}(\tau)} \mathbf{p}\otimes\mathbf{p} \ge \int_{\mathcal{D}(\tau t)\times\mathcal{D}(1)} \mathbf{p}\otimes\mathbf{p}\,,$$

which concludes this part of the proof.

$\circ$ *Step 3.3:* We continue by induction with the other transport maps in Knothe-Rosenblatt's construction. Assume that we have built $z' := (z'_k, \ldots, z'_{k^0})$ and $z := (z_k, \ldots, z_{k^0})$ for some $k > k^0$. Denote $\mu^{(k+1)}_z$ the following conditional law

$$\mu^{(k+1)}_z := \mathrm{law}\big[G_{k+1}|\{\bar\imath_1,\ldots,\bar\imath_K, F_{K+1}, \underbrace{G_k = z_k,\ldots,G_{k^0} = z_{k^0}}_{\text{denoted } G^{[k,k^0]}=z}\}\big]\,,$$

and $\mathbb{F}^{(k+1)}_z$ its cdf. Denote $\mu^{(k+1)}_{z'}$ the following conditional law

$$\mu^{(k+1)}_{z'} := \mathrm{law}\big[G_{k+1}|\{\bar\imath_1,\ldots,\bar\imath_K, F_{K+1}, \underbrace{G_k = z'_k,\ldots,G_{k^0} = z'_{k^0}}_{G^{[k,k^0]}=z'}\}\big]\,,$$

and $\mathbb{F}^{(k+1)}_{z'}$ its cdf. Note that $z \le z' = T^{(k)}(z) \le 1$. Again, we would prove that $\mathbb{F}^{(k+1)}_z \ge \mathbb{F}^{(k+1)}_{z'}$ which implies that the transport map $T^{(k+1)} := \big(\mathbb{F}^{(k+1)}_{z'}\big)^{-1}\circ\mathbb{F}^{(k+1)}_z$ satisfies $T^{(k+1)}(u) \ge u$ for all $u \in (0,1)$.

For $\mathbf{z} \in (0,1)^{k-k^0} \times (x,1)$, the conditional density of $G|\{\bar\imath_1,\ldots,\bar\imath_K, F_{K+1}, G^{[k,k^0]} = \mathbf{z}\}$ is

$$\begin{aligned}
&(\mathrm{const})\, \mathbb{1}_{g\in(0,1)^K} \mathbb{1}_{g^{[k,k^0]}=\mathbf{z}} \prod_{m<k^0} g_m^{K-m} \cosh\Big[\mu^0_{\bar\imath_m}\big((1-F_{K+1})\prod_{\ell=1}^{m} g_\ell + F_{K+1}\big)\Big] \\
&\quad\times \prod_{k^0\le m\le k} \mathbf{z}_m^{K-m} \cosh\Big[\mu^0_{\bar\imath_m}\big((1-F_{K+1}) \prod_{1\le\ell<k^0} g_\ell \prod_{n=k^0}^{m} \mathbf{z}_n + F_{K+1}\big)\Big] \\
&\quad\times \prod_{k<m} g_m^{K-m} \cosh\Big[\mu^0_{\bar\imath_m}\big((1-F_{K+1}) \prod_{1\le\ell<k^0} g_\ell \prod_{n=k^0}^{k} \mathbf{z}_n \prod_{k<\ell\le m} g_\ell + F_{K+1}\big)\Big]\,.
\end{aligned}$$

Set $\tau := \prod_{n=k^0}^{k} z'_n / \prod_{n=k^0}^{k} z_n \ge 1$ and $G'_k = \tau G_{k^0+1}$ so that

$$\Big[\prod_{n=k^0}^{k} z'_n\Big] G_{k+1} = \Big[\prod_{n=k^0}^{k} z_n\Big] G'_{k+1}\,.$$

Then the proof follows the same idea as in *Step 3.2* and we will not detail it here.

$\circ$ *Step 3.4:* This is the last step of the proof. Assume that we have built $z' := (z'_K, \ldots, z'_{k^0})$ and $z := (z_K, \ldots, z_{k^0})$. Denote $\mu^{(k^0-1)}_z$ the following conditional law

$$\mu^{(k^0-1)}_z := \mathrm{law}\big[G_{k^0-1}|\{\bar\imath_1,\ldots,\bar\imath_K, F_{K+1}, G^{[K,k^0]} = z\}\big]\,,$$

and $\mathbb{F}_z^{(k^0-1)}$ its cdf. Denote $\mu_{z'}^{(k^0-1)}$ the following conditional law

$$\mu_{z'}^{(k^0-1)} := \mathrm{law}\big[G_{k^0-1}|\{\bar{\imath}_1,\ldots,\bar{\imath}_K,F_{K+1},G^{[K,k^0]}=z'\}\big]\,,$$

and $\mathbb{F}_{z'}^{(k^0 1)}$ its cdf. Note that $z \le z' = T^{(K)}(z) \le 1$. Again, we would prove that $\mathbb{F}_z^{(k^0-1)} \ge \mathbb{F}_{z'}^{(k^0-1)}$ which implies that the transport map $T^{(k^0-1)} := \big(\mathbb{F}_{z'}^{(k^0-1)}\big)^{-1} \circ \mathbb{F}_z^{(k^0-1)}$ satisfies $T^{(k^0-1)}(u) \ge u$ for all $u \in (0,1)$.

For $\mathbf{z} \in (0,1)^{K-k^0} \times (x,1)$, the conditional density of $G|\{\bar{\imath}_1,\ldots,\bar{\imath}_K,F_{K+1},G^{[K,k^0]}=\mathbf{z}\}$ is

$$\begin{aligned}(\mathrm{const})\,\mathbb{1}_{g\in(0,1)^K}\mathbb{1}_{g^{[K,k^0]}=\mathbf{z}}\prod_{m<k^0} g_m^{K-m}\cosh\big[\mu^0_{\bar{\imath}_m}((1-F_{K+1})\prod_{\ell=1}^m g_\ell + F_{K+1})\big]\\ \times\prod_{k^0\le m\le K}\mathbf{z}_m^{K-m}\cosh\big[\mu^0_{\bar{\imath}_m}((1-F_{K+1})\prod_{1\le\ell<k^0} g_\ell\prod_{n=k^0}^m \mathbf{z}_n + F_{K+1})\big]\,.\end{aligned}$$

Now, let $(g_1,\ldots,g_{k^0-2}) \in (0,1)^{k^0-2}$ be fixed and denote by

$$\begin{aligned}\forall g\in(0,1),\quad w_z(g) :=& g^{K-k^0+1}\cosh\big[\mu^0_{\bar{\imath}_{k^0-1}}((1-F_{K+1})\prod_{\ell=1}^{k^0-2} g_\ell\times g + F_{K+1})\big]\\ &\times\prod_{k^0\le m\le K}\mathbf{z}_m^{K-m}\cosh\big[\mu^0_{\bar{\imath}_m}((1-F_{K+1})\prod_{n=k^0}^m \mathbf{z}_n\prod_{\ell=1}^{k^0-2} g_\ell\times g + F_{K+1})\big]\,.\end{aligned}$$

and, substituting z by z', define $w_{z'}$ as well. Let $t \in (0,1)$. Following the idea of *Step 3.2*, one can check that it is sufficient to prove that

$$\int_0^t\Big(\int_0^1 w_z(g)w_{z'}(g')\mathrm{d}g'\Big)\mathrm{d}g \ge \int_0^1\Big(\int_0^t w_z(g)w_{z'}(g')\mathrm{d}g'\Big)\mathrm{d}g\,.$$

Substituting

$$\int_0^t\Big(\int_0^t w_z(g)w_{z'}(g')\mathrm{d}g'\Big)\mathrm{d}g$$

on both parts, one is reduced to prove that

$$\int_0^t\Big(\int_t^1 w_z(g)w_{z'}(g')\mathrm{d}g'\Big)\mathrm{d}g \ge \int_0^t\Big(\int_t^1 w_{z'}(g)w_z(g')\mathrm{d}g'\Big)\mathrm{d}g\,.$$

Observe that $g \le g'$ in the last two integrals. Now, we have this lemma whose proof is omitted.

Lemma 22. *Let $0 < a \le a'$ and $b > 0$. The function*

$$z \mapsto \frac{\cosh(a\,z+b)}{\cosh(a'z+b)}$$

is non-increasing on the domain $(0,\infty)$.

Let $g \le g'$. From Lemma 22, we deduce that $\cosh(ag+b)\cosh(a'g'+b) \ge \cosh(ag'+b)\cosh(a'g+b)$, proving that $w_z(g)w_{z'}(g') \ge w_{z'}(g)w_z(g')$. It proves that $T^{(k^0-1)}(u) \ge u$ for all $u \in (0,1)$.

We then proceed by induction for $k^0-1 \to k^0-2 \to \cdots \to 1$. The proof follows the same line as above, *Step 3.4*.

Appendix D: A Quasi Monte Carlo (QMC) method: Cubature by lattice rule

Our goal is to compute the integral of some function$f$ on the hypercube of dimension d, namely

$$I := \int_{[0,1]^d} f(x)dx.$$

We want to approximate it by a finite sum over n points

$$I_n := \frac{1}{n}\sum_{i=1}^{n} f(x^{(i)}).$$

A convenient way of constructing the sequence $x^{(i)}, i = 1, \ldots, n$ is the so-called *lattice rule*: from the first point $x^{(1)}$ we deduce the others $x^{(i)}$ by

$$x^{(i)} = \{i.x^{(1)}\},$$

where the $\{\}$ brackets mean that we take the fractional part coordinate by coordinate. In such a case the error given by

$$E(f, n, x^{(1)}) = I - I_n$$

is a function, in particular, of starting point $x^{(1)}$.

The Fast-rank algorithm [21] is a fast algorithm that finds, component by component and as a function of the prime n, the sequence of coordinates of $x^{(1)}$ that minimizes the maximal error when f varies in a unit ball $\mathcal{E}$ of some *RKHS*, namely a tensorial product of *Koborov spaces*. In addition it gives an expression of its minimax error, namely

$$\max_{f\in\mathcal{E}}(f, n, x^{(1)}).$$

In practice, very few properties are known on the function f, so the result above is not directly applicable. Nevertheless for many functions f, it happens that the convergence of I_n to I is "*fast*": typically of the order $1/n$ while the Monte-Carlo method (choosing the $x^{(i)}$ *at random*) converges at rate $1/\sqrt{n}$.

A reliable estimate of the error is obtained by adding a *Monte-Carlo layer* as in [16] for instance. This can be done as follows. Let U a **unique** uniform variable on $[0, 1]^d$, we define

$$x_U^{(i)} := \{i.x^{(1)} + U\}, \quad I_{n,U} := \frac{1}{n}\sum_{i=1}^{n} f(x_U^{(i)}).$$

Classical computations show that $I_{n,U}$ is now an unbiased estimator of I. In a final step, we perform N (in practice 15-20) independent repetitions of the experiment above an we compute usual asymptotic confidence intervals for independent observations.

Appendix E: Main notation

Please see next page.

NOTATION	COMMENT
	General notation
$[a]$	the set of integers $\{1, ..., a\}$
$Y = X\beta^0 + \eta$	Linear Model (2), X is $n \times p$ design matrix with rank r
σ^2	the variance of the errors η
K	the number of knots $\lambda_1, \dots, \lambda_K$ that are considered, see (10)
n_1,n_2	number of d.o.f. used for constructing $\widehat{\sigma}_{\text{select}}, \widehat{\sigma}_{\text{test}}$
φ_{m_k, v_k^2}	standard Gaussian density with mean m_k and variance $v_k^2 := \sigma^2 \rho_k^2$
$\widetilde{\varphi}$	multivariate t-distribution with $\nu = n_2$ degrees of freedom, mean $m = (m_1, \dots, m_K)$ and variance-covariance matrix $\mathrm{Diag}(\rho_1, \dots, \rho_K)$
m_k, v_k^2	conditional mean, see (28), and conditional variance $v_k^2 = \sigma^2 \rho_k^2$, see (29)
$\widehat{\alpha}_{abc}$	the p-value of the *generalized spacing test* (GST), see (38)
$\mathcal{S}_{abc}$	$\mathbb{1}_{\{\widehat{\alpha}_{abc} \leq \alpha\}}$, the *generalized spacing test* (GST) see (39)
Λ_k	t-knots defined by (40)
$\widehat{\beta}_{abc}$	the p-value of the *generalized t-spacing test* (GtST), see (42)
$\mathcal{T}_{abc}$	$\mathbb{1}_{\{\widehat{\beta}_{abc} \leq \alpha\}}$, the *generalized t-spacing test* (GtST), see (44)
	Technical notation
$\widehat{\imath}_k$	a way of coding both indexes and signs, see (3)
$\bar{\imath}_k; \varepsilon_k$	the indexes and the signs of the variables that enter in the LAR path
$j_1, \dots, j_k$; $s_1, \dots, s_k$	a generic value of the sequences above
$i_1, \dots, i_k$	a generic value of the sequence $\widehat{\imath}_k$
Z	the vector of correlations, obtained by symmetry from $\overline{Z}$ defined by (4)
R	the variance-covariance matrix of Z, see (6)
$M_{i_1,\dots,i_\ell}$	sub-matrix of R indexed by $\{i_1, \dots, i_\ell\}$, see (12)
$\overline{S}^k$	$\{\bar{\imath}_1, \dots, \bar{\imath}_k\}$, a possible selected support (8)
S_0	the true support
$\widehat{S}$	the chosen set of variables : $\overline{S}^{\widehat{m}}$
$\widehat{m}$	the chosen size
($\mathcal{A}_{\text{Stop}}$)	stopping rule, see Section 2.4
H_k	$\mathrm{Span}(X_{\bar{\imath}_1}, \dots, X_{\bar{\imath}_k})$
$P_k(P_k^\perp)$	Orthogonal projector on (the orthogonal of) H_k
($\mathcal{A}_{\text{Irr.}}$)	Irrepresentable Check, see ($\mathcal{A}_{\text{Irr.}}$)
$\theta_j(i_1, \dots, i_k)$	see (11), and $\theta^\ell := \theta(\widehat{\imath}_1, \dots, \widehat{\imath}_\ell)$
$Z_j^{(i_1,\dots,i_k)}$	see (17)
$\Pi_{i_1,\dots,i_k}(Z_j)$	see (18)
$\lambda_k^f := Z_{i_k}^{i_1,\dots i_{k-1}}$	the k-th frozen knot, see (21)
$m_k^f, \sigma\rho_k^f$	mean (23) and standard deviation (24) of λ_k^f
$\mathbb{F}_{abc}(t)$	see (35)
$F_i; \mathcal{P}_{ij}$	$F_i := \Phi_i(\lambda_i) := \Phi(\lambda_i/(\sigma\rho_i))$ and $\mathcal{P}_{ij}$ is given by (32)
$\widetilde{\mathbb{F}}_{abc}(t)$	see (41)
$\boldsymbol{T}_k$	see (45)

Table 1
Table of notations (and commands).