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

[⋆] *CERMICS*
École des Ponts ParisTech, 77455 Marne la Vallée, France

Abstract: In this article we investigate the outcomes of the standard Least Angle Regression (LAR) algorithm in high dimensions under the Gaussian noise assumption. We give the exact law of the sequence of knots conditional on the sequence of variables entering the model, i.e., the "post-selection" law of the knots of the LAR. Based on this result, we prove an exact of the False Discovery Rate (FDR) in the orthogonal design case and an exact control of the existence of false negatives in the general design case.

First, we build a sequence of testing procedures on the variables entering the model and we give an exact control of the FDR in the orthogonal design case when the noise level can be unknown. Second, we introduce a new exact testing procedure on the existence of false negatives when the noise level can be unknown. This testing procedure can be deployed after any support selection procedure that will produce an estimation of the support (i.e., the indexes of nonzero coefficients) for any designs. The type I error of the test can be exactly controlled as long as the selection procedure follows some elementary hypotheses, referred to as "admissible selection procedures". These support selection procedures are such that the estimation of the support is given by the k first variables entering the model where the random variable k is a stopping time. 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 June 27, 2019

1. Introduction

Parsimonious models have become an ubiquitous tool to tackle high-dimensional representations with a small budget of observations. Successful applications is 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)... have shown that the existence of an (almost) sparse representations in some well chosen basis is reasonable assumption in practice. 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, we 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 estimation produced, e.g., the false discovery rate is controlled or a confidence interval on linear statistics of the estimator can be given.

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] for instance. This paper focuses on the LAR algorithm and presents three useful equivalent formulations of the LAR algorithm, see Section A.1. As far as we known, the last formulation given by Algorithm 1, based on a recursive formulation, is new.

One specific task is to estimate the support of the target sparse vector, namely identify the true positives. In particular, one may take the support of a 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 to handle this issue, see Section 2.4. Our procedure does not require IC but a much weaker condition, referred to as the "*Empirical Irrepresentable Check*" (See Section 2.1.3), that can be checked in polynomial time.

Another recent issue is to control the *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 2.3. 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

This paper presents a class of new tests issued 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 two consecutive knots has been studied for the first time by [20] referred to as the *Spacing test* [29]. Later, the article [1] proved that this test is unbiased and introduce a *studentized* version of this test. On the same note, inference after model selection has been studied in several papers, as for instance [15, 25] or [26] for *selective inference* and a joint estimation of the noise level. This line of works studies a single test on a linear statistics while one may ask for a simultaneous control of several tests as in multiple testing frame. To the best of our knowledge, this paper is the first to study the joint law of multiple spacing tests of LAR's knots in a non-asymptotic frame, see Sections 2.2 and 2.3.

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 [18] of the problem of FDR control as the sample size tends to infinity using *debiased LASSO*.

Algorithm 1: LAR algorithm ("recursive" 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 $\varepsilon_1, \varepsilon_2, \ldots$ ($\varepsilon_k = \pm 1$).

`/* 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

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

`/* The following recursions are given by (24), (25) and (26). */`

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 as in (10).

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

1.3. Outline of the paper

In Section 2, we introduce our method along with the framework and the notion of "*Empirical Irrepresentable Check*" (EIC), a condition that can be checked in polynomial time, see Section 2.1.3. We divided our contributions in three part: Section 2.2 presents the joint law of the knots of LAR under EIC; Section 2.3 gives an exact control of FDR in the *orthogonal design case*; and Section 2.4 presents a general framework to exactly detect false negatives for general designs and some selection procedures referred to as "*admissible*".

Detailed mathematical statements are given in Section 3 including a method to "studentize" all the tests by an independent estimation of variance.

Section 4 presents practical applications including an illustration on real data.

The Appendix gives details on the different algorithms to compute LAR, proofs of the statements and a list of notation.

2. Exact Controls using Least Angle Regression, an introductory presentation

2.1. Notation, LAR formulations and Assumptions

2.1.1. Least Angle Regression (LAR)

Consider the linear model in high-dimensions where the number of observations n may be less than the number of predictors p. 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 2 and we present equivalent formulations in Algorithm 3 (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. We introduce some notation that we will be useful throughout this paper.

We denote by $(\widehat{\imath}_1, \dots, \widehat{\imath}_k) \in [2p]^k$ the "signed" variables that enter the model along the LAR path with the convention that

$$\widehat{\imath}_\ell := \bar{\imath}_\ell + p\left(\frac{1-\varepsilon_\ell}{2}\right), \tag{1}$$

so that $\widehat{\imath}_\ell \in [2p]$ is a useful way of encoding both the variable $\bar{\imath}_\ell \in [p]$ and its sign $\varepsilon_\ell = \pm 1$ as used in Algorithm 1. We denote by $\overline{Z}$ the vector such that $\overline{Z}_k$ is the scalar product between the k-th predictor and the response variable, and we denote by $\overline{R}$ its *correlation* matrix, see (9) and (10) for further details.

2.1.2. Nested Models

Our analysis is based conditionally to $(\widehat{\imath}_1, \dots, \widehat{\imath}_n)$ and in this spirit it can be referred to as a "Post-Section" procedure. The selected model $\widehat{S}$ would be chosen among the family of nested models

$$\underbrace{\{\bar{\imath}_1\}}_{\overline{S}^1} \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2\}}_{\overline{S}^2} \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2, \dots, \bar{\imath}_k\}}_{\overline{S}^k} \subset \dots \subset \underbrace{\{\bar{\imath}_1, \bar{\imath}_2, \dots, \bar{\imath}_n\}}_{\overline{S}^n}. \tag{2}$$

Respectively, denote

$$E_1 \subset \dots \subset E_k := \mathrm{Span}(X_{\bar{\imath}_1}, \dots, X_{\bar{\imath}_k}) \subset \dots \subset E_n = \mathbb{R}^n,$$

the corresponding family of nested subspaces of $\mathbb{R}^n$. In the sequel, denote by $P_k(Y)$ (resp. $P_k^\perp(Y)$) the orthogonal projection of the observation Y onto E_k (resp. the orthogonal of E_k) for all $k \geq 1$.

2.1.3. Empirical Irrepresentable Check

In the sequel we will build our testing statistics on the joint law of the first K knots of the LAR. A detailled discussion on the choice of K is given in Section 2.5, in particular K will be such that the so-called "*Empirical Irrepresentable Check of order K*" holds.

This property will be defined and studied in this section. 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].

Definition 1 (Irrepresentable Condition of order K)**.** *The design matrix X satisfies the Irrepresentable Condition of order K if and only if*

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

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

Remark 1. *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 ($\mathcal{A}_{\mathrm{Irr.}}$) (for instance the Identity matrix) satisfies ($\mathcal{A}_{\mathrm{Irr.}}$) 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 ($\mathcal{A}_{\mathrm{Irr.}}$) is a strong requirement on the design X and, additionally, this condition cannot be checked in polynomial time. One important feature of our result, see Theorem 3, is that we do not require Irrepresentable Condition but a much weaker requirement referred to as the "*Empirical Irrepresentable Check*" of order K.

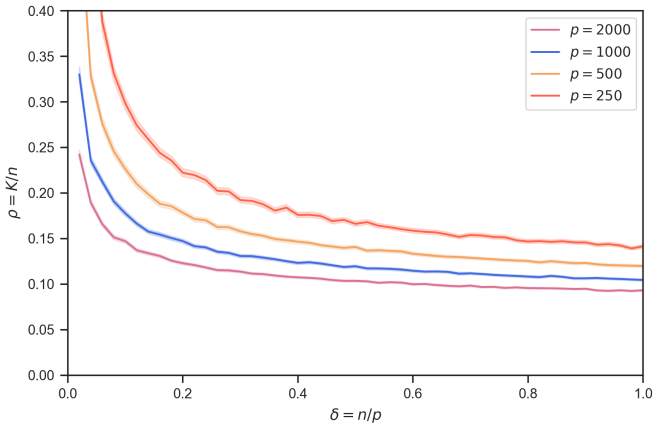

Figure 1. *Observed empirical "phase transition" over* $1,000$ *Monte-Carlo repetitions on the Empirical Irrepresentable Check of order* K *with* $\rho := K/n$ *and* $\delta := n/p$ *for different values of* p*. We considered a design* $X \in \mathbb{R}^{n\times p}$ *with independent column vectors uniformly distributed on the sphere and an independent* $y \in \mathbb{R}^n$ *with i.i.d. standard Gaussian entries, and we computed the variables* $(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)$ *entering the model with LAR. The plot represents the value* K *defined as the largest order for which* $(\widehat{\mathcal{A}}_{\text{Irr.}})$ *holds with respect to* $(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)$*.*

Definition 2 (Empirical Irrepresentable Check of order K w.r.t $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$)**.** *The design matrix* X *satisfies the Empirical Irrepresentable Check of order* K *with respect to* $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$ *if and only if for all the supports* $(\overline{S}^k)_{k\in[K]}$ *and all the signs* $(\varepsilon^k)_{k\in[K]}$ *which enter the model at steps* $k \in [K]$

$$\forall k \in [K]\,,\ \forall j \notin \overline{S}^k := \{\bar{\imath}_1, \ldots, \bar{\imath}_k\}, \quad X_j^\top X_{\overline{S}^k}\big(X_{\overline{S}^k}^\top X_{\overline{S}^k}\big)^{-1}\varepsilon^k < 1 \qquad (\widehat{\mathcal{A}}_{\text{Irr.}})$$

where $\overline{S}^k$ *is the selected support at step* k *(see (2)) and* $\varepsilon^k := \{\varepsilon_1, \ldots, \varepsilon_k\}$ *are the corresponding signs of the selected variables.*

Remark 2. *Observe that Irrepresentable Condition* $(\mathcal{A}_{\text{Irr.}})$ *of order* K *implies Empirical Irrepresentable Check* $(\widehat{\mathcal{A}}_{\text{Irr.}})$ *of order* K *with respect to any possible* $(\widehat{\imath}_1, \ldots, \widehat{\imath}_K)$*.*

Remark 3. *Note that this condition can be checked in* $\mathcal{O}(n\times p\times K + n\times K^{\mathrm{c}_{X,\text{inv}}})$ *time where* $\mathcal{O}(n\times K^{\mathrm{c}_{X,\text{inv}}})$ *accounts for the the computational time of the* K *operations* $X_S(X_S^\top X_S)^{-1}e$ *where* $e \in \mathbb{R}^k$ *is a fixed vector and* $S \subseteq [p]$ *has size* k *for* $k = 1, \ldots, K$*. A standard bound for this last term would be* $\mathrm{c}_{X,\text{inv}} \leq 6.3728639$*, see for instance [19].*

Remark 4. *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 2 or Algorithm 3 where these values are given by* θ *as shown by Proposition 2. It means that, in practice, one has for free the maximal order* K *for which Empirical Irrepresentable Check holds. Furthermore, this latter remark shows that the maximal order* K *for which Empirical Irrepresentable Check holds is a statistic of the variables* $\bar{\imath}_k \in [p]$ *and signs* $\varepsilon_k \in \{\pm 1\}$ *that enter the model along the LAR path.*

In Figure 1, in the particular case where the design matrix $X \in \mathbb{R}^{n\times p}$ is constructed from 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 maximal order K for which the Empirical Irrepresentable Check holds. Due to symmetry of the law of the design and of the independent observation y, note that the probability to observe a given path $\mathbb{P}\{\widehat{\imath}_1 = \imath_1, \ldots, \widehat{\imath}_n = \imath_n\}$ is almost the same for all possible paths, namely $\simeq 1/(n!)$. Then we can view the plot in Figure 1 as the larget value K for which $(\widehat{\mathcal{A}}_{\text{Irr.}})$ holds when the support $\{\widehat{\imath}_1 = \imath_1, \ldots, \widehat{\imath}_K = \imath_K\}$ is uniformly distributed.

In Figure 1, we chose $p = 250, 500, 1000, 2000$, different values of n and we plot a shaded blue line representing 95% of the values ρ (among $1,000$ repetitions). Interestingly, these values are

concentrated around the solid blue line referred to as the "empirical phase transition" for the Gaussian model.

Example 5. *For example, we found that for $p = 1,000$ and $n = 300$ ($\delta = n/p = 0.3$), Empirical Irrepresentable Check ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of order K holds when K is about $K \simeq \rho \times 300 = 0.13 \times 300 = 39$.*

For every couple ($\delta = n/p, \rho = K/n$) below this curve we observe that we have ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of order K with overwhelming probability as p is sufficiently large. We did not pursue this issue in the present paper but we would like to emphasize that ($\widehat{\mathcal{A}}_{\text{Irr.}}$) holds empirically for values K of the order of 10% of n at least. In addition, on real data of Section 4.2, with about $p = 200$ variables and $n = 700$ observations with found K of the order of 30.

2.2. Result one: Mixture of Gaussian Order Statistics in LAR

Recall that this article concerns the linear model in high-dimensions where the number of predictors p can be much greater than the number of samples n. We denote by $Y \in \mathbb{R}^n$ the response variable and we assume that

$$Y = X^0\beta^0 + \eta \sim \mathcal{N}_n(X^0\beta^0, \sigma^2\Sigma)\,, \tag{3}$$

where $\eta \sim \mathcal{N}_n(0, \sigma^2\Sigma)$ is some Gaussian noise, the matrix $\Sigma \succ 0$ is some known positive definite matrix, the noise level $\sigma > 0$ may be known or that has to be estimated depending on the context, and $X^0 \in \mathbb{R}^{n\times p}$. In this paper, we are interested in selecting the true support S^0 of β^0, where the support is defined by

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

Provided that Σ is known, one can always consider the homoscedastic-independent version of the responses, namely

$$\Sigma^{-\frac{1}{2}}Y = X\beta^0 + \Sigma^{-\frac{1}{2}}\eta \sim \mathcal{N}_n(X\beta^0, \sigma^2\mathrm{Id}_n)\,.$$

where the design X is given by $X := \Sigma^{-\frac{1}{2}}X^0$. From the observation $\Sigma^{-\frac{1}{2}}Y$ and the design X (or equivalently from $(\overline{Z}, \overline{R})$ defined in (9)), one can compute the Least Angle Regression (LAR) knots $(\lambda_k)_{k\geq 1}$ and the sequence $(\bar{\imath}_k, \varepsilon_k)_{k\geq 1}$ of variables $\bar{\imath}_k \in [p]$ and signs $\varepsilon_k \in \{\pm 1\}$. In the sequel, the law (resp. the conditional law) of a random vector V (resp. conditionally to W) is denoted by $\mathcal{L}(V)$ (resp. $\mathcal{L}(V|W)$). One of the main discovery of this paper (see Theorem 3) is the following:

> *Under Empirical Irrepresentable Check ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of order K, the LAR knots $(\lambda_1, \ldots, \lambda_K)$ behave as a mixture of Gaussian order statistics with heterogeneous variances, namely it holds*
>
> $$\mathcal{L}(\lambda_1, \ldots, \lambda_K \big| \lambda_{K+1}) = \sum_{(\imath_1,\ldots,\imath_K)\in[2p]^K} \pi_{(\imath_1,\ldots,\imath_K,\lambda_{K+1})} \underbrace{\mathrm{Z}^{-1}_{(\imath_1,\ldots,\imath_K,\lambda_{K+1})} \Big[\bigotimes_{k=1}^{K} \gamma_{m_k,v_k^2}\Big] \mathbb{1}_{\{\lambda_1\geq\cdots\geq\lambda_K\geq\lambda_{K+1}\}}}_{\mathcal{L}(\lambda_1,\ldots,\lambda_K\,|\,\widehat{\imath}_1=\imath_1,\ldots,\widehat{\imath}_K=\imath_K\,,\,\lambda_{K+1})}\,, \tag{4}$$
>
> *where $\pi_{(\imath_1,\ldots,\imath_K,\lambda_{K+1})} = \mathbb{P}\big\{\widehat{\imath}_1 = \imath_1, \ldots, \widehat{\imath}_K = \imath_K \big| \lambda_{K+1}\big\}$ can be interpreted as mixing probabilities, $\mathrm{Z}_{(\imath_1,\ldots,\imath_K,\lambda_{K+1})}$ is a normalizing constant, and γ_{m_k,v_k^2} is a short notation for the Gaussian law such that*
>
> - *its mean m_k is given by (15) and depends only on the covariance $\sigma^2\Sigma$, the selected support $(\bar{\imath}_1, \ldots, \bar{\imath}_k)$ at step k, and the orthogonal projection of $\overline{\mu}^0 := X^\top X\beta^0$ given by (5);*
> - *and its variance $v_k^2 := \sigma^2\rho_k^2$ is given by (13) and depends only on the covariance $\sigma^2\Sigma$ and the selected support $(\imath_1, \ldots, \imath_k)$ at step k.*

As presented in the sequel, this result is the key step to prove an exact control of the FDR, see Section 2.3, and an exact control of the False Negatives of some selection procedures, see Section 2.4. These controls are obtained from p-values $\widehat{\alpha}_{a,b,c}$ described in Theorem 5 and (19). We will see that **the p-value $\widehat{\alpha}_{abc}$ detects abnormal large values of λ_b conditional on** (λ_a, λ_c) and that, in the orthogonal design case, Theorem 6 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 $a \leq x < y < z \leq K+1$.

2.3. Result two: Control of False Discovery Rate in the Orthogonal Design case

2.3.1. Presentation in the general case

Assume that Empirical Irrepresentable Check ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of order K holds. The sequence of testing procedures under consideration is given by the sequence of p-values $(\widehat{\alpha}_{k-1,k,k+1})_{k\in[K]}$ that are referred to as "*Spacing tests*" [29] in the literature, they are recalled in Remark 10 below. These p-values account for **"abnormally large" values of λ_k conditional on** $(\lambda_{k-1}, \lambda_{k+1})$. Invoking (4), this conditional law of λ_k depends on two unobserved values: the mean m_k, given by (15), and the variance σ^2 (in the case where the variance may be unknown). As discussed in Section 2.4.1, the variance σ^2 will be estimated on residuals as described by the heuristic of Remark 6. It entails that one may efficiently estimate the variance and, as we will see in Section 3.4, one can plug this independent estimator of the variance so as to get a studentized version of the testing procedures described here. For sake of readability, we will assume that σ is known and we refer to Section 3.4 for details on how to studentize our procedure. We understand that the law of testing statistics are parametrized by the hypotheses $(m_k)_{k\in[K]}$, where m_k is given by (15).

We recall that we denote $\overline{\mu}^0 = X^\top X\beta^0$ and $\overline{\mu}^0_i$ it ith coordinate . Assuming that predictors are normalised, in the general case, this quantity is the sum of β^0_j'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 $(P^\perp_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}(\overline{\mu}^0))_{\bar{\imath}_k}$ the orthogonal projection given by

$$(P^\perp_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}(\overline{\mu}^0))_{\bar{\imath}_k} := \varepsilon_k X_i^\top \Big[\mathrm{Id}_n - X_{\overline{S}^{k-1}} \operatorname{diag}(\varepsilon_{\overline{S}^{k-1}}) \big(X^\top_{\overline{S}^{k-1}} X_{\overline{S}^{k-1}}\big)^{-1} \operatorname{diag}(\varepsilon_{\overline{S}^{k-1}}) X^\top_{\overline{S}^{k-1}}\Big] X\beta^0. \quad (5)$$

where $\operatorname{diag}(\varepsilon_{\overline{S}^{k-1}}) := \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_{k-1})$.

The tested hypotheses are conditional on the sequence of variables $(\bar{\imath}_1, \ldots, \bar{\imath}_K) \in [p]^K$ and signs $\varepsilon_1, \ldots, \varepsilon_k \in \{\pm 1\}^k$ entering the model. The p-values under consideration here are given by

$$\begin{aligned}
&\circ\ \widehat{p}_1 := \widehat{\alpha}_{0,1,2} \text{ is the } p\text{-value testing } \mathbb{H}_{0,1} : \text{“}m_1 = 0\,\text{”} \text{ namely } \overline{\mu}^0_{\bar{\imath}_1} = 0;\\
&\circ\ \widehat{p}_2 := \widehat{\alpha}_{1,2,3} \text{ is the } p\text{-value testing } \mathbb{H}_{0,1} : \text{“}m_2 = 0\,\text{”} \text{ namely } (P^\perp_1(\overline{\mu}^0))_{\bar{\imath}_2} = 0;\\
&\circ\ \widehat{p}_3 := \widehat{\alpha}_{2,3,4} \text{ is the } p\text{-value testing } \mathbb{H}_{0,1} : \text{“}m_1 = 0\,\text{”} \text{ namely } (P^\perp_2(\overline{\mu}^0))_{\bar{\imath}_3} = 0; \qquad (6)\\
&\circ\ \text{and so on...}
\end{aligned}$$

We write I_0 fo 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 $\mathrm{FP} = \operatorname{card}(\widehat{R} \cap I_0)$ and $\mathrm{TP} = \operatorname{card}(\widehat{R} \setminus I_0)$.

Denote by $\widehat{p}_{(1)} \leq \ldots \leq \widehat{p}_{(K)}$ 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 $\{k \in [K] \ :\ \widehat{p}_{(k)} \leq \alpha k/K\} = \emptyset$ and

$$\widehat{R} = \{k \in [K] \ :\ \widehat{p}_k \leq \alpha\widehat{k}/K\} \quad \text{where} \quad \widehat{k} = \max\big\{k \in [K] \ :\ \widehat{p}_{(k)} \leq \alpha k/K\big\}. \qquad (7)$$

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

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

where the expectation is unconditional to 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

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

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

2.3.2. 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 (6). Note that I_0 is simply the set of null coordinates of β. Remark also that, Irrepresentable Condition ($\mathcal{A}_{\mathrm{Irr.}}$) of order p holds and so does Empirical Irrepresentable Check ($\widehat{\mathcal{A}}_{\mathrm{Irr.}}$), see Proposition 2. Note also that I_0 is simply the set of null coordinates of β. It implies that on can consider any value $K \in [p]$ in the following result.

Theorem 1. *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 (6) and the set $\widehat{R}$ given by (7). Then*

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

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 2.3.1, then the FDR is exactly controlled by α.

2.4. Result three: Exact Testing Procedure on False Negatives

From the result of Section 2.2, one can present a method to select a model and propose an exact test on false negatives in the general design case. More precisely, we assume here for simplicity that the design X has full row rank n. In theory we are able to compute the LAR knots up to λ_n. Of course this is often out of reach due to numerical issues and one has to stop earlier to avoid bad conditioning, but we forget this possibility to avoid heavy notation. One can adapt easily the following text to take into account this possibility. Let us compute the LAR knots $(\lambda_1,\ldots,\lambda_n)$ and the "*signed*" variables $(\widehat{\imath}_1,\ldots,\widehat{\imath}_n) \in [2p]^n$ entering the model as in (1).

2.4.1. Admissible Support Selection Procedures

In order to select $\widehat{S}$, one may be interested in an estimation $\widehat{\sigma}_{\mathrm{select}}$ of the noise level since, in practice, one usually does not know it. Our procedure is valid as long as the following property ($\mathcal{P}_1$) is satisfied

> **Noise Level Estimation for Selection ($\mathcal{P}_1$):** *The estimated noise level $\widehat{\sigma}_{\mathrm{select}}$ is a measurable function of $P^{\perp}_{n_{\mathrm{select}}}(Y)$ where n_{select} may depend only on $(\widehat{\imath}_1,\ldots,\widehat{\imath}_n)$.*

In practice, one can chose a fixed $n_{\text{select}} = n - B$ giving a fixed budget of $B \geq 1$ independent observations to estimate the variance[1]. Obviously, if the noise level is known, one can simply take $n_{\text{select}} = n$ and $\widehat{\sigma}_{\text{select}} = \sigma$.

Remark 6. *Observe that if the true support S^0 is included in $\{\bar{\imath}_1, \bar{\imath}_2, \ldots, \bar{\imath}_k\}$ then $P_k^{\perp}(Y)$ is centered Gaussian vector with known covariance (up to σ^2) and it is standard to compute an estimator of the noise level σ. In particular, the information contained in $P_k^{\perp}(Y)$ does not depend on the target β^0, since $P_k^{\perp}(X^0\beta^0) = 0$, and is purely due to the "noise" part η, as defined in (3).*

Note that, for large enough n_{select}, the heuristic described in Remark 6 may be true and it entails that one may efficiently estimate the variance on $P_{n_{\text{select}}}^{\perp}(Y)$.

Now, note that choosing a model $\widehat{S}$ is equivalent to choosing a model size $\widehat{m} \in [n]$ so that

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

Our procedure is flexible on this point and it allows any choice of $\widehat{m}$ as long as the following property $(\mathcal{P}_2)$ is satisfied

> **Stopping Rule $(\mathcal{P}_2)$:** *The estimated model size $\widehat{m}$ is a "stopping time", i.e. $\mathbb{1}_{\{\widehat{m} \leq k\}}$ is a measurable function of $(P_k(Y), \widehat{\sigma}_{\text{select}})$ for all $k \in [n_{\text{select}}]$.*

In other words, the decision to select a model of size $\widehat{m} = k$ depends only on the part of the observation Y explained by the predictors $X_{\bar{\imath}_1}, \ldots, X_{\bar{\imath}_k}$. In the sequel, we will present some examples of such "stopping time" procedures, see Section 3.5. We would like to emphasize that our analysis reveal the following conditional independence

$$\mathcal{L}\big((P_{n_{\text{select}}}(Y), \widehat{\sigma}_{\text{select}})|(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)\big) = \mathcal{L}\big(P_{n_{\text{select}}}(Y)|(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)\big) \otimes \mathcal{L}\big(\widehat{\sigma}_{\text{select}}|(\widehat{\imath}_1, \ldots, \widehat{\imath}_n)\big) \tag{8}$$

which can be advantageously invoked to build Student-type testing statistics to define $\widehat{m}$.

2.4.2. Exact Testing Procedure on False Negatives

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 $\mathbb{H}_0$: "$S^0 \subseteq \widehat{S}$", namely there is no false negatives. Equivalently, one aim at testing the null hypothesis

$$\mathbb{H}_0 \ : \ \text{"} X\beta^0 \in E_{\widehat{m}} \text{"},$$

at an exact significance level $\alpha \in (0, 1)$

In this article, we introduce a new **exact** testing procedure that can be deployed when $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ hold, namely an "admissible" selection procedure is used to build $\widehat{S}$. The variance estimator $\widehat{\sigma}_{\text{select}}$ may have been used in the definition of $\widehat{m}$ and our method forbid to use it again to "studentized" our testing procedure. We need to estimate the variance again to preserve some "independence" in the spirit of (8). As in the previous variance estimation, we advocate a "budget" to this task, namely we use the "*strata*" $(n_{\text{test}}, n_{\text{select}}]$ (as in the vocabulary of analysis of variance) to build $\widehat{\sigma}^2_{\text{test}}$ an estimation of the variance. More precisely, we use the orthogonal projection $P_{(n_{\text{test}}, n_{\text{select}}]}(Y)$ which is the the orthogonal projector onto the space F defined by

$$E_{n_{\text{select}}} = E_{n_{\text{test}}} \overset{\perp}{\oplus} F.$$

Note that this space is generated by $X_{\bar{\imath}_{n_{\text{test}}+1}}, \ldots, X_{\bar{\imath}_{n_{\text{select}}}}$.

The resulting test is based on $(\lambda_{\widehat{m}}/\widehat{\sigma}_{\text{test}}, \ldots, \lambda_{n_{\text{test}}}/\widehat{\sigma}_{\text{test}})$. More precisely, we know the law of $\lambda_{\widehat{m}_{\text{select}}+1}/\widehat{\sigma}_{\text{test}}$ conditional on $\widehat{m}, \lambda_{\widehat{m}}, \lambda_{n_{\text{test}}}$ under $\mathbb{H}_0$.

[1] One can also look at the conditioning of the variance-covariance matrix of $P_k^{\perp}(Y)$ for $k = n, n-1, \ldots$ to choose n_{select}. A more detailed discussion on this issue can be found in Section 2.5.

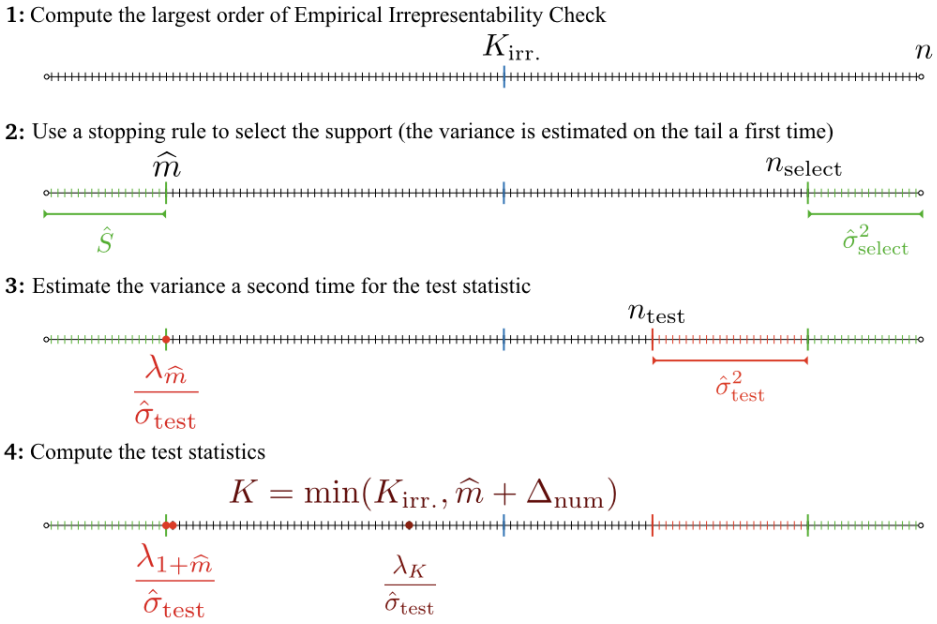

Figure 2. *The testing procedure computes the significance of knot $\lambda_{1+\widehat{m}}$ w.r.t. knots $\lambda_{\widehat{m}}$ and λ_K (see Step 4 above), where K is a compromise between the numerical limitations $\widehat{m}+\Delta_{\text{num}}$ of the cost of integrating a function on $[0,1]^{\Delta_{\text{num}}-2}$ and the statistical limitations of the design as depicted by the Empirical Irrepresentable Check.*

2.5. Practical Framework

We start by presenting the meta-algorithm in which our testing procedure can be deployed. First, we compute the sequence of LAR knots $(\lambda_1, \dots, \lambda_n)$ and "signed" variables $(\widehat{\imath}_1, \dots, \widehat{\imath}_n) \in [2p]^n$ entering the model as in (1). Conditionally to the sequence $\widehat{\imath}_1, \dots, \widehat{\imath}_n$ the quantities displayed in Figure 2 are introduced in the following order.

- Compute $K_{\text{irr.}}$ which is the maximal value such that $(\widehat{\mathcal{A}}_{\text{Irr.}})$ holds true.
- Fixing a budget B for variance estimation $\widehat{\sigma}_{\text{select}}$ we fix $n_{\text{select}} = n - B$. In most of the applications, n is large and consequently n_{select} is also large.
- From the estimation $\widehat{\sigma}_{\text{select}}$ we deduce $\widehat{m}_{\text{select}}$. Here, the practitioner is free to use any selection procedures as soon as it satisfies $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$.
- Multivariate integration programs have some dimension limitation, this implies that we can only handle simultaneously Δ_{num} knots. As a consequence K must be smaller or equal than $\widehat{m}_{\text{select}} + \Delta_{\text{num}}$.
- In most cases the quantity
$$K = \min(K_{\text{Irr.}}, \widehat{m}_{\text{select}} + \Delta_{\text{num}})$$
is convenient because $K < n_{\text{select}}$ and there is enough budget in the strata $(K, n_{\text{select}}]$ to build the second estimator of variance, namely $\widehat{\sigma}_{\text{test}}$.
- Otherwise we have to make a "trade-off": diminishing K to obtain a sufficient budget of degree of freedom in the strata $(K, n_{\text{select}}]$ to estimate the variance while maintaining K large to gain power.

An example of values of these parameters is

$$n = 200,\ p = 300,\ K_{\text{irr.}} = 35 \text{ (for a Gaussian design)},\ \text{and } \Delta_{\text{num}} = 10\,.$$

This framework has been deployed on real data in Section 4. Examples of *admissible* procedures satisfying $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ are presented in Section 3.5.

3. Main Results

Consider the correlations vector $\overline{Z}$ of the homoscedastic-independent observation $\Sigma^{-\frac{1}{2}}Y$ with the design matrix X, so that

$$\begin{aligned} \overline{Z} &:= {X^0}^\top \Sigma^{-1} Y = \overline{R}\beta^0 + {X^0}^\top \Sigma^{-1}\eta \sim \mathcal{N}_p(\overline{R}\beta^0, \sigma^2\overline{R}) \\ \text{and} \quad \overline{R} &:= X^\top X = {X^0}^\top \Sigma^{-1} X^0 \,. \end{aligned} \tag{9}$$

From $\overline{Z}$ and $\overline{R}$ we can compute the Least Angle Regression (LAR) knots $(\lambda_k)_{k\geq 1}$ and the sequence $(\bar{\imath}_k, \varepsilon_k)_{k\geq 1}$ of variables $\bar{\imath}_k$ and signs ε_k that enter the model along the LAR path. We will use these statistics as testing statistics. For sake of presentation, we may consider the $2p$-vector $Z := (\overline{Z}, -\overline{Z})$ whose mean is given by $\mu^0 := (\overline{R}\beta^0, -\overline{R}\beta^0) = (X^\top X\beta^0, -X^\top X\beta^0) = (\overline{\mu}^0, -\overline{\mu}^0)$ and its variance-covariance matrix is $\sigma^2 R$ with

$$R = \begin{bmatrix} \overline{R} & -\overline{R} \\ -\overline{R} & \overline{R} \end{bmatrix} = \begin{bmatrix} X^\top X & -X^\top X \\ -X^\top X & X^\top X \end{bmatrix} . \tag{10}$$

3.1. Problem reformulation and Assumptions

Remark that $Z = (Z_i)_i$ is a Gaussian vector of size $2p$ such that $Z_i = -Z_{i+p}$ where the indices are considered modulo $2p$. Note also that we know the covariance of Z up to some multiplicative constant σ^2, namely the variance-covariance matrix of Z is given by $\sigma^2 R$ where $R \succeq 0$ is known and σ^2 is some parameter that may be known or not.

Now, we can define

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

where $(1, \dots, 1)$ is the column vector of size k whose entries are equal to one and $\sigma^2 M_{i_1,\dots,i_k}$ 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 $M_{i_1,\dots,i_k}$ is the submatrix of R obtained by keeping the columns and the rows indexed by $\{i_1, \dots, i_k\}$. Remark that

$$\theta_j(i_1, \dots, i_k) = \mathbb{E}(Z_j \mid Z_{i_1} = 1, \dots, Z_{i_k} = 1)\,,$$

when $\mathbb{E}Z = 0$. In our context, the Irrepresentable Condition of order K is given by

$$\forall k \leq K,\ \forall (i_1, \dots, i_k) \in [2p]^k,\ \forall j \notin \{i_1, \dots, i_k\}\,, \quad \theta_j(i_1, \dots, i_k) < 1\,, \tag{12}$$

where $\theta_j(i_1, \dots, i_k)$ is given by (11). In Proposition 2 we show that (12) is equivalent to the Irrepresentable Condition ($\mathcal{A}_{\text{Irr.}}$) on the design matrix X. A proof can be found in Appendix C.2.

Proposition 2. *Assume X and R satisfy (10) then the following assumptions are equivalent:*

- *the design matrix X satisfies ($\mathcal{A}_{\text{Irr.}}$) of order K,*
- *the variance-covariance matrix R satisfies (12) of order K.*

Furthermore, if one of these two assumptions hold then "Empirical Irrepresentable Check" of order K holds, namely

$$\max\Big[\max_{j\neq\widehat{\imath}_1} \theta_j(\widehat{\imath}_1),\ \dots\ ,\ \max_{j\neq\widehat{\imath}_1,\dots,\widehat{\imath}_k} \theta_j(\widehat{\imath}_1, \dots, \widehat{\imath}_k)\Big] < 1$$

which is an equivalent formulation of ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of Definition 2.

Remark 7. *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 ($\mathcal{A}_{\text{Irr.}}$) of order $K = 1$. Hence, up to some normalization, one can always assume ($\mathcal{A}_{\text{Irr.}}$) of order $K = 1$.*

3.2. Conditional Joint Law of the Knots

In this section, we are interested in the joint law of the knots $(\lambda_1, \dots, \lambda_K)$ of the LAR conditional on λ_{K+1}. Let $(\widehat{\imath}_1, \dots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. Define the first correlation by $\rho_1 := \sqrt{R_{\widehat{\imath}_1,\widehat{\imath}_1}}$ and the others by

$$\rho_\ell := \frac{\sqrt{R_{\widehat{\imath}_\ell,\widehat{\imath}_\ell} - \big(R_{\widehat{\imath}_\ell,\widehat{\imath}_1} \cdots R_{\widehat{\imath}_\ell,\widehat{\imath}_{\ell-1}}\big) M^{-1}_{\widehat{\imath}_1,\dots,\widehat{\imath}_{\ell-1}} \big(R_{\widehat{\imath}_\ell,\widehat{\imath}_1}, \cdots, R_{\widehat{\imath}_\ell,\widehat{\imath}_{\ell-1}}\big)}}{1 - \theta^{\ell-1}_{\widehat{\imath}_\ell}}, \quad \text{for } \ell \geq 2, \tag{13}$$

where

$$\theta^{\ell-1} := \theta(\widehat{\imath}_1, \dots, \widehat{\imath}_{\ell-1}), \quad \text{for } \ell \geq 2,$$

is defined by (11) and $M_{\widehat{\imath}_1,\dots,\widehat{\imath}_{\ell-1}}$ is the submatrix of R obtained by keeping the columns and the rows indexed by $\{\widehat{\imath}_1, \dots, \widehat{\imath}_{\ell-1}\}$. Furthermore, denote

$$F_i := \Phi_i(\lambda_i) \quad \text{and} \quad \mathcal{P}_{i,j} := \Phi_i \circ \Phi_j^{-1}, \quad \text{for } i, j \in [K+1], \tag{14}$$

where $\Phi_k(\cdot) := \Phi(\cdot/(\sigma\rho_k))$ 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. The main discovery of this paper is the following theorem.

Theorem 3 (Conditional Joint Law of the LAR Knots)**.** *Let $(\lambda_1, \dots, \lambda_K, \lambda_{K+1})$ be the first knots and let $(\widehat{\imath}_1, \dots, \widehat{\imath}_K)$ be the first variables entering along the LAR path. If ($\widehat{\mathcal{A}}_{\text{Irr.}}$) holds then, conditional on $\{\widehat{\imath}_1, \dots, \widehat{\imath}_K, \lambda_{K+1}\}$, the vector $(\lambda_1, \dots, \lambda_K)$ has law with density (w.r.t. the Lebesgue measure)*

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

where φ_{m_k,v_k^2} is the standard Gaussian density with mean

$$m_k := \frac{\mu^0_{\widehat{\imath}_k} - \big(R_{\widehat{\imath}_k,\widehat{\imath}_1} \cdots R_{\widehat{\imath}_k,\widehat{\imath}_{k-1}}\big) M^{-1}_{\widehat{\imath}_1,\dots,\widehat{\imath}_{k-1}} \big(\mu^0_{\widehat{\imath}_1}, \cdots, \mu^0_{\widehat{\imath}_{k-1}}\big)}{1 - \theta^{k-1}_{\widehat{\imath}_k}} \quad \text{with} \quad \mu^0 := (\mu^0, -\mu^0), \tag{15}$$

and variance $v_k^2 := \sigma^2\rho_k^2$.

A useful corollary of this theorem is the following. It shows that one can explicitly describe the joint law of the LAR's knots after having selected a support of size $\widehat{m}$ with any procedure satisfying $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$.

Corollary 4. *Under the conditions of Theorem 3, let $\widehat{m}$ be chosen according to a procedure satisfying $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ with $n_{\text{select}} \geq K$. Then under the null hypothesis, namely*

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

and conditional on the selection event $\big\{\widehat{m} = a, F_a, F_{K+1}, \widehat{\imath}_1, \dots, \widehat{\imath}_K, \widehat{\imath}_{K+1}\big\}$ with $0 \leq a \leq K-1$, the vector $(F_{a+1}, \dots, F_K)$ is uniformly distributed on

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

where $\mathcal{P}_{i,j}$ are described in (14).

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}, $(F_1, \dots, F_K)$ is uniformly distributed on*

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

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*

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

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*.*

3.3. Testing Procedures

From Theorem 3, we deduce several testing statistics. To this end, we introduce some notation. First, define

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

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

and also

$$\mathbb{F}_{abc}(t) := \mathbb{1}_{\{\lambda_c \leq t \leq \lambda_a\}} \int_{\Phi_b(\lambda_c)}^{\Phi_b(t)} \mathcal{I}_{ab}(F_a, f_b)\, \mathcal{I}_{bc}(f_b, F_c)\, \mathrm{d}f_b \tag{16}$$

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

Then, a simple integration shows that Corollary 4 implies that

$$\mathbb{P}\big[\lambda_b \leq t \mid \lambda_a, \lambda_c, \widehat{\imath}_1, \widehat{\imath}_2, \ldots, \widehat{\imath}_{c-1}\big] = \frac{\mathbb{F}_{abc}(t)}{\mathbb{F}_{abc}(\lambda_a)}\,, \tag{17}$$

whenever $\mathbb{E}Z = 0$. Actually, this result can be refined in the case when $\mathbb{E}Z \neq 0$ by the next result.

Theorem 5. *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* $(\widehat{\mathcal{A}}_{\text{Irr.}})$ *holds and* $\widehat{m}$ *is chosen according to a procedure satisfying* $(\mathcal{P}_1)$ *and* $(\mathcal{P}_2)$ *with* $n_{\text{select}} \geq K$*, then under the null hypothesis, namely*

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

and conditional on the selection event $\big\{\widehat{m} = a\big\}$ *with* $a \leq K-1$*, and for any integers* b, c *such that* $a < b < c \leq K+1$*, it holds that*

$$\widehat{\alpha}_{abc} := 1 - \frac{\mathbb{F}_{abc}(\lambda_b)}{\mathbb{F}_{abc}(\lambda_a)} \sim \mathcal{U}(0,1)\,, \tag{18}$$

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

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 10. *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].

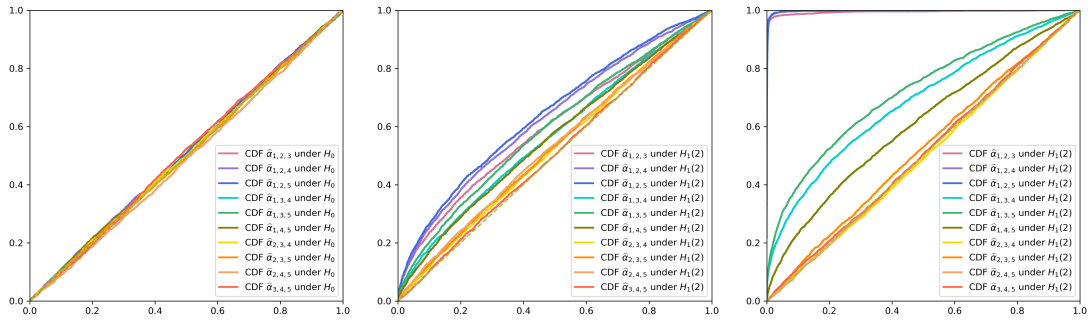

Figure 3. *CDF of p-values* $\widehat{\alpha}_{abc}$ *over* $3,000$ *Monte-Carlo iterations with* $n = 200$, $p = 300$ *and a random design matrix* X *given by* 300 *independent column vectors uniformly distributed on the Euclidean sphere* $\mathbb{S}^{299}$. *Central panel represents an alternative composed by a* 2 *sparse vector, the right panel an alternative composed by a* 2 *sparse vector* 5 *times larger while the left panel corresponds to the null.*

Given $\alpha \in (0,1)$, one can consider the following testing procedures

$$\mathcal{S}_{abc} := \mathbb{1}_{\{\widehat{\alpha}_{abc} \leq \alpha\}}\,, \tag{19}$$

that rejects if the p-value $\widehat{\alpha}_{abc}$ is less than the level α of the test. Now, recall (17) to witness that the smaller $\widehat{\alpha}_{abc}$, the bigger λ_b conditionally to (λ_a, λ_c). So **the p-value $\widehat{\alpha}_{abc}$ detects abnormal large values of λ_b conditional on (λ_a, λ_c)**.

One may investigate the power of these tests to detect false negatives, namely to detect the alternative 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 testing procedures?* A comprehensive study of the orthogonal case is given in the following theorem.

Theorem 6. *Assume that the design is orthogonal, namely* $\overline{R} = \mathrm{Id}_p$. *Let* $\widehat{m}$ *be chosen according to a procedure satisfying* $(\mathcal{P}_1)$ *and* $(\mathcal{P}_2)$ *with* $n_{\text{select}} \geq K$. *Then under the null hypothesis, namely*

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

and conditional on the selection event $\{\widehat{m} = a_0\}$ *with* $1 \leq a_0 \leq K-1$, *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.6. It shows that the best choice among the tests $\mathcal{S}_{a,b,c}$ is the test $\mathcal{S}_{a_0,a_0+1,K+1}$ with the smallest a and the largest c.

3.4. “Studentization”: Adapting to Unknown Noise Level

The previous results can be further refined to the case where σ^2 is unknown. Let K chosen following the considerations of Section 2.5, the key property is to build an estimation of the variance $\widehat{\sigma}^2_{\text{test}}$ on the strata $(n_{\text{test}}, n_{\text{select}}]$ which is independent from

$$(\lambda_{\widehat{m}}, \ldots, \lambda_{K+1}, \widehat{\sigma}_{\text{test}})\,,$$

conditional on the selection event $\{\widehat{\imath}_1 = \imath_1, \ldots, \widehat{\imath}_n = \imath_n\}$ and under $\mathbb{H}_0$: “$X\beta^0 \in E_{\widehat{m}}$”.

Because of our hypotheses, $\mathrm{rank}(P_{(n_{\text{test}},n_{\text{select}}]})$ is maximal equal to $m := n_{\text{select}} - n_{\text{test}}$. Our variance estimator

$$\widehat{\sigma}^2 := \widehat{\sigma}^2_{\text{test}} = \frac{\|P_{(n_{\text{test}},n_{\text{select}}]}(Z)\|_2^2}{m},$$

follows a $\sigma^2\chi^2(m)/m =: \sigma^2 W^2$.

We set $\widehat{\lambda}_i := \frac{\lambda_i}{\widehat{\sigma}} = \frac{\lambda_i}{\sigma W}$. Let us consider $a = m < b < c \leq K+1$ and consider, under $\mathbb{H}_0$, the distribution conditional to $\left\{\widehat{m} = a, \lambda_a, \lambda_{K+1}, \widehat{\imath}_1, \ldots, \widehat{\imath}_K, \widehat{\imath}_{K+1}\right\}$ Then the variables $(\lambda_a, \ldots, \lambda_c)$ and W are independent. More over the distribution of λ_b is given by (17). We have

$$\mathbb{P}\left\{\widehat{\lambda}_b \leq t \middle| \widehat{\lambda}_a \widehat{\lambda}_c, W = w, \widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1}\right\} = \mathbb{P}\left\{\frac{\lambda_b}{\sigma} \leq wt \middle| \widehat{\lambda}_a \widehat{\lambda}_c, W = w, \widehat{\imath}_1, \ldots, \widehat{\imath}_{K+1}\right\}.$$

Then we can use Theorem 5 in the case $\sigma = 1$ to get that the expression above is equal to

$$\frac{\overline{F}_{abc}(tw)}{\overline{F}_{abc}(\widehat{\lambda}_a w)},$$

Where $\overline{F}_{abc}$ is the function given by (17) in the case $\sigma = 1$. De-conditioning in w we get

$$\mathbb{P}\{\widehat{\lambda}_b \leq t \big| \widehat{\lambda}_a \widehat{\lambda}_c, i_1, \ldots, i_{K+1}\} = \int_{\mathbb{R}} \frac{\overline{F}_{abc}(tw)}{\overline{F}_{abc}(\widehat{\lambda}_a w)} d\,\mathbb{P}_W(w).$$

As a consequence the p-value of the test in now given by

$$\widehat{\beta}_{abc} = 1 - \int_{\mathbb{R}} \frac{\overline{F}_{abc}(\widehat{\lambda}_b w)}{\overline{F}_{abc}(\widehat{\lambda}_a w)} d\,\mathbb{P}_W(w).$$

3.5. 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 the properties $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$. We chose a level α' and we define a light modification of the Student test

$$T_{abc} = \mathbb{1}_{\widehat{\beta}_{abc} \leq \alpha'},$$

by replacing $\widehat{\sigma}^2_{\text{test}}$ by $\widehat{\sigma}^2_{\text{select}}$. The number of degrees of freedom of the χ^2 distribution is now $\overline{m} := n - n_{\text{select}}$. By a small abuse of notation, we still set

$$\widehat{\lambda}_i := \frac{\lambda_i}{\widehat{\sigma}_{\text{select}}}.$$

We limit our attention to consecutive a, b, c. In such a case it is easy to see that

$$\widehat{\beta}_{a(a+1)(a+2)} = \frac{\mathcal{T}\left(\frac{\widehat{\lambda}_{a+1}}{\sigma\rho_{a+1}}\right) - \mathcal{T}\left(\frac{\widehat{\lambda}_a}{\sigma\rho_{a+1}}\right)}{\mathcal{T}\left(\frac{\widehat{\lambda}_{a+2}}{\sigma\rho_{a+1}}\right) - \mathcal{T}\left(\frac{\widehat{\lambda}_a}{\sigma\rho_{a+1}}\right)},$$

where $\mathcal{T}$ is the cumulative distribution of the Student $T(\overline{m})$ law, 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 $T_{a,a+1,a+2}$ at the level α',
- if the test is significative, we set $a = a + 1$ and keep on going,
- if it is non-significative, we stop and set $\widehat{m} = a + 2$.

Recall that possible selected supports along the LAR's path are nested models of the form (2). Denote $k^0 \geq 1$ the smallest integer k such that the true support S^0 is contained in $\overline{S}^k$ and denote

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

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 possible admissible procedure would be the following.

- Begin with $a = 0$,
- at each step, perform the test $T_{a,a+1,a+2}$ at the level α',
- if the test is significative, 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-significatives, we stop and set $\widehat{m} = a + \gamma_{FP} + 1$.

This method has been deployed on real data in Section 4.2 with $\gamma_{FP} = 3$.

4. Numerical Experiments

4.1. Monte-Carlo experiment

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 $\mathbb{S}^{299}$. The computation of the function $\mathbb{F}_{abc}$ given by (16) is an important issue that demands multivariate integration tools, see Appendix E for a solution using cubature of integral by lattice rule. This has lead to some limitations namely $\Delta_{\text{num}} \leq 4$ 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 of $T_{(\texttt{start})(\texttt{middle})(\texttt{end})}$ of knots and indexes given by `lars` and an estimation 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. More precisely, it holds, as in the orthogonal case, that

- $\mathcal{S}_{123} \leq \mathcal{S}_{124} \leq \mathcal{S}_{125}$;
- $\mathcal{S}_{234} \leq \mathcal{S}_{245}$.

4.2. 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 3.4 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 2.4, and Section 3.5 with $\gamma_{FP} = 3$. We refer to the aforementioned Python notebook for further details.

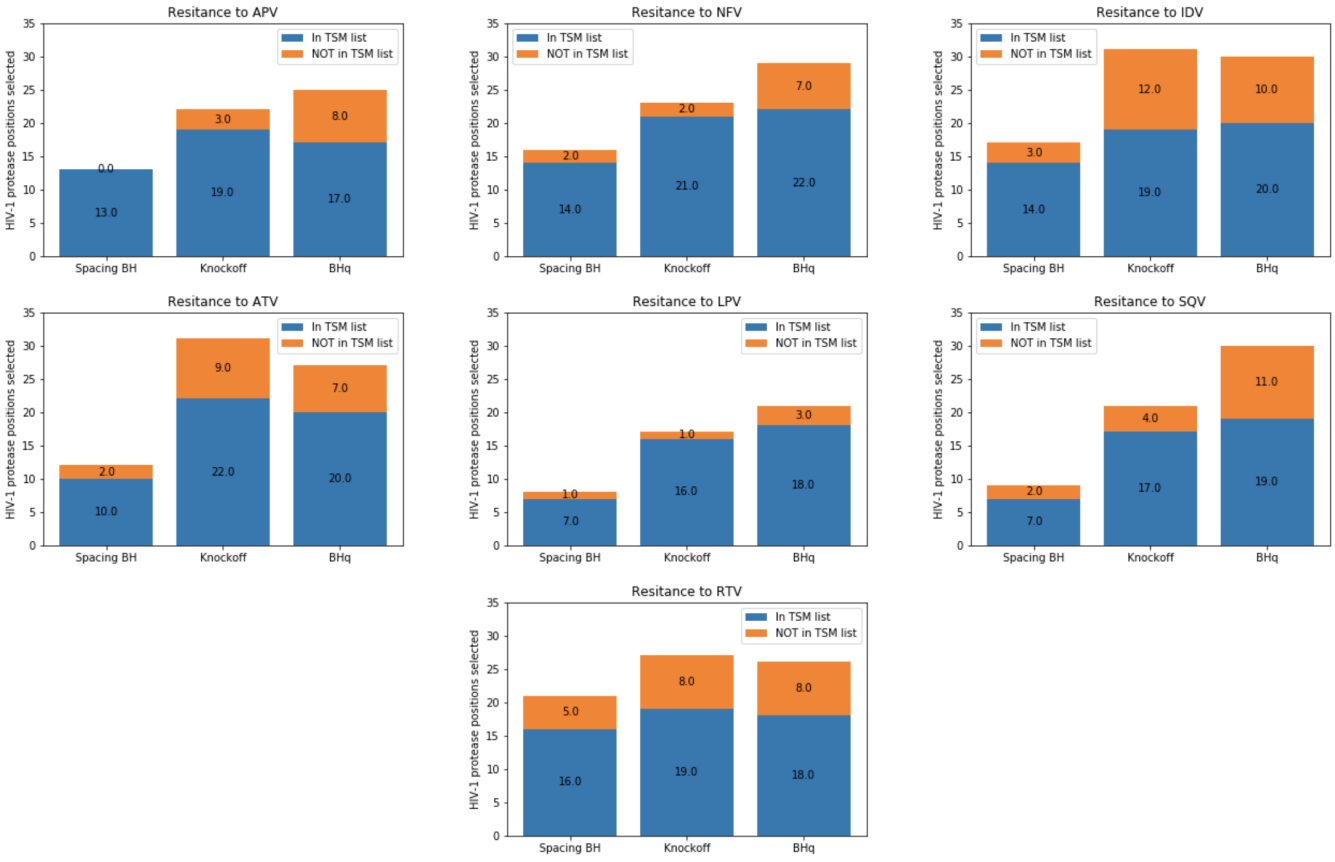

Figure 4. *Comparison of numbers true and false positives for three procedures: Spacing-BH; Knockoff [3] and Benjamini–Hochberg procedure on the coefficient of linear regression BHq. For each drug we indicate the number of true positives in blue and false positives in orange. In the three procedures 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] C. Giraud. *Introduction to high-dimensional statistics*. Chapman and Hall/CRC, 2014.

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

[19] F. Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, ISSAC '14, pages 296–303, New York, NY, USA, 2014. ACM.

[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 2: 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

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

where $M_{\bar{\imath}_1,\ldots,\bar{\imath}_{k-1}}$ is the submatrix 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

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

and pick

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

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

Then, iterate from **2**.

Algorithm 3: 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 (10), 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

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

where

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

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 2) 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 2 and note that $\widehat{\imath}_1$ and $(\bar{\imath}_1, \varepsilon_1)$ are related as in (1).

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

$$N^{(2)}(\lambda) = N^{(1)} - (\lambda_1 - \lambda)\theta(\widehat{\imath}_1)\,, \quad 0 < \lambda \leq \lambda_1\,, \tag{20}$$

and one can check that $N^{(2)}(\lambda) = (\overline{N}^{(2)}(\lambda), -\overline{N}^{(2)}(\lambda))$ where $\overline{N}(\lambda)$ is defined in Algorithm 2. 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 2. We call this value λ_2 and one may check that this definition is consistent with λ_2 in Algorithm 2.

Now, we can be more explicit on the expression of λ_2. Indeed, we may make the following discussion depending 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 (20).
- 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

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

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

A.3. Recursion: Next Knots

The loop (**2** $\rightleftarrows$ **3**) in Algorithm 2 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

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

at Step **2** and it holds that $\widehat{\imath}_\ell$ and $(\overline{\imath}_\ell, \varepsilon_\ell)$ are related as in (1). 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 2.

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

$$N^{(k)} = N^{(k-1)} - (\lambda_{k-1} - \lambda_k)\theta(\widehat{\imath}_1, \ldots, \widehat{\imath}_{k-1}), \tag{21}$$

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.3.

Lemma 7. *It holds*

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

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

Using Lemma 7 we deduce that λ_k in Algorithm 2 is consistent with

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

where $P_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_j) = \big(R_{j,\widehat{\imath}_1} \cdots R_{j,\widehat{\imath}_{k-1}}\big) M^{-1}_{\widehat{\imath}_1,\ldots,\widehat{\imath}_{k-1}}(Z_{\widehat{\imath}_1}, \ldots, Z_{\widehat{\imath}_{k-1}})$. When $\mathbb{E}Z = 0$, one may remark that $P_{i_1,\ldots,i_{k-1}}(Z_j)$ 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 2). We present this formulation in Algorithm 3.

Remark 11. *Note that Algorithm 2 implies that* $\widehat{\imath}_1, \ldots, \widehat{\imath}_k$ *are pairwise different, but also that they differ modulo p. In the rest of the paper we will limit our attention to such sequences.*

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

Given fixed $i_1, \ldots, i_k \in [2p]$, one may define $Z_j^{(i_1,\ldots,i_k)} := Z_j - P_{i_1,\ldots,i_k}(Z_j)$ for indices j such that $\theta_j(i_1, \ldots, i_k) = 1$ and

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

where we recall that $P_{i_1,\ldots,i_k}(Z_j) = \big(R_{j,i_1} \cdots R_{j,i_k}\big) M^{-1}_{i_1,\ldots,i_k}(Z_{i_1}, \ldots, Z_{i_k})$. From this point, we can introduce

$$\lambda_{k+1}^{(i_1,\ldots,i_k)} := \max_{j:\theta_j(i_1,\ldots,i_k)<1} Z_j^{(i_1,\ldots,i_k)},$$

and remark that $\lambda_{k+1} = \lambda_{k+1}^{(\widehat{\imath}_1,\ldots,\widehat{\imath}_k)}$.

B.1. Law of the First Knot

One has the following lemma governing the law of λ_1.

Lemma 8. *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*

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

- *If $\theta_j(i_1) < 1$ for all $j \neq i_1$ then, conditionally to $\{\widehat{\imath}_1 = i_1\}$ 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 clear and standard. Now, observe that

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

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.4. 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 9. *Set*

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

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

$$\frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} = 1 - \frac{1-\theta_j(i_1,\ldots,i_k)}{1-\theta_j(i_1,\ldots,i_{k-1})} \tag{22}$$

and

$$\forall j \neq i_1,\ldots,i_k\,,\quad Z_j^{(i_1,\ldots,i_k)} = \frac{Z_j^{(i_1,\ldots,i_{k-1})} - Z_{i_k}^{(i_1,\ldots,i_{k-1})}\tau_{j,i_k}/\tau_{i_k,i_k}}{1-\tau_{j,i_k}/\tau_{i_k,i_k}}. \tag{23}$$

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

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

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 $v_1, \dots, v_{k-1}$.Then

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

Using this result we deduce that

$$\begin{aligned} Z(k) &= P^{\perp}_{i_1,\dots,i_k}(Z) \\ &= P^{\perp}_{i_k}(P^{\perp}_{i_1,\dots,i_{k-1}}(Z)) \\ &= P^{\perp}_{i_k}(Z(k-1)) \\ &= Z(k-1) - P_{i_k}(Z(k-1)) \\ &= Z(k-1) - \mathbf{x}(k) Z(k-1)\,, \end{aligned} \tag{24}$$

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

$$R(k) = R(k-1) - \mathbf{x}(k) R_{i_k}(k-1)^{\top} \,. \tag{25}$$

Using (22) (or (31)), remark that

$$T(k) = T(k-1) - \mathbf{x}(k)(1 - T_{i_k}(k-1)) \,. \tag{26}$$

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

B.3. Joint Law

Regarding the joint law of knots, one has the following proposition whose proof can be found in Section C.5.

Proposition 10. *One has the following for any fixed $i_1, \dots, i_{k+1} \in [2p]$.*

- *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.
- *If ($\widehat{\mathcal{A}}_{\text{Irr.}}$) of order k holds then*

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

  for any $0 \leq a \leq k-1$ with the convention $\lambda_0 = \infty$.

Appendix C: Proofs

C.1. Proof of Theorem 3

Step 1

We prove by induction on k the following equality

$$\left\{\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_{K+1} = i_{K+1}\right\} = \left\{\lambda_{k+1}^{(i_1,\dots,i_k)} \leq \left[\begin{array}{c} \lambda_k^{(i_1,\dots,i_{k-1})} \\ Z_{i_k}^{(i_1,\dots,i_{k-1})} \end{array}\right] \leq \cdots \leq \left[\begin{array}{c} \lambda_1 \\ Z_{i_1} \end{array}\right]\right\},$$

where the bracket means that, at any position you can choose indifferently the upper- or the lower-case to get an equivalent event.

Using Lemma 8 we have the following equivalences

$$\left\{\widehat{\imath}_1 = i_1\right\} = \left\{\lambda_1 = Z_{i_1}\right\} = \left\{\lambda_2^{(i_1)} \leq Z_{i_1}\right\} = \left\{\lambda_2^{(i_1)} \leq \lambda_1\right\}$$

that give the property for $k = 1$.

Using the induction hypothesis and Proposition 9 we can pass from k to $k+1$ to get that

$$\left\{\widehat{\imath}_1 = i_1, \dots, \widehat{\imath}_{k+1} = i_{k+1}\right\} = \left\{\lambda_{k+2}^{(i_1,\dots,i_{k+1})} \leq \lambda_{k+1}^{(i_1,\dots,i_k)} \leq \left[\begin{array}{c} \lambda_k^{(i_1,\dots,i_{k-1})} \\ Z_{i_k}^{(i_1,\dots,i_{k-1})} \end{array}\right] \leq \cdots \leq \left[\begin{array}{c} \lambda_1 \\ Z_{i_1} \end{array}\right]\right\}.$$

The 2^k equivalent events above imply obviously the following 2^k events

$$\left\{\lambda_{k+2}^{(i_1,\dots,i_{k+1}} \leq Z_{i_{k+1}}^{(i_1,\dots,i_k} \leq \left[\begin{array}{c} \lambda_k^{(i_1,\dots,i_{k-1})} \\ Z_{i_k}^{(i_1,\dots,i_{k-1}} \end{array}\right] \leq \cdots \leq \left[\begin{array}{c} \lambda_1 \\ Z_{i_1} \end{array}\right]\right\}.$$

To get the implication in the other direction we remark that

$$\begin{aligned}
&\lambda_{k+2}^{(i_1,\dots,i_{k+1})} \leq Z_{i_{k+1}}^{(i_1,\dots,i_k)} \\
&\Leftrightarrow \forall j \neq i_1, \dots, i_{k+1}\,,\ Z_j^{(i_1,\dots,i_{k+1})} \leq Z_{i_{k+1}}^{(i_1,\dots,i_k)} \\
&\Leftrightarrow \forall j \neq i_1, \dots, i_{k+1}\,,\ Z_j^{(i_1,\dots,i_k)} - Z_{i_{k+1}}^{(i_1,\dots,i_k)} \frac{\tau_{j,i_{k+1}}}{\tau_{i_{k+1},i_{k+1}}} \leq Z_{i_{k+1}}^{(i_1,\dots,i_k)} - Z_{i_{k+1}}^{(i_1,\dots,i_k)} \frac{\tau_{j,i_{k+1}}}{\tau_{i_{k+1},i_{k+1}}} \\
&\Leftrightarrow \forall j \neq i_1, \dots, i_{k+1}\,,\ Z_j^{(i_1,\dots,i_k)} \leq Z_{i_{k+1}}^{(i_1,\dots,i_k)} \qquad (27) \\
&\Leftrightarrow \lambda_{k+1}^{(i_1,\dots,i_k)} = Z_{i_{k+1}}^{(i_1,\dots,i_k)}\,.
\end{aligned}$$

using that (23) and that $1 - \tau_{j,i_{k+1}}/\tau_{i_{k+1},i_{k+1}} > 0$ (which is a consequence of (22) and ($\widehat{\mathcal{A}}_{\text{Irr.}}$)) in (27).

Step 2

Using the version 3 of the LAR algorithm we deduce directly the expressions of the expectation and variance of

$$Z_{i_\ell}^{(i_1,\dots,i_{\ell-1})}$$

given by (13) (15) (Changing k into ℓ). From this same version of this algorithm we get the independence of the $Z_{i_\ell}^{(i_1,\dots,i_{\ell-1})}$ when ℓ varies.

Now suppose that we are conditional to $(a, \lambda_a, \lambda_{k+1}, \widehat{\imath}_1, \dots, \widehat{\imath}_{K+1})$, then standard properties of order statistics imply that the density of $(\lambda_{a+1}, \dots, \lambda_k)$. taken at $(\ell_{a+1}, \dots, \ell_k)$ is proportional to

$$\left(\prod_{k=a+1}^{K} \varphi_{m_k, v_k^2}(\ell_k)\right) \mathbb{1}_{\{\ell_{a+1} \geq \ell_{a+2} \geq \cdots \geq \ell_K \geq \lambda_{K+1}\}}. \qquad (28)$$

Step 3

Suppose now that we are conditional to $\{\widehat{m} = a, \lambda_a, \lambda_{K+1}, \widehat{\imath}_1, \dots, \widehat{\imath}_K, \widehat{\imath}_{K+1}\}$ with $0 \le a \le K-1$. Because of $\mathcal{P}_2$, $\lambda_{a+1}, \dots, \lambda_K$ are independent of the condition $m = a$, so their conditional distribution is equal to the unconditional distribution given by (28).

C.2. Proof of Proposition 2

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

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

Now, observe that

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

showing the equivalence between the two assumptions.

C.3. Proof of Lemma 7

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

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

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

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

By induction, using (21), we get that

$$\begin{aligned}
N^{(k-1)} &= N^{(k-2)} - (\lambda_{k-2} - \lambda_{k-1})\theta(\widehat{\imath}_1, \dots, \widehat{\imath}_{k-2})\,, \\
&= (N^{(k-2)} - \lambda_{k-2}\theta(\widehat{\imath}_1, \dots, \widehat{\imath}_{k-2})) + \lambda_{k-1}\theta(\widehat{\imath}_1, \dots, \widehat{\imath}_{k-2})\,, \\
&= Z - P_{\widehat{\imath}_1,\dots,\widehat{\imath}_{k-2}}(Z) + \lambda_{k-1}\theta(\widehat{\imath}_1, \dots, \widehat{\imath}_{k-2})\,. \qquad (29)
\end{aligned}$$

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

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

Using (29) at indices $j = \widehat{\imath}_1, \dots, \widehat{\imath}_{k-1}$, we deduce that

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

Namely

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

Using again (29) we get that

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

as claimed.

C.4. Proof of Proposition 9

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

$$\begin{aligned} x &= \frac{R_{j,i_k} - R_j^\top M^{-1} R_{i_k}}{R_{i_k,i_k} - R_{i_k}^\top M^{-1} R_{i_k}}\,, \\ \overline{M}^{-1} &= \begin{bmatrix} \mathrm{Id}_{k-1} & -M^{-1}R_{i_k} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M^{-1} & 0 \\ 0 & \big(R_{i_k,i_k} - R_{i_k}^\top M^{-1} R_{i_k}\big)^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{Id}_{k-1} & 0 \\ -R_{i_k}^\top M^{-1} & 1 \end{bmatrix}\,, \\ \overline{M}^{-1}\overline{R} &= \begin{bmatrix} M^{-1}\big(R_j - xR_{i_k}\big) \\ x \end{bmatrix}\,, \end{aligned} \tag{30}$$

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

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

To prove (23), it suffices to show that the R.H.S term above is equal to the following R.H.S term

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

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

$$\begin{aligned} &1-\theta_j(i_1,\dots,i_{k-1}) - x(1-\theta_{i_k}(i_1,\dots,i_{k-1})) \\ &= 1-\theta_j(i_1,\dots,i_{k-1}) - x + x\,\theta_{i_k}(i_1,\dots,i_{k-1}) \\ &= 1 - (\underbrace{1\cdots 1}_{k \text{ times}}) \begin{bmatrix} M^{-1}\big(R_j - xR_{i_k}\big) \\ x \end{bmatrix} \\ &= 1-\theta_j(i_1,\dots,i_k)\,, \end{aligned} \tag{31}$$

using (30). Furthermore, it proves (22). For the numerators, we use that

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

using (30).

C.5. Proof of Proposition 10

The proof of the first point can be lead by induction. The initialization of the proof is given by the first point of Lemma 8. Now, observe that $Z_{i_k}^{(i_1,\dots,i_{k-1})},\cdots,Z_{i_2}^{(i_1)},Z_{i_1}$ are measurable functions of $(Z_{i_1},\dots,Z_{i_k})$ and one may check that the vector $(Z_{i_1},\dots,Z_{i_k})$ is independent of the vector $(Z_{i_{k+1}}^{(i_1,\dots,i_k)},(Z_j^{(i_1,\dots,i_{k+1})})_{j\neq i_1,\dots,i_{k+1}})$. We deduce that $Z_{i_k}^{(i_1,\dots,i_{k-1})},\cdots,Z_{i_2}^{(i_1)},Z_{i_1}$ are independent of $(Z_{i_{k+1}}^{(i_1,\dots,i_k)},(Z_j^{(i_1,\dots,i_{k+1})})_{j\neq i_1,\dots,i_{k+1}})$. One can also check that $Z_{i_{k+1}}^{(i_1,\dots,i_k)}$ is independent of $(Z_j^{(i_1,\dots,i_{k+1})})_{j\neq i_1,\dots,i_{k+1}}$.

The second point also works by induction. The initialization of the proof is given by the second point of Lemma 8. We will use Proposition 9 to prove the second point.
Now, we have

$$\begin{aligned} & \lambda_{k+1}^{(i_1,\dots,i_k)} \le Z_{i_k}^{(i_1,\dots,i_{k-1})} \\ & \Leftrightarrow \forall j\neq i_1,\dots,i_k\,,\ Z_j^{(i_1,\dots,i_k)} \le Z_{i_k}^{(i_1,\dots,i_{k-1})} \\ & \Leftrightarrow \forall j\neq i_1,\dots,i_k\,,\ Z_j^{(i_1,\dots,i_{k-1})} - Z_{i_k}^{(i_1,\dots,i_{k-1})}\frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} \le Z_{i_k}^{(i_1,\dots,i_{k-1})} - Z_{i_k}^{(i_1,\dots,i_{k-1})}\frac{\tau_{j,i_k}}{\tau_{i_k,i_k}} \\ & \Leftrightarrow \forall j\neq i_1,\dots,i_k\,,\ Z_j^{(i_1,\dots,i_{k-1})} \le Z_{i_k}^{(i_1,\dots,i_{k-1})} \qquad (32) \\ & \Leftrightarrow \lambda_k^{(i_1,\dots,i_{k-1})} = Z_{i_k}^{(i_1,\dots,i_{k-1})}\,. \end{aligned}$$

using that (23) and that $1-\tau_{j,i_k}/\tau_{i_k,i_k}>0$ (which is a consequence of (22) and ($\widehat{\mathcal{A}}_{\text{Irr.}}$)) in (32). By induction and using (32), it holds that

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

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

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

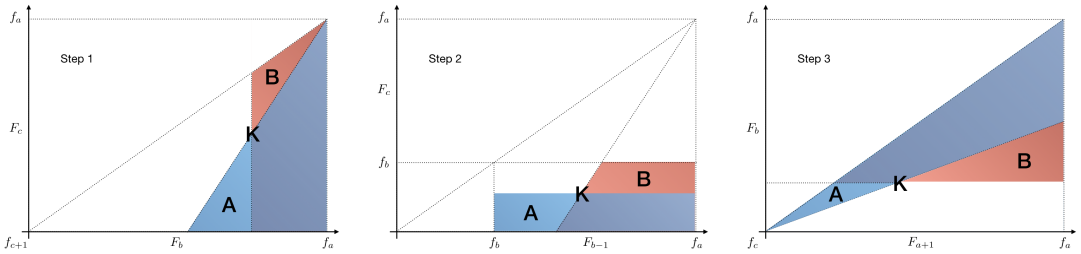

Figure 5. *Rejection domains associated to the different comparison sets appearing in steps of the proof of Theorem 6.*

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

C.6. Orthogonal Case: Proof of Theorem 6

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 $(F_{a+1}, \ldots, F_c)$ has for distribution the uniform distribution on the simplex

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

This implies by direct calculations that

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

and that

$$\frac{\mathbb{F}_{abc}(\lambda_b)}{\mathbb{F}_{abc}(\lambda_a)} = \boldsymbol{F}_{\beta((b-a),(c-b))}\left(\frac{F_b - F_c}{f_a - F_c}\right). \tag{33}$$

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

$$(F_b - F_c) \geq z_1(f_a - F_c) \Leftrightarrow F_b \geq z_1 f_a + (1 - z_1)F_c, \tag{34}$$

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

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

$$F_b \geq z_2(f_a - f_{c+1}) + f_{c+1}, \tag{35}$$

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

Lemma 11. *Let $c \leq K$. The density h_μ of $f_1, \ldots, f_c$, conditional to F_{c+1} with respect of the Lebesgue measure under the alternative is coordinate-wise non-decreasing and given by (36).*

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

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

Thus its density $p_{\mu^0, i_1, \ldots, i_c}$ does not depend on $\mu^0_{\bar{i}_1}, \ldots, \mu^0_{\bar{i}_c}$. As a consequence the following variables have the same distribution ; $\lambda_{c+1}^{i_1+\epsilon_1 p, \ldots, i_c+\epsilon_c p}$, where $\epsilon_1, \ldots, \epsilon_c$ 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

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

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

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

$$(\text{const}) \sum\nolimits' \cosh(\mu_{j_1} f_1) \dots \cosh(\mu_{j_c} f_c) \mathbb{1}_{f_1 > \dots > f_c > F_{c+1}}, \tag{36}$$

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

Lemma 12. *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 (34) and $\mathcal{R}_2$ associated to (35) 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 the 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

$$\int_{f_b}^{f_a} df_{a+1} \dots \int_{f_b}^{f_b - 2} df_{a+1} \int_{f_b}^{f_b - 2} df_{b-1} \int_{f_c}^{f_b} df_{b+1} \dots \int_{f_c}^{f_{c-2}} df_{c-1} h_{\mu^0}(f_a, \dots, f_c). \tag{37}$$

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 (37).

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 11 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 to $F_a = f_a, F_b = f_b$ 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 12 imply that the power of $\mathcal{S}_{a,(b-1),c}$ is greater of equal that that of $\mathcal{S}_{a,b,c}$.

$\circ$ **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 to $F_a = f_a, F_c = f_c$ 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 12 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.7. Proof of Theorem 1

We rely on the **Weak Positive Regression Dependency (WPRDS) property** to prove the result, one may consult [17, 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 any $k = 1, \ldots, K$, 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 $(\widehat{p}_1 = \widehat{\alpha}_{0,1,K+1}, \ldots, \widehat{p}_K = \widehat{\alpha}_{K-1,K,K+1})$ satisfy the WPRDS if for any nondecreasing set Γ and for all $k^0 \in I_0$, the function

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

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 1, note that it is sufficient [17, Chapter 8] to prove that

$$u \mapsto \overline{\mathbb{P}}\big[(\widehat{p}_1, \ldots, \widehat{p}_K) \in \Gamma \big| \widehat{p}_{k^0} \leq u\big] \text{ is nondecreasing} \tag{38}$$

where $\overline{\mathbb{E}}, \overline{\mathbb{P}}$ will denote that expectations and probabilities are conditional on $\{\bar{\imath}_1, \ldots, \bar{\imath}_K, \lambda_{K+1}\}$ 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 1.

$\circ$ **Step 1:** We start by giving the joint law of the LAR's knots under the alternative in the orthogonal design case. Lemma 11 and (36) show that, conditional on $\{\bar{\imath}_1, \ldots, \bar{\imath}_K, \lambda_{K+1}\}$, $(\lambda_1, \ldots, \lambda_K)$ is distributed on the set $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq \lambda_{K+1}$ 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 $\{\bar{\imath}_1, \ldots, \bar{\imath}_K, F_{K+1}\}$, $(F_1, \ldots, F_K)$ is distributed on the set

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

it has an **explicit density** given by (36), 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 $(G_1, \ldots, G_K)$ is supported on $[0,1]^K$. More precisely, define

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

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

$$\mathbf{p}(g):=(\text{const})\mathbb{1}_{g\in(0,1)^K}\prod_{k=1}^{K}g_k^{K-k}\cosh\big[\mu^0_{\bar\imath_k}((1-F_{K+1})\prod_{\ell=1}^{k}g_\ell+F_{K+1})\big]\,,\tag{39}$$

where we have used (36). From (18) and (33), one has

$$\widehat{p}_k=1-\boldsymbol{F}_{\beta(1,K-k+1)}\Big(\frac{F_k-F_{K+1}}{F_{k-1}-F_{K+1}}\Big)=1-\boldsymbol{F}_{\beta(1,K-k+1)}(G_k)\tag{40}$$

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 $\ell\in[K]$,

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

so that

$$\overline{\mathbb{P}}\big[(\widehat{p}_1,\dots,\widehat{p}_K)\in\Gamma\big|\widehat{p}_{k^0}\le u\big]=\overline{\mathbb{P}}\big[(G_1,\dots,G_K)\in\overline{\Gamma}\big|G_{k^0}\ge\boldsymbol{F}^{-1}_{\beta(1,K-\ell+1)}(1-u)\big]\,,\tag{41}$$

where $\overline{\Gamma}$ can be proved to be a **nonincreasing Borel set** from (40).

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

$$\mu_x:=\text{law}\big[(G_1,\dots,G_K)|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}\ge x\}\big]\,.$$

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

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

then it holds

- $\mathbb{1}_{\{T(g)\in\overline{\Gamma}\}}\le\mathbb{1}_{\{g\in\overline{\Gamma}\}}$;
- $\text{law}\big[T(G)|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}\ge x\}\big]=\text{law}\big[G|\{\bar\imath_1,\dots,\bar\imath_K,F_{K+1},G_{k^0}\ge y\}\big]$ where $G=(G_1,\dots,G_K)$.

In this case, we deduce that

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

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

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

and, in view of (41), it proves (38). 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

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

where $T^{(k^0)}, T^{(k^0+1)}, \dots, T^{(K)}, T^{(k^0-1)} \dots, T^{(1)}$ 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

- 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

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

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 (39), 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

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

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

$$\mu_{z_{k^0}}^{(k^0+1)} := \text{law}\big[G_{k^0+1} | \{\bar{\imath}_1, \dots, \bar{\imath}_K, F_{K+1}, G_{k^0} = z_{k^0}\}\big]\,,$$

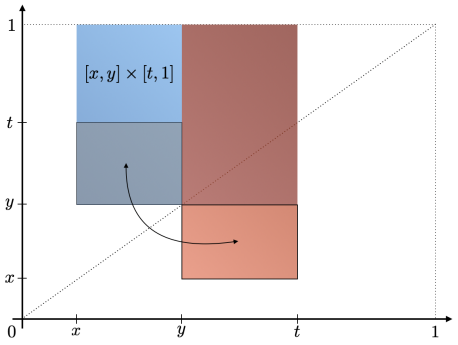

Figure 6. *Note that, by symmetry the two boxed regions have same* $\mathbf{p}\otimes\mathbf{p}$ *measure. The blue region is* $[x,t]\times[y,1]$*, its measure is the measure of the red region (namely* $cD_y(t)\times\mathcal{D}_x(1)$*) more the bluest upper left corner (namely* $[x,y]\times[t,1]$*).*

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 $x<z_{k^0}\le z'_{k^0}=T^{(k^0)}(z_{k^0})\le 1$. Again, we would like to prove that $\mathbb{F}^{(k^0+1)}_{z_{k^0}}\ge\mathbb{F}^{(k^0+1)}_{z'_{k^0}}$ 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 $T^{(k^0+1)}(u)\ge u$ 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 (39) 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

$$\begin{aligned}(\mathrm{const})\,\mathbb{1}_{g\in(0,1)^K}\mathbb{1}_{g_{k^0}=\mathbf{z}}&\prod_{k<k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\prod_{\ell=1}^{k}g_\ell+F_{K+1}\big)\Big]\\&\times\prod_{k>k^0}g_k^{K-k}\cosh\Big[\mu^0_{\bar\imath_k}\big((1-F_{K+1})\,\mathbf{z}\prod_{1\le\ell\ne k^0\le k}g_\ell+F_{K+1}\big)\Big]\,.\end{aligned}\tag{42}$$

Set $\tau:=z'_{k^0}/z_{k^0}\ge1$ 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

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

which, up to some normalising constant, is the same as (42) 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

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

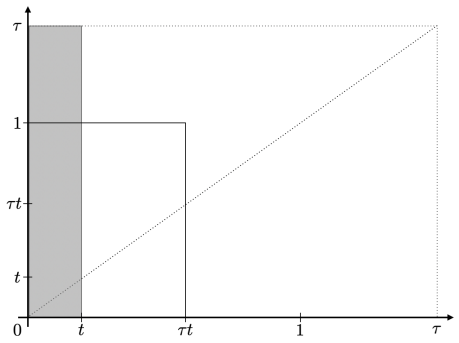

Figure 7. *The two boxed rectangles have Lebesgue measure, namely τt. The $\mathbf{p}\otimes\mathbf{p}$ measure of the grey box is greater than the $\mathbf{p}\otimes\mathbf{p}$ measure of the white box.*

We deduce that

$$\begin{aligned}
\mathbb{F}^{(k^0+1)}_{z_{k^0}}(t) \geq \mathbb{F}^{(k^0+1)}_{z'_{k^0}}(t) \quad &\Leftrightarrow \quad \overline{\mathbb{P}}(G_{k^0+1} \leq t | G_{k^0} = z_{k^0}) \geq \overline{\mathbb{P}}(G_{k^0+1} \leq t | G_{k^0} = z'_{k^0}) \\
&\Leftrightarrow \quad \overline{\mathbb{P}}(G_{k^0+1} \leq t | G_{k^0} = z_{k^0}) \geq \overline{\mathbb{P}}(G'_{k^0+1} \leq \tau t | G_{k^0} = \tau z_{k^0}) \\
&\Leftrightarrow \quad \frac{\int_{\mathcal{D}(t)} \mathbf{p}}{\int_{\mathcal{D}(1)} \mathbf{p}} \geq \frac{\int_{\mathcal{D}(\tau t)} \mathbf{p}}{\int_{\mathcal{D}(\tau)} \mathbf{p}} \\
&\Leftrightarrow \quad \int_{\mathcal{D}(t)\times\mathcal{D}(\tau)} \mathbf{p}\otimes\mathbf{p} \geq \int_{\mathcal{D}(\tau t)\times\mathcal{D}(1)} \mathbf{p}\otimes\mathbf{p}\,, \qquad (43)
\end{aligned}$$

where

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

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 13. *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 (43). 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

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

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 (43) 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

$$\begin{aligned} w'(h_1', h_2') =& \mathrm{C}_1 {h_1'}^{K-k^0-1} \cosh\left[\mu^0_{\bar\imath_k}((1-F_{K+1})\ z_{k^0} \prod_{1\le \ell<k^0} g_\ell \times t \times h_1' + F_{K+1})\right] \\ & \times \prod_{k>k^0+1} \cosh\left[\mu^0_{\bar\imath_k}((1-F_{K+1})\ z_{k^0} \prod_{1\le \ell\neq k^0, k^0+1\le k} g_\ell \times t \times h_1' + F_{K+1})\right] \\ & \times {h_2'}^{K-k^0-1} \cosh\left[\mu^0_{\bar\imath_k}((1-F_{K+1})\ z_{k^0} \prod_{1\le \ell<k^0} g_\ell \times h_2'/t + F_{K+1})\right] \\ & \times \prod_{k>k^0+1} \cosh\left[\mu^0_{\bar\imath_k}((1-F_{K+1})\ z_{k^0} \prod_{1\le \ell\neq k^0, k^0+1\le k} g_\ell \times h_2'/t + F_{K+1})\right]. \end{aligned}$$

Now, invoke Lemma 13 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.

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

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

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

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

and $\mathbb{F}_{z'}^{(k+1)}$ its cdf. Note that $z \le z' = T^{(k)}(z) \le 1$. Again, we would prove that $\mathbb{F}_z^{(k+1)} \ge \mathbb{F}_{z'}^{(k+1)}$ which implies that the transport map $T^{(k+1)} := \big(\mathbb{F}_{z'}^{(k+1)}\big)^{-1} \circ \mathbb{F}_z^{(k+1)}$ 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 $G|\{\bar\imath_1, \dots, \bar\imath_K, F_{K+1}, G^{[k,k^0]} = \mathbf{z}\}$ is

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

Set $\tau := \prod_{n=k^0}^{k} z'_n / \prod_{n=k^0}^{k} z_n \geq 1$ 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 $z' := (z'_K, \dots, z'_{k^0})$ and $z := (z_K, \dots, z_{k^0})$. Denote $\mu_z^{(k^0-1)}$ the following conditional law

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

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

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

and $\mathbb{F}_{z'}^{(k^0 1)}$ its cdf. Note that $z \leq z' = T^{(K)}(z) \leq 1$. Again, we would prove that $\mathbb{F}_z^{(k^0-1)} \geq \mathbb{F}_{z'}^{(k^0-1)}$ 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 $T^{(k^0-1)}(u) \geq u$ for all $u \in (0,1)$.

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

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

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

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

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

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

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

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

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

Lemma 14. *Let $0 < a \leq a'$ 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 $g \leq g'$. From Lemma 14, we deduce that $\cosh(ag+b)\cosh(a'g'+b) \geq \cosh(ag'+b)\cosh(a'g+b)$, proving that $w_z(g)w_{z'}(g') \geq w_{z'}(g)w_z(g')$. It proves that $T^{(k^0-1)}(u) \geq u$ 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*.

sectionProof of Theorem 5 From Corollary 4 we deduce that, under $\mathbb{H}_0$ and conditionally on the selection event $\left\{\widehat{m}=a\right\}$ with $a \leq K-1$, (17) remains true, giving the conditional cumulative distribution function of λ_b. As a consequence and under the same conditioning.

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

Since the conditional distribution doesn't depend on the condition, it is also the unconditional distribution. 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.

Appendix D: Main notation

$[a]$	the set of integers $\{1,...,a\}$,
X	a full rank $n \times p$ design matrix,
σ^2	the variance of the errors,
$\bar{\imath}_k; \varepsilon_k$	the indexes and the signs of the variables that enter in the LAR path,
$\widehat{\imath}_k$	a way of coding both indexes and signs, see (1),
$\overline{S}^k$	$\{\bar{\imath}_1, \ldots, \bar{\imath}_k\}$, a possible selected support,
S_0	the true support,
E_k	$\mathrm{Span}(X_{\bar{\imath}_1}, \ldots, X_{\bar{\imath}_k})$,
$P_k(P_k^\perp)$	projector on (the orthogonal of) E_k,
$\gamma_{m_k,v_k^2}(\phi_{m_k,v_k^2})$	The normal distribution (density) with mean m_k given (13) and variance $v_k^2 = \sigma^2\rho_k^2$ given by (15),
Z	the vector of correlations, obtained by sym. from $\overline{Z}$ defined by (9),
R	the variance-covariance matrix of Z, see (10),
K	the number of knots that are considered, see Figure 2, and Section 2.5
K_{select}	see Figure 2
$\widehat{m}$	the chosen size ,
$\widehat{S}$	the chosen set of variables : $\overline{S}^{\widehat{m}}$,
ρ_k^2	v_k^2/σ^2, see (13),
$\theta_j(i_1,\ldots,i_k)$	see (11),
θ^ℓ	$\theta(\widehat{\imath}_1,\ldots,\widehat{\imath}_\ell)$,
$M_{i_1,\ldots,i_\ell}$	the submatrix of R obtained by keeping the columns, and the rows indexed by $\{i_1,\ldots,i_\ell\}$,
$F_i; \mathcal{P}_{ij}$	$F_i := \Phi_i(\lambda_i) := \Phi(\lambda_i/(\sigma\rho_i))$ and $\mathcal{P}_{ij}$ is given by (14),
$\mathbb{F}_{abc}(t)$	see (16)
$\widehat{\alpha}_{abc}$	the p-value of the *generalized spacing test*, see (18),
$\mathcal{S}_{abc}$	$\mathbb{1}_{\{\widehat{\alpha}_{abc} \leq \alpha\}}$, the *generalized spacing test* see (19).

Appendix E: Cubature of integral 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

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

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 estimation 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

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

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.