

# Power of the spacing test for least-angle regression

JEAN-MARC AZAÏS<sup>1,\*</sup>, YOHANN DE CASTRO<sup>2</sup>  
and STÉPHANE MOURAREAU<sup>1,\*\*</sup>

<sup>1</sup>*Institut de Mathématiques de Toulouse, Université Paul Sabatier, Route de Narbonne, F-31062 Toulouse, France. E-mail: \*[jean-marc.azais@math.univ-toulouse.fr](mailto:jean-marc.azais@math.univ-toulouse.fr); \*\*[stephane.mourareau@math.univ-toulouse.fr](mailto:stephane.mourareau@math.univ-toulouse.fr)*

<sup>2</sup>*Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, F-91405 Orsay, France. E-mail: [yohann.decastro@math.u-psud.fr](mailto:yohann.decastro@math.u-psud.fr)*

Recent advances in Post-Selection Inference have shown that conditional testing is relevant and tractable in high-dimensions. In the Gaussian linear model, further works have derived unconditional test statistics such as the Kac–Rice Pivot for general penalized problems. In order to test the global null, a prominent offspring of this breakthrough is the Spacing test that accounts the relative separation between the first two knots of the celebrated least-angle regression (LARS) algorithm. However, no results have been shown regarding the distribution of these test statistics under the alternative. For the first time, this paper addresses this important issue for the Spacing test and shows that it is unconditionally unbiased. Furthermore, we provide the first extension of the Spacing test to the frame of unknown noise variance.

More precisely, we investigate the power of the Spacing test for LARS and prove that it is unbiased: its power is always greater or equal to the significance level  $\alpha$ . In particular, we describe the power of this test under various scenarios: we prove that its rejection region is optimal when the predictors are orthogonal; as the level  $\alpha$  goes to zero, we show that the probability of getting a true positive is much greater than  $\alpha$ ; and we give a detailed description of its power in the case of two predictors. Moreover, we numerically investigate a comparison between the Spacing test for LARS, the Pearson's chi-squared test (goodness of fit) and a numerical testing procedure based on the maximal correlation.

When the noise variance is unknown, our analysis unleashes a new test statistic that can be computed in cubic time in the population size and which we refer to as the  $t$ -Spacing test for LARS. The  $t$ -Spacing test involves the first two knots of the LARS algorithm and we give its distribution under the null hypothesis. Interestingly, numerical experiments witness that the  $t$ -Spacing test for LARS enjoys the same aforementioned properties as the Spacing test.

**Keywords:** hypothesis testing;  $\ell_1$ -minimization; power; spacing test

## 1. Introduction

A major development in modern statistics has been brought by the idea that one can recover a high-dimensional target  $\beta^*$  from few linear observations  $Y$  by  $\ell_1$ -minimization as soon as the target vector is “sparse” in a well-chosen basis. Undoubtedly, the notion of “sparsity” has encountered a large echo among the statistical community and many successful applications rely on  $\ell_1$ -minimization, one may consult [8–10,13,27] for some seminal works, [7,15] for a review and references therein. More precisely, some of the most popular estimators in high-dimensional statistics remain the Lasso [27] and the Dantzig selector [8]. A large amount of

interest has been dedicated to the estimation, prediction or support recovery problems using these estimators. This body of work has been developed around sufficient conditions on the design matrix  $X$  (such that *Restricted Isometry Property* [9], *Restricted Eigenvalue* [5], *Compatibility* [7,28], *Universal Distortion* [4,11],  $\mathbf{H}_{s,1}$  [17], or *Irrepresentability* [13], to name but a few) that enclose the spectral properties of the design matrix on the set of (almost) sparse vectors. Using one of these properties, one can exploits the Karush–Kuhn–Tucker conditions to get oracle inequalities or a control on the support recovery error.

Aside from those issues some recent works have been focused on hypothesis testing using penalized problems, see, for instance, [18,19,25,26] and references therein. Compared to the sparse recovery problems, very little work has been done in statistical testing in high dimensions. As a matter of fact, one of the main difficulty is that there is no tractable distribution of sparse estimators (even under the aforementioned standard conditions of high-dimensional statistics). A successful approach is then to take into account the influence of each predictor in the regression problem. More precisely, some recent works in Post-Selection Inference have shown that the selection events can be explicitly expressed as closed convex polytopes depending simply on the signs and the indices of the nonzero coefficients of the solutions of standard procedures in high-dimensional statistics (typically the solutions of the Lasso). Furthermore, an important advance has been brought by a useful parametrization of these convex polytopes under the Gaussian linear model, see for instance the book [16]. In detection testing, this is done by the first two “*knots*” of the *least-angle regression algorithm* (LARS for short) which is intimately related to the dual program of the  $\ell_1$ -minimization problem, see, for example, [12].

## 1.1. Hypothesis testing using LARS

The usual frame of the regression problems in high-dimensions is the following. Given an outcome vector  $Y \in \mathbb{R}^n$ , a matrix of predictor variables (or design matrix)  $X \in \mathbb{R}^{n \times p}$  and a variance–covariance matrix  $\Sigma$  such that

$$Y = X\beta^* + \xi \quad \text{with } \xi \sim \mathcal{N}_n(0, \Sigma),$$

we are concerned with testing whether  $\beta^*$  is equal to some known  $\beta_0^*$  or not. Notice that the response variable  $Y$  does not depend directly on  $\beta^*$  but rather on  $X\beta^*$ . We understand that a detection test may be interested in discerning between two hypothesis on the target vector, namely

$$\mathbb{H}_0 : “\beta^* \in \beta_0^* + \ker(X)” \quad \text{against} \quad \mathbb{H}_1 : “\beta^* \notin \beta_0^* + \ker(X)”,$$

where  $\ker(X)$  denotes the kernel of the design matrix  $X$ . It can be equivalently formulated (subtracting  $X\beta_0^*$ ) as a detection test whose null hypothesis is given by

$$\mathbb{H}_0 : “\beta^* \in \ker(X)” \quad \text{against} \quad \mathbb{H}_1 : “\beta^* \notin \ker(X)”.$$

To this end, we consider the vector of correlations

$$U := X^\top Y \sim \mathcal{N}_p(\mu^*, R),$$

where  $\mu^\star := X^\top X \beta^\star$  and  $R := X^\top \Sigma X$ . Observe that the hypotheses  $\mathbb{H}_0$  and  $\mathbb{H}_1$  can be equivalently written as

$$\mathbb{H}_0 : “\mu^\star = 0” \quad \text{against} \quad \mathbb{H}_1 : “\mu^\star \neq 0”,$$

and remark that the knowledge of the noise variance–covariance matrix  $\Sigma$  is equivalent to the knowledge of the correlations variance–covariance matrix  $R$ .

## 1.2. The spacing test for LARS

The test statistic we are considering was introduced in a larger context of penalization problems by the pioneering works in [25,26]. As mentioned by the authors of [26], the general test statistic “may seem complicated”. However, it can be greatly simplified in the frame of the standard regression problems under a very mild assumption, namely

$$\forall i \in \llbracket 1, p \rrbracket, \quad R_{ii} = X_i^\top \Sigma X_i =: \langle X_i, X_i \rangle_\Sigma = 1. \quad (\text{H})$$

Note that this assumption is not very restrictive because the columns  $X_i$  of  $X$  can always be scaled to get (H). In this case, the entries of  $\beta^\star$  are scaled but nor  $\mathbb{H}_0$  neither  $\mathbb{H}_1$  are changed. Hence, without loss of generality, we admit to invoke an innocuous normalization on the columns of the design matrix. Remark also that (H) is satisfied under the stronger assumption

$$\Sigma = \text{Id}_n \quad \text{and} \quad \forall i \in \llbracket 1, p \rrbracket, \quad \|X_i\|_2^2 = 1. \quad (\text{H Lasso})$$

Moreover, observe that, almost surely, there exists a unique  $(\hat{i}, \hat{\varepsilon}) \in \llbracket 1, p \rrbracket \times \{\pm 1\}$  such that  $\hat{\varepsilon} U_{\hat{i}} = \|U\|_\infty$ . Under Assumption (H), the test statistic, referred to as *Spacing test for LARS*, simplifies to

$$S := \frac{\bar{\Phi}(\lambda_1)}{\bar{\Phi}(\lambda_2)}, \quad (\text{Pivot})$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution,  $\bar{\Phi} = 1 - \Phi$  its complement,  $\lambda_1 := \hat{\varepsilon} U_{\hat{i}}$  the largest knot in the *Lasso path* [12] and

$$\lambda_2 := \bigvee_{1 \leq j \neq \hat{i} \leq p} \left\{ \frac{U_j - R_{j\hat{i}} U_{\hat{i}}}{1 - \hat{\varepsilon} R_{j\hat{i}}} \vee \frac{-U_j + R_{j\hat{i}} U_{\hat{i}}}{1 + \hat{\varepsilon} R_{j\hat{i}}} \right\},$$

with  $a \vee b := \max(a, b)$  and  $U_i$  denotes the  $i$ th entry of the vector  $U$ . Under Assumption (H Lasso), one has  $R = X^\top X$  and  $\lambda_2$  simplifies to the second largest knot in the *Lasso path*. Interestingly, the authors of [26] have shown that the test statistic  $S$  is uniformly distributed on  $[0, 1]$  under the null hypothesis  $\mathbb{H}_0$ ,

$$S \sim \text{Unif}([0, 1]).$$

Moreover, they derived the following rejection region

$$\text{Reject}_\alpha := \{S \leq \alpha\},$$

for all  $\alpha \in (0, 1)$ . In other words, the observed value of the test statistic  $S$  is the  $p$ -value of the Spacing test for LARS.

Remark that the statistic  $1 - S$  is uniformly distributed on  $[0, 1]$ , as well as many other transformations of the test statistic  $S$ . It may appear that the choice of rejection region  $\text{Reject}_\alpha$  is somehow arbitrary. Nevertheless, one can empirically witness (see Figure 1 for instance) that the *Spacing test for LARS* is an interesting test statistic that may take smaller values under the alternative hypothesis. However, no theoretical guarantees have been shown regarding its power. Furthermore, the *Spacing test for LARS* relies on the assumption that the variance–covariance matrix  $\Sigma$  of the noise is known and it should be interesting to bypass this limitation. To the best of our knowledge, this paper is the first to address these issues.

### 1.3. Power of the spacing test for LARS

Recall that the Spacing test for LARS rejects  $\mathbb{H}_0$  in favor of  $\mathbb{H}_1$  when  $\{S \leq \alpha\}$  occurs, where  $S$  is defined by (Pivot). We assume that the noise variance–covariance matrix  $\Sigma$  is known. We also assume that the columns  $(X_i)_{i=1}^p$  of the matrix  $X$  are normalized with respect to Assumption (H) and satisfy  $\forall i \neq j, |\langle X_i, X_j \rangle_\Sigma| < 1$ . The first result shows that the Spacing test for LARS is unbiased.

**Theorem 1.** *Let  $\alpha \in (0, 1)$  be a significance level. Assume that the variance–covariance matrix  $\Sigma$  of the noise is known and assume that Assumption (H) holds. Then, the Spacing test for LARS is unbiased: its power under the alternative is always greater or equal to the significance level  $\alpha$ .*

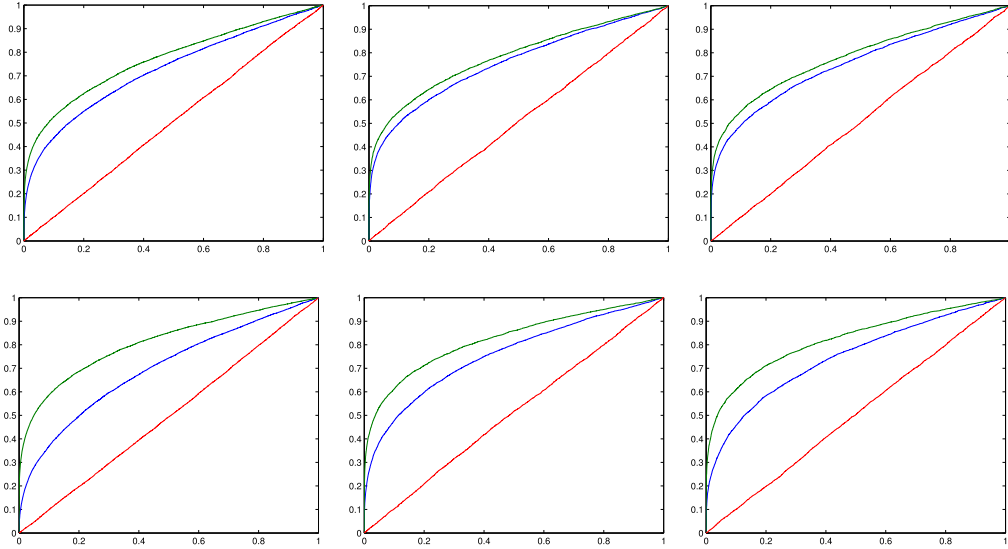
Under mild assumptions, this theorem ensures that the probability of getting a *true positive* is greater or equal to the probability of a *false positive*. Moreover, in the limit case when the significance level  $\alpha$  goes to zero, this result is refined by Theorem 5: the probability of a *true positive* is much greater than the probability of getting a *false positive*. As a matter of fact, we prove that the cumulative distribution function of  $S$  has a vertical tangent at the origin under the alternative hypothesis. The reader may consult Figure 1 which represents the empirical distribution function of  $S$  that exactly describes the uniform law.

A proof of Theorem 1 can be found in Section 2.3. Interestingly, our proof is based on Anderson’s inequality [1] for symmetric convex sets. Moreover, we derive a simple and short proof of the distribution of the test statistic (Pivot) under the null, see Corollary 1 of Proposition 4.

Theorem 1 has a stronger version in the case of orthogonal designs, for example, when the variance–covariance matrix  $\Sigma$  is  $\text{Id}_n$  and  $X^\top X = \text{Id}_p$  (which implies that  $n \geq p$ ).

**Theorem 2 (Orthogonal design).** *Assume that  $R = \text{Id}_n$  then, under any alternative in  $\mathbb{H}_1$ , the density function of  $S$  is decreasing. Hence, for all significance level  $\alpha \in (0, 1)$ , the region  $\text{Reject}_\alpha = \{S \leq \alpha\}$  is the most powerful region among all possible regions.*

This theorem may be seen as an evidence in favor of the choice of the rejection region as  $\text{Reject}_\alpha = \{S \leq \alpha\}$ . A proof of Theorem 2 can be found in Section 2.4.



**Figure 1.** On each figure, empirical distribution function of 15 000  $p$ -values coming from various scenarii. 5000  $p$ -values drawn under the null (red), 5000  $p$ -values of  $S$  under the alternative (green) and 5000  $p$ -values of  $T$  under the alternative (blue). At the top, the level of sparsity  $s$  is equal to 2. At the bottom, the sparsity  $s$  is 5. In both cases, from left to right,  $(n, p) = (50, 100)$ ,  $(100, 200)$  and  $(100, 500)$ .

## 1.4. Extension to unknown variance

Interestingly, we can derive from our analysis a *studentization* of the test statistic (Pivot). Indeed, we consider the test statistic

$$T := \frac{1 - \mathbb{F}_{m-1}(T_1)}{1 - \mathbb{F}_{m-1}(T_2)},$$

where  $2 \leq \text{rank}(X) =: m \leq n$  and  $\mathbb{F}_{m-1}$  denote the cumulative distribution function of the  $t$ -distribution with  $m - 1$  degrees of freedom. Note that  $T_1, T_2$  are statistics that can be computed in cubic time (cost of one Singular Value Decomposition (SVD) of the design matrix) from the first knots of the LARS algorithm, see Algorithm 1. In the sequel, for each  $i \in \llbracket 1, p \rrbracket$ , we may denote by  $X_{-i} \in \mathbb{R}^{n \times (p-1)}$  the sub-matrix of  $X$  where the  $i$ th column  $X_i$  has been deleted.

**Theorem 3 ( $t$ -Spacing test for LARS).** Assume that the variance-covariance matrix  $\Sigma$  is  $\sigma^2 \text{Id}_n$  where  $\sigma > 0$  is unknown and that for all  $i \neq j \in \llbracket 1, p \rrbracket$ , one has  $\|X_i\|_2 = 1$ ,  $|\langle X_i, X_j \rangle| < 1$  and the matrix  $X_{-i}$  has rank  $m := \text{rank}(X) \geq 2$ . Then, under the null  $\mathbb{H}_0$ , the statistic  $T$  described by Algorithm 1 is uniformly distributed on  $[0, 1]$ .

In particular, we derive a detection test of significance level  $\alpha$  considering the rejection region  $\text{Reject}_\alpha = \{T \leq \alpha\}$ . A proof of Theorem 3 can be found in Section 3. One can empirically witness

**Algorithm 1:** *t*-Spacing test

**Data:** An observation  $Y \in \mathbb{R}^n$  and a design matrix  $X \in \mathbb{R}^{n \times p}$  of rank  $m \geq 2$ .

**Result:** A  $p$ -value  $T \in (0, 1)$ .

Compute the first LARS knot  $\lambda_1$ ;

1. Set  $U := X^\top Y$ ;
2. Find  $(\hat{i}, \hat{\varepsilon}) \in \llbracket 1, p \rrbracket \times \{\pm 1\}$  such that  $\hat{\varepsilon}U_{\hat{i}} = \|U\|_\infty$  and set  $\lambda_1 := \hat{\varepsilon}U_{\hat{i}}$ ;

Compute the second LARS knot  $\lambda_2$ ;

3. Set  $R := X^\top X$ ;
4. Set  $\lambda_2 := \bigvee_{1 \leq j \neq \hat{i} \leq p} \left\{ \frac{U_j - R_{j\hat{i}}U_{\hat{i}}}{1 - \hat{\varepsilon}R_{j\hat{i}}} \vee \frac{-U_j + R_{j\hat{i}}U_{\hat{i}}}{1 + \hat{\varepsilon}R_{j\hat{i}}} \right\}$ ;

Compute the variance estimator  $\hat{\sigma}$ ;

5. Set  $R_{-\hat{i}} := X_{-\hat{i}}^\top (\Pi_X - X_{\hat{i}}X_{\hat{i}}^\top)X_{-\hat{i}}$  with  $\Pi_X$  the orthogonal projection onto the range of  $X$ ;
6. Compute  $R_{-\hat{i}}^{-1}$ , the pseudoinverse of  $R_{-\hat{i}}$ ;
7. Set  $\hat{\sigma}^2 := \frac{V_{-\hat{i}}^\top R_{-\hat{i}}^{-1} V_{-\hat{i}}}{m-1}$  where

$$V_{-\hat{i}} := (U_1 - R_{1\hat{i}}U_{\hat{i}}, \dots, U_{\hat{i}-1} - R_{(\hat{i}-1)\hat{i}}U_{\hat{i}}, U_{\hat{i}+1} - R_{(\hat{i}+1)\hat{i}}U_{\hat{i}}, \dots, U_p - R_{p\hat{i}}U_{\hat{i}});$$

Compute the  $p$ -value  $T$ ;

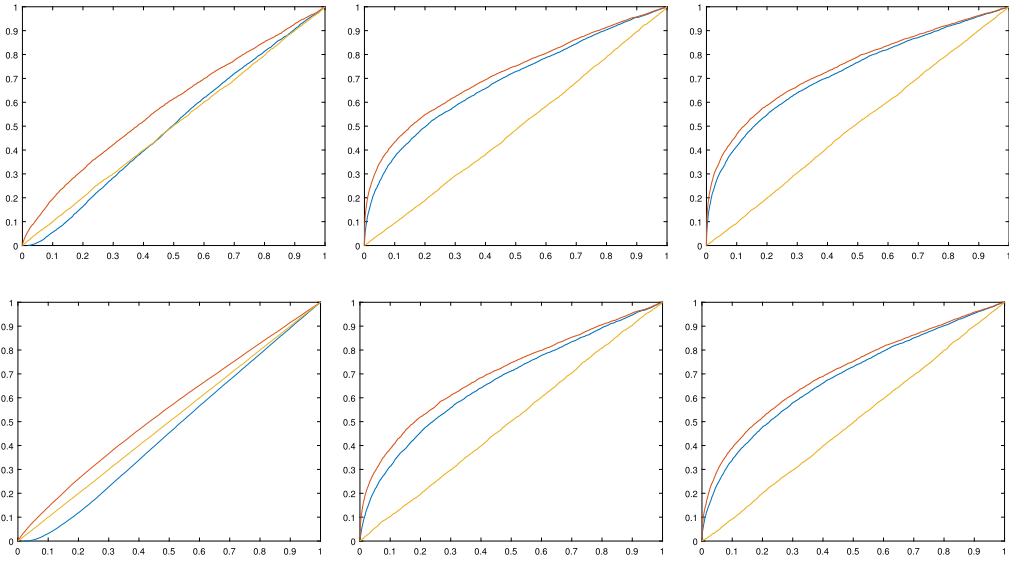
8. Set  $T_1 := \lambda_1/\hat{\sigma}$  and  $T_2 := \lambda_2/\hat{\sigma}$ ;
9. Set  $T := \frac{1 - \mathbb{F}_{m-1}(T_1)}{1 - \mathbb{F}_{m-1}(T_2)}$  where we denote by  $\mathbb{F}_{m-1}$  the cumulative distribution function of the  $t$ -distribution with  $m - 1$  degree(s) of freedom.

(see Figure 1, for instance) that the *t*-Spacing test for LARS is an interesting test statistic that may take smaller values under the alternative hypothesis.

Observe that Algorithm 1 requires the computation of one SVD at Step 6. We deduce that its computational cost is  $\mathcal{O}(p^3)$  which is reasonable in high-dimensional statistics.

**Remark 1.** In this article, we limit our attention to the hypothesis  $\Sigma = \sigma^2 \text{Id}_n$  though one may generalize Theorem 3 to weaker hypotheses  $\Sigma = \sigma^2 \Sigma_2$  where  $\Sigma_2$  is any *known* variance-covariance matrix and the variance level  $\sigma^2$  is *unknown*.

**Remark 2.** Consider now that the variance estimation was performed including not only the residuals but the full vector  $U$  and define the corresponding test statistic  $T'$ . In that case, according to simulations displayed on Figure 2, the test  $T'$  seems biased for small values of  $(n, p)$  and is anyway less powerful than  $T$ . One may argue that the situation is comparable to the ordinary regression model where the Fisher test is uniformly most powerful among unbiased tests.



**Figure 2.** Empirical distribution function of 9000  $p$ -values coming from various scenarii: orange, red and blue lines correspond respectively, to  $T$  under the null,  $T$  under the alternative and  $T'$  under the alternative (see Remark 2 for a definition). At the top, the sparsity  $s$  is equal to 2. At the bottom, the sparsity  $s$  is 5. In both cases, from left to right,  $(n, p) = (10, 50), (50, 100)$  and  $(100, 500)$ .

## 1.5. Empirical distributions of the $p$ -values

Figure 1 shows the empirical distribution of a sample of 15 000  $p$ -values constructed from standard regression problems under the global null and under the alternative for the pivots  $S$  and  $T$ . Design matrices  $X$  and the mean (under the alternative) have been drawn uniformly at random from the following cases:

- $X$  is a design matrix of size  $50 \times 100$ ,  $100 \times 200$  or  $100 \times 500$  with i.i.d.  $\mathcal{N}(0, 1)$  entries.
- $\beta$  has i.i.d.  $\mathcal{N}(0, 1)$  (small mean),  $\mathcal{N}(0, 4)$  (medium mean) or  $\mathcal{N}(\sqrt{2 \log p}, 1)$  (high mean) entries.

Under the null, the agreement with uniform is very strong. Moreover, the Spacing test for LARS is empirically more powerful than the  $t$ -Spacing test for LARS and both seem to be unbiased. However, in a context of very high-dimensional regression, the  $t$ -Spacing test for LARS is very similar to the Spacing test for LARS due to standard results on Student and chi-squared distribution.

Note that Section 4 presents a comprehensive study of the two dimensional case from both theoretical and numerical point of views, and a numerical investigation of the power of the Spacing test under various alternatives and against others testing procedures.

## 1.6. Previous works

Our test can be also referred to as the *Kac–Rice test* as introduced in the broader frame of penalization problems in the seminal paper [26]. The interested reader may consult Theorem 1 in [26] where the general “*Kac–Rice pivot*” is defined. Note that various important results on this subject have been obtained recently and we do not pursue on a comprehensive study here. The interested reader may consult Chapter 6 of the captivating book [16].

The statistic *Kac–Rice pivot* given in [26] has been used for model selection and confidence intervals on the target entries. In the frame of Lasso, the optimality of these approaches is discussed in [18,25]. Interestingly, the Spacing test is a nonasymptotic version of the covariance test [19,25], and is asymptotically equivalent to it. Note they have been intensively commented among the literature, see [6,20,21] for instance. Finally, we point that the settings for the power proof results (especially the orthogonal regressors case and the group variable selection case) are resemblant of some of the post-dating work [22] in group forward Post Selection Inference. However, the case of unknown variance is new (see Theorem 3) and has not been addressed in the literature.

## 1.7. Organization of the paper

The next section is devoted to the proof of the main results on the power of the Spacing test. In particular, the reader may find the exact formulation of Theorem 5 mentioned in the [Introduction](#). Section 3 addresses the issue of extending the Spacing test for LARS to the unknown variance frame and introduces an exact studentization of the testing procedure. Section 4.1 presents a fine description of the Spacing test for LARS’ power in the case of two predictors while the end of Section 4 gives a numerical comparison with the Pearson’s chi-squared test (goodness of fit) and the maximal correlation test (i.e., the numerical testing procedure based on  $\lambda_1$  where the rejection threshold/quantile has been set by MCQMC method).

# 2. Power of the spacing test for LARS

## 2.1. Model and notation

Recall that the vector of correlations  $U = X^\top Y$  enjoys

$$U = (U_1, \dots, U_p) \sim \begin{cases} \mathcal{N}_p(0, R), & \text{under the null hypothesis,} \\ \mathcal{N}_p(\mu^\star, R), & \text{under the alternative hypothesis,} \end{cases}$$

where  $R = X^\top \Sigma X$  and  $\mu^\star = X^\top X \beta^\star$ . Indeed, observe that

$$\{\mu^\star = 0\} \Leftrightarrow \{\mathbb{H}_0 : “\beta^\star \in \ker(X)”\}.$$

It is well known (see, for instance, the book [16]) that the first knot  $\lambda_1$  of the LARS algorithm enjoys  $\lambda_1 = \|U\|_\infty$ . Assume that the columns of  $X$  are pairwise different. It implies that, with



probability one, there exists a unique pair  $(\hat{i}, \hat{\varepsilon})$  with  $\hat{i} \in \llbracket 1, p \rrbracket$ ,  $\hat{\varepsilon} = \pm 1$  and such that

$$\hat{\varepsilon} U_{\hat{i}} = \|U\|_{\infty}. \quad (2.1)$$

Observe that the events  $\mathcal{E}_{i,\varepsilon} := \{\varepsilon U_i = \|U\|_{\infty}\}$  are almost surely disjoint, and note that

$$\lambda_1 = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \varepsilon U_i \mathbb{1}_{\mathcal{E}_{i,\varepsilon}},$$

where  $\mathbb{1}$  denotes the indicator function. Write, for all  $(i, j) \in \llbracket 1, p \rrbracket^2$ ,  $U_j = R_{ji} U_i + U_j^i$ , the regression of  $U_j$  onto  $U_i$ . Recall that the residuals  $U_j^i$  are independent of  $U_i$ . Denote, for all  $i \in \llbracket 1, p \rrbracket$  and  $\varepsilon = \pm 1$ ,

$$\lambda_2^{i,\varepsilon} = \bigvee_{1 \leq j \neq i \leq p} \left\{ \frac{U_j^i}{1 - \varepsilon R_{ji}} \vee \frac{-U_j^i}{1 + \varepsilon R_{ji}} \right\}.$$

Furthermore, remark that  $\mathcal{E}_{i,\varepsilon} = \{\lambda_2^{i,\varepsilon} < \varepsilon U_i\}$ . Indeed, for all  $i \neq j \in \llbracket 1, p \rrbracket$ ,

$$\begin{aligned} \{-\varepsilon U_i < U_j < \varepsilon U_i\} &= \{-\varepsilon U_i(1 + \varepsilon R_{ji}) < U_j - R_{ji} U_i < \varepsilon U_i(1 - \varepsilon R_{ji})\} \\ &= \left\{ \left\{ \frac{U_j^i}{1 - \varepsilon R_{ji}} \vee \frac{-U_j^i}{1 + \varepsilon R_{ji}} \right\} < \varepsilon U_i \right\}. \end{aligned}$$

Hence, define the random variable  $\lambda_2$  as

$$\lambda_2 = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \lambda_2^{i,\varepsilon} \mathbb{1}_{\mathcal{E}_{i,\varepsilon}}.$$

We deduce that

$$(\lambda_1, \lambda_2) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} (\varepsilon U_i, \lambda_2^{i,\varepsilon}) \mathbb{1}_{\{\varepsilon U_i > \lambda_2^{i,\varepsilon}\}}. \quad (2.2)$$

Denote by  $\varphi$  the probability density function of the standard normal distribution.

**Lemma 1.** *For each  $i \in \llbracket 1, p \rrbracket$  and  $\varepsilon = \pm 1$ , the random variable  $\lambda_2^{i,\varepsilon}$  has a density  $p_{\lambda_2^{i,\varepsilon}}^{\mu^*}$ . The joint density of  $(\lambda_1, \lambda_2)$  is given by*

$$\forall (\ell_1, \ell_2) \in \mathbb{R}^2, \quad p_{(\lambda_1, \lambda_2)}^{\mu^*}(\ell_1, \ell_2) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \varphi(\ell_1 - \varepsilon \mu_i^*) p_{\lambda_2^{i,\varepsilon}}^{\mu^*}(\ell_2) \mathbb{1}_{\{0 \leq \ell_2 \leq \ell_1\}}. \quad (2.3)$$

**Proof.** One can check that  $\lambda_2^{i,\varepsilon}$  has a density, the reader may also consult Ylvisaker's theorem, see Theorem 1.22 in [3], for example.

Observe that for all  $(i, \varepsilon) \in \llbracket 1, p \rrbracket \times \{\pm 1\}$ , the random variable  $\lambda_2^{i, \varepsilon}$  is a deterministic function of the random variables  $U_j^i$  for  $j \neq i$  and hence it is independent of  $\varepsilon U_i$ . We get that the density function  $p_{(\varepsilon U_i, \lambda_2^{i, \varepsilon})}^{\mu^*}$  of  $(\varepsilon U_i, \lambda_2^{i, \varepsilon})$  with respect to Lebesgue measure is given by

$$\forall (i, \varepsilon) \in \llbracket 1, p \rrbracket \times \{\pm 1\}, \forall (\ell_1, \ell_2) \in \mathbb{R}^2, \quad p_{(\varepsilon U_i, \lambda_2^{i, \varepsilon})}^{\mu^*}(\ell_1, \ell_2) = \varphi(\ell_1 - \varepsilon \mu_i^*) p_{\lambda_2^{i, \varepsilon}}^{\mu^*}(\ell_2).$$

Invoke (2.2) to complete the proof.  $\square$

**Lemma 2.** *For the study the distribution of  $S$ , we can assume, without loss of generality, that the expectations  $\mu_i^*$  are non-negative.*

**Proof.** Let  $\mu^* \in \mathbb{R}^p$  and consider the linear map  $T : \mathbb{R}^p \rightarrow \mathbb{R}^p$  that changes the signs of the coordinates of  $U$  with negative expectation. Set

$$\bar{U} := T(U) = \{t_i U_i : i \in \llbracket 1, p \rrbracket\},$$

where for all  $i \in \llbracket 1, p \rrbracket$ ,  $t_i$  is the sign of  $\mu_i^*$ . Each coordinate of  $\bar{U}$  has non-negative expectation and the variance-covariance matrix of  $\bar{U}$  is now  $\bar{R}$  with  $\bar{R}_{i,j} = t_i t_j R_{i,j}$ .

Let us check, with obvious notation, that the test statistic  $S$  enjoys  $S(\bar{U}) = S(U)$ . Indeed, it holds that the first knot  $\lambda_1$  satisfies  $\lambda_1(\bar{U}) = \lambda_1(U)$ ,  $\mathcal{E}_{i,\varepsilon}(\bar{U}) = \mathcal{E}_{i,t_i\varepsilon}(U)$  and one can note that

$$\bar{U}_j^i = t_j U_j - \bar{R}_{ij} t_i U_i = t_j U_j - t_i t_j R_{ij} t_i U_i = t_j U_j^i,$$

and

$$\lambda_2^{i,\varepsilon}(\bar{U}) = \bigvee_{j \neq i} W_{i,\varepsilon,j}(\bar{U}) \quad \text{with } W_{i,\varepsilon,j}(\bar{U}) := \frac{\varepsilon t_j U_j^i}{1 - t_i t_j R_{ij}} \vee \frac{-\varepsilon t_j U_j^i}{1 + t_i t_j R_{ij}}.$$

One may check that, whatever the signs  $t_i, t_j$  are, it holds  $W_{i,\varepsilon,j}(\bar{U}) = W_{i,(t_i\varepsilon),j}(U)$ . Thus  $\lambda_2^{i,\varepsilon}(\bar{U}) = \lambda_2^{i,(t_i\varepsilon)}(U)$  implying  $\lambda_2(\bar{U}) = \lambda_2(U)$ .  $\square$

## 2.2. Piecewise calculus of the power

We have the following useful proposition giving an exact expression of the power of Spacing test for LARS as weighed sum of Gaussian mesures of disjoint cones. Denote by  $\mathcal{C}_{i,\varepsilon}$  the cone

$$\mathcal{C}_{i,\varepsilon} := \{(u_1, \dots, u_p) \in \mathbb{R}^p \text{ such that } \forall j \neq i, |u_j| < \varepsilon u_i\},$$

recall that  $\bar{\Phi} = 1 - \Phi$  is the complement of the standard normal cumulative distribution function and define by  $\bar{\Phi}^{-1}$  its inverse function.

**Proposition 4.** *For all  $\alpha \in (0, 1)$ , define*

$$h_\alpha(\ell) := \bar{\Phi}^{-1}(\alpha \bar{\Phi}(\ell)) - \ell. \quad (2.4)$$

Then it holds,

$$\mathbb{P}_{\mu^*}\{S \leq \alpha\} = \alpha \mathbb{E}_{\mu^*} \left\{ \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \exp[\varepsilon \mu_i^* h_\alpha(\varepsilon U_i)] \mathbb{1}_{\{U \in C_{i,\varepsilon}\}} \right\}, \quad (2.5)$$

where  $\mathbb{E}_{\mu^*}$  denotes the expectation under the Gaussian distribution  $\mathcal{N}_p(\mu^*, R)$ .

**Proof.** Let  $\alpha \in (0, 1)$ . Note that

$$\{S \leq \alpha\} = \{\lambda_1 \geq \bar{\Phi}^{-1}(\alpha/2)\} \cap \{\lambda_2 \leq \bar{\Phi}^{-1}(\bar{\Phi}(\lambda_1)/\alpha)\}.$$

Using (2.2), the change of variable  $q_1 = \bar{\Phi}^{-1}(\bar{\Phi}(\ell_1)/\alpha)$  and (2.3), it holds

$$\begin{aligned} \mathbb{P}_{\mu^*}\{S \leq \alpha\} &= \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \int_{\bar{\Phi}^{-1}(\frac{\alpha}{2})}^{+\infty} d\ell_1 \varphi(\ell_1 - \varepsilon \mu_i^*) \int_0^{\bar{\Phi}^{-1}(\bar{\Phi}(\ell_1)/\alpha)} d\ell_2 p_{\lambda_2^{i,\varepsilon}}^{\mu^*}(\ell_2) \\ &= \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \int_0^{+\infty} dq_1 \frac{\varphi(q_1)}{\varphi(\ell_1)} \varphi(\ell_1 - \varepsilon \mu_i^*) \int_0^{q_1} d\ell_2 p_{\lambda_2^{i,\varepsilon}}^{\mu^*}(\ell_2) \\ &= \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \int_0^{+\infty} dq_1 e^{\varepsilon \mu_i^*(\ell_1 - q_1)} \varphi(q_1 - \varepsilon \mu_i^*) \int_0^{q_1} d\ell_2 p_{\lambda_2^{i,\varepsilon}}^{\mu^*}(\ell_2) \\ &= \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \int_0^{+\infty} dq_1 e^{\varepsilon \mu_i^*(\ell_1 - q_1)} \int_0^{q_1} d\ell_2 p_{(\varepsilon U_i, \lambda_2^{i,\varepsilon})}^{\mu^*}(q_1, \ell_2) \\ &= \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \mathbb{E}_{\mu^*} [\exp[\varepsilon \mu_i^* (\bar{\Phi}^{-1}(\alpha \bar{\Phi}(\varepsilon U_i)) - \varepsilon U_i)] \mathbb{1}_{\{\varepsilon U_i > \lambda_2^{i,\varepsilon}\}}] \\ &= \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \mathbb{E}_{\mu^*} [\exp[\varepsilon \mu_i^* (\bar{\Phi}^{-1}(\alpha \bar{\Phi}(\varepsilon U_i)) - \varepsilon U_i)] \mathbb{1}_{\{U \in C_{i,\varepsilon}\}}]. \end{aligned}$$

as claimed.  $\square$

**Remark 3.** Note the numerical evaluation of (2.5) can be performed using a  $n$ -dimensional integral, see Section 4.

**Corollary 1.** Under  $\mathbb{H}_0$ , the statistics  $S$  defined by (Pivot) follows a uniform distribution on  $[0, 1]$ .

**Proof.** The null hypothesis is equivalent to  $\mu^* = 0$  and, from (2.5), we recover that

$$\mathbb{P}_0\{S \leq \alpha\} = \alpha \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \mathbb{E}_0(C_{i,\varepsilon}) = \alpha,$$

i.e. the level of Spacing test for LARS is  $\alpha$ . This proves that, under  $\mathbb{H}_0$ , the test statistics  $S$  satisfies

$$S \sim \text{Unif}([0, 1]),$$

as claimed.  $\square$

### 2.3. Distribution under the alternative

This section is devoted to the proof of Theorem 1.

*Step 1:* By a standard approximation argument, one may assume that  $R$  is a regular matrix. Indeed, if  $R$  is singular we can approximate it by a sequence  $(R_m)_{m \geq 0}$  of regular matrices with bounded variance. If for each of these matrices we have  $\mathbb{P}_{\mu^*}\{S \leq \alpha\} \geq \alpha$  then the result will pass to  $R$  by dominated convergence in (2.5). Furthermore, using Lemma 2, we may also assume that  $\forall i \in \llbracket 1, p \rrbracket, \mu_i^* \geq 0$ .

Recall that  $\mathcal{C}_{i,\varepsilon}$  is the cone

$$\mathcal{C}_{i,\varepsilon} := \{(u_1, \dots, u_p) \in \mathbb{R}^p \text{ such that } \forall j \neq i, |u_j| < \varepsilon u_i\}$$

and denote by  $\gamma$  the non-degenerate Gaussian measure associated with the multivariate normal distribution  $\mathcal{N}_p(0, R)$ .

*Step 2:* We start from (2.5) to get that

$$\begin{aligned} \frac{1}{\alpha} \mathbb{P}_{\mu^*}\{S \leq \alpha\} &= \mathbb{E}_{\mu^*} \left\{ \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \exp[\varepsilon \mu_i^* h_\alpha(\varepsilon U_i)] \mathbb{1}_{\mathcal{C}_{i,\varepsilon}} \right\} \\ &\geq 1 + \mathbb{E}_{\mu^*} \left\{ \sum_{i=1}^p \sum_{\varepsilon=\pm 1} [\varepsilon \mu_i^* h_\alpha(\varepsilon U_i)] \mathbb{1}_{\mathcal{C}_{i,\varepsilon}} \right\}. \end{aligned}$$

Perform an integration using the fibers  $F_{\ell,i,\varepsilon} := \{u_i = \varepsilon \ell\} \cap \mathcal{C}_{i,\varepsilon}$  to obtain that

$$\frac{1}{\alpha} \mathbb{P}_{\mu^*}\{S \leq \alpha\} \geq 1 + \int_0^{+\infty} \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \varepsilon \mu_i^* h_\alpha(\ell) \sigma_{\mu^*}(\ell, i, \varepsilon) d\ell,$$

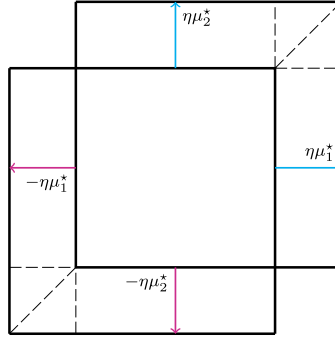
where  $\sigma_{\mu^*}(\ell, i, \varepsilon)$  is the integral of the density function  $\varphi_{\mu^*}$  of the multivariate normal distribution  $\mathcal{N}_p(\mu^*, R)$  on the fiber  $F_{\ell,i,\varepsilon}$ .

*Step 3:* Let  $\ell > 0$  and  $a \geq 0$ . Consider the hypercube  $H_\ell := [-\ell, \ell]^p$  and denote by  $H_\ell - a\mu^*$  its translation by vector  $-a\mu^*$ . Invoke Anderson's inequality (see Lemma 4) to get that

$$a \mapsto \gamma(H_\ell - a\mu^*) := \mathbb{P}\{\mathcal{N}_p(0, R) \in H_\ell - a\mu^*\},$$

is a non-increasing function on the domain  $a \geq 0$ . In particular, its derivative at point  $a = 1$  is non-positive. It reads as

$$\lim_{\eta \rightarrow 0} \frac{1}{\eta} (\gamma(H_\ell - (1 + \eta)\mu^*) - \gamma(H_\ell - \mu^*)) \leq 0,$$



**Figure 3.** Illustration of (2.6) in dimension 2. Passing to the limit, contribution of triangles (dashed lines) vanish and the derivative in  $a = 1$  is equal to the sum of each face with a weight  $\varepsilon\mu_i^*$  corresponding to its orientation.

and this quantity is simply, by Step 4,

$$\sum_{i=1}^p \mu_i^* \sigma_{\mu^*}(\ell, i, -1) - \sum_{i=1}^p \mu_i^* \sigma_{\mu^*}(\ell, i, +1) \leq 0. \quad (2.6)$$

Finally, the positivity of  $h_\alpha(\ell)$  (see Lemma 5) completes the proof.

*Step 4:* In the context of Step 3, computation on  $\gamma(H_\ell - a\mu^*)$  gives that

$$\begin{aligned} \frac{d}{da} \gamma(H_\ell - a\mu^*) &= \int_{H_\ell} \frac{\partial}{\partial a} \varphi_{a\mu^*}(z) dz = \sum_{i=1}^p -\mu_i \int_{H_\ell} \frac{\partial}{\partial z_i} \varphi_{a\mu^*}(z) dz \\ &= \sum_{i=1}^p \sum_{\varepsilon=\pm 1} -\varepsilon \mu_i^* \sigma_{a\mu^*}(\ell, i, \varepsilon), \end{aligned}$$

where, for all  $a > 0$ , we denote by  $\sigma_{a\mu^*}(\ell, i, \varepsilon)$  the integral of the density function  $\varphi_{a\mu^*}$  of the multivariate normal distribution  $\mathcal{N}_p(a\mu^*, R)$  on the fiber  $F_{\ell, i, \varepsilon}$ .

This computation might also be illustrated via finite difference method, one may see Figure 3, for instance.

## 2.4. Orthogonal case

In this section, we give the proof of Theorem 2.

Invoke (2.3) to get that, under  $\mathbb{H}_1$ ,

$$\forall (\ell_1, \ell_2) \in \mathbb{R}^2, \quad p_{(\lambda_1, \lambda_2)}^{\mu^*}(\ell_1, \ell_2) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \varphi(\ell_1 - \varepsilon \mu_i^*) p_{\lambda_2}^{\mu^*}(\ell_2) \mathbb{1}_{\{0 \leq \ell_2 \leq \ell_1\}}.$$

Recall that  $\hat{i} \in \llbracket 1, p \rrbracket$  is defined by (2.1). Since  $R = \text{Id}_p$ , remark that

$$\lambda_2 = \max_{j \neq i} |U_j|.$$

Furthermore, observe that  $\lambda_2^{i,+1} = \lambda_2^{i,-1}$  almost surely. It implies that for all  $i \in \llbracket 1, p \rrbracket$ ,

$$p_{\lambda_2^{i,+1}}^{\mu^*} = p_{\lambda_2^{i,-1}}^{\mu^*}.$$

Denote by  $p_{\lambda_2^i}^{\mu^*}$  their common value. As a consequence, it holds

$$\forall (\ell_1, \ell_2) \in \mathbb{R}^2, \quad p_{(\lambda_1, \lambda_2)}^{\mu^*}(\ell_1, \ell_2) = \sum_{i=1}^p (\varphi(\ell_1 - \mu_i^*) + \varphi(\ell_1 + \mu_i^*)) p_{\lambda_2^i}^{\mu^*}(\ell_2) \mathbb{1}_{\{0 \leq \ell_2 \leq \ell_1\}}.$$

It implies that, conditionally to  $\lambda_2 = \ell_2$ , the random variable  $\bar{\Phi}(\lambda_1)$  admits the density

$$p_{(\bar{\Phi}(\lambda_1)|\lambda_2=\ell_2)}(v) = (\text{const}) \sum_{i=1}^p \cosh(\bar{\Phi}^{-1}(v) \mu_i^*) \mathbb{1}_{\{\bar{\Phi}^{-1}(v) \geq \ell_2\}}. \quad (2.7)$$

Since  $\bar{\Phi}^{-1}(v)$  remains in the positive domain, the functions into the sum above are non-increasing and strictly decreasing for the index  $i$  such that  $\mu_i^* > 0$ . We have clearly the same result for the expression equivalent to (2.7) given the conditional density of  $S$ .

Deconditionning we obtain that the density of  $S$  is a mixture of non-increasing functions, thus non-increasing. In addition, the deconditioning formula gives positive weights to decreasing functions thus, in fact, the density is decreasing.

## 2.5. Asymptotic case

**Theorem 5.** *Under  $\mathbb{H}_1$ , it holds*

$$\alpha^{-1} \mathbb{P}_{\mu^*} \{S \leq \alpha\} \rightarrow +\infty,$$

as  $\alpha$  goes to zero, where  $\mu^* = (X^\top X) \beta^*$  and  $\mathbb{P}_{\mu^*}$  denotes the law of  $X^\top Y \sim \mathcal{N}_p(\mu^*, R)$ .

**Proof.** Recall that  $\mathbb{H}_1$  is equivalent to  $\mu^* \neq 0$ . Without loss of generality, assume  $\mu_1^* > 0$  and note that

$$\begin{aligned} \alpha^{-1} \mathbb{P}_{\mu^*} \{S \leq \alpha\} &= \mathbb{E}_{\mu^*} \left\{ \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \exp[\mu_i^* h_\alpha(\varepsilon U_i)] \mathbb{1}_{\{U \in \mathcal{C}_{i,\varepsilon}\}} \right\} \\ &\geq \mathbb{E}_{\mu^*} \{ \exp[\mu_1^* h_\alpha(U_1)] \mathbb{1}_{\{U \in \mathcal{C}_{1,1}\}} \} =: \Lambda_{1,1}(\alpha). \end{aligned}$$

Moreover, observe that

$$\forall x \in \mathbb{R}, \quad h_\alpha(x) \rightarrow +\infty,$$

as  $\alpha > 0$  goes to zero. In particular, it yields

$$\forall x \in \mathbb{R}, \quad \exp[\mu_1^\star h_\alpha(x)] \rightarrow +\infty,$$

as  $\alpha > 0$  goes to zero. Eventually, let  $(\alpha_n)_{n \in \mathbb{N}}$  be any sequence of positive reals that goes to zero as  $n$  tends to  $\infty$ . Invoke Fatou's lemma to get that

$$\begin{aligned} \lim_{n \rightarrow \infty} \alpha_n^{-1} \mathbb{P}_{\mu^\star} \{S \leq \alpha_n\} &\geq \lim_{n \rightarrow \infty} \Lambda_{1,1}(\alpha_n) \\ &\geq \liminf_{n \rightarrow \infty} \mathbb{E}_{\mu^\star} \left\{ \exp[\mu_1^\star h_\alpha(U_1)] \mathbb{1}_{\{U \in \mathcal{C}_{1,1}\}} \right\} \\ &\geq \mathbb{E}_{\mu^\star} \left\{ \liminf_{\alpha_n \rightarrow 0} \exp[\mu_1^\star h_\alpha(U_1)] \mathbb{1}_{\{U \in \mathcal{C}_{1,1}\}} \right\} \end{aligned}$$

which concludes the proof.  $\square$

### 3. Studentization of the spacing test for LARS

In this section, we give the proof of Theorem 3.

#### 3.1. Model and notation

Assume that the variance–covariance matrix  $\Sigma$  of the Gaussian noise  $\xi$  is  $\sigma^2 \text{Id}_n$  where the standard deviation  $\sigma > 0$  is unknown. Assume also that the columns  $(X_i)_{i=1}^p$  of the design matrix  $X$  enjoy  $\|X_i\|_2 = 1$  and denote by  $U := X^\top Y$  the correlation vector satisfying

$$U = (U_1, \dots, U_p) \sim \begin{cases} \mathcal{N}_p(0, \sigma^2 R), & \text{under the null hypothesis,} \\ \mathcal{N}_p(\mu^\star, \sigma^2 R), & \text{under the alternative hypothesis,} \end{cases}$$

where  $R = X^\top X$  and  $\mu^\star = R\beta^\star$ . Observe that the knots of the LARS algorithm are given by

$$(\lambda_1, \lambda_2) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} (\varepsilon U_i, \lambda_2^{i,\varepsilon}) \mathbb{1}_{\{\varepsilon U_i > \lambda_2^{i,\varepsilon}\}}.$$

For each  $i \in \llbracket 1, p \rrbracket$ , we denote by  $X_{-i} \in \mathbb{R}^{n \times (p-1)}$  the sub-matrix of  $X$  where the  $i$ th column  $X_i$  has been deleted. Also, we denote by  $U_{-i} \in \mathbb{R}^{p-1}$  (resp.  $\mu_{-i}^\star$ ) the sub-vector of  $U$  (resp.  $\mu^\star$ ) where the  $i$ th entry has been deleted. Observe that the regression of  $U_{-i}$  onto  $U_i$  reads

$$U_{-i} = (R_i)_{-i} U_i + V_{-i},$$

where  $(R_i)_{-i} \in \mathbb{R}^{p-1}$  denotes the sub-vector of the  $i$ th column  $R_i$  of the matrix  $R$  where the  $i$ th entry has been deleted. Observe that the vector  $V_{-i} \in \mathbb{R}^{p-1}$  is a Gaussian vector independent of  $U_i$  such that

$$V_{-i} \sim \mathcal{N}_{p-1}(\mu_{-i}^* - (R_i)_{-i} \mu_i^*, \sigma^2 R_{-i}), \quad (3.1)$$

where  $R_{-i} := X_{-i}^\top (\Pi_X - X_i X_i^\top) X_{-i}$  denotes its variance-covariance matrix with  $\Pi_X$  the orthogonal projection onto the range of  $X$ . Observe that  $\Pi_X$  has rank  $m$  by assumption and it follows that  $R_{-i}$  has rank  $m - 1$ . Denote  $R_{-i}^{-1/2}$  the only symmetric matrix such that  $R_{-i}^{-1/2} R_{-i} R_{-i}^{-1/2}$  is the orthogonal projection onto the range of  $R_{-i}$  (i.e.,  $R_{-i}^{-1/2}$  is the square root of the Moore–Penrose pseudoinverse of  $R_{-i}$ ).

### 3.2. Estimation of the variance

An estimation of the variance  $\sigma$  is given by

$$\hat{\sigma}_i := \frac{\|R_{-i}^{-1/2} V_{-i}\|_2}{\sqrt{m-1}}.$$

Indeed, Equation (3.1) gives that, under  $\mathbb{H}_0$ , it holds

$$\frac{\|R_{-i}^{-1/2} V_{-i}\|_2^2}{\sigma^2} \sim \chi^2(m-1),$$

where  $\chi^2(m-1)$  is the chi-squared distribution with  $m-1$  degree(s) of freedom. Since  $V_{-i}$  is independent of  $U_i$ , note that  $\varepsilon U_i$  and  $\hat{\sigma}_i$  are independent. Furthermore, since  $V_{-i}$  is Gaussian, remark that its norm and its direction are independent so that  $V_{-i}/\hat{\sigma}_i$  and  $\hat{\sigma}_i$  are independent. Recall that

$$\lambda_2^{i,\varepsilon} := \bigvee_{1 \leq j \neq i \leq p} \left\{ \frac{U_j - R_{ji} U_i}{1 - \varepsilon R_{ji}} \vee \frac{-U_j + R_{ji} U_i}{1 + \varepsilon R_{ji}} \right\},$$

and  $V_{-i} := (U_1 - R_{1i} U_i, \dots, U_{i-1} - R_{(i-1)i} U_i, U_{i+1} - R_{(i+1)i} U_i, \dots, U_p - R_{pi} U_i)$ . Eventually, remark that

$$\varepsilon U_i, \quad \frac{\lambda_2^{i,\varepsilon}}{\hat{\sigma}_i} \quad \text{and} \quad \hat{\sigma}_i \quad \text{are mutually independent.} \quad (3.2)$$

### 3.3. Distribution of the test statistic

Let  $(i, \varepsilon)$  be in  $\llbracket 1, p \rrbracket \times \{\pm 1\}$ . Recall that  $\varepsilon U_i$  and  $\lambda_2^{i,\varepsilon}$  are independent. In view of (3.2), observe that

$$T_1^{i,\varepsilon} := \frac{\varepsilon U_i}{\hat{\sigma}_i} \quad \text{and} \quad T_2^{i,\varepsilon} := \frac{\lambda_2^{i,\varepsilon}}{\hat{\sigma}_i},$$



are independent and, under  $\mathbb{H}_0$ , the random variable  $T_1^{i,\varepsilon}$  a Student random variable with  $m - 1$  degree(s) of freedom. Define  $(T_1, T_2)$  as

$$(T_1, T_2) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \left( \frac{\varepsilon U_i}{\hat{\sigma}_i}, \frac{\lambda_2^{i,\varepsilon}}{\hat{\sigma}_i} \right) \mathbb{1}_{\{\varepsilon U_i > \lambda_2^{i,\varepsilon}\}}, \quad (3.3)$$

and recall that the events  $\{\varepsilon U_i = \|U\|_\infty\} = \{\varepsilon U_i > \lambda_2^{i,\varepsilon}\}$  are almost surely disjoint.

**Lemma 3.** *Under  $\mathbb{H}_0$ , for each  $i \in \llbracket 1, p \rrbracket$  and  $\varepsilon = \pm 1$ ,  $T_2^{i,\varepsilon} = \lambda_2^{i,\varepsilon} / \hat{\sigma}_i$  has a density  $p_{T_2^{i,\varepsilon}}^0$ . Under  $\mathbb{H}_0$ , the joint density of  $(T_1, T_2)$  is given by*

$$\forall (\ell_1, \ell_2) \in \mathbb{R}^2, \quad p_{(T_1, T_2)}^0(t_1, t_2) = \mathbb{1}_{\{0 \leq t_2 \leq t_1\}} t_{m-1}(t_1) \sum_{i=1}^p \sum_{\varepsilon=\pm 1} p_{T_2^{i,\varepsilon}}^0(t_2), \quad (3.4)$$

where  $t_{m-1}$  denotes the probability density function of the  $t$ -distribution with  $m - 1$  degree(s) of freedom.

**Proof.** One can check that  $T_2^{i,\varepsilon}$  has a density, the reader may also consult Ylvisaker's theorem, see Theorem 1.22 in [3], for example. Observe that for all  $(i, \varepsilon) \in \llbracket 1, p \rrbracket \times \{\pm 1\}$ , the random variable  $T_2^{i,\varepsilon}$  is independent of the random variable  $T_1^{i,\varepsilon}$ . The result follows by (3.3).  $\square$

Recall that

$$T := \frac{1 - \mathbb{F}_{m-1}(T_1)}{1 - \mathbb{F}_{m-1}(T_2)},$$

where  $\mathbb{F}_{m-1}$  denote the cumulative distribution function of the  $t$ -distribution with  $m - 1$  degrees of freedom. The expression of the joint density (3.4) shows that, conditionally to  $T_2$ , the random variable  $T_1$  is distributed as a Student distribution conditioned to be greater than  $T_2$ . As a consequence, the conditional distribution of  $T$  is uniformly distributed on  $[0, 1]$ . We deduce that  $T$  is uniformly distributed on  $[0, 1]$ , as claimed.

## 4. Numerical experiments on the power

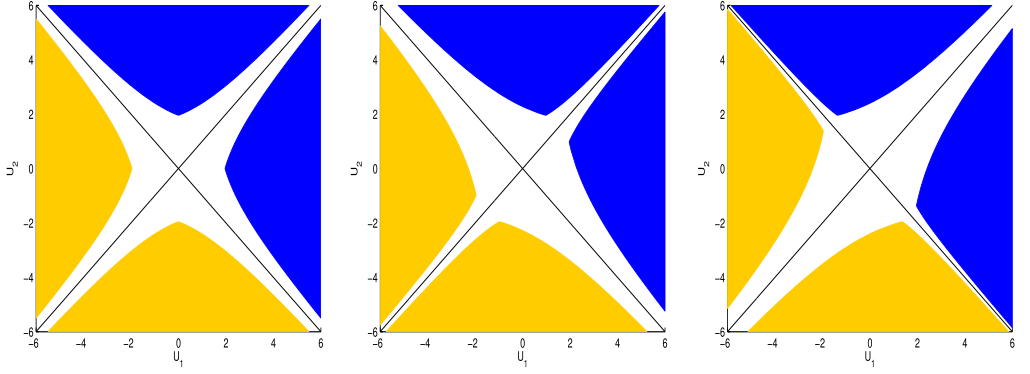
### 4.1. Overture: A theoretical analysis of the two dimensional case

In this section, we assume that  $p = 2$  and  $\Sigma = \text{Id}_2$ . Define

$$R = R(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (4.1)$$

with  $\rho = \text{Cov}(U_1, U_2) \in [-1, 1]$ . Define the rejection region  $\mathcal{R}_\alpha$  by

$$\mathbb{P}\{S \leq \alpha\} =: \mathbb{P}\{U = (U_1, U_2) \in \mathcal{R}_\alpha\}.$$



**Figure 4.** An illustration of region  $\mathcal{R}_{0.05}^{+,1} \cup \mathcal{R}_{0.05}^{+,2}$  in (dark) blue and region  $\mathcal{R}_{0.05}^{-,1} \cup \mathcal{R}_{0.05}^{-,2}$  in (bright) yellow for  $\rho = 0$ ,  $\rho = 0.5$  and  $\rho = -0.7$  (from left to right).

Note  $\mathcal{R}_\alpha$  is symmetric about the origin and it is the non-convex disjoint union of four convex regions (see Figure 4 for instance), namely

$$\mathcal{R}_\alpha^{+,1} = \{U_1 \geq \bar{\Phi}^{-1}(\alpha/2)\} \cap \{-g_\alpha(U_1)(1+\rho) + \rho U_1 \leq U_2 \leq g_\alpha(U_1)(1-\rho) + \rho U_1\},$$

$$\mathcal{R}_\alpha^{+,2} = \{U_2 \geq \bar{\Phi}^{-1}(\alpha/2)\} \cap \{-g_\alpha(U_2)(1+\rho) + \rho U_2 \leq U_1 \leq g_\alpha(U_2)(1-\rho) + \rho U_2\},$$

$$\mathcal{R}_\alpha^{-,1} = \{-U_1 \geq \bar{\Phi}^{-1}(\alpha/2)\} \cap \{-g_\alpha(-U_1)(1-\rho) + \rho U_1 \leq U_2 \leq g_\alpha(-U_1)(1+\rho) + \rho U_1\},$$

$$\mathcal{R}_\alpha^{-,2} = \{-U_2 \geq \bar{\Phi}^{-1}(\alpha/2)\} \cap \{-g_\alpha(-U_2)(1-\rho) + \rho U_2 \leq U_1 \leq g_\alpha(-U_2)(1+\rho) + \rho U_2\},$$

where  $g_\alpha(x) := \bar{\Phi}^{-1}(\bar{\Phi}(x)/\alpha) = \Phi^{-1}(1 - \frac{1-\Phi(x)}{\alpha})$ , for all  $x \in \mathbb{R}$ .

**Remark 4.** Observe this decomposition holds in any dimension  $p$ . The region  $\mathcal{R}_\alpha$  given by  $\mathbb{P}\{S \leq \alpha\} = \mathbb{P}\{U \in \mathcal{R}_\alpha\}$  is symmetric about the origin and is the non-convex disjoint union of  $2^p$  convex regions of  $\mathbb{R}^p$ .

Note Anderson's inequality (see Lemma 4) is sufficient to establish that the Gaussian measure of symmetric convex set is monotonic. Unfortunately, the region  $\mathcal{R}_\alpha^c$  is not convex. However, when  $p = 2$ , using an appropriate fibration, one may find a collection of sets (more general than  $\mathcal{R}_\alpha^c$ ) satisfying a kind of generalization of Anderson's inequality. This is the object of the following proposition.

**Proposition 6.** For  $u \in \mathbb{R}$  and  $\varepsilon = \pm 1$ , define the non-centered diagonals

$$\Delta_u^\varepsilon = \{(x, y) \in \mathbb{R}^2; y = \varepsilon x + u\}.$$

Let  $\mathcal{T}$  be a set of  $\mathbb{R}^2$  which is symmetric with respect to the two diagonals  $\Delta_0^{+1}$  and  $\Delta_0^{-1}$  and satisfies for all  $u$  and  $\varepsilon$

$$\mathcal{T} \cap \Delta_u^\varepsilon \quad \text{is an interval.} \quad (4.2)$$

Set

$$\Psi(\mu_1, \mu_2) = \mathbb{P}\{\mathcal{N}((\mu_1, \mu_2), R(\rho)) \in \mathcal{T}\}.$$

Then the function  $\Psi(u, v)$ , which is obviously symmetric with respect to centered diagonals  $\Delta_0^{+1}$  and  $\Delta_0^{-1}$ , is non-increasing along every half diagonal. More precisely, for every  $u \in \mathbb{R}$  and  $z \geq 0$ , the two functions

$$z \mapsto \Psi(u/2 - z, u/2 + z) = \Psi(u/2 + z, u/2 - z), \quad (4.3)$$

$$z \mapsto \Psi(u/2 + z, -u/2 + z) = \Psi(u/2 - z, -u/2 - z), \quad (4.4)$$

are non-increasing.

**Remark 5.** Note the result remains true if we multiply the matrix  $R(\rho)$  by a scalar.

**Proof.** The proof is given in Appendix B.1. □

**Corollary 2.**  $\forall \alpha \in (0, 1)$ ,  $\mathcal{R}_\alpha^c$  satisfies the hypothesis of Proposition 6 so the power of the Spacing test for LARS is non-decreasing along the diagonals  $\Delta_u^\varepsilon$  in the sense that it has exactly the same properties as those of the function  $\Psi$  given by (4.3) and (4.4). In particular:

- Spacing test for LARS is unbiased,
- For each  $\mu \in \mathbb{R}^2$ , the function  $t \mapsto \mathbb{P}_{t\mu}(S \leq \alpha)$  is non-increasing for  $t \geq 0$ .

**Proof.** The proof is given in Appendix B.2. □

**Remark 6.** In dimension two, note that  $\Delta_0^1 \cup \Delta_0^{-1} = \{(U_1, U_2); S(U_1, U_2) = 1\}$ . In higher dimensions, one has

$$\{(U_1, \dots, U_p); S(U_1, \dots, U_p) = 1\} = \bigcup_{i=1}^p \bigcup_{\varepsilon=\pm 1} \left\{ \lambda_1 = \varepsilon U_i = \max_{j \neq i} |U_j| \right\}.$$

Observe the aforementioned set is not a hyperplane and so no orthogonal symmetry appears. The proof given in this section cannot be generalized to higher dimensions.

## 4.2. A toolbox computing the power

To compute the power of Spacing test for LARS using (2.5), we need to perform integration in high dimensions. First, observe that (2.5) can be expressed as  $n$ -dimensional Gaussian integral. Indeed, recall that  $n = \text{rank}(R)$  and  $R = R^{1/2}(R^{1/2})^\top$  with  $R^{1/2} \in \mathbb{R}^{p \times n}$ , so that

$$\mathbb{P}_{\mu^\star}\{S \leq \alpha\} = \alpha \mathbb{E}_{\mu^\star}[W(V_1, \dots, V_n)],$$

where  $V = (V_1, \dots, V_n) \sim \mathcal{N}_n(0, \text{Id}_n)$  and

$$W(V_1, \dots, V_n) = \sum_{i=1}^p \sum_{\varepsilon=\pm 1} \exp(\varepsilon \mu_i h_\alpha(\varepsilon(R^{1/2}V)_i)) \mathbb{1}_{C_{i,\varepsilon}}.$$

The aforementioned formula is a high dimensional Gaussian integral and we use a very efficient algorithm from A. Genz [2,14], based on a reduction of the integral on the hypercube  $[0, 1]^n$  and Monte-Carlo Quasi Monte-Carlo (MCQMC) integration. In this fashion, Matlab programs *qsimvn* and *qsimvnef* provide powerful and robust numerical integration algorithms. The MC-QMC routine is based on Kronecker or lattice sequences to compute integrals. In a second step, a Monte-Carlo (MC) layer is added to ensure unbiasedness and to compute the precision. Eventually, the QMC step is nested in the MC step in order to improve the speed of convergence, see [24], for example. A Matlab toolbox computing the power of Spacing test for LARS and based on Genz' routines is available on S. Mourareau's website [23]. In addition, some practical examples are given.

### 4.3. The numerical analysis of the two-dimensional case

In dimension two, the power of Spacing test can be easily computed using numerical integration from (2.5). Consider the power function  $k_{\alpha,\rho}(\beta) = \mathbb{P}\{\mathcal{N}(R\beta, R) \in \mathcal{R}_\alpha\}$ , where  $R = R(\rho)$  is given by (4.1) and the region  $\mathcal{R}_\alpha$  is defined in Section 4.1. The aforementioned power function is monotone in  $\beta$  along the directions defined in Section 4.1 and it can be seen on Figure 5 that the variation of the power is minimal along the diagonal associated to the minimal eigenvalue of  $R(\rho)$ , see also Corollary 2.

### 4.4. Competing the spacing test for LARS

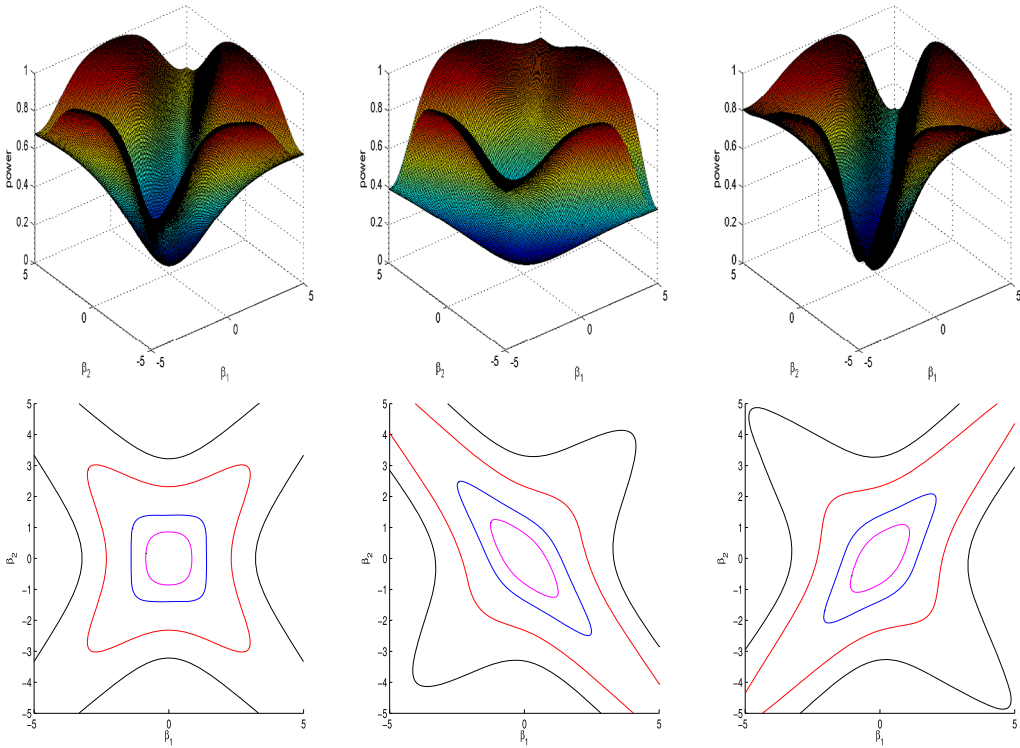
We consider the standard goodness of fit test of the hypothesis  $\mathbb{H}_0 : \beta^* \in \ker(X)$  against  $\mathbb{H}_1 : \beta^* \notin \ker(X)$ . This test is defined by the statistic  $Q = \|Y\|_2^2$  that follows a  $\chi^2(n, \|X\beta^*\|_2^2)$  distribution where  $\chi^2(a, b)$  denotes the  $\chi^2$  distribution with  $a$  degrees of freedom and non-centrality parameter  $b$ . Our aim is to compare this standard test with the Spacing test for LARS in different cases.

#### 4.4.1. The two-dimensional case

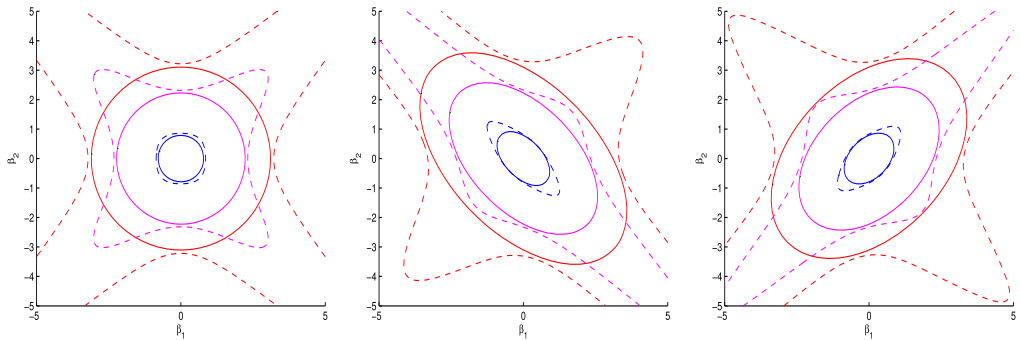
In dimension two, considering the full model  $(s, n, p) = (2, 2, 2)$ , we present a comparison of level sets of power functions for Spacing test for LARS and Pearson's chi-squared test, see Figure 6. It may suggest, from the comparison of level sets, that Pearson's chi-squared test is uniformly more powerful than Spacing test for LARS in the two-dimensional case.

#### 4.4.2. Higher dimensions and some comments

In higher dimension, our experiments have been conducted in the following frame. The design matrix  $X$  is drawn from  $n \times p$  independent standard Gaussian distribution. The target  $\beta^*$  has  $s$



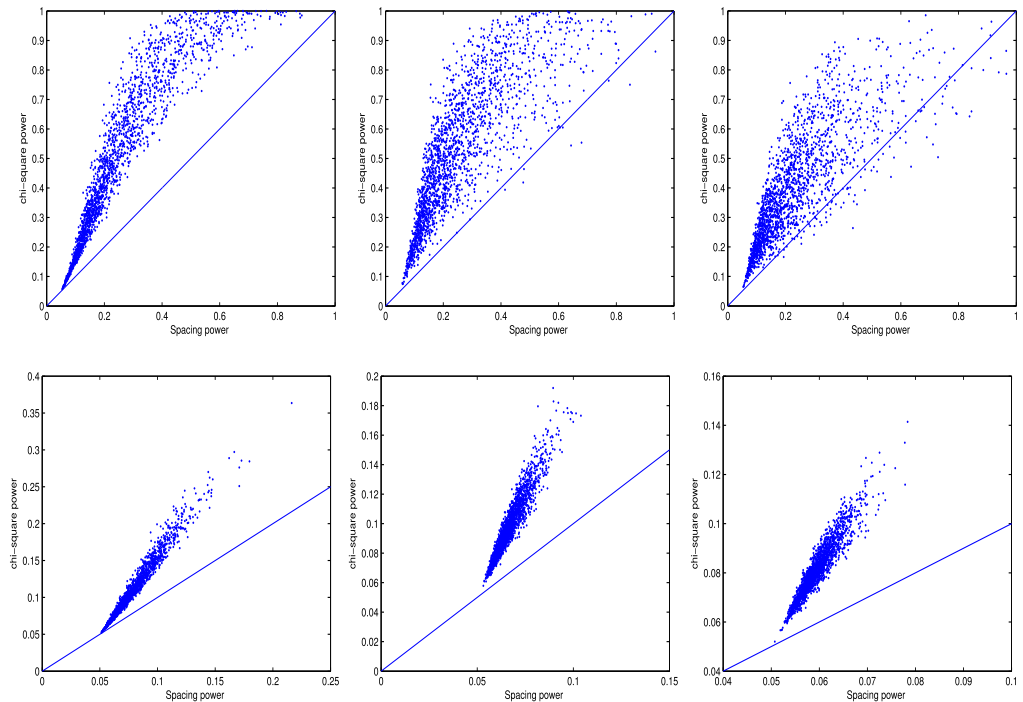
**Figure 5.** At the top, from left to right, power function  $\beta \mapsto k_{\alpha, \rho}(\beta)$  for a significance level  $\alpha = 0.05$  and correlations  $\rho = 0$ ,  $\rho = 0.5$  and  $\rho = -0.4$ . At the bottom, corresponding level sets of the power function,  $k_{\alpha, \rho}(\cdot) = 0.10, 0.20, 0.40$  and  $0.70$ .



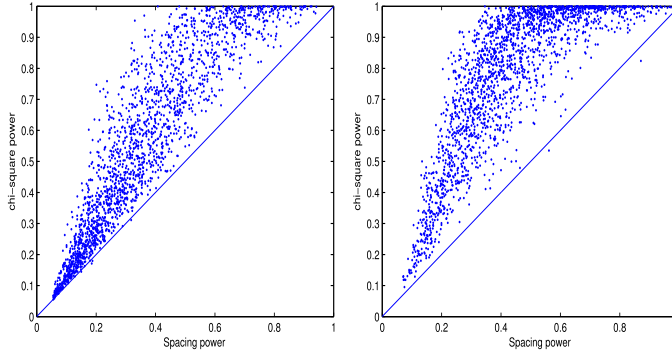
**Figure 6.** From left to right, level sets of the power functions of  $S$  (dashed lines) and  $Q$  (plain line) for  $\alpha = 0.05$  and  $\rho = 0$ ,  $\rho = 0.5$  and  $\rho = -0.4$ . We observe that the hypograph of the power function of  $S$  is included in the corresponding one of  $Q$ .

non zero entries independently and identically drawn from centered Gaussian distribution having variance 2 (referred to as the “*large mean case*”), variance 1 (“*medium mean case*”) or from uniform random distribution on  $[0, 1]$  (“*small mean case*”). The choice of  $s$ ,  $n$  and  $p$  concerns “*full*” models ( $s = n = p$ , see Figure 8), “*sparse*” models (see Figure 7) or “*very sparse*” models ( $s \ll p$ , see Figure 9). Our comments are listed below.

- As we have seen in Section 4.4.1, Figure 6 suggests, from the comparison of level sets, that the  $\chi^2$  test is uniformly more powerful than the Spacing test for LARS in the two dimensional case. Results from Figure 8 seem to confirm the interest of the  $\chi^2$  test for full models. However, the Spacing test seems much more efficient in very high dimension cases when the signal presents a major gap between the dominant component and the rest, see Figure 9 for instance. When the target entries  $\beta_i$  have the same order of magnitude (i.e., independently drawn w.r.t. the same law), even in case of sparse models, the  $\chi^2$  test seems to be more powerful than the Spacing test, see Figure 7.

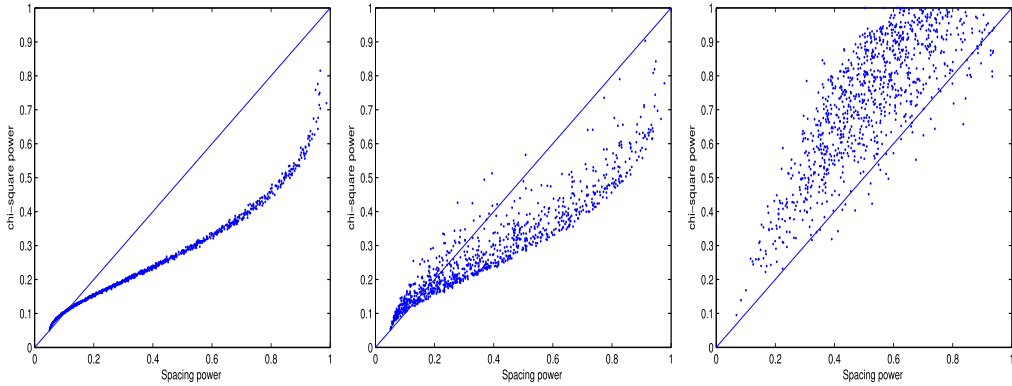


**Figure 7.** From left to right, 2000 simulations of Spacing test for LARS’s power versus  $\chi^2$  power in various sparse cases  $(s, n, p) = (5, 10, 50)$ ,  $(10, 50, 100)$  and  $(10, 100, 200)$ . At the top, the mean  $\beta$  is “large”, while, at the bottom, the mean is “small” (see Section 4.4.2 for a definition). In both case, Pearson’s chi-squared test seems more powerful in respectively 95, 94 and 99% of cases (large mean) and 91, 98 and 99% of cases.

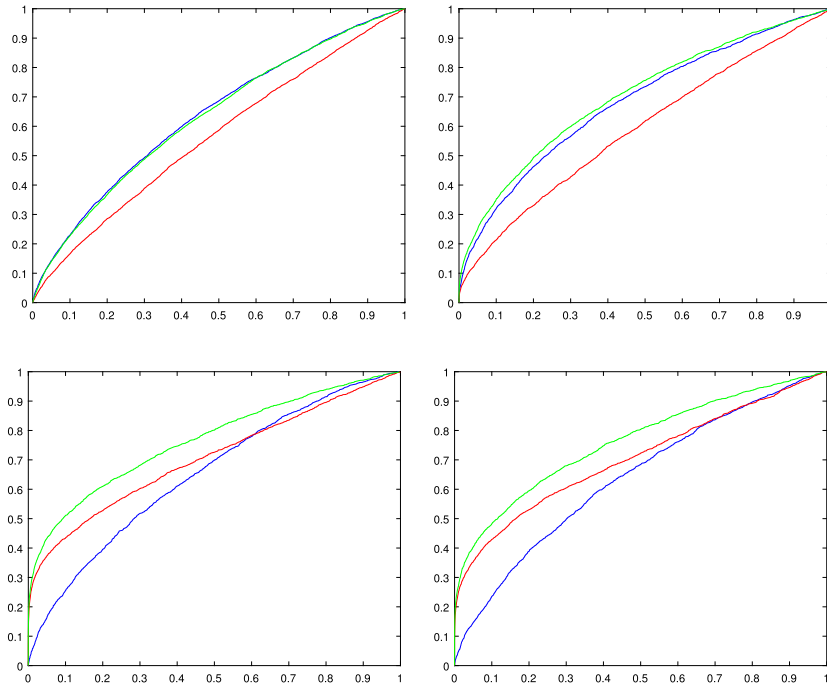


**Figure 8.** From left to right, 2000 simulations of Spacing test for LARS' power versus  $\chi^2$  power in the “full” case  $(s, n, p) = (5, 5, 5)$  and  $(10, 10, 10)$  for a mixture of small, medium and high mean. As in dimension two, the  $\chi^2$  test seems to give an improvement with respect to the Spacing test for LARS.

- On Figure 10, we have compared  $p$ -values coming from the spacing test, the  $\chi^2$  test and the “max test” (i.e., the numerical testing procedure based on  $\lambda_1$  where the rejection threshold/quantile has been set by MCQMC method) under the alternatives of our framework. On one hand, when the signal is not sparse enough, we witness the “power loss” of the Spacing test and similar power performances for the max test and the chi-squared test. On the other hand, in case of very sparse models with a gap between the largest target entry and the others, the max test and the spacing test outperform the chi-squared test.



**Figure 9.** In the first instance (left),  $(s, n, p) = (1, 100, 400)$  and the mean is drawn from  $\mathcal{N}(\sqrt{2\log(p)}, 1)$ . In the second one (center),  $(s, n, p) = (3, 100, 400)$  and  $\beta_1 \sim \mathcal{N}(\sqrt{2\log(p)}, 1)$  while  $\beta_2$  and  $\beta_3$  follow a standard Gaussian. In the third one (right),  $(s, n, p) = (3, 100, 400)$  and all means are drawn from  $\mathcal{N}(\sqrt{2\log(p)}, 1)$ . When one mean is dominant, as in the first two cases, the Spacing test for LARS seems to be more efficient. However, when the difference between the two dominant means isn't large enough, the  $\chi^2$  test seems to be more efficient.



**Figure 10.** Empirical distribution function of 6000  $p$ -values coming from various scenarii: 2000  $p$ -values of  $S$  under the alternative (red), 2000  $p$ -values of  $\lambda_1$  under the alternative (green) and 2000  $p$ -values of  $Q$  under the alternative (blue). At the top,  $(s, n, p) = (10, 10, 10)$  and  $(s, n, p) = (10, 50, 100)$  for a mixture of small, medium and high means. At the bottom line,  $(s, n, p) = (1, 100, 500)$  and the mean is drawn from  $\mathcal{N}(\sqrt{2 \log(p)}, 1)$  on the left panel; whereas  $(s, n, p) = (3, 100, 500)$  and  $\beta_1 \sim \mathcal{N}(\sqrt{2 \log(p)}, 1)$  while  $\beta_2$  and  $\beta_3$  follow a standard Gaussian on the right panel.

- Remark that the max test is very sensitive to the estimation of  $c_{1-\alpha}$  defined by  $\mathbb{P}_{\mathbb{H}_0}\{\lambda_1 \geq c_{1-\alpha}\} = \alpha$  for which we used the MCQMC program of A. Genz, *qsimvn*. In Figure 10 the noise variance  $\sigma^2$  is known with infinite precision and, in that case, we witness that the max test seems to be more powerful than the Spacing test. However, we must keep in mind that, in practice, the variance  $\sigma^2$  is not exactly known and the max test may be not robust to variance estimation errors. Furthermore, contrary to the Spacing test, no studentization of the max test exists, as far as we know. We conclude that, in perfect world, the max test would be a good opportunity though, in more realistic scenarii, one would prefer a more robust and studentized testing procedure such as the Spacing test.

## 5. Conclusion and perspectives

This article provides a new tool computing the power of the Spacing test that we leverage to obtain unbiasedness, studentization, optimal rejection regions (in the orthogonal design case) and



power behavior for small significance levels. Comprehensive numerical and theoretical studies has been given in the two dimensional case. Extensive numerical studies comparing the Spacing to naive chi-squared test and maximal correlation test have been undertaken enlightening that the “power loss” of the Spacing test when increasing sparsity and/or decreasing signal maximal amplitude.

Other interesting questions remain open. In particular, one can wonder whether the Spacing test is still unbiased when the design is not normalized as in (H). In this case, unpublished computations reveal that the fibration presented in this article will not be sufficient to invoke Anderson’s inequality. Indeed, the function  $h$  is no longer constant on opposite faces of the fibration (here the cube) and a finer fibration is needed. Moreover, one can question about the alternative, like sparsity or the separation in magnitude of the largest effect. This issue seems more tractable though, even in the orthogonal case described by (2.7), one sees that consequential work has to be done to properly exhibit the influence of the sparsity (here the sparsity is understood through  $\mu^\star := X^\top X\beta^\star = \beta^\star$ ) and the magnitude of the coefficients of the target  $\mu^\star$ . We feel that it might be carried out through explicit (and nontrivial) computations but we did not pursue this lead here. This would be an interesting question to address in future work.

## Appendix A: Technical lemmas

**Lemma 4 (Anderson’s inequality for Gaussian measure [1]).** *Let  $E$  be a convex set in  $\mathbb{R}^p$ , symmetric around the origin, and let  $Z \sim \mathcal{N}_p(0, V)$ . For all  $t \geq 0$  and  $\mu \in \mathbb{R}^p$  define*

$$\gamma_{E,\mu}(t) := \mathbb{P}(Z + t\mu \in E).$$

*Then  $t \mapsto \gamma_{E,\mu}(t)$  is a non-increasing function.*

**Lemma 5.** *For all  $\alpha \in [0, 1]$  and for all  $u \in \mathbb{R}$  such that  $u \geq \bar{\Phi}(\alpha/2)$ , it holds that the function  $h_\alpha$  defined by (2.4) enjoys  $h_\alpha \geq 0$ ,  $h_\alpha$  is non-increasing, and  $h_\alpha$  goes to zero at infinity.*

**Proof.** First, note  $\alpha\bar{\Phi}(u) \leq \bar{\Phi}(u)$  and  $\bar{\Phi}$  is non-increasing, to get that  $h_\alpha \geq 0$ . Compute the derivative and use that  $h_\alpha \geq 0$  to show that  $h'_\alpha(u) \leq \alpha - 1 \leq 0$ . Eventually, we get that

$$\forall \alpha \in ]0, 1], \forall u \geq u_0, \quad \bar{\Phi}(u + a) \leq \alpha\bar{\Phi}(u).$$

As  $\bar{\Phi}$  is non-increasing, it implies that  $u + a \geq \bar{\Phi}^{-1}(\alpha\bar{\Phi}(u))$  so that  $a \geq h_\alpha(u) \geq 0$  which concludes the proof.  $\square$

**Lemma 6.**  $\forall \alpha \in [0, 1], \forall u \geq \bar{\Phi}^{-1}(\alpha/2), \forall v \geq 0$ , it holds

$$g_\alpha(u + v) \geq g_\alpha(u) + v,$$

where  $g_\alpha(u) = \Phi^{-1}(1 - \frac{1-\Phi(u)}{\alpha}) = \bar{\Phi}^{-1}(\bar{\Phi}(u)/\alpha)$ .

**Proof.** To prove (6), we show

$$\frac{\partial g_\alpha}{\partial u}(u, \alpha) = \frac{\varphi(u)}{\alpha \varphi(g_\alpha(u))} =: \frac{\varphi(u)}{j(u, \alpha)} \geq 1.$$

Use the fact that  $\forall u \in \mathbb{R}$ ,  $\varphi(u) \geq u \bar{\Phi}(u)$  to compute

$$\frac{\partial j}{\partial \alpha}(u, \alpha) = \varphi(g_\alpha(u)) - \frac{\bar{\Phi}(u)}{\alpha} g_\alpha(u) \geq 0$$

so

$$\frac{\partial g_\alpha}{\partial u}(u, \alpha) \geq \frac{\varphi(u)}{j(u, 1)} = 1,$$

as claimed. □

## Appendix B: Proofs of Section 4.1

### B.1. Proof of Proposition 6

Set  $R = R(\rho)$  for short. The proof relies on the fact that the eigenvectors of a two dimensional correlation matrix (such a  $R$ ) are fixed and coincide with the diagonals of  $\mathbb{R}^2$ . Symmetry of  $\mathcal{T}$  with respect to these diagonals is a key point in the proof.

First, we can use a  $\pi/4$  rotation and consider a variance–covariance matrix  $\bar{R}$  which is diagonal and a set  $\bar{\mathcal{T}}$  which is symmetric with respect to the two axis and whose intersections with a line parallel to one axe is an interval.

Let  $\bar{\Psi}$  the expression of the function  $\Psi$  after this rotation. We have to prove that  $\bar{\Psi}$  is non-increasing along the relevant half-axis. Fix, for example,  $\mu_1$  and consider, for  $\mu_2 > 0$  the function

$$\mu_2 \mapsto \bar{\Psi}(\mu) = \mathbb{P}(\mathcal{N}(\mu, \bar{R}) \in \bar{\mathcal{T}}) = \int_{-\infty}^{+\infty} \varphi\left(\frac{u - \mu_1}{\sigma_1}\right) \mathbb{P}(\mathcal{N}(\mu_2, \sigma_2^2) \in I_u) du,$$

where  $\sigma_1^2, \sigma_2^2$  are the diagonal elements of  $\bar{R}$ ,  $\mu = (\mu_1, \mu_2)$  and

$$I_u = \{v \in \mathbb{R} : (u, v) \in \bar{\mathcal{T}}\}.$$

Our hypotheses imply that for all  $u$ ,  $I_u$  is an interval that is symmetrical with respect to zero. Anderson's inequality (Lemma 4) implies directly that the Gaussian measure of  $I_u$  is non-increasing as a function of  $\mu_2$  so the function  $\mu_2 \mapsto \bar{\Psi}(\mu_1, \mu_2)$  is non-increasing. This gives half of the statement, the other statement is obtained exactly in the same fashion by exchanging the roles of  $\mu_1$  and  $\mu_2$ .

## B.2. Proof of Corollary 2

*Step 1:* If  $(U_1, U_2)$  has distribution  $\mathcal{N}(0, R)$ , then it is also the case of  $(U_1, -U_2)$ ,  $(-U_1, U_2)$  or  $(-U_1, -U_2)$ . This implies that  $\mathcal{R}_\alpha^c$ , which is computed under the null hypothesis, has the two required symmetry properties of Proposition 6.

*Step 2:* We consider now hypothesis (4.2). Consider  $(u_1, u_2) \in \mathcal{R}_\alpha^{+,1}$ . By definition of this region, it holds

$$u_1 \geq \bar{\Phi}^{-1}(\alpha/2),$$

and

$$g_\alpha(u_1)(1 - \rho) + \rho u_1 \geq u_2 \geq -g_\alpha(u_1)(1 + \rho) + \rho u_1.$$

Let  $r > 0$  and consider the points  $(u_1 + r, u_2 + r)$  and  $(u_1 + r, u_2 - r)$ . It is proven in Lemma 6 in the appendix that

$$g_\alpha(u_1 + r) \geq g_\alpha(u_1) + r.$$

As a consequence, for example,  $g_\alpha(u_1 + r)(1 - \rho) + \rho(u_1 + r) \geq g_\alpha(u_1)(1 - \rho) + \rho(u_1) + r$  and this implies directly that  $(u_1 + r, u_2 + r)$  and  $(u_1 + r, u_2 - r)$  belong to  $\mathcal{R}_\alpha^{+,1}$ . The intersections of  $\mathcal{R}_\alpha^{+,1}$  with the diagonals  $\Delta_\mu^\varepsilon$  are half lines or empty sets. We have the same results for the three other regions in the same fashion and this implies that the intersections of  $\mathcal{R}_\alpha^c$  with the diagonals  $\Delta_\mu^\varepsilon$  are intervals. Finally, this result is true in  $\beta$  because  $R$  preserves symmetry properties along the diagonals.

## Acknowledgment

We thank Pr. Franck Barthe for precious discussions. We would like to thank anonymous referees for their interesting and useful comments on early version of this work (especially regarding Section 4 which has been improved from valuable remarks and questions).

## References

- [1] Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. [MR0069229](#)
- [2] Azaïs, J.-M. and Genz, A. (2013). Computation of the distribution of the maximum of stationary Gaussian processes. *Methodol. Comput. Appl. Probab.* **15** 969–985. [MR3117635](#)
- [3] Azaïs, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Hoboken, NJ: Wiley. [MR2478201](#)
- [4] Bertin, K., Le Pennec, E. and Rivoirard, V. (2011). Adaptive Dantzig density estimation. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 43–74. [MR2779396](#)
- [5] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- [6] Bühlmann, P., Meier, L. and van de Geer, S. (2014). Discussion: “A significance test for the lasso”. *Ann. Statist.* **42** 469–477. [MR3210971](#)

- [7] Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Heidelberg: Springer. [MR2807761](#)
- [8] Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- [9] Candes, E.J. and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory* **52** 5406–5425. [MR2300700](#)
- [10] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. [MR1639094](#)
- [11] de Castro, Y. (2013). A remark on the lasso and the Dantzig selector. *Statist. Probab. Lett.* **83** 304–314. [MR2998757](#)
- [12] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#)
- [13] Fuchs, J.J. (2005). Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inform. Theory* **51** 3601–3608. [MR2237526](#)
- [14] Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.* **1** 141–149.
- [15] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed. Springer Series in Statistics. New York: Springer. [MR2722294](#)
- [16] Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton: CRC Press.
- [17] Juditsky, A. and Nemirovski, A. (2011). Accuracy guarantees for  $\ell_1$ -recovery. *IEEE Trans. Inform. Theory* **57** 7818–7839. [MR2895363](#)
- [18] Lee, J.D., Sun, D.L., Sun, Y. and Taylor, J.E. (2013). Exact post-selection inference with the lasso. Preprint. Available at [arXiv:1311.6238](#).
- [19] Lockhart, R., Taylor, J., Tibshirani, R.J. and Tibshirani, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- [20] Lockhart, R., Taylor, J., Tibshirani, R.J. and Tibshirani, R. (2014). Correction: Rejoinder to “A significance test for the lasso”. *Ann. Statist.* **42** 2138–2139. [MR3269975](#)
- [21] Lockhart, R., Taylor, J., Tibshirani, R.J. and Tibshirani, R. (2014). Rejoinder: “A significance test for the lasso”. *Ann. Statist.* **42** 518–531. [MR3210977](#)
- [22] Loftus, J.R. and Taylor, J.E. (2015). Selective inference in regression models with groups of variables. Preprint. Available at [arXiv:1511.01478](#).
- [23] Mourareau, S. (2015). Available at [arXiv:http://www.math.univ-toulouse.fr/~smourare/power\\_test.m](#). March 2015.
- [24] Nuyens, D. and Cools, R. (2006). Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces. *Math. Comp.* **75** 903–920 (electronic). [MR2196999](#)
- [25] Taylor, J.E., Lockhart, R., Tibshirani, R.J. and Tibshirani, R. (2014). Exact post-selection inference for forward stepwise and least angle regression. Preprint. Available at [arXiv:1401.3889](#).
- [26] Taylor, J.E., Loftus, J.R. and Tibshirani, R.J. (2014). Tests in adaptive regression via the Kac–Rice formula. Preprint. Available at [arXiv:1308.3020v3](#).
- [27] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- [28] van de Geer, S.A. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. [MR2576316](#)

Received October 2015 and revised May 2016