



A Rice method proof of the Null-Space Property over the Grassmannian

Jean-Marc Azaïs, Yohann de Castro, Stéphane Mourareau

► To cite this version:

Jean-Marc Azaïs, Yohann de Castro, Stéphane Mourareau. A Rice method proof of the Null-Space Property over the Grassmannian. 2014. hal-00995902v2

HAL Id: hal-00995902

<https://hal.science/hal-00995902v2>

Preprint submitted on 16 Jun 2014 (v2), last revised 8 Oct 2015 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A RICE METHOD PROOF OF THE NULL-SPACE PROPERTY OVER THE GRASSMANNIAN

J.-M. AZAÏS, Y. DE CASTRO, AND S. MOURAREAU

ABSTRACT. The Null-Space Property (NSP) is a necessary and sufficient condition for the recovery of the largest coefficients of solutions to an under-determined system of linear equations. Interestingly, this property governs also the success and the failure of recent developments in high-dimensional statistics, signal processing, error-correcting codes and the theory of polytopes.

Although this property is the keystone of ℓ_1 -minimization techniques, it is an open problem to construct deterministic matrices satisfying NSP. One can only prove that NSP holds, with an overwhelming probability, when the matrix is chosen at random, in general with a rotation invariant distribution.

In this article, we provide the first proof of NSP using random processes theory and the Rice method. As a matter of fact, our analysis gives non-asymptotic bounds for NSP with respect to rotation invariant distributions. Moreover, we derive a simple and explicit sufficient condition for NSP.

1. INTRODUCTION

1.1. Null-Space Property. One of the simplest inverse problem can be described as follows: given a matrix $X \in \mathbb{R}^{n \times p}$ and $y \in \text{Im}(X)$, can we faithfully recover β^* such that the identity $y = X\beta^*$ holds? In the ideal case where $n \geq p$ and the matrix X is one to one (namely, the model is identifiable), this problem is trivial. However, in view of recent applications in genetics, signal processing, or medical imaging, the frame of high-dimensional statistics is governed by the opposite situation where $n < p$. To bypass the limitations due to the lack of identifiability, one usually assumes that the matrix X is at random and one considers the ℓ_1 -minimization procedure [17]:

$$(P_{\ell_1}) \quad \Delta_X(\beta^*) \in \arg \min_{X\beta = X\beta^*} \|\beta\|_1,$$

where $\beta^* \in \mathbb{R}^p$ is a “target” vector we aim to recover. Interestingly, Program (P_{ℓ_1}) can be solved efficiently using linear programming, e.g. [14]. Furthermore, the high-dimensional models often assume that the target vector β^* belongs to the space Σ_s of s -sparse vectors:

$$\Sigma_s = \{\beta \in \mathbb{R}^p, \|\beta\|_0 \leq s\},$$

where $\|\beta\|_0$ denotes the size of the support of β . Note that this framework is the baseline of the flourishing Compressed Sensing (CS), see [13, 21, 18, 16] and references therein. A breakthrough brought by CS states that if the matrix X is drawn at random (e.g. X has i.i.d. standard Gaussian entries) then, with overwhelming probability, one can faithfully recover $\beta^* \in \Sigma_s$ using (P_{ℓ_1}) . More precisely, the interplay between randomness and ℓ_1 -minimization shows that with only $n = \mathcal{O}(s \log(p/s))$, one can faithfully reconstruct any s -sparse vector β^* from the

Date: June 16, 2014.

Key words and phrases. Rice Method; High-dimensional statistics; ℓ_1 -minimization; Null-Space Property; Random processes theory.

knowledge of X and $y := X\beta^*$. Notably, this striking fact is governed by the Null-Space Property (NSP).

Definition (Null-Space Property of order s and dilatation C) — *Let $0 < s < p$ be two integers and G be a sub-space of \mathbb{R}^p . One says that the sub-space G satisfies $\text{NSP}(s, C)$, the Null-Space Property of order s and dilatation $C \geq 1$, if and only if:*

$$\forall \gamma \in G, \forall S \subset \{1, \dots, p\} \text{ s.t. } |S| \leq s, \quad C \|\gamma_S\|_1 \leq \|\gamma_{S^c}\|_1,$$

where S^c denotes the complement of S , the vector γ_S has entry equal to γ_i if $i \in S$ and 0 otherwise, and $|S|$ is the size of the set S .

As a matter of fact, one can prove [18] that the operator Δ_X is the identity on Σ_s if and only if the kernel of X satisfies $\text{NSP}(s, C)$ for some $C > 1$.

Theorem 1 ([18]) — *For all $\beta^* \in \Sigma_s$ there is a unique solution to (P_{ℓ_1}) and $\Delta_X(\beta^*) = \beta^*$ if and only if the nullspace $\ker(X)$ of the matrix X enjoys $\text{NSP}(s, C)$ for some $C > 1$. Moreover, if $\ker(X)$ enjoys $\text{NSP}(s, C)$ for some $C > 1$ then for all $\beta^* \in \mathbb{R}^p$,*

$$\|\beta^* - \Delta_X(\beta^*)\|_1 \leq \frac{2(C+1)}{C-1} \min_{|S| \leq s} \|\beta^* - \beta_S^*\|_1.$$

Additionally, NSP suffices to show that any solution to (P_{ℓ_1}) is comparable to the s -best approximation of the target vector β^* . Theorem 1 demonstrates that NSP is a natural property that should be required in CS and High-dimensional statistics. This analysis can be lead a step further considering Lasso [32] or Dantzig selector [15]. Indeed, in the frame of noisy observations, ℓ_1 -minimization procedures are based on sufficient conditions like Restricted Isometry Property (RIP) [15], Restricted Eigenvalue Condition (REC) [12], Compatibility Condition (CC) [33], Universal Distortion Property (UDP) [20], or $H_{s,1}$ condition [26]. Note that all of these properties imply that the kernel of the matrix X satisfies NSP. While there exists pleasingly ingenious and simple proofs of RIP, see [16] for instance, a direct proof of NSP (without the help of RIP) remains a challenging issue.

1.2. Direct proofs of NSP with dilatation $C = 1$. To the best of our knowledge, all the direct proofs of NSP with dilatation $C = 1$ are based either on integral convex geometry theory, Gaussian widths, the approximate kinematic formula, or empirical process theory.

1.2.1. Grassmann angles. In a captivating series of papers [23, 22, 25, 24], Donoho and Tanner have proved that the kernel of a matrix $X(n, p_n) \in \mathbb{R}^{n \times p_n}$ with i.i.d. centered Gaussian entries enjoys a phase transition, i.e. there exists a function $\rho_S:]0, 1[\rightarrow]0, 1[$ such that for all $(\rho, \delta) \in]0, 1[^2$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\ker(X(n, p_n)) \text{ enjoys } \text{NSP}(s_n, 1)] = \begin{cases} 0 & \text{if } \rho > \rho_S(\delta), \\ 1 & \text{if } \rho < \rho_S(\delta), \end{cases}$$

where $s_n = \lfloor \rho n \rfloor$, $p_n = \lfloor \frac{n}{\delta} \rfloor$ and $\lfloor \cdot \rfloor$ denotes the integer part. Moreover, they have characterized implicitly and computed numerically the function ρ_S (note that the subscript S stands for “Strong” since ρ_S is often named the “strong threshold”). Observe their approach is based on computation of Grassmann angles of a polytope due to Affentranger and Schneider [4] and Vershik and Sporyshev [34]. Furthermore, note their phase transition is characterized implicitly using an equation involving inverse Mills ratio of the standard normal density. However, they have derived a nice explicit expression of the phase transition for small values of δ , i.e. when $\delta \rightarrow 0$. Hence, they uncover that, in the regime $n \ll p$, $\text{NSP}(s, 1)$ holds when:

$$n \gtrsim 2s \log\left(\frac{p}{s}\right),$$

for n large enough.

1.2.2. Gaussian widths. In recent works [30, 31], Stojnic have shown a simple characterization of the sign of the exponent appearing in the expression of the “weak threshold” given by Donoho and Tanner. Note the weak threshold governs the exact reconstruction by ℓ_1 -minimization of s -sparse vectors with prescribed support and signs, while NSP characterizes the exact reconstruction of all s -sparse vectors. In the paper [30], using Gordon’s escape through a mesh theorem, Stojnic have derived a simpler implicit characterization of the strong threshold ρ_S . As in Donoho and Tanner’s work, observe this implicit characterization involves inverse Mill’s ratio of the normal distribution and no explicit formulation of ρ_S can be given.

1.2.3. Approximate kinematic formula. In the papers [28, 5], the authors present appealing and rigorous quantitative estimates of weak thresholds appearing in convex optimization, including the location and the width of the transition region. Recall that NSP is characterized by the strong threshold. Nevertheless, the weak threshold describes a region where NSP cannot be satisfied, i.e.

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\ker(X(n, p_n)) \text{ enjoys NSP}(s_n, 1)] = 0.$$

Based on the approximate kinematic formula, the authors have derived recent fine estimates of the weak threshold. Although their result has not been stated for the strong threshold, their work should provide, invoking a simple union bound argument, a direct proof of NSP with dilatation $C = 1$.

1.2.4. Empirical process theory. Using empirical process theory under weak moment assumptions, a recent result of Lecué and Mendelson [27] gives a direct proof of NSP with dilatation $C = 1$ for matrices X with sub-exponential rows (not necessarily rotation invariant). Observe that, in this case, the kernel $\ker(X)$ can be drawn uniformly on the Grassmannian. Although the authors do not pursue an expression of the strong threshold, their work shows that NSP with dilatation $C = 1$ holds, with overwhelming probability, for parameters satisfying:

$$n \geq c_0 s \log\left(\frac{ep}{n}\right),$$

with $c_0 > 0$ a universal (unknown) constant. Observe we provide the same flavor of results in Theorem 3 with explicit constants.

1.3. A previous direct proof of NSP with dilatation $C \geq 1$. Using integral convex geometry theory as in Donoho and Tanner’s works [23, 22, 25, 24], Xu and Hassibi have investigated [35, 36] the property $\text{NSP}(s, C)$ for values $C \geq 1$. However, their result uses an implicit equation involving inverse Mill’s ratio of the normal distribution and no explicit formulation of their thresholds can be derived. To the best of our knowledge, this is the only proof of $\text{NSP}(s, C)$ for values $C > 1$ predating this paper.

2. NEW SIMPLE AND EXPLICIT BOUND FOR NSP WITH DILATATION $C \geq 1$

In this paper, we follow a different path using stochastic processes method and more precisely the Rice method [9, 11]. This latter is specially design to study the tail of the maximum of differentiable random processes or random fields. As it is done for a deterministic function, it consists of studying the maximum through zeroes of the derivative in quadratic mean. For the tail of a stationary Gaussian process defined on the real line, it is known from the work of Piterbarg [29] that it is super-exponentially sharp. More precisely, if the process $X(t)$ is normalized by:

$$\mathbb{E}(X(t)) = 0, \quad \text{Var}(X(t)) = 1, \quad \text{and} \quad \text{Var}(X'(t)) = 1,$$

(which can be always achieved considering an appropriate scaling, dilatation and translation of the original stochastic process), then it holds:

$$\forall u > 0, \quad \mathbb{P}\left\{\sup_{t \in [0, T]} X(t) > u\right\} = \frac{e^{-\frac{u^2}{2}}}{2\pi} T + 1 - \Phi(u) + \mathcal{O}(\phi(u(1+c))),$$

where ϕ denotes the density function of the standard Gaussian distribution with respect to the Lebesgue measure, Φ the cumulative distribution function of the standard Gaussian, and c is some positive real number. Besides, when the parameter t is multi-dimensional, i.e. for random fields, the Rice method [8, 10] is equivalent to the Euler Characteristic method of Adler and Taylor [3] and gives, under some mild conditions, super exponentially sharp results. As we will see later, it can be a powerful tool to establish results on Compressed Sensing and High-dimensional statistics properties as NSP.

However, the situation is more complicated than in the aforementioned papers since the considered process $X(t)$ is:

- defined on the sphere (as in the recent work [7] for example),
- non Gaussian,
- and, last but not least, non differentiable (as in [19] nevertheless, a contrario to this last paper, we don't use a smoothing argument but a simple partition of the sphere).

Interestingly, our analysis relies on the Rice formula applied to locally linear processes defined on the sphere and it provides a short and direct proof of NSP(s, C) for any value $C \geq 1$.

2.1. An explicit sufficient condition. In its more involved form, our result reads as follows.

Theorem 2 (Explicit lower bound) — *Let $0 < s < n < p$. Let $X(n, p)$ a $(n \times p)$ random matrix with i.i.d. centered Gaussian entries. For all $C \geq 1$, it holds:*

$$\mathbb{P}[\ker(X(n, p)) \text{ enjoys NSP}(s, C)] = 1 - \Pi,$$

where:

$$(1) \quad \Pi \leq \psi(C) \left[\sum_{k=0}^{p-n-1} \binom{p}{k} \sqrt{\pi} (C^2 - 1 + \frac{p-k}{s})^{\frac{n+k+1-p}{2}} \frac{\Gamma(\frac{2p-2k-n-1}{2})}{\Gamma(\frac{p-k}{2})\Gamma(\frac{p-n-k}{2})} \right. \\ \left. \times C^{p-n-k-1} \left[\frac{H_k!}{H_k^{H_k+k-p+n} (p-n-k)!} \right]^{\frac{1}{2}} \right]$$

with:

$$\psi(C) = 2 \left[\frac{s \log(eC_0 \frac{p}{s})}{C_0 p} \right]^{\frac{p}{2}}, \quad C_0 = \frac{\pi}{2e^2(1+C)^2},$$

$$\text{and } H_k = \lfloor (\frac{\pi}{2}((C^2 - 1)s + p - k) \wedge (p - n)) \rfloor.$$

Since $\psi(C)$ assesses the concentration of one Gaussian vector (see Lemma 4), the expression into bracket in (1) can be seen as a complexity bound. Each terms of the aforementioned sum are treated in Appendix B. In particular, an overall bound is given in Lemma 6. From (1), we can describe (numerically, see Figure 1) a region (ρ, δ) where NSP is satisfied, i.e.

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\ker(X(n, p_n)) \text{ enjoys NSP}(s_n, 1)] = 1,$$

where $s_n = \lfloor \rho n \rfloor$, $p_n = \lfloor \frac{n}{\delta} \rfloor$ and $X(n, p_n)$ is a $(n \times p_n)$ random matrix with i.i.d. centered Gaussian entries. Interestingly, our lower bound compares (up to a multiplicative factor less than 14 for $\delta > 0.7$) to the phase transition of Donoho and Tanner [23], see Figure 1.

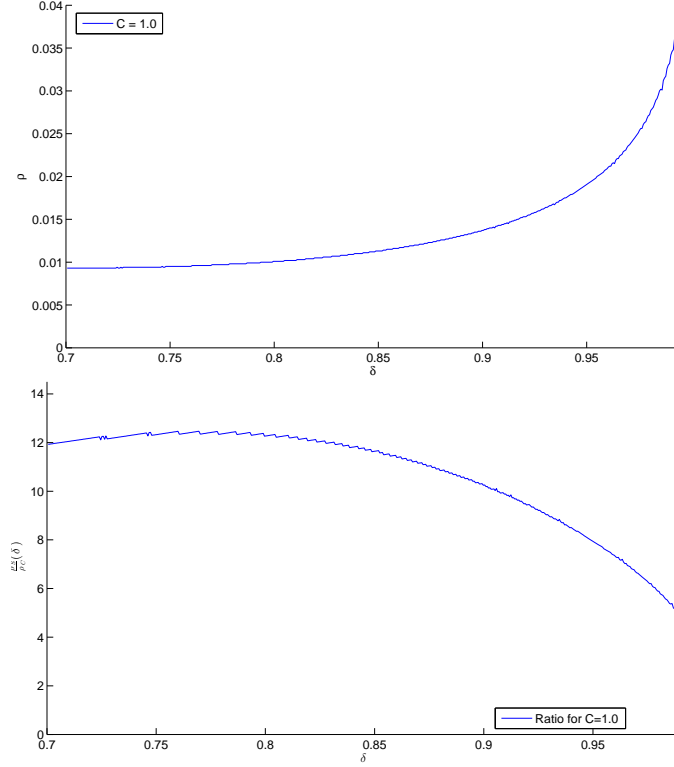


FIGURE 1. The top panel illustrates numerically the threshold $\delta \rightarrow \rho_{C,D}(\delta)$ appearing in Theorem 2 for $C = 1$ and $D = 0.01$. Note that the region (ρ, δ) such that $\Pi \simeq 0$, i.e. NSP($s, 1$) holds, is located below the curve. The bottom panel shows the ratio between the threshold of Theorem 2 and the strong phase transition of Donoho and Tanner [23].

2.2. A simpler and explicit sufficient condition. In its simpler (and weaker) form, our result gives a simple and explicit lower bound $\rho_{C,D}$ on the phase transition for NSP with dilatation $C > 1$.

Theorem 3 (Simple and explicit lower bound) — *For all $C \geq 1$ and $D > 0$, for all $(\rho, \delta) \in]0, 1[\times]0, 1[$ and $\rho < \rho_{C,D}(\delta)$,*

$$(2) \quad \mathbb{P}[\ker(X(n, p_n)) \text{ enjoys NSP}(s_n, C)] \geq 1 - \frac{9.7}{C} \left(\frac{n}{\delta}\right)^{\frac{3}{2}} \exp(-nD),$$

where $s_n = \lfloor \rho n \rfloor$, $p_n = \lfloor \frac{n}{\delta} \rfloor$, $X(n, p_n)$ is a $(n \times p_n)$ matrix with i.i.d. centered Gaussian entries, and $\rho_{C,D}$ is characterized by:

$$\rho_{C,D}(\delta) = \begin{cases} \zeta_{(\delta,C)}^{-1}(-D - \frac{4-\delta}{2\delta} \log 2 - \frac{1-\delta}{\delta} \log C) & \forall \delta \in]0, 1/2], \\ \zeta_{(\delta,C)}^{-1}(-D - \frac{2-\delta}{2\delta} \log 2 - \frac{1}{\delta} \mathbf{H}_e(\delta) - \frac{1-\delta}{\delta} \log C) & \forall \delta \in]1/2, 1[, \end{cases}$$

where $\mathbf{H}_e(\delta) = -\delta \log(\delta) - (1-\delta) \log(1-\delta)$,

$$\zeta_{(\delta,C)}(\rho) = \log\left(-\frac{\rho\delta}{C_0} \log\left(\frac{\rho\delta}{eC_0}\right)\right) \quad \text{and} \quad C_0 := \frac{\pi}{2e^2(1+C)^2}.$$

Observe that (2) depends only on the kernel of X . In particular, the result holds for any value of the variance of the entries of $X(n, p)$. As a matter of fact, the result holds for any kernel drawn at random uniformly on the Grassmannian $\text{Gr}_m(\mathbb{R}^p)$

whith $m = p - n$, namely the compact group of m -dimensional sub-spaces of \mathbb{R}^p . Note that this is the case when the lines of X are drawn independently from (possibly distinct) rotation invariant distributions.

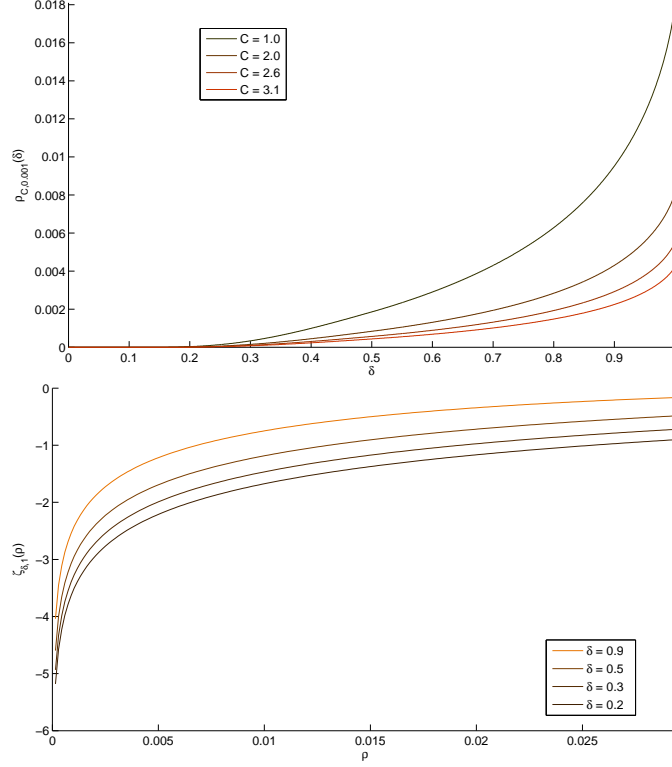


FIGURE 2. The functions $\rho_{(C,0.1)}$ (left panel) and $\zeta_{(\delta,1)}$ (right panel) of Theorem 3.

Since the inverse function of $\zeta_{(\delta,C)}$ is not explicit, we apply the useful identity (see Figure 3):

$$\forall x \in]0, 1[, \quad \log(-x \log(\frac{x}{e})) \leq \frac{1}{2} \log(x(2-x)),$$

to get that:

$$\forall u \in]-\infty, 0[, \quad \zeta_{(\delta,C)}^{-1}(u) \geq 1 - (1 - \exp(2u))^{\frac{1}{2}}.$$

We deduce the following corollary:

Corollary (Explicit lower bound on the phase transition) — *Theorem 3 is still true if one replaces $\rho_{(\delta,C)}$ by $\rho'_{(\delta,C)}$ defined by:*

$$\rho'_{C,D}(\delta) = \begin{cases} 1 - (1 - e^{-2D}(\frac{1}{2})^{\frac{4-\delta}{2}}(\frac{1}{C})^{\frac{2(1-\delta)}{2}})^{\frac{1}{2}} & \forall \delta \in]0, 1/2], \\ 1 - (1 - e^{-2D}(\frac{1}{2})^{\frac{2-\delta}{2}}(\frac{1}{C})^{\frac{2(1-\delta)}{2}}e^{-\frac{2}{\delta}\mathbf{H}_e(\delta)})^{\frac{1}{2}} & \forall \delta \in]1/2, 1[. \end{cases}$$

We see that in this last form, the lower bound on the phase transition is totally explicit and can be expressed in terms of simple functions.

3. PROOF OF THE MAIN RESULT

3.1. Model and notation. Let $0 < s < n < p$, let $C > 1$ and set $m = p - n$. Let G be uniformly distributed on the Grassmannian $\text{Gr}_m(\mathbb{R}^p)$. Observe that it

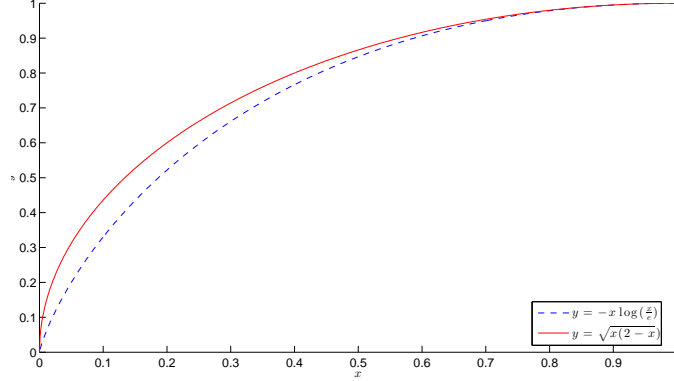


FIGURE 3. Comparison between $x \mapsto -x \log(\frac{x}{e})$ (dotted line) and $x \mapsto \sqrt{x(2-x)}$ (plain line). Observe that this latter has an explicit and simple inverse function on $[0, 1]$.

can be generated by m standard Gaussian vectors $g_i \in \mathbb{R}^p$ for $i = 1, \dots, m$. Define $\{Z(t); t \in \mathbb{S}^{m-1}\}$ the process with values in \mathbb{R}^p given by:

$$Z(t) := \sum_{i=1}^m t_i g_i.$$

Note this process spans G and it can be written as:

$$\text{for } j = 1, \dots, p, \quad Z_j(t) = \langle t, g^j \rangle,$$

where $(g^j)_{j=1}^p$ are independent Gaussian random vectors of size m with standard distribution in \mathbb{R}^m . Let $O \in \mathbb{R}^{p \times p}$ be an orthogonal matrix. Remark that, due to rotation invariance of the Gaussian distribution, it holds:

$$\forall t \in \mathbb{S}^{m-1}, \quad OZ(t) \sim Z(t).$$

Consider the order statistic of the absolute values of the coordinates of $Z(t)$:

$$|Z_{(1)}(t)| \geq \dots \geq |Z_{(p)}(t)|.$$

Given a sparsity s , a degree of freedom m , and a degree of constraint p , consider the real valued process $\{X(t); t \in \mathbb{S}^{m-1}\}$:

$$(3) \quad X(t) = C|Z_{(1)}(t)| + \dots + C|Z_{(s)}(t)| - [|Z_{(s+1)}(t)| + \dots + |Z_{(p)}(t)|].$$

NSP is equivalent to the fact that this process is always non positive. We will prove that it happens with an overwhelming probability, see Section 3.3.

3.2. Cutting the sphere out. Let $A \subseteq \{1, \dots, p\}$, define the random subsets \mathbb{S}_A and $\dot{\mathbb{S}}_A$ of the unit sphere \mathbb{S}^{m-1} by:

$$\mathbb{S}_A = \{t \in \mathbb{S}^{m-1}; \quad Z_A(t) = 0\},$$

$$\dot{\mathbb{S}}_A = \{t \in \mathbb{S}^{m-1}; \quad Z_A(t) = 0 \text{ and } \forall j \notin A, \quad Z_j(t) \neq 0\}.$$

One can check that \mathbb{S}_A is the unit sphere of the orthogonal of $\text{Span}\{g^j; j \in A\}$. This implies that a.s. \mathbb{S}_A is a random sphere of dimension $m-1-|A|$ if $|A| \leq m-1$ and is almost surely empty if $|A| > m-1$. It follows that the quantities $|Z_{(1)}(t)|, \dots, |Z_{(n+1)}(t)|$ are a.s. positive and that a.s.

$$\mathbb{S}^{m-1} = \bigcup_{|A| \leq m-1} \dot{\mathbb{S}}_A,$$

giving a partition of the sphere. We define also, for later use, the random subset W by $W := \{t \in \mathbb{S}^{m-1}; |Z_{(s)}(t)| = |Z_{(s+1)}(t)|\}$. Observe that, conditionally to g^j , the set W is closed with empty interior.

3.3. Probability of failure. We consider the probability:

$$\Pi := \mathbb{P}\{\mathcal{M} > 0\} \leq \sum_{|A| \leq m-1} \mathbb{P}\{\mathcal{M}_{\dot{\mathbb{S}}_A} > 0\},$$

where \mathcal{M} and $\mathcal{M}_{\dot{\mathbb{S}}_A}$ are respectively the number of local maxima of $X(t)$ along \mathbb{S}^{m-1} satisfying $X(t) > 0$ and living in \mathbb{S}^{m-1} , respectively in $\dot{\mathbb{S}}_A$. The baseline of our proof is to upper-bound each right hand side probability, using the expectation of the number of local maximum above zero and the Markov inequality. The first element is Lemma 4 that proves that:

$$\forall t \in \mathbb{S}^{m-1}, \quad \mathbb{P}\{X(t) > 0\} \leq \psi(C),$$

where $\psi(C) = 2 \left[\frac{s \log(\epsilon C_0 \frac{p}{2})}{C_0 p} \right]^{\frac{p}{2}}$ with $C_0 = \frac{\pi}{2e^2(1+C)^2}$. The second element is that $X(t)$ admits a density p_X . To check that, note that $|Z_{(1)}|, \dots, |Z_{(p)}|$ are the order statistics of the absolute values of i.i.d. Gaussian variables and thus they have a known joint density on the simplex $|Z_{(1)}| \geq \dots \geq |Z_{(p)}|$. Formula (3) implies the existence of a density for $X(t)$. Moreover, this density does not depend on t due to invariance of Gaussian distribution.

3.4. Initialization: local maxima on $\dot{\mathbb{S}}_\emptyset$. By considering the symmetry properties of the sphere $\dot{\mathbb{S}}_\emptyset$, we have

$$\mathbb{P}\{\mathcal{M}_{\dot{\mathbb{S}}_\emptyset} > 0\} \leq \frac{1}{2} \mathbb{E}(\mathcal{M}_{\dot{\mathbb{S}}_\emptyset}).$$

In this part, our aim will be to give bound to the expectation using a Kac-Rice formula. One can check that if t belongs to $\dot{\mathbb{S}}_\emptyset$ and does not belong to W , $X(t)$ is locally the sum of the absolute values of some s coordinates multiplied by C minus the sum of the absolute values of the other coordinates. It can be written as:

$$X(t) = C\epsilon_1 Z_{(1)}(t) + \dots + C\epsilon_s Z_{(s)}(t) + \epsilon_{s+1} Z_{(s+1)}(t) + \dots + \epsilon_p Z_{(p)}(t),$$

where $\epsilon_1, \dots, \epsilon_p$ are random variables taking values ± 1 .

Lemma 1 — *Let $t \in \mathbb{S}^{m-1}$ and denote by P_{t^\perp} the orthogonal projection onto the orthogonal of t . Then the random vector*

$$(4) \quad P_{t^\perp}(C\epsilon_1 g^{(1)} + \dots + C\epsilon_s g^{(s)} + \epsilon_{s+1} g^{(s+1)} + \dots + \epsilon_p g^{(p)}),$$

is a centered isotropic Gaussian random vector onto t^\perp with variance $(sC^2 + (p-s))$ that is independent of $(\epsilon_1, \dots, \epsilon_p, (1), \dots, (p))$ where ϵ_i , resp. (k) , denotes the sign, resp. the indice, of the k -th largest entries (taking the natural order in case of ties) in absolute value of $Z(t)$.

Proof. Remark $Z(t)$ is a function of $(P_t(g^1), \dots, P_t(g^p)) = (Z_1(t)t, \dots, Z_p(t)t)$ with obvious notation. Note the orthogonal projection $(P_{t^\perp}(g^1), \dots, P_{t^\perp}(g^p))$ onto the orthogonal of t is independent of $(\epsilon_1, \dots, \epsilon_p, (1), \dots, (p))$. Moreover, observe that the random vector (4) can be written as:

$$\mathbf{A}(P_{t^\perp}(g^1), \dots, P_{t^\perp}(g^p)),$$

where $\mathbf{A} : (t^\perp)^p \rightarrow t^\perp$ is a random linear operator which is a function of the random variables $(\epsilon_1, \dots, \epsilon_p, (1), \dots, (p))$. In particular, \mathbf{A} is independent of the random vector $(P_{t^\perp}(g^1), \dots, P_{t^\perp}(g^p))$ which is a standard Gaussian vector on $(t^\perp)^p$. We deduce that the random vector (4) is a Gaussian random vector on t^\perp with variance operator $\mathbf{A}^* \mathbf{A} = (sC^2 + (p-s))\text{Id}_{t^\perp}$, where Id_{t^\perp} is the identity operator on t^\perp . \square

Since the distribution of the process $Z(t)$ is invariant by isometries of \mathbb{R}^p , it turns that all the 2^p possible signs $\epsilon_1, \dots, \epsilon_p$ and all the $p!$ orderings $(1), \dots, (p)$ play the same role. To it suffices consider the particular case (that we will call the condition \mathcal{C}) where:

$$(C) \quad X(t) = CZ_1(t) + \dots + CZ_s(t) - Z_{s+1} - \dots - Z_p(t),$$

an observe that $X(t)$ is locally linear and thus differentiable.

The next step is to prove that a.s. there is no local maximum on W . The case where there are tied among the $|Z_i(t)|$ has to be considered though it happens with probability 0 for a fixed t . Note that in that case, the order statistics remain uniquely defined while the ordering $(1), \dots, (p)$ is no more uniquely defined.

Suppose that $t \in W$ and that, to simplify notation and without loss of generality, similarly to condition \mathcal{C} , we have: $Z_1(t) \geq \dots \geq Z_{s-h-1}(t) > Z_{s-h}(t) = \dots = Z_{s+k}(t) > Z_{s+k+1}(t) \geq \dots \geq Z_p(t)$ then, for w in some neighborhood N , (not included in W) of t :

$$X(w) = CZ_1(w) + \dots + CZ_{s-h-1}(w) + (1+C)\text{Max}_h(Z_{s-h}(w) + \dots + Z_{s+k}(w)) - (Z_{s-h}(w) + \dots + Z_p(w)),$$

where Max_h is the sum of the h largest element of a set. As being the maximum of $\binom{h}{(s+k)+1}$ linear forms the function Max_h is convex

Let us consider in detail the vectors g^{s-h}, \dots, g^{s+k} . With probability 1, they are pairwise different. The point t is such that their projection on t coincide. As a consequence the derivatives of the linear forms $Z_\ell(w) = \langle g^\ell, w \rangle$, $\ell = (s-h) \dots (s+k)$ on the tangent space orthogonal to t are pairwise different. This implies that the function Max_h has some direction in which it is strictly convex and as a consequence t cannot be a local maxima.

Suppose that $t \notin W$, conditionally the g^j 's the function $X(t)$ is locally linear and hence differentiable. Without loss of generality we can assume we are under the condition \mathcal{C} :

$$X(t) = CZ_1(t) + \dots + CZ_s(t) - [Z_{s+1}(t) + \dots + Z_p(t)].$$

The gradient in \mathbb{R}^m of this function is given by:

$$\dot{X}(t) = Cg^1 + \dots + Cg^s - [g^{s+1} + \dots + g^p].$$

As a consequence of formula of geometry (it can be proved by local parametrization), the derivative along the sphere $X'(t)$ is the tangent projection of the free gradient. In our case, it reads as:

$$X'(t) = P_{t^\perp} \dot{X}(t),$$

where P_{t^\perp} is the projector onto the orthogonal of t . The second derivative $X''(t)$ along the unit sphere is given by the tangent projection of the second derivative (that vanishes) minus identity multiplied by the normal derivative (with the well chosen orientation). It holds:

$$X''(t) = -\langle \dot{X}(t), t \rangle I_{m-1}.$$

It follows that $X'(t) = P_{t^\perp} [Cg^1 + \dots + Cg^s - (g^{s+1} + \dots + g^p)]$ and $X''(t) = -X(t)I_m$. This has several implications:

- First, we see that $X'(t)$ and $(X(t), X''(t))$ are independent.
- Second, the distribution of $P_{t^\perp} g^1, \dots, P_{t^\perp} g^p$ is not affected by the conditioning on $P_t g^1, \dots, P_t g^p$ implied by condition \mathcal{C} . This implies that on t^\perp , $X'(t)$ is centered Gaussian vector with variance:

$$q_C := (C^2 - 1)s + p.$$

This is what we obtain under the condition \mathcal{C} but since every condition plays the same role and since the distribution does not depend on the condition, it shows clearly that for every t , $X'(t)$ is Gaussian though $X(t)$ is obviously not, see Lemma 1.

- Third, if we limit our attention to points t such that $X(t) > 0$, $X''(t)$ cannot be singular.

This last condition implies that we can apply Theorem 5.1.1 of [2]. This lemma is a Kac type formula that shows that the zeroes of the derivative $X'(t)$ are isolated and thus in finite number. In addition recalling that $\mathcal{M}_{\dot{\mathbb{S}}_0}$ is the number of local maxima of $X(t)$ satisfying $X(t) > 0$ and belonging to $\dot{\mathbb{S}}_0$, this number satisfies

$$\mathcal{M}(\dot{\mathbb{S}}_0) = \lim_{\delta \rightarrow 0} \frac{1}{V(\delta)} \int_{\mathbb{S}^{m-1}} \mathbb{E}(|\det X''(t)| \mathbf{1}_{|X'(t)-0|<\delta} \mathbf{1}_{t \in \dot{\mathbb{S}}_0} \mathbf{1}_{X(t)>0}) \sigma(dt),$$

where σ is the surfacic measure on \mathbb{S}^{m-1} and $V(\delta)$ is the volume of the ball $B(\delta)$ with radius δ . Passing to the limit using the Fatou lemma gives:

$$\begin{aligned} \mathbb{E}(\mathcal{M}(\dot{\mathbb{S}}_0)) &\leq \liminf_{\delta \rightarrow 0} \int_0^\infty dx \int_{\dot{\mathbb{S}}_0} dt p_{X(t)}(x) \\ &\quad \frac{1}{V(\delta)} \int_{B(\delta)} dx' p_{X'(t)}(x') \mathbb{E}(|\det(X''(t))| \mid X(t) = x, X'(t) = x'), \\ &\leq (2\pi q_C)^{\frac{1-m}{2}} 2 \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} \int_0^\infty x^{m-1} p_{X(t)}(x) dx, \end{aligned}$$

where $p_{X(t)}(x)$ denotes the density of $X(t)$ at x and Γ denotes the Gamma function. Note that we have used:

- the fact that every point t is equivalent so we can replace the integral on the unit sphere by the volume of the unit sphere $2\pi^{\frac{m}{2}}/\Gamma(\frac{m}{2})$ and the value at a given point.
- $\mathbb{E}(|\det(X''(t))| \mid X(t) = x, X'(t) = x') = X^{m-1}$
- the gaussian density $p_{X'(t)}(x')$ is bounded by $(2\pi q_C)^{\frac{1-m}{2}}$.

So it remains to bound:

$$\mathbb{E}[(X(t)^+)^{m-1}].$$

For that purpose we write $X(t)$ as the independent product $\|Z(t)\|_2 Y(t)$, where the process $Y(t)$ is constructed exactly as the process $X(t)$ but starting now from a uniform distribution U on the unit sphere \mathbb{S}^{p-1} instead of the standard Gaussian distribution of $Z(t)$. Using standard results on the moments of the χ^2 distribution we have:

$$\mathbb{E}((X(t)^+)^{m-1}) = 2^{\frac{m-1}{2}} \frac{\Gamma(\frac{m-1+p}{2})}{\Gamma(\frac{p}{2})} \mathbb{E}((Y(t)^+)^{m-1}).$$

We use now the fact that $Y(t) \leq C\sqrt{s}$ to get that:

$$\mathbb{E}((Y(t)^+)^{m-1}) \leq (C\sqrt{s})^{m-1} \mathbb{P}\{Y(t) > 0\}.$$

Moreover, Lemma 4 shows that, with probability greater than $1 - \psi(C)$, a standard Gaussian vector g in \mathbb{R}^p enjoys:

$$(C+1)\sqrt{s}\|g\|_2 \leq \|g\|_1.$$

This implies that:

$$(5) \quad \mathbb{P}\{Y(t) > 0\} \leq \psi(C),$$

and consequently the probability of having a local maxima above 0 on $\dot{\mathbb{S}}_0$ is bounded by:

$$\begin{aligned}
& \sigma(\dot{\mathbb{S}}_\emptyset) (2\pi q_C)^{\frac{1-m}{2}} 2^{\frac{m-1}{2}} \frac{\Gamma(\frac{m-1+p}{2})}{\Gamma(\frac{p}{2})} C^{m-1} \psi(C) \\
(6) \quad & \leq 2\sqrt{\pi} \left(C^2 - 1 + \frac{p}{s} \right)^{\frac{1-m}{2}} \frac{\Gamma(\frac{m-1+p}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{m}{2})} C^{m-1} \psi(C) =: h(s, m, p).
\end{aligned}$$

3.5. Induction: maxima on smaller spheres. We turn now to the study of the first term that corresponds to existence of a maximum on $\dot{\mathbb{S}}_A$. It leads exactly to the same computations as the case $\dot{\mathbb{S}}_\emptyset$ because conditionally to the $g^j, j \in A$, we can replace the space \mathbb{R}^m by the orthogonal of V and obtain the same result replacing m by $m - k$ and p by $p - k$:

$$\begin{aligned}
h(s, m - k, p - k) &:= 2\pi^{\frac{1}{2}} \left(C^2 - 1 + \frac{p - k}{s} \right)^{\frac{k+1-m}{2}} \\
&\times \frac{\Gamma(\frac{m-1+p-2k}{2})}{\Gamma(\frac{p-k}{2})\Gamma(\frac{m-k}{2})} C^{m-k-1} \psi(C).
\end{aligned}$$

Observe that $\psi(C)$ does not depend on k , see (5). However, we have to pay attention to the value $k = m - 1$. In such a case \mathbb{S}_A consists of two points giving the same value of $X(t)$ and the first term corresponds only to the probability of $X(t)$ to be positive which is $\psi(C)$.

3.6. Refinements: localization of the sub-gradient. Let us consider now the case of a maximum on $\dot{\mathbb{S}}_A$, $A \neq \emptyset$ and set $k = |A|$. First we have to draw the g^i , $i \in A$ and define the space $V = \text{Span}\{g^i; i \in A\}$ and then conditionally to these vectors we consider \mathbb{S}_A , the random sphere orthogonal to these vectors. Note that the behavior of $X(t)$ on \mathbb{S}_A is independent of $\{g^i, i \in A\}$ since for all $t \in \mathbb{S}_A$, $X(t)$ depends only on the $g^j, j \notin A$ and that for such j ,

$$Z_j(t) = \langle g^j, t \rangle = \langle \Pi_{V^\perp} g^j, t \rangle.$$

A point $t \in \dot{\mathbb{S}}_A \setminus \mathcal{W}$ is a local maximum on \mathbb{S}^{m-1} if it is a local maximum along \mathbb{S}_A and its super-gradient in the orthogonal space contains zero. Indeed, locally around t , the behavior of $X(w)$ along V is the sum of some linear forms (for $j \notin A$) and of absolute value of linear forms (for $j \in A$) thus it is locally concave and we can define its super-gradient as the opposite of the sub-gradient of $-X$.

As we have seen, the two conditions are independent. Let us look to the second one. Thus we have to compute the probability of the super-gradient to contain zero. More precisely for s in a neighborhood of $t \in \dot{\mathbb{S}}_A \setminus \mathcal{W}$,

$$X(w) = X_A(w) + X_{A^c}(w),$$

where, because $s \leq n$:

$$X_A(w) = - \sum_{i \in A} |Z_i(w)|.$$

Around t , $X_{A^c}(s)$ is differentiable and, with a possible harmless change of sign (see Lemma 1), its gradient is given by

$$\sum_{i \in A} C_i g^i,$$

where the coefficient C_i takes the value C for s of them and -1 for the others. This gradient is distributed as an isotropic normal variable $\xi \in V$ with variance:

$$\tilde{p}_k := (C^2 - 1)s + p - k.$$

By this we mean that the distribution of ξ , in a convenient basis, is $\mathcal{N}(0, \tilde{p}_k I_k)$.

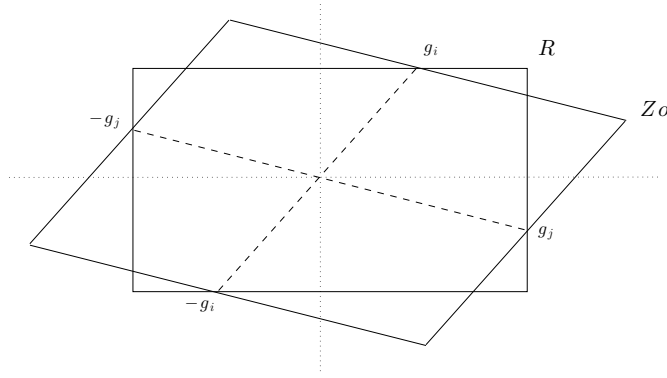


FIGURE 4. The standard Gaussian measure of the zonotope Zo is greater than the standard Gaussian measure of the rectangle R .

Let us consider now the case $i \in A$. Observe that the super-gradient of the concave function $-|Z_i(t)|$ at point t is the segment $[-g^i, g^i]$ and thus the super-gradient of $X_A(t)$ is the zonotope:

$$(7) \quad Zo = \sum_{i \in A} [-g^i, g^i],$$

where the sum means Minkowsky addition.

In conclusion, the probability of the super-gradient to contain zero is equal to $\mathcal{P}(k, \tilde{p}, m)$ where $\mathcal{P}(k, \tilde{p}, m)$ is the probability of the following event:

- draw k standard Gaussian variables g^1, \dots, g^k in \mathbb{R}^m and consider the zonotope Zo given by formula (7),
- draw in the space V generated by g^1, \dots, g^k an independent isotropic normal variable ξ of variance \tilde{p} ,
- define $\mathcal{P}(k, \tilde{p}, m)$ as the probability of ξ to be in Zo .

Lemma 2 — Define the orthonormal basis e^1, \dots, e^k obtained by Gram-Schmidt orthogonalization of the vectors g^1, \dots, g^k . Then:

- (a) $\mathcal{P}(k, \tilde{p}, m)$ is less than the probability $\mathcal{Q}(k, \tilde{p}, m)$ of ξ to be in the hyper-rectangle:

$$R = \sum_{i \in A} [-\langle e^i, g^i \rangle, \langle e^i, g^i \rangle],$$

- (b) this last probability satisfies:

$$(\mathcal{Q}(k, \tilde{p}_k, m))^2 \leq \frac{H_k!}{H_k^{H_k + k - m} (m - k)!},$$

with $\tilde{p}_k := (C^2 - 1)s + p - k$ and $H_k = \lfloor (\frac{\pi}{2} \tilde{p}_k) \wedge m \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part.

Proof. (a) We prove the result conditionally to the g^i 's and by induction on k . When $k = 1$ the result is trivial since the zonotope and the rectangle are simply the same segment.

Let Φ_h be the standard Gaussian distribution on \mathbb{R}^h , $\mathcal{P}(k, \tilde{p}, m)$ is equal to:

$$\Phi_k((\tilde{p})^{-1/2} \cdot Zo) =: \Phi_k(\tilde{Zo}).$$

We can compute this probability using the Fubini Theorem:

$$\mathcal{P}(k, \tilde{p}_k, m) = \int_{-\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}}}}^{\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}}}} \Phi_{k-1}(\widetilde{Z}o_{k-1} + vz) \phi(z) dz = \mathcal{Q}(k, \tilde{p}_k, m),$$

where ϕ is the standard Gaussian density; $\widetilde{Z}o_{k-1}$ is the zonotope generated by g^1, \dots, g^{k-1} and normalized by $(\tilde{p})^{-1/2}$ and v is some vector. By use of the Anderson inequality [6], the non-centered zonotope $(\widetilde{Z}o_{k-1} + vz)$ has a smaller standard Gaussian measure than the centered one.

$$\mathcal{P}(k, \tilde{p}, m) \leq \int_{-\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}}}}^{\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}}}} \Phi_{k-1}(\widetilde{Z}o_{k-1}) \phi(z) dz \leq \int_R \phi(z_1) \dots \phi(z_k) dz_1 \dots dz_k.$$

The last inequality is due to the induction hypothesis. It achieves the proof.

(b) We use the relation above and deconditionning on the g^i . Note the dimension of the edges of the rectangle R are independent with distribution:

$$2\chi(m), 2\chi(m-1), \dots, 2\chi(m-k+1),$$

where the law $\chi(d)$ is defined as the square root of a $\chi^2(d)$. As a consequence, using the independence of the components of ξ in the basis e^1, \dots, e^k and the fact that a Student density T is uniformly bounded by $(2\pi)^{-1/2}$, we get that:

$$\mathcal{Q}(k, \tilde{p}, m) = \prod_{\ell=0}^{k-1} \mathbb{P}\left[|T(m-\ell)| \leq \sqrt{\frac{m-\ell}{\tilde{p}}}\right].$$

Suppose that $\pi\tilde{p}_k \geq 2m$, then a convenient bound is obtained by using the fact that a Student density is uniformly bounded by $(2\pi)^{-1/2}$:

$$(\mathcal{Q}(k, \tilde{p}, m))^2 \leq \left(\frac{2}{\pi\tilde{p}_k}\right)^k \frac{m!}{(m-k)!}.$$

In the other case, set $H_k = \lfloor (\pi\tilde{p}_k)/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part. Removing factors that are greater than 1 in the computation above gives

$$(\mathcal{Q}(k, \tilde{p}, m))^2 \leq \left(\frac{2}{\pi\tilde{p}_k}\right)^{H_k+k-m} \frac{H_k!}{(m-k)!}.$$

and the desired general form is obtained by setting $H_k = m$ in the first case. Observe that $H_k > m - k + 1$ for $k \geq 1$. \square

3.7. Conclusion: gathering the pieces. Finally summing up over the $\binom{p}{k}$ sets of size k , we get that:

$$\frac{1}{2} \sum_{k=0}^{p-n-1} \binom{p}{k} h(s, p-n-k, p-k) \left[\frac{H_k!}{H_k^{H_k-p+n+k} (p-n-k)!} \right]^{\frac{1}{2}}$$

Observe that if $\delta := \frac{n}{p} > \delta_0 := \frac{2}{2+\pi}$ then for all $0 \leq k \leq p-n$, $H_k = p-n$. It follows that the aforementioned equation simplifies to:

$$\frac{1}{2} \sum_{k=0}^{p-n-1} \binom{p}{k} h(s, p-n-k, p-k) \left[\frac{(p-n)!}{(p-n)^k (p-n-k)!} \right]^{\frac{1}{2}}$$

This proves Theorem 2.

3.8. Simplification: an explicit expression of a lower bound on the phase transition. While we can compute the upper bound (1) for given s , n and p and derive a “phase transition” from numerical experiments, we choose to present a more simpler (and weaker) upper bound in Theorem 3. Let $1 \leq k \leq m-2$ and set $\rho := \frac{s}{n}$, $\delta := \frac{n}{p}$. First, using Lemma 5, note that for all $1 \leq u \leq v$, it holds:

$$\binom{v}{u} \leq e^{v\mathbf{H}_e(\frac{u}{v})} \quad \text{and} \quad \frac{\Gamma(u + \frac{1}{2})}{\Gamma(u)} \leq e(u+1)^{\frac{1}{2}}.$$

where $\mathbf{H}_e(x) := -x \log(x) - (1-x) \log(1-x)$ is the Shannon entropy in basis e . Moreover, observe that:

$$\frac{H_k!}{H_k^{H_k+k-m}(m-k)!} \leq 1$$

Hence, Equation (6) gives that:

$$\binom{p}{k} h(s, m-k, p-k) \left[\frac{H_k!}{H_k^{H_k+k-m}(m-k)!} \right]^{\frac{1}{2}} \leq \frac{9.7}{C} \left(\frac{n}{\delta} \right)^{\frac{1}{2}} \exp(n\mathbb{A}(\rho, \delta, C, n, k)),$$

where the exponent \mathbb{A} is given by:

$$\begin{aligned} \mathbb{A}(\rho, \delta, C, n, k) := & \frac{1}{2} \left(\frac{k+1}{n} - \frac{1-\delta}{\delta} \right) \log(C^2 - 1) \\ & + \frac{n-k\delta}{n\rho\delta} + \frac{1}{2\delta} \log \left[-\frac{\rho\delta}{C_0} \log \left(\frac{\rho\delta}{eC_0} \right) \right] \\ & + \left(\frac{1-\delta}{2\delta} - \frac{k}{n} \right) \log C \\ & + \left(\frac{2-\delta}{2\delta} - \frac{k}{n} \right) \mathbf{H}_e \left(\left(2 - \frac{n}{(n/\delta - k - 2)} \right)^{-1} \right) + \frac{1}{\delta} \mathbf{H}_e \left(\delta \frac{k}{n} \right). \end{aligned}$$

where C_0 is defined as in Lemma 4. We make the following observations:

- (1) The map $\zeta_{(\delta, C)} : \rho \mapsto \log \left[-\frac{\rho\delta}{C_0} \log \left(\frac{\rho\delta}{eC_0} \right) \right]$ is an increasing \mathcal{C}^1 -diffeomorphism from $]0, \frac{C_0}{\delta}[$ to $] -\infty, 0[$.
- (2) It holds:

$$\left(\frac{k+1}{n} - \frac{1-\delta}{\delta} \right) \log(C^2 - 1 + \frac{n-k\delta}{n\rho\delta}) \leq -\frac{2}{n} \log(C^2 - 1 + \frac{1}{\rho}).$$

- (3) Observe that:

$$\left(\frac{1-\delta}{\delta} - \frac{k}{n} \right) \log C \leq \frac{1-\delta}{\delta} \log C.$$

- (4) Note that for all $1 \leq k \leq m = \frac{1-\delta}{\delta}n$, the last term:

$$\left(\frac{2-\delta}{2\delta} - \frac{k}{n} \right) \mathbf{H}_e \left(\left(2 - \frac{n}{(n/\delta - k - 2)} \right)^{-1} \right) + \frac{1}{\delta} \mathbf{H}_e \left(\delta \frac{k}{n} \right)$$

is less than:

$$\begin{cases} \left(\frac{4-\delta}{2\delta} \right) \log 2 & \text{if } \delta \in]0, 1/2], \\ \left(\frac{2-\delta}{\delta} \right) \log 2 + \frac{1}{\delta} \mathbf{H}_e(\delta) & \text{if } \delta \in]1/2, 1]. \end{cases}$$

It yields that:

$$\mathbb{A}(\rho, \delta, C, n, k) \leq \zeta_{(\delta, C)}(\rho) + \frac{4-\delta}{2\delta} \log 2 + \frac{1-\delta}{\delta} \log C,$$

for $\delta \in]0, 1/2]$ and:

$$\mathbb{A}(\rho, \delta, C, n, k) \leq \zeta_{(\delta, C)}(\rho) + \frac{2-\delta}{\delta} \log 2 + \frac{1}{\delta} \mathbf{H}_e(\delta) + \frac{1-\delta}{\delta} \log C,$$

for $\delta \in]1/2, 1]$. We see that, if $\rho < \rho_{C,D}(\delta)$ then $\mathbb{A}(\rho, \delta, C, n, k) < -D$, where:

$$\rho_{C,D}(\delta) = \begin{cases} \zeta_{(\delta,C)}^{-1}(-D - \frac{4-\delta}{2\delta} \log 2 - \frac{1-\delta}{\delta} \log C) & \text{if } \delta \in]0, 1/2], \\ \zeta_{(\delta,C)}^{-1}(-D - \frac{2-\delta}{\delta} \log 2 - \frac{1}{\delta} \mathbf{H}_e(\delta) - \frac{1-\delta}{\delta} \log C) & \text{if } \delta \in]1/2, 1]. \end{cases}$$

Moreover, Lemma 4 shows that if $\rho < \rho_{C,D}(\delta)$ then:

$$\binom{p}{n+1} \psi(C) \left[\frac{H_{m-1}!}{H_{m-1}^{H_{m-1}-1}} \right]^{\frac{1}{2}} \leq 2 \exp(-Dn).$$

We conclude using (1).

APPENDIX A. GAUSSIAN CONCENTRATION INEQUALITIES

Lemma 3 — *Let g be a standard Gaussian vector in \mathbb{R}^p and $t_1, t_2 > 0$ such that $t_1^2 \geq 1 + (\pi(p-2))^{-\frac{1}{2}}$ then:*

$$(8) \quad \mathbb{P}\{\|g\|_2 > t_1 \sqrt{p}\} \leq \exp \left[-\frac{p}{2}(t_1^2 - 1 - 2 \log(t_1)) \right],$$

$$(9) \quad \mathbb{P}\{\|g\|_1 < t_2 p\} \leq \exp \left[(p \log(e \sqrt{\frac{2}{\pi}} t_2)) \right].$$

Proof. • Proof of (8): By integrating in polar coordinates:

$$\begin{aligned} \mathbb{P}\{\|g\|_2 > t \sqrt{p}\} &= \int_{t\sqrt{p}}^{+\infty} v^{p-1} \sigma(\mathbb{S}^{p-1}) (2\pi)^{-p/2} e^{-v^2/2} dv, \\ &= 2^{1-\frac{p}{2}} \left(\Gamma\left(\frac{p}{2}\right) \right)^{-1} \int_{t\sqrt{p}}^{+\infty} v^{p-1} e^{-\frac{v^2}{2}} dv, \end{aligned}$$

where t stands for t_1 and $\sigma(\mathbb{S}^{p-1}) = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}$ is the $(p-1)$ -dimensional volume of the unit sphere. Consider the case p even. By integration by parts, if $w^2 > p$,

$$\begin{aligned} \int_w^{+\infty} v^{p-1} e^{-\frac{v^2}{2}} dv &= [w^{p-2} + (p-2)w^{p-4} + \dots + (p-2)(p-4)\dots 2] e^{-\frac{w^2}{2}}, \\ &\leq w^{p-2} \left(1 - \frac{p}{w^2}\right)^{-1} e^{-\frac{w^2}{2}}. \end{aligned}$$

Using this result, the Stirling formula (see Lemma 5 for instance), and the fact that $(1-x)^{1-\frac{1}{x}} < e$ for all $0 < x < 1$ (see See [1] Eq. 4.1.34), we get that:

$$\begin{aligned} \mathbb{P}\{\|g\|_2 > t \sqrt{p}\} &\leq \left(\frac{p}{2}\right)^{\frac{p-2}{2}} \frac{t^p}{t^2-1} \left(\Gamma\left(\frac{p}{2}\right)\right)^{-1} \exp(-t^2 \frac{p}{2}), \\ &\leq \exp \left[-\frac{p}{2}(t^2 - 1 - 2 \log t) \right]. \end{aligned}$$

The proof for the case p odd is very similar using that $\overline{\Phi}(u) \leq \frac{\phi(u)}{u}$.

• Proof of (9): We use the following facts:

- The density of g is bounded by $(2\pi)^{-\frac{p}{2}}$,
- The volume of the L^1 unit ball is $\frac{2^p}{p!}$,

to get that:

$$\mathbb{P}\{\|g\|_1 \leq t_2 p\} \leq \left(\frac{2t_2 p}{\sqrt{2\pi}}\right)^p \frac{1}{p!} \leq \exp \left[(p \log(e \sqrt{\frac{2}{\pi}} t_2)) \right],$$

using the Stirling formula. \square

Lemma 4 — *Let $C > 0$, then, except with a probability smaller than:*

$$\psi(C) := 2 \left(\frac{s \log(eC_0 \frac{p}{s})}{C_0 p} \right)^{\frac{p}{2}} \quad \text{with} \quad C_0 = \frac{\pi}{2e^2(1+C)^2},$$

a standard Gaussian vector $g \in \mathbb{R}^p$ enjoys for all $S \subset \{1, \dots, p\}, |S| \leq s$

$$C \|g_S\|_1 \leq \|g_{S^c}\|_1,$$

which is equivalent to $(C+1)\|g_S\|_1 \leq \|g\|_1$.

Proof. Define:

$$t_1 = (1 + \log(C_0 \frac{p}{s}))^{\frac{1}{2}} \quad \text{and} \quad t_2 = t_1(C+1) \left(\frac{s}{p} \right)^{\frac{1}{2}}.$$

Lemma 3 shows that:

$$\begin{aligned} -\frac{1}{p} \log(\mathbb{P}(\|g\|_2 > t_1 \sqrt{p})) &\geq \frac{1}{2} \log(C_0 p / (s \log(eC_0 \frac{p}{s}))), \\ -\frac{1}{p} \log(\mathbb{P}(\|g\|_1 < t_2 p)) &\geq \frac{1}{2} \log(C_0 p / (s \log(eC_0 \frac{p}{s}))). \end{aligned}$$

Hence, except with a probability smaller than $\psi(C)$, it holds:

$$(C+1)\|g_S\|_1 \leq s^{\frac{1}{2}}(C+1)\|g\|_2 \leq (C+1) \frac{t_1}{t_2} \left[\frac{s}{p} \right]^{\frac{1}{2}} \|g\|_1 = \|g\|_1.$$

as claimed. \square

APPENDIX B. TECHNICAL LEMMAS

B.1. Stirling's formula.

Lemma 5 — *Let $z > 0$ then there exists $\theta \in (0, 1)$ such that:*

$$\Gamma(z+1) = (2\pi z)^{\frac{1}{2}} \left(\frac{z}{e} \right)^z \exp\left(\frac{\theta}{12z} \right).$$

Proof. See [1] Eq. 6.1.38. \square

B.2. Overall expression.

Lemma 6 — *Set $\rho := \frac{s}{n}$, $\delta := \frac{n}{p}$. Then for all $1 \leq k \leq m-2$,*

$$\binom{p}{k} h(s, m-k, p-k) \left[\frac{H_k!}{H_k^{H_k+k-m} (m-k)!} \right]^{1/2} \leq \exp(\mathbb{B}(\rho, \delta, C, n, k)),$$

where:

$$\begin{aligned} \mathbb{B}(\rho, \delta, C, n, k) &:= \frac{\delta}{12n} + \frac{1}{12((\frac{1}{\delta} - \frac{1}{2})n - k - \frac{3}{2})} + \frac{1}{24H_k} + \log\left[\frac{1}{\sqrt{\pi e}} \right] \\ &+ \frac{1}{2} \left(k+1 - \frac{1-\delta}{\delta} n \right) \log(C^2 - 1 + \frac{n-k\delta}{n\rho\delta}) + \left(\frac{1-\delta}{\delta} n - k - 1 \right) \log(C) \\ &+ \left(\frac{n}{\delta} + \frac{1}{2} \right) \log \frac{n}{\delta} + \left[\left(\frac{1}{\delta} - \frac{1}{2} \right) n - k - 1 \right] \log \left(\left(\frac{1}{\delta} - \frac{1}{2} \right) n - k - \frac{3}{2} \right) \\ &- \left(\frac{n}{\delta} - k + \frac{1}{2} \right) \log \left(\frac{n}{\delta} - k \right) - \left(k + \frac{1}{2} \right) \log k \\ &- \left(\frac{n}{2\delta} - \frac{k+1}{2} \right) \log \left(\frac{n}{2\delta} - \frac{k+2}{2} \right) \\ &- \left(\frac{1-\delta}{2\delta} n - \frac{k+1}{2} \right) \log \left(\frac{1-\delta}{2\delta} n - \frac{k+2}{2} \right) \\ &+ \frac{1}{2} \left(\frac{1-\delta}{\delta} n - k + \frac{1}{2} \right) \log(H_k) - \frac{1}{2} \left(\frac{1-\delta}{\delta} n - k \right) \log \left(\frac{1-\delta}{\delta} n - k \right) \\ &+ \frac{1}{2} \left(\frac{1-\delta}{\delta} n - k - H_k \right) + \log(\psi(C)), \end{aligned}$$

with $H_k = \lfloor (\frac{1-\delta}{\delta})n \wedge [\frac{\pi}{2}((C^2-1)\rho n + \frac{n}{\delta} - k)] \rfloor$.

Proof. Set $P = \binom{p}{k} h(s, m-k, p-k)$. Using (6) it holds:

$$P \leq 2\pi^{\frac{1}{2}} \left[C^2 - 1 + \frac{n-k\delta}{n\delta\rho} \right]^{\frac{k}{2} + \frac{1}{2} - \frac{1-\delta}{2\delta}n} C^{\frac{1-\delta}{\delta}n-k-1} \psi(C) R,$$

where:

$$R = \frac{\Gamma(\frac{n}{\delta} + 1) \Gamma((\frac{1}{\delta} - \frac{1}{2})n - (k + \frac{1}{2}))}{\Gamma(\frac{n}{\delta} - k + 1) \Gamma(k + 1) \Gamma(\frac{n}{2\delta} - \frac{k}{2}) \Gamma(\frac{1-\delta}{2\delta}n - \frac{k}{2})}.$$

Using Lemma 5, this last expression can be written as:

$$\begin{aligned} \log R \leq & -\log(2\pi) - \frac{1}{2} + \frac{\delta}{12n} + \frac{1}{12((\frac{1}{\delta} - \frac{1}{2})n - k - \frac{3}{2})} + (\frac{n}{\delta} + \frac{1}{2}) \log \frac{n}{\delta} \\ & + [(\frac{1}{\delta} - \frac{1}{2})n - k - 1] \log((\frac{1}{\delta} - \frac{1}{2})n - k - \frac{3}{2}) - (\frac{n}{\delta} - k + \frac{1}{2}) \log(\frac{n}{\delta} - k) \\ & - (k + \frac{1}{2}) \log k - (\frac{n}{2\delta} - \frac{k+1}{2}) \log(\frac{n}{2\delta} - \frac{k+2}{2}) \\ (10) \quad & - (\frac{1-\delta}{2\delta}n - \frac{k+1}{2}) \log(\frac{1-\delta}{2\delta}n - \frac{k+2}{2}). \end{aligned}$$

Set:

$$T = \left[\frac{H_k!}{H_k^{H_k+k-m}(m-k)!} \right]^{\frac{1}{2}} \leq \left[\frac{H_k^{m-k}}{(m-k)!} e^{-H_k} \sqrt{2\pi H_k} e^{\frac{1}{12H_k}} \right]^{\frac{1}{2}}.$$

Invoking Lemma 5, it follows that:

$$\begin{aligned} 2 \log T \leq & \frac{1}{12H_k} + (\frac{1-\delta}{\delta}n - k + \frac{1}{2}) \log(H_k) - (\frac{1-\delta}{\delta}n - k) \log(\frac{1-\delta}{\delta}n - k) \\ (11) \quad & + \frac{1-\delta}{\delta}n - k - H_k. \end{aligned}$$

From (10) and (11), we deduce that $\log(P \times T) \leq \mathbb{B}(\rho, \delta, n, k)$. \square

Remark. In the case $k = 0$, note that:

$$p \times h(s, m, p) \leq \exp \left[\log(\psi(C)) - \frac{1}{3} + \log p + \frac{m}{2} \log(1 + \frac{p}{m}) + \frac{p}{2} \log(1 + \frac{m}{p}) \right].$$

Remark. For $C = 1$, using that $\binom{p}{k} \leq (\frac{ep}{k})^k / (2\pi k)^{\frac{1}{2}}$, it holds for $1 \leq k \leq m-1$:

$$\begin{aligned} \exp(\mathbb{B}(\rho, \delta, C, n, k)) & \leq \psi(1) \frac{2\sqrt{\pi}}{\sqrt{2\pi}} \frac{e^{-5/12}}{\sqrt{2\pi}} \left(\frac{p}{k} \right)^k \frac{e^k}{\sqrt{k}} e^{\frac{m-k}{2}} \\ & \times \left(1 + \frac{p-1-k}{m-2-k} \right)^{\frac{m-1-k}{2}} \left(\frac{p}{s} \right)^{\frac{1-m+k}{2}}, \\ & \leq \psi(1) \frac{e^{-\frac{5}{12}}}{\sqrt{\pi}} \left(\frac{p}{s} \right)^{\frac{1-m}{2}} e^{\frac{m}{2}} \left(\frac{pe}{k} \right)^k e^{\frac{k}{2}} \left(\frac{p}{s} \right)^{\frac{k}{2}} \\ & \times \left(1 + \frac{p-1-k}{m-2-k} \right)^{\frac{m-1-k}{2}}, \\ & \leq \psi(1) \frac{e^{-\frac{5}{12}}}{\sqrt{\pi}} s^{\frac{m-1}{2}} e^{\frac{m}{2}} \left(\frac{pe^{\frac{3}{2}}}{k\sqrt{s}} \right)^k, \\ & \leq \psi(1) e^{-\frac{1}{6}} (se)^{\frac{m}{2}} \left(\frac{pe^{\frac{3}{2}}}{k\sqrt{s}} \right)^k, \\ & \leq \left(\frac{C_0}{2} \right) \rho \delta \log(eC_0 \rho \delta) (\rho ne)^{\frac{1-\delta}{2\delta}} \left(\frac{\sqrt{ne}^{\frac{3}{2}}}{Ck\delta\sqrt{\rho}} \right)^k, \end{aligned}$$

where C_0 is defined as in Lemma 4.

REFERENCES

- [1] M. Abramowitz and I. Stegun. Handbook of mathematical functions. *National Bureau of Standards, Washington DC*, 1965.
- [2] R. J. Adler. *The geometry of random fields*. Siam, 1981.
- [3] R. J. Adler and J. E. Taylor. *Random fields and geometry*, volume 115. Springer, 2009.
- [4] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete and Computational Geometry*, 7(1):219–226, 1992.
- [5] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: A geometric theory of phase transitions in convex optimization. *arXiv preprint arXiv:1303.6672*, 2013.
- [6] T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955.
- [7] A. Auffinger and G. B. Arous. Complexity of random smooth functions on the high-dimensional sphere. *The Annals of Probability*, 41(6):4214–4247, 2013.
- [8] J.-M. Azaïs and C. Delmas. Asymptotic expansions for the distribution of the maximum of gaussian random fields. *Extremes*, 5(2):181–212, 2002.
- [9] J.-M. Azaïs and M. Wschebor. On the distribution of the maximum of a gaussian field with d parameters. *The Annals of Applied Probability*, 15(1A):254–278, 2005.
- [10] J.-M. Azaïs and M. Wschebor. A general expression for the distribution of the maximum of a gaussian field and the approximation of the tail. *Stochastic Processes and their Applications*, 118(7):1190 – 1218, 2008.
- [11] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields*. John Wiley and Sons, 2009.
- [12] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705–1732, 2009.
- [13] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [14] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [15] E. J. Candes and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n . *Ann. Statist.*, 35(6):2313–2351, 2007.
- [16] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interaction between Compressed Sensing, Random matrices and High dimensional geometry*. Number 37 in Panoramas et synthèses. SMF, 2012.
- [17] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
- [18] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.
- [19] F. Cucker and M. Wschebor. On the expected condition number of linear programming problems. *Numerische Mathematik*, 94(3):419–478, 2003.
- [20] Y. de Castro. A remark on the lasso and the dantzig selector. *Statistics and Probability Letters*, 2012.
- [21] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [22] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete and Computational Geometry*, 35(4):617–652, 2006.
- [23] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9452–9457, 2005.
- [24] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.
- [25] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–4293, 2009.
- [26] A. Juditsky and A. Nemirovski. Accuracy guarantees for-recovery. *Information Theory, IEEE Transactions on*, 57(12):7818–7839, 2011.

- [27] G. Lecué and S. Mendelson. Compressed sensing under weak moment assumptions. *arXiv preprint arXiv:1401.2188*, 2014.
- [28] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex deconvolution, with applications. *arXiv preprint arXiv:1205.1580*, 2012.
- [29] V. Piterbarg. Comparison of distribution functions of maxima of gaussian processes. *Teoriya Veroyatnostei i ee Primeneniya*, 26(4):702–719, 1981.
- [30] M. Stojnic. Various thresholds for ℓ_1 -optimization in compressed sensing. *arXiv preprint arXiv:0907.3666*, 2009.
- [31] M. Stojnic. A rigorous geometry-probability equivalence in characterization of ℓ_1 -optimization. *arXiv preprint arXiv:1303.7287*, 2013.
- [32] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.
- [33] S. A. van de Geer and P. Bühlmann. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.
- [34] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Math. Soviet*, 11(2):181–201, 1992.
- [35] W. Xu and B. Hassibi. Compressed sensing over the grassmann manifold: A unified analytical framework. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 562–567. IEEE, 2008.
- [36] W. Xu and B. Hassibi. Precise stability phase transitions for minimization: A unified geometric framework. *Information Theory, IEEE Transactions on*, 57(10):6894–6919, 2011.

JMA IS WITH THE INSTITUT DE MATHÉMATIQUES DE TOULOUSE (CNRS UMR 5219). UNIVERSITÉ PAUL SABATIER, 118 ROUTE DE NARBONNE, 31062 TOULOUSE, FRANCE.

E-mail address: `jean-marc.azais@math.univ-toulouse.fr`

YdC IS WITH LABORATOIRE DE MATHÉMATIQUES D’ORSAY (CNRS UMR 8628), UNIVERSITÉ PARIS-SUD, FACULTÉ DES SCIENCES D’ORSAY, 91405 ORSAY, FRANCE.

E-mail address: `yohann.decastro@math.u-psud.fr`

SM IS WITH THE INSTITUT DE MATHÉMATIQUES DE TOULOUSE (CNRS UMR 5219). UNIVERSITÉ PAUL SABATIER, 118 ROUTE DE NARBONNE, 31062 TOULOUSE, FRANCE.

E-mail address: `stephane.mourareau@math.univ-toulouse.fr`