

A rice method proof of the null-space property over the Grassmannian

Jean-Marc Azaïs^a, Stéphane Mourareau^a and Yohann De Castro^b

^a*Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.
E-mail: jean-marc.azais@math.univ-toulouse.fr; stephane.mourareau@math.univ-toulouse.fr*

^b*Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.
E-mail: yohann.decastro@math.u-psud.fr*

Received 30 September 2015; revised 19 February 2016; accepted 5 June 2016

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 derive a closed form for the phase transition on NSP. 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 unitarily invariant distributions. Furthermore, we derive a simple sufficient condition for NSP.

Résumé. La propriété du noyau (NSP en anglais) est une condition nécessaire et suffisante pour estimer les plus grands coefficients d'un système linéaire sous-déterminé d'équations. De manière intéressante, cette propriété gouverne aussi le succès ou l'échec de récentes approches en statistique en grandes dimensions, traitement du signal, codes correcteurs d'erreurs et la théorie des polytopes.

Bien que cette propriété soit au centre des techniques de minimisation L1, un problème ouvert reste l'obtention d'une forme explicite de la transition de phase de la propriété NSP. Dans cet article, nous donnons la première preuve de la propriété NSP du point de vue de la théorie des processus aléatoires et de la méthode de Rice. Ainsi, notre analyse conduit à de nouvelles bornes non asymptotiques pour la propriété NSP pour toute distribution invariante par rotation. De plus, nous déduisons une condition suffisante simple pour établir la propriété NSP.

MSC: 62J05; 62H12; 62H20

Keywords: Rice Method; High-dimensional statistics; ℓ_1 -minimization; Null-Space Property; Random processes theory

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 elementary. 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 [14]:

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

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. [11]. 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 [10,13,15,19] 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 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 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 [15] that the operator Δ_X is the identity on Σ_s if and only if the kernel of X satisfies NSP(s, C) for some $C > 1$.

Theorem 1 ([15]). *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 NSP(s, C) for some $C > 1$. Moreover, if $\ker(X)$ enjoys 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 [33] or Dantzig selector [12]. Indeed, in the frame of noisy observations, ℓ_1 -minimization procedures are based on sufficient conditions like Restricted Isometry Property (RIP) [12], Restricted Eigenvalue Condition (REC) [9], Compatibility Condition (CC) [34], Universal Distortion Property (UDP) [18], or $H_{s,1}$ condition [25]. 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 [13] for instance, a direct proof of NSP (without the use of RIP) remains a challenging issue.

1.2. Contribution

Given $(\rho, \delta) \in]0, 1[^2$, set $s_n = \lfloor \rho n \rfloor$ and $p_n = \lfloor \frac{n}{\delta} \rfloor$ where $\lfloor \cdot \rfloor$ denotes the integer part. Consider $G(n, p_n)$ be uniformly distributed on the Grassmannian $\text{Gr}_m(\mathbb{R}^{p_n})$ where $m = p_n - n$ or equivalently $G(n, p_n) = \ker(X(n, p_n))$ where $X(n, p_n) \in \mathbb{R}^{n \times p_n}$ denotes a matrix with i.i.d. centered Gaussian entries. In this paper, we describe a region of parameters (ρ, δ) such that the probability $\mathbb{P}[G(n, p_n) \text{ enjoys NSP}(s_n, C)]$ tends to one as n goes to infinity. Our result provides a new and simple description of such region of parameters (ρ, δ) .

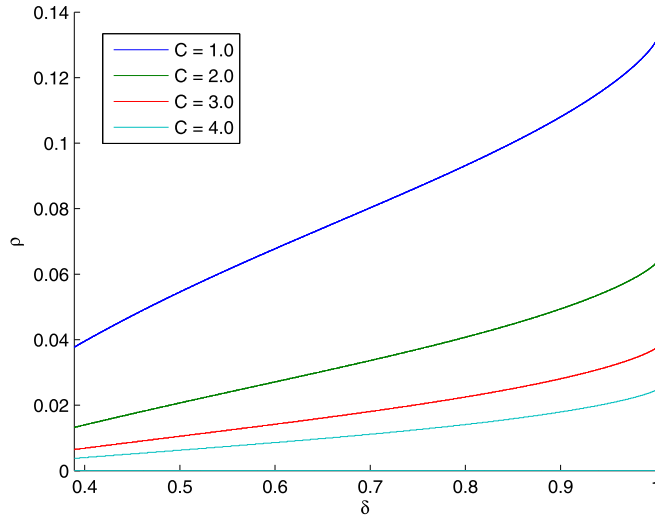


Fig. 1. The phase transition of Theorem 2. From top to bottom, $C = 1, 2, 3, 4$.

Theorem 2. Let $C \geq 1$. For all $n \geq 1$, set $s_n = \lfloor \rho n \rfloor$ and $p_n = \lfloor \frac{n}{\delta} \rfloor$. Let $G(n, p_n)$ be uniformly distributed on the Grassmannian $\text{Gr}_m(\mathbb{R}^{p_n})$ where $m = p_n - n$. If $\delta \geq (1 + \pi/2)^{-1} \simeq 0.389$ and:

$$\begin{aligned} &\rho \log \left(\sqrt{\frac{\pi}{2eC^2}} \frac{(1-\rho)^2}{\rho^2} \right) + \log \left(C e \frac{\sqrt{\rho(1-\delta)(1+(C^2-1)\rho)}}{(1-\rho)(1+(2C^2-1)\rho)\sqrt{\delta}} \right) \\ &+ \frac{1}{\delta} \log \left(\sqrt{\frac{2}{e\pi}} \frac{1+(2C^2-1)\rho}{(1-\rho)\sqrt{\delta(1-\delta)(1+(C^2-1)\rho)}} \right) \leq 0 \end{aligned}$$

then $\mathbb{P}[G(n, p_n) \text{ enjoys NSP}(s_n, C)]$ tends exponentially to one as n goes to infinity.

In Figure 1, we display the lower bound on NSP given Theorem 2 for various choices of constant $C \geq 1$. We emphasize that the condition $\{\delta \geq (1 + \pi/2)^{-1}\}$ is due to technical considerations depicting an union bound on $\mathcal{O}(e^{1/\delta})$ spheres in our proof. Note that Theorem 4 gives a lower bound for any values of δ . Interestingly, we can compare our result with the strong phase transition on NSP of Donoho and Tanner [21]. Although both transitions are defined by implicit equations (which makes it difficult to give a comparison), they can be faithfully computed by standard solvers with prescribed precision. In Figure 2, we present this numerical comparison which unveils the following behaviors:

- As $\delta \rightarrow 0$ (depicting the $n \ll p$ framework), Donoho and Tanner [21] uncover a $(s \log(p/(\sqrt{\pi}n)))$ -bound (see page 305 in [24] for instance). In this regime, our bound compares unfavorably and we cannot obtain the standard $s \log(p/s)$ -bound of Compressed Sensing with our argument.
- In the regime $n \sim p$, our results compare favorably with the strong phase transition up to a multiplicative constant less than 1.8, see Figure 3 and the example below.

Example 1. In the case $C = 1$, we can compare our result (see Theorem 2) given by the set of $(\rho, \delta) \in]0, 1]^2$ such that $\delta \geq (1 + \pi/2)^{-1}$ and such that

$$\rho \log \left[\sqrt{\frac{\pi}{2e}} \frac{(1-\rho)^2}{\rho^2} \right] + \log \left[e \frac{\sqrt{\rho(1-\delta)}}{(1-\rho)(1+\rho)\sqrt{\delta}} \right] + \frac{1}{\delta} \log \left[\sqrt{\frac{2}{e\pi}} \frac{1+\rho}{(1-\rho)\sqrt{\delta(1-\delta)}} \right]$$

is non-positive to the work of Donoho and Tanner [21] (see Figure 3). Observe that, up to a constant bounded by 1.8, we recover the strong phase transition on NSP.

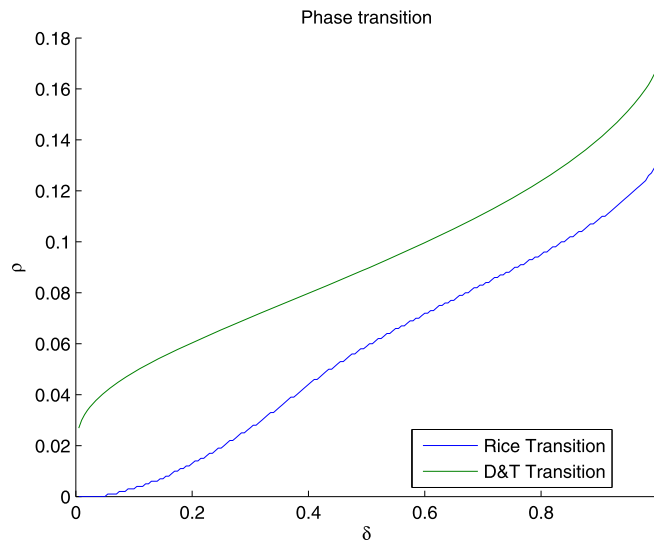


Fig. 2. The panel illustrates numerically the border of the region described by Theorem 4 (blue line) for which NSP holds ($\Pi \simeq 0$) and the strong phase transition of Donoho and Tanner (green line). Note that the region (ρ, δ) such that $\Pi \simeq 0$, i.e. NSP($s, 1$) holds, is located below the curve. Simulations have been performed with $n = 200,000$.

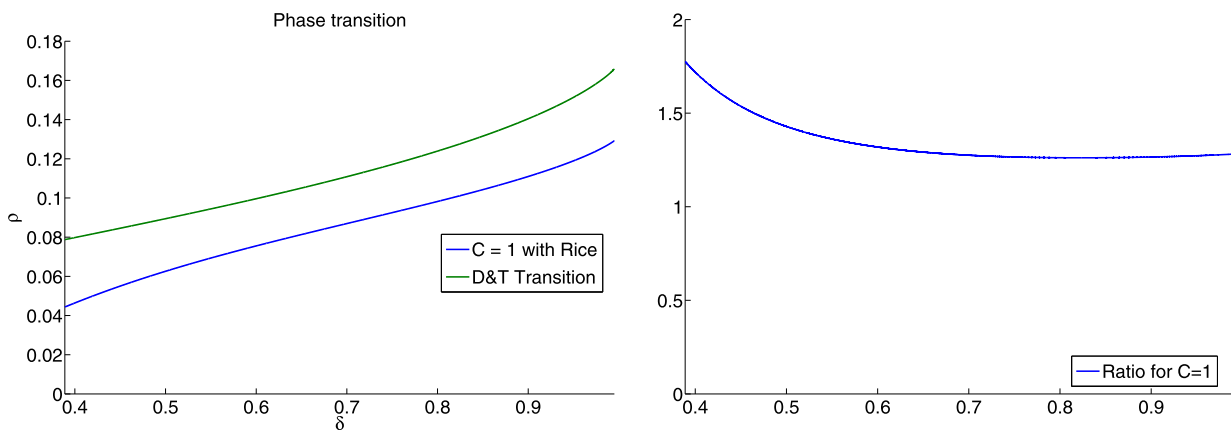


Fig. 3. On the left, comparison between the border of the region described by Theorem 2 (blue line) and the strong phase transition of Donoho and Tanner (green line) for $\delta \geq 0.39$. On the right, ratio between the green and the blue line.

We outline that explicit expressions of lower bounds on the phase transition can be found in Section 2.

1.3. 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. This section is devoted to a short review of some state-of-the-art results on direct proofs of NSP.

1.3.1. Grassmann angles

In a captivating series of papers [20–23], 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 $(\rho, \delta) \in]0, 1[^2$,

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

where we recall that $s_n = \lfloor \rho n \rfloor$ and $p_n = \lfloor \frac{n}{\delta} \rfloor$. 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 [3] and Vershik and Sporyshev [35]. 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 $n \geq C_S \log(\frac{p}{s})$ for n large enough.

1.3.2. Gaussian widths

In recent works [31,32], Stojnic has 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 [31], using “Gordon’s escape through a mesh” theorem, Stojnic has 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.

Predating Stojnic’s work, Rudelson and Vershynin (Theorem 4.1 in [30]) were the first to use “Gordon’s escape through the mesh” theorem to derive a non-asymptotic bound on sparse recovery. A similar result can found in the astonishing book of Foucart and Rauhut, see Theorem 9.29 in [24]. Observe that these results hold with a probability at least $1 - \alpha$ and their bounds depend on $\log(\alpha)$ so one needs one more step to derive a lower bound on the strong phase transition. We did not pursue in this direction.

1.3.3. Approximate kinematic formula

In the papers [4,27], 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.

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

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 $C = 1$.

1.3.4. Empirical process theory

Using empirical process theory, Lecué and Mendelson [26] gives a direct proof of NSP for matrices X with sub-exponential rows. 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, when:

$$n \geq c_0 s \log\left(\frac{ep}{s}\right), \tag{1}$$

with $c_0 > 0$ a universal (unknown) constant.

1.3.5. A previous direct proof of NSP with dilatation $C \geq 1$

Using integral convex geometry theory as in Donoho and Tanner’s works [20–23], Xu and Hassibi have investigated [36,37] the property $\text{NSP}(s, C)$ for values $C \geq 1$. 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.

1.4. Simple bounds on the phase transition

As mentioned in Proposition 2.2.17 of [13], if NSP holds then

$$n \geq c_1 s \log\left(\frac{c_2 p}{s}\right), \tag{2}$$

with $c_1, c_2 > 0$ are universal (unknown) constants. The result of Section 1.3.4 shows that a similar bound is also sufficient to get NSP. What can be understood is that the true phase transition (as presented in [20–23]) lies between the two bounds described by (1) (lower bound) and (2) (upper bound). Observe that these bounds can be equivalently expressed in terms of $\rho = s/n$ and $\delta = n/p$. Indeed, one has:

$$\left\{ n \geq c_1 s \log\left(\frac{c_2 p}{s}\right) \right\} \Leftrightarrow \{ A_\star \rho \delta \log(A_\star \rho \delta) \geq -B_\star \delta \}, \tag{3}$$

where $A_\star = c_2^{-1} > 0$ and $1/e \geq B_\star = c_1^{-1} c_2^{-1} > 0$. Denote by W_0 (resp. W_{-1}) the first (resp. the second) Lambert W function, see [16] for a definition. We deduce that (3) is equivalent to:

$$\rho \leq \frac{\exp(W_{-1}(-B_\star \delta))}{A_\star \delta} \quad \text{or} \quad \rho \geq \frac{\exp(W_0(-B_\star \delta))}{A_\star \delta}. \tag{4}$$

Furthermore, the papers [20–23] show that NSP enjoys a phase transition that can be described as a region $\rho \leq \rho_S(\delta)$, see Section 1.3. In particular, one can check that the region described by the right hand term of (4) cannot be a region of solutions of the phase transition problem. We deduce from [13,26] that ρ_S , the phase transition of Donoho and Tanner [20–23], can be bounded by the left hand term of (4). Hence, it holds the following result.

Theorem 3. *The strong threshold ρ_S (phase transition of NSP) of Donoho and Tanner [20–23] is bounded by:*

$$\forall \delta \in]0, 1[, \quad \frac{\exp(W_{-1}(-B_1 \delta))}{A_1 \delta} \leq \rho_S(\delta) \leq \frac{\exp(W_{-1}(-B_2 \delta))}{A_2 \delta}, \tag{5}$$

where $A_1, A_2 > 0$ and $1/e \geq B_1, B_2 > 0$ are universal (unknown) constants.

Although bounds (1) (lower bound) and (2) (upper bound) are known, their expressions as exponential of second Lambert W functions remain overlooked in the literature. As a matter of fact, Figure 4 depicts a comparison between ρ_S and:

$$\delta \mapsto \frac{\exp(W_{-1}(-0.3394\delta))}{1.38\delta}, \tag{6}$$

where the strong threshold curve has been taken from [20–23]. Roughly speaking, the curve (6) shows empirically that NSP holds when:

$$n \geq 4s \log(0.7p/s),$$

for large values of s, n, p . Recall that it is still an open problem to find a closed form for the weak and the strong thresholds. In the regime $\delta \rightarrow 0$, Donoho and Tanner [20–23] have proved that the phase transition enjoys

$$n \geq 2es \log(p/(\sqrt{\pi}s)) \simeq 5.4s \log(0.6p/s),$$

in the asymptotic.

1.5. Outline of the paper

The main theorem (Theorem 4) is stated in the next section and Section 3 proves it. Section 4 is devoted to the proof of Theorem 2. All the numerical experiments can be reproduced using the codes available at [28].

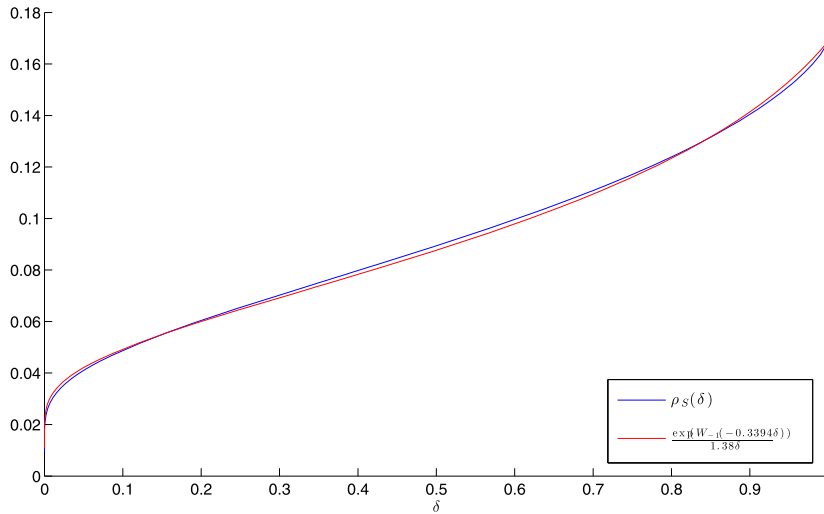


Fig. 4. The strong threshold ρ_S and the mapping $\delta \mapsto \frac{\exp(W_{-1}(-0.3394\delta))}{1.38\delta}$.

2. Rice method bound for NSP with dilatation $C \geq 1$

In this paper, we prove NSP following a new path based on stochastic processes theory and more precisely on the Rice method [7,8]. This latter is specially design to study the tail of the maximum of differentiable random processes or random fields. Similarly to the case of a deterministic function, it consists of studying the maximum through the zeros of the derivative. 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.

However, the situation here is more involved than in the aforementioned papers since the considered process $X(t)$ is defined on the sphere (as in the recent work [6] for example), non Gaussian and, last but not least, non differentiable. Note that the paper [17] studies the maximum of locally linear process by a smoothing argument. A contrario to this paper, we will use a partition of the sphere and directly the Rice method. This provides a short and direct proof of NSP(s, C) for any value $C \geq 1$. Our main result reads as follows.

Theorem 4 (Explicit lower bound). *Let $0 < s < n < p$ and $m = p - n$. Let $G(n, p)$ be the Kernel of $X(n, p)$, a $(n \times p)$ random matrix with i.i.d. centered Gaussian entries, then for all $C \geq 1$, it holds:*

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

with Π satisfying

$$\Pi \leq \sqrt{\pi} \left[\sum_{k=0}^{p-n-1} \binom{p}{k} \left(\frac{C^2 s}{\tilde{p}_{C,k}} \right)^{\frac{p-n-1-k}{2}} \frac{\Gamma(\frac{2p-2k-n-1}{2})}{\Gamma(\frac{p-k}{2})\Gamma(\frac{p-n-k}{2})} \psi_{p-k}(C) \mathbf{Q}(k, \tilde{p}_{C,k}, m) \right], \tag{7}$$

where Γ denotes the Gamma function, $\psi_{p-k}(C)$ is defined by Lemma 4, $\mathbf{Q}(k, \tilde{p}_{C,k}, m)$ is defined by Lemma 2 and $\tilde{p}_{C,k} := (C^2 - 1)s + p - k$.

From this bound we can deduce a lower bound on the phase transition depending on ρ and δ as presented in Theorem 2. Recall that the case $C = 1$ has been compared to the strong phase transition below Theorem 2. We have seen that the bound of Theorem 4 is informative and almost sharp in the regime $n \sim p$. However, investigating the last term of the sum (7) (i.e. when $k = p - n - 1$) as $\delta \rightarrow 0$, one can check that the bound (7) is at least of the order of $\mathcal{O}((1/\delta)^n)$, which is not informative. The bound of Theorem 4 is illustrated in Figure 5 for various $C \geq 1$.

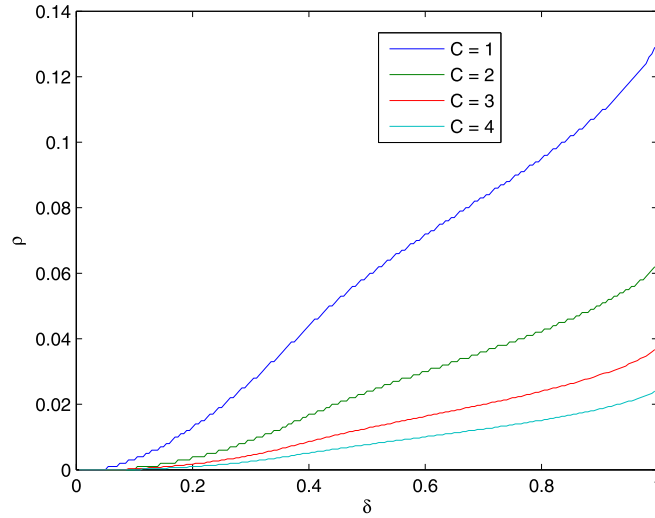


Fig. 5. The region $\Pi \simeq 0$ (NSP holds) for $C = 1, 2, 3, 4$ (from bottom to top) based on Theorem 4. Simulations have been performed with $n = 200,000$.

3. Proof of Theorem 4

3.1. Model and notation

Let $0 < s < n < p$, let $C > 1$ and set $m = p - n$. Let $G(n, p)$ be uniformly distributed on the Grassmannian $\text{Gr}_m(\mathbb{R}^p)$. Observe that it can be generated by m independent 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 $G(n, p)$ 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 with standard distribution in \mathbb{R}^m . Let O_p and O_m two orthogonal matrices of size, respectively, $(p \times p)$ and $(m \times m)$. Thanks to unitarily invariance of the Gaussian distribution, remark that:

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

Consider now the ordered statistics of the absolute values of the coordinates of $Z(t)$:

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

where the ordering $((1), \dots, (p))$ is always uniquely defined if we adopt the convention of keeping the natural order in case of ties. Given a sparsity s , a degree of freedom m , and a degree of constraint p , consider the real valued process $\{X(t); t \in \mathbb{S}^{m-1}\}$ such that:

$$X(t) = C|Z_{(1)}(t)| + \dots + C|Z_{(s)}(t)| - [|Z_{(s+1)}(t)| + \dots + |Z_{(p)}(t)|]. \tag{8}$$

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

3.2. Cutting the sphere out

As we will see later, the process $X(\cdot)$ is locally linear over some subsets and to take benefit of that, we need to consider a particular partition of the sphere.

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:

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

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

where we denote $Z_A(t) = (Z_j(t))_{j \in A}$. One can check that \mathbb{S}_A is the unit sphere of the orthogonal of $V_A := \text{Span}\{g^j; j \in A\}$. 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 \mathcal{W} by:

$$\mathcal{W} := \{t \in \mathbb{S}^{m-1}; |Z_{(s)}(t)| = |Z_{(s+1)}(t)|\}.$$

Observe that, conditionally to g^j , the set \mathcal{W} is closed with empty interior.

3.3. Probability of failure

We consider the probability:

$$\Pi = \mathbb{P}\{\mathcal{M} > 0\} \leq \sum_{|A| \leq m-1} \mathbb{P}\{\mathcal{M}_{\dot{\mathbb{S}}_A} > 0\}, \tag{9}$$

where \mathcal{M} and $\mathcal{M}_{\dot{\mathbb{S}}_A}$ are respectively the number of positive local maximum of $X(\cdot)$ along \mathbb{S}^{m-1} and $\dot{\mathbb{S}}_A$. The baseline of our proof is to upper-bound each right hand side probabilities, using the expected number of positive local maximum above zero and Markov inequality. The first element is Lemma 4 proving that:

$$\forall t \in \dot{\mathbb{S}}_A, \quad \mathbb{P}\{X(t) > 0\} \leq \psi_{p-k}(C),$$

where $k := |A|$ and:

$$\psi_{p-k}(C) = \binom{p-k}{s} \left(\frac{C^2 4s}{\pi}\right)^{\frac{p-k-s}{2}} \frac{\Gamma((p-k)/2)}{\Gamma(s/2)\Gamma(p-k-s+1)},$$

where Γ denotes the Gamma function. The second element is that $X(t)$ admits a density $p_{X(t)}$. To check that, note that $|Z_{(1)}(t)|, \dots, |Z_{(p)}(t)|$ 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 (8) 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 \mathbb{S}_\emptyset and does not belong to \mathcal{W} , $X(\cdot)$ 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:

$$X(u) = C\varepsilon_1 Z_{(1)}(u) + \dots + C\varepsilon_s Z_{(s)}(u) + \varepsilon_{s+1} Z_{(s+1)}(u) + \dots + \varepsilon_p Z_{(p)}(u),$$

where $\varepsilon_1, \dots, \varepsilon_p$ are random variables taking values ± 1 .

Lemma 1. *Let $t \in \mathbb{S}^{m-1}$ then, almost surely, it holds $t \in \mathbb{S}_\emptyset$ and $t \notin \mathcal{W}$. Furthermore, the spherical gradient $X'(t)$ and the spherical Hessian $X''(t)$ of $X(\cdot)$ along \mathbb{S}^{m-1} at t exist and:*

- $X''(t) = -X(t)I_{m-1}$.
- $(X(t), X''(t))$ and $X'(t)$ are independent.
- $X'(t)$ has a Gaussian centered isotropic distribution onto t^\perp with variance $\tilde{p}_{C,0} = (sC^2 + (p - s))$.

Proof. The fact that, with probability 1, $t \in \mathbb{S}_\emptyset$ and $t \notin \mathcal{W}$ implies that the process $X(\cdot)$ is locally given by

$$X(u) = C\varepsilon_1 Z_{(1)}(u) + \dots + C\varepsilon_s Z_{(s)}(u) + \varepsilon_{s+1} Z_{(s+1)}(u) + \dots + \varepsilon_p Z_{(p)}(u),$$

where the signs $(\varepsilon_1, \dots, \varepsilon_p)$ and the ordering $(1), \dots, (p)$ are those of t . The process $X(\cdot)$ is locally linear and thus differentiable around t and its gradient in \mathbb{R}^m at t , denoted $\dot{X}(t)$, is given by

$$C\varepsilon_1 g^{(1)} + \dots + C\varepsilon_s g^{(s)} + \varepsilon_{s+1} g^{(s+1)} + \dots + \varepsilon_p g^{(p)}.$$

Moreover, note that its Hessian on \mathbb{R}^m vanishes.

Let us consider now the spherical gradient $X'(t)$ and the spherical Hessian $X''(t)$. It is well known that $X'(t) = P_{t^\perp} \dot{X}(t)$, where P_{t^\perp} is the orthogonal projection onto the orthogonal of t . As for the spherical Hessian, it is defined on the tangent space t^\perp and is equal to the projection of the Hessian in \mathbb{R}^m , which vanishes, minus the product of the normal derivative by the identity matrix. This is detailed in Lemma 5. In the case of the unit sphere, the vector normal to the sphere at t is t itself and

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

In the case of $X'(t)$, remark that $Z(t)$ and thus $X(t)$, $(\varepsilon_1, \dots, \varepsilon_p, (1), \dots, (p))$ are functions of $(P_t(g^1), \dots, P_t(g^p)) = (Z_1(t)t, \dots, Z_p(t)t)$ (with obvious notation). They are therefore independent of $X'(t)$ which is a function of $(P_{t^\perp}(g^1), \dots, P_{t^\perp}(g^p))$. Conditionally to $(\varepsilon_1, \dots, \varepsilon_p, (1), \dots, (p))$, $X'(t)$ can be written as

$$X'(t) = (C\varepsilon_1 P_{t^\perp} g^{(1)} + \dots + C\varepsilon_s P_{t^\perp} g^{(s)} + \varepsilon_{s+1} P_{t^\perp} g^{(s+1)} + \dots + \varepsilon_p P_{t^\perp} g^{(p)}),$$

which implies that the conditional distribution of $X'(t)$ is Gaussian with variance-covariance matrix $(sC^2 + (p - s))\text{Id}_{t^\perp}$, where Id_{t^\perp} is the identity operator on t^\perp . Since $X'(t)$ is independent of $(\varepsilon_1, \dots, \varepsilon_p, (1), \dots, (p))$ this conditional distribution is in fact equal to the unconditional distribution. □

The next step is to prove that a.s. there is no local maximum on \mathcal{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 the order statistics and the ordering remain uniquely defined because of our convention.

Suppose that $t \in \mathcal{W}$. Since all the possible ordering $((1), \dots, (p))$ and signs $(\varepsilon_1, \dots, \varepsilon_p)$ play the same role by unitarily invariance of the distribution of $Z(t)$ for all t , we make the proof in the particular case where $((1), \dots, (p))$ is the identity and all the signs $(\varepsilon_1, \dots, \varepsilon_p)$ are positive:

$$Z_1(t) \geq \dots \geq Z_{s-h-1}(t) > Z_{s-h}(t) = \dots = Z_{s+k}(t) > Z_{s+k+1}(t) \geq \dots \geq Z_p(t) > 0.$$

Then, for w in some neighborhood N of t (not included in \mathcal{W}), we have:

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

where Max_h is the sum of the h largest element of its $(h + k + 1)$ arguments. 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 chosen 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 t^\perp 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 maximum.

Suppose that $t \notin W$, and suppose that we limit our attention to points t such that $X(t) > 0$, then Lemma 1 implies that $X''(t)$ 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 zeros of the derivative $X'(t)$ are isolated and thus in finite number. In addition recalling that $\mathcal{M}_{\dot{S}_\emptyset}$ is the number of positive local maximum of $X(\cdot)$ and belonging to \dot{S}_\emptyset , this number satisfies

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

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:

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

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,
- $\mathbb{E}(|\det(X''(t))| | X(t) = x, X'(t) = x') = x^{m-1}$,
- the Gaussian density $p_{X'(t)}(x')$ is bounded by $(2\pi \tilde{p}_{C,0})^{\frac{1-m}{2}}$.

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 $1 - \psi_p(C)$, a standard Gaussian vector g in \mathbb{R}^p enjoys:

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

This implies that:

$$\mathbb{P}\{Y(t) > 0\} \leq \psi_p(C), \tag{10}$$

and consequently the probability of having a local maximum above 0 on \dot{S}_\emptyset is bounded by:

$$\sigma(\dot{S}_\emptyset) (2\pi \tilde{p}_0)^{\frac{1-m}{2}} 2^{\frac{m-1}{2}} \frac{\Gamma(\frac{m-1+p}{2})}{\Gamma(\frac{p}{2})} (C\sqrt{s})^{m-1} \psi_p(C) \leq 2\sqrt{\pi} \left(\frac{C^2 s}{\tilde{p}_{C,0}}\right)^{\frac{m-1}{2}} \frac{\Gamma(\frac{m-1+p}{2})}{\Gamma(\frac{p}{2}) \Gamma(\frac{m}{2})} \psi_p(C). \tag{11}$$

Denote the right hand side of this last inequality by $h_C(s, m, p)$.

3.5. Maximum on smaller spheres

Let us now consider the case of a maximum on \mathbb{S}_A , $A \neq \emptyset$. A point $t \in \mathbb{S}_A \setminus \mathcal{W}$ is a local maximum on \mathbb{S}^{m-1} if it satisfies the following conditions:

- it is a local maximum along \mathbb{S}_A ,
- its super-gradient along the orthogonal space V_A contains zero,

where the super-gradient is defined as the opposite of the sub-gradient. One can easily check that the two conditions are independent. Indeed, recall that $k = |A|$ and $V_A = \text{Span}\{g^i; i \in A\}$ (see Section 3.2) and consider the process $X(\cdot)$ conditionally to V_A . In that case, \mathbb{S}_A becomes a deterministic sphere of dimension $m - k - 1$. Moreover, note that the behavior of $X(\cdot)$ on \mathbb{S}_A depends only on the g^j , $j \notin A$ and that for such j ,

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

so, conditionally to V_A , the distribution of $X(t)$ corresponds to the case \mathbb{S}_\emptyset in the space of dimension $m - k$ instead of m and with $p - k$ vectors. In conclusion, the first condition leads to the same computations as the case \mathbb{S}_\emptyset and is bounded by

$$h_C(s, m - k, p - k) = 2\sqrt{\pi} \left(\frac{C^2 s}{\tilde{p}_{C,k}} \right)^{\frac{p-n-1-k}{2}} \frac{\Gamma(\frac{2p-2k-n-1}{2})}{\Gamma(\frac{p-k}{2})\Gamma(\frac{p-n-k}{2})} \psi_{p-k}(C).$$

Let us look to the second one which depends only on the g^j , $j \in A$. Thus we have to compute the probability of the super-gradient to contain zero. Indeed, locally around t , the behavior of $X(w)$ along V_A 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. More precisely, for w in a neighborhood of $t \in \mathbb{S}_A \setminus \mathcal{W}$,

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

where, because $k \leq p - s$:

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

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

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

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 $\xi \in V_A$ with variance:

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

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

Let us now consider the case $i \in A$. Observe that the super-gradient along V_A 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:

$$Z_0 = \sum_{i \in A} [-g^i, g^i], \tag{12}$$

where the sum denotes the Minkowsky addition. Recall that the distribution of $X(t)$ does not depend on t .

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

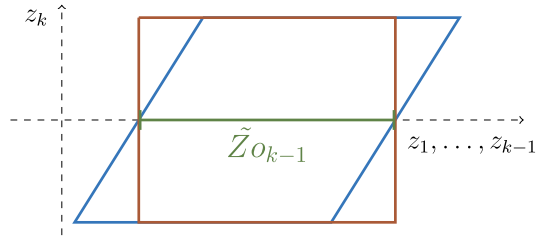


Fig. 6. The standard Gaussian measure of the zonotope (in blue) is smaller than that of the rectangle (in red) with basis $\tilde{Z}o_{k-1}$ (in green).

- draw k standard Gaussian variables g^1, \dots, g^k in \mathbb{R}^m and consider the zonotope Zo given by formula (12),
- draw in the space V_A generated by g^1, \dots, g^k an independent isotropic normal variable ξ of variance $\tilde{p}_{C,k}$,
- define $\mathbf{P}(k, \tilde{p}_{C,k}, m)$ 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) $\mathbf{P}(k, \tilde{p}_{C,k}, m)$ is less than the probability $\mathbf{Q}(k, \tilde{p}_{C,k}, m)$ of ξ to be in the hyper-rectangle:

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

(b) this last probability satisfies:

$$(\mathbf{Q}(k, \tilde{p}_{C,k}, m))^2 \leq \left(\frac{2}{\pi \tilde{p}_{C,k}} \right)^{H_k + k - m} \frac{H_k!}{(m - k)!},$$

with $H_k = \lfloor (\frac{\pi}{2} \tilde{p}_{C,k}) \wedge m \rfloor$, 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 , $\mathbf{P}(k, \tilde{p}_{C,k}, m)$ is equal to:

$$\varphi_k((\tilde{p}_{C,k})^{-1/2} \cdot Zo) =: \varphi_k(\tilde{Zo}).$$

Via Gram-Schmidt orthogonalization at step k , we can compute this probability using the Fubini theorem:

$$\mathbf{P}(k, \tilde{p}_{C,k}, m) = \int_{-\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}_{C,k}}}}^{\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}_{C,k}}}} \varphi_{k-1}(\tilde{Zo}_{k-1} + vz) \varphi(z) dz,$$

where φ is the standard Gaussian density on \mathbb{R} , \tilde{Zo}_{k-1} is the zonotope generated by g^1, \dots, g^{k-1} and normalized by $(\tilde{p}_{C,k})^{-1/2}$ and v is some vector in \mathbb{R}^m . By use of the Anderson inequality [5], the non-centered zonotope $(\tilde{Zo}_{k-1} + vz)$ has a smaller standard Gaussian measure than the centered one (see Figure 6), so

$$\begin{aligned} \mathbf{P}(k, \tilde{p}_{C,k}, m) &\leq \int_{-\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}_{C,k}}}}^{\frac{\langle e^k, g^k \rangle}{\sqrt{\tilde{p}_{C,k}}}} \varphi_{k-1}(\tilde{Zo}_{k-1}) \varphi(z) dz \\ &\leq \int_{\mathbb{R}} \varphi(z_1) \cdots \varphi(z_k) dz_1 \cdots dz_k =: \mathbf{Q}(k, \tilde{p}_{C,k}, m). \end{aligned}$$

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

(b) We use the relation above and deconditioning 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:

$$\begin{aligned} \mathbf{Q}(k, \tilde{p}_{C,k}, m) &= \mathbb{P}(\xi \in R) = \prod_{\ell=0}^{k-1} \mathbb{P}\left[|T(m - \ell)| \leq \sqrt{\frac{m - \ell}{\tilde{p}_{C,k}}}\right] \\ &= \prod_{\ell=m-k+1}^m \mathbb{P}\left[|T(\ell)| \leq \sqrt{\frac{\ell}{\tilde{p}_{C,k}}}\right]. \end{aligned}$$

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

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

In the other case, set $H_k = \lfloor (\pi \tilde{p}_{C,k})/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part. Observe that $H_k > m - k + 1$ for $k \geq 1$ to remove factors that are greater than 1 in the computation and obtain

$$(\mathbf{Q}(k, \tilde{p}_{C,k}, m))^2 \leq \left(\frac{2}{\pi \tilde{p}_{C,k}}\right)^{H_k+k-m} \frac{H_k!}{(m - k)!},$$

which conclude the proof. □

Eventually, summing up over the $\binom{p}{k}$ sets of size k , we get Theorem 4.

4. Influence of smaller spheres

4.1. General bound on the sum

In this part, we simplify the general bound of Theorem 4 to derive a simpler one, exponentially decreasing in n , as presented in Theorem 2. Considering (7), we have:

$$\Pi \leq \sqrt{\pi} \sum_{k=1}^m B_k(s, n, p), \tag{13}$$

where:

$$\begin{aligned} B_k(s, n, p) &= \binom{p}{n+k} \binom{n+k}{s} \left(\frac{C^2 s}{\tilde{p}_{m-k}}\right)^{\frac{k+1}{2}} \frac{\Gamma(\frac{n}{2} + k - \frac{1}{2})}{\Gamma(\frac{n+k}{2})\Gamma(\frac{k}{2})} \\ &\quad \times \left(\frac{C^2 4s}{\pi}\right)^{\frac{n+k-s}{2}} \frac{\Gamma(\frac{n+k}{2})\sqrt{H_{m-k}!}}{\Gamma(\frac{s}{2})(n+k-s)!\sqrt{k}!} \left(\frac{2}{\pi \tilde{p}_{C,m-k}}\right)^{\frac{H_{m-k}-k}{2}}. \end{aligned}$$

In order to derive a lower bound (the aforementioned bound goes exponentially fast towards zero), we limit our attention to the case described by

- (H1) $\rho \leq 1/2$,
- (H2) $\frac{1}{8} \leq 1 + \pi/2(1 + \rho(C^2 - 1))$.

Observe that (H1) is not a restriction since we know that NSP does not hold for $\rho \geq 0.2$. Under (H2), note that $\forall k, H_k = m$, and hence

$$\begin{aligned} \Pi &\leq \mathbf{R}(s, n, p) \frac{p!}{s! \sqrt{m!} \Gamma(s/2)} \left(\frac{4C^2s}{\pi} \right)^{\frac{n-s}{2}} \sum_{i=1}^m \binom{m}{k} \frac{(2C^2s)^k \Gamma(\frac{n}{2} + k)}{(n-s+k)!^2} \left(\frac{2}{\pi \tilde{p}_{C,m-k}} \right)^{\frac{m}{2}} \\ &\leq \mathbf{R}(s, n, p) \frac{p!}{s!(n-s)! \sqrt{m!} \Gamma(s/2)} \left(\frac{4C^2s}{\pi} \right)^{\frac{n-s}{2}} \\ &\quad \times \sum_{i=1}^m \binom{m}{k} \frac{(2C^2s)^k \Gamma(\frac{n}{2} + k)}{(n-s)^k (n-s+k)!} \left(\frac{2}{\pi \tilde{p}_{C,m-k}} \right)^{\frac{m}{2}}, \end{aligned}$$

where $\mathbf{R}(s, n, p)$ is a polynomial term in (s, n, p) . Consider now the quantity

$$\alpha(k) := \frac{\Gamma(\frac{n}{2} + k)}{(n-s+k)!}$$

which is a decreasing function of k under assumption (H1) and the fact that $\tilde{p}_{C,k}$ is an increasing function of k , then we obtain

$$\Pi \leq \mathbf{R}(s, n, p) \frac{p! \Gamma(n/2)}{s!(n-s)!^2 \sqrt{m!} \Gamma(s/2)} \left(\frac{4C^2s}{\pi} \right)^{\frac{n-s}{2}} \sum_{i=1}^m \binom{m}{k} \frac{(2C^2s)^k}{(n-s)^k} \left(\frac{2}{\pi \tilde{p}_{C,m-1}} \right)^{\frac{m}{2}}.$$

At last, using Stirling Formula (see Lemma 3), it yields

$$\Pi \leq \mathbf{R}(s, n, p) \frac{p! \Gamma(n/2)}{s!(n-s)!^2 \Gamma(s/2)} \left(\frac{4C^2s}{\pi} \right)^{\frac{n-s}{2}} \left(\frac{2e}{\pi(n+(C^2-1)s)m} \right)^{\frac{m}{2}} \left(1 + \frac{2C^2s}{n-s} \right)^m.$$

Gathering the piece, one has:

$$\begin{aligned} \Pi &\leq \mathbf{R}(s, n, p) \left(\sqrt{\frac{\pi}{2eC^2}} \frac{(n-s)^2}{s^2} \right)^s \left(C e^{\frac{\sqrt{ns m(n+(C^2-1)s)}}{(n-s)(n+(2C^2-1)s)}} \right)^n \\ &\quad \times \left(\sqrt{\frac{2}{e\pi}} \frac{p(n+(2C^2-1)s)}{(n-s)\sqrt{m(n+(C^2-1)s)}} \right)^p, \end{aligned}$$

which gives the result of Theorem 2.

Remark. The upper bound on $\alpha(k)$ and the lower bound on $\tilde{p}_{C,k}$ may seem weak but they do not change the result on the phase transition because first terms of the sum give the right order on ρ and δ . To ensure that, see Figure 7, which compare the numerical bound with the one of Theorem 2.

Appendix A: Stirling’s formula

Lemma 3. Let $z > 0$ then there exists $\theta \in (0, 1)$ such that:

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

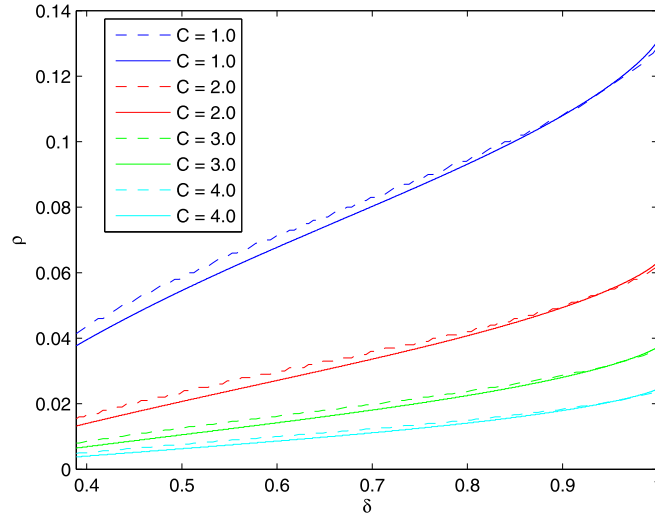


Fig. 7. Comparison between the phase transition of Theorem 2 and the numerical approximation given by Theorem 4 ($\Pi \simeq 0$ with $n = 200,000$, dashed line). From top to bottom, $C = 1, 2, 3, 4$.

In particular, if $z > 1/12$,

$$\left(\frac{z}{e}\right)^z \leq \Gamma(z + 1) \leq \sqrt{2\pi z} \left(\frac{z}{e}\right)^z.$$

Proof. See [1], Eq. 6.1.38. □

Appendix B: Concentration $\psi_l(C)$

Lemma 4. Let $C \geq 1$, then, except with a probability smaller than:

$$\psi_l(C) := \binom{l}{s} \left(\frac{C^2 4s}{\pi}\right)^{\frac{l-s}{2}} \frac{\Gamma(l/2)}{\Gamma(s/2)\Gamma(l-s+1)},$$

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

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

Proof. Let $\xi_C := \{v \in \mathbb{R}^l \text{ such that } \text{Span}(v) \text{ does not satisfy NSP}(s, C)\}$ and consider the joint law of standard Gaussian ordered statistics $(W_{(1)}, \dots, W_{(l)})$, then

$$\begin{aligned} \mathbb{P}(\xi_C) &= 2^l l! \int_{\mathbb{R}^l} \mathbb{1}_{t \in \xi_C} \mathbb{1}_{t_1 \geq \dots \geq t_l} \varphi(t_1) \cdots \varphi(t_l) dt_1 \cdots dt_l \\ &= \frac{2^l l!}{(l-s)!} \int_{\mathbb{R}^{l-s} \times \mathbb{R}^s} \mathbb{1}_{t \in \xi_C} \mathbb{1}_{t_1 \geq \dots \geq t_s} \varphi(t_1) \cdots \varphi(t_s) \varphi(t_{s+1}) \cdots \varphi(t_l) dt_1 \cdots dt_l \\ &\leq \frac{2^l l!}{(l-s)!} \left(\frac{1}{2\pi}\right)^{\frac{l-s}{2}} \int_{\mathbb{R}^s} \mathbb{1}_{t_1 \geq \dots \geq t_s} \frac{\lambda_{l-s}(B_1(C(t_1 + \dots + t_l)))}{2^{l-s}} \varphi(t_1) \cdots \varphi(t_s) dt_1 \cdots dt_s, \end{aligned}$$

where the last inequality relies on $\mathbb{P}(\mathcal{N}(0, I_{l-s}) \in B_1(t_1 + \dots + t_s))$ is bounded by the density function of $\mathcal{N}(0, I_{l-s})$ in 0 times the Lebesgue measure of the l_1 ball of radius $C(t_1 + \dots + t_s)$ in \mathbb{R}^{l-s} . Finally, as

$$\lambda_{l-s}(B_1(R)) = \frac{(2R)^{l-s}}{(l-s)!},$$

it implies,

$$\begin{aligned} \mathbb{P}(\xi_C) &\leq \left(\frac{2C^2}{\pi}\right)^{\frac{l-s}{2}} \frac{2^s l!}{(l-s)!^2} \int_{\mathbb{R}^s} (t_1 + \dots + t_s)^{l-s} \mathbb{1}_{t_1 \geq \dots \geq t_s} \varphi(t_1) \dots \varphi(t_s) dt_1 \dots dt_s \\ &= \left(\frac{2C^2}{\pi}\right)^{\frac{l-s}{2}} \frac{l!}{(l-s)!^2 s!} \mathbb{E}((|W_1| + \dots + |W_s|)^{l-s}) \\ &= \left(\frac{2C^2}{\pi}\right)^{\frac{l-s}{2}} \binom{l}{s} \frac{1}{(l-s)!} \mathbb{E}(\|W\|_1^{l-s}), \end{aligned}$$

where W is a standard Gaussian vector in \mathbb{R}^s . At last, using bound on l_1 norm, it comes,

$$\begin{aligned} \mathbb{P}(\xi_C) &\leq \left(\frac{2C^2}{\pi}\right)^{\frac{l-s}{2}} \binom{l}{s} \frac{1}{(l-s)!} s^{\frac{l-s}{2}} \mathbb{E}(\|W\|_2^{l-s}) \\ &= \left(\frac{2C^2}{\pi}\right)^{\frac{l-s}{2}} \binom{l}{s} \frac{1}{(l-s)!} s^{\frac{l-s}{2}} 2^{\frac{l-s}{2}} \frac{\Gamma(l/2)}{\Gamma(s/2)}, \end{aligned}$$

where the last equality follows from classical results on the moment of the χ distribution. □

Appendix C: Spherical Hessian

Lemma 5. Denote $X''(\cdot)$ the Hessian of $X(\cdot)$ along the sphere \mathbb{S}^{m-1} then

$$X''(t) = -X(t)I_{m-1}.$$

Proof. To compute the spherical Hessian, since every point plays the same role, we can compute it at the “east pole” $t = e_1$, the first vector of the canonical basis. Consider a basis (w_2, \dots, w_n) of the tangent space at $t = e_1$ and use, as a chart of the sphere, the orthogonal projection on this space.

Let $Y(t_2, \dots, t_m)$ be the process $X(\cdot)$ written in this chart in some neighborhood of e_1 . By the Pythagorean theorem,

$$Y(t_2, \dots, t_m) = X\left(\sqrt{1 - t_2^2 - \dots - t_m^2}, t_2, \dots, t_m\right).$$

Plugging this into the order two Taylor expansion of the process $X(\cdot)$ at $t = e_1$ gives

$$\begin{aligned} Y(t_2, \dots, t_m) &= X(e_1) + t_2 X'_2(e_1) + \dots + t_m X'_m(e_1) + \sum_{2 \leq i, j \leq m} \frac{t_i t_j}{2} X''_{ij}(e_1) \\ &\quad - X'_1(e_1) \frac{t_2^2 + \dots + t_m^2}{2} + o(t_2^2 + \dots + t_m^2), \end{aligned}$$

where $X'_k(e_1) = \frac{\partial X}{\partial w_k}(e_1)$ and $X''_{ij}(e_1) = \frac{\partial^2 X}{\partial w_i \partial w_j}(e_1)$. As the process is locally linear, in a small enough neighborhood of e_1 , X''_{ij} is equal to zero and, by identification,

$$X''(e_1) = -X'_1(e_1)I_{m-1} = -X(e_1)I_{m-1}$$

giving the desired result. □

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, Philadelphia, 1981. MR0611857
- [3] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete Comput. Geom.* **7** (1) (1992) 219–226. MR1149653
- [4] D. Amelunxen, M. Lotz, M. B. McCoy and J. A. Tropp. Living on the edge: A geometric theory of phase transitions in convex optimization. Preprint, 2013. Available at arXiv:1303.6672.
- [5] T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** (2) (1955) 170–176. MR0069229
- [6] A. Auffinger and G. B. Arous. Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.* **41** (6) (2013) 4214–4247. MR3161473
- [7] J.-M. Azaïs and M. Wschebor. On the distribution of the maximum of a Gaussian field with d parameters. *Ann. Appl. Probab.* **15** (1A) (2005) 254–278. MR2115043
- [8] J.-M. Azaïs and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. Wiley, New York, 2009. MR2478201
- [9] P. J. Bickel, Y. Ritov and A. B. Tsybakov. Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** (4) (2009) 1705–1732. MR2533469
- [10] E. J. Candès, J. Romberg and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52** (2) (2006) 489–509. MR2236170
- [11] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory* **51** (12) (2005) 4203–4215. MR2243152
- [12] E. J. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** (6) (2007) 2313–2351. MR2382644
- [13] D. Chafaï, O. Guédon, G. Lecué and A. Pajor. *Interaction Between Compressed Sensing, Random Matrices and High Dimensional Geometry. Panoramas et synthèses*, **37**. SMF, Paris, 2012. MR3113826
- [14] S. S. Chen, D. L. Donoho and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** (1) (1998) 33–61. MR1639094
- [15] A. Cohen, W. Dahmen and R. DeVore. Compressed sensing and best k -term approximation. *J. Amer. Math. Soc.* **22** (1) (2009) 211–231. MR2449058
- [16] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth. On the Lambert W function. *Adv. Comput. Math.* **5** (1) (1996) 329–359. MR1414285
- [17] F. Cucker and M. Wschebor. On the expected condition number of linear programming problems. *Numer. Math.* **94** (3) (2003) 419–478. MR1981163
- [18] Y. De Castro. A remark on the lasso and the Dantzig selector. *Statist. Probab. Lett.* **83** (1) (2013) 304–314. MR2998757
- [19] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory* **52** (4) (2006) 1289–1306. MR2241189
- [20] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.* **35** (4) (2006) 617–652. MR2225676
- [21] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA* **102** (27) (2005) 9452–9457. MR2168716
- [22] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22** (1) (2009) 1–53. MR2449053
- [23] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** (1906) (2009) 4273–4293. MR2546388
- [24] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, Berlin, 2013. MR3100033
- [25] A. Juditsky and A. Nemirovski. Accuracy guarantees for-recovery. *IEEE Trans. Inform. Theory* **57** (12) (2011) 7818–7839. MR2895363
- [26] G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. Preprint, 2014. Available at arXiv:1401.2188.
- [27] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex deconvolution, with applications. Preprint, 2012. Available at arXiv:1205.1580.
- [28] S. Mourareau, 2015. Available at <http://www.math.univ-toulouse.fr/~smourare/Software.html>.
- [29] V. Piterbarg. Comparison of distribution functions of maxima of Gaussian processes. *Teor. Veroyatn. Primen.* **26** (4) (1981) 702–719. MR0636766
- [30] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.* **61** (8) (2008) 1025–1045. MR2417886
- [31] M. Stojnic. Various thresholds for ℓ_1 -optimization in compressed sensing. Preprint, 2009. Available at arXiv:0907.3666.
- [32] M. Stojnic. A rigorous geometry-probability equivalence in characterization of ℓ_1 -optimization, Preprint, 2013. Available at arXiv:1303.7287.
- [33] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** (1) (1996) 267–288. MR1379242
- [34] S. A. van de Geer and P. Bühlmann. On the conditions used to prove oracle results for the lasso. *Electron. J. Stat.* **3** (2009) 1360–1392. MR2576316
- [35] 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) (1992) 181–201. MR1166627
- [36] W. Xu and B. Hassibi. Compressed sensing over the Grassmann manifold: A unified analytical framework. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing* 562–567. IEEE, New York, 2008.
- [37] W. Xu and B. Hassibi. Precise stability phase transitions for minimization: A unified geometric framework. *IEEE Trans. Inform. Theory* **57** (10) (2011) 6894–6919. MR2882270