



**HAL**  
open science

# Restricted Isometry Constants for Gaussian and Rademacher matrices

Sandrine Dallaporta, Yohann de Castro

► **To cite this version:**

Sandrine Dallaporta, Yohann de Castro. Restricted Isometry Constants for Gaussian and Rademacher matrices. 2016. hal-01309439v1

**HAL Id: hal-01309439**

**<https://hal.science/hal-01309439v1>**

Preprint submitted on 29 Apr 2016 (v1), last revised 12 Nov 2018 (v4)

**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.

# RESTRICTED ISOMETRY CONSTANTS FOR GAUSSIAN AND RADEMACHER MATRICES

SANDRINE DALLAPORTA AND YOHANN DE CASTRO

ABSTRACT. Restricted Isometry Constants (RICs) are a pivotal notion in Compressed Sensing (CS) as these constants finely assess how a linear operator is conditioned on the set of sparse vectors and hence how it performs in stable and robust sparse regression (SRSR). While it is an open problem to construct deterministic matrices with apposite RICs, one can prove that such matrices exist using random matrices models. One of the most popular model may be the sub-Gaussian matrices since it encompasses random matrices with Gaussian or Rademacher i.i.d. entries.

In this paper, we provide a description of the phase transition on SRSR for those matrices using state-of-the-art (small) deviation estimates on their extreme eigenvalues. In particular, we show new upper bounds on RICs for Gaussian and Rademacher matrices. This allows us to derive a new lower bound on the probability of getting SRSR. One of the benefit of this novel approach is to broaden the scope of phase transition on RICs and SRSR to the quest of universality results in Random Matrix Theory.

## 1. INTRODUCTION

**1.1. Stable and Robust Sparse Recovery (SRSR).** A popular problem addressed in recent researches aim at solving under-determined systems of linear equations (with an additive error term  $\mathbf{e}$ ) such that

$$(1) \quad \mathbf{y} = \mathbf{M}\mathbf{x}_0 + \mathbf{e}$$

where  $\mathbf{M}$  is a known  $n \times p$  matrix,  $\mathbf{x}_0$  a unknown vector in  $\mathbb{R}^p$ ,  $\mathbf{y}$  and  $\mathbf{e}$  are vectors in  $\mathbb{R}^n$  and  $n$  is (much) smaller than  $p$ . This frame fits many interests across various fields of research, e.g. in statistics one would estimate  $p$  parameters  $\mathbf{x}_0$  from a sample  $\mathbf{y}$  of size  $n$ ,  $\mathbf{M}$  being the design matrix and  $\mathbf{e}$  some random centered noise. Although the matrix  $\mathbf{M}$  is not injective, recent advances have shown that one can recover an interesting estimate  $\hat{\mathbf{x}}$  of  $\mathbf{x}_0$  considering  $\ell_1$ -minimization solutions as

$$(2) \quad \hat{\mathbf{x}} \in \arg \min \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2 \leq \eta$$

where  $\eta > 0$  is a tuning parameter. Then, a standard goal is to prove that

$$(\ell_1\text{-SRSR}) \quad \|\mathbf{x}_0 - \hat{\mathbf{x}}\|_1 \leq C\sigma_s(\mathbf{x}_0)_1 + D\sqrt{s}\eta$$

$$(\ell_2\text{-SRSR}) \quad \|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x}_0)_1 + D\eta$$

where  $C, D > 0$  are constants and  $\sigma_s(\mathbf{x}_0)_1$  denotes the approximation error in  $\ell_1$ -norm by  $s$  coefficients, namely  $\sigma_s(\mathbf{x}_0)_1 := \min \|x_0 - x\|_1$  where the minimum is taken over the space  $\Sigma_s$  of sparse vectors  $x$ , i.e. the set of vectors with at most  $s$  nonzero coordinates. The important feature described by  $(\ell_1\text{-SRSR})$  and  $(\ell_2\text{-SRSR})$  may be referenced as the Stable and Robust Sparse Recovery (SRSR) property of order  $s$ , see [FR13, Page 88]. It

---

*Date:* April 29, 2016.

*Key words and phrases.* Restricted Isometry Property; Gaussian; Rademacher; Small Deviations;

shows that  $\ell_1$ -minimization recovers the  $s$  largest coefficients of a target vector  $x_0$  in a stable<sup>1</sup> and robust (to additive errors  $\mathbf{e}$ ) manner. Interestingly, it has been shown that SRSR holds whenever the matrix  $X$  satisfies some properties, see for instance [CRT06, CT06, FL09, BRT09, vdGB09, BLPR11, JN11, DC13] or [CGLP12, FR13] for valuable books on this subject.

**1.2. Restricted Isometry Property and Restricted Isometry Constants.** One of the most important of these properties is undoubtedly the Restricted Isometry Property [CRT06, CT06] of order  $s$  and parameter  $c$ , referred to as  $\text{RIP}(s, c, \mathbf{M})$ . It is defined by

$$\forall x \in \Sigma_s, \quad (1 - c)\|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \leq (1 + c)\|x\|_2^2.$$

Denote by  $c(s, \mathbf{M})$  the minimum of such  $c$ 's. One can prove (see Theorem 6.12 in [FR13] for instance) that, if RIP such that

$$\text{(CT-}c(2s)\text{)} \quad c(2s, \mathbf{M}) < \frac{4}{\sqrt{41}} \simeq 0.625,$$

holds and  $\hat{x}$  is any solution to (2) then SRSR of order  $s$  holds with constants  $C, D$  depending only on  $c(2s, \mathbf{M})$ . A slightly modified RIP was introduced by Foucart and Lai in [FL09]. They introduce two constants, called Restricted Isometry Constants (RICs).

**Definition** (Restricted Isometry Constants (RICs)). *For a matrix  $\mathbf{M}$  of size  $(n \times p)$ , the restricted isometry constants (RICs)  $c_{\min}(s, \mathbf{M})$  and  $c_{\max}(s, \mathbf{M})$  are defined as*

$$\begin{aligned} c_{\min} &:= \min_{c_- \geq 0} c_- \quad \text{subject to} \quad (1 - c_-)\|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \quad \text{for all } x \in \Sigma_s, \\ c_{\max} &:= \min_{c_+ \geq 0} c_+ \quad \text{subject to} \quad (1 + c_+)\|x\|_2^2 \geq \|\mathbf{M}x\|_2^2 \quad \text{for all } x \in \Sigma_s. \end{aligned}$$

Hence, it holds  $(1 - c_{\min})\|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \leq (1 + c_{\max})\|x\|_2^2$  for all  $x \in \Sigma_s$ , where we recall that  $\Sigma_s$  denotes the set of vectors with at most  $s$  nonzero coordinates.

Interestingly, Foucart and Lai proved the following result.

**Theorem 1** (Theorem 2.1 in [FL09]). *If  $\mathbf{M}$  satisfies this asymmetric Restricted Isometry Property with RICs such that*

$$\text{(FL-}\gamma(2s)\text{)} \quad \gamma(2s, n, p) = \frac{1 + \sqrt{2}}{4} \left[ \frac{1 + c_{\max}}{1 - c_{\min}} - 1 \right] < 1,$$

then the Stable and Robust Sparse Recovery (SRSR) property of order  $s$  holds with positive constants  $C$  and  $D$  depending only on  $c_{\min}(2s, \mathbf{M})$  and  $c_{\max}(2s, \mathbf{M})$ .

One interesting feature of Condition (FL- $\gamma(2s)$ ) is that it report the influence of both extreme eigenvalues of covariance matrices build from  $2s$  columns of  $\mathbf{M}$ .

**1.3. Behavior of extreme eigenvalues of covariance matrices.** Let  $\mathbf{X}$  be  $s \times n$  random matrix such that its entries are i.i.d. random (real or complex) variables with mean 0 and variance 1. Define

$$\mathbf{C} := n^{-1} \mathbf{X} \mathbf{X}^*,$$

its covariance matrix. Note that  $\mathbf{C}$  is positive semidefinite real symmetric (or Hermitian). Furthermore, remark that, in our settings,  $s \leq n$  and that  $\mathbf{C}$  has  $s$  non trivial eigenvalues. These eigenvalues are real non-negative and are denoted by  $0 \leq \lambda_s \leq \dots \leq \lambda_1$ .

<sup>1</sup>In an idealized situation one would assume that  $x_0$  is sparse. Nevertheless, in practice, we can only claim that  $x_0$  is close to sparse vectors. The stability is the ability to control the estimation error  $\|x_0 - \hat{x}\|$  by the distance between  $x_0$  and the sparse vectors. The reader may consult [FR13, Page 82] for instance.

These matrices were introduced by Wishart in 1928 to model tables of random data in multivariate statistics. The spectral properties of such matrices were soon conjectured to be universal in the sense that they do not depend on the distribution of the entries. An important example is the case when the entries are Gaussian where the matrix  $\mathbf{C}$  is then said to belong to the Laguerre Unitary Ensemble (or Laguerre Orthogonal Ensemble if the entries are real). Due to the unitary invariance of the law of these matrices, the eigenvalue joint distribution is explicitly known, allowing for a complete study of the asymptotic spectral properties, see for example [AGZ10, BS10, PS11].

The following important results describe the behavior of the spectrum in the proportional regime  $\frac{s}{n} \rightarrow \rho \leq 1$  as  $n$  goes to infinity. Note that, in our settings, the sparsity  $s$  is much smaller than the number of observations  $n$ . Therefore, the ratio  $\rho$  will be supposed to be less than a constant  $\rho_0 < 1$ .<sup>2</sup>

- First among universality results, the classical Marchenko-Pastur theorem states that the empirical spectral measure  $(1/s) \sum \delta_{\lambda_j}$  converges almost surely to a measure, called the Marchenko-Pastur distribution of parameter  $\rho$ . This measure is compactly supported and is absolutely continuous with respect to Lebesgue measure, with density

$$d\mu_{MP(\rho)}(x) = \frac{1}{2\pi\rho x} \sqrt{(b_\rho - x)(x - a_\rho)} 1_{[a_\rho, b_\rho]}(x) dx,$$

where  $a_\rho := (1 - \sqrt{\rho})^2$  and  $b_\rho := (1 + \sqrt{\rho})^2$ , see for example [BS10].

- The behavior of extreme eigenvalues was more difficult to achieve. Bai *et al* [BS10] proved that, under a condition on the fourth moment of the entries,

$$\lambda_1 \xrightarrow[n \rightarrow \infty]{} b_\rho \quad \text{and} \quad \lambda_s \xrightarrow[n \rightarrow \infty]{} a_\rho \quad \text{a.s.}$$

The convergence of the smallest eigenvalue  $\lambda_s$  towards  $a_\rho$  was recently proved to hold for all covariance matrices without the assumption on the fourth moment, see [Tik15].

- At the fluctuation level, Borodin and Forrester [BF03] proved that, for LUE matrices,

$$(3) \quad \left[ \frac{n\rho^{1/4}}{a_\rho} \right]^{\frac{2}{3}} (a_\rho - \lambda_s) \xrightarrow[n \rightarrow \infty]{(d)} F_2,$$

where  $F_2$  is the so-called Tracy-Widom law. As for the largest eigenvalue, Johansson [Joh00] proved that, for LUE matrices,

$$(4) \quad \left[ \frac{n\rho^{1/4}}{b_\rho} \right]^{\frac{2}{3}} (\lambda_1 - b_\rho) \xrightarrow[n \rightarrow \infty]{(d)} F_2.$$

Similar results were obtained by the same authors for LOE matrices. As the proofs rely heavily on the Gaussian structure, universality was really hard to achieve. It was first reached by Soshnikov [Sos02] and Pécché [Péc09] for the largest eigenvalue and then by Feldheim and Sodin [FS10] for the smallest. These works rely on combinatorial techniques and require the symmetry of the law of the entries. Universality for more general models was obtained simultaneously by Pillai and Yin [PY14] on one side and Wang [Wan12] on the other side.

---

<sup>2</sup>When  $\rho = 1$ , the behavior of the smallest eigenvalues is different, the reader may consult [PS11] for more information.

Note that  $\lambda_1$  (resp.  $\lambda_s$ ) typically fluctuates around  $(1 + \sqrt{\rho})^2$  (resp.  $(1 - \sqrt{\rho})^2$ ) at a scale given by  $n^{-2/3}(1 + \sqrt{\rho})^{4/3}\rho^{-1/6}$  (resp.  $n^{-2/3}(1 - \sqrt{\rho})^{4/3}\rho^{-1/6}$ ).

We understand that the benefit of the RIC constants  $c_{\min}(s, \mathbf{M})$  and  $c_{\max}(s, \mathbf{M})$  upon the RIP constant  $c(s, \mathbf{M})$  is that they precisely account for typical values of both the smallest and the largest eigenvalues of empirical covariance matrices. Indeed, suppose that  $x$  belongs to  $\Sigma_s$  and denote by  $S$  a subset of size  $s$  of  $\{1, \dots, p\}$  such that it contains the support (i.e. the set of the indices of the nonzero coefficients) of  $x$ . Denote by  $\mathbf{M}_S$  the  $n \times s$  matrix obtained from  $\mathbf{M}$  by keeping only the columns indexed by  $S$ . Then  $\|\mathbf{M}x\|_2^2 = x^\top \mathbf{M}_S^\top \mathbf{M}_S x$  where  $\mathbf{M}_S^\top \mathbf{M}_S$  is a  $s \times s$  real symmetric matrix. It has  $s$  real non-negative eigenvalues and therefore

$$\lambda_{\min}^S \|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \leq \lambda_{\max}^S \|x\|_2^2,$$

where  $\lambda_{\min}^S$  and  $\lambda_{\max}^S$  are the extreme eigenvalues of  $\mathbf{M}_S^\top \mathbf{M}_S$ . Getting bounds depending only on  $s$ ,  $p$  and  $n$  on such extreme eigenvalues leads to bounds on the RICs. Indeed, it holds

$$(5) \quad \forall x \in \Sigma_s, \quad \inf_{S \subset \{1, \dots, p\}, \#S=s} \lambda_{\min}^S \|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \leq \sup_{S \subset \{1, \dots, p\}, \#S=s} \lambda_{\max}^S \|x\|_2^2.$$

As we have seen above, the couple  $(\inf \lambda_{\min}^S, \sup \lambda_{\max}^S)$  is not necessarily symmetric with respect to 1. Observe that RIP does not seize this asymmetry while RICs are meant to portray it.

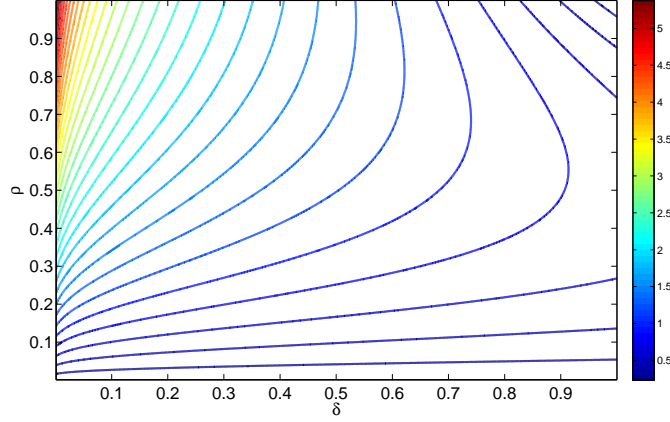
The convergence in distribution of the extreme eigenvalues gives intuition on what is expected for their asymptotic behavior. However getting bounds on RICs requires some non asymptotic counterparts of the preceding results. Precise deviation inequalities for the extreme eigenvalues will be given in Section 2.

**1.4. From deviation inequalities to RICs and SRSR bounds.** This paper provides a simple tool to derive lower bounds on SRSR or upper bounds on RICs from deviation inequalities on extreme eigenvalues (or extreme singular values) of covariance matrices  $\mathbf{C}_{s,n} = \frac{1}{n} \mathbf{X} \mathbf{X}^*$  where  $\mathbf{X} \in \mathbb{R}^{s \times n}$  has i.i.d. entries drawn with respect to  $\mathcal{L}$ . In the asymptotic proportional growth model where  $s/n \rightarrow \rho$  and  $n/p \rightarrow \delta$ , we assume that we have access to a deviation inequality on extreme eigenvalues with rate function  $t \mapsto \mathbb{W}(\rho, t)$  depending on the ratio  $\rho$ . For instance, we will consider that for all  $n \geq n_0(\rho)$ ,

$$\forall 0 \leq t < \tau_1, \quad \mathbb{P}\left\{(\lambda_1 - (1 + \sqrt{\rho})^2) \vee ((1 - \sqrt{\rho})^2 - \lambda_s) \geq t\right\} \leq c(\rho) e^{-n\mathbb{W}(\rho, t)}$$

where  $\tau_1 \in \overline{\mathbb{R}}$ ,  $n_0(\rho) \geq 2$  and  $c(\rho) > 0$  may both depend on the ratio  $\rho$ , the function  $t \mapsto \mathbb{W}(\rho, t)$  is continuous and increasing on  $[0, \tau_1)$  such that  $\mathbb{W}(\rho, 0) = 0$ . Notably, it appears along our analysis that bounds on SRSR and RICs are extremely dependent on the behavior, for fixed  $t$ , of the rate function  $\rho \mapsto \mathbb{W}(\rho, t)$  when  $\rho$  is small, and possibly tending to zero. Unfortunately, this dependence is overlooked in the literature and we have to take another look at state-of-the-art results in this field. Revisiting the captivating paper of Feldheim and Sodin [FS10] on sub-Gaussian matrices, Appendix A reveals the dependency on  $\rho$  as well as bounds on the constant appearing in their rate function  $\mathbb{W}_{FS}$  for the special case of Rademacher entries. Other important results due to Ledoux and Rider [LR10], and Davidson and Szarek [DS01] are investigated in Section 3.1.

The rate function  $\mathbb{W}$  at hand, our paper provides a simple tool to derive bounds on RICs and SRSR as shown in the following two subsections.

FIGURE 1. The combinatorial term  $\delta^{-1}\mathbf{H}_e(\rho\delta)$ .

**1.5. Discrepancy of the RICs.** Following the analysis of [BCT11] we provide asymptotic bounds on RICs in the proportional growth model. We aim at controlling uniformly the extreme eigenvalues as in (5), the combinatorial complexity is standardly [BCT11] controlled by the quantity  $\delta^{-1}\mathbf{H}_e(\rho\delta)$  where  $\mathbf{H}_e(t) = -t \log t - (1-t) \log(1-t)$  for  $t \in (0, 1)$  denotes the Shannon entropy, see Figure 1. This quantity governs the value of the deviation  $t_0 := \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$  in the rate function  $\mathbb{W}(\rho, t)$  when bounding the extreme eigenvalues uniformly over all possible supports  $S$  of size  $s$  among the set of indices  $\{1, \dots, p\}$ , see (5). The upper bounds on RICs given by the rate functions  $\mathbb{W}_{LR}$  of [LR10] and  $\mathbb{W}_{DS}$  of [DS01] are presented in Figure 2.

**Theorem 2.** *The following holds for each couple  $(\mathbb{W}, \mathcal{L})$  defined by, for all  $t > 0$ , for all  $\rho \in (0, 1)$ ,*

$$\begin{aligned} \mathbb{W}_{LR}(\rho, t) &:= \frac{\rho^{\frac{1}{2}}}{C_{LR}(1 + \sqrt{\rho})^3} t^{\frac{3}{2}} 1_{t \leq \sqrt{\rho}(1 + \sqrt{\rho})^2} & \text{and } \mathcal{L}_{LR} &:= \mathcal{N}(0, 1) \\ &+ \frac{\rho^{\frac{1}{2}}}{C_{LR}(1 + \sqrt{\rho})^2} t 1_{t > \sqrt{\rho}(1 + \sqrt{\rho})^2} \\ \mathbb{W}_{FS}(\rho, t) &:= \frac{\rho \log(1 + \frac{t}{2\sqrt{\rho}})^{\frac{3}{2}}}{C_{FS}(1 + \sqrt{\rho})^2} & \text{and } \mathcal{L}_{FS} &:= \text{Rademacher} \end{aligned}$$

where  $C_{LR} > 0$  and  $837 \geq C_{FS} > 0$  are numerical constants. For any  $\varepsilon > 0$ , any  $\delta \in (0, 1)$  and any  $\rho \in (0, 1)$ , it holds

$$\begin{aligned} \mathbb{P}\left\{\epsilon_{\min} \geq \min\{1, \sqrt{\rho}(2 - \sqrt{\rho}) + t_0\} + \varepsilon\right\} &\leq c(\rho) e^{-nD(\rho, \delta, \varepsilon)}, \\ \mathbb{P}\left\{\epsilon_{\max} \geq \sqrt{\rho}(2 + \sqrt{\rho}) + t_0 + \varepsilon\right\} &\leq c(\rho) e^{-nD(\rho, \delta, \varepsilon)}, \end{aligned}$$

where  $D(\rho, \delta, \varepsilon) > 0$  and  $t_0 := \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$ .

The proof is given in Section 3.1, see Corollary 1. Roughly speaking, the RICs are upper bounded by  $\sqrt{\rho}(2 \pm \sqrt{\rho}) + \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$  with overwhelming probability, where  $\mathbb{W}^{-1}(\rho, \cdot)$  denotes the inverse function of  $\mathbb{W}(\rho, \cdot)$ .

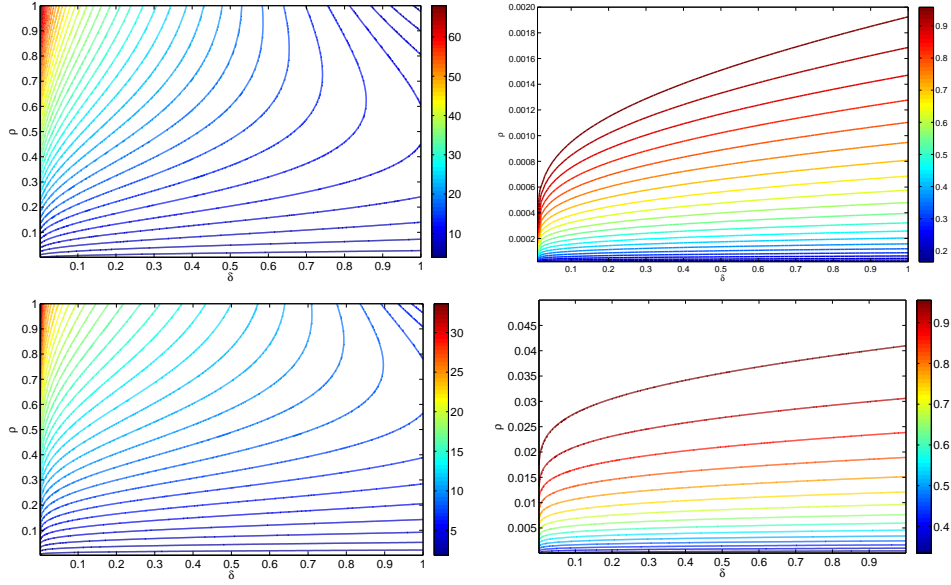


FIGURE 2. Ledoux and Rider bounds on RICs (the constant  $c_{\max}$  is on the top left panel and the constant  $c_{\min}$  is on the top right panel). The constant in Ledoux and Rider lower bound is arbitrarily set to  $C_{LR} = 3$  as we did not track an upper bound on this constant. Davidson and Szarek bounds on RICs (the constant  $c_{\max}$  is on the bottom left panel and the constant  $c_{\min}$  is on the bottom right panel).

**Theorem 3.** Consider the couple  $(\mathbb{W}, \mathcal{L})$  defined by, for all  $t > 0$ ,

$$\mathbb{W}_{DS}(t) := t^2/2 \quad \text{and} \quad \mathcal{L}_{DS} := \mathcal{N}(0, 1).$$

For any  $\varepsilon > 0$ , any  $\delta \in (0, 1)$  and any  $\rho \in (0, 1)$ , it holds

$$\mathbb{P}\{c_{\max} \geq (\sqrt{\rho} + t_0)(2 + \sqrt{\rho} + t_0) + \varepsilon\} \leq c(\rho)e^{-nD(\rho, \delta, \varepsilon)},$$

and if further  $t_0 + \sqrt{\rho} < 1$  then

$$\mathbb{P}\{c_{\min} \geq (\sqrt{\rho} + t_0)(2 - \sqrt{\rho} - t_0) + \varepsilon\} \leq c(\rho)e^{-nD(\rho, \delta, \varepsilon)},$$

where  $D(\rho, \delta, \varepsilon) > 0$  and  $t_0 := \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$ .

These upper bounds on RICs are depicted in Figure 2. The bound of Davidson and Szarek (with rate function  $\mathbb{W}_{DS}$ ) is rather close to the one obtained in [BCT11] which gives an exact computation of RICs bounds using the joint law of the Gaussian Wishart eigenvalues. Based on  $\mathbb{W}_{DS}$ , a similar upper bound on the RIP constant can be found in [CT05, Lemma 3.1]. Though our work provides a simple bounds on RICs in the Rademacher case, we did not illustrate these bounds since it compares unfavorably to the ones obtained in the Gaussian case. In particular, the dependance in  $\rho$  in  $\mathbb{W}_{FS}$  is far from being optimal and leads to extremely large bounds on RICs. The “ideal” rate function is presented in Section 3.6 while the dependence of  $\mathbb{W}_{FS}$  in  $\rho$  is studied in Appendix A.

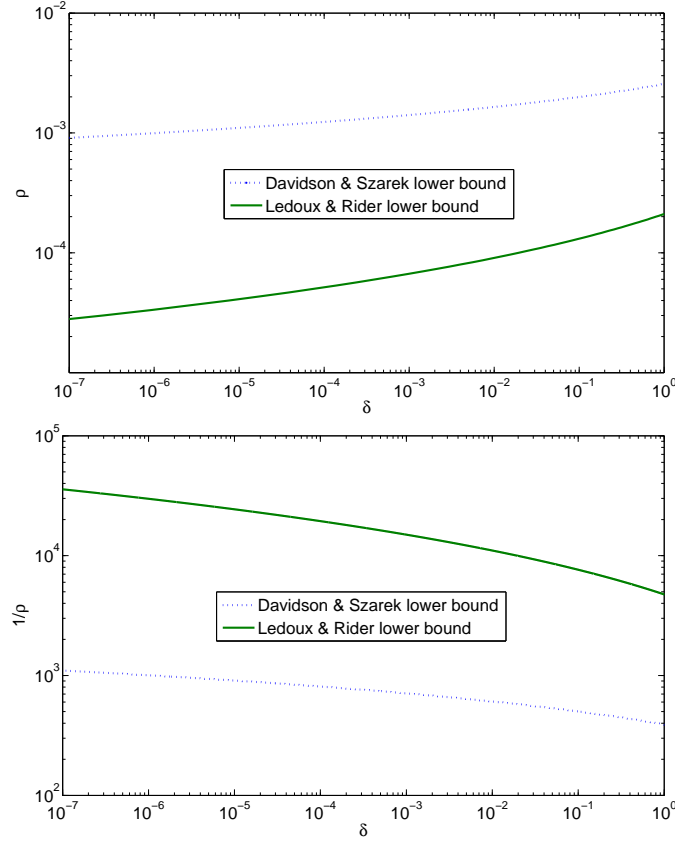


FIGURE 3. Various lower bounds on SRSR using Condition  $(FL-\gamma(2s))$ . The constant in Ledoux and Rider lower bound is arbitrarily set to  $C_{LR} = 3$  as we did not track an upper bound on this constant.

All figures can be reproduced using the [following code](#).

**1.6. Explicit lower bounds on SRSR.** In this paper, we prove that SRSR holds for Gaussian and Rademacher matrices using Condition  $(FL-\gamma(2s))$  and explicit small deviations estimates, see Section 2 for more detail on this latter point. Actually these results provide a general method to prove SRSR for matrices with entries satisfying small deviations estimates. Set

$$\tau_0 := \frac{2}{3 + \sqrt{2}} \simeq 0.4531 \quad \text{and} \quad \rho_0 := \frac{[3 + \sqrt{2} - \sqrt{7 + 6\sqrt{2}}]^2}{4} \simeq 0.0574,$$

and note that  $\sqrt{\rho_0} = 0.2395$ . The bounds given by the rate functions  $\mathbb{W}_{LR}$  of [LR10] and  $\mathbb{W}_{DS}$  of [DS01] are presented in Figure 3.



**Theorem 4.** *The following holds for each couple  $(\mathbb{W}, \mathcal{L})$  defined by, for all  $t \in (0, \tau_0)$ , for all  $\rho \in (0, \rho_0)$ ,*

$$\begin{aligned} \mathbb{W}_{LR}(\rho, t) &:= \frac{\rho^{\frac{1}{4}}}{C_{LR}(1 + \sqrt{\rho})^3} t^{\frac{3}{2}} 1_{t \leq \sqrt{\rho}(1 + \sqrt{\rho})^2} & \text{and } \mathcal{L}_{LR} &:= \mathcal{N}(0, 1) \\ &+ \frac{\rho^{\frac{1}{2}}}{C_{LR}(1 + \sqrt{\rho})^2} t 1_{t > \sqrt{\rho}(1 + \sqrt{\rho})^2} \\ \mathbb{W}_{FS}(\rho, t) &:= \frac{\rho \log(1 + \frac{t}{2\sqrt{\rho}})^{\frac{3}{2}}}{C_{FS}(1 + \sqrt{\rho})^2} & \text{and } \mathcal{L}_{FS} &:= \text{Rademacher} \end{aligned}$$

where  $C_{LR} > 0$  and  $837 \geq C_{FS} > 0$  are numerical constants. For any  $\delta \in (0, 1)$  and any  $\rho \in (0, \rho_0)$  such that

$$\delta > \frac{1}{\rho} \exp \left[ 1 - \frac{\mathbb{W}[\rho, \tau_0(\sqrt{\rho} - \sqrt{\rho_0})(\sqrt{\rho} - 1/\sqrt{\rho_0})]}{\rho} \right],$$

it holds that any sequence of  $n \times p$  matrices  $(\mathbf{M}^{(n)})_{n \geq 2}$  with i.i.d. entries with respect to  $\mathcal{L}$  and such that  $n/p \rightarrow \delta$  satisfy

$$\mathbb{P} \left\{ \frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ conforms to } (\text{FL-}\gamma(2s)) \text{ with } 2s \leq \lfloor \rho n \rfloor \right\} \rightarrow 1$$

as  $n$  tends to infinity.

The  $\mathbb{W}$  function appearing in Theorem 5 below arises from deviation inequalities on singular values of Gaussian covariance matrices, see Lemma 7. Details will be given in Section 3.2.

**Theorem 5.** *Consider the couple  $(\mathbb{W}, \mathcal{L})$  defined by, for all  $t \in (0, \sqrt{\rho_0})$ ,*

$$\mathbb{W}_{DS}(t) := t^2/2 \quad \text{and} \quad \mathcal{L}_{DS} := \mathcal{N}(0, 1).$$

For any  $0 < \delta < 1$  and  $0 < \rho < \rho_0$  such that

$$\delta > \frac{1}{\rho} \exp \left[ 1 - \frac{\mathbb{W}[\sqrt{\rho_0} - \sqrt{\rho}]}{\rho} \right],$$

it holds that any sequence of  $n \times p$  matrices  $(\mathbf{M}^{(n)})_{n \geq 2}$  with i.i.d. entries with respect to  $\mathcal{L}$  and such that  $n/p \rightarrow \delta$  satisfy

$$\mathbb{P} \left\{ \frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ conforms to } (\text{FL-}\gamma(2s)) \text{ with } 2s \leq \lfloor \rho n \rfloor \right\} \rightarrow 1$$

as  $n$  tends to infinity.

Together with Theorem 1, the lower bounds given by the aforementioned theorems are also lower bounds on the SRSR property (i.e. they guarantee that SRSR holds).

**1.7. Previous works on bounding RIP and RICs.** The existence of RIP matrices with bounded RIP constant such as  $(\text{CT-c}(2s))$  has been proved using random matrix models, see [MPTJ08, ALPTJ11, CGLP12] for instance. This approach has encountered a large echo and it might be seen as a pillar of the theory of Compressed Sensing. Popular results show that  $(\text{CT-c}(2s))$  holds with overwhelming probability for a large class of random matrix models as soon as the interplay between sparsity  $s$ , number of measurements  $n$  and number of unknown parameters  $p$  satisfies

$$(6) \quad n \geq c_1 s \log(c_2 p/s)$$

for some universal constants  $c_1$  and  $c_2$  (that might depend on the random matrix model). It should be mentioned that finding deterministic matrices satisfying (CT-c(2s)) with  $n = \mathcal{O}(s \log(p/s))$  is one of the most prominent open problem in Compressed Sensing, see [FR13] for instance. Furthermore, it has been shown in [CGLP12, Proposition 2.2.17] that the converse is true for any matrix  $\mathbf{M}$ . If the SRSR recovery ( $\ell_1$ -SRSR) or ( $\ell_2$ -SRSR) (with  $\eta = 0$ ) holds then necessarily  $n \geq c'_1 s \log(c'_2 p/s)$  for some universal constants  $c'_1$  and  $c'_2$ . Since we have lower and upper bounds of the same flavor, it seems that the condition (6) captures all we need to know about  $\ell_1$ -recovery schemes. In reality, there is a gap between the constants appearing in the upper and lower bounds. A simple way to witness it is to consider the companion problem when there is no additive errors. In this case  $\mathbf{e} = 0$  in (1) and  $\eta = 0$  in (2), then stable recovery occurs for all target vector  $x_0$  if and only a property called “Null-Space Property” (NSP) holds. As for RIP one can prove that (6) depicts a necessary and sufficient condition on NSP up to a change of constants, see for instance [CGLP12, ADCM14]. Nevertheless, a better description of this property is offered in the works [DT05, DT09a, DT09b] since the authors provide a phase transition on NSP for large Gaussian matrices with i.i.d. entries.

Following this outbreaking result, one can wonder whether a phase transition holds for properties guaranteeing SRSR such as (CT-c(2s)) or the asymmetric (FL- $\gamma$ (2s)). To the best of our knowledge, the only work looking for a phase transition on SRSR can be found in the captivating paper [BCT11] where the authors considered matrices with independent standard Gaussian entries and used an upper bound on the joint density of the eigenvalues to derive a region where (FL- $\gamma$ (2s)) holds. Their lower bound is not explicit but one can witness in [BCT11, Page 119]. In the present paper, we provide explicit lower bounds on SRSR using Condition (FL- $\gamma$ (2s)) and small deviation techniques. One of the benefits of this novel approach is to broaden the scope of RICs to the frame of universality results in Random Matrix Theory.

## 2. SMALL DEVIATIONS ON EXTREME EIGENVALUES OF COVARIANCE MATRICES

Recall Eqs. (3) and (4) where  $F_\beta$  is the so-called Tracy-Widom law (with  $\beta = 1$  if the entries are real and  $\beta = 2$  if they are complex). Deviation inequalities for  $\lambda_1$  and  $\lambda_s$  are needed to get a bound on RICs with high probability, i.e. control on  $\mathbb{P}(\lambda_1 \geq b_\rho + t)$  and  $\mathbb{P}(\lambda_s \leq a_\rho - t)$  for  $t > 0$ . We focus on the largest eigenvalue  $\lambda_1$  and write:

$$\mathbb{P}(\lambda_1 \geq b_\rho + t) = \mathbb{P}\left(\frac{n^{2/3} \rho^{1/6}}{b_\rho^{2/3}} (\lambda_1 - b_\rho) \geq \frac{n^{2/3} \rho^{1/6}}{b_\rho^{2/3}} t\right).$$

This deviation probability is therefore expected to be close to

$$1 - F_\beta\left(\frac{n^{2/3} \rho^{1/6}}{b_\rho^{2/3}} t\right),$$

where  $F_\beta$  is the cdf of the Tracy-Widom distribution. Thus it is expected to be close to the tail behavior of  $F_\beta$  at  $\infty$ . This tail behavior is actually known:

$$1 - F_\beta(x) \underset{x \rightarrow \infty}{\sim} C \exp(-cx^{3/2}),$$

with  $C, c > 0$ . As a consequence, deviation inequalities for the largest eigenvalue are expected to conform to

$$(7) \quad \mathbb{P}(\lambda_1 \geq b_\rho + t) \leq C \exp\left(-c \frac{\rho^{1/4}}{(1 + \sqrt{\rho})^2} nt^{3/2}\right),$$

at least for bounded  $t$  (possibly depending on  $\rho$ ). Similarly, deviation inequalities for the smallest eigenvalue are expected to be close to

$$(8) \quad \mathbb{P}(\lambda_s \leq a_\rho - t) \leq C \exp\left(-c \frac{\rho^{1/4}}{(1 - \sqrt{\rho})^2} nt^{3/2}\right).$$

For bigger  $t$ , the expected behavior is sub-exponential

$$(9) \quad \mathbb{P}(\lambda_1 \geq b_\rho + t) \leq C e^{-cnt}.$$

See [Led07] for a detailed survey on this subject. Unfortunately, such precise small deviation inequalities are really hard to achieve, even for Gaussian matrices.

- Davidson and Szarek consider a  $s \times n$  matrix  $\mathbf{X}$  with i.i.d. standard Gaussian entries. Relying on the concentration of measure phenomenon, they establish the following:

$$\forall t > 0, \quad \mathbb{P}\left\{(\sigma_1 - (1 + \sqrt{\rho})) \vee ((1 - \sqrt{\rho}) - \sigma_s) \geq t\right\} \leq 2e^{-nt^2/2},$$

where  $\sigma_i(\mathbf{X})$  denotes the singular values of  $\frac{1}{\sqrt{n}}\mathbf{X}$ . See for instance [FR13] p. 291. It can be written in terms of eigenvalues of the covariance matrix  $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ : for all  $u > 0$ ,

$$\mathbb{P}\left\{\lambda_1 - b_\rho \geq u\right\} \leq 2 \exp\left[-nb_\rho\left(1 - \sqrt{1 + \frac{u}{b_\rho} + \frac{u}{2b_\rho}}\right)\right].$$

Note that, when  $u$  is small, the left-hand side becomes  $2e^{-nu^2/(8b_\rho)}$ , which is not the expected order for small deviations. For large  $u$ , it provides the expected sub-exponential behavior. A similar inequality can be deduced for  $\lambda_s$ . Note that the constant  $\frac{1}{2}$  in the exponent is explicit and rather large, which is very good.

- In [LR10], Ledoux and Rider give unified proofs of small deviation inequalities for  $\beta$  ensembles, which include Gaussian covariance matrices  $\mathbf{C}_{s,n} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$  ( $\mathbf{X}$  is of size  $s \times n$ ). They establish the following: for all  $0 < \rho < 1$  and for all  $n \geq 2$ , setting  $s = \lfloor \rho n \rfloor$ ,  $\forall 0 < t \leq \sqrt{\rho}(1 + \sqrt{\rho})^2$ ,

$$\mathbb{P}\left\{\lambda_1 - (1 + \sqrt{\rho})^2 \geq t\right\} \leq c(\rho) \exp\left(-n \frac{\rho^{1/4}}{C_{LR}(1 + \sqrt{\rho})^3} t^{3/2}\right).$$

For bigger  $t$ , they get the expected sub-exponential behavior. In particular, for all  $t > \sqrt{\rho}(1 + \sqrt{\rho})^2$ ,

$$\mathbb{P}\left\{\lambda_1 - (1 + \sqrt{\rho})^2 \geq t\right\} \leq c(\rho) \exp\left(-n \frac{\rho^{1/2}}{C_{LR}(1 + \sqrt{\rho})^2} t\right).$$

Similar inequalities are established for  $\lambda_s$ . These inequalities are exactly what is expected for small  $t$ : they reflect the fluctuation rate and the tail behavior of the Tracy-Widom distribution. The authors did not mention it but it seems possible to bound explicitly the constant  $C_{LR}$ . However, we did not track an upper bound for this constant.

- In the paper [FS10], Feldheim and Sodin consider sub-Gaussian covariance matrices  $\mathbf{C}_{s,n} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$  ( $\mathbf{X}$  is of size  $s \times n$ ), with symmetric distribution. They establish the following: for all  $0 < \rho < 1$  and for all  $n \geq 2$ , setting  $s = \lfloor \rho n \rfloor$ ,

$$\forall t > 0, \quad \mathbb{P}\left\{\lambda_1 - (1 + \sqrt{\rho})^2 \geq t\right\} \leq c(\rho) \exp\left\{-n \frac{\rho \log\left(1 + \frac{t}{2\sqrt{\rho}}\right)^{\frac{3}{2}}}{C_{FS}(1 + \sqrt{\rho})^2}\right\},$$

and a similar inequality for  $\lambda_s$ . Note that, when  $\frac{t}{\sqrt{\rho}}$  is small, the left-hand side has exactly the expected order:

$$c(\rho) \exp\left(-n \frac{\rho^{1/4}}{C_{FS}(1 + \sqrt{\rho})^2} t^{\frac{3}{2}}\right).$$

In the special case of independent Rademacher entries, it is possible to bound the constant  $C_{FS}$ . This will be the object of Appendix A. However, the bound we obtain by following carefully the proof is really large, so that  $\frac{1}{C_{LR}}$  is very small.

As far as we know, these are the only precise small deviation inequalities which are established for extreme eigenvalues of covariance matrices.

### 3. EXPLICIT LOWER BOUNDS

**3.1. Key lemmas using eigenvalues.** We present key lemmas allowing us to derive a lower bound on the probability of getting Condition (FL- $\gamma(2s)$ ). We recall that  $\tau_0 := \frac{2}{3+\sqrt{2}} \simeq 0.4531$  and  $\sqrt{\rho_0} := (3 + \sqrt{2} - \sqrt{7 + 6\sqrt{2}})/2 \simeq 0.2395$ .

**Lemma 6.** *Assume that for all  $0 < \rho < 1$ , the largest eigenvalue  $\lambda_1$  and the smallest eigenvalue  $\lambda_s$  of the covariance matrix*

$$\mathbf{C}_{s,n} := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^{(s)} (\mathbf{X}_i^{(s)})^*,$$

where  $s := \lfloor \rho n \rfloor$  and  $\mathbf{X}_i^{(s)}$  are random vectors in  $\mathbb{R}^s$  with i.i.d. entries with respect to a law  $\mathcal{L}$ , satisfy for all  $n \geq n_0(\rho)$ ,

$$\forall 0 \leq t < \tau_0, \quad \mathbb{P}\left\{(\lambda_1 - (1 + \sqrt{\rho})^2) \vee ((1 - \sqrt{\rho})^2 - \lambda_s) \geq t\right\} \leq c(\rho) e^{-n\mathbb{W}(\rho,t)}$$

where  $n_0(\rho) \geq 2$  and  $c(\rho) > 0$  may both depend on  $\rho$ , the function  $t \mapsto \mathbb{W}(\rho, t)$  is continuous and increasing on  $[0, \tau_0)$  such that  $\mathbb{W}(\rho, 0) = 0$ . Then for any  $0 < \delta < 1$  and  $0 < \rho < \rho_0$  such that

$$(10) \quad \delta > \frac{1}{\rho} \exp\left[1 - \frac{\mathbb{W}\left(\rho, \frac{2}{3+\sqrt{2}}(1 - (3 + \sqrt{2})\sqrt{\rho} + \rho)\right)}{\rho}\right],$$

any sequence of  $n \times p$  matrices  $(\mathbf{M}^{(n)})_{n \geq 2}$  with i.i.d. entries with respect to  $\mathcal{L}$  and such that  $n/p \rightarrow \delta$  satisfy

$$\mathbb{P}\left\{\frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ satisfies (FL-}\gamma(2s)\text{) with } 2s \leq \lfloor \rho n \rfloor\right\} \rightarrow 1$$

as  $n$  tends to infinity.

*Proof.* Let  $t \in (0, \tau_0)$ . A simple calculation gives that, on the event

$$\{c_{\max} < \sqrt{\rho}(2 + \sqrt{\rho}) + t\} \cup \{c_{\min} < \sqrt{\rho}(2 - \sqrt{\rho}) + t\},$$

the Condition (FL- $\gamma(2s)$ ) is satisfied whenever

$$(11) \quad -2 + 2(3 + \sqrt{2})\sqrt{\rho} - 2\rho + (3 + \sqrt{2})t < 0.$$

Indeed, observe that

$$\frac{1 + c_{\max}}{1 - c_{\min}} - 1 < \frac{(1 + \sqrt{\rho})^2 + t}{(1 - \sqrt{\rho})^2 - t} - 1 = \frac{4\sqrt{\rho} + 2t}{(1 - \sqrt{\rho})^2 - t}.$$

Fix  $\rho$  and  $\delta$  as in (10) and consider parameters  $s$  and  $n$  so that  $s/n \rightarrow \rho$  as  $n$  goes to infinity. Choosing  $s$  columns over  $p$  in  $\mathbf{M}$  and considering the covariance matrix  $\mathbf{C}_{s,n}$  of those columns, it holds for  $n \geq n_0(\rho)$ ,

$$\begin{aligned} \mathbb{P}\left\{\exists x \in \Sigma_s \text{ s.t. } n^{-1}\|\mathbf{M}x\|^2 \geq ((1 + \sqrt{\rho})^2 + t)\|x\|^2\right\} &\leq \sum_{\mathbf{C}_{s,n}} \mathbb{P}\left\{\lambda_1(\mathbf{C}_{s,n}) \geq (1 + \sqrt{\rho})^2 + t\right\} \\ &\leq \binom{p}{s} c(\rho) e^{-n\mathbb{W}(\rho,t)} \\ &\leq c(\rho) e^{-n\mathbb{W}(\rho,t) + p\mathbf{H}_e(s/p)} \\ &\leq c(\rho) e^{-nD}, \end{aligned}$$

with  $\mathbf{H}_e(t) = -t \log t - (1-t) \log(1-t)$  for  $t \in (0, 1)$  and  $D = \mathbb{W}(\rho, t) - \frac{1}{\delta} \mathbf{H}_e(\rho\delta)$ . Denote  $\mathbb{W}^{-1}(\rho, \cdot)$  the inverse of the function  $\mathbb{W}(\rho, \cdot)$ . Now, recall that  $D + \frac{1}{\delta} \mathbf{H}_e(\rho\delta)$  belongs to the range of the function  $\mathbb{W}(\rho, \cdot)$  and consider

$$t_D := \mathbb{W}^{-1}\left(\rho, D + \frac{1}{\delta} \mathbf{H}_e(\rho\delta)\right).$$

Note that  $\mathbb{W}(\rho, t_D) = D + \frac{1}{\delta} \mathbf{H}_e(\rho\delta)$ . We deduce that it holds

$$\mathbb{P}\{c_{\max} \geq \sqrt{\rho}(2 + \sqrt{\rho}) + t_D\} \leq c(\rho) e^{-nD},$$

for all  $D$  that can be written as  $D = \mathbb{W}(\rho, t) - \frac{1}{\delta} \mathbf{H}_e(\rho\delta)$  with  $0 \leq t < \tau_0$ . Following the same arguments, we get a similar inequality for  $c_{\min}$ .

Furthermore, using  $\mathbf{H}_e(t) \leq -t \log t + t$  we get that the condition (10) implies

$$(12) \quad 0 = \mathbb{W}(\rho, 0) < \frac{1}{\delta} \mathbf{H}_e(\rho\delta) < \mathbb{W}\left(\rho, \frac{2}{3 + \sqrt{2}}(1 - (3 + \sqrt{2})\sqrt{\rho} + \rho)\right)$$

Note that  $\mathbb{W}^{-1}(\rho, \cdot)$  is continuous and increasing on the range of  $\mathbb{W}(\rho, \cdot)$  and set

$$t_0 := \mathbb{W}^{-1}\left(\rho, \frac{1}{\delta} \mathbf{H}_e(\rho\delta)\right).$$

Eq. (12) shows that  $\frac{1}{\delta} \mathbf{H}_e(\rho\delta)$  belongs to the interior of the range of  $\mathbb{W}(\rho, \cdot)$ . Invoke the continuity of  $\mathbb{W}^{-1}(\rho, \cdot)$  at point  $\frac{1}{\delta} \mathbf{H}_e(\rho\delta)$  to see that  $t_D$  tends to  $t_0$  as  $D$  goes to 0. Applying  $\mathbb{W}^{-1}(\rho, \cdot)$ , we deduce that (12) is equivalent to

$$(13) \quad -2 + 2(3 + \sqrt{2})\sqrt{\rho} - 2\rho + (3 + \sqrt{2})t_0 < 0.$$

Then for  $D > 0$  small enough it holds

$$(14) \quad -2 + 2(3 + \sqrt{2})\sqrt{\rho} - 2\rho + (3 + \sqrt{2})t_D < 0,$$

by continuity. It follows that Condition (FL- $\gamma(2s)$ ) fails with a probability smaller than  $2c(\rho)e^{-nD}$  for some small enough  $D > 0$  (that may depend on  $\delta$  and  $\rho$ ). We conclude

that the Condition (FL- $\gamma(2s)$ ) holds under the Condition (14) on  $\rho$  and  $\delta$  which is implied by (10).  $\square$

From the previous proof, we deduce the following result.

**Corollary 1.** *Assume that for all  $0 < \rho < 1$ , the largest eigenvalue  $\lambda_1$  and the smallest eigenvalue  $\lambda_s$  of the covariance matrix*

$$\mathbf{C}_{s,n} := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^{(s)} (\mathbf{X}_i^{(s)})^*,$$

where  $s := \lfloor \rho n \rfloor$  and  $\mathbf{X}_i^{(s)}$  are random vectors in  $\mathbb{R}^s$  with i.i.d. entries with respect to a law  $\mathcal{L}$ , satisfy for all  $n \geq n_0(\rho)$ ,

$$\forall 0 \leq t < \tau_1, \quad \mathbb{P}\left\{(\lambda_1 - (1 + \sqrt{\rho})^2) \vee ((1 - \sqrt{\rho})^2 - \lambda_s) \geq t\right\} \leq c(\rho) e^{-n\mathbb{W}(\rho,t)}$$

where  $\tau_1 \in \overline{\mathbb{R}}$ ,  $n_0(\rho) \geq 2$  and  $c(\rho) > 0$  may both depend on  $\rho$ , the function  $t \mapsto \mathbb{W}(\rho, t)$  is continuous and increasing on  $[0, \tau_1)$  such that  $\mathbb{W}(\rho, 0) = 0$ . Then for all  $\varepsilon > 0$  and for all  $\rho$  and  $\delta$  such that  $\delta^{-1} \mathbf{H}_e(\rho\delta)$  belongs to the range of  $\mathbb{W}(\rho, \cdot)$ , it holds

$$\begin{aligned} \mathbb{P}\left\{c_{\min} \geq \sqrt{\rho}(2 - \sqrt{\rho}) + \mathbb{W}^{-1}(\rho, \delta^{-1} \mathbf{H}_e(\rho\delta)) + \varepsilon\right\} &\leq c(\rho) e^{-nD(\rho, \delta, \varepsilon)}, \\ \mathbb{P}\left\{c_{\max} \geq \sqrt{\rho}(2 + \sqrt{\rho}) + \mathbb{W}^{-1}(\rho, \delta^{-1} \mathbf{H}_e(\rho\delta)) + \varepsilon\right\} &\leq c(\rho) e^{-nD(\rho, \delta, \varepsilon)}, \end{aligned}$$

where  $D(\rho, \delta, \varepsilon) > 0$ .

### 3.2. Key lemmas using singular values.

**Lemma 7.** *Assume that for all  $0 < \rho < 1$ , the largest singular value  $\sigma_1$  and the smallest singular value  $\sigma_s$  of a  $s \times n$  matrix (where  $s := \lfloor \rho n \rfloor$ ) with i.i.d. entries with respect to a law  $\mathcal{L}$ , satisfy for all  $n \geq n_0(\rho)$ ,*

$$\forall 0 < t < \sqrt{\rho_0}, \quad \mathbb{P}\left\{\left(\frac{\sigma_1}{\sqrt{n}} - (1 + \sqrt{\rho})\right) \vee \left((1 - \sqrt{\rho}) - \frac{\sigma_s}{\sqrt{n}}\right) \geq t\right\} \leq c(\rho) e^{-n\mathbb{W}(\rho,t)}$$

where  $n_0(\rho) \geq 2$  and  $c(\rho) > 0$  may both depend on  $\rho$ , the function  $t \mapsto \mathbb{W}(\rho, t)$  is continuous and increasing on  $[0, \sqrt{\rho_0})$  such that  $\mathbb{W}(\rho, 0) = 0$ . Then for any  $0 < \delta < 1$  and  $0 < \rho < \rho_0$  such that

$$\delta > \frac{1}{\rho} \exp\left[1 - \frac{\mathbb{W}\left[\rho, \frac{1}{2}(3 + \sqrt{2} - \sqrt{7 + 6\sqrt{2}}) - \sqrt{\rho}\right]}{\rho}\right],$$

any sequence of  $n \times p$  matrices  $(\mathbf{M}^{(n)})_{n \geq 2}$  with i.i.d. entries with respect to  $\mathcal{L}$  and such that  $n/p \rightarrow \delta$  satisfy

$$\mathbb{P}\left\{\frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ satisfies (FL-}\gamma(2s)\text{) with } 2s \leq \lfloor \rho n \rfloor\right\} \rightarrow 1$$

as  $n$  tends to infinity.

*Proof.* The proof follows the same lines as in Lemma 6. In this case, the conditioning event is  $\{c_{\max} < (1 + \sqrt{\rho} + t)^2 - 1\} \cup \{c_{\min} < 1 - (1 - \sqrt{\rho} - t)^2\}$ . Eq (11) becomes

$$-(3 + \sqrt{2} - \sqrt{7 + 6\sqrt{2}}) + 2\sqrt{\rho} + 2t < 0.$$

Indeed, observe that

$$\frac{1 + c_{\max}}{1 - c_{\min}} < \left(\frac{1 + \sqrt{\rho} + t}{1 - \sqrt{\rho} - t}\right)^2,$$

so that Condition (FL- $\gamma(2s)$ ) is implied by

$$\frac{1 + \sqrt{\rho} + t}{1 - \sqrt{\rho} - t} < \sqrt{4\sqrt{2} - 3}.$$

Eventually, notice that

$$\mathbb{P}\left\{\exists x \in \Sigma_s \text{ s.t. } n^{-1}\|\mathbf{M}x\|^2 \geq (1 + \sqrt{\rho} + t)^2\|x\|^2\right\} \leq \sum_{\mathbf{C}_{s,n}} \mathbb{P}\left\{n^{-1/2}\sigma_1 \geq 1 + \sqrt{\rho} + t\right\},$$

where the sum is over all choice of  $s$  columns over the  $p$  in  $\mathbf{M}$ .  $\square$

**Corollary 2.** *Assume that for all  $0 < \rho < 1$ , the largest singular value  $\sigma_1$  and the smallest singular value  $\sigma_s$  of a  $s \times n$  matrix (where  $s := \lfloor \rho n \rfloor$ ) with i.i.d. entries with respect to a law  $\mathcal{L}$ , satisfy for all  $n \geq n_0(\rho)$ ,*

$$\forall 0 < t < \tau_1, \quad \mathbb{P}\left\{\left(\frac{\sigma_1}{\sqrt{n}} - (1 + \sqrt{\rho})\right) \vee \left((1 - \sqrt{\rho}) - \frac{\sigma_s}{\sqrt{n}}\right) \geq t\right\} \leq c(\rho)e^{-n\mathbb{W}(\rho,t)}$$

where  $\tau_1 \in \overline{\mathbb{R}}$ ,  $n_0(\rho) \geq 2$  and  $c(\rho) > 0$  may both depend on  $\rho$ , the function  $t \mapsto \mathbb{W}(\rho, t)$  is continuous and increasing on  $[0, \tau_1]$  such that  $\mathbb{W}(\rho, 0) = 0$ . Then for all  $\varepsilon > 0$  and for all  $\rho$  and  $\delta$  such that  $\delta^{-1}\mathbf{H}_\varepsilon(\rho\delta)$  belongs to the range of  $\mathbb{W}(\rho, \cdot)$ , it holds

$$\mathbb{P}\left\{c_{\max} \geq (\sqrt{\rho} + t_0)(2 + \sqrt{\rho} + t_0) + \varepsilon\right\} \leq c(\rho)e^{-nD(\rho,\delta,\varepsilon)},$$

and if further  $t_0 + \sqrt{\rho} < 1$  then

$$\mathbb{P}\left\{c_{\min} \geq (\sqrt{\rho} + t_0)(2 - \sqrt{\rho} - t_0) + \varepsilon\right\} \leq c(\rho)e^{-nD(\rho,\delta,\varepsilon)},$$

where  $D(\rho, \delta, \varepsilon) > 0$  and  $t_0 := \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_\varepsilon(\rho\delta))$ .

**3.3. Davidson and Szarek's deviations.** Consider a  $s \times n$  matrix  $\mathbf{X}$  with i.i.d. standard Gaussian entries. In the paper [DS01], Davidson and Szarek have shown that for all  $0 < \rho < 1$  it holds

$$\forall t > 0, \quad \mathbb{P}\left\{\left(\frac{\sigma_1(\mathbf{X})}{\sqrt{n}} - (1 + \sqrt{\rho})\right) \vee \left((1 - \sqrt{\rho}) - \frac{\sigma_s(\mathbf{X})}{\sqrt{n}}\right) \geq t\right\} \leq 2e^{-n\mathbb{W}_{DS}(\rho,t)}$$

where  $\sigma_i(\mathbf{X})$  denotes the singular values of  $\mathbf{X}$  and  $\mathbb{W}_{DS}(\rho, t) := t^2/2$ , see [FR13, Page 291] for instance. This inequality relies on the concentration of measure phenomenon. Invoke Lemma 7 to get Theorem 5. This deviation has been used in the paper [CT05, Lemma 3.1] to bound the RIP constant.

**3.4. Ledoux and Rider's deviations.** Ledoux and Rider proved in [LR10] small deviation inequalities for  $\beta$  Hermite and Laguerre Ensembles. Their work rely on the tridiagonal model for these matrix ensembles and on a variational formulation of the Tracy-Widom distribution. For real covariance matrices, their deviation inequality for the largest eigenvalue is the following. For all  $0 < \rho < 1$  and for all  $n \geq 2$ , setting  $s = \lfloor \rho n \rfloor$ ,

$$\forall t > 0, \quad \mathbb{P}\left(\lambda_1 - (1 + \sqrt{\rho})^2 \geq t\right) \leq c(\rho)e^{-n\mathbb{W}_{LR}^{\max}(\rho,t)}$$

where  $\lambda_i$  denotes the eigenvalues of a  $s \times n$  covariance matrix  $\mathbf{C}$  with i.i.d. standard Gaussian entries and

$$\mathbb{W}_{LR}^{\max}(\rho, t) := \frac{\rho^{\frac{1}{4}}}{C_{LR}(1 + \sqrt{\rho})^3} t^{\frac{3}{2}} \mathbf{1}_{t \leq \sqrt{\rho}(1 + \sqrt{\rho})^2} + \frac{\rho^{\frac{1}{2}}}{C_{LR}(1 + \sqrt{\rho})^2} t \mathbf{1}_{t > \sqrt{\rho}(1 + \sqrt{\rho})^2},$$

where  $C_{LR} > 0$  may be bounded explicitly from [LR10]. This inequality is of the expected order in  $t$ , i.e.  $t^{3/2}$ , for  $t \leq \mathcal{O}(\sqrt{\rho})$  and of linear order in  $t$  for larger  $t$ . The authors discuss this change of behavior in [LR10] p.1322. As explained in Section 1.4, the dependency of function  $\mathbb{W}$  in parameter  $\rho$  is of crucial importance in our analysis. Therefore, we choose to write the most precise deviation inequalities the paper reached, even in the case when  $s/n$  is bounded. For  $\lambda_s$ , we follow the procedure explained in [LR10, Section 5, Page 1338] to write the following

$$\forall t > 0, \quad \mathbb{P}\left((1 - \sqrt{\rho})^2 - \lambda_s \geq t\right) \leq c(\rho)e^{-n\mathbb{W}_{LR}^{\min}(\rho, t)}$$

where

$$(15) \quad \mathbb{W}_{LR}^{\min}(\rho, t) := \frac{\rho^{\frac{1}{4}}}{C_{LR}(1 - \sqrt{\rho})^3} t^{\frac{3}{2}} 1_{t \leq \sqrt{\rho}(1 - \sqrt{\rho})^2} + \frac{\rho^{\frac{1}{2}}}{C_{LR}(1 - \sqrt{\rho})^2} t 1_{t > \sqrt{\rho}(1 - \sqrt{\rho})^2}.$$

In order to simplify the analysis of the phase transition, observe that  $\mathbb{W}_{LR}^{\max}(\rho, t) \leq \mathbb{W}_{LR}^{\min}(\rho, t)$  for all  $\rho$  and  $t$ . This yields

$$\forall t > 0, \quad \mathbb{P}\left\{\left(\lambda_1 - (1 + \sqrt{\rho})^2\right) \vee \left((1 - \sqrt{\rho})^2 - \lambda_s\right) \geq t\right\} \leq c(\rho)e^{-n\mathbb{W}_{LR}(\rho, t)}$$

where

$$(16) \quad \mathbb{W}_{LR}(\rho, t) := \mathbb{W}_{LR}^{\max}(\rho, t).$$

Invoke Lemma 6 to get Theorem 4. Note that, in Figure 2,  $c_{\min}$  is plotted in the top-right panel using (15) instead of (16) to get a more precise plot.

**3.5. Feldheim and Sodin's deviations.** For all  $0 < \rho < 1$  and for all  $n \geq n_0$ , setting  $s = \lfloor \rho n \rfloor$  it follows from [FS10] that

$$\forall t > 0, \quad \mathbb{P}\left\{\left(\lambda_1 - (1 + \sqrt{\rho})^2\right) \vee \left((1 - \sqrt{\rho})^2 - \lambda_s\right) \geq t\right\} \leq c(\rho)e^{-n\mathbb{W}_{FS}(\rho, t)}$$

where  $\lambda_i$  denotes the eigenvalues of a  $s \times n$  covariance matrix  $\mathbf{C}$  with i.i.d. Rademacher entries and

$$\mathbb{W}_{FS}(\rho, t) := \frac{\rho \log\left(1 + \frac{t}{2\sqrt{\rho}}\right)^{\frac{3}{2}}}{C_{FS}(1 + \sqrt{\rho})^2},$$

where  $0 < C_{FS} < 837$ , as shown in Proposition 8. Invoke Lemma 6 to get Theorem 4.

**3.6. Ideal deviations.** Considering the expected small deviation inequalities (7) due to the convergence to the Tracy-Widom distribution, and the expected sub-exponential behavior (9), it may be possible to prove the following for sub-Gaussian random matrices.

$$\forall t > 0, \quad \mathbb{P}\left\{\lambda_1 - (1 + \sqrt{\rho})^2 \geq t\right\} \leq c(\rho)e^{-n\mathbb{W}_{TW}(\rho, t)}$$

where  $\lambda_i$  denotes the eigenvalues of a  $s \times n$  covariance matrix  $\mathbf{C}$  with i.i.d. sub-Gaussian entries and

$$\mathbb{W}_{TW}(\rho, t) := \frac{\rho^{\frac{1}{4}}}{C_{TW}(1 + \sqrt{\rho})^2} t^{\frac{3}{2}} 1_{t \leq t(\rho)} + \frac{1}{C_{TW}(1 + \sqrt{\rho})^2} t 1_{t > t(\rho)},$$

where  $C_{TW} > 0$  and  $t(\rho) > 0$ . A similar deviation inequality may be established for the smallest eigenvalue  $\lambda_s$  with almost the same  $\mathbb{W}$  function. Indeed, using (8) instead of (7), the only change would be to replace  $(1 + \sqrt{\rho})^2$  by  $(1 - \sqrt{\rho})^2$ . Lemma 6 could then be invoked to get a theorem similar to Theorem 4. Ledoux and Rider's results combined to Davidson and Szarek's ones actually achieved this ideal lower bound on



the phase transition (except for a  $(1 + \sqrt{\rho})^3$  instead of  $(1 + \sqrt{\rho})^2$ , which is not of great importance for small  $\rho$ ) for Gaussian covariance matrices.

#### APPENDIX A. SMALL DEVIATIONS FOR THE RADEMACHER MODEL

In this section we follow the steps of the work [FS10] to get small deviation inequalities on the extreme eigenvalues of Gram matrices built from the Rademacher law. The paper [FS10] focuses on the asymptotic distribution of the fluctuations of the extreme eigenvalues, and it proved that the extreme eigenvalues of the sample covariance matrices built from sub-Gaussian matrices asymptotically fluctuate around their limiting values (with proper scaling) with respect to the Tracy-Widom distribution. Their results follow from an interesting estimation of the moments of the fluctuations. While their estimation is interestingly of the right order (namely  $\varepsilon^{3/2}$ ), the authors of [FS10] did not pursue on giving an upper bound of the constant appearing in their rate function, see Claim (a) and (b) of Point 2 in [FS10, Corollary V.2.1].

Unfortunately, the constant  $C_{FS}$  appearing in the rate function is of crucial importance when deriving phase transitions, see Section 2 for instance. Hence, we need to track the proof of [FS10] in order to provide an upper bound on  $C_{FS}$  and its dependence on the ratio  $\rho$  of the sizes of the Rademacher matrix. This strenuous hunt necessitates to recast all the asymptotic bounds appearing in [FS10] into non asymptotic ones as sharp as possible. The benefit of this elementary but non trivial task is two fold. First it gives, for the first time, an explicit expression of small deviations of extreme eigenvalues of the sample covariance matrices at the sharp rate  $\varepsilon^{3/2}$ . This section is devoted to prove the following result.

**Proposition 8.** *Let  $N > M \geq 54$  and consider*

$$\mathbf{C} := \mathbf{X}\mathbf{X}^\top \quad \text{where } \mathbf{X} \in \{\pm 1\}^{M \times N} \text{ with i.i.d. Rademacher entries}$$

then

$$\begin{aligned} \mathbb{P}\left\{\lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N\right\} &\leq \frac{\mathbb{W}_0(\rho, \varepsilon)}{1 - \rho} M \exp(-N \mathbb{W}_{FS}(\rho, \varepsilon)) \\ \mathbb{P}\left\{\lambda_1(\mathbf{C}) \leq (\sqrt{M} - \sqrt{N})^2 - \varepsilon N\right\} &\leq \frac{\mathbb{W}_0(\rho, \varepsilon)}{1 - \rho} M \exp(-N \mathbb{W}_{FS}(\rho, \varepsilon)) \end{aligned}$$

where  $\rho = M/N$  and

$$\begin{aligned} \mathbb{W}_0(\rho, \varepsilon) &:= c_0 \exp\left[c_0 \sqrt{\log\left(1 + \frac{\varepsilon}{2\sqrt{\rho}}\right)}\right] \\ \mathbb{W}_{FS}(\rho, \varepsilon) &:= \frac{\rho \log\left(1 + \frac{\varepsilon}{2\sqrt{\rho}}\right)^{\frac{3}{2}}}{C_{FS}(1 + \sqrt{\rho})^2} \end{aligned}$$

for some universal constants  $c_0 > 0$  and  $837 > C_{FS} > 0$ . Furthermore, for any  $C > 3242$  there exists a constant  $\nu := \nu(\rho, C) > 0$  that depends only on  $\rho = M/N$  and  $C$  such that, for all  $0 < \varepsilon < \sqrt{\rho}$ ,

$$\begin{aligned} \mathbb{P}\left\{\lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N\right\} &\leq \nu \exp\left(-C^{-1} N \frac{\rho^{1/4}}{(1 + \sqrt{\rho})^2} \varepsilon^{\frac{3}{2}}\right) \\ \mathbb{P}\left\{\lambda_1(\mathbf{C}) \leq (\sqrt{M} - \sqrt{N})^2 - \varepsilon N\right\} &\leq \nu \exp\left(-C^{-1} N \frac{\rho^{1/4}}{(1 + \sqrt{\rho})^2} \varepsilon^{\frac{3}{2}}\right). \end{aligned}$$

**A.1. Sketch of the proof.** The result of [FS10] is based on a combinatorial proof. Interestingly, this approach is suited for the Rademacher model since, in this case, traces of polynomials of the covariance matrix  $\mathbf{C}$  can be expressed as the number of non-backtracking paths of given length. In this section, we change notation and we use the notation of the paper [FS10] to ease readability when referring to this latter. Hence, we consider a Rademacher matrix of size  $M \times N$  with  $M < N$  (referred to as  $s \times n$  with  $s < n$  in the rest of this paper). We draw this proof into the following points.

- (1) The proof [FS10] is based on a moment method that captures the influence of the largest and the smallest eigenvalues considering a new centering

$$\tilde{\mathbf{C}} := \frac{\mathbf{C} - (M + N - 2)}{2\sqrt{(M - 1)(N - 1)}}.$$

The authors [FS10] then use the trace of  $\tilde{\mathbf{C}}^{2m} + \tilde{\mathbf{C}}^{2m-1}$  (resp.  $\tilde{\mathbf{C}}^{2m} - \tilde{\mathbf{C}}^{2m-1}$ ) to estimate the moments of the largest (resp. smallest) eigenvalue.

- (2) The control of

$$A_m := \mathbb{E}\text{tr}[\tilde{\mathbf{C}}^{2m}] + \mathbb{E}\text{tr}[\tilde{\mathbf{C}}^{2m-1}]$$

(resp.  $B_m := \mathbb{E}\text{tr}[\tilde{\mathbf{C}}^{2m}] - \mathbb{E}\text{tr}[\tilde{\mathbf{C}}^{2m-1}]$ ) is given by a control of traces of polynomials  $Q_n(\mathbf{C})$  of  $\mathbf{C}$ . Up to a proper scaling, these polynomials are the orthogonal polynomials of the Marchenko-Pastur law which can be expressed by Chebyshev polynomials  $U_n$  of the second kind.

- (3) In the Rademacher model, the aforementioned traces, namely  $\mathbb{E}\text{tr}[Q_n(\mathbf{C})]$ , are exactly the number  $\hat{\Sigma}_1^1(n)$  of non-backtracking paths on the complete bi-partite graph that cross an even number of times each edge and end at their starting vertex. This claim can be generalized to general random sub-Gaussian matrices, up to technicalities.
- (4) To estimate the number of non-backtracking paths  $\hat{\Sigma}_1^1(n)$ , the article [FS10] begins with a mapping from the collection of non-backtracking paths into the collection of weighted diagrams. Then it provides an automaton which constructs all possible diagrams. The number of diagrams constructed by the automaton ending in  $s$  steps is denoted  $D_1(s)$ . Lemma 9 provides an upper bound on this quantity. Summing over  $s$ , it yields an upper bound on  $\hat{\Sigma}_1^1(n)$ , see (17) in Lemma 10.
- (5) In the Rademacher model,  $\hat{\Sigma}_1^1(n)$  is the expectation of the trace of  $Q_n$ . Hence, we deduce an upper bound on these traces.
- (6) Using Markov inequality and optimizing over the parameters, we deduce small deviation inequalities on the smallest and largest eigenvalues.

**A.2. Number of diagrams.** Recall that  $D_1(s)$  denotes the number of diagrams constructed by the automaton ending in  $s$  steps. The description of the automaton can be found in [FS10] Section II.2 page 101.

**Lemma 9.** *It holds, for all  $s \geq 1$ ,*

$$D_1(s) \leq C_{0,D} C_D^{s-1} s^{s-1/2}$$

where  $C_{0,D}$  and  $C_D$  can be chosen as  $C_{0,D} = 8.31$  and  $C_D = 53.8$ .

*Proof.* We follow Proposition II.2.3 of [FS10] but we focus on the case (of sample covariance matrices) corresponding to  $\beta = 1$ . In this case, there are three types of transitions

from one state to the following one. Let  $s = 2g + h$  be the number of steps in the automaton at the end, where  $h$  is the number of transition of type 3 and  $g$  the number of transition of type 1.

- If  $h = 0$  then the number of ways to order the transitions of the type 1 and 2 is exactly  $\frac{(2g)!}{g!(g+1)!}$ . Informally, the state of the automaton can be seen as a “thread” made of straight pieces and loops. The total length of this thread changes at each step. These changes of length are encoded by non-negative integers  $m_i$ . For precise definition of these numbers, see [FS10] Section II.2 page 103. In the present case, the number of ways to choose the numbers  $m_i$  is at most  $\binom{6g-1}{4g}$ . The number of diagrams corresponding to a fixed order of transitions and fixed  $m_i$  is at most  $(6g-1)^{2g}$  (indeed, the following state is then determined by choosing an edge and there are  $6g-1$  edges in the diagram). As in [FS10], we deduce that an upper on  $D_1$  is

$$\frac{(2g)!}{g!(g+1)!} \binom{6g-1}{4g} (6g-1)^{2g} = \frac{2g(6g-1)!(6g-1)^{2g}}{g!(g+1)!(4g)!}.$$

Using Lemma 12, this number is upper bounded by

$$\frac{e^{2+1/60}}{\pi} g \frac{(6g-1)^{8g-1/2}}{g^{g+1/2}(g+1)^{g+3/2}(4g-1)^{4g-1/2}}.$$

Writing  $\theta = \frac{(6g-1)^{8g-1/2}}{g^{g+1/2}(g+1)^{g+3/2}(4g-1)^{4g-1/2}}$  in exponential form, we get

$$\theta = \exp\left(2g \log g + g(8 \log(6) - 4 \log(4)) - 3 \log g - \frac{1}{2}(3 \log 2 + \log 3) + \gamma(g)\right),$$

with

$$\gamma(g) = \left(8g - \frac{1}{2}\right) \log\left(1 - \frac{1}{6g}\right) - \left(g + \frac{3}{2}\right) \log\left(1 + \frac{1}{g}\right).$$

Note that  $\gamma$  is non decreasing on  $(1, \infty)$  and goes to  $-\frac{7}{3}$  when  $g \rightarrow \infty$ . Therefore, the number of diagrams in this case is upper bounded by (recall that  $s = 2g$  here)

$$\frac{1}{\pi} e^{7/2 \log(3) - 3/2 \log(2) + 1/60 - 1/3} (40.5)^{s-1} s^{s-2} \leq 3.84 (40.5)^{s-1} s^{s-2}.$$

- If  $g = 0$  then there are only transitions of the third kind. The number of ways to choose the numbers  $m_i$  is at most  $\binom{2h-1}{h-1}$ . The number of diagrams corresponding to a fixed order of transitions and fixed  $m_i$  is at most  $(3h-1)^h$  (indeed, recall that the number of edges of the diagram is  $3h-1$ ). We deduce that an upper on  $D_1$  is

$$\frac{(2h-1)!}{h!(h-1)!} (3h-1)^h.$$

Note that this number is 2 when  $h = 1$ . For  $h \geq 2$ , using Lemma 12, this number is upper bounded by

$$\frac{e^{1/12}}{\sqrt{2\pi}} \frac{(2h-1)^{2h-1/2} (3h-1)^h}{h^{h+1/2} (h-1)^{h-1/2}}.$$

Once again, we write  $\theta = \frac{(2h-1)^{2h-1/2} (3h-1)^h}{h^{h+1/2} (h-1)^{h-1/2}}$  in exponential form. This yields

$$\theta = \exp\left[\left(h - \frac{1}{2}\right) \log h + (2 \log(2) + \log(3))(h-1) + \frac{3}{2} \log 2 + \log(3) + \gamma(h)\right],$$

with

$$\gamma(h) = \left(2h - \frac{1}{2}\right) \log\left(1 - \frac{1}{2h}\right) + h \log\left(1 - \frac{1}{3h}\right) - \left(h - \frac{1}{2}\right) \log\left(1 - \frac{1}{h}\right).$$

Note that  $\gamma$  is non increasing on  $(2, h^*)$  and non decreasing on  $(h^*, \infty)$  for some  $h^* > 2$ . Therefore,  $\gamma(h)$  is bounded by  $\max(\gamma(2), \lim_{h \rightarrow \infty} \gamma(h))$ . This yields  $\gamma(h) \leq -0.33$  for all  $h \geq 2$ . Finally, the number of diagrams in this case is upper bounded by (recall that  $s = h$  here)

$$\frac{e^{1/12}}{\sqrt{2\pi}} e^{3/2 \log(2) + \log(3) - 0.33} (12)^{s-1} s^{s-1/2} \leq 2.65 (12)^{s-1} s^{s-1/2}.$$

• If  $h \neq 0$  and  $g \neq 0$  then the number of ways to order the transitions of the three types is exactly

$$\binom{2g+h}{h} \frac{(2g)!}{g!(g+1)!} = \frac{(2g+h)!}{h!g!(g+1)!}.$$

The number of ways to choose the numbers  $m_i$  is at most  $\binom{6g+2h-1}{2g+h-1}$ . The number of diagrams corresponding to a fixed order of transitions and fixed  $m_i$  is at most  $(6g+3h-1)^{2g+h}$  (indeed, recall that the number of edges of the diagram is  $6g+3h-1$ ).

We deduce that an upper bound on  $D_1$  is

$$\frac{(2g+h)!}{h!g!(g+1)!} \binom{6g+2h-1}{2g+h-1} (6g+3h-1)^{2g+h}.$$

Using the fact that  $s = 2g+h$  and Lemma 12, this number is bounded by

$$\frac{e^{131/126}}{(2\pi)^{3/2}} \frac{s^{s+1/2} (3s-h-1)^{3s-h-1/2} (3s-1)^s}{h^{h+1/2} g^{g+1/2} (g+1)^{g+3/2} (s-1)^{s-1/2} (2s-h)^{2s-h+1/2}}.$$

Let  $t = h/s \in [1/s, 1-2/s]$  so that an upper bound is

$$\frac{e^{131/126}}{(2\pi)^{3/2}} \frac{s^{s+1/2} (3s-ts-1)^{3s-ts-1/2} (3s-1)^s}{(ts)^{ts+1/2} (s-\frac{1-t}{2})^{s(1-t)/2+1/2} (s\frac{1-t}{2}+1)^{s(1-t)/2+3/2} (s-1)^{s-1/2} (2s-ts)^{2s-ts+1/2}}.$$

Once again, we write this in exponential form and get

$$\frac{e^{131/126}}{(2\pi)^{3/2}} \exp\left(s \log s - \frac{5}{2} \log s + \beta(t)s + \alpha(t) + \gamma(s, t)\right),$$

with

$$\begin{aligned} \alpha(t) &= 2 \log 2 - \frac{1}{2} \log(3-t) - \frac{1}{2} \log t - 2 \log(1-t) - \frac{1}{2} \log(2-t), \\ \beta(t) &= (3-t) \log(3-t) + \log(3) \\ &\quad - t \log t - (1-t) \log(1-t) + \log(2)(1-t) - (2-t) \log(2-t), \\ \gamma(s, t) &= \left((3-t)s - \frac{1}{2}\right) \log\left(1 - \frac{1}{(3-t)s}\right) + s \log\left(1 - \frac{1}{3s}\right) \\ &\quad - \frac{1}{2} \left(s(1-t) + 3\right) \log\left(1 + \frac{2}{s(1-t)}\right) - \left(s - \frac{1}{2}\right) \log\left(1 - \frac{1}{s}\right). \end{aligned}$$

◦ We focus first on  $\beta$ . This function is non decreasing on  $(0, t^*)$  and non increasing on  $(t^*, 1)$ , with  $t^* = \frac{3}{2} - \frac{\sqrt{57}}{6} \approx 0.24$ . Therefore, it reaches its maximum at  $t^*$ . Computing it yields  $\beta(t) \leq 3.985$  for all  $t \in (0, 1)$ .

◦ We focus now on  $\alpha$ . This function is non increasing on  $(0, t')$  and non decreasing on  $(t', 1)$  with  $t' \in (0, 1)$ . Recall that  $t \in (1/s, 1-2/s)$ . Therefore,  $\alpha(t) \leq \max(\alpha(1/s), \alpha(1-2/s))$ . Computing these two values and using the fact that  $s \geq 3$  leads to  $\alpha(t) \leq \alpha(1-2/s)$  for all  $t \in (1/s, 1-2/s)$ . Consequently

$$\alpha(t) \leq 2 \log s - \frac{1}{2} \log\left(2 + \frac{2}{s}\right) - \frac{1}{2} \log\left(1 - \frac{2}{s}\right) - \frac{1}{2} \log\left(1 + \frac{2}{s}\right).$$

◦ Let's turn to  $\gamma$ . Recall that  $t \in (1/s, 1-2/s)$ . Dealing separately with the two terms  $((3-t)s - \frac{1}{2}) \log(1 - \frac{1}{(3-t)s})$  and  $\frac{1}{2}(s(1-t) + 3) \log(1 + \frac{2}{s(1-t)})$  yields

$$\begin{aligned} \gamma \leq & \left(3s - \frac{3}{2}\right) \log\left(1 - \frac{1}{3s-1}\right) + s \log\left(1 - \frac{1}{3s}\right) \\ & - \frac{1}{2}(s+2) \log\left(1 + \frac{2}{s-1}\right) - \left(s - \frac{1}{2}\right) \log\left(1 - \frac{1}{s}\right). \end{aligned}$$

Going back to the number of diagrams in this case, it is bounded by

$$\frac{e^{131/126}}{(2\pi)^{3/2}} \exp\left(s \log s - \frac{1}{2} \log s + 3.985s + \delta(s)\right),$$

with

$$\begin{aligned} \delta(s) = & -\frac{1}{2} \log\left(2 + \frac{2}{s}\right) - \frac{1}{2} \log\left(1 - \frac{2}{s}\right) - \frac{1}{2} \log\left(1 + \frac{2}{s}\right) \\ & + \left(3s - \frac{3}{2}\right) \log\left(1 - \frac{1}{3s-1}\right) + s \log\left(1 - \frac{1}{3s}\right) \\ & - \frac{1}{2}(s+2) \log\left(1 + \frac{2}{s-1}\right) - \left(s - \frac{1}{2}\right) \log\left(1 - \frac{1}{s}\right). \end{aligned}$$

This function is non decreasing on  $(3, \infty)$  and goes to  $-\frac{4}{3} - \frac{\log 2}{2} \leq -1.67$  when  $s$  goes to  $\infty$ . Therefore, there are at most

$$\frac{e^{131/126-1.67}}{(2\pi)^{3/2}} (e^{3.985})^{s-1} s^{s-1/2} \leq 1.82(53.8)^{s-1} s^{s-1/2}$$

diagrams in this case. This leads to the result.  $\square$

**A.3. Number of paths.** Let  $n \geq 1$  be fixed. Recall that  $\mathbb{E} \text{tr}[Q_n(\mathbf{C})]$  is equal to the number  $\hat{\Sigma}_1^1(n)$  of non-backtracking paths, see Page 115 in [FS10]. Recall that  $M \leq N$  denotes the sizes of the Rademacher matrix.

**Lemma 10.** *It holds*

$$(17) \quad \hat{\Sigma}_1^1(n) \leq C_{0, \hat{\Sigma}} (MN)^{n/2} \exp\left[\frac{C_{\hat{\Sigma}}(1 + \sqrt{M/N})n^{3/2}}{\sqrt{M}}\right]$$

where  $C_{0, \hat{\Sigma}} = 160.4$  and  $C_{\hat{\Sigma}} = 13.3$ . As a consequence,

$$\mathbb{E}[\text{tr}[Q]_n(\mathbf{C})] \leq C_{0, \hat{\Sigma}} (MN)^{n/2} n \exp\left(C_{\hat{\Sigma}}(1 + \sqrt{M/N}) \frac{n^{3/2}}{M^{1/2}}\right).$$

*Proof.* The number of diagrams is  $D_1(s)$  for  $1 \leq s \leq n$ . The number of ways to choose the vertices on a diagram constructed in  $s$  steps by the automaton is at most

$$\frac{1}{2}(MN)^{n/2} \left[ (1 + \sqrt{M/N})(M^{-1/2} + N^{-1/2})^{2s-2} + (1 - \sqrt{M/N})(M^{-1/2} - N^{-1/2})^{2s-2} \right],$$

see [FS10, Page 117]. The number of ways to choose the weights on a diagram constructed in  $s$  steps by the automaton is at most

$$\frac{(3s+1)}{(3s-2)!} \left( \frac{n-3s+1}{2} + 3s-2 \right)^{3s-2}.$$

We deduce that the number  $\hat{\Sigma}_1^1(n)$  of non-backtracking paths is at most

$$\hat{\Sigma}_1^1(n) \leq \frac{1}{2}(MN)^{n/2} \left[ \left(1 + \sqrt{\frac{M}{N}}\right) T_1 + \left(1 - \sqrt{\frac{M}{N}}\right) T_2 \right]$$

where

$$T_1 := \sum_{s=1}^n D_1(s)(M^{-1/2} + N^{-1/2})^{2s-2} \frac{(3s+1)}{(3s-2)!} \left( \frac{n-3s+1}{2} + 3s-2 \right)^{3s-2}$$

$$T_2 := \sum_{s=1}^n D_1(s)(M^{-1/2} - N^{-1/2})^{2s-2} \frac{(3s+1)}{(3s-2)!} \left( \frac{n-3s+1}{2} + 3s-2 \right)^{3s-2}$$

We can bound each term. It reads as follows.

$$T_1 \leq C_{0,D} \sum_{s=1}^n C_D^{s-1} s^{s-1/2} (M^{-1/2} + N^{-1/2})^{2s-2} \frac{(3s+1)}{(3s-2)!} \left( \frac{n-3s+1}{2} + 3s-2 \right)^{3s-2}$$

$$\leq C_{0,D} \sum_{s=1}^n C_D^{s-1} \left[ \frac{1 + \sqrt{M/N}}{\sqrt{M}} \right]^{2(s-1)} \frac{(3s+1)(n+3s-3)^{3s-2} s^{s-1/2}}{(3s-2)! 2^{3s-2}}$$

using Lemma 9. Invoke Lemma 12 to get that

$$\frac{(2(s-1))! (3s+1)(n+3s-3)^{3s-2} s^{s-1/2}}{n^{3(s-1)} (3s-2)! 2^{3s-2}}$$

$$\leq n \frac{e^{s+1/12}}{2^{3s-2}} (3s+1) \sqrt{\frac{2s-2}{3s-2}} \left( 1 + \frac{3s-3}{n} \right)^{3s-2} \frac{(2s-2)^{2s-2} s^{s-1/2}}{(3s-2)^{3s-2}}$$

$$\leq ne^{1/12} \sqrt{\frac{2s-2}{3s-2}} 2^{3s-2} (3s+1) \frac{(2s-2)^{2s-2} s^{s-1/2}}{(3s-2)^{3s-2}} e^s.$$

But  $2^{3s-2} (3s+1) \frac{(2s-2)^{2s-2} s^{s-1/2}}{(3s-2)^{3s-2}} e^s \leq \exp(s + f(s))$  where

$$f(s) = (3s-2)\log(2) + (2s-2)\log(2s-2) + \log(3s+1) + \left(s - \frac{1}{2}\right)\log(s) - (3s-2)\log(3s-2).$$

Some elementary computations give the following:

$$f(s) = \frac{1}{2} \log s + (5 \log 2 - 3 \log 3)s + 3 \log 3 - 4 \log 2 + (2s-2) \log \left( 1 - \frac{1}{s} \right)$$

$$- (3s-2) \log \left( 1 - \frac{2}{3s} \right) + \log \left( 1 + \frac{1}{3s} \right)$$

$$= (5 \log 2 - 3 \log 3)s + 3 \log 3 - 4 \log 2 + g(s),$$

with  $g(s) = \frac{1}{2} \log s + (2s-2) \log \left( 1 - \frac{1}{s} \right) - (3s-2) \log \left( 1 - \frac{2}{3s} \right) + \log \left( 1 + \frac{1}{3s} \right)$ . We have

$$g'(s) = \frac{3s-1}{2s(3s+1)} + 2 \log \left( 1 - \frac{1}{s} \right) - 3 \log \left( 1 - \frac{2}{3s} \right),$$

$$g''(s) = \frac{-27s^4 + 99s^3 - 21s^2 + 11s + 2}{2s^2(s-1)(3s-2)(3s+1)^2}.$$

It may be shown that there exists  $s_* > 2$  such that  $g''$  is positive on  $(1, s_*)$  and negative on  $(s_*, \infty)$ . Therefore,  $g$  is strictly concave on  $(s_*, \infty)$  and its curve is below its tangents, which write  $y = g'(s_0)(s - s_0) + f(s_0)$ . For  $s \in [1, s_*]$ ,  $g(s) \leq g(1) = 2 \log 2$ . As a consequence, we are looking for the point  $s_0 \in (s_*, \infty)$  such that the tangent at  $s_0$  goes through the point  $(1, 2 \log 2)$ . This tangent goes through the point  $(1, g(s_0) + (1-s_0)g'(s_0))$ . Set  $h(s) = g(s) + (1-s)g'(s)$ . This function is non decreasing and there is a unique point  $s_0 \in (s_*, \infty)$  such that  $h(s_0) = 2 \log 2$ . It may be shown that  $s_0 \in (39.66; 39.67)$ . As  $g'$  is non increasing on this interval,  $g'(s_0) \leq g'(39.66) \leq 0.013$ . This leads to

$$g(s) \leq 0.013(s-1) + 2 \log 2.$$

Then

$$\begin{aligned} & \frac{(2(s-1))! (3s+1)(n+3s-3)^{3s-2} s^{s-1/2}}{n^{3(s-1)} (3s-2)! 2^{3s-2}} \\ & \leq ne^{1/12} \sqrt{\frac{2}{3}} \exp\left((5\log 2 - 3\log 3 + 1.013)s + 3\log 3 - 2\log 2 - 0.013\right) \\ & \leq ne^{1/12} \sqrt{\frac{2}{3}} \exp(1 + 3\log 2) \exp\left((5\log 2 - 3\log 3 + 1.013)(s-1)\right) \\ & \leq 19.3 n (3.27)^{s-1}. \end{aligned}$$

As a consequence,

$$\begin{aligned} T_1 & \leq 19.3 C_{0,D} n \sum_{s=1}^n \frac{1}{(2(s-1))!} \left[ \frac{1.81(1 + \sqrt{M/N})n^{3/2}}{\sqrt{C_D^{-1}M}} \right]^{2(s-1)} \\ & \leq 19.3 C_{0,D} n \exp\left(1.81\sqrt{C_D}(1 + \sqrt{M/N})\frac{n^{3/2}}{M^{1/2}}\right). \end{aligned}$$

Similarly, one gets

$$\begin{aligned} T_2 & \leq 19.3n C_{0,D} \sum_{s=1}^n \frac{1}{(2(s-1))!} \left[ \frac{1.81(1 - \sqrt{M/N})n^{3/2}}{\sqrt{C_D^{-1}M}} \right]^{2(s-1)} \\ & \leq 19.3n C_{0,D} \exp\left[\frac{1.81(1 - \sqrt{M/N})\sqrt{C_D}n^{3/2}}{\sqrt{M}}\right]. \end{aligned}$$

This yields the result.  $\square$

#### A.4. Bound on the traces.

**Lemma 11.** *It holds that*

$$(18) \quad \left(\mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m}]] + \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m-1}]]\right) \vee \left(\mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m}]] - \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m-1}]]\right) \leq \Delta_m$$

where

$$\Delta_m = \frac{C_{0,\text{Rad}}}{1 - \frac{M}{N}} m \left[ \left(\frac{MN}{(M-1)(N-1)}\right)^m + \frac{M}{m} \right] \exp\left(C_{\text{Rad}}(1 + \sqrt{M/N})^4 \frac{m^3}{M^2}\right),$$

and

$$\begin{aligned} C_{0,\text{Rad}} & = 594 C_{0,\hat{\Sigma}} = 95,278 \\ C_{\text{Rad}} & = 355.7 C_D^2 = 830,415. \end{aligned}$$

*Proof.* Invoke Lemma IV.1.1 Page 115 in [FS10] and Lemma 10 to get that

$$(19) \quad \mathbb{E}[\text{tr}[V_{n, \frac{(M-2)^2}{(M-1)(N-1)}}(\tilde{\mathbf{C}})]] \leq C_{0,\hat{\Sigma}} \left(\frac{MN}{(M-1)(N-1)}\right)^{n/2} n \exp\left(C_{\hat{\Sigma}}(1 + \sqrt{M/N})\frac{n^{3/2}}{M^{1/2}}\right).$$

Set  $s := \frac{(M-2)^2}{(M-1)(N-1)}$ . For  $m \geq 1$ , let  $A_m = \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m}]] + \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m-1}]]$ . Following Pages 95-96 in [FS10] yields:

$$\begin{aligned} (20) \quad A_m & = \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^m (2n+1) \binom{2m+1}{m-n} \mathbb{E}[\text{tr}[U_{2n}(\tilde{\mathbf{C}})]] \\ & + \frac{1}{2m2^{2m}} \sum_{n=1}^m 2n \binom{2m}{m-n} \mathbb{E}[\text{tr}[U_{2n-1}(\tilde{\mathbf{C}})]]. \end{aligned}$$

Using the fact that  $V_{k,s} = U_k + \sqrt{s}U_{k-1}$ , it holds

$$\begin{aligned} A_m &= \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^m (2n+1) \binom{2m+1}{m-n} \sum_{k=0}^{2n} (-1)^k s^{k/2} \mathbb{E}[\text{tr}[V_{2n-k,s}(\tilde{\mathbf{C}})]] \\ &\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^m 2n \binom{2m}{m-n} \sum_{k=0}^{2n-1} (-1)^k s^{k/2} \mathbb{E}[\text{tr}[V_{2n-k-1,s}(\tilde{\mathbf{C}})]]. \end{aligned}$$

Note that the expectation  $\mathbb{E}[\text{tr}[V_{k,s}(\tilde{\mathbf{C}})]]$  is non-negative. Indeed, one can check that  $\mathbb{E}[\text{tr}[V_{k,s}(\tilde{\mathbf{C}})]] = \mathbb{E}[\text{tr}[Q_k(\tilde{\mathbf{C}})]] = \hat{\Sigma}_1^1(k)$  up to a multiplicative positive constant. It follows that

$$\begin{aligned} A_m &\leq \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^m (2n+1) \binom{2m+1}{m-n} \sum_{k=0}^n s^k \mathbb{E}[\text{tr}[V_{2(n-k),s}(\tilde{\mathbf{C}})]] \\ &\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^m 2n \binom{2m}{m-n} \sum_{k=0}^{n-1} s^k \mathbb{E}[\text{tr}[V_{2n-2k-1,s}(\tilde{\mathbf{C}})]] \\ &\leq \frac{1}{(2m+1)2^{2m}} \sum_{n=1}^m (2n+1) \binom{2m+1}{m-n} \left( \sum_{k=0}^{n-1} s^k \mathbb{E}[\text{tr}[V_{2(n-k),s}(\tilde{\mathbf{C}})]] + s^n M \right) \\ (21) \quad &\quad + \frac{1}{2m2^{2m}} \sum_{n=0}^m 2n \binom{2m}{m-n} \sum_{k=0}^{n-1} s^k \mathbb{E}[\text{tr}[V_{2n-2k-1,s}(\tilde{\mathbf{C}})]] \end{aligned}$$

Invoke (19) to get with  $C_{M,N} = C_{\hat{\Sigma}}(1 + \sqrt{M/N})$ ,

$$\begin{aligned} A_m &\leq \sum_{n=1}^m \frac{2n+1}{(2m+1)2^{2m}} \binom{2m+1}{m-n} \\ &\quad \times \sum_{k=0}^{n-1} s^k C_{0,\hat{\Sigma}} \left( \frac{MN}{(M-1)(N-1)} \right)^{n-k} 2(n-k) \exp \left[ C_{M,N} \frac{2^{\frac{3}{2}}(n-k)^{\frac{3}{2}}}{M^{\frac{1}{2}}} \right] \\ &\quad + \sum_{n=1}^m \frac{n}{m2^{2m}} \binom{2m}{m-n} \\ &\quad \times \sum_{k=0}^{n-1} s^k C_{0,\hat{\Sigma}} \left( \frac{MN}{(M-1)(N-1)} \right)^{n-k-\frac{1}{2}} 2 \left( n-k-\frac{1}{2} \right) \exp \left[ C_{M,N} \frac{2^{\frac{3}{2}}(n-k-1/2)^{\frac{3}{2}}}{M^{\frac{1}{2}}} \right] \\ &\quad + \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^m (2n+1) \binom{2m+1}{m-n} s^n M \\ &\leq \frac{2C_{0,\hat{\Sigma}}}{1 - \frac{(M-2)^2}{MN}} \sum_{n=1}^m \left[ \frac{2n+1}{(2m+1)2^{2m}} \binom{2m+1}{m-n} + \frac{n}{m2^{2m-1}} \binom{2m}{m-n} \right] \\ &\quad \times n \left( \frac{MN}{(M-1)(N-1)} \right)^n \exp \left[ C_{M,N} \frac{2^{\frac{3}{2}} n^{\frac{3}{2}}}{M^{\frac{1}{2}}} \right] \\ &\quad + \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^m (2n+1) \binom{2m+1}{m-n} s^n M. \end{aligned}$$

From Lemma 13 it holds

$$\log \left[ \frac{n+1/2}{2^{2m}} \binom{2m+1}{m-n} \right] \vee \log \left[ \frac{n}{2^{2m}} \binom{2m}{m-n} \right] \leq -c_1 - c_2 \frac{n^2}{m}$$



where  $c_1 = -5$  and  $c_2 = 0.6321$ . We deduce that

$$\begin{aligned} A_m &\leq \frac{4C_{0,\hat{\Sigma}}}{1 - \frac{(M-1)(N-1)s}{MN}} \frac{\exp(-c_1)}{m} \sum_{n=1}^m n \left( \frac{MN}{(M-1)(N-1)} \right)^n \exp\left(-c_2 \frac{n^2}{m} + C_{M,N} \frac{2^{3/2} n^{3/2}}{M^{1/2}}\right) \\ &\quad + \frac{M \exp(-c_1)}{m} \sum_{n=0}^m s^n \exp\left(-c_2 \frac{n^2}{m}\right), \\ &\leq \frac{4C_{0,\hat{\Sigma}} \exp(-c_1)}{1 - \frac{M}{N}} \left[ \left( \frac{MN}{(M-1)(N-1)} \right)^m + \frac{M}{m} \right] \sum_{n=1}^m \exp\left(-c_2 \frac{n^2}{m} + C_{M,N} \frac{2^{3/2} n^{3/2}}{M^{1/2}}\right). \end{aligned}$$

Observe that the maximum of  $-ax^4 + bx^3$  is  $\frac{27b^4}{256a^3}$ . We deduce that

$$-c_2 \frac{n^2}{m} + C_{M,N} \frac{2^{3/2} n^{3/2}}{M^{1/2}} \leq C_{\text{Rad}} (1 + \sqrt{M/N})^4 \frac{m^3}{M^2}$$

where

$$C_{\text{Rad}} = \frac{27}{4} \frac{C_{M,N}^4}{c_2^3 (1 + \sqrt{M/N})^4} = \frac{27}{4} \frac{C_{\hat{\Sigma}}^4}{c_2^3} = \frac{27}{4} \frac{1.81^4 C_D^2}{c_2^3} = 286.9 C_D^2,$$

as claimed.

The bound on  $B_m := \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m}]] - \mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m-1}]]$  follows the same lines. The minus in front of  $\mathbb{E}[\text{tr}[\tilde{\mathbf{C}}^{2m-1}]]$  change the line (20) to its opposite. The change of indices  $k$  leads to the term  $s^{k+1/2} \mathbb{E}[\text{tr}[V_{2(n-k-1),s}(\tilde{\mathbf{C}})]]$  in (21). Since we uniformly bound  $n-k-1$  by  $n$  in the rest of the proof and  $s^{1/2} < 1$ , we get the same result.  $\square$

**A.5. Small deviation on the largest eigenvalue.** Observe that

$$\mathbb{P}\{\lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N\} = \mathbb{P}\{\lambda_M(\tilde{\mathbf{C}}) \geq \varepsilon_{M,N}\},$$

with

$$\frac{55}{53} \left(1 + \frac{\varepsilon}{2\sqrt{M/N}}\right) \geq \varepsilon_{M,N} := \frac{\sqrt{MN} + 1}{\sqrt{(M-1)(N-1)}} + \frac{\varepsilon N}{2\sqrt{(M-1)(N-1)}} \geq 1 + \frac{\varepsilon}{2\sqrt{M/N}},$$

for all  $N > M \geq 54$ . Set  $f(x) := x^{2m} + x^{2m-1}$  and note that  $f$  is non-increasing on  $(-\infty, -1 + \frac{1}{2m}]$  and non-decreasing on  $[-1 + \frac{1}{2m}, \infty)$ . Furthermore, its minimum is  $-e_m$  where

$$e_m := \frac{(2m-1)^{2m-1}}{(2m)^{2m}} = \frac{(1 - \frac{1}{2m})^{2m}}{2m-1} \leq \frac{1}{2em},$$

and it is non-negative on  $(-\infty, -1] \cup [0, \infty)$ . Using Markov inequality, we deduce that

$$\begin{aligned} \mathbb{P}(\lambda_M(\tilde{\mathbf{C}}) \geq \varepsilon_{M,N}) &\leq \mathbb{P}(f(\lambda_M(\tilde{\mathbf{C}})) + e_m \geq f(\varepsilon_{M,N}) + e_m) \\ &\leq \frac{\mathbb{E}[f(\lambda_M(\tilde{\mathbf{C}})) + e_m]}{f(\varepsilon_{M,N}) + e_m} \\ &\leq \frac{\sum_{k=1}^M (\mathbb{E}[f(\lambda_k(\tilde{\mathbf{C}})) + e_m])}{f(\varepsilon_{M,N})} \\ (22) \quad &= \frac{A_m + M e_m}{f(\varepsilon_{M,N})} \end{aligned}$$

Invoke Lemma 11 to get that

$$\mathbb{P}(\lambda_M(\tilde{\mathbf{C}}) \geq \varepsilon_{M,N}) \leq \frac{C_{0,\text{Rad}} m \left[ \left( \frac{MN}{(M-1)(N-1)} \right)^m + \frac{M}{m} \right] \exp \left( C_{\text{Rad}} (1 + \sqrt{M/N})^4 \frac{m^3}{M^2} \right) + \frac{M}{2em}}{\left(1 - \frac{M}{N}\right) f(\varepsilon_{M,N})},$$

for all  $m \in \mathbb{N}$ . Using that  $M \geq 54$  and  $\log(1+x) \leq x$ , we get

$$\mathbb{P}(\lambda_M(\tilde{\mathbf{C}}) \geq \varepsilon_{M,N}) \leq \frac{C_{0,\text{Rad}} \left[ m + \frac{1+2e}{2e} M \right] e^{C_{\text{Rad}} (1 + \sqrt{M/N})^4 \frac{m^3}{M^2} + 54m \left( \frac{1}{M} + \frac{1}{N} \right) \log(\frac{54}{53})}}{\left(1 - \frac{M}{N}\right) f(\varepsilon_{M,N})}$$

for all  $m \in \mathbb{N}$ . Optimizing on  $m$  yields the choice  $m = \sqrt{\frac{2 \log(\varepsilon_{M,N})}{3C_{\text{Rad}}(1 + \sqrt{M/N})^4}} M$  and

$$\mathbb{P}\left\{ \lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N \right\} \leq \frac{\mathbb{W}_0(\rho, \varepsilon)}{1 - \rho} M \exp(-N \mathbb{W}_1(\rho, \varepsilon))$$

where  $\rho = M/N$  and

$$\begin{aligned} \mathbb{W}_0(\rho, \varepsilon) &:= \frac{C_{0,\text{Rad}}(1+2e)\sqrt{3C_{\text{Rad}}}(1+\sqrt{\rho})^2 + 2e\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{2e\sqrt{3C_{\text{Rad}}}(1+\sqrt{\rho})^2} \\ &\quad \times \exp \left[ 54 \log\left(\frac{54}{53}\right)(1+\rho) \frac{\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{(1+\sqrt{\rho})^2\sqrt{3C_{\text{Rad}}}} \right] \\ \mathbb{W}_1(\rho, \varepsilon) &:= \frac{4\sqrt{2}\rho \log(1+\frac{\varepsilon}{2\sqrt{\rho}})^{\frac{3}{2}}}{3\sqrt{3}(1+\sqrt{\rho})^2\sqrt{C_{\text{Rad}}}} \end{aligned}$$

Using that  $\rho \leq 1$ , we derive that

$$\begin{aligned} \mathbb{W}_0(\rho, \varepsilon) &\leq \frac{4C_{0,\text{Rad}}(1+2e)\sqrt{3C_{\text{Rad}}} + 2e\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{2e\sqrt{3C_{\text{Rad}}}} \\ &\quad \times \exp \left[ 108 \log\left(\frac{54}{53}\right) \frac{\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{\sqrt{3C_{\text{Rad}}}} \right] \\ &\leq c_0 \exp \left[ c_0 \sqrt{\log\left(1 + \frac{\varepsilon}{2\sqrt{\rho}}\right)} \right] \end{aligned}$$

for some universal constant  $c_0 > 0$ . We deduce the following useful bound

$$(23) \quad \mathbb{P}\left\{ \lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N \right\} \leq \frac{c_0 M e^{c_0 \sqrt{\log\left(1 + \frac{\varepsilon}{2\sqrt{\rho}}\right)}}}{1 - \rho} e^{-N \mathbb{W}_1(\rho, \varepsilon)}.$$

For  $\varepsilon \leq \sqrt{\rho}$  we can deduce a small deviation inequality as follows. Observe that for any  $\eta > 0$  one can pick a constant  $c_1(\eta) > 0$ , that depends only on  $\eta$ , such that for all  $M \geq 1$ , it holds  $M \leq c_1(\eta) \exp(\eta M)$ . Note that  $\log(3/2) \frac{\varepsilon}{\sqrt{\rho}} \leq \log\left(1 + \frac{\varepsilon}{2\sqrt{\rho}}\right)$  and set

$$\mathbb{V}_{\text{Rad}} := \frac{3\sqrt{3C_{\text{Rad}}}}{4\sqrt{2}\log(3/2)^{3/2}}.$$

We deduce that for any  $C > \mathbb{V}_{\text{Rad}} \approx 3242$  there exists a constant  $\nu := \nu(\rho, C) > 0$  that depends only on  $\rho = M/N$  and  $C$  such that, for all  $0 \leq \varepsilon \leq \sqrt{\rho}$ ,

$$(24) \quad \mathbb{P}\{\lambda_M(\mathbf{C}) \geq (\sqrt{M} + \sqrt{N})^2 + \varepsilon N\} \leq \nu \exp\left(-C^{-1}N \frac{\rho^{1/4}}{(1 + \sqrt{\rho})^2} \varepsilon^{\frac{3}{2}}\right).$$

**A.6. Small deviation on the smallest eigenvalue.** Observe that

$$\mathbb{P}\{\lambda_1(\mathbf{C}) \leq (\sqrt{M} - \sqrt{N})^2 - \varepsilon'N\} = \mathbb{P}\{\lambda_1(\tilde{\mathbf{C}}) \leq -\varepsilon'_{M,N}\},$$

with

$$1 + \frac{\varepsilon'27}{53\sqrt{M/N}} \geq \varepsilon'_{M,N} := \frac{\sqrt{MN} - 1}{\sqrt{(M-1)(N-1)}} + \frac{\varepsilon'N}{2\sqrt{(M-1)(N-1)}} \geq \frac{53}{54} \left(1 + \frac{\varepsilon'27}{53\sqrt{M/N}}\right),$$

for all  $N > M \geq 54$ . Set  $g(x) := x^{2m} - x^{2m-1}$  and note that  $g(x) = f(-x)$ . It holds

$$\begin{aligned} \mathbb{P}\{\lambda_1(\tilde{\mathbf{C}}) \leq -\varepsilon'_{M,N}\} &\leq \mathbb{P}\{g(\lambda_1(\tilde{\mathbf{C}})) + e_m \geq g(-\varepsilon'_{M,N}) + e_m\} \\ &\leq \frac{\mathbb{E}[g(\lambda_1(\tilde{\mathbf{C}})) + e_m]}{f(\varepsilon'_{M,N}) + e_m} \\ &\leq \frac{\sum_{k=1}^M (\mathbb{E}[g(\lambda_k(\tilde{\mathbf{C}})) + e_m])}{f(\varepsilon'_{M,N})} \\ &= \frac{B_m + M e_m}{f(\varepsilon'_{M,N})} \end{aligned}$$

and we recover an upper bound of the form (22) for which Lemma 11 can also be applied and we get that

$$\mathbb{P}\{\lambda_1(\tilde{\mathbf{C}}) \leq -\varepsilon'_{M,N}\} \leq \frac{C_{0,\text{Rad}} m \left[ \left( \frac{MN}{(M-1)(N-1)} \right)^m + \frac{M}{m} \right] \exp\left( C_{\text{Rad}} (1 + \sqrt{M/N})^4 \frac{m^3}{M^2} \right) + \frac{M}{2em}}{\left(1 - \frac{M}{N}\right) f(\varepsilon'_{M,N})},$$

for all  $m \in \mathbb{N}$ . The rest of the proof follows the same lines as in Section A.5 where we change  $\varepsilon_{M,N}$  by  $\varepsilon'_{M,N}$ , we choose  $m = \sqrt{\frac{2 \log(54\varepsilon'_{M,N}/53)}{3C_{\text{Rad}}(1 + \sqrt{M/N})^4}} M$  and may have changed the harmless constant  $c_0$  in  $\mathbb{W}_0$ . Eventually, note that (24) has been obtained from (23) and we can use the same argument for the deviation on the smallest eigenvalue. This proves Proposition 8.

#### APPENDIX B. STIRLING'S FORMULA AND BOUNDS ON BINOMIAL COEFFICIENTS

**Lemma 12.** *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 [AS65] Eq. 6.1.38. □

**Lemma 13.** *It holds, for all  $1 \leq n \leq m$ ,*

$$\begin{aligned} \log \left[ \frac{n}{2^{2m}} \binom{2m}{m-n} \right] &\leq 5 - 0.6321 \frac{n^2}{m} \\ \log \left[ \frac{n+1/2}{2^{2m}} \binom{2m+1}{m-n} \right] &\leq 2 - 0.6555 \frac{n^2}{m} \end{aligned}$$

*Proof.* If  $n = m$  then the result is clear. Otherwise, using Lemma 12, one has

$$\begin{aligned} \log \left[ \frac{n}{2^{2m}} \binom{2m}{m-n} \right] &\leq -0.364 + \log n + (2m + 1/2) \log m \\ &\quad - (m - n + 1/2) \log(m - n) - (m + n + 1/2) \log(m + n), \\ &\leq -0.364 - 1/2 \log((m^2 - n^2)/(mn^2)) \\ &\quad + m \left[ \frac{n}{m} \log \left( 1 - \frac{2n/m}{1 + n/m} \right) - \log \left( 1 - \left( \frac{n}{m} \right)^2 \right) \right]. \end{aligned}$$

The last term in the right hand side can be upper bounded thanks to the identity  $x \log(1 - 2x/(1 + x)) - \log(1 - x^2) \leq -x^2$  for all  $0 < x < 1$ . It yields

$$m \left[ \frac{n}{m} \log \left( 1 - \frac{2n/m}{1 + n/m} \right) - \log \left( 1 - \left( \frac{n}{m} \right)^2 \right) \right] \leq -\frac{n^2}{m}.$$

Let  $x = n/m$  and observe that  $x \leq 1 - 1/m$ . It holds that the middle term of the aforementioned right hand side can be expressed as

$$-1/2 \log((m^2 - n^2)/(mn^2)) = 1/2 \log(mx^2/(1 - x^2)).$$

If  $x \leq 0.99995$  then, using that  $\log(z) \leq z/e$ , we have

$$1/2 \log(mx^2/(1 - x^2)) \leq 4.6052 + (1/(2e))mx^2.$$

If  $0.99995 < x \leq 1 - 1/m$  then

$$1/2 \log(mx^2/(1 - x^2)) \leq \log m \leq m/e < 0.3679mx^2.$$

In all cases, we get that

$$1/2 \log(mx^2/(1 - x^2)) \leq 4.6052 + 0.3679mx^2$$

We deduce that

$$\log \left[ \frac{n}{2^{2m}} \binom{2m}{m-n} \right] \leq 4.24 - 0.6321n^2/m,$$

as claimed.  $\square$

**Acknowledgments.** We would like to thank Sasha Sodin for his patient answers to our many questions.

#### REFERENCES

- [ADCM14] J.-M. Azaïs, Y. De Castro, and S. Mourareau, *A rice method proof of the null-space property over the grassmannian*, arXiv preprint arXiv:1405.6417 (2014).
- [AGZ10] G. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
- [ALPTJ11] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, *Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling*, *Constructive Approximation* **34** (2011), no. 1, 61–88.
- [AS65] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, National Bureau of Standards, Washington DC (1965).
- [BCT11] J. D. Blanchard, C. Cartis, and J. Tanner, *Compressed sensing: How sharp is the restricted isometry property?*, *SIAM review* **53** (2011), no. 1, 105–125.
- [BF03] A. Borodin and P. Forrester, *Increasing subsequences and the hard-to-soft edge transition in matrix ensembles*, *J. Phys. A* **36** (2003), no. 12, 2963–2981, *Random matrix theory*.
- [BLPR11] K. Bertin, E. Le Pennec, and V. Rivoirard, *Adaptive dantzig density estimation*, *Annales de l’IHP, Probabilités et Statistiques* **47** (2011), no. 1, 43–74.
- [BRT09] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, *Simultaneous analysis of lasso and Dantzig selector*, *Ann. Statist.* **37** (2009), no. 4, 1705–1732. MR 2533469 (2010j:62118)

- [BS10] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR 2567175 (2011d:60014)
- [CGLP12] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor, *Interaction between compressed sensing, random matrices and high dimensional geometry*, Panoramas et synthèses, no. 37, SMF, 2012.
- [CRT06] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509. MR 2236170 (2007e:94020)
- [CT05] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215. MR 2243152 (2007b:94313)
- [CT06] ———, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425. MR 2300700 (2008c:94009)
- [DC13] Y. De Castro, *A remark on the lasso and the dantzig selector*, Statistics and Probability Letters **83** (2013), no. 1, 304–314.
- [DS01] K. R Davidson and S. J. Szarek, *Local operator theory, random matrices and banach spaces*, Handbook of the geometry of Banach spaces **1** (2001), no. 317-366, 131.
- [DT05] 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** (2005), no. 27, 9452–9457.
- [DT09a] ———, *Counting faces of randomly projected polytopes when the projection radically lowers dimension*, Journal of the American Mathematical Society **22** (2009), no. 1, 1–53.
- [DT09b] ———, *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** (2009), no. 1906, 4273–4293.
- [FL09] S. Foucart and M.-J. Lai, *Sparsest solutions of underdetermined linear systems via  $l_q$ -minimization for  $0 < q \leq 1$* , Applied and Computational Harmonic Analysis **26** (2009), no. 3, 395–407.
- [FR13] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Springer, 2013.
- [FS10] O. N. Feldheim and S. Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geometric And Functional Analysis **20** (2010), no. 1, 88–123.
- [JN11] A. Juditsky and A. Nemirovski, *Accuracy guarantees for  $l_1$ -recovery*, Information Theory, IEEE Transactions on **57** (2011), no. 12, 7818–7839.
- [Joh00] K. Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. **209** (2000), no. 2, 437–476.
- [Led07] M. Ledoux, *Deviation inequalities on largest eigenvalues*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1910, Springer, Berlin, 2007, pp. 167–219.
- [LR10] M. Ledoux and B. Rider, *Small deviations for beta ensembles*, Electron. J. Probab **15** (2010), no. 41, 1319–1343.
- [MPTJ08] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, *Uniform uncertainty principle for Bernoulli and subgaussian ensembles*, Constr. Approx. **28** (2008), no. 3, 277–289. MR 2453368 (2009k:46020)
- [Péc09] S. Péché, *Universality results for the largest eigenvalues of some sample covariance matrix ensembles*, Probab. Theory Related Fields **143** (2009), no. 3-4, 481–516.
- [PS11] L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, vol. 171, American Mathematical Society, Providence, RI, 2011.
- [PY14] N.S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), no. 3, 935–1001.
- [Sos02] A. Soshnikov, *A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices*, J. Statist. Phys. **108** (2002), no. 5-6, 1033–1056.
- [Tik15] K. Tikhomirov, *The limit of the smallest singular value of random matrices with i.i.d. entries*, Adv. Math. **284** (2015), 1–20.
- [vdGB09] S. A. van de Geer and P. Bühlmann, *On the conditions used to prove oracle results for the lasso*, Electronic Journal of Statistics **3** (2009), 1360–1392.
- [Wan12] K. Wang, *Random covariance matrices: universality of local statistics of eigenvalues up to the edge*, Random Matrices Theory Appl. **1** (2012), no. 1, 1150005, 24.

SD IS WITH CMLA, ENS CACHAN, CNRS, UNIVERSITÉ PARIS-SACLAY, 94235 CACHAN, FRANCE.  
*E-mail address:* `sandrine.dallaporta@cmla.ens-cachan.fr`

YDC IS WITH THE LABORATOIRE DE MATHÉMATIQUES D'ORSAY, UNIV. PARIS-SUD, CNRS, UNIVERSITÉ PARIS-SACLAY, 91405 ORSAY, FRANCE.  
*E-mail address:* `yohann.decastro@math.u-psud.fr`