# 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 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. In this paper, we show upper bounds on RICs for Gaussian and Rademacher matrices using state-of-the-art small deviation estimates on their extreme eigenvalues. This allows us to derive a lower bound on the probability of getting SRSR. One of the benefits of this approach is to introduce a simple tool from Random Matrix Theory to derive upper bounds on RICs and phase transition on SRSR from small deviations on the extreme eigenvalues.

## 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

$$y = \mathbf{M}x_0 + \mathbf{e} \tag{1}$$

where $\mathbf{M}$ is a known $n \times p$ matrix, $x_0$ a unknown vector in $\mathbb{R}^p$, $y$ 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 $x_0$ from a sample $y$ 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 $\hat{x}$ of $x_0$ considering $\ell_1$-minimization solutions as

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

where $\eta > 0$ is a tuning parameter such that the experimenter believes it holds $||\mathbf{e}||_2 \leqslant \eta$ with high probability. Then, a standard goal is to prove that

$$\|x_0 - \hat{x}\|_1 \leqslant C\sigma_s(x_0)_1 + D\sqrt{s}\eta \tag{$\ell_1$-SRSR}$$

$$\|x_0 - \hat{x}\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x_0)_1 + D\eta \tag{$\ell_2$-SRSR}$$

where $C, D > 0$ are constants and $\sigma_s(x_0)_1$ denotes the approximation error in $\ell_1$-norm by $s$ coefficients, namely $\sigma_s(x_0)_1 := \min \|x_0 - x\|_1$ 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 ($\ell_1$-**SRSR**) and ($\ell_2$-**SRSR**) may

*Date*: March 1, 2017.
*Key words and phrases.* Restricted Isometry Property; Gaussian; Rademacher; Small Deviations;

be referenced as the Stable and Robust Sparse Recovery (SRSR) property of order $s$, see [FR13, Page 88]. It shows that $\ell_1$-minimization recovers the $s$ largest coefficients of a target vector $x_0$ in a stable[1] 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 RIP$(s, c, \mathbf{M})$. It is defined by

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

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

$$\mathfrak{c}(2s, \mathbf{M}) < 4/\sqrt{41} \simeq 0.625\,, \qquad \text{(FR-}\mathfrak{c}(2s))$$

holds and $\hat{x}$ is any solution to (2) then SRSR of order $s$ holds with $C, D > 0$ depending only on $\mathfrak{c}(2s, \mathbf{M})$. A slightly modified RIP was introduced by Foucart and Lai in [FL09, BCT11]. They introduce two constants, called Restricted Isometry Constants (RICs). For a matrix $\mathbf{M}$ of size $(n \times p)$, the RICs, $\mathfrak{c}_{\min}(s, \mathbf{M})$ and $\mathfrak{c}_{\max}(s, \mathbf{M})$, are defined as

$$\mathfrak{c}_{\min} := \min_{c_- \geqslant 0} c_- \quad \text{subject to} \quad (1-c_-)\|x\|_2^2 \leqslant \|\mathbf{M}x\|_2^2 \quad \text{for all} \quad x \in \Sigma_s,$$

$$\mathfrak{c}_{\max} := \min_{c_+ \geqslant 0} c_+ \quad \text{subject to} \quad (1+c_+)\|x\|_2^2 \geqslant \|\mathbf{M}x\|_2^2 \quad \text{for all} \quad x \in \Sigma_s.$$

Hence, it holds $(1-\mathfrak{c}_{\min})\|x\|_2^2 \leqslant \|\mathbf{M}x\|_2^2 \leqslant (1+\mathfrak{c}_{\max})\|x\|_2^2$ for all $x \in \Sigma_s$, where we recall that $\Sigma_s$ denotes the set of vectors with at most $s$ nonzero coordinates. Reporting the influence of both extreme eigenvalues of covariance matrices built from $2s$ columns of $\mathbf{M}$, one can weaken (FR-$\mathfrak{c}(2s)$), see for instance Theorem 2.1 in [FL09]. Revisiting [FL09] and [FR13, Proof of Theorem 6.13 (Page 145)], this paper provides the weakest condition to get SRSR, see Appendix A.1 for a proof.

**Theorem 1.** *If $\mathbf{M}$ satisfies this asymmetric Restricted Isometry Property with RICs such that*

$$\gamma(2s, n, p) := \frac{1+\mathfrak{c}_{\max}(2s, \mathbf{M})}{1-\mathfrak{c}_{\min}(2s, \mathbf{M})} < \frac{(4+\sqrt{41})^2}{25} \simeq 4.329, \qquad \text{(SRSR-}\gamma(2s))$$

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

*Remark.* Remark that the condition to get SRSR described in [FL09, Theorem 2.1] can be equivalently written as $\gamma(2s, n, p) < (5+\sqrt{2})/(1+\sqrt{2}) \simeq 2.657$ which is a stronger requirement than (SRSR-$\gamma(2s)$). Also, note that Condition (SRSR-$\gamma(2s)$) leads to $(1+\mathfrak{c}(2s, \mathbf{M}))/(1-\mathfrak{c}(2s, \mathbf{M})) < (4+\sqrt{41})^2/25$ and one can check that this is exactly Condition (FR-$\mathfrak{c}(2s)$). From this remark, one can view (SRSR-$\gamma(2s)$) as a generalization of (FR-$\mathfrak{c}(2s)$) to the frame of asymmetric isometry constants.

[1] In a 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.

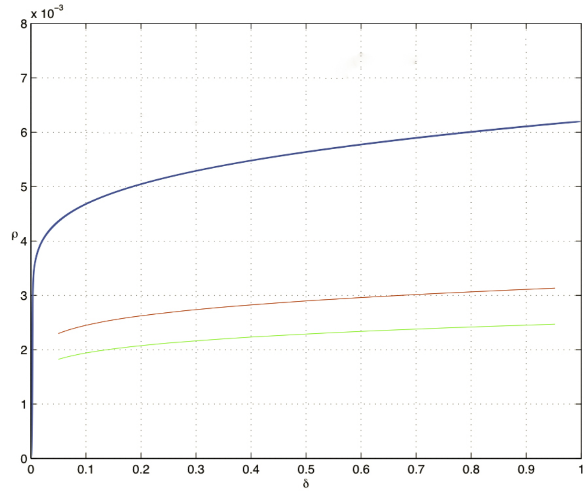


FIGURE 1. Our new bound in blue (derived from (3)) improves upon the lower bound of [BCT11] in red (derived using Foucart and Lai condition [FL09, Theorem 2.1]) and upon [Can08] (derived from symmetric RICs bounds as in (FR-$\mathfrak{c}(2s)$)). This figure is an update of Figure 3.2 in [BCT11].

1.3. **New RICs and SRSR bounds.** In this paper, we provide a new tool to derive upper bounds on RICs (with overwhelming probability) from deviation inequalities on extreme eigenvalues (or extreme singular values) of covariance matrices $\mathbf{C}_{s,n} = \frac{1}{n}\mathbf{X}\mathbf{X}^\star$ where $\mathbf{X} \in \mathbb{R}^{s\times n}$ has i.i.d. entries drawn with respect to a law $\mathcal{L}$. We consider the asymptotic proportional growth model where $s/n \to \rho$ (size of the sparse vectors over number of equations) and $n/p \to \delta$ (number of equations over number of unknowns) as in [DT05, DT09a, DT09b, BT10, BCT11, BT14]. Using Theorem 1, these results on RICs give new lower bounds on SRSR. More precisely, we establish a new sufficient condition on SRSR that improves previous state-of-the-art results [BCT11]. Indeed, using Davidson-Szarek's deviation [DS01], we prove that if

$$\delta > \frac{1}{\rho}\exp\left[1 - \frac{1}{2\rho}\left(\sqrt{\frac{33-5\sqrt{41}}{8}} - \sqrt{\rho}\right)^2\right] \tag{3}$$

then SRSR holds with overwhelming probability when $\mathbf{X}$ has i.i.d. Gaussian entries, see Section 2.3.2. It improves previous state-of-the-art result [BCT11, RV08], see Figure 1. On a more general note, 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 $n \geqslant n_0(\rho)$,

$$\forall t \geqslant 0, \quad \mathbb{P}\Big\{\big(\lambda_1 - (1+\sqrt{\rho})^2\big) \vee \big((1-\sqrt{\rho})^2 - \lambda_s\big) \geqslant t\Big\} \leqslant c(\rho)e^{-n\mathbb{W}(\rho,t)} \tag{4}$$

where $n_0(\rho) \geqslant 2$ and $c(\rho) > 0$ may depend on the ratio $\rho$, the function $t \mapsto \mathbb{W}(\rho, t)$ is continuous and increasing on $[0, \tau_1)$ such that $\mathbb{W}(\rho, 0) = 0$. The known asymptotic behavior of these extreme eigenvalues provides an expected behavior for $W(\rho, t)$ in both variables $t$ and $\rho$. 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. More details will be given in Section 2.3. Unfortunately, this dependence is sometimes unclear 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 B reveals the dependency on $\rho$ as well as bounds on the constant appearing in their rate function $\mathbb{W}_{\mathrm{FS}}$ 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 A.2.

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.

1.4. **Previous works on bounding RIP and RICs.** The existence of RIP matrices with bounded RIP constant such as (FR-$\mathfrak{c}(2s)$) 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 (FR-$\mathfrak{c}(2s)$) 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

$$n \geqslant c_1\, s \log(c_2 p/s) \tag{5}$$

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 (FR-$\mathfrak{c}(2s)$) 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 $n \geqslant c_1'\, s \log(c_2' p/s)$ 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 (5) 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 (5) depicts a necessary and sufficient condition on NSP up to a change of constants, see for instance [CGLP12, ADCM17]. 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. Let us also mention the important papers [MT14, ALMT14] that give appealing and rigorous quantitative estimates of "weak" thresholds appearing in convex optimization, including the location and the width of the transition region for NSP.

Following this outbreaking result, one can wonder whether a phase transition holds for properties guaranteeing SRSR such as Condition (FR-$\mathfrak{c}(2s)$) or the asymmetric (SRSR-$\gamma(2s)$). To the best of our knowledge, the first 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 (SRSR-$\gamma(2s)$) holds. Their lower bound is not explicit but one can witness in [BCT11, Page 119]. Furthermore they provide web forms for the calculation of bounds on RICs, which are available at Jared Tanner's webpage. Shortly after, Bah and Tanner improved these bounds in [BT10] preventing the use of union bound over all sub-matrices built from $2s$ columns of $\mathbf{M}$ by grouping those which share a substantial number of columns. Their bounds are still implicit but web forms for their calculation are available at the same place. The same authors provided later [BT14] explicit bounds for the RICs in extreme asymptotic regime:

(a) when $\rho \to 0$ and $\delta > 0$ is fixed,
(b) when $\delta \to 0$ and $\rho > 0$ is fixed,
(c) when $\rho = \frac{-1}{\gamma \log \delta}$ ($\gamma$ is a fixed parameter) and $\delta \to 0$.

In the sequel, we may refer to these regimes as Regime $(a)$, $(b)$ and $(c)$ respectively.

1.5. **Outline.** The paper is organized as follows. Section 2 states the main results: it provides a general method to derive bounds on RICs and phase transition in Condition (SRSR-$\gamma(2s)$) from deviation inequalities on eigenvalues or singular values. Subsection 2.3 begins with a discussion on what is expected for such deviation inequalities. The general method described in Subsections 2.1 and 2.2 is then applied to previously known inequalities. Section 2 ends with a summary of the obtained bounds.

The proofs are contained in the appendix. Appendix A.1 provides the proof of Theorem 1, while Appendix A.2 and Appendix A.3 contain the proofs of Theorems 2 and 3. In Appendix B, we follow the steps of [FS10] to provide an upper bound on the constant in the deviation inequality for extreme singular values of Rademacher matrices.

**Acknowledgments.** We would like to thank Sasha Sodin for his patient answers to our many questions. Moreover, this paper greatly benefited from the comments of anonymous referees.

## 2. From small deviations to RICs and SRSR bounds

Following the framework of [BCT11], we provide asymptotic bounds on RICs in the proportional growth model. As previously explained, we suppose that we are able to control the deviation of extreme eigenvalues or singular values (4). We aim at controlling uniformly the extreme eigenvalues, the combinatorial complexity is standardly [BCT11] controlled by the quantity $\delta^{-1}\mathbf{H}_e(\rho\delta)$ where

$$\mathbf{H}_e(t) := -t \log t - (1-t)\log(1-t) \quad \text{for } t \in (0,1)\,,$$

denotes the Shannon entropy. The improvement introduced in [BT10] to deal with this combinatorial complexity could be used here but we chose not to do so as it would have turned our explicit bounds into implicit ones. One may remark that the quantity $t_0 := \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$ governs the order of the deviation in the rate function $t \mapsto \mathbb{W}(\rho, t)$ when bounding the extreme eigenvalues uniformly over all possible supports $S$ of size $s$ among the set of indices $\{1, \dots, p\}$, see the functions $\Psi_{\min / \max}$ in the next theorems. Here $\mathbb{W}^{-1}(\rho, .)$ denotes the inverse of $\mathbb{W}$ with respect to its second variable.

2.1. **Using extreme eigenvalues small deviations.** A useful constant is

$$\rho_0 := (33 - 5\sqrt{41})/8 \simeq 0.1230$$

and note that $\sqrt{\rho_0} \simeq 0.3508$. Also, we denote $\tau_0 := 4/\sqrt{41} \simeq 0.6247$. The key main result is the following theorem proved in Section A.2.

**Theorem 2.** *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 $n \geqslant n_0(\rho)$,*

$$\forall 0 \leqslant t < \tau_0, \quad \mathbb{P}\Big\{ \big(\lambda_1 - (1+\sqrt{\rho})^2\big) \vee \big((1-\sqrt{\rho})^2 - \lambda_s\big) \geqslant t \Big\} \leqslant c(\rho) e^{-n\mathbb{W}(\rho,t)}$$

*where $n_0(\rho) \geqslant 2$ 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*

$$\delta > \Psi_0^{(1)}(\rho, \mathbb{W}) := \rho^{-1} \exp\Big(1 - \rho^{-1}\mathbb{W}\big[\rho, \tau_0(\sqrt{\rho} - \sqrt{\rho_0})(\sqrt{\rho} - 1/\sqrt{\rho_0})\big]\Big)\,, \tag{6}$$

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

$$\mathbb{P}\Big\{ \frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ satisfies (SRSR-}\gamma(2s)\text{) with } 2s \leqslant \lfloor \rho n \rfloor \Big\} \geqslant 1 - 2c(\rho)e^{-nD_1(\delta,\rho)} \to 1$$

*for some $D_1(\delta, \rho) > 0$ that may depend on $\delta$ and $\rho$.*

*Furthermore, 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*

$$\mathbb{P}\Big\{ \mathfrak{c}_{\min} \geqslant \Psi_{\min}^{(1)}(\delta, \rho, \mathbb{W}) + \varepsilon \Big\} \leqslant c(\rho)e^{-nD_2(\rho,\delta,\varepsilon)},$$

$$\mathbb{P}\Big\{ \mathfrak{c}_{\max} \geqslant \Psi_{\max}^{(1)}(\delta, \rho, \mathbb{W}) + \varepsilon \Big\} \leqslant c(\rho)e^{-nD_2(\rho,\delta,\varepsilon)},$$

*where $D_2(\rho, \delta, \varepsilon) > 0$, $\Psi_{\min}^{(1)}(\delta, \rho, \mathbb{W}) := \sqrt{\rho}(2 - \sqrt{\rho}) + \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$ and $\Psi_{\max}^{(1)}(\delta, \rho, \mathbb{W}) := \sqrt{\rho}(2 + \sqrt{\rho}) + \mathbb{W}^{-1}(\rho, \delta^{-1}\mathbf{H}_e(\rho\delta))$.*

2.2. **Using extreme singular values small deviations.** A similar result can be derived from deviations on singular values, a proof is given in Section A.3.

**Theorem 3.** *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 $n \geqslant n_0(\rho)$,*

$$\forall 0 < t < \sqrt{\rho_0}, \quad \mathbb{P}\Big\{ \big(\frac{\sigma_1}{\sqrt{n}} - (1+\sqrt{\rho})\big) \vee \big((1-\sqrt{\rho}) - \frac{\sigma_s}{\sqrt{n}}\big) \geqslant t \Big\} \leqslant c(\rho)e^{-n\mathbb{W}(\rho,t)}$$

*where $n_0(\rho) \geqslant 2$ 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*

$$\delta > \Psi_0^{(2)}(\rho, \mathbb{W}) := \rho^{-1} \exp\big(1 - \rho^{-1}\mathbb{W}\big[\rho, \sqrt{\rho_0} - \sqrt{\rho}\big]\big)\,, \tag{7}$$

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

$$\mathbb{P}\Big\{\frac{\mathbf{M}^{(n)}}{\sqrt{n}} \text{ satisfies (SRSR-}\gamma(2s)\text{) with } 2s \leqslant \lfloor \rho n \rfloor\Big\} \geqslant 1 - 2c(\rho)e^{-nD_1(\delta,\rho)} \to 1$$

*for some $D_1(\delta,\rho) > 0$ that may depend on $\delta$ and $\rho$.*

*Furthermore, 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*

$$\mathbb{P}\Big\{\mathfrak{c}_{\min} \geqslant \Psi^{(2)}_{\min}(\delta,\rho,\mathbb{W}) + \varepsilon\Big\} \leqslant c(\rho)e^{-nD_2(\rho,\delta,\varepsilon)},$$
$$\mathbb{P}\Big\{\mathfrak{c}_{\max} \geqslant \Psi^{(2)}_{\max}(\delta,\rho,\mathbb{W}) + \varepsilon\Big\} \leqslant c(\rho)e^{-nD_2(\rho,\delta,\varepsilon)},$$

*where $D_2(\rho,\delta,\varepsilon) > 0$ and we denote $\Psi^{(2)}_{\min}(\delta,\rho,\mathbb{W}) := \min\{1, (\sqrt{\rho}+t_0)(2-\sqrt{\rho}-t_0)\}$ and $\Psi^{(2)}_{\max}(\delta,\rho,\mathbb{W}) := (\sqrt{\rho}+t_0)(2+\sqrt{\rho}+t_0)$ with $t_0 := \mathbb{W}^{-1}(\rho,\delta^{-1}\mathbf{H}_e(\rho\delta))$.*

Theorems 2 and 3 give a general method to derive bounds on RICs from deviation inequalities satisfied by the extreme eigenvalues or singular values of a random covariance matrix. In the following subsection, three known deviation inequalities are used to provide such bounds for Gaussian and Rademacher matrices.

2.3. **State-Of-The-Art deviation inequalities.** The asymptotic behavior of extreme eigenvalues of random covariance matrices with iid entries has been known for some years. From this behavior and the concentration of measure phenomenon, we present what is expected for deviation inequalities for extreme eigenvalues of such matrices with sub-Gaussian entries. This is what we call "ideal deviations". The next two paragraphs are devoted to deviation inequalities for Gaussian matrices due to Davidson and Szarek [DS01], and Ledoux and Rider [LR10]. The last paragraph focuses on a deviation inequality for Rademacher matrices, proved by Feldheim and Sodin [FS10].

2.3.1. *Ideal deviations.* The asymptotic behavior of extreme eigenvalues for random covariance matrices was first established for matrices with Gaussian entries [Joh00, BF03] and extended to ones with more general entries in [Sos02, Péc09, FS10, PY14, Wan12]. The largest eigenvalue fluctuations are described by the following:

$$\left[\frac{n\rho^{1/4}}{(1+\sqrt{\rho})^2}\right]^{\frac{2}{3}}(\lambda_1 - (1+\sqrt{\rho})^2) \underset{n\to\infty}{\overset{(d)}{\to}} F_1, \tag{8}$$

where $F_1$ is the so-called Tracy-Widom law. As for the smallest eigenvalue, when $\rho < 1$ (which is true in our setting),

$$\left[\frac{n\rho^{1/4}}{(1-\sqrt{\rho})^2}\right]^{\frac{2}{3}}((1-\sqrt{\rho})^2 - \lambda_s) \underset{n\to\infty}{\overset{(d)}{\to}} F_1. \tag{9}$$

We focus on the largest eigenvalue $\lambda_1$ and write:

$$\mathbb{P}(\lambda_1 \geqslant (1+\sqrt{\rho})^2 + t) = \mathbb{P}\bigg(\left[\frac{n\rho^{1/4}}{(1+\sqrt{\rho})^2}\right]^{\frac{2}{3}}(\lambda_1 - (1+\sqrt{\rho})^2) \geqslant \frac{n^{2/3}\rho^{1/6}}{(1+\sqrt{\rho})^{4/3}}t\bigg).$$

This deviation probability is therefore expected to be close to

$$1 - F_1\bigg(\frac{n^{2/3}\rho^{1/6}}{(1+\sqrt{\rho})^{4/3}}t\bigg),$$

where $F_1$ is the cdf of the Tracy-Widom distribution. Thus it is expected to be close to the tail behavior of $F_1$ at $\infty$, which is actually known:

$$1 - F_1(x) \underset{x\to\infty}{\sim} e^{-\frac{2}{3}x^{3/2}}.$$

As a consequence, deviation inequalities for the largest eigenvalue are expected to conform to

$$\mathbb{P}(\lambda_1 \geqslant (1+\sqrt{\rho})^2 + t) \leqslant C\exp\Big(-c\frac{\rho^{1/4}}{(1+\sqrt{\rho})^2} n t^{3/2}\Big), \tag{10}$$

at least for $t$ of the order of the spectrum width (which behaves asymptotically as $\mathcal{O}(\sqrt{\rho})$). For bigger $t$, due to the concentration of measure phenomenon, the expected behavior is the following:

$$\mathbb{P}(\lambda_1 \geqslant (1+\sqrt{\rho})^2 + t) \leqslant C e^{-cn\min(t,t^2)}. \tag{11}$$

Similar results should hold for the smallest eigenvalue, except that $\lambda_s \geqslant 0$ almost surely and therefore only moderate deviations can occur. See [Led07] for a detailed survey on this subject and [LR10] p.1322 for a specific discussion on the change of behavior occurring around $t = \mathcal{O}(\sqrt{\rho})$.

Considering these expected deviation inequalities, it may be possible to prove the following for sub-Gaussian random matrices.

$$\forall t > 0, \quad \mathbb{P}\Big\{\lambda_1 - (1+\sqrt{\rho})^2 \geqslant t\Big\} \leqslant c(\rho) e^{-n\mathbb{W}_{\mathrm{TW}}(\rho,t)}$$

where $\lambda_1$ denotes the largest eigenvalue of a $s \times n$ covariance matrix $\mathbf{C}$ with i.i.d. sub-Gaussian entries and

(12)

$$\mathbb{W}_{\mathrm{TW}}(\rho,t) := \frac{1}{C}\Bigg\{\frac{\rho^{\frac{1}{4}}}{(1+\sqrt{\rho})^2} t^{\frac{3}{2}} \mathbb{1}_{t\leqslant\sqrt{\rho}} + \frac{t^2}{(1+\sqrt{\rho})^2}\mathbb{1}_{\sqrt{\rho}<t\leqslant 1} + \frac{t}{(1+\sqrt{\rho})^2}\mathbb{1}_{t>1}\Bigg\},$$

where $C > 0$. A similar deviation inequality may be established for the smallest eigenvalue $\lambda_s$ with almost a similar $\mathbb{W}$ function (the $(1+\sqrt{\rho})^2$ terms should be replaced by $(1-\sqrt{\rho})^2$). We should obtain

$$\begin{aligned}\mathbb{W}_{\mathrm{TW}}^{-1}(\rho,u) = {} & \frac{C^{2/3}(1+\sqrt{\rho})^{4/3}}{\rho^{\frac{1}{6}}} u^{\frac{2}{3}} \mathbb{1}_{u\leqslant\frac{\rho}{C(1+\sqrt{\rho})^2}} \\ & + \quad C^{1/2}(1+\sqrt{\rho})\sqrt{u}\,\mathbb{1}_{\frac{\rho}{C(1+\sqrt{\rho})^2}<u\leqslant\frac{1}{C(1+\sqrt{\rho})^2}} \\ & + \quad C(1+\sqrt{\rho})^2 u \mathbb{1}_{u>\frac{1}{C(1+\sqrt{\rho})^2}}.\end{aligned}$$

Theorem 2 could then be invoked to get bounds on RICs,

$$\begin{aligned}\mathfrak{c}_{\max} &\leqslant \sqrt{\rho}(2+\sqrt{\rho}) + t_0, \\ \mathfrak{c}_{\min} &\leqslant \sqrt{\rho}(2-\sqrt{\rho}) - t_0,\end{aligned}$$

where $t_0 = \mathbb{W}_{\mathrm{TW}}^{-1}\big(\rho, \frac{1}{\delta}\mathbf{H}_e(\rho\delta)\big)$. Note that in the two asymptotic Regimes (a) and (c) of [BT14], $t_0$ will be larger than $\frac{\rho}{C(1+\sqrt{\rho})^2}$. Therefore it seems that the most important part in the rate function (12) for our present use is the moderate and large deviation part, arising from the concentration of measure phenomenon.

2.3.2. *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

$$\forall t > 0, \quad \mathbb{P}\Big\{\big(\frac{\sigma_1(\mathbf{X})}{\sqrt{n}} - (1+\sqrt{\rho})\big) \vee \big((1-\sqrt{\rho}) - \frac{\sigma_s(\mathbf{X})}{\sqrt{n}}\big) \geqslant t\Big\} \leqslant 2e^{-n\mathbb{W}_{\mathrm{DS}}(\rho,t)}$$

where $\sigma_i(\mathbf{X})$ denotes the singular values of $\mathbf{X}$ and $\mathbb{W}_{\mathrm{DS}}(\rho,t) := t^2/2$, see [FR13, Page 291] for instance. This inequality relies on the concentration of measure phenomenon. Note that

$$\mathbb{W}_{\mathrm{DS}}^{-1}(\rho,u) = \sqrt{2u}.$$

Theorem 3 applied here gives the following high probability bounds on RICs:

$$\mathfrak{c}_{\max} \leqslant (\sqrt{\rho}+t_0)(2+\sqrt{\rho}+t_0),$$
$$\mathfrak{c}_{\min} \leqslant (\sqrt{\rho}+t_0)(2-\sqrt{\rho}-t_0),$$

where $t_0 = \sqrt{\frac{2}{\delta}\mathbf{H}_e(\rho\delta)}$. In the three asymptotic Regimes (a), (b) and (c), these bounds on RICs behave similarly to the ones obtained by Bah and Tanner in [BT14], except that constants are better in [BT14]. Note that this deviation has been used in the paper [CT05, Lemma 3.1] to bound the RIP constant.

Furthermore, Theorem 3 states that Condition (SRSR-$\gamma(2s)$) is satisfied with high probability whenever

$$\delta > \frac{1}{\rho}\exp\Big[1 - \frac{1}{2\rho}(\sqrt{\rho_0}-\sqrt{\rho})^2\Big].$$

When $\rho$ is small (which is the case in the Regimes (a) and (c)), this condition approximately writes

$$\delta > \frac{1}{\rho}\exp\Big[-\frac{\rho_0}{2\rho}\Big].$$

2.3.3. *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 $n \geqslant 2$, setting $s = \lfloor \rho n \rfloor$,

$$\forall t > 0, \quad \mathbb{P}\Big(\lambda_1 - (1+\sqrt{\rho})^2 \geqslant t\Big) \leqslant c(\rho)e^{-n\mathbb{W}_{\mathrm{LR}}^{\max}(\rho,t)}$$

where $\lambda_1$ denotes the largest eigenvalue of a $s \times n$ covariance matrix $\mathbf{C}$ with i.i.d. standard Gaussian entries and

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

where $C_{\mathrm{LR}} > 0$ may be bounded explicitly from [LR10]. As explained in Section 1.3, 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

$$\forall t > 0, \quad \mathbb{P}\Big((1-\sqrt{\rho})^2 - \lambda_s \geqslant t\Big) \leqslant c(\rho)e^{-n\mathbb{W}_{\mathrm{LR}}^{\min}(\rho,t)}$$

where

(13)
$$\mathbb{W}_{\mathrm{LR}}^{\min}(\rho,t):=\frac{\rho^{\frac14}}{C_{\mathrm{LR}}(1-\sqrt{\rho})^3}t^{\frac32}\mathbb{1}_{t\leqslant\sqrt{\rho}(1-\sqrt{\rho})^2}+\frac{\rho^{\frac12}}{C_{\mathrm{LR}}(1-\sqrt{\rho})^2}t\mathbb{1}_{t>\sqrt{\rho}(1-\sqrt{\rho})^2}\,.$$

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

$$\forall t>0,\quad \mathbb{P}\Big\{\big(\lambda_1-(1+\sqrt{\rho})^2\big)\vee\big((1-\sqrt{\rho})^2-\lambda_s\big)\geqslant t\Big\}\leqslant c(\rho)e^{-n\mathbb{W}_{\mathrm{LR}}(\rho,t)}$$

where

$$\mathbb{W}_{\mathrm{LR}}(\rho,t):=\mathbb{W}_{\mathrm{LR}}^{\max}(\rho,t)\,. \tag{14}$$

Therefore

$$\mathbb{W}_{\mathrm{LR}}^{-1}(\rho,u)=C_{\mathrm{LR}}^{2/3}\frac{(1+\sqrt{\rho})^2}{\rho^{1/6}}u^{2/3}\mathbb{1}_{u\leqslant\frac{\rho}{C_{\mathrm{LR}}}}+C_{\mathrm{LR}}\frac{(1+\sqrt{\rho})^2}{\sqrt{\rho}}u\mathbb{1}_{u>\frac{\rho}{C_{\mathrm{LR}}}}\,.$$

Theorem 2 applied here gives the following high probability bounds on RICs:

$$\mathfrak{c}_{\max}\leqslant\sqrt{\rho}(2+\sqrt{\rho})+t_0,$$
$$\mathfrak{c}_{\min}\leqslant\sqrt{\rho}(2-\sqrt{\rho})-t_0,$$

where $t_0=\mathbb{W}_{\mathrm{LR}}^{-1}(\rho,\frac{1}{\delta}\mathbf{H}_e(\rho\delta))$. In the three asymptotic Regimes (a), (b) and (c), it may be shown that $t_0\geqslant\frac{\rho}{C_{\mathrm{LR}}}$. These bounds on RICs behave similarly to the ones obtained by Bah and Tanner in [BT14] in Regime (a), except that their constants are better. In Regimes (b) and (c), they behave badly compared to those of [BT14].

Furthermore, Theorem 2 states that Condition (SRSR-$\gamma(2s)$) is satisfied with high probability whenever

$$\delta>\frac{1}{\rho}\exp\Big[1-\frac{1}{\rho}\mathbb{W}_{\mathrm{LR}}\Big(\rho,\tau_0(\sqrt{\rho}-\sqrt{\rho_0})(\sqrt{\rho}-\frac{1}{\sqrt{\rho_0}})\Big)\Big].$$

When $\rho$ is small (which is the case in Regimes (a) and (c)), the second argument in $\mathbb{W}_{\mathrm{LR}}$ is approximately $\tau_0$ and this condition approximately writes

$$\delta>\frac{1}{\rho}\exp\Big[-\frac{\tau_0/C_{\mathrm{LR}}}{\rho^{3/2}}\Big].$$

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

$$\forall t>0,\quad \mathbb{P}\Big\{\big(\lambda_1-(1+\sqrt{\rho})^2\big)\vee\big((1-\sqrt{\rho})^2-\lambda_s\big)\geqslant t\Big\}\leqslant c(\rho)e^{-n\mathbb{W}_{\mathrm{FS}}(\rho,t)}$$

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

$$\mathbb{W}_{\mathrm{FS}}(\rho,t):=\frac{\rho\log(1+\frac{t}{2\sqrt{\rho}})^{\frac32}}{C_{\mathrm{FS}}(1+\sqrt{\rho})^2}\,,$$

where $0<C_{\mathrm{FS}}<837$, as shown in Proposition 6. Furthermore

$$\mathbb{W}_{\mathrm{FS}}^{-1}(\rho,u)=2\sqrt{\rho}\Big\{\exp\Big(C_{\mathrm{FS}}^{2/3}\frac{(1+\sqrt{\rho})^{4/3}}{\rho^{2/3}}u^{2/3}\Big)-1\Big\}.$$

Theorem 2 applied here gives the following high probability bounds on RICs:

$$\mathfrak{c}_{\max} \leqslant \sqrt{\rho}(2+\sqrt{\rho}) + t_0,$$
$$\mathfrak{c}_{\min} \leqslant \sqrt{\rho}(2-\sqrt{\rho}) - t_0,$$

where $t_0 = \mathbb{W}_{\mathrm{FS}}^{-1}(\rho, \frac{1}{\delta}\mathbf{H}_e(\rho\delta))$. In the three asymptotic regimes (a), (b) and (c), these bounds behave really badly compared to the ones by Bah and Tanner in [BT14] but note that we consider here entries which are not Gaussian anymore.

Furthermore, Theorem 2 states that Condition (SRSR-$\gamma(2s)$) is satisfied with high probability whenever

$$\delta > \frac{1}{\rho}\exp\Big[1 - \frac{1}{\rho}\mathbb{W}_{\mathrm{FS}}\Big(\rho, \tau_0(\sqrt{\rho}-\sqrt{\rho_0})(\sqrt{\rho}-\frac{1}{\sqrt{\rho_0}})\Big)\Big].$$

When $\rho$ is small (which is the case in regimes (a) and (c)), the second argument in $\mathbb{W}_{\mathrm{FS}}$ is approximately $\tau_0$ and this condition approximately writes

$$\delta > \frac{1}{\rho}\exp\Big[-\frac{|\log\rho|^{3/2}}{2C_{\mathrm{FS}}}\Big].$$

2.4. **Bounds on RICs and SRSR.** We summarize the bounds we obtained in the previous subsections. For sake of readability, we focus on the asymptotic Regime (a), in which $\rho \to 0$ and $\delta > 0$ is fixed, so that the functions $\Psi_{\min}^{(1)}$, $\Psi_{\max}^{(1)}$, $\Psi_{\min}^{(2)}$ and $\Psi_{\max}^{(2)}$ have a simplest expression.

| Inequality by | $\Psi_{\max}^{(1,2)}$ | $\Psi_{\min}^{(1,2)}$ |
|---|---|---|
| Davidson-Szarek | $2\sqrt{2}\sqrt{\rho\lvert\log\rho\rvert} + 2\sqrt{\rho} + \sqrt{2}(1-\log\delta)\sqrt{\frac{\rho}{\lvert\log\rho\rvert}} + o\Big(\sqrt{\frac{\rho}{\lvert\log\rho\rvert}}\Big)$ | $= \Psi_{\max}^{(2)}$ |
| Ledoux-Rider | $C_{\mathrm{LR}}\sqrt{\rho}\lvert\log\rho\rvert + \Big[2 + C_{\mathrm{LR}}\big(1-\log\delta\big)\Big]\sqrt{\rho} + o(\sqrt{\rho})$ | $= \Psi_{\max}^{(1)}$ |
| Feldheim-Sodin | $2\sqrt{\rho}\exp\Big[C_{\mathrm{FS}}^{2/3}\lvert\log\rho\rvert^{2/3}\Big] + o\Big(\sqrt{\rho}\exp\Big[C_{\mathrm{FS}}^{2/3}\lvert\log\rho\rvert^{2/3}\Big]\Big)$ | $= \Psi_{\max}^{(1)}$ |

We summarize next the conditions we obtained in the previous subsections on $\delta$ and $\rho$ so that Condition (SRSR-$\gamma(2s)$) is satisfied with high probability. For sake of readability again, this condition is written assuming that $\rho$ is small.

| Inequality by | Condition (SRSR-$\gamma(2s)$) |
|---|---|
| Davidson-Szarek | $\delta > \frac{1}{\rho}\exp\big[-\frac{\rho_0}{2\rho}\big]$ |
| Ledoux-Rider | $\delta > \frac{1}{\rho}\exp\big[-\frac{\tau_0/C_{\mathrm{LR}}}{\rho^{3/2}}\big]$ |
| Feldheim-Sodin | $\delta > \frac{1}{\rho}\exp\big[-\frac{\lvert\log\rho\rvert^{3/2}}{2C_{\mathrm{FS}}}\big]$ |

## Appendix A. Proofs of the main results

A.1. **Proof of Theorem 1.** The proof follows the same guidelines as [FR13, Proof of Theorem 6.13, Page 145]. A sufficient condition for SRSR is the $\ell_2$-robust null space property, see [FR13, Theorem 4.22, Page 88]. Namely, we need to find constants $\rho \in (0,1)$ and $\tau > 0$ such that, for any $v \in \mathbb{R}^p$ and any $S \subset \{1,\ldots,p\}$ such that $|S| = s$,

$$||v_S||_2 \leqslant \frac{\kappa}{\sqrt{s}}||v_{S^c}||_2 + \tau||\mathbf{M}v||_2\,.$$

Given $v \in \mathbb{R}^p$, it is enough to consider $S = S_0$ the set of the $s$ largest (in magnitude) entries of $v$, $S_1$ the set of the $s$ largest (in magnitude) entries of $v$ in $S_0^c$, $S_2$ the set of the $s$ largest (in magnitude) entries of $v$ in $(S_0 \cup S_1)^c$, etc. By definition of the RICs, one has

$$||\mathbf{M}v_{S_0}||_2^2 = (1{+}t)||v_{S_0}||_2^2$$
$$\text{with} \quad -\mathfrak{c}_{\min}(2s,\mathbf{M}) \leqslant -\mathfrak{c}_{\min}(s,\mathbf{M}) \leqslant t \leqslant \mathfrak{c}_{\max}(s,\mathbf{M}) \leqslant \mathfrak{c}_{\max}(2s,\mathbf{M})\,.$$

We begin with a first lemma. For sake of readability and from now on, $\mathfrak{c}_{\min}$ denotes $\mathfrak{c}_{\min}(2s,\mathbf{M})$ and $\mathfrak{c}_{\max}$ denotes $\mathfrak{c}_{\max}(2s,\mathbf{M})$.

**Lemma 4.** *For all $k \geqslant 1$, it holds*

$$|\langle \mathbf{M}v_{s_0}, \mathbf{M}v_{s_k}\rangle| \leqslant \sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}||v_{s_0}||_2||v_{s_k}||_2\,.$$

*Proof.* Set $u = v_{s_0}/||v_{s_0}||_2$ and $w = \pm v_{s_k}/||v_{s_k}||_2$ where the sign of $w$ is chosen so that $|\langle \mathbf{M}u, \mathbf{M}w\rangle| = \langle \mathbf{M}u, \mathbf{M}w\rangle$. For $\alpha, \beta > 0$ to be chosen later, it holds

$$\begin{aligned}
2|\langle \mathbf{M}u, \mathbf{M}w\rangle| &= \frac{1}{\alpha+\beta}\Big[||\mathbf{M}(\alpha u + w)||_2^2 - ||\mathbf{M}(\beta u - w)||_2^2 - (\alpha^2-\beta^2)||\mathbf{M}u||_2^2\Big]\\
&\leqslant \frac{1}{\alpha+\beta}\Big[(1+\mathfrak{c}_{\max})||\alpha u + w||_2^2 - (1-\mathfrak{c}_{\min})||\beta u - w||_2^2 - (\alpha^2-\beta^2)(1+t)||u||_2^2\Big]\\
&= \frac{1}{\alpha+\beta}\Big[(1+\mathfrak{c}_{\max})(\alpha^2+1) - (1-\mathfrak{c}_{\min})(\beta^2+1) - (\alpha^2-\beta^2)(1+t)\Big]\\
&= \frac{1}{\alpha+\beta}\Big[\alpha^2(\mathfrak{c}_{\max}-t) + \beta^2(\mathfrak{c}_{\min}+t) + \mathfrak{c}_{\max} + \mathfrak{c}_{\min}\Big]\,.
\end{aligned}$$

Then, chose $\alpha = \sqrt{(\mathfrak{c}_{\min}+t)/(\mathfrak{c}_{\max}-t)}$ and $\beta = \sqrt{(\mathfrak{c}_{\max}-t)/(\mathfrak{c}_{\min}+t)}$ to get the desired inequality. □

Using Lemma 4, observe that

$$\begin{aligned}
(1+t)||v_{S_0}||_2^2 &= ||\mathbf{M}v_{S_0}||_2^2\\
&= \langle \mathbf{M}v_{S_0}, \mathbf{M}v\rangle - \sum_{k\geqslant 1}\langle \mathbf{M}v_{S_0}, \mathbf{M}v_{S_k}\rangle\\
&\leqslant ||\mathbf{M}v_{S_0}||_2||\mathbf{M}v||_2 + \sum_{k\geqslant 1}\sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}||v_{s_0}||_2||v_{s_k}||_2\\
&= ||v_{S_0}||_2\Big[\sqrt{1+t}||\mathbf{M}v||_2 + \sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}\sum_{k\geqslant 1}||v_{s_k}||_2\Big]\,.
\end{aligned}$$

Now, Lemma 6.14 in [FR13] gives that

$$\sum_{k\geqslant 1}||v_{s_k}||_2 \leqslant \frac{1}{\sqrt{s}}||v_{S_0^c}||_1 + \frac{1}{4}||v_{S_0}||_2\,.$$

We deduce that

$$||v_{S_0}||_2 \leqslant \frac{b}{4}||v_{S_0}||_2 + \frac{b}{\sqrt{s}}||v_{S_0^c}||_1 + \frac{||\mathbf{M}v||_2}{\sqrt{1+t}}\,,$$

where $b := \sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}/(1+t)$. It follows that

$$||v_{S_0}||_2 \leqslant \frac{1}{\sqrt{s}}\frac{4b}{4-b}||v_{S_0^c}||_1 + \frac{||\mathbf{M}v||_2}{\sqrt{1+t}}\,.$$

It suffices that $\kappa = 4b/(4-b) < 1$ to get the $\ell_2$-robust null space property and hence SRSR. This is equivalent to $b = \sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}/(1+t) < 4/5$. We have the following lemma.

**Lemma 5.** *For any $t \in [-\mathfrak{c}_{\min}, \mathfrak{c}_{\max}]$, it holds*

$$\frac{\sqrt{(\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)}}{(1+t)} \leqslant \frac{\mathfrak{c}_{\min}+\mathfrak{c}_{\max}}{2\sqrt{(1-\mathfrak{c}_{\min})(1+\mathfrak{c}_{\max})}}\,.$$

*Proof.* Define $f(t) = (\mathfrak{c}_{\max}-t)(\mathfrak{c}_{\min}+t)/(1+t)^2$ whose derivative is given by

$$f'(t) = \frac{\mathfrak{c}_{\max}-\mathfrak{c}_{\min}-2\mathfrak{c}_{\max}\mathfrak{c}_{\min}-t(2+\mathfrak{c}_{\max}-\mathfrak{c}_{\min})}{(1+t)^3}$$

We easily deduce that the function $f$ is upper bounded by the quantity $f(t^\star)$ where we denote $t^\star = (\mathfrak{c}_{\max}-\mathfrak{c}_{\min}-2\mathfrak{c}_{\max}\mathfrak{c}_{\min})/(2+\mathfrak{c}_{\max}-\mathfrak{c}_{\min})$. Now, remark that it holds $f(t^\star) = (\mathfrak{c}_{\min}+\mathfrak{c}_{\max})^2/(4(1-\mathfrak{c}_{\min})(1+\mathfrak{c}_{\max}))$. This gives the desired inequality. □

It shows that SRSR holds whenever $(\mathfrak{c}_{\min}+\mathfrak{c}_{\max})/\sqrt{(1-\mathfrak{c}_{\min})(1+\mathfrak{c}_{\max})} < 8/5$. This last condition reads $\sqrt{\gamma}-1/\sqrt{\gamma} < 8/5$ which is equivalent to $\sqrt{\gamma} < (4+\sqrt{41})/5$, where we denote $\gamma = \gamma(2s, \mathbf{M})$. The desired condition follows.

A.2. **Proof of Theorem 2.** Let $t \in (0, \tau_0)$. A simple calculation gives that, on the event $\{\mathfrak{c}_{\max} < \sqrt{\rho}(2+\sqrt{\rho})+t\} \cap \{\mathfrak{c}_{\min} < \sqrt{\rho}(2-\sqrt{\rho})+t\}$, Condition (SRSR-$\gamma(2s)$) is satisfied whenever

$$(1-\gamma_0)+2(1+\gamma_0)\sqrt{\rho}+(1-\gamma_0)\rho+(1+\gamma_0)t < 0\,, \tag{15}$$

where $\gamma_0 := (4+\sqrt{41})^2/25$. Indeed, observe that, for all $0 < t < (1-\sqrt{\rho})^2$,

$$\begin{aligned}\gamma(2s,\mathbf{M}) - \frac{(4+\sqrt{41})^2}{25} &= \frac{1+\mathfrak{c}_{\max}}{1-\mathfrak{c}_{\min}} - \gamma_0 \\ &< \frac{(1+\sqrt{\rho})^2+t}{(1-\sqrt{\rho})^2-t} - \gamma_0 \\ &= \frac{(1-\gamma_0)+2(1+\gamma_0)\sqrt{\rho}+(1-\gamma_0)\rho+(1+\gamma_0)t}{(1-\sqrt{\rho})^2-t}\,.\end{aligned}$$

Fix $\rho$ and $\delta$ as in (6) and consider parameters $s$ and $n$ so that $s/n \to \rho$ 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 $n \geqslant n_0(\rho)$,

$$\begin{aligned}&\mathbb{P}\Big\{1+\mathfrak{c}_{\max} \geqslant (1+\sqrt{\rho})^2+t\Big\} \\ &\quad= \mathbb{P}\Big\{\exists x \in \Sigma_s \text{ s.t. } n^{-1}\|\mathbf{M}x\|^2 \geqslant \big((1+\sqrt{\rho})^2+t\big)\|x\|^2\Big\} \\ &\quad\leqslant \sum_{\mathbf{C}_{s,n}} \mathbb{P}\Big\{\lambda_1(\mathbf{C}_{s,n}) \geqslant (1+\sqrt{\rho})^2+t\Big\} \\ &\quad\leqslant \binom{p}{s} c(\rho) e^{-n\mathbb{W}(\rho,t)} \\ &\quad\leqslant c(\rho)\Theta\, e^{-n\mathbb{W}(\rho,t)+p\mathbf{H}_e(s/p)} \\ &\quad\leqslant c(\rho)\Theta\, e^{-nD},\end{aligned}$$

with $\mathbf{H}_e(t) = -t\log t - (1-t)\log(1-t)$ for $t \in (0,1)$, $\Theta^2 := e^{1/2}/(2\pi[s(1-s/p)]^{1/p})$ and $D = \mathbb{W}(\rho, t) - \frac{1}{\delta}\mathbf{H}_e(\rho\delta)$. Indeed, note that Stirling formula (see Lemma 10) leads to

$$\binom{p}{s} \leqslant \frac{e^{1/4}}{\sqrt{2\pi}[s(1-s/p)]^{1/(2p)}} e^{-s\log(s/p)-(p-s)\log(1-s/p)} = \Theta e^{p\mathbf{H}_e(s/p)}\,.$$

where $\Theta \to e^{1/4}/\sqrt{2\pi}$ when $0 < s/p < 1$ and $p$ goes to infinity.
Denote $\mathbb{W}^{-1}(\rho, .)$ the inverse of the function $\mathbb{W}(\rho, .)$. Now, recall that $D + \frac{1}{\delta}\mathbf{H}_e(\rho\delta)$ belongs to the range of the function $\mathbb{W}(\rho, .)$ and consider

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

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

$$\mathbb{P}\{\mathfrak{c}_{\max} \geqslant \sqrt{\rho}(2+\sqrt{\rho}) + t_D\} \leqslant c(\rho)\Theta e^{-nD},$$

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

We now prove that (SRSR-$\gamma(2s)$) holds. Note that $\mathbb{W}^{-1}(\rho, .)$ is continuous and increasing on the range of $\mathbb{W}(\rho, .)$ and set

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

Consider $\rho$ and $\delta$ such that

$$t_0 < t_D < -\frac{(1-\gamma_0) + 2(1+\gamma_0)\sqrt{\rho} + (1-\gamma_0)\rho}{1+\gamma_0}\,.$$

Applying the increasing function $\mathbb{W}(\rho, \cdot)$, we get that

$$0 < \frac{1}{\delta}\mathbf{H}_e(\rho\delta) < \mathbb{W}(\rho, t_D) < \mathbb{W}\Big(\rho, -\frac{(1-\gamma_0) + 2(1+\gamma_0)\sqrt{\rho} + (1-\gamma_0)\rho}{1+\gamma_0}\Big) \tag{16}$$

Using $\mathbf{H}_e(t) \leqslant -t\log t + t$ , this last inequality is implied by

$$\delta > \frac{1}{\rho}\exp\Big[1 - \frac{\mathbb{W}\Big(\rho, -\frac{(1-\gamma_0)+2(1+\gamma_0)\sqrt{\rho}+(1-\gamma_0)\rho}{1+\gamma_0}\Big)}{\rho}\Big]\,,$$

which is (6) observing that $\tau_0 = (\gamma_0 - 1)/(\gamma_0 + 1)$ and $\tau_0(\sqrt{\rho_0} + 1/\sqrt{\rho_0}) = 2$. Furthermore, (16) shows that $\frac{1}{\delta}\mathbf{H}_e(\rho\delta)$ belongs to the interior of the range of $\mathbb{W}(\rho, .)$. Invoke the continuity of $\mathbb{W}^{-1}(\rho, .)$ 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 $\mathbb{W}^{-1}(\rho, .)$, we deduce that (16) is equivalent to

$$(1-\gamma_0) + 2(1+\gamma_0)\sqrt{\rho} + (1-\gamma_0)\rho + (1+\gamma_0)t_0 < 0. \tag{17}$$

Then for $D > 0$ small enough it holds

$$(1-\gamma_0) + 2(1+\gamma_0)\sqrt{\rho} + (1-\gamma_0)\rho + (1+\gamma_0)t_D < 0\,, \tag{18}$$

by continuity. It follows that Condition (SRSR-$\gamma(2s)$) fails with a probability smaller than $2c(\rho)\Theta e^{-nD}$ for some small enough $D > 0$ that may depend on $\delta$ and $\rho$. Observe that $D - \log(\Theta)/n \geqslant D/2$ for large enough $n$ then, changing $D$ by $D/2$, we can prove that the probability of failure can be as small as $2c(\rho)e^{-nD}$. We conclude that Condition (SRSR-$\gamma(2s)$) holds under Condition (18) on $\rho$ and $\delta$ which is implied by (6).

A.3. **Proof of Theorem 3.** We follow the same lines as in the previous proof. Here, the conditioning event is $\{\mathfrak{c}_{\max} < (1+\sqrt{\rho}+t)^2-1\}\cap\{\mathfrak{c}_{\min} < 1-(1-\sqrt{\rho}-t)^2\}$. First, note that, by a similar argument as in the previous proof,

$$\begin{aligned}\mathbb{P}\Big\{1+\mathfrak{c}_{\max} \geqslant (1+\sqrt{\rho}+t)^2\Big\}\\ &= \mathbb{P}\Big\{\exists x \in \Sigma_s \text{ s.t. } n^{-1}\|\mathbf{M}x\|^2 \geqslant (1+\sqrt{\rho}+t)^2\|x\|^2\Big\}\\ &\leqslant \sum_{\mathbf{C}_{s,n}} \mathbb{P}\Big\{\sigma_1 \geqslant 1+\sqrt{\rho}+t\Big\}\,,\\ &\leqslant c(\rho)\Theta\, e^{-nD},\end{aligned}$$

where the sum is over all choices of $s$ columns over the $p$ in $\mathbf{M}$.
Then, (15) becomes

$$1-\sqrt{\gamma_0}+(1+\sqrt{\gamma_0})\sqrt{\rho}+(1+\sqrt{\gamma_0})t<0\,,$$

where $\sqrt{\gamma_0}=(4+\sqrt{41})/5$. Indeed, observe that

$$\frac{1+\mathfrak{c}_{\max}}{1-\mathfrak{c}_{\min}} < \Big(\frac{1+\sqrt{\rho}+t}{1-\sqrt{\rho}-t}\Big)^2\,,$$

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

$$\frac{1+\sqrt{\rho}+t}{1-\sqrt{\rho}-t} < \sqrt{\gamma_0}\,.$$

This inequality can be equivalently written as

$$t < \frac{\sqrt{\gamma_0}-1}{\sqrt{\gamma_0}+1}-\sqrt{\rho}\,.$$

From this, Eq. (16) becomes

$$0=\mathbb{W}(\rho,0) < \frac{1}{\delta}\mathbf{H}_e(\rho\delta) < \mathbb{W}\big(\rho, \frac{\sqrt{\gamma_0}-1}{\sqrt{\gamma_0}+1}-\sqrt{\rho}\big)\,.$$

Using again $\mathbf{H}_e(t) \leqslant -t\log t + t$ , this last inequality is implied by

$$\delta > \frac{1}{\rho}\exp\Big[1-\frac{\mathbb{W}\Big(\rho, \frac{\sqrt{\gamma_0}-1}{\sqrt{\gamma_0}+1}-\sqrt{\rho}\Big)}{\rho}\Big]\,,$$

which is (7) noticing that $(\sqrt{\gamma_0}-1)/(\sqrt{\gamma_0}+1)=\sqrt{\rho_0}$.

## Appendix B. 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 $C_{\mathrm{FS}}$ 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 $C_{\mathrm{FS}}$ 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 the following. 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 6.** *Let $N > M \geqslant 54$ 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*

$$\mathbb{P}\Big\{\lambda_M(\mathbf{C}) \geqslant (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\Big\} \leqslant \frac{\mathbb{W}_0(\rho,\varepsilon)}{1-\rho} M \exp(-N\mathbb{W}_{\mathrm{FS}}(\rho,\varepsilon))$$
$$\mathbb{P}\Big\{\lambda_1(\mathbf{C}) \leqslant (\sqrt{M}-\sqrt{N})^2 - \varepsilon N\Big\} \leqslant \frac{\mathbb{W}_0(\rho,\varepsilon)}{1-\rho} M \exp(-N\mathbb{W}_{\mathrm{FS}}(\rho,\varepsilon))$$

*where $\rho = M/N$ and*

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

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

$$\mathbb{P}\Big\{\lambda_M(\mathbf{C}) \geqslant (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\Big\} \leqslant v \exp\Big(-C^{-1}N\frac{\rho^{1/4}}{(1+\sqrt{\rho})^2}\varepsilon^{\frac{3}{2}}\Big)$$
$$\mathbb{P}\Big\{\lambda_1(\mathbf{C}) \leqslant (\sqrt{M}-\sqrt{N})^2 - \varepsilon N\Big\} \leqslant v \exp\Big(-C^{-1}N\frac{\rho^{1/4}}{(1+\sqrt{\rho})^2}\varepsilon^{\frac{3}{2}}\Big).$$

B.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

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

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

(2) The control of

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

(resp. $B_m := \mathbb{E}\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m}] - \mathbb{E}\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m-1}]$) 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 $\mathbb{E}\mathfrak{tr}[Q_n(\mathbf{C})]$, 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 7 provides an upper bound on this quantity. Summing over $s$, it yields an upper bound on $\hat{\Sigma}_1^1(n)$, see (19) in Lemma 8.

(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.

B.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 7.** *It holds, for all $s \geqslant 1$,*

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

*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 10, 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

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

with

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

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

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

• 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 bound on $D_1$ is

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

Note that this number is 2 when $h = 1$. For $h \geqslant 2$, using Lemma 10, 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

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

with

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

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 $\max(\gamma(2), \lim_{h\to\infty}\gamma(h))$. This yields $\gamma(h) \leqslant -0.33$ for all $h \geqslant 2$. Finally, the number of diagrams in this case is upper bounded by (recall that $s = h$ here)

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

• 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 10, 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

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

with

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

∘ 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 $\beta(t) \leqslant 3.985$ 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, $\alpha(t) \leqslant \max(\alpha(1/s), \alpha(1-2/s))$. Computing these two values and using the fact that $s \geqslant 3$ leads to $\alpha(t) \leqslant \alpha(1-2/s)$ for all $t \in (1/s, 1-2/s)$. Consequently

$$\alpha(t) \leqslant 2\log s - \frac{1}{2}\log\Big(2+\frac{2}{s}\Big) - \frac{1}{2}\log\Big(1-\frac{2}{s}\Big) - \frac{1}{2}\log\Big(1+\frac{2}{s}\Big).$$

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

$$\begin{aligned}
\gamma &\leqslant \Big(3s - \frac{3}{2}\Big)\log\Big(1-\frac{1}{3s-1}\Big) + s\log\Big(1-\frac{1}{3s}\Big)\\
&\quad - \frac{1}{2}(s+2)\log\Big(1+\frac{2}{s-1}\Big) - \Big(s-\frac{1}{2}\Big)\log\Big(1-\frac{1}{s}\Big).
\end{aligned}$$

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

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

with

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

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

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

diagrams in this case. This leads to the result. □

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

**Lemma 8.** *It holds*

$$\hat{\Sigma}_1^1(n)\leqslant C_{0,\hat{\Sigma}}n(MN)^{n/2}\exp\Big[\frac{C_{\hat{\Sigma}}(1+\sqrt{M/N})n^{3/2}}{\sqrt{M}}\Big] \tag{19}$$

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

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

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

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

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

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

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

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

where

$$T_1:=\sum_{s=1}^{n}D_1(s)(M^{-1/2}+N^{-1/2})^{2s-2}\frac{(3s+1)}{(3s-2)!}\Big(\frac{n-3s+1}{2}+3s-2\Big)^{3s-2}$$
$$T_2:=\sum_{s=1}^{n}D_1(s)(M^{-1/2}-N^{-1/2})^{2s-2}\frac{(3s+1)}{(3s-2)!}\Big(\frac{n-3s+1}{2}+3s-2\Big)^{3s-2}$$

We can bound each term. It reads as follows.

$$\begin{aligned}T_1 &\leqslant C_{0,D}\sum_{s=1}^{n} C_D^{s-1}s^{s-1/2}(M^{-1/2}+N^{-1/2})^{2s-2}\frac{(3s+1)}{(3s-2)!}\Big(\frac{n-3s+1}{2}+3s-2\Big)^{3s-2}\\ &\leqslant C_{0,D}\sum_{s=1}^{n} C_D^{s-1}\Big[\frac{1+\sqrt{M/N}}{\sqrt{M}}\Big]^{2(s-1)}\frac{(3s+1)(n+3s-3)^{3s-2}s^{s-1/2}}{(3s-2)!2^{3s-2}}\end{aligned}$$

using Lemma 7. Invoke Lemma 10 to get that

$$\begin{aligned}&\frac{(2(s-1))!}{n^{3(s-1)}}\frac{(3s+1)(n+3s-3)^{3s-2}s^{s-1/2}}{(3s-2)!2^{3s-2}}\\ &\qquad\leqslant n\frac{e^{s+1/12}}{2^{3s-2}}(3s+1)\sqrt{\frac{2s-2}{3s-2}}\Big(1+\frac{3s-3}{n}\Big)^{3s-2}\frac{(2s-2)^{2s-2}s^{s-1/2}}{(3s-2)^{3s-2}}\\ &\qquad\leqslant ne^{1/12}\sqrt{\frac{2s-2}{3s-2}}2^{3s-2}(3s+1)\frac{(2s-2)^{2s-2}s^{s-1/2}}{(3s-2)^{3s-2}}e^s.\end{aligned}$$

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

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

Some elementary computations give the following:

$$\begin{aligned}f(s) &= \frac{1}{2}\log s+(5\log 2-3\log 3)s+3\log 3-4\log 2+(2s-2)\log\Big(1-\frac{1}{s}\Big)\\ &\qquad -(3s-2)\log\Big(1-\frac{2}{3s}\Big)+\log\Big(1+\frac{1}{3s}\Big)\\ &= (5\log 2-3\log 3)s+3\log 3-4\log 2+g(s),\end{aligned}$$

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

$$\begin{aligned}g'(s) &= \frac{3s-1}{2s(3s+1)}+2\log\Big(1-\frac{1}{s}\Big)-3\log\Big(1-\frac{2}{3s}\Big),\\ g''(s) &= \frac{-27s^4+99s^3-21s^2+11s+2}{2s^2(s-1)(3s-2)(3s+1)^2}.\end{aligned}$$

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 $s\in[1,s_*]$, $g(s)\leqslant g(1)=2\log 2$. 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, $g'(s_0)\leqslant g'(39.66)\leqslant 0.013$. This leads to

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

Then

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

As a consequence,

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

Similarly, one gets

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

This yields the result. □

B.4. **Bound on the traces.**

**Lemma 9.** *It holds that*

$$\Big(\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m}]]+\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m-1}]]\Big)\vee\Big(\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m}]]-\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m-1}]]\Big)\leqslant\Delta_m \tag{20}$$

*where*

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

*and*

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

*Proof.* Invoke Lemma IV.1.1 Page 115 in [FS10] and Lemma 8 to get that

(21)

$$\mathbb{E}[\mathfrak{tr}[V_{n,\frac{(M-2)^2}{(M-1)(N-1)}}(\widetilde{\mathbf{C}})]]\leqslant C_{0,\hat\Sigma}\Big(\frac{MN}{(M-1)(N-1)}\Big)^{n/2}n\exp\Big(C_{\hat\Sigma}(1+\sqrt{M/N})\frac{n^{3/2}}{M^{1/2}}\Big).$$

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

$$
\begin{aligned}
A_m &= \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^{m} (2n+1) \binom{2m+1}{m-n} \mathbb{E}[\mathfrak{tr}[U_{2n}(\widetilde{\mathbf{C}})]] \\
&\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^{m} 2n \binom{2m}{m-n} \mathbb{E}[\mathfrak{tr}[U_{2n-1}(\widetilde{\mathbf{C}})]].
\end{aligned} \tag{22}
$$

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

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

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

$$
\begin{aligned}
A_m &\leqslant \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^{m} (2n+1) \binom{2m+1}{m-n} \sum_{k=0}^{n} s^k \mathbb{E}[\mathfrak{tr}[V_{2(n-k),s}(\widetilde{\mathbf{C}})]] \\
&\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^{m} 2n \binom{2m}{m-n} \sum_{k=0}^{n-1} s^k \mathbb{E}[\mathfrak{tr}[V_{2n-2k-1,s}(\widetilde{\mathbf{C}})]] \\
&\leqslant \frac{1}{(2m+1)2^{2m}} \sum_{n=1}^{m} (2n+1) \binom{2m+1}{m-n} \Big( \sum_{k=0}^{n-1} s^k \mathbb{E}[\mathfrak{tr}[V_{2(n-k),s}(\widetilde{\mathbf{C}})]] + s^n M \Big) \\
&\quad + \frac{1}{2m2^{2m}} \sum_{n=0}^{m} 2n \binom{2m}{m-n} \sum_{k=0}^{n-1} s^k \mathbb{E}[\mathfrak{tr}[V_{2n-2k-1,s}(\widetilde{\mathbf{C}})]]
\end{aligned} \tag{23}
$$

Invoke (21) to get with $C_{M,N}=C_{\hat\Sigma}(1+\sqrt{M/N})$,

$$\begin{aligned}
A_m &\leqslant \sum_{n=1}^{m}\frac{2n+1}{(2m+1)2^{2m}}\binom{2m+1}{m-n}\\
&\quad\times\sum_{k=0}^{n-1}s^kC_{0,\hat\Sigma}\Big(\frac{MN}{(M-1)(N-1)}\Big)^{n-k}2(n-k)\exp\Big[C_{M,N}\frac{2^{\frac32}(n-k)^{\frac32}}{M^{\frac12}}\Big]\\
&\quad+\sum_{n=1}^{m}\frac{n}{m2^{2m}}\binom{2m}{m-n}\\
&\quad\times\sum_{k=0}^{n-1}s^kC_{0,\hat\Sigma}\Big(\frac{MN}{(M-1)(N-1)}\Big)^{n-k-\frac12}2\Big(n-k-\frac12\Big)\exp\Big[C_{M,N}\frac{2^{\frac32}(n-k-1/2)^{\frac32}}{M^{\frac12}}\Big]\\
&\quad+\frac{1}{(2m+1)2^{2m}}\sum_{n=0}^{m}(2n+1)\binom{2m+1}{m-n}s^nM\\
&\leqslant\frac{2C_{0,\hat\Sigma}}{1-\frac{(M-2)^2}{MN}}\sum_{n=1}^{m}\Big[\frac{2n+1}{(2m+1)2^{2m}}\binom{2m+1}{m-n}+\frac{n}{m2^{2m-1}}\binom{2m}{m-n}\Big]\\
&\quad\times n\Big(\frac{MN}{(M-1)(N-1)}\Big)^{n}\exp\Big[C_{M,N}\frac{2^{\frac32}n^{\frac32}}{M^{\frac12}}\Big]\\
&\quad+\frac{1}{(2m+1)2^{2m}}\sum_{n=0}^{m}(2n+1)\binom{2m+1}{m-n}s^nM.
\end{aligned}$$

From Lemma 11 it holds

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

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

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

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

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

where

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

as claimed.

The bound on $B_m:=\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m}]]-\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m-1}]]$ follows the same lines. The minus in front of $\mathbb{E}[\mathfrak{tr}[\widetilde{\mathbf{C}}^{2m-1}]]$ change the line (22) to its opposite. The change of indices $k$ leads to the term $s^{k+1/2}\mathbb{E}[\mathfrak{tr}[V_{2(n-k-1),s}(\widetilde{\mathbf{C}})]]$ in (23). 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. $\square$

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

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

with

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

for all $N>M\geqslant 54$. 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

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

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

$$\begin{aligned}
\mathbb{P}(\lambda_M(\widetilde{\mathbf{C}}) \geqslant \varepsilon_{M,N}) &\leqslant \mathbb{P}(f(\lambda_M(\widetilde{\mathbf{C}}))+e_m \geqslant f(\varepsilon_{M,N})+e_m)\\
&\leqslant \frac{\mathbb{E}[f(\lambda_M(\widetilde{\mathbf{C}})]+e_m}{f(\varepsilon_{M,N})+e_m}\\
&\leqslant \frac{\sum_{k=1}^M(\mathbb{E}[f(\lambda_k(\widetilde{\mathbf{C}})]+e_m)}{f(\varepsilon_{M,N})}\\
&= \frac{A_m+Me_m}{f(\varepsilon_{M,N})} \qquad (24)
\end{aligned}$$

Invoke Lemma 9 to get that

$$\mathbb{P}(\lambda_M(\widetilde{\mathbf{C}}) \geqslant \varepsilon_{M,N}) \leqslant \frac{C_{0,\mathrm{Rad}}m\Big[\Big(\frac{MN}{(M-1)(N-1)}\Big)^m+\frac{M}{m}\Big]\exp\Big(C_{\mathrm{Rad}}(1+\sqrt{M/N})^4\frac{m^3}{M^2}\Big)+\frac{M}{2em}}{(1-\frac{M}{N})f(\varepsilon_{M,N})}\,,$$

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

$$\mathbb{P}(\lambda_M(\widetilde{\mathbf{C}}) \geqslant \varepsilon_{M,n}) \leqslant \frac{C_{0,\mathrm{Rad}}\big[m+\frac{1+2e}{2e}M\big]\,e^{C_{\mathrm{Rad}}(1+\sqrt{M/N})^4\frac{m^3}{M^2}+54m(\frac{1}{M}+\frac{1}{N})\log(\frac{54}{53})}}{(1-\frac{M}{N})f(\varepsilon_{M,N})}$$

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

$$\mathbb{P}\Big\{\lambda_M(\mathbf{C}) \geqslant (\sqrt{M}+\sqrt{N})^2+\varepsilon N\Big\} \leqslant \frac{\mathbb{W}_0(\rho,\varepsilon)}{1-\rho}M\exp(-N\mathbb{W}_1(\rho,\varepsilon))$$

where $\rho = M/N$ and

$$\mathbb{W}_0(\rho,\varepsilon) := \frac{C_{0,\mathrm{Rad}}(1+2e)\sqrt{3C_{\mathrm{Rad}}}(1+\sqrt{\rho})^2 + 2e\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{2e\sqrt{3C_{\mathrm{Rad}}}(1+\sqrt{\rho})^2} \times \exp\left[54\log(\frac{54}{53})(1+\rho)\frac{\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{(1+\sqrt{\rho})^2\sqrt{3C_{\mathrm{Rad}}}}\right]$$

$$\mathbb{W}_1(\rho,\varepsilon) := \frac{4\sqrt{2}}{3\sqrt{3}}\frac{\rho\log(1+\frac{\varepsilon}{2\sqrt{\rho}})^{\frac{3}{2}}}{(1+\sqrt{\rho})^2\sqrt{C_{\mathrm{Rad}}}}$$

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

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

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

$$\mathbb{P}\Big\{\lambda_M(\mathbf{C}) \geqslant (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\Big\} \leqslant \frac{c_0 M e^{c_0\sqrt{\log\left(1+\frac{\varepsilon}{2\sqrt{\rho}}\right)}}}{1-\rho} e^{-N\mathbb{W}_1(\rho,\varepsilon)}\,. \tag{25}$$

For $\varepsilon \leqslant \sqrt{\rho}$ 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 $M \geqslant 1$, it holds $M \leqslant c_1(\eta)\exp(\eta M)$. Note that $\log(3/2)\frac{\varepsilon}{\sqrt{\rho}} \leqslant \log\left(1+\frac{\varepsilon}{2\sqrt{\rho}}\right)$ and set

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

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

$$\mathbb{P}\Big\{\lambda_M(\mathbf{C}) \geqslant (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\Big\} \leqslant v\exp\Big(-C^{-1}N\frac{\rho^{1/4}}{(1+\sqrt{\rho})^2}\varepsilon^{\frac{3}{2}}\Big)\,. \tag{26}$$

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

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

with

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

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

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

and we recover an upper bound of the form (24) for which Lemma 9 can also be applied and we get that

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

for all $m \in \mathbb{N}$. The rest of the proof follows the same lines as in Section B.5 where we change $\varepsilon_{M,N}$ by $\varepsilon'_{M,N}$, we choose $m = \sqrt{\frac{2\log(54\varepsilon'_{M,N}/53)}{3C_{\mathrm{Rad}}(1+\sqrt{M/N})^4}}M$ and may have changed the harmless constant $c_0$ in $\mathbb{W}_0$. Eventually, note that (26) has been obtained from (25) and we can use the same argument for the deviation on the smallest eigenvalue. This proves Proposition 6.

## Appendix C. Stirling's formula and bounds on binomial coefficients

**Lemma 10.** *Let $z > 0$ then there exists $\theta \in (0,1)$ such that:*

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

*Proof.* See [AS65] Eq. 6.1.38. □

**Lemma 11.** *It holds, for all $1 \leqslant n \leqslant m$,*

$$\log\Big[\frac{n}{2^{2m}}\binom{2m}{m-n}\Big] \leqslant 5 - 0.6321\frac{n^2}{m}$$
$$\log\Big[\frac{n+1/2}{2^{2m}}\binom{2m+1}{m-n}\Big] \leqslant 2 - 0.6555\frac{n^2}{m}$$

*Proof.* If $n = m$ then the result is clear. Otherwise, using Lemma 10, one has

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

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

$$m\Big[\frac{n}{m}\log(1-\frac{2n/m}{1+n/m})-\log(1-(\frac{n}{m})^2)\Big]\leqslant -\frac{n^2}{m}\,.$$

Let $x=n/m$ and observe that $x\leqslant 1-1/m$. 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 $x\leqslant 0.99995$ then, using that $\log(z)\leqslant z/e$, we have

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

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

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

In all cases, we get that

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

We deduce that

$$\log\Big[\frac{n}{2^{2m}}\binom{2m}{m-n}\Big]\leqslant 4.24-0.6321n^2/m\,,$$

as claimed. □

## References


[ADCM17] J.-M. Azaïs, Y. De Castro, and S. Mourareau, *A rice method proof of the null-space property over the grassmannian*, Bernoulli (2017).

[ALMT14] D. Amelunxen, M. Lotz, M.B. McCoy, and J. A. Tropp, *Living on the edge: phase transitions in convex programs with random data*, Inf. Inference **3** (2014), no. 3, 224–294.

[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)

[BT10] Bubacarr Bah and Jared Tanner, *Improved bounds on restricted isometry constants for gaussian matrices*, SIAM Journal on Matrix Analysis and Applications **31** (2010), no. 5, 2882–2898.

[BT14] B. Bah and J. Tanner, *Bounds of restricted isometry constants in extreme asymptotics: formulae for Gaussian matrices*, Linear Algebra Appl. **441** (2014), 88–109.

[Can08] Emmanuel J. Candes, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 9-10, 589–592. MR 2412803 (2009b:65104)

[CGLP12] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor, *Interaction between compressed sensing, random matrices and high dimensional geometry*, Panoramas et synthéses, 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 lq-minimization for 0<q<=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 l1-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)

[MT14] M. B. McCoy and J. A. Tropp, *Sharp recovery bounds for convex demixing, with applications*, Found. Comput. Math. **14** (2014), no. 3, 503–567.

[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.

[PY14] N.S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), no. 3, 935–1001.

[RV08] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), no. 8, 1025–1045. MR MR2417886 (2009e:94034)

[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.

[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