SPARSE RECOVERY FROM EXTREME EIGENVALUES DEVIATION INEQUALITIES

SANDRINE DALLAPORTA AND YOHANN DE CASTRO

ABSTRACT. This article provides a new toolbox to derive sparse recovery guarantees—that is referred to as "stable and robust sparse regression" (SRSR)—from deviations on extreme singular values or extreme eigenvalues obtained in Random Matrix Theory. This work is based on Restricted Isometry Constants (RICs) which are a pivotal notion in Compressed Sensing and High-Dimensional Statistics as these constants finely assess how a linear operator is conditioned on the set of sparse vectors and hence how it performs in 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 deviation estimates on their extreme eigenvalues. This allows us to derive a lower bound on the probability of getting SRSR. One benefit of this paper is a direct and explicit derivation of upper bounds on RICs and lower bounds on SRSR from deviations on the extreme eigenvalues given by Random Matrix theory.

1. INTRODUCTION

1.1. **Stable and Robust Sparse Recovery (SRSR).** The recent breakthrough of Compressed Sensing [14, 16, 15] has shown that one can acquire and compress simultaneously a signal from few linear measurements. This methodology has an important impact in practice since it may be deployed in applied contexts where the time of acquisition is limited—*e.g.* medical imaging (MRI and functional MRI)—and/or costly—*e.g.* reflection seismology, one may consult [18, 27] and references therein.

More precisely, the problem addressed in recent researches aims at solving under-determined systems of linear equations (with an additive error term $\mathbf{e}$) of the form

$$(1.1) \qquad y = \mathbf{M}x_0 + \mathbf{e}$$

where $\mathbf{M}$ is a known $(n \times p)$ matrix, x_0 an 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 ℓ_1-minimization solutions as

$$(1.2) \qquad \hat{x} \in \arg \min \|x\|_1 \quad \text{s.t.} \quad \|y - \mathbf{M}x\|_2 \leq \eta \,,$$

where $\eta > 0$ is a tuning parameter such that the experimenter believes it holds $\|\mathbf{e}\|_2 \leq \eta$ with high probability.

A standard goal is to prove that the estimate $\hat{x}$ is close to x_0 or more precisely that the norm of the error $\hat{x} - x_0$ is small. To this purpose, one says that $\hat{x}$ satisfies the Stable and Robust Sparse Recovery (SRSR, see [27, page 88]) if the following two error bounds hold

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

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

where $C, D > 0$ are constants and $\sigma_s(x_0)_1$ denotes the approximation error in ℓ_1-norm by s coefficients, namely

$$\sigma_s(x_0)_1 := \min \|x_0 - x\|_1 \,,$$

where the minimum is taken over the space Σ_s of sparse vectors x, *i.e.* the set of vectors with at most s nonzero coordinates. SRSR shows that the estimate $\hat{x}$ recovers the s largest coefficients of the target vector x_0 in a stable[1] and robust (to additive errors $\mathbf{e}$) manner. One can prove that the $\sqrt{s}$ terms are optimal in the sense of Approximation Theory, one may consult [27, Remark 4.2.3 page 88] for further details. Sufficient condition for SRSR holds whenever the matrix X satisfies some properties, see for instance [15, 17, 26, 10, 44, 9, 29, 20] or [18, 27].

1.2. **Main Result: a Toolbox to get SRSR from Deviation Inequalities.** In this paper, we want to guarantee SRSR when the matrix $\mathbf{M}$ is chosen at random and $\hat{x}$ is produced by an ℓ_1-minimization estimate such as (1.2).

The first random model we assume for $\mathbf{M}$ is as follows. Let R be a subset of $\{1, \ldots, p\}$ of size r and denote $\mathbf{M}_R$ the $(n \times r)$ matrix obtained by keeping the columns of $\mathbf{M}$ that belongs to R and $\mathbf{M}_R^\star$ its hermitian adjoint. Assume that all the Gram matrices $\mathbf{M}_R^\star \mathbf{M}_R$ satisfies

(1.3) $\forall r \in \{1, \ldots, \lceil 0.1230\, n \rceil\}, \ \forall R \subseteq \{1, \ldots, p\} \text{ s.t. } |R| = r, \quad \mathbf{M}_R^\star \mathbf{M}_R \sim \mathbf{C}_{r,n}\,,$

namely $\mathbf{M}_R^\star \mathbf{M}_R$ are identically distributed with respect to the law of the random covariance matrix $\mathbf{C}_{r,n}$ of size $(r \times r)$ that may depend on the parameter n also. As we will see in the sequel, standard models of covariance matrices $\mathbf{C}_{r,n}$ are given by $\mathbf{C}_{r,n} = (1/n)\,\mathbf{X}\mathbf{X}^\star$ where $\mathbf{X} \in \mathbb{R}^{r \times n}$ has *iid* entries drawn with respect to a law $\mathcal{L}$. Note that we do not require that $\mathbf{M}$ has independent entries here, we only assume that the matrices $\mathbf{M}_R \mathbf{M}_R^\star$ are identically distributed and that we have access to the rate function of the deviations on its extreme eigenvalues.

Guaranteeing SRSR with sparsity parameter s requires to bound the eigenvalues of $\mathbf{M}_R^\star \mathbf{M}_R$ when $r = 2s$, see for instance Property (SRSR-$\gamma(2s)$). We consider the asymptotic proportional growth model where $s/n \to \rho$, *i.e.* we assume that size of the sparse vectors over number of equations tends to a constant. We also consider the parameter $\bar{\rho} = r/n \in (0, 0.1230)$ and since $r = 2s$ we remark that $\bar{\rho} = 2\rho$. We assume that we have access to a deviation inequality on extreme eigenvalues of $\mathbf{C}_{r,n}$ with rate function $t \mapsto \mathbb{W}(\bar{\rho}, t)$ depending on this ratio. For instance, we will consider that for all $n \geq n_0(\bar{\rho})$,

(1.4) $\forall t \in [0, 0.6247), \quad \mathbb{P}\big\{\big(\lambda_1 - (1 + \sqrt{\bar{\rho}})^2\big) \vee \big((1 - \sqrt{\bar{\rho}})^2 - \lambda_r\big) \geq t\big\} \leq c(\bar{\rho})e^{-n\mathbb{W}(\bar{\rho},t)}$

where $n_0(\bar{\rho}) \geq 2$ and $c(\bar{\rho}) > 0$ may depend on the ratio $\bar{\rho}$, the function $t \mapsto \mathbb{W}(\bar{\rho}, t)$ is continuous and increasing on $[0, 0.6247)$ such that $\mathbb{W}(\bar{\rho}, 0) = 0$, λ_1 and λ_r are respectively the largest and the least eigenvalues of $\mathbf{C}_{r,n}$. Remark that the "bulk" bounds $(1 \pm \sqrt{\bar{\rho}})^2$ are prescribed by the Marchenko-Pastur law. Consider the asymptotic proportional growth model for $\mathbf{M}$ so that

---

[1] 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 [27, page 82] for instance.
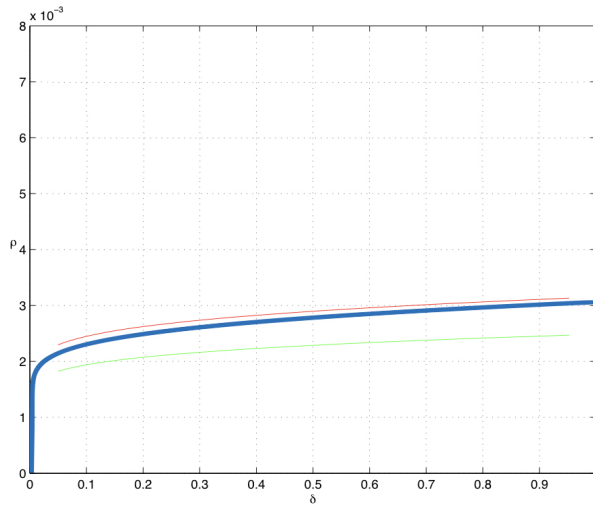
FIGURE 1. **[Bounds on SRSR in the *iid* Gaussian case]** The region below the curves gives pairs of $(\delta = n/p, \rho = s/n)$ for which SRSR holds with overwhelming probability when $\mathbf{M}$ has *iid* Gaussian entries. Our new bound (1.6) in blue is comparable to the one of [11] in red (derived using Foucart and Lai condition [26, Theorem 2.1]) and the one of [14] in green (derived from symmetric RICs bounds as in (FR-$\mathfrak{c}(2s)$)). This figure is an update of Figure 3.2 in [11].

$n/p \to \delta$, *i.e.* the number of equations over number of unknowns tends also to a constant. The main contribution of this paper is to give a bound on (δ, ρ) from the rate function $\mathbb{W}$ so that SRSR holds with overwhelming probability.

The second random model covered by this paper is given by matrices $\mathbf{M}$ satisfying the following property

$$(1.5) \qquad \forall r \in \{1, \dots, \lceil 0.1230\,n \rceil\}, \ \forall R \subseteq \{1, \dots, p\} \ \text{s.t.} \ |R| = r, \quad \mathbf{M}_R \sim \mathbf{X}_{r,n},$$

namely $\mathbf{M}_R$ are identically distributed with respect to the law of the random rectangular matrix $\mathbf{X}_{r,n}$ of size $(n \times r)$. Standard models of rectangular matrices $\mathbf{X}_{r,n}$ are given by $\mathbf{X}_{r,n} = (1/\sqrt{n})\,\mathbf{X}^\star$ where $\mathbf{X} \in \mathbb{R}^{r \times n}$ has *iid* entries drawn with respect to a law $\mathcal{L}$. As in the first model, we do not require that $\mathbf{M}$ has independent entries here, we only assume that the matrices $\mathbf{M}_R$ are identically distributed. In this case, we assume that we have a deviation inequality with rate function $\mathbb{W}$ on the extreme singular values of $\mathbf{X}_{r,n}$. Again, our result gives a bound on (δ, ρ) from the rate function $\mathbb{W}$ so that SRSR holds with overwhelming probability.

The detailed results are presented in Section 2 and we give here the bound one can get in the case where $\mathbf{M}$ has *iid* Gaussian entries. More precisely, we establish a new sufficient condition on SRSR that offers the same lower bound as previous state-of-the-art results such as the results presented in [11]. Indeed, using Davidson-Szarek's deviation [19], we prove that if

$$(1.6) \qquad \delta > \frac{1}{2\rho} \exp\left[1 - \frac{1}{4\rho}\left(\sqrt{\frac{33 - 5\sqrt{41}}{8}} - \sqrt{2\rho} \right)^2 \right],$$

then SRSR holds with overwhelming probability when $\mathbf{M}$ has *iid* Gaussian entries, see Section 2.3.2. This bound is comparable to previous state-of-the-art result [11, 40], see Figure 1.

1.3. **Byproduct Result: New Bounds on the Restricted Isometry Constants.** One property for assessing SRSR is the Restricted Isometry Property [15, 17] of order s and parameter c, referred to as $\mathrm{RIP}(s, c, \mathbf{M})$. 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 $\mathfrak{c}(s, \mathbf{M})$ the minimum of such c's. One can prove (see Theorem 6.12 in [27] for instance) that, if RIP such that

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

holds and $\hat{x}$ is any solution to (1.2) 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 [26, 11]. 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

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

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

Hence, it holds $(1-\mathfrak{c}_{\min})\|x\|_2^2 \leq \|\mathbf{M}x\|_2^2 \leq (1+\mathfrak{c}_{\max})\|x\|_2^2$ for all $x \in \Sigma_s$, where we recall that Σ_s denotes the set of vectors with at most s nonzero coordinates. Reporting the influence of both extreme eigenvalues of covariance matrices built from $r = 2s$ columns of $\mathbf{M}$, one can weaken (FR-$\mathfrak{c}(2s)$), see for instance Theorem 2.1 in [26]. Revisiting [26] and [27, Proof of Theorem 6.13 (page 145)], this paper provides the weakest condition to get SRSR in the following theorem, see Appendix A.1 for a proof.

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

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

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. *The condition to get SRSR described in [26, 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 Condition (SRSR-$\gamma(2s)$). Also, remark that Condition (SRSR-$\gamma(2s)$) leads to the inequality $(1 + \mathfrak{c}(2s, \mathbf{M}))/(1 - \mathfrak{c}(2s, \mathbf{M})) < (4 + \sqrt{41})^2/25$ 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.*

Note that RIP constants are involved in other frameworks such as low-rank or group-sparse recovery, see for instance [43, 13, 4]. Our results do not directly apply to those frameworks but the control on RIP constants given by the methodology presented in this paper (namely (2.2), (2.3), (2.6) and (2.7)) may be invoked in these settings.

1.4. **From rate functions to SRSR.** In this paper, we provide a toolbox to derive upper bounds on RICs (with overwhelming probability) from deviation inequalities on extreme eigenvalues (or extreme singular values) of covariance matrices $\mathbf{C}_{r,n}$. The known asymptotic behavior of these extreme eigenvalues provides an expected behavior for $\mathbb{W}(\bar{\rho}, t)$ in both variables t and $\bar{\rho}$. 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 $\bar{\rho} \mapsto \mathbb{W}(\bar{\rho}, t)$ when $\bar{\rho}$ 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 paper of Feldheim and Sodin [25] on sub-Gaussian matrices, Appendix B reveals the dependency on $\bar{\rho}$ 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 [33], and Davidson and Szarek [19] are investigated in Appendix A.2.

1.5. **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 [8, 37, 2, 18] for instance. It has been proved 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

$$(1.7) \qquad n \geq c_1 \, s \log(c_2 p / s)$$

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 (FR-$\mathfrak{c}(2s)$) with $n = \mathcal{O}(s \log(p/s))$ is one of the most prominent open problem in Compressed Sensing, see [27] for instance. Furthermore, it has been shown in [18, Proposition 2.2.17] that the converse is true for any matrix $\mathbf{M}$. If the SRSR recovery (ℓ_1-**SRSR**) or (ℓ_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 (1.7) captures all we need to know about ℓ_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 (1.1) and $\eta = 0$ in (1.2), 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 (1.7) depicts a necessary and sufficient condition on NSP up to a change of constants, see for instance [18, 5]. Nevertheless, a better description of this property is offered in the works [22, 23, 24] since the authors provide a phase transition on NSP for large Gaussian matrices with *iid* entries. Let us also mention the important papers [36, 3] that give 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 [11] 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 [11, 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 [6] by preventing the use of union bound over all sub-matrices built from $r = 2s$ 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 [7] 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 $\rho = \frac{-1}{\gamma \log \delta}$ (γ is a fixed parameter) and $\delta \to 0$.

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

We would like to point out recent "off-the-shelf" concentration results, often based on generic chaining, that imply bounds on RIP constants, at least for subgaussian random matrices. The reader may consult for instance [21, Section 5] or [34]. From these results, classical bounds on gaussian widths of sparse vectors lead to null space properties or RIP constants, see [34, Section 2.6]. We would also point out the interesting results on the "Small Ball Method", see [30, 31] for instance, that requires only weak moment assumptions to get sparse recovery guarantees.

1.6. **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 [25] 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 on previous version of this paper.

2. From deviations to RICs and SRSR bounds

Following the framework of [11], 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. We aim at controlling uniformly the extreme eigenvalues, the combinatorial complexity is standardly [11] controlled by the quantity $\delta^{-1}\mathbf{H}_e(\bar{\rho}\delta)$ 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 [6] 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

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

governs the order of the deviation in the rate function $t \mapsto \mathbb{W}(\bar{\rho}, t)$ when bounding the extreme eigenvalues uniformly over all possible supports S of size r among the set of indices $\{1, \ldots, p\}$, see (2.4) and (2.8) in the functions $\Psi_{\min/\max}$. Here $\mathbb{W}^{-1}(\bar{\rho}, .)$ denotes the inverse of $\mathbb{W}$ with respect to its second variable.

The next theorems gives the probability that the matrix $\mathbf{M}$ satisfies (SRSR-$\gamma(2s)$). From Theorem 1 note that this event is included in the event such that SRSR holds, namely

$$\big\{\mathbf{M} \text{ satisfies (SRSR-}\gamma(2s)) \text{ with } s \le s_0\big\} \subset \big\{\text{SRSR holds with } s \le s_0\big\}$$

where the right hand event means that for any $x_0 \in \mathbb{R}^p$, any solution $\hat{x}$ to (1.2) satisfies SRSR for all parameters s such that $s \le s_0$.

2.1. **Using extreme eigenvalues deviations inequalities.** Useful constants are $\rho_0 := (33 - 5\sqrt{41})/16 \simeq 0.0615$ and $\sqrt{\rho_0} \simeq 0.3508$. 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 $0 < \bar{\rho} < 2\rho_0$, the largest eigenvalue λ_1 and the smallest eigenvalue λ_r of a $(r \times r)$ covariance matrix $\mathbf{C}_{r,n}$ where $r := \lfloor \bar{\rho} n \rfloor$ satisfy for all $n \geq n_0(\bar{\rho})$,*

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

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

$$(2.1) \qquad \delta > \Psi_0^{(1)}(\rho, \mathbb{W}) := \frac{1}{2\rho}\exp\left\{ 1 - \frac{1}{2\rho}\mathbb{W}\left[2\rho, 2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})\left(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}\right)\right] \right\},$$

and for any sequence of $(n \times p)$ matrices $(\mathbf{M}^{(n)})_{n \geq 2}$ with $n/p \to \delta$ satisfying (1.3), it holds that

$$\mathbb{P}\left\{ \mathbf{M}^{(n)} \text{ satisfies } (\text{SRSR-}\gamma(2s)) \text{ with } s \leq s_0 = \lfloor \rho n \rfloor \right\} \geq 1 - 2c(2\rho)e^{-nD_1(\delta,\rho)} \to 1$$

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

Furthermore, for all $\varepsilon > 0$ and for all $\bar{\rho}$ and δ such that $\delta^{-1}\mathbf{H}_e(\bar{\rho}\delta)$ belongs to the range of $\mathbb{W}(\bar{\rho}, \cdot)$, it holds

$$(2.2) \qquad \mathbb{P}\left\{ \mathfrak{c}_{\min}(\lfloor \bar{\rho} n \rfloor, \mathbf{M}^{(n)}) \geq \Psi_{\min}^{(1)}(\delta, \bar{\rho}, \mathbb{W}) + \varepsilon \right\} \leq c(\bar{\rho})e^{-nD_2(\bar{\rho},\delta,\varepsilon)},$$

$$(2.3) \qquad \mathbb{P}\left\{ \mathfrak{c}_{\max}(\lfloor \bar{\rho} n \rfloor, \mathbf{M}^{(n)}) \geq \Psi_{\max}^{(1)}(\delta, \bar{\rho}, \mathbb{W}) + \varepsilon \right\} \leq c(\bar{\rho})e^{-nD_2(\bar{\rho},\delta,\varepsilon)},$$

where $D_2(\bar{\rho}, \delta, \varepsilon) > 0$ and we denote

$$(2.4) \qquad \Psi_{\min}^{(1)}(\delta, \bar{\rho}, \mathbb{W}) := \sqrt{\bar{\rho}}(2 - \sqrt{\bar{\rho}}) + t_0,$$
$$\Psi_{\max}^{(1)}(\delta, \bar{\rho}, \mathbb{W}) := \sqrt{\bar{\rho}}(2 + \sqrt{\bar{\rho}}) + t_0,$$

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

2.2. **Using extreme singular values 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 $0 < \bar{\rho} < 2\rho_0$, the largest singular value σ_1 and the smallest singular value σ_r of a $(n \times r)$ matrix $\mathbf{X}_{r,n}$ where $r := \lfloor \rho n \rfloor$ satisfy for all $n \geq n_0(\bar{\rho})$,*

$$\forall 0 < t < \sqrt{2\rho_0}, \quad \mathbb{P}\left\{ \left(\sigma_1 - (1 + \sqrt{\bar{\rho}})\right) \vee \left((1 - \sqrt{\bar{\rho}}) - \sigma_r\right) \geq t \right\} \leq c(\bar{\rho})e^{-n\mathbb{W}(\bar{\rho},t)}$$

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

$$(2.5) \qquad \delta > \Psi_0^{(2)}(\rho, \mathbb{W}) := \frac{1}{2\rho}\exp\left\{ 1 - \frac{1}{2\rho}\mathbb{W}\left[2\rho, \sqrt{2\rho_0} - \sqrt{2\rho}\right] \right\},$$

and for any sequence of $(n \times p)$ matrices $(\mathbf{M}^{(n)})_{n \geq 2}$ with $n/p \to \delta$ satisfying (1.5), it holds that

$$\mathbb{P}\left\{ \mathbf{M}^{(n)} \text{ satisfies } (\text{SRSR-}\gamma(2s)) \text{ with } s \leq s_0 = \lfloor \rho n \rfloor \right\} \geq 1 - 2c(2\rho)e^{-nD_1(\delta,\rho)} \to 1$$

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

Furthermore, for all $\varepsilon > 0$ and for all $\bar{\rho}$ and δ such that $\delta^{-1}\mathbf{H}_e(\bar{\rho}\delta)$ belongs to the range of $\mathbb{W}(\bar{\rho}, \cdot)$, it holds

$$(2.6) \qquad \mathbb{P}\left\{ \mathfrak{c}_{\min}(\lfloor\bar{\rho}n\rfloor, \mathbf{M}^{(n)}) \geq \Psi_{\min}^{(2)}(\delta, \bar{\rho}, \mathbb{W}) + \varepsilon \right\} \leq c(\bar{\rho})e^{-nD_2(\bar{\rho}, \delta, \varepsilon)},$$

$$(2.7) \qquad \mathbb{P}\left\{ \mathfrak{c}_{\max}(\lfloor\bar{\rho}n\rfloor, \mathbf{M}^{(n)}) \geq \Psi_{\max}^{(2)}(\delta, \bar{\rho}, \mathbb{W}) + \varepsilon \right\} \leq c(\bar{\rho})e^{-nD_2(\bar{\rho}, \delta, \varepsilon)},$$

where $D_2(\bar{\rho}, \delta, \varepsilon) > 0$ and we denote

$$(2.8) \qquad \Psi_{\min}^{(2)}(\delta, \bar{\rho}, \mathbb{W}) := \min\{1, (\sqrt{\bar{\rho}} + t_0)(2 - \sqrt{\bar{\rho}} - t_0)\},$$
$$\Psi_{\max}^{(2)}(\delta, \bar{\rho}, \mathbb{W}) := (\sqrt{\bar{\rho}} + t_0)(2 + \sqrt{\bar{\rho}} + t_0),$$

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

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 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 [19], and Ledoux and Rider [33]. The last paragraph focuses on a deviation inequality for Rademacher matrices, proved by Feldheim and Sodin [25].

In what follows, the bounds on RICs are written in terms of parameters $\bar{\rho} = \frac{r}{n}$ and $\delta = \frac{n}{p}$. When expressing Condition (2.1) or (2.5), we use parameter $\rho = \frac{s}{n}$, with $s = r/2$. Note that throughout this paper, it holds that $\bar{\rho} = 2\rho$.

2.3.1. *Ideal deviations.* The asymptotic behavior of extreme eigenvalues for random covariance matrices was first established for matrices with Gaussian entries [28, 12] and extended to ones with more general entries in [42, 38, 25, 39, 45]. The largest eigenvalue fluctuations are described by the following:

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

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

$$\left[\frac{n\bar{\rho}^{1/4}}{(1 - \sqrt{\bar{\rho}})^2}\right]^{\frac{2}{3}}((1 - \sqrt{\bar{\rho}})^2 - \lambda_r) \xrightarrow[n\to\infty]{(d)} F_1.$$

We focus on the largest eigenvalue λ_1 and write:

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

This deviation probability is therefore expected to be close to

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

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 ∞, 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

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

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

$$\mathbb{P}\Big(\lambda_1 \geq (1 + \sqrt{\bar{\rho}})^2 + t\Big) \leq Ce^{-cn\min(t,t^2)}.$$

Similar results should hold for the smallest eigenvalue, except that $\lambda_r \geq 0$ almost surely and therefore only moderate deviations can occur. See [32] for a detailed survey on this subject and [33, Page 1322] for a specific discussion on the change of behavior occurring around $t = \mathcal{O}(\sqrt{\bar{\rho}})$.

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

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

where λ_1 denotes the largest eigenvalue of a $(r \times r)$ covariance matrix $\mathbf{C}_{r,n}$ with *iid* sub-Gaussian entries and

$$(2.9) \qquad \mathbb{W}_{\mathrm{TW}}(\bar{\rho},t) := \frac{1}{C}\Big\{\frac{\bar{\rho}^{\frac{1}{4}}}{(1+\sqrt{\bar{\rho}})^2}t^{\frac{3}{2}}\mathbb{1}_{t \leq \sqrt{\bar{\rho}}} + \frac{t^2}{(1+\sqrt{\bar{\rho}})^2}\mathbb{1}_{\sqrt{\bar{\rho}} < t \leq 1} + \frac{t}{(1+\sqrt{\bar{\rho}})^2}\mathbb{1}_{t > 1}\Big\},$$

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

$$\mathbb{W}_{\mathrm{TW}}^{-1}(\bar{\rho},u) = \underbrace{\frac{C^{2/3}(1 + \sqrt{\bar{\rho}})^{4/3}}{\bar{\rho}^{\frac{1}{6}}}u^{\frac{2}{3}}}_{\text{small deviation part}}\mathbb{1}_{u \leq \frac{\bar{\rho}}{C(1+\sqrt{\bar{\rho}})^2}}$$

$$+ \underbrace{C^{1/2}(1 + \sqrt{\bar{\rho}})\sqrt{u}}_{\text{moderate deviation part}}\mathbb{1}_{\frac{\bar{\rho}}{C(1+\sqrt{\bar{\rho}})^2} < u \leq \frac{1}{C(1+\sqrt{\bar{\rho}})^2}} + \underbrace{C(1 + \sqrt{\bar{\rho}})^2 u}_{\text{large deviation part}}\mathbb{1}_{u > \frac{1}{C(1+\sqrt{\bar{\rho}})^2}}.$$

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

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

where $t_0 = \mathbb{W}_{\mathrm{TW}}^{-1}\big(\bar{\rho}, \frac{1}{\delta}\mathbf{H}_e(\bar{\rho}\delta)\big)$.

Remark. *Note that in the two asymptotic Regimes (a) and (c) of [7], $\frac{1}{\delta}\mathbf{H}_e(\bar{\rho}\delta)$ will be larger than $\frac{\bar{\rho}}{C(1+\sqrt{\bar{\rho}})^2}$. Therefore it seems that the most important part in the rate function (2.9) for our present use is the moderate and large deviation parts, arising from the concentration of measure phenomenon.*

2.3.2. *Davidson and Szarek's deviations.* Consider a $(r \times n)$ matrix $\mathbf{X}$ with *iid* standard Gaussian entries. In the paper [19], Davidson and Szarek have shown that for all $0 < \bar{\rho} < 1$ it holds

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

where $\sigma_i(\mathbf{X})$ denotes the singular values of $\mathbf{X}$ and $\mathbb{W}_{\mathrm{DS}}(\bar{\rho}, t) := t^2/2$, see [27, page 291] for instance. This inequality relies on the concentration of measure phenomenon. Note that

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

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

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

where $t_0 = \sqrt{\frac{2}{\delta}\mathbf{H}_e(\bar{\rho}\delta)}$.

Remark. *In the three asymptotic Regimes* (a), (b) *and* (c), *these bounds on RICs behave similarly to the ones obtained by Bah and Tanner in* [7], *except that constants are better in* [7]. *Note that this deviation has been used in the paper* [16, Lemma 3.1] *to bound the RIP constant.*

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

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

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

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

2.3.3. *Ledoux and Rider's deviations.* Ledoux and Rider proved in [33] small deviation inequalities for β 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 $0 < \bar{\rho} < 1$ and for all $n \geq 2$, setting $r = \lfloor \bar{\rho}n \rfloor$,

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

where λ_1 denotes the largest eigenvalue of a $(r \times r)$ covariance matrix $\mathbf{C}_{r,n}$ with *iid* standard Gaussian entries and

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

where $C_{\mathrm{LR}} > 0$ may be bounded explicitly from [33]. As explained in Section 1.4, the dependency of function $\mathbb{W}$ in parameter $\bar{\rho}$ 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 r/n is bounded. For λ_r, we follow the procedure explained in [33, Section 5, page 1338] to write the following

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

where

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

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

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

where

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

Therefore

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

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

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

where $t_0 = \mathbb{W}_{\mathrm{LR}}^{-1}(\bar{\rho}, \frac{1}{\delta} \mathbf{H}_e(\bar{\rho}\delta))$.

Remark. *In the three asymptotic Regimes* (a), (b) *and* (c), *it may be shown that* $\frac{1}{\delta} \mathbf{H}_e(\bar{\rho}\delta) \ge \frac{\bar{\rho}}{C_{\mathrm{LR}}}$. *These bounds on RICs behave similarly to the ones obtained by Bah and Tanner in* [7] *in Regime* (a), *except that their constants are better. In Regimes* (b) *and* (c), *they behave badly compared to those of* [7].

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

$$\delta > \frac{1}{2\rho} \exp\Big\{ 1 - \frac{1}{2\rho} \mathbb{W}_{\mathrm{LR}}\Big[2\rho, 2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})\big(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}\big) \Big] \Big\}.$$

When ρ is small (which is the case in Regimes (a) and (c)), the second argument in $\mathbb{W}_{\mathrm{LR}}$ is approximately τ_0 and this condition approximately writes

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

2.3.4. *Feldheim and Sodin's deviations.* For all $0 < \bar{\rho} < 1$ and for all $n \ge n_0$, setting $r = \lfloor \bar{\rho} n \rfloor$ it follows from [25] that

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

where λ_i denotes the eigenvalues of a $(r \times r)$ covariance matrix $\mathbf{C}_{r,n}$ with *iid* Rademacher entries and

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

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

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

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

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

where $t_0 = \mathrm{W}_{\mathrm{FS}}^{-1}(\bar{\rho}, \frac{1}{\delta}\mathbf{H}_e(\bar{\rho}\delta))$.

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

$$\delta > \frac{1}{2\rho}\exp\Big\{1 - \frac{1}{2\rho}\mathrm{W}_{\mathrm{FS}}\Big[2\rho, 2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})\big(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}\big)\Big]\Big\}.$$

When ρ is small (which is the case in regimes (a) and (c)), the second argument in W_{FS} is approximately τ_0 and this condition approximately writes

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

Remark. *In the three asymptotic regimes (a), (b) and (c), these bounds behave badly compared to the ones by Bah and Tanner in [7] (but note that we consider here entries which are not Gaussian anymore). This is indeed not surprising: from Section 2.3.1, it seems that the most important part in the rate function (2.9) is the moderate and large deviation behavior, whereas Feldheim and Sodin's inequality focuses only on small deviations. For matrices with independent Rademacher entries, concentration of measure phenomenon still occurs and one can use Talagrand's inequality to control large deviations of the largest singular value σ_1, which is a convex function of the entries. A similar inequality holds for matrices with independent sub-Gaussian entries:*

$$(2.10) \qquad \mathbb{P}\Big(\frac{\sigma_1(\mathbf{X})}{\sqrt{n}} \geq t(1 + \sqrt{\bar{\rho}})\Big) \leq e^{-cnt^2}, \quad \text{for all } t \geq C,$$

where $\mathbf{X}$ is a $(r \times n)$ matrix with iid sub-Gaussian entries, with mean zero and variance 1. Here $c, C > 0$ depend on the sub-Gaussian moment of the entries and on $\bar{\rho}$ (see for example [35]). Note that these constants might be computed explicitly but we did not pursue on this laborious task here. As far as we know, the more precise moderate and large deviation inequality for the smallest singular value of sub-Gaussian matrices was established by Rudelson and Vershynin (see for example [41]):

$$(2.11) \qquad \mathbb{P}\Big(\frac{\sigma_r(\mathbf{X})}{\sqrt{n}} \leq t(1 - \sqrt{\bar{\rho}})\Big) \leq (Ct)^{n(1-\bar{\rho})} + e^{-cn}, \quad \text{for all } t \geq 0,$$

where $c, C > 0$ depend only on the sub-Gaussian moment of the entries. Again, these constants may be computed explicitly. Note that the e^{-cn} term quantifies the fact that $\mathbf{X}$ is non singular with probability strictly less than 1.

Combining Feldheim and Sodin's inequality with the preceding ones (2.10) and (2.11) would probably lead to a more accurate deviation inequality. However constants should be computed explicitly and the proof of Theorem 3 modified in order to deal with the additive term e^{-cn} in (2.11). We will not pursue this task here.

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 $\bar{\rho} \to 0$ 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	$2\sqrt{2}\sqrt{\bar{\rho}\lvert\log\bar{\rho}\rvert} + 2\sqrt{\bar{\rho}} + \sqrt{2}(1-\log\delta)\sqrt{\frac{\bar{\rho}}{\lvert\log\bar{\rho}\rvert}} + o\!\left(\sqrt{\frac{\bar{\rho}}{\lvert\log\bar{\rho}\rvert}}\right)$	$=\Psi_{\max}^{(2)}$
Ledoux-Rider	$C_{\mathrm{LR}}\sqrt{\bar{\rho}\lvert\log\bar{\rho}\rvert} + \left[2 + C_{\mathrm{LR}}\left(1-\log\delta\right)\right]\sqrt{\bar{\rho}} + o(\sqrt{\bar{\rho}})$	$=\Psi_{\max}^{(1)}$
Feldheim-Sodin	$2\sqrt{\bar{\rho}}\exp\left[C_{\mathrm{FS}}^{2/3}\lvert\log\bar{\rho}\rvert^{2/3}\right] + o\!\left(\sqrt{\bar{\rho}}\exp\left[C_{\mathrm{FS}}^{2/3}\lvert\log\bar{\rho}\rvert^{2/3}\right]\right)$	$=\Psi_{\max}^{(1)}$

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

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

APPENDIX A. PROOFS OF THE MAIN RESULTS

A.1. **Proof of Theorem 1.** The proof follows the same guidelines as [27, Proof of Theorem 6.13, page 145]. A sufficient condition for SRSR is the ℓ_2-robust null space property, see [27, Theorem 4.22, page 88]. Namely, we need to find constants $\kappa \in (0,1)$ and $\tau > 0$ such that, for any $v \in \mathbb{R}^p$ and any $S \subset \{1,\ldots,p\}$ such that $\lvert S \rvert = s$,

$$\lVert v_S \rVert_2 \leq \frac{\kappa}{\sqrt{s}}\lVert v_{S^c} \rVert_1 + \tau \lVert \mathbf{M}v \rVert_2 .$$

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

$$\lVert \mathbf{M}v_{S_0} \rVert_2^2 = (1+t)\lVert v_{S_0} \rVert_2^2$$
$$\text{with} \quad -\mathfrak{c}_{\min}(2s,\mathbf{M}) \leq -\mathfrak{c}_{\min}(s,\mathbf{M}) \leq t \leq \mathfrak{c}_{\max}(s,\mathbf{M}) \leq \mathfrak{c}_{\max}(2s,\mathbf{M}) .$$

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 $k \geq 1$, it holds*

$$\lvert \langle \mathbf{M}v_{S_0}, \mathbf{M}v_{S_k} \rangle \rvert \leq \sqrt{(\mathfrak{c}_{\max} - t)(\mathfrak{c}_{\min} + t)}\lVert v_{S_0} \rVert_2 \lVert v_{S_k} \rVert_2 .$$

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

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

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. $\square$

Using Lemma 4, observe that

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

Now, Lemma 6.14 in [27] gives that

$$
\sum_{k \geq 1}\|v_{S_k}\|_2 \leq \frac{1}{\sqrt{s}}\|v_{S_0^c}\|_1 + \frac{1}{4}\|v_{S_0}\|_2.
$$

We deduce that

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

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

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

It suffices that $\kappa = 4b/(4-b) < 1$ to get the ℓ_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*

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

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. $\qquad\square$

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.** We first prove (2.3). Set $\varepsilon > 0$ and suppose that $\bar{\rho}$ and δ are such that $\frac{1}{\delta}\mathbf{H}_e(\bar{\rho}\delta)$ is in the range of $\mathbb{W}(\bar{\rho}, \cdot)$. Set $t_0 = \mathbb{W}^{-1}\big(\bar{\rho}, \frac{1}{\delta}\mathbf{H}_e(\bar{\rho}\delta)\big)$. Then

$$
\begin{aligned}
\mathbb{P}\Big\{\mathfrak{c}_{\max} \geq \sqrt{\bar{\rho}}(2 + \sqrt{\bar{\rho}}) + t_0 + \varepsilon\Big\} &= \mathbb{P}\Big\{1 + \mathfrak{c}_{\max} \geq (1 + \sqrt{\bar{\rho}})^2 + t_0 + \varepsilon\Big\} \\
&= \mathbb{P}\Big\{\exists x \in \Sigma_r \text{ s.t. } \|\mathbf{M}x\|^2 \geq \big((1 + \sqrt{\bar{\rho}})^2 + t_0 + \varepsilon\big)\|x\|^2\Big\} \\
&\leq \sum_{\mathbf{C}_{r,n}} \mathbb{P}\Big\{\lambda_1(\mathbf{C}_{r,n}) \geq (1 + \sqrt{\bar{\rho}})^2 + t_0 + \varepsilon\Big\} \\
&\leq \binom{p}{r} c(\bar{\rho}) e^{-n\mathbb{W}(\bar{\rho}, t_0 + \varepsilon)} \\
&\leq c(\bar{\rho})\Theta\, e^{-n\mathbb{W}(\bar{\rho}, t + \varepsilon) + p\mathbf{H}_e(r/p)} \\
&\leq c(\bar{\rho})\Theta\, e^{-nD},
\end{aligned}
$$

with $\mathbf{H}_e(t) = -t\log t - (1 - t)\log(1 - t)$ for $t \in (0, 1)$, $\Theta^2 := e^{1/2}/(2\pi[r(1 - r/p)]^{1/p})$ and $D = \mathbb{W}(\bar{\rho}, t_0 + \varepsilon) - \mathbb{W}(\bar{\rho}, t_0) > 0$. Indeed, note that Stirling formula (see Lemma 10) leads to

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

where $\Theta \to e^{1/4}/\sqrt{2\pi}$ when $0 < r/p < 1$ and p goes to infinity. Set $D_2(\bar{\rho}, \delta, \varepsilon) = \frac{D}{2}$. Observe that $D - \log(\Theta)/n \geq D_2(\bar{\rho}, \delta, \varepsilon)$ for large enough n. Then, for n large enough (depending only on $\bar{\rho}$, δ and ε), it holds

$$
\mathbb{P}\Big\{\mathfrak{c}_{\max} \geq \sqrt{\bar{\rho}}(2 + \sqrt{\bar{\rho}}) + t_0 + \varepsilon\Big\} \leq c(\bar{\rho}) e^{-nD_2(\bar{\rho}, \delta, \varepsilon)}.
$$

Following the same arguments, we get a similar inequality (2.2) for $\mathfrak{c}_{\min}$.

We suppose now that $(\mathbf{M}^{(n)})_{n \geq 2}$ is a sequence of $(n \times p)$ matrices with $n/p \to \delta$ satisfying (1.3) and that Condition (2.1) is satisfied, namely

$$
\delta > \Psi_0^{(1)}(\rho, \mathbb{W}) := \frac{1}{2\rho}\exp\Big\{1 - \frac{1}{2\rho}\mathbb{W}\Big[2\rho, 2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})\big(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}\big)\Big]\Big\}.
$$

Using the fact that $\mathbf{H}_e(t) \leq -t\log t + t$, this condition implies

$$
(\text{A.1}) \qquad 0 < \frac{1}{\delta}\mathbf{H}_e(2\rho\delta) < \mathbb{W}\Big[2\rho, 2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})\big(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}\big)\Big].
$$

Note that $2\tau_0(\sqrt{\rho} - \sqrt{\rho_0})(\sqrt{\rho} - \frac{1}{2\sqrt{\rho_0}}) = \frac{8}{\sqrt{41}}\rho - 2\sqrt{2}\sqrt{\rho} + \frac{4}{\sqrt{41}}$ belongs to $(0, \tau_0)$. Recall that $\mathbb{W}(2\rho, .)$ is increasing on this interval. Applying $\mathbb{W}^{-1}(2\rho, .)$ to (A.1) leads to

$$
0 < t_0 < \frac{8}{\sqrt{41}}\rho - 2\sqrt{2}\sqrt{\rho} + \frac{4}{\sqrt{41}}.
$$

Let $t_0 < t < t_* = \frac{8}{\sqrt{41}}\rho - 2\sqrt{2}\sqrt{\rho} + \frac{4}{\sqrt{41}}$. Remark that Condition (SRSR-$\gamma(2s)$) is satisfied on the event $\{\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) < \sqrt{2\rho}(2 + \sqrt{2\rho}) + t\} \cap \{\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) < \sqrt{2\rho}(2 - \sqrt{2\rho}) + t\}$. Indeed, denoting $\gamma_0 = \frac{(4+\sqrt{41})^2}{25}$,

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

Note that $t < t_* = \frac{8}{\sqrt{41}}\rho - 2\sqrt{2}\sqrt{\rho} + \frac{4}{\sqrt{41}}$ implies that the denominator is positive. Moreover $1 - \gamma_0 = -\frac{8}{25}(4 + \sqrt{41})$ and $1 + \gamma_0 = \frac{2\sqrt{41}}{25}(4 + \sqrt{41}) > 0$. Therefore

$$
(1 - \gamma_0) + 2\sqrt{2}(1 + \gamma_0)\sqrt{\rho} + 2(1 - \gamma_0)\rho + (1 + \gamma_0)t = (1 + \gamma_0)(t - t_*) < 0\,.
$$

As a consequence,

$$
\gamma(2s, \mathbf{M}) < \frac{(4 + \sqrt{41})^2}{25}\,.
$$

Invoking (2.3) and (2.2) with $\varepsilon = t - t_0$ and $\bar{\rho} = 2\rho$ yields the following

$$
\begin{aligned}
\mathbb{P}\Big(\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) \geq \sqrt{2\rho}(2 + \sqrt{2\rho}) + t\Big) &\leq c(2\rho)e^{-nD_2(2\rho, \delta, t - t_0)}, \\
\mathbb{P}\Big(\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) \geq \sqrt{2\rho}(2 - \sqrt{2\rho}) + t\Big) &\leq c(2\rho)e^{-nD_2(2\rho, \delta, t - t_0)},
\end{aligned}
$$

with $D_2(2\rho, \delta, t - t_0) = \big(\mathbb{W}(2\rho, t) - \mathbb{W}(2\rho, t_0)\big)/2 > 0$. Note that $t \mapsto D_2(2\rho, \delta, t - t_0)$ is continuous and increasing from $[t_0, t_*) \subset [0, \tau_0)$ onto $\big[0, D_2(2\rho, \delta, t_* - t_0)\big)$. Choose $D_1(\rho, \delta) \in \big(0, D_2(2\rho, \delta, t_* - t_0)\big)$ (remark that t_* and t_0 depend only on ρ and δ) and define $t_{\rho, \delta}$ in (t_0, t_*) so that $D_2(2\rho, \delta, t_{\rho, \delta} - t_0) = D_1(\rho, \delta)$. Therefore

$$
\begin{aligned}
&\mathbb{P}\Big\{\mathbf{M}^{(n)} \text{ does not satisfy (SRSR-}\gamma(2s)\text{) with } s \leq \lfloor \rho n \rfloor\Big\} \\
&\leq \mathbb{P}\Big(\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) \geq \sqrt{2\rho}(2 + \sqrt{2\rho}) + t_{\rho, \delta}\Big) + \mathbb{P}\Big(\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) \geq \sqrt{2\rho}(2 - \sqrt{2\rho}) + t_{\rho, \delta}\Big) \\
&\leq 2c(2\rho)e^{-nD_1(\rho, \delta)},
\end{aligned}
$$

which concludes the proof.

A.3. **Proof of Theorem 3.** We follow the same lines as in the previous proof. We first prove (2.7). For $\varepsilon > 0$, we get the following.

$$\mathbb{P}\Big\{\mathfrak{c}_{\max} \geq (\sqrt{\bar{\rho}} + t_0)(2 + \sqrt{\bar{\rho}} + t_0) + \varepsilon\Big\}$$

$$= \mathbb{P}\Big\{1 + \mathfrak{c}_{\max} \geq (1 + \sqrt{\bar{\rho}} + t_0)^2 + \varepsilon\Big\}$$

$$= \mathbb{P}\Big\{\exists x \in \Sigma_r \text{ s.t. } \|\mathbf{M}x\|^2 \geq \big((1 + \sqrt{\bar{\rho}} + t_0)^2 + \varepsilon\big)\|x\|^2\Big\}$$

$$\leq \sum_{\mathbf{C}_{r,n}} \mathbb{P}\Big\{\sigma_1 \geq \sqrt{(1 + \sqrt{\bar{\rho}} + t_0)^2 + \varepsilon}\Big\}$$

$$\leq \sum_{\mathbf{C}_{r,n}} \mathbb{P}\Big\{\sigma_1 \geq 1 + \sqrt{\bar{\rho}} + t_0 + f(\bar{\rho}, \delta, \varepsilon)\Big\},$$

where $f(\bar{\rho}, \delta, \varepsilon) = \sqrt{(1 + \sqrt{\bar{\rho}} + t_0)^2 + \varepsilon} - (1 + \sqrt{\bar{\rho}} + t_0) > 0$. Then

$$\mathbb{P}\Big(\mathfrak{c}_{\max} \geq (\sqrt{\bar{\rho}} + t_0)(2 + \sqrt{\bar{\rho}} + t_0) + \varepsilon\Big) \leq \binom{p}{r} c(\bar{\rho}) e^{-n\mathbb{W}\big(\bar{\rho}, t_0 + f(\delta, \bar{\rho}, \varepsilon)\big)}$$

$$\leq c(\bar{\rho})\Theta \, e^{-n\mathbb{W}\big(\rho, t_0 + f(\delta, \bar{\rho}, \varepsilon)\big) + p\mathbf{H}_e(r/p)}$$

$$\leq c(\bar{\rho})\Theta \, e^{-nD},$$

with $\mathbf{H}_e(t) = -t \log t - (1 - t) \log(1 - t)$ for $t \in (0, 1)$, $\Theta^2 := e^{1/2}/(2\pi[r(1 - r/p)]^{1/p})$ and $D = \mathbb{W}\big(\bar{\rho}, t_0 + f(\delta, \bar{\rho}, \varepsilon)\big) - \mathbb{W}(\bar{\rho}, t_0)$. Set $D_2(\bar{\rho}, \delta, \varepsilon) = \frac{D}{2}$. Observe that $D - \log(\Theta)/n \geq D_2(\bar{\rho}, \delta, \varepsilon)$ for large enough n. Then, for n large enough (depending only on $\bar{\rho}$, δ and ε), it holds

$$\mathbb{P}\Big\{\mathfrak{c}_{\max} \geq (\sqrt{\bar{\rho}} + t_0)(2 + \sqrt{\bar{\rho}} + t_0) + \varepsilon\Big\} \leq c(\bar{\rho}) e^{-nD_2(\bar{\rho}, \delta, \varepsilon)}.$$

Following the same arguments, we get a similar inequality (2.6) for $\mathfrak{c}_{\min}$.

We suppose now that $(\mathbf{M}^{(n)})_{n \geq 2}$ is a sequence of $(n \times p)$ matrices with $n/p \to \delta$ satisfying (1.5) and that Condition (2.5) is satisfied, namely

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

Using the fact that $\mathbf{H}_e(t) \leq -t \log t + t$, this condition implies

$$0 < \frac{1}{\delta}\mathbf{H}_e(2\rho\delta) < \mathbb{W}\Big(2\rho, \sqrt{2\rho_0} - \sqrt{2\rho}\Big).$$

Recall that $\mathbb{W}(2\rho, .)$ is increasing on $[0, \sqrt{2\rho_0})$. Applying $\mathbb{W}^{-1}(2\rho, .)$ leads to

$$0 < t_0 < \sqrt{2\rho_0} - \sqrt{2\rho}.$$

Let $t_0 < t < t_* = \sqrt{2\rho_0} - \sqrt{2\rho}$. Remark that Condition (SRSR-$\gamma(2s)$) is satisfied on the event $\{\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) < (\sqrt{2\rho} + t)(2 + \sqrt{2\rho} + t)\} \cap \{\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) < (\sqrt{2\rho} + t)(2 - \sqrt{2\rho} - t)\}$. Indeed

$$\sqrt{\gamma(2s, \mathbf{M})} - \frac{4 + \sqrt{41}}{5} < \frac{1 + \sqrt{2\rho} + t}{1 - \sqrt{2\rho} - t} - \sqrt{\gamma_0}$$

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

Note that $t < t_* = \sqrt{2\rho_0} - \sqrt{2\rho}$ implies that the denominator is positive. Moreover

$$1 - \sqrt{\gamma_0} + \sqrt{2}(1 + \sqrt{\gamma_0})\sqrt{\rho} + (1 + \sqrt{\gamma_0})t = (1 + \sqrt{\gamma_0})(t - t_*) < 0.$$

As a consequence,

$$\gamma(2s, \mathbf{M}) < \frac{(4 + \sqrt{41})^2}{25}.$$

Similarly to (2.7) and (2.6), it can be proved that for $t > t_0$,

$$\mathbb{P}\Big(\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) \geq (\sqrt{2\rho} + t)(2 + \sqrt{2\rho} + t)\Big) \leq c(2\rho)e^{-nD_3(2\rho, \delta, t)},$$

$$\mathbb{P}\Big(\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) \geq (\sqrt{2\rho} + t)(2 - \sqrt{2\rho} - t)\Big) \leq c(2\rho)e^{-nD_3(2\rho, \delta, t)},$$

with $D_3(2\rho, \delta, t) > 0$. Note that $t \mapsto D_3(2\rho, \delta, t)$ is continuous and increasing from $[t_0, t_*) \subset [0, \tau_0)$ onto $\big[D_3(2\rho, \delta, t_0), D_3(2\rho, \delta, t_*)\big)$. Choose $D_1(\rho, \delta) \in \big(D_2(2\rho, \delta, t_0); D_3(2\rho, \delta, t_*)\big)$ (note that t_* and t_0 depend only on ρ and δ) and define $t_{\rho,\delta}$ so that $D_3(2\rho, \delta, t_{\rho,\delta}) = D_1(\rho, \delta)$. Therefore

$$\mathbb{P}\Big\{\mathbf{M}^{(n)} \text{ does not satisfy (SRSR-}\gamma(2s)\text{) with } s \leq \lfloor \rho n \rfloor\Big\}$$

$$\leq \mathbb{P}\Big(\mathfrak{c}_{\max}(2s, \mathbf{M}^{(n)}) \geq \big(\sqrt{2\rho} + t_{\rho,\delta}\big)\big(2 + \sqrt{2\rho} + t_{\rho,\delta}\big)\Big)$$

$$+ \mathbb{P}\Big(\mathfrak{c}_{\min}(2s, \mathbf{M}^{(n)}) \geq \big(\sqrt{2\rho} + t_{\rho,\delta}\big)\big(2 - \sqrt{2\rho} - t_{\rho,\delta}\big)\Big)$$

$$\leq 2c(2\rho)e^{-nD_1(\rho, \delta)},$$

which concludes the proof.

References

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions*. National Bureau of Standards, Washington DC, 1965.

[2] 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(1):61–88, 2011.

[3] D. Amelunxen, M. Lotz, M. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, 3(3):224–294, 2014.

[4] U. Ayaz, S. Dirksen, and H. Rauhut. Uniform recovery of fusion frame structured sparse signals. *Applied and Computational Harmonic Analysis*, 41(2):341–361, 2016.

[5] J.-M. Azaïs, Y. De Castro, and S. Mourareau. A rice method proof of the null-space property over the grassmannian. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 53(4):1821–1838, 2017.

[6] B. Bah and J. Tanner. Improved bounds on restricted isometry constants for gaussian matrices. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2882–2898, 2010.

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

[8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[9] K. Bertin, E. Le Pennec, and V. Rivoirard. Adaptive dantzig density estimation. *Annales de l'IHP, Probabilités et Statistiques*, 47(1):43–74, 2011.

[10] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705–1732, 2009.

[11] J. D. Blanchard, C. Cartis, and J. Tanner. Compressed sensing: How sharp is the restricted isometry property? *SIAM review*, 53(1):105–125, 2011.

[12] A. Borodin and P. Forrester. Increasing subsequences and the hard-to-soft edge transition in matrix ensembles. *J. Phys. A*, 36(12):2963–2981, 2003. Random matrix theory.

[13] T. T. Cai and A. Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Information Theory*, 60(1):122–132, 2014.

[14] E. J. Candes. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.

[15] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[16] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[17] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[18] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interaction between Compressed Sensing, Random matrices and High dimensional geometry*. Number 37 in Panoramas et synthéses. SMF, 2012.

[19] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. *Handbook of the geometry of Banach spaces*, 1(317-366):131, 2001.

[20] Y. De Castro. A remark on the lasso and the dantzig selector. *Statistics and Probability Letters*, 83(1):304–314, January 2013.

[21] S. Dirksen. Tail bounds via generic chaining. *Electronic Journal of Probability*, 20(53):29pp, 2015.

[22] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9452–9457, 2005.

[23] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.

[24] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–4293, 2009.

[25] O. N. Feldheim and S. Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geometric And Functional Analysis*, 20(1):88–123, 2010.

[26] 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(3):395–407, 2009.

[27] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[28] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.

[29] A. Juditsky and A. Nemirovski. Accuracy guarantees for l1-recovery. *Information Theory, IEEE Transactions on*, 57(12):7818–7839, 2011.

[30] G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. *Journal of the European Mathematical Society*, 19(3):881–904, 2017.

[31] G. Lecué and S. Mendelson. Regularization and the small-ball method i: sparse recovery. *The Annals of Statistics*, 46(2):611–641, 2018.

[32] M. Ledoux. Deviation inequalities on largest eigenvalues. In *Geometric aspects of functional analysis*, volume 1910 of *Lecture Notes in Math.*, pages 167–219. Springer, Berlin, 2007.

[33] M. Ledoux and B. Rider. Small deviations for beta ensembles. *Electron. J. Probab*, 15(41):1319–1343, 2010.

[34] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin. A simple tool for bounding the deviation of random matrices on geometric sets. In *Geometric aspects of functional analysis*, pages 277–299. Springer, 2017.

[35] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491–523, 2005.

[36] M. B. McCoy and J. A. Tropp. Sharp recovery bounds for convex demixing, with applications. *Found. Comput. Math.*, 14(3):503–567, 2014.

[37] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and sub-gaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.

[38] S. Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields*, 143(3-4):481–516, 2009.

[39] N. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014.

[40] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[41] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1576–1602, New Delhi, 2010. Hindustan Book Agency.

[42] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002.

[43] Y. Traonmilin and R. Gribonval. Stable recovery of low-dimensional cones in hilbert spaces: One rip to rule them all. *Applied and Computational Harmonic Analysis*, 45:170–205, 2016.

[44] S. A. van de Geer and P. Bühlmann. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

[45] K. Wang. Random covariance matrices: universality of local statistics of eigenvalues up to the edge. *Random Matrices Theory Appl.*, 1(1):1150005, 24, 2012.

APPENDIX B. SUPPLEMENT : DEVIATIONS FOR THE RADEMACHER MODEL

In this section we follow the steps of the work [25] to get small deviation inequalities on the extreme eigenvalues of Gram matrices built from the Rademacher law. The paper [25] 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 [25] 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 [25, 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 [25] in order to provide an upper bound on C_{FS} and its dependence on the ratio ρ of the sizes of the Rademacher matrix. This strenuous hunt necessitates to recast all the asymptotic bounds appearing in [25] 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 $N > M \geq 54$ 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

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

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

where $\rho = M/N$ and

$$\mathbb{W}_0(\rho, \varepsilon) := c_0 \exp \left[c_0 \sqrt{\log \left(1 + \frac{\varepsilon}{2\sqrt{\rho}} \right)} \right]$$

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

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 $0 < \varepsilon < \sqrt{\rho}$,

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

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

B.1. **Sketch of the proof.** The result of [25] 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 [25] 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 [25] 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 [25] 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 [25] 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 (B.1) 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 [25] Section II.2 page 101.

Lemma 7. *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 [25] 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 [25] 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 [25], 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

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

Writing $\theta = \frac{(6g-1)^{8g-1/2}}{g^{g+1/2}(g+1)^{g+3/2}(4g)^{4g+1/2}}$ 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 γ 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)

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

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

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

• 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

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

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 β. 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 α. 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 γ. Recall that $t \in (1/s, 1 - 2/s)$. Dealing separately with the two terms $\left((3-t)s - \frac{1}{2}\right) \log \left(1 - \frac{1}{(3-t)s}\right)$ and $\frac{1}{2}(s(1-t) + 3) \log(1 + \frac{2}{s(1-t)})$ yields

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

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

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

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 ∞. 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. □

B.3. **Number of paths.** Let $n \geq 1$ 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 [25]. Recall that $M \leq N$ denotes the sizes of the Rademacher matrix.

Lemma 8. *It holds*

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

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

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

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 [25, page 117]. The number of ways to choose the weights on a diagram constructed in s steps by the automaton is at most [25]

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

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 7. Invoke Lemma 10 to get that

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

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

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

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_*, ∞). Therefore, g is strictly concave on (s_*, ∞) 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

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

$$\leq ne^{1/12}\sqrt{\frac{2}{3}} \exp\left((5\log 2 - 3\log 3 + 1.013)s + 3\log 3 - 2\log 2 - 0.013\right)$$

$$\leq ne^{1/12}\sqrt{\frac{2}{3}} \exp(1 + 3\log 2) \exp\left((5\log 2 - 3\log 3 + 1.013)(s - 1)\right)$$

$$\leq 19.3\, n\, (3.27)^{s-1}.$$

As a consequence,

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

$$\leq 19.3 C_{0,D} n \exp\left(1.81\sqrt{C_D}(1 + \sqrt{M/N})\frac{n^{3/2}}{M^{1/2}}\right).$$

Similarly, one gets

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

$$\leq 19.3 n C_{0,D} \exp\left[\frac{1.81(1 - \sqrt{M/N})\sqrt{C_D}n^{3/2}}{\sqrt{M}}\right].$$

This yields the result. $\qquad \square$

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

Lemma 9. *It holds that*

(B.2) $$\left(\mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m}]] + \mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m-1}]]\right) \vee \left(\mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m}]] - \mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m-1}]]\right) \leq \Delta_m$$

where

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

and

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

Proof. Invoke Lemma IV.1.1 page 115 in [25] and Lemma 8 to get that

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

Set $s := \frac{(M-2)^2}{(M-1)(N-1)}$. For $m \geq 1$, let $A_m = \mathbb{E}[\text{tr}[\widetilde{\mathbf{C}}^{2m}]] + \mathbb{E}[\text{tr}[\widetilde{\mathbf{C}}^{2m-1}]]$. Following pages 95-96 in [25] yields:

$$
\begin{aligned}
A_m &= \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^{m}(2n+1)\binom{2m+1}{m-n}\mathbb{E}[\text{tr}[U_{2n}(\widetilde{\mathbf{C}})]] \\
(\text{B.4}) \qquad &\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^{m} 2n\binom{2m}{m-n}\mathbb{E}[\text{tr}[U_{2n-1}(\widetilde{\mathbf{C}})]].
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
A_m &\leq \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^{m}(2n+1)\binom{2m+1}{m-n} \sum_{k=0}^{n} s^k\mathbb{E}[\text{tr}[V_{2(n-k),s}(\widetilde{\mathbf{C}})]] \\
&\quad + \frac{1}{2m2^{2m}} \sum_{n=1}^{m} 2n\binom{2m}{m-n} \sum_{k=0}^{n-1} s^k\mathbb{E}[\text{tr}[V_{2n-2k-1,s}(\widetilde{\mathbf{C}})]] \\
&\leq \frac{1}{(2m+1)2^{2m}} \sum_{n=1}^{m}(2n+1)\binom{2m+1}{m-n}\Big(\sum_{k=0}^{n-1} s^k\mathbb{E}[\text{tr}[V_{2(n-k),s}(\widetilde{\mathbf{C}})]] + s^n M\Big) \\
(\text{B.5}) \qquad &\quad + \frac{1}{2m2^{2m}} \sum_{n=0}^{m} 2n\binom{2m}{m-n} \sum_{k=0}^{n-1} s^k\mathbb{E}[\text{tr}[V_{2n-2k-1,s}(\widetilde{\mathbf{C}})]]
\end{aligned}
$$

Invoke (B.3) to get with $C_{M,N} = C_{\hat\Sigma}(1 + \sqrt{M/N})$,

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

Then

$$
A_m \leq \frac{2C_{0,\hat\Sigma}}{1 - \frac{(M-2)^2}{MN}} \sum_{n=1}^{m} \Big[\frac{2n+1}{(2m+1)2^{2m}}\binom{2m+1}{m-n} + \frac{n}{m 2^{2m-1}}\binom{2m}{m-n}\Big]
$$
$$
\times n\Big(\frac{MN}{(M-1)(N-1)}\Big)^n \exp\Big[C_{M,N}\frac{2^{\frac{3}{2}}n^{\frac{3}{2}}}{M^{\frac{1}{2}}}\Big]
$$
$$
+ \frac{1}{(2m+1)2^{2m}} \sum_{n=0}^{m}(2n+1)\binom{2m+1}{m-n} s^n M.
$$

From Lemma 11 it holds

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

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

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

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

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

where

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

as claimed.

The bound on $B_m := \mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m}]] - \mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m-1}]]$ follows the same lines. The minus in front of $\mathbb{E}[\mathrm{tr}[\widetilde{\mathbf{C}}^{2m-1}]]$ change the line (B.4) to its opposite. The change of indices k leads to the term $s^{k+1/2}\mathbb{E}[\mathrm{tr}[V_{2(n-k-1),s}(\widetilde{\mathbf{C}})]]$ in (B.5). 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. $\qquad\square$

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

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

with

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

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

$$\mathbb{P}(\lambda_M(\widetilde{\mathbf{C}}) \geq \varepsilon_{M,N}) \leq \mathbb{P}(f(\lambda_M(\widetilde{\mathbf{C}})) + e_m \geq f(\varepsilon_{M,N}) + e_m)$$

$$\leq \frac{\mathbb{E}[f(\lambda_M(\widetilde{\mathbf{C}})] + e_m}{f(\varepsilon_{M,N}) + e_m}$$

$$\leq \frac{\sum_{k=1}^M (\mathbb{E}[f(\lambda_k(\widetilde{\mathbf{C}})] + e_m)}{f(\varepsilon_{M,N})}$$

(B.6) $$= \frac{A_m + M e_m}{f(\varepsilon_{M,N})}$$

Invoke Lemma 9 to get that

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

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

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

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

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

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 $\rho \leq 1$, we derive that

$$
\mathbb{W}_0(\rho, \varepsilon) \leq \frac{4C_{0,\mathrm{Rad}}(1+2e)\sqrt{3C_{\mathrm{Rad}}} + 2e\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{2e\sqrt{3C_{\mathrm{Rad}}}}
$$

$$
\times \exp\left[108\log(\frac{54}{53})\frac{\sqrt{2\log(\frac{55}{53}(1+\frac{\varepsilon}{2\sqrt{\rho}}))}}{\sqrt{3C_{\mathrm{Rad}}}}\right]
$$

$$
\leq c_0 \exp\left[c_0\sqrt{\log\left(1+\frac{\varepsilon}{2\sqrt{\rho}}\right)}\right]
$$

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

(B.7) $\qquad \mathbb{P}\left\{\lambda_M(\mathbf{C}) \geq (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\right\} \leq \frac{c_0 M e^{c_0\sqrt{\log\left(1+\frac{\varepsilon}{2\sqrt{\rho}}\right)}}}{1-\rho} e^{-N\mathbb{W}_1(\rho,\varepsilon)}$.

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 η, 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

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

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 $0 \leq \varepsilon \leq \sqrt{\rho}$,

(B.8) $\qquad \mathbb{P}\left\{\lambda_M(\mathbf{C}) \geq (\sqrt{M}+\sqrt{N})^2 + \varepsilon N\right\} \leq v\exp\left(-C^{-1}N\frac{\rho^{1/4}}{(1+\sqrt{\rho})^2}\varepsilon^{\frac{3}{2}}\right).$

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

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

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

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

and we recover an upper bound of the form (B.6) for which Lemma 9 can also be applied and we get that

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

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 (B.8) has been obtained from (B.7) 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:*

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

Proof. See [1] Eq. 6.1.38. □

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

$$
\log\left[\frac{n}{2^{2m}}\binom{2m}{m-n}\right] \leq 5 - 0.6321\frac{n^2}{m}
$$

$$
\log\left[\frac{n+1/2}{2^{2m}}\binom{2m+1}{m-n}\right] \leq 2 - 0.6555\frac{n^2}{m}
$$

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

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

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

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

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

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

as claimed. $\qquad\square$

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

YDC is with the MOKAPLAN team at INRIA Paris, 2 rue Simone Iff, 75012 Paris, France.
E-mail address: yohann.de-castro@inria.fr

YDC is with the CERMICS Laboratory at Ponts ParisTech, 6 et 8 avenue Blaise Pascal, 77455 Marne la Vallée Cedex 2, France.
E-mail address: yohann.de-castro@enpc.fr