Smallest singular value and limit eigenvalue distribution of a class of non-Hermitian random matrices with statistical application

Arup Bose[(1)], Walid Hachem[(2)]

[(1)] Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata, India.
[(2)] LIGM, CNRS, Univ Gustave Eiffel, ESIEE Paris, F-77454 Marne-la-Vallée, France.

Abstract

Suppose X is an $N \times n$ complex matrix whose entries are centered, independent, and identically distributed random variables with variance $1/n$ and whose fourth moment is of order $\mathcal{O}(n^{-2})$. Suppose A is a deterministic matrix whose smallest and largest singular values are bounded below and above respectively, and $z \neq 0$ is a complex number. First we consider the matrix $XAX^* - z$, and obtain asymptotic probability bounds for its smallest singular value when N and n diverge to infinity and $N/n \to \gamma$, $0 < \gamma < \infty$. Then we consider the special case where $A = J = [\mathbb{1}_{i-j=1 \mod n}]$ is a circulant matrix. Using the above result, we show that the limit spectral distribution of XJX^* exists when $N/n \to \gamma$, $0 < \gamma < \infty$ and describe the limit explicitly. Assuming that X represents a $\mathbb{C}^N$-valued time series which is observed over a time window of length n, the matrix XJX^* represents the one-step sample autocovariance matrix of this time series. A whiteness test against an MA correlation model for this time series is introduced based on the above limit result. Numerical simulations show the excellent performance of this test.

Keywords: Large non-Hermitian matrix theory, Limit spectral distribution, Smallest singular value, Whiteness test in multivariate time series.
MSC 2020: Primary 60B20, 62H15, Secondary 15A18, 15B52.

1 Introduction and the main results

Let $(N^{(n)})_{n\geq 1}$ be a sequence of positive integers, such that $\lim_{n\to\infty} N^{(n)}/n = \gamma$, $0 < \gamma < \infty$. Let $(X^{(n)} = [x_{ij}^{(n)}]_{i,j=0}^{N^{(n)}-1,n-1})_{n\geq 1}$ be a sequence of complex random matrices and $(A^{(n)})$ be a sequence of $n \times n$ deterministic matrices with complex entries. Suppose that

Assumption 1. For each $n \geq 1$, the complex random variables $\{x_{ij}^{(n)}\}_{i,j=0}^{N^{(n)}-1,n-1}$ are i.i.d. with $\mathbb{E}x_{00}^{(n)} = 0$, $\mathbb{E}|x_{00}^{(n)}|^2 = 1/n$. Moreover, there exists a constant m_4 such that $\sup_n n^2 \mathbb{E}|x_{00}^{(n)}|^4 \leq m_4 < \infty$.

Assumption 2.

$$0 < \inf_n s_{n-1}(A^{(n)}) \leq \sup_n s_0(A^{(n)}) < \infty,$$

where $s_0(M) \geq \cdots \geq s_{n-1}(M)$ will refer hereinafter to the singular values of the matrix $M \in \mathbb{C}^{n\times n}$.

We shall first study the probabilistic behavior of the smallest singular value of the matrix $X^{(n)}A^{(n)}X^{(n)*} - zI_N$, where M^* is the Hermitian adjoint of the matrix M, and where z is any non-zero complex number. We shall then use this result to obtain the limiting spectral behavior of the matrix $X^{(n)}J^{(n)}X^{(n)*}$ where $J^{(n)}$ is as in Equation (1) below. Finally, we shall discuss a statistical application of this last result.

The behavior of the smallest singular value of large random matrices has recently aroused an intense research effort in the field of random matrix theory. One of the main motivations for this interest is its close connections with the spectral behavior of large square non-Hermitian random matrices. It is well-known that probabilistic control of the smallest singular value of the matrix $Y - z$ is a key step towards understanding the behavior of the spectral measure of Y [7]. Starting with the fundamental model where Y has i.i.d. elements, most of the contributions dealing with the smallest singular value assume the independence between the entries of Y, as seen in [1, 27, 33, 18, 34, 35, 10] among many others. An increasing degree of generality on the probability law on the independent entries of Y has been considered in these contributions. On the other side, more structured models, such as the one dealt with in this paper, have received comparatively much less attention. We can however cite in this respect the works of Girko (see, *e.g.*, his treatise [15]) or [6, 31], which all deal with quite different models than ours.

Our results will be proved under the following additional assumption on the elements of $X^{(n)}$.

Assumption 3. The random variables $x_{00}^{(n)}$ satisfy $\sup_n |n\mathbb{E}(x_{00}^{(n)})^2| < 1$.

Assumption 3 essentially amounts to demanding that $x_{ij} \notin \mathbb{R}$. Indeed, suppose that for some n, $|n\mathbb{E}(x_{00}^{(n)})^2| = 1$. Drop the superscript $^{(n)}$, and write $x_{00} = \Re x_{00} + \imath \Im x_{00}$. Then, $1/n = |\mathbb{E}x_{00}^2| = \mathbb{E}|x_{00}|^2$. This implies $(\mathbb{E}\Re x_{00}\Im x_{00})^2 = \mathbb{E}(\Re x_{00})^2\mathbb{E}(\Im x_{00})^2$. Suppose $\Re x_{00} \not\equiv 0$. Then clearly $\Im x_{00} = \alpha \Re x_{00}$ with probability one (w.p.1) for some constant α. Thus, denoting as $\stackrel{\mathscr{L}}{=}$ the equality in law, $x_{00} \stackrel{\mathscr{L}}{=} \exp(\imath\theta)Z$, where Z is a real random variable and θ is a constant. This amounts to x_{00} being real since the factor $\exp(\imath\theta)$ has no influence on XJX^*.

We can now state our first result. We denote as $\|\cdot\|$ the spectral norm of a matrix. Events are expressed in the forms $[\ldots]$ or $\{\ldots\}$.

Theorem 1. Let Assumptions 1, 2, and 3 hold true. Then, there exist $\alpha, \beta > 0$ such that for each $C > 0$, $t > 0$, and $z \in \mathbb{C} \setminus \{0\}$,

$$\mathbb{P}\big[s_{N-1}(X^{(n)}A^{(n)}X^{(n)^*} - z) \le t,\ \|X\| \le C\big] \le c\big(n^{\alpha}t^{1/2} + n^{-\beta}\big),$$

where the constant $c > 0$ depends on C, z, and m_4 only.

The first step to prove this theorem is to "linearize" (similar linearizations have been used elsewehere, see, *e.g.*, [21]), and consider the matrix

$$H^{(n)} = \begin{pmatrix} A^{(n)^{-1}} & X^{(n)^*} \\ X^{(n)} & z \end{pmatrix} \in \mathbb{C}^{(N+n)\times(N+n)}.$$

The inversion formula for partitioned matrices implies that $\|(X^{(n)}A^{(n)}X^{(n)^*} - z)^{-1}\| \le \|H^{(n)^{-1}}\|$. Thus, it is enough to deal with $s_{N+n-1}(H^{(n)})$. A similar problem was tackled in [30, 38]. We follow closely the approach of [38] but there, the author had a real symmetric matrix with i.i.d. elements above the diagonal. Our matrix $H^{(n)}$ is more structured, and so we need appropriate modification in the arguments.

Motivated by the statistical application described in Section 2, we consider the following choice of $A^{(n)}$. Since it is an orthogonal circulant matrix, it automatically satisfies Assumption 2.

$$A^{(n)} = J^{(n)} = \begin{pmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{1}$$

Theorem 1 can be used to study the limit eigenvalue distribution of the matrix $X^{(n)}J^{(n)}X^{(n)^*}$ (see [7] or Section 4 for more information). Let $\{\lambda_0^{(n)}, \ldots, \lambda_{N^{(n)}-1}^{(n)}\}$ be its (complex) eigenvalues. The spectral distribution or measure of $X^{(n)}J^{(n)}X^{(n)^*}$ is the random probability measure:

$$\mu_n = \frac{1}{N^{(n)}} \sum_{i=0}^{N^{(n)}-1} \delta_{\lambda_i^{(n)}}.$$

When $N^{(n)}/n \to \gamma$, $0 < \gamma < \infty$, we shall identify a deterministic probability measure μ such that $\mu_n \Rightarrow \mu$ in probability, where $\Rightarrow$ refers to weak convergence. The limit μ is called the limiting spectral distribution or measure (LSD) of the sequence of matrices. To describe μ, we need the following function. For any $0 < \gamma < \infty$, let

$$g(y) = \frac{y}{y+1}(1-\gamma+2y)^2,\ (0 \vee (\gamma-1)) \le y \le \gamma. \tag{2}$$

Then g^{-1} exists on the interval $[0 \vee ((\gamma-1)^3/\gamma), \gamma(\gamma+1)]$ and maps it to $[0 \vee (\gamma-1), \gamma]$. It is an analytic increasing function on the interior of the interval.

Theorem 2. Suppose Assumptions 1 and 3 hold. Then, there exists a deterministic rotationally invariant probability measure μ on $\mathbb{C}$ such that $\mu_n \Rightarrow \mu$ in probability. Let the distribution function of its radial component be $F(r) = \mu(\{z \in \mathbb{C} : |z| \le r\})$, $0 \le r < \infty$. If $\gamma \le 1$, then

$$F(r) = \begin{cases} \gamma^{-1}g^{-1}(r^2) & \text{if } 0 \le r \le \sqrt{\gamma(\gamma+1)}, \\ 1 & \text{if } r > \sqrt{\gamma(\gamma+1)}. \end{cases}$$

If $\gamma > 1$, then

$$F(r) = \begin{cases} 1-\gamma^{-1} & \text{if } 0 \le r \le (\gamma-1)^{3/2}/\sqrt{\gamma}, \\ \gamma^{-1}g^{-1}(r^2) & \text{if } (\gamma-1)^{3/2}/\sqrt{\gamma} < r \le \sqrt{\gamma(\gamma+1)}, \\ 1 & \text{if } r > \sqrt{\gamma(\gamma+1)}. \end{cases}$$

The support of μ is the disc $\{z : |z| \leq \sqrt{\gamma(\gamma+1)}\}$ when $\gamma \leq 1$, and when $\gamma > 1$, it is the ring $\{z : (\gamma-1)^{3/2}/\sqrt{\gamma} \leq |z| \leq \sqrt{\gamma(\gamma+1)}\}$ together with the point $\{0\}$ which has mass $1-\gamma^{-1}$. Moreover, $F(r)$ has a positive and analytical density on the open interval $(0 \vee \text{sign}(\gamma-1)|\gamma-1|^{3/2}/\sqrt{\gamma}, \sqrt{\gamma(\gamma+1)})$. A closer inspection of g shows that this density is bounded if $\gamma \neq 1$. If $\gamma = 1$, then the density is bounded everywhere except when $r \downarrow 0$.

In the next section, we consider a statistical application of this result, belonging to the domain of hypothesis testing. We then turn to the proofs of our results. Theorem 1 is proven in Section 3, while Sections 4–6 are devoted to the proof of Theorem 2. Specifically, the main steps of this proof are provided in Section 4. One of these steps is to analyze the singular value spectrum of $X^{(n)}J^{(n)}X^{(n)*} - z$ for $z \in \mathbb{C}$. This will be done in Section 5. Finally, Section 6 is devoted to the identification of the measure μ.

2 Application to statistical hypothesis testing

Consider the high dimensional linear moving average time series model

$$y_t^{(n)} = \sum_{i=0}^{p} B_i^{(n)} w_{t-i}^{(n)}, \tag{3}$$

where $\{B_i^{(n)}\}_{i=0}^p$ are $\mathbb{C}^{N\times N}$ deterministic parameter matrices, and $\{w_i^{(n)}\}_i$ are random vectors such that the random matrix $W^{(n)} = (w_0^{(n)} \cdots w_{n-1}^{(n)})$ is equal in distribution to $n^{1/2}X^{(n)}$. Such models have found increasing attention in, *e.g.*, the fields of signal processing, wireless communications, Radar, Sonar, and wideband antenna array processing [22, 37]. The sample autocovariance matrices $\{n^{-1}\sum_{t=k+1}^n y_t^{(n)}(y_{t-k}^{(n)})^*\}, k \geq 0$, ($k$ is called the lag or the step) carry useful information about the model (3), specially through their spectral distributions. Some of the works that deal with limit spectral distributions, mostly for high-dimensional real-valued time series, and their use in statistical inference are, [2, 3, 4, 8, 28, 39, 26, 25, 5].

The k-step sample autocovariance matrices ($k \neq 0$), are non-Hermitian. To the best of our knowledge LSD results are known only for certain symmetrized versions of these matrices. All references cited above rely on symmetrization. The result of Theorem 2 is a beginning towards deriving the LSD of the sample autocovariance matrices in the general model (3) by considering the simplest case where $B_0^{(n)} = I_N$ and $p = 0$. This will be called the white noise model.

Consider the problem of testing the white noise model against an MA correlated model. To this end, we explore the idea of designing a test which is based on the eigenvalue distribution of the one-step sample autocovariance matrix, in contrast to more classical tests that are based on its singular value distribution. A non-rigorous justification of this idea is that when performing an eigenvalue-based test, we take advantage of the higher sensitivity of the eigenvalues of a matrix with respect to perturbations as compared to its singular values.

Assuming for simplicity that $p = 1$, our purpose is to test the null (white noise) hypothesis H0: $B_0^{(n)} = I, B_1^{(n)} = 0$ against the alternative H1: $B_0^{(n)} = I, B_1^{(n)} \neq 0$. Consider the one-step sample autocovariance matrix

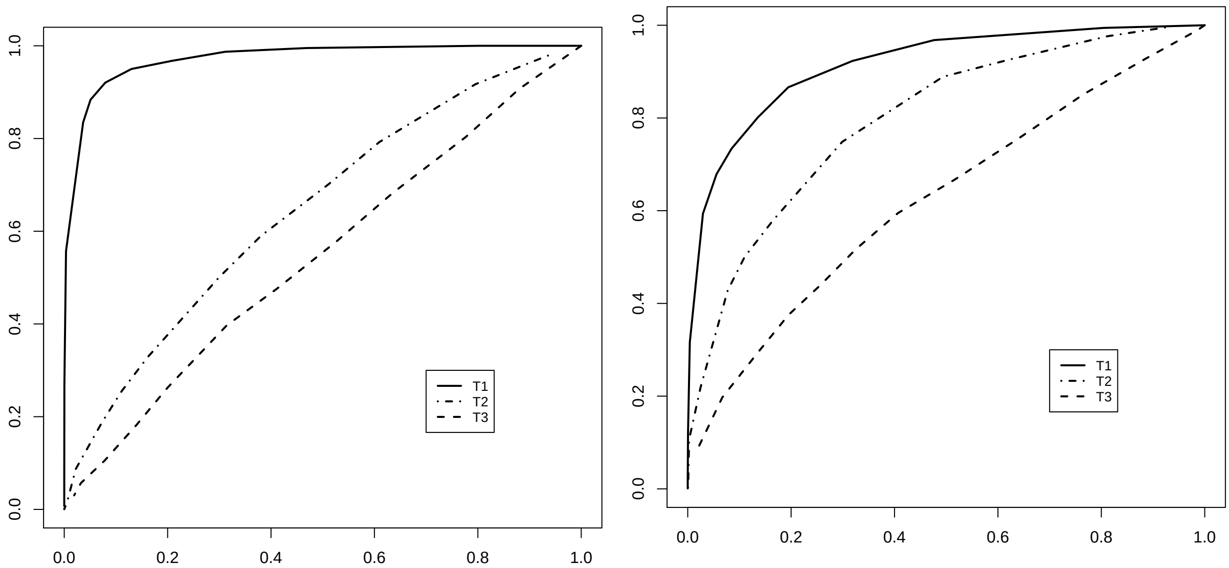
$$\widehat{R}_1^{(n)} = \frac{1}{n}\sum_{t=0}^{n-1} y_t^{(n)} {y_{t-1}^{(n)}}^*,$$

where the sum is taken modulo n, and observe that under H0, this matrix coincides with $X^{(n)}J^{(n)}X^{(n)*}$. We shall consider the asymptotic regime where $n \to \infty$ and $N/n \to \gamma > 0$. By Theorem 2, the spectral measure of $\widehat{R}_1^{(n)}$ converges weakly in probability to the measure μ. This suggests the use of a white noise test based on a distance between the spectral measure of $\widehat{R}_1^{(n)}$ and μ. We consider herein a test based on the 2-Wasserstein distance between these two distributions, which is known to produce the same topology as the weak convergence topology. For the sake of comparison, we also considered the more classical singular value based test which consists in comparing $N^{-1}\operatorname{tr}\widehat{R}_1^{(n)}(\widehat{R}_1^{(n)})^*$ to a threshold. We denote these two tests as T_1 and T_2 respectively. To get a more complete picture of the problem, we also considered a third test which is based on the eigenvalue distribution of the Hermitian sample covariance matrix

$$\widehat{R}_{0,1}^{(n)} = \frac{1}{n}\sum_{t=0}^{n-1} \begin{pmatrix} y_t^{(n)} \\ y_{t-1}^{(n)} \end{pmatrix} \begin{pmatrix} y_t^{(n)*} & {y_{t-1}^{(n)}}^* \end{pmatrix}.$$

Its spectral distribution is known to converge weakly almost surely under H0 to the Marchenko-Pastur distribution $\text{MP}_{2\gamma}$ with parameter 2γ (see [29], which deals with the Gaussian case). This suggests the use of the 2-Wasserstein distance between the spectral measure of $\widehat{R}_{0,1}$ and $\text{MP}_{2\gamma}$. We denote the resulting test as T_3.

Figures 1a and 1b give the ROC curves for these tests. Tests T_1 and T_3 were implemented by sampling μ and $\text{MP}_{2\gamma}$ from the spectra of two large random matrices and by using the `transport` library of the `R` software. These figures clearly show that T_1 outperforms T_2 and T_3. This tends to corroborate the intuition that the eigenvalue sensitivity alluded to earlier, can be beneficial when it comes to designing white noise tests.



(a) $B_1^{(n)} = \alpha I_N$ with $\alpha^2 = 10^{-2.5}$.

(b) $B_1^{(n)} = (b_{ij}^{(n)})$ Toeplitz with $b_{ij}^{(n)} \propto \exp(-8|i-j|/N)$ and $\operatorname{tr} B_1^{(n)} (B_1^{(n)})^* / N = 10^{-2}$.

Fig. 1: ROC curves for the tests T_1, T_2 and T_3 described in Section 2. Setting: $(N, n) = (50, 100)$

To better understand the behavior of the eigenvalue-based tests, the next step would be to study the large dimensional behavior of the spectral distribution of $\widehat{R}_1^{(n)}$ under H1. It would be also useful to evaluate the test performance for Wasserstein distances of orders higher than two. These tasks appear to be quite non-trivial and are left for future research.

Before entering our proofs, let us present some notations. The dimension of a subspace V of some vector space will be denoted $\dim(V)$. The orthogonal subspace to V will be $V^\perp$. The column span of a matrix M will be denoted as $\operatorname{span}(M)$. Similarly, $\operatorname{span}(V, d)$ is the span of the vector space V and the vector d.

The indices of the elements of a vector or a matrix start from zero. Given a positive integer m, we write $[m] = \{0, \ldots, m-1\}$. For $i \in [m]$, we denote as $e_{m,i}$ the i^{th} canonical vector of $\mathbb{C}^m$, with 1 at the mth place and 0 elsewhere. Given a matrix $M \in \mathbb{C}^{m \times n}$ and two sets $\mathscr{I} \subset [m]$ and $\mathscr{J} \subset [n]$, $M_{\mathscr{I},\mathscr{J}}$ denotes the $|\mathscr{I}| \times |\mathscr{J}|$ supmatrix of M that is obtained by retaining the rows and columns of M whose indices belong to $\mathscr{I}$ and $\mathscr{J}$ respectively. We also write $M_{\cdot,k} = M_{[m],\{k\}}$ and $M_{k,\cdot} = M_{\{k\},[n]}$. We define as $\Pi_{\mathscr{I}} : \mathbb{C}^m \to \mathbb{C}^m$ the projection operator such that $\Pi_{\mathscr{I}} u$ is the vector obtained by setting to zero the elements of u whose indices are in $\mathscr{I}^{\mathrm{c}}$. We also denote as $u_{\mathscr{I}}$ the vector of $\mathbb{C}^{|\mathscr{I}|}$ obtained by removing the elements of u whose indices are in $\mathscr{I}^{\mathrm{c}}$. When M is a matrix, Π_M refers to the orthogonal projector on $\operatorname{span}(M)$.

As mentioned above, $\|\cdot\|$ denotes the spectral norm. It will also denote the Euclidean norm of a vector. The Hilbert-Schmidt norm of a matrix will be denoted as $\|\cdot\|_{\mathrm{HS}}$. The unit-sphere of $\mathbb{C}^n$ will be denoted as $\mathbb{S}^{n-1}$.

The notations $\mathbb{P}_x$ and $\mathbb{E}_x$ will refer respectively to the probability and the expectation with respect to the law of the vector x.

3 Proof of Theorem 1: Smallest singular value

To simplify the notations, from now on, we omit the superscript $^{(n)}$. Writing $\mathbf{s}_{\inf} = \inf_n s_{n-1}(A^{-1})$ and $\mathbf{s}_{\sup} = \sup_n s_0(A^{-1})$, Assumption 2 is rewritten as $0 < \mathbf{s}_{\inf} \leq \mathbf{s}_{\sup} < \infty$. We also assume that $z \neq 0$ without further mention.

3.1 General context and outline of proof

It is enough to establish Theorem 1 under the assumption that the entries have densities. This is because we can replace the matrix X with, the independent sum $(1 - n^{-20})^{-1/2}(X + n^{-10}X')$ where X' is a properly chosen matrix whose elements have densities, and use a standard perturbation argument. Hence, we assume throughout this section that the elements of X have densities.

Suppose $E \in \mathbb{C}^{N\times n}$ is such that $\|E\|^2\|A\| < |z|$. Then $\det(z - EAE^*) \neq 0$. This implies that the multivariate polynomial $\det(z - XAX^*)$ in the variables $(\Re x_{ij}, \Im x_{ij})_{i,j}$ is not identically zero. Since X has a density, we conclude that $z - XAX^*$ is invertible w.p. 1. Define the matrix

$$H = \begin{pmatrix} A^{-1} & X^* \\ X & z \end{pmatrix} \in \mathbb{C}^{(N+n)\times(N+n)}.$$

By the inversion formula for partitioned matrices, the $N\times N$ lower-right block of H^{-1} coincides with $(z - XAX^*)^{-1}$. Thus, $\|(XAX^* - z)^{-1}\| \le \|H^{-1}\|$. Therefore, it is enough to prove that

$$\mathbb{P}\left[s_{N+n-1}(H) \le t,\ \|X\| \le C\right] \le c\left(n^{\alpha} t^{1/2} + n^{-\beta}\right), \tag{4}$$

where $c > 0$ depends on C, z, and m_4 only.

As mentioned earlier, we shall follow here the argument of [38] but we shall need substantial modifications. Here is a description of the general approach. First recall that

$$s_{N+n-1}(H) = \min_{u\in\mathbb{S}^{N+n-1}} \|Hu\|.$$

Invoking an idea that has been frequently used in the literature since [27, 33], we partition $\mathbb{S}^{N+n-1}$ into two sets of so-called compressible and incompressible vectors as follows.

Let $\theta, \rho \in (0,1)$ be fixed. A vector in $\mathbb{S}^{N+n-1}$ is said to be θ-sparse if it does not have more than $\lfloor \theta(N+n) \rfloor$ non-zero elements. Let $\mathbb{S}^{N+n-1}_{\mathscr{I}}$ be the set of vectors of $\mathbb{S}^{N+n-1}$ that are supported by the (index) set $\mathscr{I} \subset [N+n]$. Given $S \subset \mathbb{C}^{N+n}$, let $\mathscr{N}_\delta(S)$ denote the δ-neighborhood of S in $\mathbb{C}^{N+n}$ in the Euclidean metric. Given $\theta, \rho \in (0,1)$, the set of (θ,ρ)-compressible vectors is

$$\mathrm{comp}(\theta,\rho) = \mathbb{S}^{N+n-1} \cap \bigcup_{\substack{\mathscr{I}\subset[N+n] \\ |\mathscr{I}| = \lfloor\theta(N+n)\rfloor}} \mathscr{N}_\rho(\mathbb{S}^{N+n-1}_{\mathscr{I}}).$$

This is the set of all unit vectors at a distance at most ρ from the set of the θ-sparse unit vectors. The complementary set $\mathrm{incomp}(\theta,\rho) = \mathbb{S}^{N+n-1} \setminus \mathrm{comp}(\theta,\rho)$ is the set of incompressible vectors.

With these notations, for judiciously chosen $\theta, \rho \in (0,1)$, we write

$$s_{N+n-1}(H) = \inf_{u\in\mathrm{comp}(\theta,\rho)} \|Hu\| \wedge \inf_{u\in\mathrm{incomp}(\theta,\rho)} \|Hu\|. \tag{5}$$

The infimum over $\mathrm{comp}(\theta,\rho)$ is relatively easier to handle. Given a fixed vector $u \in \mathbb{S}^{N+n-1}$, we first show that $\mathbb{P}[\|Hu\| \le c]$ for some $c > 0$ is exponentially small in n. Recall that an ε-net is a set of points that are separated from each other by a distance of at most ε. Now, since the vectors of $\mathrm{comp}(\theta,\rho)$ are close to being sparse, it has an ε-net of controlled cardinality for a well-chosen $\varepsilon > 0$. Using this, along with a simple union bound, we will be able to infer the smallness of the probability that $\inf_{u\in\mathrm{comp}(\theta,\rho)} \|Hu\|$ is small.

The infimum over the set of incompressible vectors poses a much bigger challenge since the ε-net argument fails. In this case the argument is more geometric. Observe that when u is incompressible, Hu is close to a sum of $\mathscr{O}(n)$ columns of H with comparable weights. This helps to reduce the problem of controlling $\inf_{u\in\mathrm{incomp}(\theta,\rho)} \|Hu\|$ to the problem of controlling the distance between an arbitrary column of H and the subspace generated by the other columns.

Let h_0 be the first column of H, and let $H_{-0} \in \mathbb{C}^{(N+n)\times(N+n-1)}$ be the supmatrix left after extracting this column. Partition H accordingly as

$$H = \begin{pmatrix} b & g_{01} \\ g_{10} & G \end{pmatrix},$$

with $b \in \mathbb{C}$ and $G \in \mathbb{C}^{(N+n-1)\times(N+n-1)}$. In Lemma 11 below, we show that G^{-1} exists w.p. 1. Noting that the distance $\mathrm{dist}(h_0, H_{-0})$ between h_0 and the column span of H_{-0} satisfies $\mathrm{dist}(h_0,H_{-0})^2 = h_0^*(I - H_{-0}(H_{-0}^* H_{-0})^{-1} H_{-0}^*) h_0$, and making use of the Sherman-Morrison-Woodbury formula [16, Chap. 2], we get after a small derivation that

$$\mathrm{dist}(h_0, H_{-0}) = \frac{\left|b - g_{01} G^{-1} g_{10}\right|}{\sqrt{1 + \|g_{01} G^{-1}\|^2}}. \tag{6}$$

Our purpose is to bound the probability that this distance is small. If we write

$$A^{-1} = \begin{pmatrix} b & b_{01} \\ b_{10} & B \end{pmatrix}, \quad X = \begin{pmatrix} x & W \end{pmatrix},$$

where $B \in \mathbb{C}^{(n-1)\times(n-1)}$, and $x \in \mathbb{C}^N$ is the first column of X, then

$$g_{01} = \begin{pmatrix} b_{01} & x^* \end{pmatrix},\ g_{10} = \begin{pmatrix} b_{10} \\ x \end{pmatrix},\ \text{and } G = \begin{pmatrix} B & W^* \\ W & zI_N \end{pmatrix}. \tag{7}$$

Assuming that the inverse exists, partition G^{-1} as

$$G^{-1} = \begin{pmatrix} E & F \\ P & R \end{pmatrix}, \quad E \in \mathbb{C}^{(n-1)\times(n-1)},\ R \in \mathbb{C}^{N\times N}. \tag{8}$$

Then using Equation (6), we have

$$\mathrm{dist}(h_0, H_{-0}) = \frac{\mathrm{Num}}{\mathrm{Den}} \quad \text{w.p. } 1, \text{ where,} \tag{9}$$

$$\begin{aligned} \mathrm{Num} &= |b - b_{01}Eb_{10} - x^*Pb_{10} - b_{01}Fx - x^*Rx|, \quad \text{and} \\ \mathrm{Den} &= \left(1 + \|b_{01}E + x^*P\|^2 + \|b_{01}F + x^*R\|^2\right)^{1/2}. \end{aligned} \tag{10}$$

To control the behavior of Num, we need an anti-concentration result. Loosely speaking, we show that conditionally on the matrix W and for most of these matrices, the probability that a properly normalized version of the random variable $x^*Pb_{10} + b_{01}Fx + x^*Rx$ lives in an arbitrary ball of $\mathbb{C}$ of small radius is itself small.

Small ball probabilities are captured by the so-called Lévy's concentration function. Given a constant vector $a \in \mathbb{C}^n$ and a random vector $Z \in \mathbb{C}^n$, Lévy's concentration function of the inner product $\langle a, Z\rangle$ at $\varepsilon > 0$ is

$$\mathscr{L}_Z(\langle a, Z\rangle, \varepsilon) = \sup_{w\in\mathbb{C}} \mathbb{P}_Z\left[|\langle a, Z\rangle - w| \leq \varepsilon\right].$$

When the elements of Z are i.i.d. random variables with finite third moment, the behavior of $\mathscr{L}_Z$ can be controlled by the Berry-Esséen theorem, whose use in random matrix theory dates back to [27]. Berry-Esséen theorem is a refinement of the Central Limit Theorem and implies that when $a \in \mathbb{S}^{n-1}$ has $\mathscr{O}(n)$ elements with magnitudes of order $1/\sqrt{n}$, it holds that $\mathscr{L}_Z(\langle a, Z\rangle, \varepsilon) \lesssim \varepsilon + 1/\sqrt{n}$.

Our plan now is to apply this theorem after replacing Z with the random vector x. Unfortunately, this theorem cannot be used as is on the random variable $x^*Pb_{10} + b_{01}Fx + x^*Rx$ because of the presence of the quadratic form x^*Rx. To circumvent this problem, we use a decoupling argument that replaces $x^*Pb_{10} + b_{01}Fx + x^*Rx$ with an inner product that can be processed by the Berry-Esséen theorem. This decoupling idea that dates back to [17] has also been used in [38].

3.2 Technical results

The following proposition is an easy variation of [35, Proposition 5.1], see also [7, Lemma A2] and [19].

Proposition 3 (Distance of a random vector to a constant subspace)**.** Let $Z = (Z_0, \ldots, Z_{n-1}) \in \mathbb{C}^n$ be a vector of i.i.d. centered unit-variance random variables such that for some $\kappa > 0$, $\mathbb{E}|Z_0|^{2+\kappa} \leq C_\kappa < \infty$. Then, there exist $c, c' > 0$ and $\delta \in (0,1)$ that depend only on κ and C_κ and that satisfy the following property. For all $n \gg 1$, and for any deterministic subspace V of $\mathbb{C}^n$ such that $0 \leq \dim(V) \leq \delta n$,

$$\mathbb{P}[\mathrm{dist}(Z, V) \leq c\sqrt{n}] \leq \exp(-c'n).$$

This result leads to the following lemma:

Lemma 4. Let the matrix X satisfy Assumption 1. Then, there exist constants $c, c' > 0$ and a constant $\delta \in (0,1)$ that depend on m_4 only and that satisfy the following property. For each deterministic vector $u \in \mathbb{S}^{n-1}$ and each deterministic subspace $V \subset \mathbb{C}^N$ with $0 \leq \dim(V) \leq \delta N$,

$$\mathbb{P}\left[\mathrm{dist}(Xu, V) \leq c\right] \leq \exp(-c'n). \tag{11}$$

In particular, for each deterministic vector $a \in \mathbb{C}^N$, it holds that $\mathbb{P}\left[\|Xu - a\| \leq c\right] \leq \exp(-c'n)$. Similar conclusions hold if X is replaced with X^*.

Proof. Let $\tilde{x}_0, \ldots, \tilde{x}_{N-1} \in \mathbb{C}^{1\times n}$ be the rows of X, and define the random variables $Z_k = \sqrt{n}\tilde{x}_k u$ for $k \in [n]$. These random variables are i.i.d., centered, and have unit-variance. Furthermore, writing $u = (u_0, \ldots, u_{n-1})^\mathsf{T}$, we get by Rosenthal's inequality that for some universal constant C,

$$\mathbb{E}|Z_1|^4 \leq C\left((n^2\mathbb{E}|X_{11}|^4 \sum |u_i|^4) \ \vee\ 1\right) \leq Cm_4.$$

Writing $Z = (Z_0, \ldots, Z_{N-1})^\mathsf{T}$, we note that $\mathrm{dist}(Xu, V) = \mathrm{dist}(Z, V)/\sqrt{n}$. Applying Proposition 3 with $\kappa = 2$, we obtain (11). The rest of the claims follow immediately. □

The two following results regarding Lévy's concentration functions will be needed.

Lemma 5 (Lemma 2.1 of [33])**.** Let $Z \in \mathbb{C}^n$ be a vector of independent random variables. Then, for each non-empty $\mathscr{I} \subset [n]$, we have $\mathscr{L}_Z(\langle a, Z\rangle, \varepsilon) \leq \mathscr{L}_{Z_\mathscr{I}}(\langle a_\mathscr{I}, Z_\mathscr{I}\rangle, \varepsilon)$.

Proposition 6. There exists a constant $c > 0$ such that for any vector $Z = (Z_0, \ldots, Z_{n-1})$ of complex centered independent random variables with finite third moments,

$$\mathscr{L}_Z\Big(\sum Z_i, t\Big) \le \frac{ct}{\sqrt{\sum \mathbb{E}|Z_i|^2}} + \frac{c\sum \mathbb{E}|Z_i|^3}{(\sum \mathbb{E}|Z_i|^2)^{3/2}}.$$

For a proof, see [7, Lemma A6]. In particular, if there exist two positive constants c_2 and c_3 such that $\mathbb{E}|Z_i|^2 \ge c_2$ and $\mathbb{E}|Z_i|^3 \le c_3$ for each $i \in [n]$, then

$$\mathscr{L}_Z\Big(\sum Z_i, t\sqrt{n}\Big) \le c' t + \frac{c''}{\sqrt{n}}, \tag{12}$$

where $c' = c/\sqrt{c_2}$ and $c'' = cc_3/c_2^{3/2}$.

We now enter the proof of Theorem 1 via proving Inequality (4). Recall that we have written $X = \begin{pmatrix} x & W \end{pmatrix}$ where x is the first column of X. Given $C > 0$, we denote as $\mathscr{E}_{\rm op}(C)$ the probability event

$$\mathscr{E}_{\rm op}(C) = [\|W\| \le C].$$

In the remainder of this section, the constants that do not depend on n will be referred to by the letter c, possibly with primes or numerical indices referring to the propositions or lemmas where these constants appear for the first time. In all statements of the type $\mathbb{P}[[\cdots \le c] \cap \mathscr{E}] \le \exp(-c'n) + c_1 n^{-\gamma}$, where $\mathscr{E} = [\|X\| \le C]$ or $\mathscr{E}_{\rm op}(C)$, the constants such as c, c', or c_1 depend on C, z, and m_4 at most.

3.3 Compressible vectors

Recalling (5), we start with the compressible vectors.

Proposition 7. Let Assumption 1 hold true. Then, there exists $\theta_7 \in (0,1)$, $\rho_7 > 0$, $c > 0$ and $c' > 0$ such that

$$\mathbb{P}\Big[\big[\inf_{u \in \mathrm{comp}(\theta_7, \rho_7)} \|Hu\| \le c\big] \cap [\|X\| \le C]\Big] \le \exp(-c'n) \quad \text{for large enough} \quad n.$$

Proof. We first show that there exist $c_0, c_1 > 0$ such that for each deterministic vector $u \in \mathbb{S}^{N+n-1}$,

$$\mathbb{P}[\|Hu\| \le c_0] \le \exp(-c_1 n). \tag{13}$$

Write $u = (v^{\mathsf{T}}, w^{\mathsf{T}})^{\mathsf{T}}$, where $v \in \mathbb{C}^n$ and $w \in \mathbb{C}^N$. Since $\|u\| = 1$, either $\|v\| \ge 1/\sqrt{2}$ or $\|w\| \ge 1/\sqrt{2}$. Assume that $\|w\| \ge 1/\sqrt{2}$, and note that $[\|Hu\| \le c_0] \subset [\|A^{-1}v + X^* w\| \le c_0]$. Writing $\tilde{w} = w/\|w\|$, we have

$$\mathbb{P}\left[\|X^* w + A^{-1} v\| \le c_0\right] = \mathbb{P}\left[\|X^*\tilde{w} + A^{-1}v/\|w\|\| \le c_0/\|w\|\right] \le \mathbb{P}\left[\|X^*\tilde{w} + A^{-1}v/\|w\|\| \le c_0\sqrt{2}\right] \le \exp(-c_1 n)$$

by applying Lemma 4 and choosing c_0 and c_1 judiciously. When $\|v\| \ge 1/\sqrt{2}$, we can use a similar argument (with possibly different $c_0$and c_1) after observing that $[\|Hu\| \le c_0] \subset [\|Xv + zw\| \le c_0]$. This establishes (13).

Now, on the event $[\|X\| \le C]$, we have

$$\|H\| \le \left\| \begin{pmatrix} & X^* \\ X & \end{pmatrix} \right\| + \left\| \begin{pmatrix} A^{-1} & \\ & z \end{pmatrix} \right\| \le C_H \triangleq C + |z| \vee \mathbf{s}_{\rm sup}.$$

On this event, assume that there exists $y \in \mathscr{N}_{c_0/(2C_H)}(\{u\})$ such that $\|Hy\| \le c_0/2$. Then $\|Hu\| \le \|H(u-y)\| + \|Hy\| \le c_0$. In other words,

$$\left[\exists y \in \mathscr{N}_{c_0/(2C_H)}(\{u\}) : \|Hy\| \le c_0/2\right] \cap [\|X\| \le C] \subset [\|Hu\| \le c_0].$$

We now use a ε-net argument. Let $\theta_7 \in (0,1)$ to be fixed in a moment, and choose $\mathscr{I} \subset [N+n]$ so that that $|\mathscr{I}| = \lfloor \theta_7(n+N) \rfloor$. The unit-sphere $\mathbb{S}^{N+n-1}_{\mathscr{I}}$ of the subspace of the vectors of $\mathbb{C}^{N+n}$ that are supported by $\mathscr{I}$ has a $(c_0/(2C_H))$-net of cardinality bounded by $(6C_H/c_0)^{2|\mathscr{I}|}$ (see, *e.g.*, [10, Lemma 2.2]). Applying the previous results and making use of the union bound, we get that

$$\mathbb{P}\left[\left[\exists y \in \mathscr{N}_{c_0/(2C_H)}(\mathbb{S}^{N+n-1}_{\mathscr{I}}) : \|Hy\| \le c_0/2\right] \cap [\|X\| \le C]\right] \le (6C_H/c_0)^{2\theta_7(N+n)} \exp(-c_1 n).$$

Finally, considering all the sets $\mathscr{I} \subset [N+n]$ such that $|\mathscr{I}| = \lfloor \theta_7(N+n) \rfloor$, recalling the elementary bound on the binomial coefficients $\binom{m}{k} \le (em/k)^k$, and using the union bound, we get that

$$\mathbb{P}[[\exists y \in \mathrm{comp}(\theta_7, c_0/(2C_H)) : \|Hy\| \le c_0/2 \cap [\|X\| \le C]] \le \left(\frac{36eC_H^2}{\theta_7 c_0^2}\right)^{\theta_7(N+n)} \exp(-c_1 n).$$

A small calculation shows that the right hand side is of the form $\exp(-c'n)$ for large enough n when θ_7 is chosen small enough. By taking $\rho_7 = c_0/(2C_H)$, the proposition is proven. $\square$

3.4 Incompressible vectors

One main feature of incompressible vectors of $\mathbb{C}^n$ is that they contain $\mathcal{O}(n)$ elements of absolute values of order $\mathcal{O}(n^{-1/2})$, as shown in [33, Lemma 3.4]. We shall need the following slightly stronger version of this lemma with a similar proof.

Lemma 8. Let $u = (u_0,\dots,u_{n-1})^\mathsf{T} \in \operatorname{incomp}(\theta,\rho)$, and let $\tilde{u} = (\tilde{u}_0,\dots,\tilde{u}_{n-1})^\mathsf{T} \in \mathbb{S}^{n-1}$. Then, $|J| \geq \theta n/2$ where,

$$J = \{i \in [n] : \frac{\rho}{\sqrt{n}} \leq |u_i| \leq \frac{2}{\sqrt{\theta n}} \text{ and } |\tilde{u}_i| \leq \frac{2}{\sqrt{\theta n}}\}.$$

One consequence of [33, Lemma 3.4] is the following lemma, which implies that the infimum of $\|Hu\|$ over a set of incompressible vectors can be handled by controlling the distance between an arbitrary column of H and the subspace generated by the other columns:

Lemma 9 (Invertibility via mean distance, Lemma 3.5 of [33]). Let $M \in \mathbb{C}^{n\times n}$ be a random matrix. Let m_k be the kth column of M and let $M_{-k} \in \mathbb{C}^{n\times(n-1)}$ be the supmatrix left after removing this column. Then,

$$\mathbb{P}\left[\inf_{u\in\operatorname{incomp}(\theta,\rho)} \|Mu\| \leq \frac{\rho t}{\sqrt{n}}\right] \leq \frac{2}{\theta n}\sum_{k=0}^{n-1} \mathbb{P}\left[\operatorname{dist}(m_k, M_{-k}) \leq t\right].$$

Using Lemma 9, we need to control the distance between a column of H and the subspace generated by the other columns. Denote as x_k the k^{th} column of X (thus, $x_0 = x$). Let b_k and $\tilde{x}_\ell$ denote the k^{th} column of A^{-1} and the ℓ^{th} row of X respectively. Then the columns of H are one of the two types: $\begin{pmatrix} b_k \\ x_k \end{pmatrix}$, or $\begin{pmatrix} \tilde{x}_\ell^* \\ z e_{N,\ell} \end{pmatrix}$. Due to the fact that A is not necessarily a diagonal matrix, it will be more difficult to control the distances involving columns of the first type.

Recall the partition $H = \begin{pmatrix} h_0 & H_{-0} \end{pmatrix}$, where h_0 is the first column of H. We have

Proposition 10. Let Assumptions 1, 2, and 3 hold true. Then

$$\mathbb{P}[[\operatorname{dist}(h_0,H_{-0}) \leq t] \cap [\|X\| \leq C]] \leq c_1(n^{59/88}t^{1/2} + n^{-1/22}) + \exp(-c_2 n).$$

Since $[\|X\| \leq C]$ is obviously included in $\mathcal{E}_{\text{op}}(C)$, it will be enough to establish the inequality

$$\mathbb{P}[[\operatorname{dist}(h_0,H_{-0}) \leq t] \cap \mathcal{E}_{\text{op}}(C)] \leq c_1(n^{59/88}t^{1/2} + n^{-1/22}) + \exp(-c_2 n)$$

to obtain Proposition 10. Replacing $[\|X\| \leq C]$ with $\mathcal{E}_{\text{op}}(C)$ will be more convenient due to the independence of x and $\mathcal{E}_{\text{op}}(C)$. The remainder of this section is devoted towards proving this inequality. Our starting point will be Equation (6). To be able to use it, we need to check that G defined in (7) is invertible. Recall that X is assumed to have a density.

Lemma 11. The matrix G is invertible with probability one.

Proof. Since $z \neq 0$, the matrix zI_N is invertible. Thus, to show that G is invertible with the probability one, we need to show that the Schur complement $\Delta = B - z^{-1}W^*W$ of zI_N in G is invertible with probability one.

Since $A^{-1} = \begin{pmatrix} b & b_{01} \\ b_{10} & B \end{pmatrix}$, it follows that $\operatorname{rank}(B) \geq n-2$. Thus, either B is invertible or $\operatorname{rank}(B) = n-2$.

Assume it is invertible. Then on the set $\{W \in \mathbb{C}^{N\times(n-1)} : \|z^{-1}W^*W\| \leq s_{n-2}(B)/2\}$, it holds that $s_{n-2}(\Delta) \geq s_{n-2}(B) - \|z^{-1}W^*W\| \geq s_{n-2}(B)/2 > 0$. Thus, $\det(\Delta)$ is a non-zero multivariate polynomial in the real and imaginary parts of the elements of W. Since W has a density, $\det(\Delta) \neq 0$ w.p. 1.

Assume now that $\operatorname{rank}(B) = n-2$. Then we can write $B = UV^*$ where $U, V \in C^{(n-1)\times(n-2)}$ are full column-rank matrices. Writing $W^* = \begin{pmatrix} w & Y \end{pmatrix}$ where $w \in \mathbb{C}^{n-1}$, we get that

$$B - z^{-1}W^*W = \begin{pmatrix} U & z^{-1}w \end{pmatrix}\begin{pmatrix} V & -w \end{pmatrix}^* - z^{-1}YY^* = D - z^{-1}YY^*.$$

Given a vector $u \perp \operatorname{span}(U)$ where $\perp$ denotes the orthogonality, the inner product u^*w is a continuous random variable, thus $u^*w \neq 0$ w.p. 1. Consequently, $w \notin \operatorname{span}(U)$ w.p. 1., which implies that $\begin{pmatrix} U & z^{-1}w \end{pmatrix}$ is invertible w.p. 1. The same argument holds for $\begin{pmatrix} V & -w \end{pmatrix}$, and thus the matrix D is invertible w.p. 1. To obtain that Δ is invertible, it remains to apply the previous argument after replacing B with D and W^* with Y, and making use of the independence of w and Y along with the Fubini-Tonelli theorem. □

From Equation (6), $\operatorname{dist}(h_0,H_{-0}) = \text{Num}/\text{Den}$ w.p. 1, where Num and Den are as given in (10). To study the behavior of these random variables, we first need to show that the image of each deterministic vector by the matrix R at the right hand side of (8) is incompressible with high probability. This will be stated in the corollary of Proposition 13 below.

Lemma 12. $s_{n-3}(B) \geq \mathbf{s}_{\inf}$.

Proof. The matrix $b_{10}b_{10}^* + BB^*$ is a principal supmatrix of the Hermitian matrix $A^{-1}(A^{-1})^*$. Using the variational representation of the eigenvalues of $A^{-1}(A^{-1})^*$, we get that $s_{n-2}(b_{10}b_{10}^* + BB^*) \geq \mathbf{s}_{\inf}^2$. By Weyl's interlacing inequalities [23, Chap. 4], $s_{n-3}(BB^*) \geq s_{n-2}(b_{10}b_{10}^* + BB^*)$, hence the result. □

Proposition 13. There exist $\theta_{13} \in (0,1)$, $\rho_{13} > 0$, and $c_{13} > 0$ such that for each $d \in \mathbb{C}^N$,

$$\mathbb{P}\left[\left[\inf_{\substack{v\in\mathbb{C}^{n-1},\\ w\in\mathrm{comp}(\theta_{13},\rho_{13})}} \mathrm{dist}\left(G\begin{pmatrix} v \\ w\end{pmatrix}, \mathrm{span}\left(\begin{pmatrix} 0 \\ d\end{pmatrix}\right)\right) \le \rho_{13}\right] \cap \mathscr{E}_{\mathrm{op}}(C)\right] \le \exp(-c_{13}n).$$

Proof. Let $\theta_{13} \in (0,1)$ and $t \in (0,1)$ to be fixed later. Let $\mathscr{I} \in [N]$such that $|\mathscr{I}| = \lfloor \theta_{13}N \rfloor$. Fix an element w of the unit-sphere $\mathbb{S}^{N-1}_{\mathscr{I}}$. In this first part of the proof, we shall control the probability of the event

$$\left[\inf_{v\in\mathbb{C}^{n-1}} \mathrm{dist}\left(G\begin{pmatrix} v \\ w\end{pmatrix}, \mathrm{span}\left(\begin{pmatrix} 0 \\ d\end{pmatrix}\right)\right) \le t\right] \cap \mathscr{E}_{\mathrm{op}}(C). \tag{14}$$

Given two elements a and b of some vector space on $\mathbb{C}$, it holds that $\mathrm{dist}(a, \mathrm{span}\, b) = \inf_{\alpha\in\mathbb{C}} \|a - \alpha b\|$. Thus, the event between [ ] brackets in (14) is included in the event

$$\mathscr{E}_w(t) = \left[\exists v \in \mathbb{C}^{n-1}, \exists \alpha \in \mathbb{C} : \|Bv + W^*w\| \le t, \|Wv + zw + \alpha d\| \le t\right]. \tag{15}$$

Let

$$B = \begin{pmatrix} P & p \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & s_{n-2}(B) \end{pmatrix} \begin{pmatrix} Q^* \\ q^* \end{pmatrix} \tag{16}$$

be a singular value decomposition of B, where p and q are respectively the last columns of the unitary matrices $\begin{pmatrix} P & p \end{pmatrix}$ and $\begin{pmatrix} Q & q \end{pmatrix}$. Given any vector $y \in \mathbb{C}^{n-1}$, we shall use in the remainder of the proof the notations $y_Q = \Pi_Q y$ and $y_q = \Pi_q y$, making $y = y_Q + y_q$ an orthogonal sum. As is well-known [16, Chap. 5], the vector $u = -B^\dagger W^* w$ where $B^\dagger$ is the Moore-Penrose pseudo-inverse of B, minimizes $\|By + W^*w\|$ with respect to y. Assume that there is a solution $v \in \mathbb{C}^{n-1}$ of the inequality $\|By + W^*w\| \le t$ in y. Then, since u is also a solution, we get that $\|B(u_Q - v_Q) + B(u_q - v_q) + Bv + W^*w\| \le t$, and hence, $\|B(u_Q - v_Q) + B(u_q - v_q)\| \le \|Bv + W^*w\| + t \le 2t$. Noting that $B(u_Q - v_Q)$ and $B(u_q - v_q)$ are orthogonal, we get that $\|B(u_Q - v_Q)\| \le 2t$. By Lemma 12, the smallest singular value of the operator B restricted to the subspace $\mathrm{span}(Q)$ is bounded below by $\mathbf{s}_{\inf}$. Hence

$$\|v_Q - u_Q\| \le \frac{2t}{\mathbf{s}_{\inf}}.$$

The vector v also satisfies the inequality $\|Wv + zw + \alpha d\| \le t$ for some $\alpha \in \mathbb{C}$. Thus, $\|W(v_Q - u_Q) + Wv_q + Wu_Q + zw + \alpha d\| \le t$, which implies that on the event $\mathscr{E}_{\mathrm{op}}(C)$,

$$\|Wv_q + Wu_Q + zw + \alpha d\| \le \|W(v_Q - u_Q)\| + t \le \left(1 + \frac{2C}{\mathbf{s}_{\inf}}\right) t.$$

Observing that v_q is collinear with q, we get at this stage of the proof that

$$\mathscr{E}_w(t) \cap \mathscr{E}_{\mathrm{op}}(C) \subset \left[\exists \alpha, \beta \in \mathbb{C}, : \|\beta Wq + Wu_Q + zw + \alpha d\| \le \left(1 + \frac{2C}{\mathbf{s}_{\inf}}\right) t\right] \cap \mathscr{E}_{\mathrm{op}}(C). \tag{17}$$

To proceed, we need to control the Euclidean norm of u_Q. For $m, M > 0$, consider the event

$$\mathscr{E}_{u_Q}(m, M) = [m \le \|u_Q\| \le M].$$

Since $u_Q = -\Pi_Q B^\dagger W^* w$, we get from Lemma 12 that $\mathbf{s}_{\sup}^{-1}\|W^*w\| \le \|u_Q\| \le \mathbf{s}_{\inf}^{-1}\|W^*w\|$. By Lemma 4, there exist $c_0 > 0$ and $c_1 > 0$ such that $\mathbb{P}[\|W^*w\| \le c_0] \le \exp(-c_1 n)$. We thus obtain

$$\mathbb{P}\left[\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0, \mathbf{s}_{\inf}^{-1}C)^{\mathrm{c}} \cap \mathscr{E}_{\mathrm{op}}(C)\right] \le \exp(-c_1 n). \tag{18}$$

To bound the probability of the event at the right hand side of the inclusion (17), we consider separately the situations where $|\beta|$ is large and where $|\beta|$is bounded above. Consider the event

$$\mathscr{E}_{|\beta|>}(m, M) = [\exists \alpha, \beta \in \mathbb{C} : \|\beta Wq + Wu_Q + zw + \alpha d\| \le m, |\beta| \ge M].$$

On $\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0, \mathbf{s}_{\inf}^{-1}C) \cap \mathscr{E}_{\mathrm{op}}(C)$, it holds that

$$\|\beta Wq + Wu_Q + zw + \alpha d\| \ge \|\beta Wq + zw + \alpha d\| - \mathbf{s}_{\inf}^{-1}C^2 \ge |\beta|\, \mathrm{dist}(Wq, \mathrm{span}[w, d]) - \mathbf{s}_{\inf}^{-1}C^2$$

From Lemma 4, there exist $c_2, c_3 > 0$ such that $\mathbb{P}[\mathrm{dist}(Wq, \mathrm{span}[w, d]) \le c_2] \le \exp(-c_3 n)$. Writing $s = (1 + 2C/\mathbf{s}_{\inf})t$, we have

$$\begin{aligned}\mathscr{E}_{|\beta|>}(s, M) \cap \mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0, \mathbf{s}_{\inf}^{-1}C) \cap \mathscr{E}_{\mathrm{op}}(C) &\subset \left[\exists \beta \in \mathbb{C} : |\beta|\, \mathrm{dist}(Wq, \mathrm{span}[w, d]) - \mathbf{s}_{\inf}^{-1}C^2 \le s, |\beta| \ge M\right] \\ &\subset \left[\mathrm{dist}(Wq, \mathrm{span}[w, d]) \le \frac{s + \mathbf{s}_{\inf}^{-1}C^2}{M}\right].\end{aligned}$$

Thus, setting $C' = (s+\mathbf{s}_{\inf}^{-1}C^2)/c_2$, we get that

$$\mathbb{P}\left[\mathscr{E}_{|\beta|>}(s,C')\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\cap\mathscr{E}_{\mathrm{op}}(C)\right]\le\exp(-c_3n). \tag{19}$$

Now consider the case $|\beta|<C'$. We discretize this ball as follows. Consider the event

$$\mathscr{E}_{|\beta|<}(s,C') = \left[\exists\alpha,\beta\in\mathbb{C}:\|\beta Wq+Wu_Q+zw+\alpha d\|\le s,\,|\beta|<C'\right].$$

Given $k,\ell\in\mathbb{Z}$, define the event

$$\mathscr{E}_q(k,\ell,s,C) = \left[\exists\alpha\in\mathbb{C}:\left\|\frac{s}{C\sqrt{2}}(k+\imath\ell)Wq+Wu_Q+zw+\alpha d\right\|\le s\right].$$

For $\beta\in\mathbb{C}$, let $k_\beta=\lfloor C\sqrt{2}\Re\beta/s\rfloor$ and $\ell_\beta=\lfloor C\sqrt{2}\Im\beta/s\rfloor$. Then $\left|\beta-(k_\beta+\imath\ell_\beta)s/(C\sqrt{2})\right|\le s/C$. Therefore,

$$\mathscr{E}_{|\beta|<}(s,C')\cap\mathscr{E}_{\mathrm{op}}(C)\subset\bigcup_{\substack{k,\ell\in\mathbb{Z},\\|k+\imath\ell|\le CC'\sqrt{2}/s}}\mathscr{E}_q(k,\ell,2s,C).$$

Let us bound the probability of the event $\mathscr{E}_q(k,\ell,2s,C)\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)$. Recalling that $u_Q=-\Pi_Q B^\dagger W^*w$ and that w is supported by $\mathscr{I}$, we observe that u_Q and $W_{\mathscr{I}^c,\cdot}$ are independent. Writing $r=\frac{s(k+\imath\ell)}{C\sqrt{2}}q+u_Q$ and $\tilde{r}=r/\|r\|$, we have

$$\begin{aligned}
&\mathscr{E}_q(k,\ell,2s,C)\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\subset[\exists\alpha\in\mathbb{C},:\|Wr+zw+\alpha d\|\le 2s]\cap\left[\|u_Q\|\ge\mathbf{s}_{\sup}^{-1}c_0\right]\\
&\subset\left[\exists\alpha\in\mathbb{C},:\|W_{\mathscr{I}^c,\cdot}r+\alpha d_{\mathscr{I}^c}\|\le 2s\right]\cap\left[\|u_Q\|\ge\mathbf{s}_{\sup}^{-1}c_0\right]\subset\left[\|r\|\operatorname{dist}\left(W_{\mathscr{I}^c,\cdot}\tilde{r},\operatorname{span}d_{\mathscr{I}^c}\right)\le 2s\right]\cap\left[\|u_Q\|\ge\mathbf{s}_{\sup}^{-1}c_0\right]\\
&\subset\left[\operatorname{dist}\left(W_{\mathscr{I}^c,\cdot}\tilde{r},\operatorname{span}d_{\mathscr{I}^c}\right)\le 2s\mathbf{s}_{\sup}/c_0\right].
\end{aligned}$$

By Lemma 4 once again, $\mathbb{P}[\operatorname{dist}\left(W_{\mathscr{I}^c,\cdot}\tilde{r},\operatorname{span}d_{\mathscr{I}^c}\right)\le c_2]\le\exp(-c_3|\mathscr{I}^c|)$. Thus, if we choose t small enough so that $\left(2+4\frac{C}{\mathbf{s}_{\inf}}\right)\frac{\mathbf{s}_{\sup}}{c_0}t\le c_2$, we get that

$$\mathbb{P}\left[\mathscr{E}_q(k,\ell,2s,C)\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\right]\le\exp(-(1-\theta_{13})c_3n). \tag{20}$$

Putting things together, we get

$$\begin{aligned}
&\mathbb{P}\left[\left[\inf_{v\in\mathbb{C}^{n-1}}\operatorname{dist}\left(G\begin{pmatrix}v\\w\end{pmatrix},\operatorname{span}\left(\begin{pmatrix}0\\d\end{pmatrix}\right)\right)\le t\right]\cap\mathscr{E}_{\mathrm{op}}(C)\right]\\
&\le\mathbb{P}\left[\mathscr{E}_w(t)\cap\mathscr{E}_{\mathrm{op}}(C)\right]\ \text{(using (15))}\\
&\le\mathbb{P}\left[\mathscr{E}_w(t)\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\cap\mathscr{E}_{\mathrm{op}}(C)\right]+\mathbb{P}\left[\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)\right]\\
&\le\mathbb{P}\left[\mathscr{E}_{|\beta|>}(s,C')\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\cap\mathscr{E}_{\mathrm{op}}(C)\right]\\
&\quad+\mathbb{P}\left[\mathscr{E}_{|\beta|<}(s,C')\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\cap\mathscr{E}_{\mathrm{op}}(C)\right]+\exp(-c_1n)\ \text{(using (18))}\\
&\le\exp(-c_3n)+\sum_{|k+\imath\ell|\le CC'\sqrt{2}/s}\mathbb{P}\left[\mathscr{E}_q(k,\ell,2s,C)\cap\mathscr{E}_{u_Q}(\mathbf{s}_{\sup}^{-1}c_0,\mathbf{s}_{\inf}^{-1}C)\right]+\exp(-c_1n)\ \text{(using (19))}\\
&\le\exp(-c_1n)+C''\exp(-(1-\theta_{13})c_3n)\ \text{(using (20))},
\end{aligned}$$

where $C''=C''(m_4,C)>0$.

Now, let Σ_t be a t-net of $(\mathbb{S}_{\mathscr{I}}^{N-1})$. Given an element y of $\mathscr{N}_t(\mathbb{S}_{\mathscr{I}}^{N-1})\cap\mathbb{S}^{N-1}$, there exists $y'\in\mathbb{S}_{\mathscr{I}}^{N-1}$ such that $\|y-y'\|\le t$, and there exists $w\in\Sigma_t$ such that $\|w-y'\|\le t$. Thus, $\|y-w\|\le 2t$ by the triangle inequality. Assume that there exist $\alpha\in\mathbb{C}$ and $v\in\mathbb{C}^{n-1}$ such that the inequality

$$\left\|G\begin{pmatrix}v\\y\end{pmatrix}+\alpha\begin{pmatrix}0\\d\end{pmatrix}\right\|\le t$$

holds true. Then on the set $\mathscr{E}_{\mathrm{op}}(C)$, we have

$$\left\|G\begin{pmatrix}v\\w\end{pmatrix}+\alpha\begin{pmatrix}0\\d\end{pmatrix}\right\|=\left\|G\left(\begin{pmatrix}v\\w\end{pmatrix}-\begin{pmatrix}v\\y\end{pmatrix}\right)+G\begin{pmatrix}v\\y\end{pmatrix}+\alpha\begin{pmatrix}0\\d\end{pmatrix}\right\|\le 2(C+|z|)t+t.$$

Observe that $|\Sigma_t|\le(3/t)^{2|\mathscr{I}|}$. Adjusting t again in such a way that $\left(2C+2|z|+1)(2+4\frac{C}{\mathbf{s}_{\inf}}\right)\frac{\mathbf{s}_{\sup}}{c_0}t\le c_2$, we obtain that

$$\mathbb{P}\left[\left[\inf_{\substack{v\in\mathbb{C}^{n-1},\\y\in\mathscr{N}_t(\mathbb{S}_{\mathscr{I}}^{N-1})\cap\mathbb{S}^{N-1}}}\operatorname{dist}\left(G\begin{pmatrix}v\\y\end{pmatrix},\operatorname{span}\left(\begin{pmatrix}0\\d\end{pmatrix}\right)\right)\le t\right]\cap\mathscr{E}_{\mathrm{op}}(C)\right]\le(3/t)^{2\theta_{13}n}\left(\exp(-c_1n)+C''\exp(-(1-\theta_{13})c_3n)\right).$$

Finally, considering all the sets $\mathscr{I} \subset [N]$ such that $|\mathscr{I}| = \lfloor \theta_{13} N \rfloor$, and using the bound $\binom{m}{k} \le (em/k)^k$ along with the union bound, we get that

$$\mathbb{P}\Big[\Big[\inf_{\substack{v\in\mathbb{C}^{n-1},\\ w\in\mathrm{comp}(\theta_{13},t)}} \mathrm{dist}\left(G\begin{pmatrix} v \\ w \end{pmatrix}, \mathrm{span}\left(\begin{pmatrix} 0 \\ d \end{pmatrix} \right) \right) \le t \Big] \cap \mathscr{E}_{\mathrm{op}}(C) \Big] \le (e/\theta_{13})^{\theta_{13}N} (3/t)^{2\theta_{13}n} \left(\exp(-c_1 n) + C'' \exp(-(1-\theta_{13})c_3 n) \right).$$

Choosing θ_{13} small enough, we get the result with $\rho_{13} = t$ and c_{13} small enough. $\square$

Corollary 14. For each deterministic vector $d \in \mathbb{C}^N \setminus \{0\}$,

$$\mathbb{P}\left[[Rd/\|Rd\| \in \mathrm{comp}(\theta_{13}, \rho_{13})] \cap \mathscr{E}_{\mathrm{op}}(C) \right] \le \exp(-c_{13} n).$$

Proof. Write $\begin{pmatrix} u \\ y \end{pmatrix} = G^{-1} \begin{pmatrix} 0 \\ d \end{pmatrix} = \begin{pmatrix} Fd \\ Rd \end{pmatrix}$ with $y \in \mathbb{C}^N$, and let $\tilde{y} = y/\|y\|$, which can be shown to be defined w.p. 1 as in the proof of Lemma 11. Considering the event $\mathscr{E}_{\tilde{y}} = [\tilde{y} \text{ defined}, \tilde{y} \in \mathrm{comp}(\theta_{13}, \rho_{13})]$, our purpose is to show that $\mathbb{P}[\mathscr{E}_{\tilde{y}} \cap \mathscr{E}_{\mathrm{op}}(C)] \le \exp(-c_{13} n)$. Since $G \begin{pmatrix} u/\|y\| \\ \tilde{y} \end{pmatrix} = \|y\|^{-1} \begin{pmatrix} 0 \\ d \end{pmatrix}$, it holds that $\mathscr{E}_{\tilde{y}} \subset \Big[\inf_{\substack{v\in\mathbb{C}^{n-1},\\ w\in\mathrm{comp}(\theta_{13},\rho_{13})}} \mathrm{dist}\left(G\begin{pmatrix} v \\ w \end{pmatrix}, \mathrm{span}\left(\begin{pmatrix} 0 \\ d \end{pmatrix} \right) \right) \le \rho_{13} \Big]$, and the result follows from Proposition 13. $\square$

We now get back to the expression (9) of $\mathrm{dist}(h_0, H_{-0})$, handling the denominator Den given by (10).

Lemma 15. For $M = F$, P, or E, there exist positive constants c_{15} and C_{15} such that

$$\mathbb{P}\left[[\|M\| \ge C_{15} \|R\|] \cap \mathscr{E}_{\mathrm{op}}(C) \right] \le \exp(-c_{15} n).$$

Proof. We reuse the notations of the singular value decomposition (16) of B. For any matrix M with $n-1$ rows, we also write $M_Q = \Pi_Q M$ and $M_q = \Pi_q M$. We first consider $M = F$.

From Lemma 4, we know that there exist $c_0, c > 0$ such that $\mathbb{P}[\|Wq\| \le c_0] \le \exp(-cn)$. We shall show that on the event $[\|Wq\| \ge c_0] \cap \mathscr{E}_{\mathrm{op}}(C)$, there exists some $C_1 > 0$, such that $\forall u \in \mathbb{S}^{N-1}$, $\|Fu\| \le C_1(1 + \|Ru\|)$. This will establish that

$$\mathbb{P}\left[[\|F\| \ge C_1(1 + \|R\|)] \cap \mathscr{E}_{\mathrm{op}}(C) \right] \le \exp(-cn). \tag{21}$$

Recall that $G^{-1} \begin{pmatrix} 0 \\ u \end{pmatrix} = \begin{pmatrix} Fu \\ Ru \end{pmatrix} = \begin{pmatrix} v \\ w \end{pmatrix}$ say, or equivalently,

$$Bv + W^* w = 0, \tag{22a}$$
$$Wv + zw = u. \tag{22b}$$

Since $Bv_q \perp Bv_Q$, Lemma 12 and (22a) imply that $\mathbf{s}_{\inf} \|v_Q\| \le \|Bv_Q\| \le \|W^* w\|$. Thus, $\|v_Q\| \le (C/\mathbf{s}_{\inf}) \|w\|$ on $\mathscr{E}_{\mathrm{op}}(C)$. Writing $v_q = \beta q$, Equation (22b) can be rewritten as $\beta Wq = u - zw - Wv_Q$, which gives that $|\beta| \le (1 + (|z| + C^2/\mathbf{s}_{\inf})\|w\|)/c_0$ on $[\|Wq\| \ge c_0] \cap \mathscr{E}_{\mathrm{op}}(C)$. Since $\|v\|^2 = |\beta|^2 + \|v_Q\|^2$, there exists $C_1 > 0$ such that $\|v\| \le C_1(1 + \|w\|)$, and the inequality (21) follows.

Our next step is to show that there exists a constant C_2 such that $\mathscr{E}_{\mathrm{op}}(C) \subset [\|R\| \ge C_2]$. It is then easy to deduce from (21) that $\mathbb{P}\left[[\|F\| \ge C' \|R\|] \cap \mathscr{E}_{\mathrm{op}}(C) \right] \le \exp(-cn)$ with $C' = C_1(C_2^{-1} + 1)$. We shall assume that $\|R\| < C_2$ on $\mathscr{E}_{\mathrm{op}}(C)$ and obtain a contradiction if C_2 is chosen small enough. From the equation $GG^{-1} = I_{N+n-1}$, we have

$$BF + W^* R = 0, \tag{23a}$$
$$WF + zR = I. \tag{23b}$$

By Equation (23a), $\|BF\| \le CC_2$ on $\mathscr{E}_{\mathrm{op}}(C)$. Writing $BF = BF_Q + BF_q$ and observing from (16) that $\mathrm{span}(BF_Q)$ and $\mathrm{span}(BF_q)$ are orthogonal, we obtain that $\|BF_Q\| \le \|BF_Q + BF_q\| \le CC_2$. Turning to (16) again and using Lemma 12, we also have

$$\|BF_Q\|^2 = \|F^* Q \Sigma^2 Q^* F\| \ge \mathbf{s}_{\inf}^2 \|F^* Q Q^* F\| = \mathbf{s}_{\inf}^2 \|F_Q\|^2,$$

thus, $\|F_Q\| \le CC_2/\mathbf{s}_{\inf}$. Now, rewriting Equation (23b) as $WF_q - I = -zR - WF_Q$ and using the triangle inequality, we get that $\|WF_q - I\| \le |z|\|R\| + \|WF_Q\| \le (|z| + C^2/\mathbf{s}_{\inf})C_2$. Since WF_q is a rank-one matrix, the set of vectors $u \in \mathbb{S}^{N-1}$ such that $WF_q u = 0$ is not empty. For any such vectors, we have

$$(|z| + C^2/\mathbf{s}_{\inf})C_2 \ge \|WF_q - I\| \ge \|(WF_q - I)u\| = 1,$$

which raises a contradiction if we choose $C_2 < (|z| + C^2/\mathbf{s}_{\inf})^{-1}$. The lemma is proven for $M = F$.

The case $M = P$ can be shown similarly. To handle the case $M = E$, we first show an analogue of (21) where (F, R) is replaced with (E, F), and then we combine the obtained inequality with (21) to get that

$$\mathbb{P}\left[[\|E\| \ge C_1(1 + \|R\|)] \cap \mathscr{E}_{\mathrm{op}}(C) \right] \le \exp(-cn)$$

with possibly different constants. The rest of the proof is unchanged. $\square$

The following lemma is very close to [38, Proposition 8.2], with the difference that the bound on the probability in Statement (iii) is a Berry-Esséen type bound.

Lemma 16. The following hold true:

(i) There exist $c_{16}, C_{16} > 0$ such that
$$\mathbb{P}[[\|g_{01} G^{-1}\| \leq C_{16}] \cap \mathscr{E}_{\mathrm{op}}(C)] \leq \exp(-c_{16} n).$$

(ii) Let $y = (y_0, \ldots, y_{N-1})^\top \in \mathbb{C}^N$ be a random vector with independent elements such that $\mathbb{E} y_i = 0$ and $\mathbb{E}|y_i|^2 = 1/n$ for all $i \in [N]$, and let $M \in \mathbb{C}^{N \times N}$ be deterministic. Then for each $\eta > 0$,
$$\mathbb{P}\left[\|y^* M\| \leq \frac{1}{\sqrt{\eta}} \frac{\|M\|_{\mathrm{HS}}}{\sqrt{n}}\right] \geq 1 - \eta.$$

(iii) There exists $c > 0$ such that for each $\varepsilon \geq 0$,
$$\mathbb{P}\left[\left[\|x^* R\| \leq \varepsilon \frac{\|R\|_{\mathrm{HS}}}{\sqrt{n}}\right] \cap \mathscr{E}_{\mathrm{op}}(C)\right] \leq c\varepsilon + \frac{c}{\sqrt{n}}.$$

The following result is needed to prove this lemma:

Lemma 17 (Lemma 8.3 of [38])**.** Let $Z_0, \ldots, Z_{N-1}$ be arbitrary non-negative random variables, and let $p_0, \ldots, p_{N-1}$ be non-negative numbers such that $\sum p_k = 1$. Then, $\mathbb{P}[\sum p_k Z_k \leq t] \leq 2 \sum p_k \mathbb{P}[Z_k \leq 2t]$ for each $t \geq 0$.

Proof of Lemma 16. To prove the first statement, we write $\|g_{01}\| = \|g_{01} G^{-1} G\| \leq \|g_{01} G^{-1}\| \, \|G\|$. By Lemma 4, there exist two constants $c, c_{16} > 0$ such that $\|g_{01}\| \geq \|x\| \geq c$ with a probability larger than $1 - \exp(-c_{16} n)$. Moreover, $\|G\| \leq (C + |z| \vee \mathbf{s}_{\sup})$ on $\mathscr{E}_{\mathrm{op}}(C)$, hence the result.

We have $\mathbb{E}\|y^* M\|^2 = \mathbb{E} y^* M M^* y = \|M\|_{\mathrm{HS}}^2 / n$, thus, $\mathbb{P}[\|y^* M\| \geq \|M\|_{\mathrm{HS}} / \sqrt{\eta n}] \leq \eta$ by Markov's inequality. This proves Statement (ii).

Turning to the third statement, we start by writing
$$\|x^* R\|^2 = \sum_{k \in [N]} |\langle R^* x, e_{N,k} \rangle|^2 = \sum_{k \in [N]} |\langle x, R e_{N,k} \rangle|^2 = \sum_{k \in [N]} \|R e_{N,k}\|^2 |\langle x, \frac{R e_{N,k}}{\|R e_{N,k}\|} \rangle|^2. \tag{24}$$

Define $u_k = R e_{N,k} / \|R e_{N,k}\| = (u_{0,k}, \ldots, u_{N-1,k})^\top$. The idea of the proof is the following. By Corollary 14, u_k is incompressible with high probability. Moreover, x and u_k are independent. Therefore, we can use the Berry-Esséen theorem (Proposition 6) to control the behavior of the inner products $\langle x, u_k \rangle$. We now apply Lemma 17. Specifically, define for each $k \in [N]$ the set of indices
$$\mathscr{I}_k = \left\{ i \in [N] : \frac{\rho_{13}}{\sqrt{N}} \leq |u_{i,k}| \leq \frac{2}{\sqrt{\theta_{13} N}} \right\}.$$
After a small calculation using the independence of x and u_k, Lemma 5 and Proposition 6,
$$\mathbb{P}_x\left[|\langle x, u_k \rangle| \leq \varepsilon \sqrt{2/n}\right] \leq \mathscr{L}_x\Big(\sum_{i \in \mathscr{I}_k} \bar{x}_{i,0} u_{i,k}, \varepsilon \sqrt{2/n}\Big) \leq V_k \wedge 1,$$

where
$$V_k = \frac{c \varepsilon \sqrt{2}}{\rho_{13} \sqrt{|\mathscr{I}_k| N^{-1}}} + \frac{8 c m_4^{3/4}}{\theta_{13}^{3/2} \rho_{13}^3} \frac{1}{\sqrt{|\mathscr{I}_k|}},$$
and $c > 0$ is the constant that appears in the statement of Proposition 6. Observing that $\sum_{k \in [N]} \|R e_{N,k}\|^2 = \|R\|_{\mathrm{HS}}^2$ and using Lemma 17, we get that
$$\mathbb{P}_x\Big[\sum_{k \in [N]} \frac{\|R e_{N,k}\|^2}{\|R\|_{\mathrm{HS}}^2} |\langle x, u_k \rangle|^2 \leq \frac{\varepsilon^2}{n}\Big] \leq 2 \sum_{k \in [N]} \frac{\|R e_{N,k}\|^2}{\|R\|_{\mathrm{HS}}^2} (V_k \wedge 1).$$

Defining the event $\mathscr{E}_{\mathrm{incomp}} = \cap_{k \in [N]} [u_k \in \mathrm{incomp}(\theta_{13}, \rho_{13})]$, we know from Corollary 14 that $\mathbb{P}[\mathscr{E}_{\mathrm{incomp}}^{\mathrm{c}} \cap \mathscr{E}_{\mathrm{op}}(C)] \leq N \exp(-c_{13} n)$. Moreover, $|\mathscr{I}_k| \geq \theta_{13} N/2$ on $\mathscr{E}_{\mathrm{incomp}}$ for each $k \in [N]$ by Lemma 8. Thus, by changing the value of the constant c above we get that $V_k \leq c\varepsilon + c/\sqrt{n}$ on $\mathscr{E}_{\mathrm{incomp}}$ for each $k \in [N]$. Putting things together, we conclude that
$$\mathbb{P}\left[\left[\|x^* R\| \leq \varepsilon \frac{\|R\|_{\mathrm{HS}}}{\sqrt{n}}\right] \cap \mathscr{E}_{\mathrm{op}}(C)\right] = \mathbb{E}_W\left[\mathbb{P}_x\Big[\sum_{k \in [N]} \frac{\|R e_{N,k}\|^2}{\|R\|_{\mathrm{HS}}^2} |\langle x, u_k \rangle|^2 \leq \frac{\varepsilon^2}{n}\Big] \mathbb{1}_{\mathscr{E}_{\mathrm{op}}(C)}\right]$$
$$\leq 2 \mathbb{E}_W\Big[\sum_{k \in [N]} \frac{\|R e_{N,k}\|^2}{\|R\|_{\mathrm{HS}}^2} (V_k \wedge 1) \mathbb{1}_{\mathscr{E}_{\mathrm{incomp}}}\Big] + 2 \mathbb{E}_W[\mathbb{1}_{\mathscr{E}_{\mathrm{incomp}}^{\mathrm{c}}} \mathbb{1}_{\mathscr{E}_{\mathrm{op}}(C)}] \leq 2c\varepsilon + 2c/\sqrt{n} + 2N \exp(-c_{13} n),$$

which leads to the required result after changing once again the value of c. $\square$

Lemmas 15 and 16 lead to the following control on the denominator:

Lemma 18. There exist positive constants c_{18} and C_{18} such that for each $\eta > 0$,

$$\mathbb{P}\left[\left[\mathrm{Den}^2 \geq C_{18}(1+\eta^{-1})\|R\|_{\mathrm{HS}}^2\right] \cap \mathscr{E}_{\mathrm{op}}(C)\right] \leq 2\eta + \exp(-c_{18}n).$$

Proof. Starting with the expression $\mathrm{Den}^2 = 1 + \|g_{01}G^{-1}\|^2$, and using Lemma 16–(i), we get that

$$\mathbb{P}\left[\left[\mathrm{Den}^2 \geq (C_{16}^{-2}+1)\|g_{01}G^{-1}\|^2\right] \cap \mathscr{E}_{\mathrm{op}}(C)\right] \leq \exp(-c_{16}n). \tag{25}$$

Since $\|g_{01}G^{-1}\|^2 \leq 2(\|b_{01}E\|^2 + \|b_{01}F\|^2 + \|x^*P\|^2 + \|x^*R\|^2)$, the event

$$\mathscr{E} = \left[\|g_{01}G^{-1}\|^2 \geq 2(\|b_{01}E\|^2 + \|b_{01}F\|^2 + \|P\|_{\mathrm{HS}}^2/(\eta n) + \|R\|_{\mathrm{HS}}^2/(\eta n))\right]$$

is included in the event

$$\mathscr{E}' = \left[\|x^*P\|^2 \geq \|P\|_{\mathrm{HS}}^2/(\eta n)\right] \cup \left[\|x^*R\|^2 \geq \|R\|_{\mathrm{HS}}^2/(\eta n)\right].$$

Thus, $\mathbb{P}[\mathscr{E}] \leq \mathbb{P}[\mathscr{E}'] = \mathbb{P}_W \otimes \mathbb{P}_x[\mathscr{E}'] \leq 2\eta$ by Lemma 16–(ii). Furthermore, the event

$$\mathscr{E}'' = \left[\|g_{01}G^{-1}\|^2 \geq 4\mathbf{s}_{\sup}^2 C_{15}^2\|R\|^2 + 2C_{15}^2\|R\|^2/\eta + 2\|R\|_{\mathrm{HS}}^2/(\eta n))\right]$$

is included in the event

$$\mathscr{E} \cup [\|E\| \geq C_{15}\|R\|] \cup [\|F\| \geq C_{15}\|R\|] \cup [\|P\| \geq C_{15}\|R\|],$$

since $\|P\|_{\mathrm{HS}}^2/n \leq \|P\|^2$. Thus, $\mathbb{P}\left[\mathscr{E}'' \cap \mathscr{E}_{\mathrm{op}}(C)\right] \leq 2\eta + 3\exp(-c_{15}n)$ by Lemma 15. The proof is completed by combining this inequality with (25) and using the inequality $\|R\| \leq \|R\|_{\mathrm{HS}}$. □

We now turn to the numerator Num in (10).

We shall use the idea of decoupling that will allow us to replace the term $-x^*Pb_{10} - b_{01}Fx - x^*Rx$ in Num with an inner product whose concentration function is manageable by means of the Berry-Esséen theorem. This decoupling idea that dates back to [17] has been used many times in the literature.

Lemma 19 (See in [38])**.** Let Y and Z be independent random vectors, and let Z' be an independent copy of Z. Let $\mathscr{E}(Y,Z)$ be an event that depends on Y and Z. Then $\mathbb{P}[\mathscr{E}(Y,Z)]^2 \leq \mathbb{P}[\mathscr{E}(Y,Z) \cap \mathscr{E}(Y,Z')]$.

Lemma 20. Let $a \in \mathbb{C}$, $u, v \in \mathbb{C}^N$ and $M \in \mathbb{C}^{N\times N}$ be deterministic. Let $\mathscr{I} \subset [N]$. Then for each $t > 0$,

$$\mathbb{P}\left[|x^*Mx + u^*x + x^*v + a| \leq t\right]^2 \leq \mathbb{E}_{x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}} \mathscr{L}_{x_{\mathscr{I}}}\left((x_{\mathscr{I}^c} - x'_{\mathscr{I}^c})^* M_{\mathscr{I}^c,\mathscr{I}} x_{\mathscr{I}} + x_{\mathscr{I}}^* M_{\mathscr{I},\mathscr{I}^c}(x_{\mathscr{I}^c} - x'_{\mathscr{I}^c}), 2t\right),$$

where x' is an independent copy of x (here we assume that the right hand side is equal to one if $\mathscr{I} = \emptyset$ or $[N]$).

Proof. Assume without loss of generality that $\mathscr{I} = [|\mathscr{I}|]$. Write

$$x = \begin{pmatrix} x_{\mathscr{I}} \\ x_{\mathscr{I}^c} \end{pmatrix}, \quad \text{and} \quad \tilde{x} = \begin{pmatrix} x_{\mathscr{I}} \\ x'_{\mathscr{I}^c} \end{pmatrix}.$$

Using Lemma 19 with $Y = x_{\mathscr{I}}$, $Z = x_{\mathscr{I}^c}$, and $Z' = x'_{\mathscr{I}^c}$, we get

$$\begin{aligned}
\mathbb{P}\left[|x^*Mx + u^*x + x^*v + a| \leq t\right]^2 &\leq \mathbb{P}_{x_{\mathscr{I}}, x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}}\left[|x^*Mx + u^*x + x^*v + a| \leq t, |\tilde{x}^*M\tilde{x} + u^*\tilde{x} + \tilde{x}^*v + a| \leq t\right] \\
&\leq \mathbb{P}_{x_{\mathscr{I}}, x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}}\left[|x^*Mx - \tilde{x}^*M\tilde{x} + u^*(x-\tilde{x}) + (x-\tilde{x})^*v| \leq 2t\right],
\end{aligned}$$

where the second inequality is due to the triangle inequality. Developing, we get that

$$\begin{aligned}
&\mathbb{P}_{x_{\mathscr{I}}, x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}}\left[|x^*Mx - \tilde{x}^*M\tilde{x} + u^*(x-\tilde{x}) + (x-\tilde{x})^*v| \leq 2t\right] \\
&= \mathbb{P}_{x_{\mathscr{I}}, x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}}\Big[\big|(x_{\mathscr{I}^c} - x'_{\mathscr{I}^c})^* M_{\mathscr{I}^c,\mathscr{I}} x_{\mathscr{I}} + x_{\mathscr{I}}^* M_{\mathscr{I},\mathscr{I}^c}(x_{\mathscr{I}^c} - x'_{\mathscr{I}^c}) \\
&\qquad + u_{\mathscr{I}^c}^*(x_{\mathscr{I}^c} - x'_{\mathscr{I}^c}) + (x_{\mathscr{I}^c} - x'_{\mathscr{I}^c})^* v_{\mathscr{I}^c} + x_{\mathscr{I}^c}^* M_{\mathscr{I}^c,\mathscr{I}^c} x_{\mathscr{I}^c} - (x'_{\mathscr{I}^c})^* M_{\mathscr{I}^c,\mathscr{I}^c} x'_{\mathscr{I}^c}\big| \leq 2t\Big] \\
&\leq \mathbb{E}_{x_{\mathscr{I}^c}, x'_{\mathscr{I}^c}} \mathscr{L}_{x_{\mathscr{I}}}\left((x_{\mathscr{I}^c} - x'_{\mathscr{I}^c})^* M_{\mathscr{I}^c,\mathscr{I}} x_{\mathscr{I}} + x_{\mathscr{I}}^* M_{\mathscr{I},\mathscr{I}^c}(x_{\mathscr{I}^c} - x'_{\mathscr{I}^c}), 2t\right).
\end{aligned}$$

□

We now have all the ingredients to prove Proposition 10.

Proof of Proposition 10. In the remainder, we write

$$\mathscr{E}_{\mathrm{Den}}(\eta) = [\mathrm{Den} \leq C_\eta \|R\|_{\mathrm{HS}}],$$

where $C_\eta = C_{18}^{1/2}(1+\eta^{-1})^{1/2}$. Given $t > 0$, we have

$$\begin{aligned}
\mathbb{P}\left[[\mathrm{dist}(h_0, H_{-0}) \leq t] \cap \mathscr{E}_{\mathrm{op}}(C)\right]^2 &= \mathbb{P}[[\mathrm{Num} \leq t\mathrm{Den}] \cap \mathscr{E}_{\mathrm{op}}(C)]^2 \\
&\leq 2\mathbb{P}\left[[\mathrm{Num} \leq t\mathrm{Den}] \cap \mathscr{E}_{\mathrm{Den}}(\eta) \cap \mathscr{E}_{\mathrm{op}}(C)\right]^2 + 2\mathbb{P}[\mathscr{E}_{\mathrm{Den}}(\eta)^c \cap \mathscr{E}_{\mathrm{op}}(C)]^2,
\end{aligned}$$

and

$$\mathbb{P}\left[[\text{Num} \le t\text{Den}] \cap \mathscr{E}_{\text{Den}}(\eta) \cap \mathscr{E}_{\text{op}}(C)\right]^2 \le \mathbb{P}\left[[\text{Num}/\|R\|_{\text{HS}} \le tC_\eta] \cap \mathscr{E}_{\text{op}}(C)\right]^2$$
$$= \mathbb{E}_W\left[\mathbb{E}_x[\mathbb{1}_{[\text{Num}/\|R\|_{\text{HS}} \le tC_\eta]}]\mathbb{1}_{\mathscr{E}_{\text{op}}(C)}\right]^2 \le \mathbb{E}_W\left[(\mathbb{E}_x\mathbb{1}_{[\text{Num}/\|R\|_{\text{HS}} \le tC_\eta]})^2\mathbb{1}_{\mathscr{E}_{\text{op}}(C)}\right].$$

Given an arbitrary $\mathscr{I} \subset [n]$, we denote as $u \in \mathbb{C}^{|\mathscr{I}|}$, $v \in \mathbb{C}^{|\mathscr{I}^c|}$, and $w \in \mathbb{C}^{|\mathscr{I}^c|}$ three independent vectors, independent of everything else, such that $u \stackrel{\mathscr{L}}{=} x_{\mathscr{I}}$ and $v, w \stackrel{\mathscr{L}}{=} x_{\mathscr{I}^c}$. Recalling the expression of Num in (10) and using Lemma 20, we get that for each $s > 0$,

$$\mathbb{P}_x[\text{Num} \le s]^2 \le \mathbb{E}_{v,w}\mathscr{L}_u\left((v-w)^* R_{\mathscr{I}^c,\mathscr{I}} u + u^* R_{\mathscr{I},\mathscr{I}^c}(v-w), 2s\right) = \mathbb{E}_{v,w}\mathscr{L}_u\left((v-w)^* P_{\mathscr{I}^c}^* R P_{\mathscr{I}} u + u^* P_{\mathscr{I}}^* R P_{\mathscr{I}^c}(v-w), 2s\right), \tag{26}$$

where $P_{\mathscr{I}}$ the $\mathbb{C}^{|\mathscr{I}|} \to \mathbb{C}^N$ linear mapping such that if $\mathscr{I} = \{i_1 < \cdots < i_{|\mathscr{I}|}\}$, then $P_{\mathscr{I}} u = (0, \ldots, 0, u_1, 0, \ldots, 0, u_{|\mathscr{I}|}, 0, \ldots)$, where u_j is at the position i_j.

Let $\xi = (\xi_0, \ldots, \xi_{N-1})$ be a vector of N i.i.d. Bernoulli random variables valued in $\{0,1\}$ such that $\mathbb{P}[\xi_0 = 1] = p$, where the probability $p$will be fixed below. This vector is assumed to be independent of everything else. Since (26) is true for each $\mathscr{I} \subset [N]$, we can randomize $\mathscr{I}$ by setting $\mathscr{I} = \{i \in [n] : \xi_i = 1\}$. Setting $s = \|R\|_{\text{HS}} C_\eta t$, we obtain

$$\begin{aligned}(\mathbb{E}_x\mathbb{1}_{[\text{Num}/\|R\|_{\text{HS}} \le tC_\eta]})^2 &\le \mathbb{E}_\xi \mathbb{E}_{v,w}\mathscr{L}_u\left(\frac{(v-w)^* P_{\mathscr{I}^c}^* R}{\|R\|_{\text{HS}}} P_{\mathscr{I}} u + u^* P_{\mathscr{I}}^* \frac{R P_{\mathscr{I}^c}(v-w)}{\|R\|_{\text{HS}}}, 2C_\eta t\right) \\ &= \mathbb{E}_{\xi,x,x'}\mathscr{L}_u\left(\frac{(x-x')^* \Pi_{\mathscr{I}^c} R}{\|R\|_{\text{HS}}} P_{\mathscr{I}} u + u^* P_{\mathscr{I}}^* \frac{R \Pi_{\mathscr{I}^c}(x-x')}{\|R\|_{\text{HS}}}, 2C_\eta t\right).\end{aligned} \tag{27}$$

where x' is a vector that has the same law as x and is independent of all other random variables.

Write

$$y = \frac{R\Pi_{\mathscr{I}^c}(x-x')}{\|R\Pi_{\mathscr{I}^c}(x-x')\|} = \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix}, \quad \tilde{y}^* = \frac{(x-x')^*\Pi_{\mathscr{I}^c} R}{\|(x-x')^*\Pi_{\mathscr{I}^c} R\|} = (\bar{\tilde{y}}_0, \ldots, \bar{\tilde{y}}_{N-1}),$$

and let

$$\alpha = \frac{\sqrt{n}\|R\Pi_{\mathscr{I}^c}(x-x')\|}{\sqrt{2(1-p)}\|R\|_{\text{HS}}}, \quad \tilde{\alpha} = \frac{\sqrt{n}\|(x-x')^*\Pi_{\mathscr{I}^c} R\|}{\sqrt{2(1-p)}\|R\|_{\text{HS}}}.$$

For $i \in \mathscr{I}$, let

$$Z_i = \tilde{\alpha}\bar{\tilde{y}}_i [P_{\mathscr{I}} u]_i + \alpha \overline{[P_{\mathscr{I}} u]}_i y_i.$$

Then the concentration function $\mathscr{L}_u$ at the right hand side of (27) can be rewritten as

$$\mathscr{L}_u\left(\tilde{\alpha}\tilde{y}^* P_{\mathscr{I}} u + \alpha u^* P_{\mathscr{I}}^* y, \sqrt{2/(1-p)} C_\eta t\sqrt{n}\right) = \mathscr{L}_u\Big(\sum_{i \in \mathscr{I}} Z_i, \sqrt{2/(1-p)} C_\eta t\sqrt{n}\Big).$$

We wish to control this by using the Berry-Esséen theorem (Proposition 6). Recalling Proposition 13, define the set

$$\mathscr{J} = \left\{ i \in [N] : \frac{\rho_{13}}{\sqrt{N}} \le |y_i| \le \frac{2}{\sqrt{\theta_{13} N}}, \ |\tilde{y}_i| \le \frac{2}{\sqrt{\theta_{13} N}} \right\}.$$

By the restriction lemma 5, we have

$$\mathscr{L}_u\Big(\sum_{i \in \mathscr{I}} Z_i, \sqrt{2/(1-p)} C_\eta t\sqrt{n}\Big) \le \mathscr{L}_u\Big(\sum_{i \in \mathscr{I} \cap \mathscr{J}} Z_i, \sqrt{2/(1-p)} C_\eta t\sqrt{n}\Big).$$

Informally, we expect that $|\mathscr{I} \cap \mathscr{J}| = \mathscr{O}(n)$ with high probability, the $\mathbb{E}_u|Z_i|^2$ to be lower bounded with high probability, and the $\mathbb{E}_u|Z_i|^3$ to be upper bounded with high probability for $i \in \mathscr{I} \cap \mathscr{J}$, in order to benefit from the effect of the Berry-Esséen theorem in a manner similar to Inequality (12).

More rigorously, for each $i \in \mathscr{I}$, we have

$$\mathbb{E}_u|Z_i|^2 = \mathbb{E}_{x_{00}}|\tilde{\alpha}\bar{\tilde{y}}_i x_{00} + \alpha y_i \overline{x_{00}}|^2 = \mathbb{E}|x_{00}|^2\left(\tilde{\alpha}^2|\tilde{y}_i|^2 + \alpha^2|y_i|^2\right) + 2\alpha\tilde{\alpha}\Re\left(\mathbb{E}x_{00}^2 \bar{\tilde{y}}_i \overline{y_i}\right) \ge n^{-1}\vartheta(\tilde{\alpha}^2|\tilde{y}_i|^2 + \alpha^2|y_i|^2)$$

for all large enough n, where $\vartheta = 1 - \limsup_n |n\mathbb{E}x_{00}^2|$ is positive by Assumption 3. Focusing on the set $\mathscr{I} \cap \mathscr{J}$, we get that

$$\sum_{i \in \mathscr{I} \cap \mathscr{J}} \mathbb{E}_u|Z_i|^2 \ge n^{-1}\vartheta \sum_{i \in \mathscr{I} \cap \mathscr{J}} \alpha^2|y_i|^2 \ge \alpha^2\vartheta\rho_{13}^2 \frac{|\mathscr{I} \cap \mathscr{J}|}{nN}. \tag{28}$$

Moreover,

$$\sum_{i \in \mathscr{I} \cap \mathscr{J}} \mathbb{E}_u|Z_i|^3 \le 32\mathbb{E}|x_{00}|^3(\alpha^3 + \tilde{\alpha}^3)\frac{|\mathscr{I} \cap \mathscr{J}|}{\theta_{13}^{3/2} N^{3/2}}.$$

Then, by the Berry-Esséen theorem,

$$\mathscr{L}_u\left(\sum_{i\in_c I} Z_i, \sqrt{2/(1-p)}C_\eta t\sqrt{n}\right) \le \mathscr{L}_u\Big(\sum_{i\in\mathscr{I}\cap\mathscr{J}} Z_i, \sqrt{2/(1-p)}C_\eta t\sqrt{n}\Big)$$
$$\le \Big(\sqrt{2/(1-p)}c\frac{\sqrt{N}}{\alpha\rho_{13}\sqrt{\vartheta|\mathscr{I}\cap\mathscr{J}|}}C_\eta nt + \frac{32cm_4^{3/4}(\alpha^3+\tilde{\alpha}^3)}{\rho_{13}^3\vartheta^{3/2}\theta_{13}^{3/2}\alpha^3}\frac{1}{\sqrt{|\mathscr{I}\cap\mathscr{J}|}}\Big)\wedge 1 \triangleq V\wedge 1$$

(here, we assume that $\mathscr{L}_u(\sum_{\mathscr{I}\cap\mathscr{J}}\cdots) = V = 1$ if $\mathscr{I}\cap\mathscr{J} = \emptyset$). The constant $c>0$ in the term after the second inequality is the one that appears in the statement of Proposition 6. In the remainder of the proof, the value of this constant may change without mention.

At this stage of the calculation, we have

$$\mathbb{P}[[\mathrm{dist}(h_0,H_{-0})\le t]\cap\mathscr{E}_{\mathrm{op}}(C)]^2 \le 2\mathbb{E}_{W,\xi,x,x'}[(V\wedge 1)\mathbb{1}_{\mathscr{E}_{\mathrm{op}}(C)}] + 2\mathbb{P}[\mathscr{E}_{\mathrm{Den}}(\eta)^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)]^2. \tag{29}$$

Now, take $p = 1-\theta_{13}/8$, and consider the event

$$\mathscr{E}_\xi = [|\mathscr{I}| > N(1-\theta_{13}/4)] = \left[\sum \xi_i > N(1-\theta_{13}/4)\right].$$

Since $\mathscr{E}_\xi = [|\mathscr{I}| > N(p-\theta_{13}/8)]$, we get by Hoeffding's concentration inequality [24] that

$$\mathbb{P}[\mathscr{E}_\xi^{\mathrm{c}}] \le \exp(-N\theta_{13}^2/32).$$

Consider also the event

$$\mathscr{E}_{\mathrm{incomp}} = [y\in\mathrm{incomp}(\theta_{13},\rho_{13})].$$

By Corollary 14, there exists a constant $c>0$ such that

$$\mathbb{P}[\mathscr{E}_{\mathrm{incomp}}^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)] \le \exp(-cn).$$

On $\mathscr{E}_{\mathrm{incomp}}$, we have that $|\mathscr{J}|\ge\theta_{13}N/2$ by Lemma 8. Therefore, on $\mathscr{E}_\xi\cap\mathscr{E}_{\mathrm{incomp}}$, it holds that

$$|\mathscr{I}\cap\mathscr{J}| = N - |\mathscr{I}^{\mathrm{c}}\cup\mathscr{J}^{\mathrm{c}}| \ge N - |\mathscr{I}^{\mathrm{c}}| - |\mathscr{J}^{\mathrm{c}}| \ge N\theta_{13}/4.$$

It remains to control the terms α and $\tilde{\alpha}$ in the expression of V. Given a small $\beta>0$, consider the event

$$\mathscr{E}_\alpha(\beta) = \left[\beta\|R\|_{\mathrm{HS}}/\sqrt{n} \le \frac{\|R\Pi_{\mathscr{J}^{\mathrm{c}}}(x-x')\|}{\sqrt{2(1-p)}} \le \beta^{-1/2}\|R\|_{\mathrm{HS}}/\sqrt{n}\right] \cap \left[\frac{\|(x-x')^*\Pi_{\mathscr{J}^{\mathrm{c}}}R\|}{\sqrt{2(1-p)}} \le \beta^{-1/2}\|R\|_{\mathrm{HS}}/\sqrt{n}\right].$$

Note that $\alpha\in[\beta,\beta^{-1/2}]$ and $\tilde{\alpha}\le\beta^{-1/2}$, thus $(\alpha^3+\tilde{\alpha}^3)/\alpha^3\le 2\beta^{-9/2}$ on $\mathscr{E}_\alpha(\beta)$. Applying Lemma 16 after setting the vector y in its statement to $((1-\xi_0)(x_{00}-x_0'),\ldots,(1-\xi_{N-1})(x_{N-1,0}-x_{N-1}'))^{\mathsf{T}}/\sqrt{2(1-p)}$, we get that there exists a constant $c>0$ for which

$$\mathbb{P}[\mathscr{E}_\alpha(\beta)^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)] \le c\beta + \frac{c}{\sqrt{n}}.$$

Turning back to (29), we can now conclude by writing

$$\begin{aligned}
&\mathbb{P}[[\mathrm{dist}(h_0,H_{-0})\le t]\cap\mathscr{E}_{\mathrm{op}}(C)]^2\\
&\le 2\mathbb{E}_{W,\xi,x,x'}[V\mathbb{1}_{\mathscr{E}_\xi}\mathbb{1}_{\mathscr{E}_{\mathrm{incomp}}}\mathbb{1}_{\mathscr{E}_\alpha(\beta)}\mathbb{1}_{\mathscr{E}_{\mathrm{op}}(C)}] + 2\mathbb{P}[\mathscr{E}_\xi^{\mathrm{c}}] + 2\mathbb{P}[\mathscr{E}_{\mathrm{incomp}}^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)] + 2\mathbb{P}[\mathscr{E}_\alpha(\beta)^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)] + 2\mathbb{P}[\mathscr{E}_{\mathrm{Den}}(\eta)^{\mathrm{c}}\cap\mathscr{E}_{\mathrm{op}}(C)]^2\\
&\le c\left(\frac{n}{\beta\sqrt{\eta}}t + \frac{\beta^{-9/2}}{\sqrt{n}} + \beta + \eta + \frac{1}{\sqrt{n}}\right) + \exp(-c'n).
\end{aligned}$$

If we take $\eta\propto n^{-1/2}$ and $\beta\propto n^{-1/11}$ (without further optimization of these exponents), then we get that

$$\mathbb{P}[[\mathrm{dist}(h_0,H_{-0})\le t]\cap\mathscr{E}_{\mathrm{op}}(C)]^2 \le c(n^{59/44}t + n^{-1/11}) + \exp(-c'n),$$

which proves Proposition 10.

Theorem 1: end of proof. First note that for any $k\in[n]$, Proposition 10 continues to hold when $\mathrm{dist}(h_0,H_{-0})$ is replaced by $\mathrm{dist}(h_k,H_{-k})$, by the same proof. When $n\le k<N+n$, too, the proof continues to be valid once the roles of A and z are interchanged. Indeed, one can check that the argument is simpler and hence is omitted. Applying Lemma 9, we obtain that

$$\mathbb{P}\left[\left[\inf_{u\in\mathrm{incomp}(\theta_7,\rho_7)}\|Hu\|\le t\right]\cap\mathscr{E}_{\mathrm{op}}(C)\right] \le c(n^{81/88}t^{1/2} + n^{-1/22}) + \exp(-c'n).$$

Using Proposition 7 along with the characterization (5) of the smallest singular value, we obtain Theorem 1 with $\alpha = 81/88$ and $\beta = 1/22$.

Remark 1. The proof of Proposition 10 shows that the origin of the slow decreasing term $n^{-\beta} = n^{-1/22}$ at the right hand side of the last inequality is the $\mathscr{O}(1/\sqrt{n})$ decay that is optimal while using the Berry-Esséen theorem, as shown by Inequality (12). To obtain a better decay rate of the concentration functions, one can use the so-called Littlewood-Offord theory [12] instead. This was the approach of [33, 35, 36, 38] to solve small singular value problems.

Remark 2. Assumption 3 was needed in the proof of Proposition 10 to ensure that the variance at the left hand side of (28) is bounded away from zero.

4 Proof of Theorem 2

A well established technique for studying the spectral behavior of large random non-Hermitian matrices is Girko's so-called hermitization technique [14]. This is intimately tied to the logarithmic potential of their spectral measures. In all the remainder, we shall write $Y = XJX^*$, and recall that $\{\lambda_0, \ldots, \lambda_{N-1}\}$ are the eigenvalues of this matrix. The logarithmic potential $U_{\mu_n} : \mathbb{C} \to (-\infty, \infty]$ of the spectral measure μ_n can then be written as

$$\begin{aligned} U_{\mu_n}(z) &= -\int_{\mathbb{C}} \log|\lambda - z|\, \mu_n(d\lambda) = -\frac{1}{N}\sum_{i=0}^{N-1} \log|\lambda_i - z| = -\frac{1}{N}\log|\det(Y-z)| \\ &= -\frac{1}{2N}\log\det(Y-z)(Y-z)^* = -\int \log\lambda\ \nu_{n,z}(d\lambda), \end{aligned}$$

where the probability measure $\nu_{n,z}$ is the singular value distribution of $Y - z$, given as

$$\nu_{n,z} = \frac{1}{N}\sum_{i=0}^{N-1} \delta_{s_i(Y-z)}.$$

The above observation is at the heart of the hermitization technique. It transforms the eigenvalue problem into a problem of singular values. To study the asymptotic behavior of μ_n, we need to study the asymptotic behavior of $U_{\mu_n}(z)$ for Lebesgue almost all $z \in \mathbb{C}$. To that end, we need to perform the two following steps, see, *e.g.*, [7, Lemma 4.3]:

Step 1: Show that for almost all $z \in \mathbb{C}$, $\nu_{n,z} \Rightarrow \nu_z$ (a deterministic probability measure) in probability.

Step 2: Show that the function log is uniformly integrable with respect to the measure $\nu_{n,z}$ for almost all $z \in \mathbb{C}$ in probability. That is,

$$\forall \varepsilon > 0, \quad \lim_{T\to\infty} \limsup_{n\geq 1} \mathbb{P}\left[\int_0^\infty |\log\lambda|\, \mathbb{1}_{|\log\lambda|\geq T}\ \nu_{n,z} d(\lambda) > \varepsilon\right] = 0. \tag{30}$$

By achieving these two steps, we conclude that there exists a probability measure μ such that $\mu_n \Rightarrow \mu$ in probability, and such that $U_\mu(z) = -\int \log|\lambda|\ \check{\nu}_z(d\lambda)$ $\mathbb{C}$-almost everywhere.

Step 3: Identify μ. This can be done by relying on the generic relation $\mu = -(2\pi)^{-1}\Delta U_\mu$, where Δ is the Laplace operator defined on $\mathscr{D}'(\mathbb{C})$, the space of Schwarz distributions on $\mathbb{C}$.

Step 1: Weak convergence of $\nu_{n,z}$ Going a bit further than what is required for this step, we shall show that for each $z \in \mathbb{C}$, there exists a probability measure ν_z such that $\nu_{n,z} \Rightarrow \nu_z$ almost surely. As is usual in random matrix theory, this convergence will be established through the convergence of the associated Stieltjes transforms. For this, it will be convenient to consider the Hermitian matrix

$$\Sigma(z) = \begin{pmatrix} 0 & Y - z \\ Y^* - \bar{z} & 0 \end{pmatrix}$$

whose spectral measure

$$\check{\nu}_{n,z} = \frac{1}{2N}\sum_{i=0}^{N-1}\left(\delta_{s_i(Y-z)} + \delta_{-s_i(Y-z)}\right)$$

is the symmetrized version of $\nu_{n,z}$ ($\check{\nu}_{n,z}$ is symmetric in the sense that $\check{\nu}_{n,z}(S) = \check{\nu}_{n,z}(-S)$ for each Borel set $S \subset \mathbb{R}$). It is enough to show that $\check{\nu}_{n,z}$ converges weakly a.s. to a probability measure $\check{\nu}_z$. Given $\eta \in \mathbb{C}_+ = \{w \in \mathbb{C}, \Im w > 0\}$, let us write

$$Q(z,\eta) = (\Sigma(z) - \eta)^{-1} = \begin{pmatrix} \eta[(Y-z)(Y-z)^* - \eta^2]^{-1} & (Y-z)[(Y-z)^*(Y-z) - \eta^2]^{-1} \\ [(Y-z)^*(Y-z) - \eta^2]^{-1}(Y-z)^* & \eta[(Y-z)^*(Y-z) - \eta^2]^{-1} \end{pmatrix} = \begin{pmatrix} Q_{00}(z,\eta) & Q_{01}(z,\eta) \\ Q_{10}(z,\eta) & Q_{11}(z,\eta) \end{pmatrix}, \tag{31}$$

which is the resolvent of $\Sigma(z)$ in the complex variable η. The a.s. convergence $\nu_{n,z} \Rightarrow \nu_z$ is a consequence of the following theorem, whose proof is provided in Section 5. Note that the Stieltjes transform of a symmetric probability measure is purely imaginary with a positive imaginary part on the positive imaginary axis.

Theorem 21. Let Assumption 1 hold true. Then

$$\frac{1}{2N}\operatorname{tr} Q(z,\eta) \xrightarrow[n\to\infty]{\text{a.s.}} \gamma^{-1}p(z,\eta), \quad \frac{1}{N}\operatorname{tr} Q_{01}(z,\eta) \xrightarrow[n\to\infty]{\text{a.s.}} \gamma^{-1}d(z,\eta), \quad \eta \in \mathbb{C}_+, \tag{32}$$

where for each $z \in \mathbb{C}$, $(p(z,\cdot), d(z,\cdot))$ is a pair of holomorphic functions on $\mathbb{C}_+$ such that $\gamma^{-1}p(z,\cdot)$ is the Stieltjes transform of a symmetric probability measure, $|d(z,\eta)| \le \gamma/\Im\eta$, and writing $p(z,\imath t) = \imath h(z,t)$ for $t>0$, the pair $(h(z,t), d(z,\imath t)) \in (0,\infty)\times\mathbb{C}$ uniquely solves the equations

$$-th(z,t) + \bar{z}d(z,\imath t) = u(h(z,t), d(z,\imath t), t) - \gamma, \tag{33a}$$
$$zh(z,t) + td(z,\imath t) = v(h(z,t), d(z,\imath t), t), \tag{33b}$$

where

$$\begin{aligned} u(h,d) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{h^2 + |d|^2 + d\exp(\imath\theta)}{h^2 + |1 + d\exp(\imath\theta)|^2} d\theta, \\ v(h,d) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{h\exp(-\imath\theta)}{h^2 + |1 + d\exp(\imath\theta)|^2} d\theta. \end{aligned} \tag{34}$$

It is well known that the convergence of $(2N)^{-1}\operatorname{tr} Q(z,\eta)$ in (32) implies that $\check{\nu}_{n,z} \Rightarrow \check{\nu}_z$ (symmetric) a.s., with Stieltjes transform $\gamma^{-1}p(z,\cdot)$. The system of equations (33) which provides the values of $p(z,\cdot)$ on the positive imaginary axis completely determines the measure $\check{\nu}_z$. The function $d(z,\imath t)$ will be used below to identify the limit measure μ.

Step 2: uniform integrability Noting that log is unbounded near both 0 and ∞, we start with the uniform integrability near zero.

Proposition 22. Let Assumption 1 hold true, and assume that $z \neq 0$. Then, there exist two constants $\alpha, C > 0$ such that

$$\frac{-\imath}{N}\mathbb{E}\operatorname{tr} Q(z,\imath t) \le C(1 + t^{-\alpha}n^{-1/2}).$$

The proof is sketched at the end of Section 5. We just point out that by using this proposition and by making some elementary Stieltjes transform calculations (see [20, Proposition 14]), one can show that there exist constants $K, \rho > 0$ such that $\mathbb{E}\check{\sigma}_n([-x,x]) \le K(x \vee n^{-\rho})$. This is a so-called local Wegner estimate on the number of intermediate singular values [7].

The smallest singular value of $Y - z$ is controlled by Theorem 1 with $A = J$. Thanks to the boundedness of the fourth moment specified by Assumption 1, we know from [40] that $\|X\| \xrightarrow[n\to\infty]{\text{a.s.}} 1 + \sqrt{\gamma}$. Thus, the probability of the event $[\|X\| \le C]$ in the statement of Theorem 1 converges to 1 by setting $C = 2 + \sqrt{\gamma}$.

Thanks to these controls, we get that for all $z \in \mathbb{C}\setminus\{0\}$,

$$\forall \varepsilon > 0, \quad \lim_{\delta\to 0}\limsup_{n\ge 1} \mathbb{P}\Big[\Big|\int_0^\delta \log\lambda \ \nu_{n,z}(d\lambda)\Big| > \varepsilon\Big] = 0, \tag{35}$$

see [20, Proposition 14] for a proof that results of the type of Proposition 22 and Theorem 1 lead together to (35).

Remark 3. The uniform integrability only in probability and the convergence only in probability in Theorem 2 are due to the slow rate $\beta = 1/22$ in the statement of Theorem 1.

To complete Step 2, it only remains to establish the uniform integrability of log near infinity. But this result follows immediately from the identity $\int_{\lambda \ge T} \log\lambda \ \nu_{n,z}(d\lambda) = 0$ a.s. for all large n, which is valid for $T > (1+\sqrt{\gamma})^2 + |z| + 1$.

Step 3: Identification of μ We use an idea that dates back to [13] and that has been frequently used in the literature devoted to large non-Hermitian matrices. Define on $\mathbb{C}\times(0,\infty)$ the regularized versions of $U_{\mu_n}(z)$ and $U_\mu(z)$ respectively:

$$\begin{aligned} \mathscr{U}_n(z,t) &= -\frac{1}{2N}\log\det((Y-z)^*(Y-z) + t^2), \quad \text{and} \\ \mathscr{U}(z,t) &= -\frac{1}{2}\int \log(\lambda^2 + t^2) \ \check{\nu}_z(d\lambda). \end{aligned}$$

In parallel, let us get back to the resolvent $Q(z,\eta)$ defined in (31). By Jacobi's formula,

$$\partial_{\bar{z}}\mathscr{U}_n(z,t) = \frac{1}{2N}\operatorname{tr}(Y-z)((Y-z)^*(Y-z) + t^2)^{-1} = \frac{1}{2N}\operatorname{tr} Q_{01}(z,\imath t).$$

Letting $n \to \infty$ we know from Theorem 21 that $\partial_{\bar{z}} \mathscr{U}_n(z,t) \to (2\gamma)^{-1} d(z,\imath t)$ a.s. At the same time, $\mathscr{U}_n(z,t) \to \mathscr{U}(z,t)$ a.s. since $\nu_{n,z} \Rightarrow \nu_z$. We can therefore assert that $\partial_{\bar{z}} \mathscr{U}(z,t) = (2\gamma)^{-1} d(z,\imath t)$ in $\mathscr{D}'(\mathbb{C})$, and then extract the properties of μ from the equation

$$\mu = -\frac{1}{2\pi}\Delta U_\mu = -\frac{2}{\pi}\lim_{t\to 0}\partial_z\partial_{\bar{z}}\mathscr{U}(z,t) = -\frac{1}{\gamma\pi}\lim_{t\to 0}\partial_z d(z,\imath t) \quad \text{in } \mathscr{D}'(\mathbb{C}).$$

This line of thought leads to the following proposition, which proof can be done along the lines of [11, Section 7], and is omitted.

Proposition 23. As $t \to 0$, the function $(2\gamma)^{-1} d(\cdot, \imath t)$ converges to $\partial_{\bar{z}} U_\mu(\cdot)$ in $\mathscr{D}'(\mathbb{C})$.

The following lemma specifies the properties of the function g defined in (2), that we shall need. Its proof is straightforward.

Lemma 24. Consider the function g on the interval $[0 \vee (\gamma-1), \gamma]$. It is analytical and increasing on $(0 \vee (\gamma-1), \gamma)$. Moreover, $g(0 \vee (\gamma-1)) = 0 \vee (\gamma-1)^3/\gamma$, and $g(\gamma) = \gamma(\gamma+1)$.

By this lemma, g has an inverse g^{-1} on $[0 \vee (\gamma-1)^3/\gamma, \gamma(\gamma+1)]$ that takes this interval to $[0 \vee (\gamma-1), \gamma]$. On $(0 \vee (\gamma-1)^3/\gamma, \gamma(\gamma+1))$, the function g^{-1} is analytical and increasing.

By showing that $d(z,\imath t)$ converges as $t \to 0$ point-wise for each $z \neq 0$ and by identifying the limit function $b(z)$, we get the following proposition whose proof is given in Section 6.

Proposition 25. Let $b(z)$ be the function defined on $\mathbb{C} \setminus \{0\}$ as follows: If $\gamma \leq 1$, then

$$b(z) = \begin{cases} -g^{-1}(|z|^2)/\bar{z} & \text{if } 0 < |z| \leq \sqrt{\gamma(\gamma+1)}, \\ -\gamma/\bar{z} & \text{if } |z| \geq \sqrt{\gamma(\gamma+1)}. \end{cases}$$

If $\gamma > 1$, then

$$b(z) = \begin{cases} -(\gamma-1)/\bar{z} & \text{if } 0 < |z| \leq (\gamma-1)^{3/2}/\sqrt{\gamma}, \\ -g^{-1}(|z|^2)/\bar{z} & \text{if } (\gamma-1)^{3/2}/\sqrt{\gamma} \leq |z| \leq \sqrt{\gamma(\gamma+1)}, \\ -\gamma/\bar{z} & \text{if } |z| \geq \sqrt{\gamma(\gamma+1)}. \end{cases}$$

Then $\partial_{\bar{z}} U_\mu(z) = (2\gamma)^{-1} b(z)$ in $\mathscr{D}'(\mathbb{C})$.

By Lemma 24, $b(z) = b(u+\imath v)$ is continuously differentiable as a function of u, v on the open set $S = \{z \in \mathbb{C} : z \neq 0, |z|^2 \neq (\gamma-1)^3/\gamma, |z|^2 \neq \gamma(\gamma+1)\}$. Therefore, $\Delta U_\mu = 4\partial_z\partial_{\bar{z}} U_\mu$ coincides with $2\gamma^{-1}\partial_z b$ in $\mathscr{D}'(S)$, where $\partial_z b$ is the pointwise derivative of b w.r.t. z. Specifically, for each test function $\varphi \in C_c^\infty(S)$, the set of compactly supported real smooth functions on $\mathbb{C}$, we have

$$\int_{\mathbb{C}} \varphi \, d\mu = -\frac{1}{2\pi}\int_{\mathbb{C}} \varphi(z)\Delta U_\mu(z)\, dz = -\frac{1}{\gamma\pi}\int_{\mathbb{C}} \varphi(z)\partial_z b(z)\, dz = \int_{\mathbb{C}} \varphi(z) f(z)\, dz,$$

where, by Proposition 25, the density $f(z)$ of μ on S is given by

$$f(z) = \begin{cases} \dfrac{1}{\gamma\pi}\partial_z \dfrac{g^{-1}(|z|^2)}{\bar{z}} = \dfrac{1}{\gamma\pi}\partial_{|z|^2} g^{-1}(|z|^2) & \text{if } 0 \vee ((\gamma-1)^3/\gamma) < |z|^2 < \gamma(\gamma+1), \\ 0 & \text{elsewhere} \end{cases} \tag{36}$$

Hence the density f depends on z through $|z|$ only, and thus μ is rotationally invariant on S.

Now we consider μ on the boundary ∂S. We deal separately with the cases $\gamma \leq 1$ and $\gamma > 1$. First suppose $\gamma \leq 1$. Let $0 < s < r < \sqrt{\gamma(\gamma+1)}$. Changing to polar co-ordinates, we get

$$\mu(\{z : |z| \in [s,r]\}) = \frac{1}{\gamma\pi}\int_{\{z:|z|\in[s,r]\}} \partial_{|z|^2} g^{-1}(|z|^2)\, dz = \gamma^{-1} g^{-1}(r^2) - \gamma^{-1} g^{-1}(s^2).$$

But since $\gamma^{-1} g^{-1}(0) = 0$ and $\gamma^{-1} g^{-1}(\gamma(\gamma+1)) = 1$, we get that $\mu(\{0\}) = \mu(\{z : |z| = \sqrt{\gamma(\gamma+1)}\}) = 0$, establishing the formula in Theorem 2 for $\gamma \leq 1$.

Now suppose $\gamma > 1$. Put $a = (\gamma-1)^{3/2}/\sqrt{\gamma}$. If we set $0 < s < r < a$, we obtain from (36) that $\mu(\{z : |z| \in [s,r]\}) = 0$. If $a < s < r < \sqrt{\gamma(\gamma+1)}$, then $\mu(\{z : |z| \in [s,r]\}) = \gamma^{-1} g^{-1}(r^2) - \gamma^{-1} g^{-1}(s^2)$ by the same derivation as for $\gamma \leq 1$.

Now we claim that $\mu(\{z : |z| = a\}) = 0$. To show this, let $\phi : [-1,1] \to [0,1]$ be a smooth function such that $\phi(0) = 1$ and $\phi(-1) = \phi(1) = 0$. Given $\varepsilon > 0$, define the $\mathbb{C} \to [0,1]$ function $\psi_\varepsilon(z) = \phi((|z|-a)/\varepsilon)$, which is supported on the ring $\{z : a-\varepsilon \le |z| \le a+\varepsilon\}$. It is then enough to show that $\int \psi_\varepsilon \, d\mu \to 0$ as $\varepsilon \to 0$. Indeed, by an integration by parts, we get that

$$\int \psi_\varepsilon \, d\mu = -\frac{1}{\gamma\pi}\int \psi_\varepsilon(z)\,\partial_{\bar z} b(z)\, dz = \frac{1}{\gamma\pi}\int \partial_{\bar z}\psi_\varepsilon(z)\, b(z)\, dz = \frac{1}{2\varepsilon\gamma\pi}\int \frac{1}{|z|}\phi'\left(\frac{|z|-a}{\varepsilon}\right) c(|z|)\, dz,$$

where the function $c(\rho) = \bar z b(z)$ for $\rho = |z|$ is a real bounded function near $\rho = a$ that satisfies $c(a) = 1-\gamma$ by Proposition 25. Making a Cartesian to polar variable change and using the dominated convergence theorem, we get that $\int \psi_\varepsilon \, d\mu \to_{\varepsilon\to 0} \frac{1-\gamma}{\gamma}(\phi(1)-\phi(-1)) = 0$. Since $\gamma^{-1}g^{-1}(a^2) = 1-\gamma^{-1}$, we can infer now that $\mu(\{z : s \le |z| \le r\}) = \gamma^{-1}g^{-1}(r^2) - (1-\gamma^{-1})$ for each $s \in (0,a)$ and each $r \in [a, \sqrt{\gamma(\gamma+1)})$. Letting $s \downarrow 0$ and $r \uparrow \sqrt{\gamma(\gamma+1)}$, and recalling that $g^{-1}(\gamma(\gamma+1)) = \gamma$, we get that $\mu(\{z : |z| < \sqrt{\gamma(\gamma+1)}\}) = 1-(1-\gamma^{-1}) + \mu(\{0\})$. Similarly to $\mu(\{z : |z| = a\}) = 0$, we can show that $\mu(\{z : |z| = \sqrt{\gamma(\gamma+1)}\}) = 0$. We therefore get that $\mu(\{0\}) = 1-\gamma^{-1}$, and hence the formula in Theorem 2 is verified also for $\gamma > 1$.

5 Limit singular value distribution

Given $(z,\eta) \in \mathbb{C}\times\mathbb{C}_+$, $\alpha \in \mathbb{R}$, and a sequence $(a_n(z,\eta))_n$ of complex numbers, the notation $a_n = \mathcal{O}_\eta(n^\alpha)$ (or $a_n = \mathcal{O}_t(n^\alpha)$ when $\eta = \imath t$) will refer in this section to the existence of a constant $C > 0$ and two non-negative integers k and ℓ such that $|a_n(z,\eta)| \le \frac{C|\eta|^k}{(\Im\eta)^\ell} n^\alpha$. The constants C, k, and ℓ may depend on z but not on η or n. If $a_n(z,\eta)$ is a matrix, then the notations $a_n = \mathcal{O}_\eta(n^\alpha)$ and $a_n = \mathcal{O}_t(n^\alpha)$, are to be understood in a uniform entry-wise sense.

Proof of Theorem 21. We first state that the $n^{-1}\operatorname{tr} Q_{ij}(z,\eta)$, $i,j = 0,1$ concentrate around their means, and that the elements of X can be replaced by complex Gaussian random variables. The proof of the following proposition is standard and is omitted.

Proposition 26. Under Assumption 1, for each $(z,\eta) \in \mathbb{C}\times\mathbb{C}_+$, if $n \to \infty$,

$$\frac{1}{n}\begin{pmatrix} \operatorname{tr} Q_{00}(z,\eta) & \operatorname{tr} Q_{01}(z,\eta) \\ \operatorname{tr} Q_{10}(z,\eta) & \operatorname{tr} Q_{11}(z,\eta)\end{pmatrix} - \frac{1}{n}\begin{pmatrix} \operatorname{tr} \mathbb{E}Q_{00}(z,\eta) & \operatorname{tr} \mathbb{E}Q_{01}(z,\eta) \\ \operatorname{tr} \mathbb{E}Q_{10}(z,\eta) & \operatorname{tr} \mathbb{E}Q_{11}(z,\eta)\end{pmatrix} \xrightarrow{\text{a.s.}} 0.$$

Let $x^{\mathcal N} = (U+\imath V)/\sqrt{2n}$, where U and V are real independent standard Gaussian random variables. Define $X^{\mathcal N} = \left(x^{\mathcal N}_{ij}\right)_{i,j=0}^{N-1,n-1}$, where the $x^{\mathcal N}_{ij}$ are independent copies of $x^{\mathcal N}$. Let $Q^{\mathcal N}_{ij}(z,\eta)$ be the analogues of the $Q_{ij}(z,\eta)$, obtained by replacing the matrix X with $X^{\mathcal N}$. Then,

$$\frac{1}{n}\begin{pmatrix} \operatorname{tr} \mathbb{E}Q_{00}(z,\eta) & \operatorname{tr} \mathbb{E}Q_{01}(z,\eta) \\ \operatorname{tr} \mathbb{E}Q_{10}(z,\eta) & \operatorname{tr} \mathbb{E}Q_{11}(z,\eta)\end{pmatrix} - \frac{1}{n}\begin{pmatrix} \operatorname{tr} \mathbb{E}Q^{\mathcal N}_{00}(z,\eta) & \operatorname{tr} \mathbb{E}Q^{\mathcal N}_{01}(z,\eta) \\ \operatorname{tr} \mathbb{E}Q^{\mathcal N}_{10}(z,\eta) & \operatorname{tr} \mathbb{E}Q^{\mathcal N}_{11}(z,\eta)\end{pmatrix} = \mathcal{O}_\eta(n^{-1/2}).$$

Thanks to Proposition 26, we reduce our problem to the study of $n^{-1}\operatorname{tr}\mathbb{E}Q_{ij}(z,\eta)$ in the complex Gaussian case. We now invoke the so-called Integration by Parts (IP) formula for Gaussian variables [32]. Let $w = (w_0,\ldots,w_{n-1})^\top$ be a complex Gaussian random vector with $\mathbb{E}w = 0$, $\mathbb{E}ww^\top = 0$, and $\mathbb{E}[ww^*] = \Xi$. Let $\varphi = \varphi(w_0,\ldots,w_{n-1},\bar w_0,\ldots,\bar w_{n-1})$ be a C^1 complex function which is polynomially bounded together with its derivatives. Then, the IP formula reads as

$$\mathbb{E}w_k\varphi(w) = \sum_{\ell=0}^{n-1}[\Xi]_{k\ell}\,\mathbb{E}\left[\frac{\partial\varphi(w)}{\partial\bar w_\ell}\right]. \tag{37}$$

We shall apply this formula to the case $w \equiv X$ and $\varphi \equiv u^*Qv$ where $Q = Q(z,\eta)$ is the resolvent given by Eq. (31) (seen as a function of X), and u and v are deterministic vectors in $\mathbb{C}^{2N}$. By a standard derivation, we have

$$\frac{\partial u^*Qv}{\partial\bar x_{ij}} = -u^*Q\begin{pmatrix} 0 & XJe_{n,j}e^*_{N,i} \\ XJ^{-1}e_{n,j}e^*_{N,i} & 0\end{pmatrix}Qv.$$

In particular, by taking $u = e_{2N,k}$ and $v = e_{2N,\ell}$ for $k,\ell \in [N]$, we obtain that

$$\frac{\partial[Q_{00}]_{k,\ell}}{\partial\bar x_{ij}} = -[Q_{01}XJ^{-1}]_{kj}[Q_{00}]_{i\ell} - [Q_{00}XJ]_{kj}[Q_{10}]_{i\ell}, \tag{38}$$

and by taking $u = e_{2N,k}$ and $v = e_{2N,N+\ell}$ for $k,\ell \in [N]$, we get

$$\frac{\partial[Q_{01}]_{k,\ell}}{\partial\bar x_{ij}} = -[Q_{00}XJ]_{kj}[Q_{11}]_{i\ell} - [Q_{01}XJ^{-1}]_{kj}[Q_{01}]_{i\ell}. \tag{39}$$

Given $M \in \mathbb{C}^{n\times n}$, we shall also use the trivial relations $[MJ^k]_{\cdot,j} = [M]_{\cdot,j+k}$ and $[J^kM]_{i,\cdot} = [M]_{i-k,\cdot}$, where both the sum $j+k$ and the difference $i-k$ are taken modulo-n.

We can now start our calculations. Recalling that x_j refers to the j^{th} column of X for $j \in [n]$, our first task is to study quadratic forms of the type $x_k^* Q_{00} x_\ell$ and $x_k^* Q_{01} x_\ell$. Define the matrices

$$A_{00} = \mathbb{E}\left(x_k^* Q_{00} x_\ell\right)_{k,\ell=0}^{n-1} \quad \text{and} \quad A_{01} = \mathbb{E}\left(x_k^* Q_{01} x_\ell\right)_{k,\ell=0}^{n-1}.$$

It is obvious that $X \stackrel{\mathscr{L}}{=} XJ^m$ for each $m \in \mathbb{Z}$. Thus, given a measurable function $f : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$ and the integers $k, \ell, m \in [n]$, it holds that

$$x_{k+m}^* f(XJX^*) x_{\ell+m} = e_{n,k+m}^* X^* f(XJX^*) X e_{n,\ell+m} = e_{n,k}^* J^{-m} X^* f(XJ^m J J^{-m} X^*) X J_m e_{n,\ell} \stackrel{\mathscr{L}}{=} x_k^* f(XJX^*) x_\ell,$$

where the index summations are taken modulo-n. As a consequence, the matrices A_{00} and A_{01} are circulant matrices, a fact very useful to us.

Starting with A_{00}, we have by the IP formula (37),

$$\begin{aligned}
\mathbb{E} x_k^* Q_{00} x_\ell &= \sum_{i,j=0}^{N-1} \mathbb{E}[(\bar{x}_{ik}[Q_{00}]_{ij}) x_{j\ell}] = \frac{1}{n}\sum_{i,j} \mathbb{E}\left[\frac{\partial(\bar{x}_{ik}[Q_{00}]_{ij})}{\partial \bar{x}_{j\ell}}\right] \\
&= \frac{1}{n}\sum_{i,j} \mathbb{1}_{i=j}\mathbb{1}_{k=\ell}\mathbb{E}[Q_{00}]_{ij} - \frac{1}{n}\sum_{i,j}\mathbb{E}[Q_{01}XJ^{-1}]_{i\ell}\bar{x}_{ik}[Q_{00}]_{jj} - \frac{1}{n}\sum_{i,j}\mathbb{E}[Q_{00}XJ]_{i\ell}\bar{x}_{ik}[Q_{10}]_{jj} \quad (\text{using } 38) \\
&= \mathbb{1}_{k=\ell}\mathbb{E}\operatorname{tr} Q_{00}/n - \mathbb{E}\left[[X^*Q_{01}XJ^{-1}]_{k\ell} \operatorname{tr} Q_{00}/n\right] - \mathbb{E}\left[[X^*Q_{00}XJ]_{k\ell}\operatorname{tr} Q_{10}/n\right].
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E} x_k^* Q_{01} x_\ell &= \sum_{i,j=0}^{N-1} \mathbb{E}[(\bar{x}_{ik}[Q_{01}]_{ij}) x_{j\ell}] = \frac{1}{n}\sum_{i,j}\mathbb{E}\left[\frac{\partial(\bar{x}_{ik}[Q_{01}]_{ij})}{\partial \bar{x}_{j\ell}}\right] \quad (\text{using } 39) \\
&= \mathbb{1}_{k=\ell}\mathbb{E}[\operatorname{tr} Q_{01}/n] - \mathbb{E}\left[[X^*Q_{00}XJ]_{k\ell}\operatorname{tr} Q_{11}/n\right] - \mathbb{E}\left[[X^*Q_{01}XJ^{-1}]_{k\ell}\operatorname{tr} Q_{01}/n\right].
\end{aligned}$$

In the right side of the above two expressions we have terms of the type $\mathbb{E}\left[[\cdots]_{k\ell}\operatorname{tr} Q_{ij}/n\right]$. We now need to decouple $[\cdots]_{k\ell}$ from $\operatorname{tr} Q_{ij}/n$. Specifically, we have the following lemma.

Lemma 27. For each $i, j \in \{0,1\}$ and each $k, \ell \in [n]$,

$$\mathbb{V}\mathrm{ar}(\operatorname{tr} Q_{ij}/n) = \mathscr{O}_\eta(n^{-2}) \quad \text{and} \quad \mathbb{V}\mathrm{ar}(x_k^* Q_{ij} x_\ell) = \mathscr{O}_\eta(n^{-1}).$$

This lemma can be proven with the help of, *e.g.*, the so-called Poincaré-Nash inequality [9], [32], which is a particular case of the Brascamp-Lieb inequality [24]. A result of this sort is standard in random matrix theory. We omit its proof for lack of space.

Let us write $q_{ij} = q_{ij}(z,\eta) = n^{-1}\mathbb{E}\operatorname{tr} Q_{ij}(z,\eta)$ for $i, j \in \{0,1\}$. Using Lemma 27, and applying the Cauchy-Schwartz inequality, it is easy to see that

$$\begin{aligned}
\mathbb{E} x_k^* Q_{00} x_\ell &= \mathbb{1}_{k=\ell} q_{00} - \mathbb{E}\left[[X^*Q_{01}XJ^{-1}]_{k\ell}\right] q_{00} - \mathbb{E}\left[[X^*Q_{00}XJ]_{k\ell}\right] q_{10} + \mathscr{O}_\eta(n^{-3/2}), \\
\mathbb{E} x_k^* Q_{01} x_\ell &= \mathbb{1}_{k=\ell} q_{01} - \mathbb{E}\left[[X^*Q_{00}XJ]_{k\ell}\right] q_{11} - \mathbb{E}\left[[X^*Q_{01}XJ^{-1}]_{k\ell}\right] q_{01} + \mathscr{O}_\eta(n^{-3/2}).
\end{aligned}$$

Since Y is a square matrix, we see from (31) that $q_{00} = q_{11}$. Thus, the equations above can be written in a matrix form as

$$A_{00}(I_n + q_{10}J) + q_{00}A_{01}J^{-1} = q_{00}I_n + \mathscr{O}_\eta(n^{-3/2}), \text{ and} \tag{40}$$

$$q_{00}A_{00}J + A_{01}(I_n + q_{01}J^{-1}) = q_{01}I_n + \mathscr{O}_\eta(n^{-3/2}). \tag{41}$$

Let us give these equations a more symmetric form. Developing $(41) \times q_{00}J^{-1} - (40) \times (I + q_{01}J^{-1})$, we get that

$$A_{00}\left[q_{00}^2 - (I_n + q_{10}J)(I_n + q_{01}J^{-1})\right] = -q_{00} + \mathscr{O}_\eta(n^{-3/2}). \tag{42}$$

Similarly, taking $(40) \times q_{00}J - (41) \times (I + q_{10}J)$,

$$A_{01}\left[q_{00}^2 - (I_n + q_{10}J)(I_n + q_{01}J^{-1})\right] = q_{00}^2 J - q_{01}(I_n + q_{10}J) + \mathscr{O}_\eta(n^{-3/2}). \tag{43}$$

Now, by using the obvious identity $Q(\Sigma - \eta) = I_{2N}$ we obtain

$$\begin{aligned}
-\eta Q_{00} - \bar{z}Q_{01} + Q_{01}XJ^{-1}X^* &= I_N, \\
-zQ_{00} - \eta Q_{01} + Q_{00}XJX^* &= 0,
\end{aligned}$$

(the similar equations involving the terms Q_{10} and Q_{11} will not be used). Taking the traces of the expectations, we get

$$-\eta q_{00} - \bar{z} q_{01} + n^{-1} \operatorname{tr} A_{01} J^{-1} = \gamma_n, \tag{44}$$

$$-z q_{00} - \eta q_{01} + n^{-1} \operatorname{tr} A_{00} J = 0, \tag{45}$$

where $\gamma_n = N/n$.

Recalling that $q_{00}^{(n)}(z,\eta) = n^{-1} \mathbb{E} \operatorname{tr} Q^{(n)}(z,\eta)$, the function $\gamma_n^{-1} q_{00}^{(n)}(z,\cdot)$ is the Stieltjes transform of the probability measure $\mathbb{E}\check{\nu}_{n,z}$. Hence, $|\gamma_n^{-1} q_{00}^{(n)}(z,\eta)| \le 1/\Im\eta$. So, $\{q_{00}^{(n)}(z,\cdot)\}_{n\in\mathbb{N}}$ is a normal family of holomorphic functions on $\mathbb{C}_+$. Similarly, $q_{01}^{(n)}(z,\cdot) = n^{-1}\mathbb{E} \operatorname{tr} Q_{01}^{(n)}(z,\cdot)$ and $q_{10}^{(n)}(z,\eta) = n^{-1}\mathbb{E}\operatorname{tr} Q_{10}^{(n)}(z,\eta)$ are holomorphic functions in $\eta \in \mathbb{C}_+$ whose absolute values are bounded by $\sup_n \gamma_n / \Im\eta$.

Using the normal family theorem, let us extract from the sequence (n) a subsequence (still denoted as (n)) such that $q_{00}^{(n)}(z,\cdot)$, $q_{01}^{(n)}(z,\cdot)$, and $q_{10}^{(n)}(z,\cdot)$ converge to holomorphic functions in the sense of uniform convergence on the compact subsets of $\mathbb{C}_+$. Denote these functions respectively as $p(z,\cdot)$, $d(z,\cdot)$ and $\tilde{d}(z,\cdot)$. We shall show that they uniquely solve a system of equations on the line segment $\imath[C,\infty)$ of the positive imaginary axis, where C is some positive constant. This will show that $p(z,\cdot)$ is uniquely defined on $\mathbb{C}_+$, and that $q_{00}(z,\cdot) \to_n p(z,\cdot)$ and $q_{01}(z,\cdot) \to_n d(z,\cdot)$ on $\mathbb{C}_+$. We then show that $t\Im p(z,\imath t) \to \gamma$ as $t\to\infty$. This will lead to the fact that $\gamma^{-1} p(z,\cdot)$ is the Stieltjes transform of a symmetric probability measure $\check{\nu}_z$.

Assume that $\eta = \imath t$ where $t > 0$. Then, since the measure $\mathbb{E}\check{\nu}_{n,z}$ is symmetric, $q_{00}(z,\imath t) = \imath s(z,t)$ with $s(z,t) > 0$. Moreover, we notice from the expressions of Q_{01} and Q_{10} in (31) that $q_{10}(z,\imath t) = \bar{q}_{01}(z,\imath t)$.

Recall that A_{00} and A_{01} are circulant matrices. Writing $\mathsf{F}_n = n^{-1/2}[\exp(2\imath\pi k\ell/n)]_{k,\ell=0}^{n-1}$, the circulant matrix J can be written as $J = \mathsf{F}_n \operatorname{diag}(\exp(-2\imath\pi k/n))_{k=0}^{n-1} \mathsf{F}_n^*$. Notice that the matrices A_{00}, A_{01} and J commute, since they are circulant. Now, (42) can be rewritten as $A_{00} P = \imath s + E$ where $E = \mathcal{O}_t(n^{-3/2})$ is a circulant matrix, and

$$P = s^2 + (I_n + \bar{q}_{01} J)(I_n + q_{01} J^*) = \mathsf{F}_n \operatorname{diag}\left(s^2 + |1 + q_{01}\exp(2\imath\ell/n)|^2\right)_{\ell=0}^{n-1} \mathsf{F}_n^*. \tag{46}$$

If $t \ge 2\sup_n \gamma_n$, then $|q_{01}| \le 1/2$, and thus, the positive definite matrix P satisfies $P \ge (1/4) I_n$ in the semi-definite positive ordering. In view of Equation (45), we need an expression for $n^{-1}\operatorname{tr} A_{00} J$. We can write

$$\frac{\operatorname{tr} A_{00} J}{n} = \frac{\imath s \operatorname{tr} P^{-1} J}{n} + \frac{\operatorname{tr} P^{-1} J E}{n} = \frac{\imath s}{n} \sum_{\ell=0}^{n-1} \frac{\exp(-2\imath\pi\ell/n)}{s^2 + |1 + q_{01}\exp(2\imath\pi\ell/n)|^2} + \frac{\operatorname{tr} P^{-1} J E}{n}. \tag{47}$$

Given two square matrices M_1 and M_2 of the same size, it is well known that $|\operatorname{tr} M_1 M_2| \le (\operatorname{tr} M_1 M_1^*)^{1/2} (\operatorname{tr} M_2 M_2^*)^{1/2}$. Thus, since $E = \mathcal{O}_t(n^{-3/2})$, we get that

$$\frac{|\operatorname{tr} P^{-1} J E|}{n} \le \frac{1}{n}\sqrt{\operatorname{tr} P^{-2}}\sqrt{\operatorname{tr} E E^*} \le \frac{1}{n} 2 n^{1/2} \mathcal{O}_t(n^{-1/2}) = \mathcal{O}_t(n^{-1}).$$

By a similar derivation, and in view of Equation (44), we also get from Equation (43) that

$$\frac{\operatorname{tr} A_{01} J^{-1}}{n} = \frac{1}{n}\sum_{\ell=0}^{n-1} \frac{s^2 + |q_{01}|^2 + q_{01}\exp(2\imath\pi\ell/n)}{s^2 + |1 + q_{01}\exp(2\imath\pi\ell/n)|^2} + \mathcal{O}_t(n^{-1}). \tag{48}$$

Now, taking n to infinity along the subsequence (n) in Equations (44), (45), (47), and (48), writing $p(\imath t) = \imath h(z,t)$ where $h(z,t) \ge 0$, and noting that $\tilde{d}(z,\imath t) = \bar{d}(z,\imath t)$, the pair $(h(z,t), d(z,\imath t))$ satisfies the system of Equations (33) of the statement of Theorem 21 for $t \ge 2\sup_n \gamma_n$.

Let us consider the system of equations in $(h,d) \in (0,\infty)\times\mathbb{C}$

$$-th + \bar{z} d = u(h,d) - \gamma, \tag{49a}$$

$$zh + td = v(h,d), \tag{49b}$$

where $u(h,d)$ and $v(h,d)$ are given by Equations (34). Writing

$$I(a,u) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{a^2 + |1 + u\exp(\imath\theta)|^2} d\theta \quad \text{and} \quad J(a,u) = \frac{1}{2\pi}\int_0^{2\pi} \frac{\exp(\imath\theta)}{a^2 + |1 + u\exp(\imath\theta)|^2} d\theta,$$

the system (49) can be rewritten as

$$-th + \bar{z} d = (h^2 + |d|^2) I(h,d) + d J(h,d) - \gamma, \tag{50a}$$

$$zh + td = h\bar{J}(h,d). \tag{50b}$$

By the residue theorem (derivations omitted), the integrals are given by the expressions

$$I(a,u) = \frac{1}{\sqrt{(a^2 + |u|^2 + 1)^2 - 4|u|^2}}, \quad \text{and} \quad J(a,u) = \frac{1}{2u}\left(1 - \frac{a^2 + |u|^2 + 1}{\sqrt{(a^2 + |u|^2 + 1)^2 - 4|u|^2}}\right) \tag{51}$$

for each $a \in \mathbb{R}$ and $u \in \mathbb{C}$ such that $a \ne 0$ or $|u| \ne 1$.

Lemma 28. There exists $C > 0$ (depending on z and γ) such that for each $t \in [C, \infty)$, the system (49) has a unique solution (h, d) such that $h \in (0, \gamma/t)$ and $|d| < \gamma/t$.

Sketch of proof. Using (50b), we show that we can assume without generality loss that z and d are real. Rewriting the system (49) as $(h, d)^\mathsf{T} = f((h, d)^\mathsf{T})$, we show by computing the Jacobian matrix of f that if C is large enough, f is a Banach contraction on $[0, 2\gamma/t] \times [-\gamma/t, \gamma/t]$. □

We now prove that $th(z,t) = t\Im p(z, \imath t) \to \gamma$ as $t \to \infty$. The functions $h(z,t)$ and $d(t)$ satisfy (33a), and furthermore, $0 \le h(z,t), |d(t)| \le \gamma/t$. From the expressions (51), it is clear that $(h^2 + d^2)I(h(z,t), d(t))$ and $d(t)J(h(z,t), d(t))$ converge to zero as $t \to \infty$. The result is then obtained from Equation (50a).

We also need to prove that $(h(z,t), d(z, \imath t))$ satisfy the system (33) for each $t > 0$. By the convergence $q_{00}(z, \cdot) \to p(z, \cdot)$, we get that $\mathbb{E}\check{\nu}_{n,z} \Rightarrow \check{\nu}_z$. In particular, $\mathbb{E}\check{\nu}_{n,z}$ is tight. Let $a > 0$ be such that $\inf_n \mathbb{E}\check{\nu}_{n,z}([-a, a]) \ge 1/2$. By an easy derivation involving the expression of a Stieltjes transform, we then get that $s(z, \imath t) \ge \gamma_n t/(2(a^2 + t^2))$. Therefore, for each $t > 0$, the matrix P defined in (46) satisfies $P \ge \gamma t/(4(a^2 + t^2))I$ in the semidefinite ordering for all large enough n. By repeating the argument that follows Equation (46), we obtain that $(p(z, \imath t), d(z, \imath t))$ solve the system (33).

The proof of Theorem 21 is completed by combining Proposition 26 with the convergence of $q_{00}(z, \cdot)$ to $p(z, \cdot)$.

Proposition 22: Sketch of proof. Assume first that $X \stackrel{\mathscr{L}}{=} X^{\mathscr{N}}$, where $X^{\mathscr{N}}$ was defined in the statement of Proposition 26. Fixing $z \neq 0$, and writing $q_{00}(z, \imath t) = \imath s$, we obtain from Equations (44), (45), (47), and (48) that (s, q_{01}) satisfy a system which is a finite-n analogue to Equations (49) with a $\mathscr{O}_t(n^{-1})$ error. This system can be used to show that there exist constants $\alpha, C > 0$ such that $s \in (0, C(1 + n^{-1}t^{-\alpha})]$ for $t \in (0, 1]$. The Gaussian assumption is then removed with the help of Proposition 26.

6 Identification of μ. Proof of Proposition 25

The following preliminary lemma can be proven by inspecting Equations (49) and by using that $p(z, \cdot)/\gamma$ is the Stieltjes transform of $\check{\nu}_z$.

Lemma 29. For each $z \neq 0$, the function $h(z,t)$ is bounded for $t \in (0, \infty)$, and $h(z,t)/t$ is lower-bounded by a positive constant for $t \in (0, 1]$. Moreover, $|d(z, \imath t)| \le C/|z|$ for $C > 0$.

In the proof of Proposition 25, we shall use the fact that $(h(z,t), d(z, \imath t))$ satisfies the system of equations (50). We rewrite Equation (50a) as $\gamma = (h^2 + |d|^2)I(h,d) + dJ(h,d) - \bar{z}d + th$, and Equation (50b) as $\bar{z}h + t\bar{d} = hJ(h,d)$, or equivalently, as $\bar{z}d = dJ(h,d) - t|d|^2/h$. Since $h(z,t) > 0$ for $t > 0$, we can use the expressions (51) of the integrals $I(h,d)$ and $J(h,d)$ to obtain

$$\begin{aligned} \gamma &= \frac{h^2 + |d|^2}{\sqrt{\Delta(h,d)}} + \frac{t}{h}\left(h^2 + |d|^2\right), \\ 2\bar{z}d &= 1 - \frac{h^2 + |d|^2 + 1}{\sqrt{\Delta(h,d)}} - 2|d|^2\frac{t}{h} = 1 - \gamma - \frac{\gamma}{h^2 + |d|^2} + \frac{t}{h}\left(h^2 + 1 - |d|^2\right), \end{aligned} \tag{52}$$

where $\Delta(h,d) = (h^2 + |d|^2 + 1)^2 - 4|d|^2$. We now let $t \to 0$. Here, each sequence $t_k \to 0$ satisfies one of two cases : either $t_k/h(z, t_k) \to 0$, or $t_k/h(z, t_k) \to \alpha$ where α is a positive number. Indeed, Lemma 29 shows that $t_k/h(z, t_k) \to \infty$ is excluded.

The case $t_k/h(z, t_k) \to 0$. Using Lemma 29, and taking a further subsequence, still denoted as (k), we can assume that $d(t_k) \to b \in \mathbb{C}$ and $h(z, t_k) \to r \ge 0$. The pair (r, b) satisfies the equations

$$\gamma^2\Delta(r,b) = (r^2 + |b|^2)^2, \quad \text{and} \tag{53}$$

$$2\bar{z}b = 1 - \gamma - \frac{\gamma}{r^2 + |b|^2}. \tag{54}$$

By Equation (54), the number $y = -\bar{z}b$ is real and satisfies

$$r^2 + |b|^2 = r^2 + \frac{y^2}{|z|^2} = \frac{\gamma}{1 - \gamma + 2y}. \tag{55}$$

Moreover, we have $\Delta(r,b) = ((\gamma/(1 - \gamma + 2y) + 1)^2 - 4y^2/|z|^2$. Replacing in (53), we get

$$\left(\frac{\gamma}{1 - \gamma + 2y} + 1\right)^2 - 4\frac{y^2}{|z|^2} = \frac{1}{(1 - \gamma + 2y)^2}.$$

Reducing to the same denominator, we get after some simple manipulations that $|z|^2 = g(y)$, where g is the function given in the statement of Theorem 2. Let us delineate the domain of variation of y. Equation $|z|^2 = g(y) = (1 - \gamma + 2y)^2 y/(y+1)$ shows that $y(y+1) > 0$, thus $y < -1$ or $y > 0$. By Equation (55), $\frac{\gamma}{1-\gamma+2y} \ge \frac{y^2}{|z|^2} = \frac{y(y+1)}{(1-\gamma+2y)^2}$. We therefore get that

$2y+1-\gamma>0$ and furthermore, by rearranging the terms of the inequality above, that $y^2+(1-2\gamma)y+\gamma(\gamma-1)\leq 0$. The last inequality implies that $\gamma-1\leq y\leq\gamma$. In conclusion, we get that $y\in[0\vee(\gamma-1),\gamma]\setminus\{0\}$.

The case $t_k/h(z,t_k)\to\alpha>0$. Here we get of course that $h(z,t_k)\to 0$. Taking a subsequence if necessary, we shall assume that $d(t_k)\to b$. Getting back to the system (52) and taking t_k to zero, we get that

$$\gamma\left|1-|b|^2\right| = |b|^2+\alpha|b|^2\left|1-|b|^2\right|,$$
$$2\bar{z}b|b|^2 = (1-\gamma)|b|^2-\gamma+\alpha|b|^2(1-|b|^2).$$

The first equation implies that $|b|\notin\{0,1\}$, and that $\alpha=\frac{\gamma}{|b|^2}-\frac{1}{|1-|b|^2|}$. Replacing α by its value in the second equation, we get after a simple calculation that $2\bar{z}b=1-2\gamma-\frac{1-|b|^2}{|1-|b|^2|}$. Here we need to consider two cases: either $|b|<1$ or $|b|>1$. If $|b|<1$, we get from the last equation that $b=-\gamma/\bar{z}$ (thus, $|z|\geq\gamma$). Plugging in the expression of α, we get that $\alpha=|z|^2\left(\frac{1}{\gamma}-\frac{1}{|z|^2-\gamma^2}\right)$. Since $\alpha>0$, this implies that $|z|>\sqrt{\gamma(\gamma+1)}$. If $|b|>1$, we obtain that $b=(1-\gamma)/\bar{z}$, thus, $|z|<|1-\gamma|$ and $\alpha=|z|^2\left(\frac{\gamma}{(1-\gamma)^2}-\frac{1}{(1-\gamma)^2-|z|^2}\right)$. Using again that $\alpha>0$, we get after a small calculation that $\gamma>1$ and $|z|^2\leq(\gamma-1)^3/\gamma$. Let us summarize our conclusions for clarity.

- If $t_k/h(z,t_k)\to 0$, let b be an arbitrary accumulation point of $d(z,t_k)$, and let $y=-\bar{z}b$. If $\gamma\leq 1$, then $y\in(0,\gamma]$, and $|z|^2=g(y)\in(0,\gamma(\gamma+1)]$. If $\gamma>1$, then $y\in[\gamma-1,\gamma]$, and $|z|^2=g(y)\in[(\gamma-1)^3/\gamma,\gamma(\gamma+1)]$.
- If $t_k/h(z,t_k)$ converges to a positive constant, let b be an arbitrary accumulation point of $d(z,t_k)$. If $\gamma\leq 1$, then $|z|^2>\gamma(\gamma+1)$, and $b=-\gamma/\bar{z}$. If $\gamma>1$, thein either $|z|^2>\gamma(\gamma+1)$ in which case $b=-\gamma/\bar{z}$, or $|z|^2<(\gamma-1)^3/\gamma$, in which case $b=(1-\gamma)/\bar{z}$.

These statements show that given $z\neq 0$, the accumulation points b reduce to a genuine limit. Moreover, the behavior of this limit $b(z)$ is as described in the statement of Proposition 25.

From the point-wise convergence $d(z,\imath t)\to_{t\to 0} b(z)$ for $z\neq 0$ and Lemma 29, we get that $d(\cdot,\imath t)\to_{t\to 0} b(\cdot)$ in $\mathscr{D}'(\mathbb{C})$. Thus, $(2\gamma)^{-1}b(z)=\partial_{\bar{z}}U_\mu(z)$ in $\mathscr{D}'(\mathbb{C})$ by Proposition 23.

Acknowledgments

The visit of AB to France has been funded by the Indo-French Center for Applicable Mathematics project High Dimensional Random Matrix Models with Applications. This work has been partially supported by the French ANR grant HIDITSA (ANR-17-CE40-0003). The authors would like to thank Monika Bhattacharjee, Nicholas Cook, and David Renfrew for fruitful discussions.

References

[1] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[2] M. Bhattacharjee and A. Bose. Estimation of autocovariance matrices for infinite dimensional vector linear process. *J. Time Series Anal.*, 35(3):262–281, 2014.

[3] M. Bhattacharjee and A. Bose. Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.*, 44(2):598–628, 2016.

[4] M. Bhattacharjee and A. Bose. Polynomial generalizations of the sample variance-covariance matrix when $pn^{-1}\to 0$. *Random Matrices Theory Appl.*, 5(4), 2016.

[5] M. Bhattacharjee and A. Bose. Joint convergence of sample autocovariance matrices when $p/n\to 0$ with application. *Ann. Statist.*, 47(6):3470–3503, 2019.

[6] C. Bordenave, P. Caputo, and D. Chafaï. Circular law theorem for random Markov matrices. *Probab. Theory Related Fields*, 152(3-4):751–779, 2012.

[7] Ch. Bordenave and D. Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.

[8] A. Bose and M. Bhattacharjee. *Large covariance and autocovariance matrices*, volume 162 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2019.

[9] S. Chatterjee and A. Bose. A new method for bounding rates of convergence of empirical spectral distributions. *J. Theoret. Probab.*, 17(4):1003–1019, 2004.

[10] N. Cook. Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.*, 46(6):3442–3500, 11 2018.

[11] N. Cook, W. Hachem, J. Najim, and D. Renfrew. Non-Hermitian random matrices with a variance profile (I): deterministic equivalents and limiting ESDs. *Electron. J. Probab.*, 23:Paper No. 110, 61, 2018.

[12] P. Erdös. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[13] J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

[14] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29:669–679, 1984. English translation: Theory Probab. Appl. 29 (1984), 694-706.

[15] V. L. Girko. *Theory of stochastic canonical equations. Vol. I and II*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[16] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

[17] F. Götze. Asymptotic expansions for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete*, 50(3):333–355, 1979.

[18] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 07 2010.

[19] F. Götze and A. Tikhomirov. On the asymptotic spectrum of products of independent random matrices. *arXiv preprint arXiv:1012.2710*, 2010.

[20] A. Guionnet, M. Krishnapur, and O. Zeitouni. The single ring theorem. *Ann. of Math. (2)*, 174:1189–1217, 2011.

[21] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.

[22] S. S. Haykin and A. O. Steinhardt. *Adaptive radar detection and estimation*, volume 11. Wiley-Interscience, 1992.

[23] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[24] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[25] W. Li, Z. Li, and J. Yao. Joint central limit theorem for eigenvalue statistics from several dependent large dimensional sample covariance matrices with application. *Scand. J. Stat.*, 45(3):699–728, 2018.

[26] Z. Li, C. Lam, J. Yao, and Q. Yao. On testing for high-dimensional white noise. *arXiv preprint arXiv:1808.03545*, 2018.

[27] 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.

[28] H. Liu, A. Aue, and D. Paul. On the Marčenko-Pastur law for linear time series. *Ann. Statist.*, 43(2):675–712, 2015.

[29] Ph. Loubaton. On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices. *J. Theoret. Probab.*, 29(4):1339–1443, 2016.

[30] H. H. Nguyen. On the least singular value of random symmetric matrices. *Electron. J. Probab.*, 17:53, 2012.

[31] S. O'Rourke, D. Renfrew, A. Soshnikov, and V. Vu. Products of independent elliptic random matrices. *J. Stat. Phys.*, 160(1):89–119, 2015.

[32] L. A. Pastur. A simple approach to the global regime of Gaussian ensembles of random matrices. *Ukrain. Mat. Zh.*, 57(6):790–817, 2005.

[33] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[34] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[35] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[36] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169:595–632, 2009.

[37] H. L. Van Trees. *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[38] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures & Algorithms*, 44(2):135–182, 2014.

[39] L. Wang, A. Aue, and D. Paul. Spectral analysis of linear time series in moderately high dimensions. *Bernoulli*, 23(4A):2181–2209, 2017.

[40] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab Theory Relat Fields*, 78(4):509–521, 1988.