On eigenvalue distributions of large auto-covariance matrices

Jianfeng Yao and Wangjun Yuan

Department of Statistics and Actuarial Science
The University of Hong Kong
e-mail: jeffyao@hku.hk

Department of Mathematics
The University of Hong Kong
e-mail: ywangjun@connect.hku.hk

Abstract: In this article, we establish a limiting distribution for eigenvalues of a class of auto-covariance matrices. The same distribution has been found in the literature for a regularized version of these auto-covariance matrices. The original non-regularized auto-covariance matrices are non invertible, thus introducing supplementary difficulties for the study of their eigenvalues through Girko’s Hermitization scheme. The key result in this paper is a new polynomial lower bound for a specific family of least singular values associated to a rank-defective quadratic function of a random matrix with independent and identically distributed entries. Another innovation from the paper is that the lag of the auto-covariance matrices can grow to infinity with the matrix dimension.

AMS 2000 subject classifications: Primary 60B20, 15B52.
Keywords and phrases: Girko’s Hermitization principle, Eigenvalue distribution, Large auto-covariance matrix, Least singular value.

1. Introduction

Most of matrix ensembles studied in random matrix theory are either Hermitian or unitary. General random matrices without such symmetry or invariance are much less studied as their eigenvalues can lay everywhere in the complex plane. Typically, eigenvalues of non-Hermitian matrices are much more unstable than those of the Hermitian ones, and their study needs new mathematical tools. These tools have taken long time to emerge as illustrated by the history of the circular law. The law states that the empirical spectral distribution of a $n \times n$ random matrix with i.i.d entries of mean 0 and variance $1/n$ converges to the uniform distribution on the unit disk of the plane. When the common distribution is complex Gaussian and as early as in 1964, Ginibre [1964] established the circular law in expectation (although the abstract of the paper ended with a quite confusing statement that “the limit of the eigenvalue density as $n \to \infty$ is constant over the whole complex plane”). The real Gaussian equivalent of Ginibre’s result was later established by Edelman [1997]. The general non-Gaussian case was first tackled by Girko [1984] who coined the name of circular law, and more importantly, introduced a powerful analytic tool, known as Girko’s Hermitization, which is followed in all the subsequent papers. His own result remains however controversial as few key steps in his argument, were deemed not fully justified at the time [Tao, 2012, Section 2.8]. Bai [1997] provided a rigorous proof of the circular law assuming a few additional conditions on the moments and density functions of the entries. These non necessary conditions are afterwards removed in several subsequent papers before reaching the final circular law with minimal conditions established in Tao et al. [2010]. We refer to Bordenave and Chafaï [2012] for a
detailed description of these different episodes of the circular law.

One immediately notes that such multi-decade efforts were only about the simplest non-Hermitian matrix filled with i.i.d. entries. Recent progress is made with a more involved matrix ensemble, known as structured random matrices, which has the form $A_n \circ X^{(n)} + B_n$ where the random matrix $X^{(n)}$ has i.i.d. entries as in the circular law, and $(A_n)$ and $(B_n)$ are two sequences of deterministic matrices (the $\circ$ denotes the Hadamard product), see Cook [2018]. This model includes the class of random matrices with profile Cook et al. [2018] with $B_n \equiv 0$, and $A_n(i, j) = \sigma(i/n, j/n)$ for $(i, j) \in [n]^2$ driven by a scalar profile function $\sigma : (0, 1)^2 \to (0, \infty)$. The model also contains the class of band matrices with $B_n \equiv 0$, and $A_n(i, j) = 0$ if $|i - j| > k$ for some given bandwidth $k \geq 1$. The limiting distribution of random matrices with profiles is obtained in Cook et al. [2018] where a key ingredient is a polynomial lower bounds for a specific family of least singular values established in Cook [2018].

In this paper we study a particular random matrix of the following form. Let $N = N(n)$ and $1 \leq k = k(n) < n$ be sequences of positive integers varying with $n \in \mathbb{N}_+$. For each $n$, consider a rectangular $N \times n$ random matrix $X^{(n)} = (X^{(n)}_{ij})_{1 \leq i \leq N, 1 \leq j \leq n} \in \mathbb{C}^{N \times n}$ with i.i.d. complex-valued entries with mean 0 and variance $1/n$. The matrix of interest is

$$Y^{(n)} = X^{(n)} A^{(n)} (X^{(n)})^*, \quad (1.1)$$

where $(A^{(n)})$ is the sequence of deterministic matrices with entries $A^{(n)}_{ij} = \delta_{i=j+k}$, i.e. $A^{(n)}$ is of the form

$$A^{(n)} = \begin{pmatrix} 0 & 0 \\ I_{n-k} & 0 \end{pmatrix}. \quad (1.2)$$

The non-Hermitian matrix $Y^{(n)}$ originates from high-dimensional time series analysis. Write $X^{(n)}$ in function of its $n$ column vectors of dimension $N$, as in $X^{(n)} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Clearly,

$$Y^{(n)} = \mathbf{x}_{k+1} \mathbf{x}_1^* + \cdots + \mathbf{x}_n \mathbf{x}_{n-k}^*.$$ 

In this form, the matrix $Y^{(n)}$ is seen as the so-called lag-$k$ auto-covariance matrix of the time series $(\mathbf{x}_j)_{1 \leq j \leq n}$ in the space $\mathbb{C}^N$ and observed at time $1 \leq j \leq n$. The case of $k = 1$ would be of special interest and we denote it as $Y_1^{(n)}$. The spectral property of the matrix has a fundamental role for the analysis of the series. For example it helps to test the hypothesis whether the series is a white noise, that is the series is indeed an i.i.d. sequence. The limiting distribution of the singular values of $Y_1^{(n)}$ has been found in Li et al. [2015], Wang and Yao [2016]; it has been applied to high-dimensional statistics in Li et al. [2017, 2019].

To fix the discussions, throughout the paper the dimension parameters are taken to satisfy the following asymptotic scheme:

$$\lim_{n \to \infty} \frac{N}{n} = \gamma_0 \in (0, \infty), \quad \lim_{n \to \infty} \frac{k}{n} = \gamma_1 \in [0, 1). \quad (1.3)$$

Very few is known on the eigenvalue distribution of the matrix $Y^{(n)}$. Simulation and plots are given in the book Bose and Bhattacharjee [2018] for (see Figure 8.1 there). In a recent paper Bose and Hachem [2020], the authors consider a variant of $Y_1^{(n)}$, namely

$$Z^{(n)} = x_2 x_1^* + \cdots + x_n x_{n-1}^* + x_1 x_n^* = Y_1^{(n)} + x_1 x_n^*. \quad (1.4)$$
They established that the empirical spectral distribution of $Z^{(n)}$ converges to a deterministic limiting distribution $\mu^{(n)}$ in probability (this distribution will be detailed later). Therefore the two matrices $Y^{(n)}_1$ and $Z^{(n)}$ differ only by the rank-one matrix $x_1x_1^\ast$. However due to the mentioned high spectral instability, rank-one perturbations can preserve or destroy the spectrum of the original matrix depending on their nature. In other words, the existence of the LSD $\mu^{(n)}$ for $Z^{(n)}$ does not imply anything a priori on the asymptotic properties of the lag-1 auto-covariance matrix $Y^{(n)}_1$.

Actually the present paper aims at establishing the LSD for the general lag-$k$ auto-covariance matrix $Y^{(n)}$ under specific conditions. Note that while in Bose and Hachem [2020], the matrix $Z^{(n)}$ corresponds to the case $k = 1$, we allow $k$ growing with $n$ in this paper. Technically, by mimicking the methodology introduced in the development of the circular law, the main technical challenge here is to establish a polynomial lower bounds for the least singular value of the matrix $Y^{(n)} - zI_N$ for almost all $z \in \mathbb{C}$. Consider for a moment the method employed in Bose and Hachem [2020] for the establishment of the LSD $\mu^{(n)}$ for the matrix $Z^{(n)}$. Note that this matrix can be rewritten as

$$Z^{(n)} = X^{(n)}J^{(n)}(X^{(n)})^\ast,$$

with the permutation matrix

$$J^{(n)} = \begin{pmatrix} 0 & 1 \\ I_{n-1} & 0 \end{pmatrix}.$$

In a setting where the entries of $X^{(n)}$ have a (common) density and noting that $J^{(n)}$ is of full rank, the matrix $Z^{(n)}$ is of full rank almost surely. This is a main ingredient for the method in Bose and Hachem [2020] to establish a polynomial lower bound for the least singular value of the matrices $Z^{(n)} - zI_N$ ($z \in \mathbb{C}$). (Note that in the reference, such polynomial lower bound is established for general degree two monomials of the form $X^{(n)}C^{(n)}(X^{(n)})^\ast$ where $C^{(n)}$ is asymptotically non degenerated). This method is broken in our case of $Y^{(n)}$ since the inner matrix $A^{(n)}$ in $Y^{(n)}$ is nilpotent and has rank $n-k$. We thus introduce a specially designed non-degenerated auxiliary matrix $H(z) \in \mathbb{C}^{(N+n-k) \times (N+n-k)}$ for the matrix $Y^{(n)} - zI_N$, which has a smaller least singular value. A careful analysis leads to a manageable polynomial bound for the least singular value of $H(z)$ which is thus easily transferred to $Y^{(n)} - zI_N$. This construction of a lower bound for the matrix $Y^{(n)} - zI_N$ is indeed the main technical innovation of the paper. It is developed in Section 3. Note that in the case of the circular law or the structured matrices of Cook [2018], the matrix is linear in its independent entries. In contrary, the matrix $Y^{(n)}$, as well as the matrix $Z^{(n)}$ is a more involved quadratic function of these independent entries. Note that complex Gaussian valued auto-covariance matrices were also considered in Nowak and Tarnowski [2017].

The rest of the paper is as follows. Section 2 recalls a few preliminaries and useful results from the literature. Section 3 presents the main result of the paper, that is, a polynomial bound on the least singular value of the matrix $Y^{(n)} - zI_N$. Applying this bound leads to the LSD for the matrix $Y^{(n)}$ in Section 4. The two appendices collect a few useful but standard lemmas from linear algebra and probability theory.

Below are some useful notations.

- A ball with center $z \in \mathbb{C}$ and radius $r \geq 0$ is denoted as $B(z, r)$. Let $T$ be a set of complex numbers, then $B(T, r) = \cup_{z \in T} B(z, r)$.
For an integer $n$, define $[n] = [1, n] \cap \mathbb{N}_+$. For a vector $u \in \mathbb{C}^n$ and a set $I \subseteq [n]$, $u_I$ is the sub-vector of $u$ with indexes in $I$. Similarly for a matrix $M \in \mathbb{C}^{n \times n}$ and index sets $I, J \subseteq [n]$, $M_{I,J}$ denotes the submatrix of $M$ restricted to rows with index in $I$ and columns with index in $J$. In the case that the set $I$ contains one element $i_0$ only, we may write $M_{i_0,J}$. The same abbreviation also applies to the column index set $J$.

For any index set $I \subset [n]$, let $\Pi_I : \mathbb{C}^n \to \mathbb{C}^n$ be a projection such that $(\Pi_I u)_i = u_{i_1 i \in I}$.

Without further indication, all vectors in this paper are column vectors. The canonical basis of $\mathbb{C}^n$ is denoted by $\{e^{(n)}_1, ..., e^{(n)}_n\}$.

For a given matrix $M \in \mathbb{C}^{p \times q}$, let $s_1(M) \geq \cdots \geq s_r(M)$ be its ordered singular values, where $r = \min\{p, q\}$. We use the convention that $s_j(M) = 0$ for $j > r$. We also denote by $\nu_M = \frac{1}{p} \sum_{j=1}^r \delta_{s_j(M)}$ the singular values empirical distribution of the matrix. If $p = q$, then $\{\lambda_1(M), \ldots, \lambda_p(M)\}$ denote the set of eigenvalues of $M$, and $\mu_M = \frac{1}{p} \sum_{j=1}^p \delta_{\lambda_j(M)}$ the corresponding empirical spectral distribution.

For a given matrix $M$, $\|M\|$ denotes its operator norm, and $\|M\|_{HS}$ its Hilbert-Schmidt norm of $M$.

Denote $\iota = \sqrt{-1}$.

2. Preliminaries

To ease the reading of the proofs in Section 3 and 4, we collect a few concepts and results from the literature that will be used afterwards.

2.1. Compressible vectors and incompressible vectors

For $\theta, \rho \in (0, 1)$, we define the set of compressible vectors

$$\text{Comp}(\theta, \rho) = \mathbb{S}^{n-1} \cap \bigcup_{I \subseteq [n], |I| = \theta n} B(\mathbb{S}_I^{n-1}, \rho)$$

where $\mathbb{S}_I^{n-1}$ is the $(n-1)$-dimensional sphere centered at the origin in $\mathbb{C}^n$ with radius $\rho$, and

$$= \{u \in \mathbb{S}^{n-1} : \exists J \subseteq [n], |J| = \theta n, \exists v \in \mathbb{S}^{n-1}, \text{ s.t. } v_{[n] \setminus J} = 0 \text{ and } \|u - v\| \leq \rho\},$$

and the set of incompressible vectors

$$\text{Incomp}(\theta, \rho) = \mathbb{S}^{n-1} \setminus \text{Comp}(\theta, \rho).$$

The following lemma describes the structure of the set of incompressible vectors, which can be found in [Rudelson and Vershynin, 2008, Lemma 3.4] or [Cook, 2018, Lemma 2.1].

**Lemma 2.1.** [Rudelson and Vershynin, 2008, Lemma 3.4] For $\theta, \rho \in (0, 1)$, for any $u = (u_1, \ldots, u_n)^T \in \text{Incomp}(\theta, \rho)$, the set

$$J = \left\{i \in [n] : \frac{\rho}{\sqrt{n}} \leq |u_i| \leq \frac{2}{\sqrt{\theta n}} \right\}$$

has cardinal number $|J| \geq 3\theta n/4$.

Lemma 2.1 can be slightly extend to the following lemma, which can be found in Bose and Hachem [2020].
Lemma 2.2. [Bose and Hachem, 2020, Lemma 8] For \(\theta, \rho \in (0, 1)\), for any \(u = (u_1, \ldots, u_n)^T \in \text{Incomp}(\theta, \rho)\), and \(\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_n)^T \in S^{n-1}\), the set
\[ J' = \left\{ i \in [n] : \frac{\rho}{\sqrt{n}} \leq |u_i| \leq \frac{2}{\sqrt{\theta n}}, |\tilde{u}_i| \leq \frac{2}{\sqrt{\theta n}} \right\}, \]
has cardinal number \(|J'| \geq \theta n/2\).

Proof. Denote
\[ J'' = \left\{ i \in [n] : |\tilde{u}_i| \leq \frac{2}{\sqrt{\theta n}} \right\}, \]
then \(|(J'')^c| \leq \theta n/4\). As \(J' = J \cap J''\), \(|J'| \geq |J| - |(J'')^c| \geq \theta n/2\), by Lemma 2.1. \(\square\)

The following lemma describes the invertibility of a matrix via distance, which can be found in Rudelson and Vershynin [2008].

Lemma 2.3. [Rudelson and Vershynin, 2008, Lemma 3.5] Let \(A \in \mathbb{C}^{n \times n}\) be any random matrix. Let \(A_{-k}\) be the span of all column vectors of \(A\) except the \(k\)-th column. Then for every \(\theta, \rho \in (0, 1)\) and every \(\epsilon > 0\), one has
\[ \mathbb{P} \left( \inf_{x \in \text{Incomp}(\theta, \rho)} \|Ax\| < \frac{\rho t}{\sqrt{n}} \right) \leq \frac{\epsilon}{\theta n} \sum_{k=1}^{n} \mathbb{P} \left( \text{dist}(A_{[n],k}, A_{-k}) \leq t \right). \]

The following lemma deals with the distance of the columns of a matrix, which can be found in Bose and Hachem [2020].

Lemma 2.4. [Bose and Hachem, 2020, page 5] Let \(A \in \mathbb{C}^{n \times n}\). For \(k \in [n]\), \(A_{-k}\) be the space spanned by \(A_{[n],i}\) with \(i \neq k\). Then for \(k \in [n]\)
\[ \text{dist}(A_{[n],k}, A_{-k}) = \frac{|A_{kk} - A_{k,[n]\{k\}} (A_{[n]\{k\},[n]\{k\}})^{-1} A_{[n]\{k\},k}|}{\sqrt{1 + \|A_{k,[n]\{k\}} (A_{[n]\{k\},[n]\{k\}})^{-1}\|^2}}. \]

The following lemma estimates the metric entropy of the sphere, which is introduced in Cook [2018].

Lemma 2.5. [Cook, 2018, Lemma 2.2] Let \(V \subseteq \mathbb{C}^n\) be a subspace of (complex) dimension \(k\), and let \(T \subseteq V \cap S^{n-1}\). For \(\rho \in (0, 1)\), \(T\) has a \(\rho\)-net \(\Sigma \subseteq T\) of cardinality \(|\Sigma| \leq (3/\rho)^{2k}\).

2.2. Small ball probability

The following definition is the small ball probability, which can be found in Tao and Vu [2008].

Definition 2.1. [Tao and Vu, 2008, Definition 3.1] Let \(Z = (Z_1, \ldots, Z_n)\) be a complex random vector with independent entries. For any \(r \geq 0\), we define the small ball probability
\[ S \left( \sum_{j=1}^{n} Z_j, r \right) = \sup_{z \in \mathbb{C}} \mathbb{P} \left( \sum_{j=1}^{n} Z_j \in B(z, r) \right). \]
The following lemma states that the small ball probability is monotone with respect to the dimension.

**Lemma 2.6.** [Rudelson and Vershynin, 2008, Lemma 2.1] For any \( r \geq 0 \) and any index set \( I \subseteq [n] \),
\[
S \left( \sum_{j \in [n]} Z_j, r \right) \leq S \left( \sum_{j \in I} Z_j, r \right).
\]

The following lemma is the Berry-Esseen theorem for small ball probability, which can be found in Bordenave and Chafaï [2012].

**Lemma 2.7.** [Bordenave and Chafaï, 2012, Lemma A.6] Suppose that the independent complex random variables \( Z_1, \ldots, Z_n \) are centered with finite third moments. Then there exists a constant \( c' > 0 \), such that
\[
S \left( \sum_{j=1}^{n} Z_j, r \right) \leq c' r \sqrt{\sum_{j=1}^{n} \mathbb{E}[|Z_j|^2]} + c' \sum_{j=1}^{n} \mathbb{E}[|Z_j|^3] \left( \sum_{j=1}^{n} \mathbb{E}[|Z_j|^2] \right)^{3/2}.
\]

The following lemma estimates the probability that a quadratic form is bounded, which is introduced in Bose and Hachem [2020].

**Lemma 2.8.** [Bose and Hachem, 2020, Lemma 20] Let \( a \in \mathbb{C}, u,v \in \mathbb{C}^n, M \in \mathbb{C}^{n \times n} \) be deterministic. Let \( Z \in \mathbb{C}^n \) be a random variable with independent entries, and \( Z' \) be an independent copy of \( Z \). Let \( I \subseteq [n] \), then for each \( t > 0 \),
\[
\mathbb{P}(\left| Z^* M Z + u^* Z + Z^* v + a \right| \leq t)^2 \leq \mathbb{E}_{Z_{I_c}, Z'_{I_c}} \left[ S_{Z_1} \left( (Z_{I_c} - Z'_{I_c})^* M_{I_c,I_c} Z_1 + Z_{I_c} M_{I_c,I_c} (Z_{I_c} - Z'_{I_c}) \right), 2t \right] .
\]

Here, \( S_{Z_1} \) is the small ball probability defined in Definition 2.1, where the expectation is taken with respect to \( Z_1 \). We also use the convention that the right hand side is 1 if \( I = \emptyset \) or \([n] \).

### 2.3. Logarithmic potential

Let \( \mathcal{P}(\mathbb{C}) \) be the set of probability measures on \( \mathbb{C} \) which \( \ln |x| \) is integrable in a neighbourhood of infinity.

**Definition 2.2.** The logarithmic potential \( U_\mu \) of \( \mu \in \mathcal{P}(\mathbb{C}) \) is the function defined by
\[
\mathcal{L}_\mu(z) = -\int_{\mathbb{C}} \ln |z - \lambda| d\mu(\lambda), z \in \mathbb{C}.
\]

The following lemmas are from Bordenave and Chafaï [2012]

**Lemma 2.9.** [Bordenave and Chafaï, 2012, Lemma 4.1] For \( \mu, \nu \in \mathcal{P}(\mathbb{C}) \), if \( \mathcal{L}_\mu = \mathcal{L}_\nu \) almost everywhere, then \( \mu = \nu \).

**Lemma 2.10.** [Bordenave and Chafaï, 2012, Lemma 4.3] Let \( B_n \in \mathbb{C}^{n \times n} \) be a complex random matrix. Suppose that there exists a family of non-random probability measures \( \{\nu_z : z \in \mathbb{C}\} \) on \( \mathbb{R}_+ \), such that \( \nu_{B_n - z I_n} \) converges to \( \nu_z \) in probability as \( n \to \infty \), and the function \( \ln(x) \) is uniformly integrable for the family \( \{\nu_{B_n - z I_n} : n \in \mathbb{N}_+\} \) in probability, for almost all
We use the formula of inverse blocking matrices to find an invertible matrix $H \in \mathbb{C}^{(N+n-k)\times(N+n-k)}$, such that $s_{N+n-k}(H) \leq s_N(XAX^* - zI_N)$. Then we estimate the infimum over compressible vectors in Section 3.1 by using an $\epsilon$-net argument, and over incompressible vectors, by using Lemma 2.3, in Section 3.2 (the case $2k+1 \leq n$), and in Section 3.3 (the case $2k+1 > n$), respectively. In Section 3.2, due to the structure of $H$, we estimate the distance $\text{dist}(H_{[N+n-k+l],H_{-l}})$ for $N+k+1 \leq l \leq N+n-k$ in Section 3.2.1, for $N+1 \leq l \leq N+k$ in Section 3.2.2, and for $1 \leq l \leq N$ in Section 3.2.3, respectively. In Section 3.3, we estimate the distance $\text{dist}(H_{[N+n-k+l],H_{-l}})$ for $N+1 \leq l \leq N+n-k$ in Section 3.3.1, and for $1 \leq l \leq N$ in Section 3.3.2, respectively.
To start the proof, we can assume that $X_{11}$ has density by a perturbation argument (see Bose and Hachem [2020]). We fix arbitrary $z \in \mathbb{C} \setminus \{0\}$. All constants in the proof may depend on $z$. Denote $X = (X_1, \ldots, X_n)$ then we have

$$XAX^* = (X_1, \ldots, X_n)A \begin{pmatrix} X_1^* \\ \vdots \\ X_n^* \end{pmatrix} = (X_{k+1}, \ldots, X_n, 0, \ldots, 0) \begin{pmatrix} X_1^* \\ \vdots \\ X_n^* \end{pmatrix} = (X_{k+1}, \ldots, X_n) \begin{pmatrix} X_1^* \\ \vdots \\ X_{n-k}^* \end{pmatrix}. $$

We first show that $zI_N - XAX^*$ is invertible with probability 1. Indeed, we consider the function $f(W) = \det(zI_N - WAW^*)$ for $W = (W_1, \ldots, W_n) \in \mathbb{C}^{N \times n}$, which is a polynomial of the entries of $W$. We have $f \left( (e_1^{(N)}, \ldots, e_1^{(N)}) \right) = (z - n + k)z^{N-1} \neq z^N = f(0)$, which implies that $f(W)$ is a non-zero polynomial of the entries of $W$. Thus, the polynomial hypersurface $\{W : f(W) = 0\}$ has zero Lebesgue measure in $\mathbb{C}^{N \times N}$. Since the entries of $X$ have density, $f(X) \neq 0$ almost surely. Hence, if we denote

$$H' = \begin{pmatrix} zI_N \\ (X_1, \ldots, X_{n-k})^* \\ I_{n-k} \end{pmatrix} \begin{pmatrix} X_{k+1}, \ldots, X_n \end{pmatrix} \in \mathbb{C}^{(N+n-k) \times (N+n-k)},$$

then by Lemma A.1, we have

$$(H')^{-1} = \begin{pmatrix} (zI_N - XAX^*)^{-1} \\ * \\ * \end{pmatrix}. $$

Thus, by Lemma A.3,

$$s_N \left( XAX^* - zI_N \right) = \frac{1}{s_1 \left( (XAX^* - zI_N)^{-1} \right)} \geq \frac{1}{s_1 \left( (H')^{-1} \right)} = s_{N+n-k} \left( H' \right) \tag{3.2}$$

Define $H = H'$ if $2k + 1 > n$, and

$$H = \begin{pmatrix} zI_N \\ (X_1, \ldots, X_{n-k})^* \\ (X_1^{(n-k)}, \ldots, X_1^{(n-k)}, X_{k+1}, \ldots, X_{n-k}) \end{pmatrix} \in \mathbb{C}^{(N+n-k) \times (N+n-k)},$$

if $2k + 1 \leq n$. When $2k + 1 \leq n$, the two blocks $(X_1, \ldots, X_{n-k})^*$ and $(X_{k+1}, \ldots, X_n)$ in $H'$ have overlapping variables $(X_{k+1}, \ldots, X_{n-k})$, which are in the middle of the matrix. These variables are moved to the last columns in $H$ (they are already at the last rows of $H$). This is useful when we apply later Lemma 2.4 to get a quadratic form (see the numerator of the distance in Lemma 2.4): the central inverse matrix in this quadratic form becomes independent of the neighboring vectors.

Note that in the case $2k + 1 \leq n$, one can obtain $H$ from $H'$ by permuting the columns. Thus, $H$ and $H'$ have same singular values. Hence, by (3.2), it is enough to show

$$\mathbb{P} \left( s_{N+n-k}(H) \leq n^{-37/22}, \|X\| \leq C_0 \right) \leq C n^{-1/22}, \tag{3.3}$$

for all $z \in \mathbb{C} \setminus \{0\}$.

Note that for any $\theta, \rho \in (0, 1)$,

$$s_{N+n-k}(H) = \inf_{w \in S^{N+n-k}} \|Hw\| = \inf_{w \in \text{Comp}(\theta, \rho)} \|Hw\| \wedge \inf_{w \in \text{Incomp}(\theta, \rho)} \|Hw\|. $$
Then the conclusion of the theorem will follow from the following two key estimates:

\[
\mathbb{P} \left( \inf_{w \in \text{Comp}(\theta, \rho)} \|Hw\| \leq c, \|X\| \leq C_0 \right) \leq \exp(-cn), \tag{3.4}
\]

and

\[
\mathbb{P} \left( \inf_{w \in \text{Incomp}(\theta, \rho)} \|Hw\| \leq n^{-37/22}, \|X\| \leq C_0 \right) \leq Cn^{-1/22}, \tag{3.5}
\]

separably for a pair of special \(\theta\) and \(\rho\). These estimates are established in next subsections, respectively.

### 3.1. Estimate (3.4) for compressible vectors

We first derive the proof of (3.4) for the case \(2k + 1 \leq n\). The investigation of compressible vectors can be transferred to the study of deterministic vectors by the \(\epsilon\)-net argument (Lemma 2.5).

For a deterministic vector \(w = (u^T, v^T)^T \in \mathbb{S}^{N + n - k - 1}\), where \(u \in \mathbb{C}^N\) and \(v \in \mathbb{C}^{n - k}\), for \(0 \leq t \leq c\), by Lemma B.2, we have

\[
\begin{align*}
\mathbb{P} \left( \|Hw\| \leq t, \|u\| > 1/2 \right) & \leq \mathbb{P} \left( \left\| (X_1, \ldots, X_{n-k})^* u + e_{n-k+1}, \ldots, e_{n-k+1} e_{1} \ldots e_{n-k} v \right\| \leq t, \|u\| > 1/2 \right) \\
& \leq \mathbb{P} \left( \left\| (X_1, \ldots, X_{n-k})^* \frac{u}{\|u\|} + \frac{1}{\|u\|} (v_{k+1}, \ldots, v_{n-k}, v_{1}, \ldots, v_{k})^T \right\| \leq 2t, \|u\| > 1/2 \right) \\
& \leq \mathbb{P} \left( \text{dist} \left( (X_1, \ldots, X_{n-k})^* \frac{u}{\|u\|}, \text{Span} \{v_{k+1}, \ldots, v_{n-k}, v_{1}, \ldots, v_{k}\} \right) \leq 2t \right) \\
& \leq \exp(-cn). \tag{3.6}
\end{align*}
\]

Similarly, by Lemma B.2, we have

\[
\begin{align*}
\mathbb{P} \left( \|Hw\| \leq t, \|v\| > 1/2 \right) & \leq \mathbb{P} \left( \|zu + (X_{n-k+1}, \ldots, X_n, X_{k+1}, \ldots, X_{n-k})v\| \leq t, \|v\| > 1/2 \right) \\
& \leq \mathbb{P} \left( \left\| \frac{z}{\|v\|} u + (X_{n-k+1}, \ldots, X_n, X_{k+1}, \ldots, X_{n-k}) \frac{v}{\|v\|} \right\| \leq 2t, \|v\| > 1/2 \right) \\
& \leq \mathbb{P} \left( \text{dist} \left( (X_{n-k+1}, \ldots, X_n, X_{k+1}, \ldots, X_{n-k}) \frac{v}{\|v\|}, \text{Span} \{u\} \right) \leq 2t \right) \\
& \leq \exp(-cn). \tag{3.7}
\end{align*}
\]

Thus, by (3.6) and (3.7), for \(0 < t < c\),

\[
\mathbb{P} \left( \|Hw\| \leq t \right) \leq \mathbb{P} \left( \|Hw\| \leq t, \|u\| > 1/2 \right) + \mathbb{P} \left( \|Hw\| \leq t, \|v\| > 1/2 \right) \leq \exp(-cn). \tag{3.8}
\]

Note that on the event \(\{\|X\| \leq C_0\}\),

\[
\|H\| = \left\| \begin{pmatrix} zI_N & (X_{n-k+1}, \ldots, X_n, X_{k+1}, \ldots, X_{n-k}) \\ (X_1, \ldots, X_{n-k})^* & e_{n-k+1}, \ldots, e_{n-k+1} e_{1} \ldots e_{n-k} \end{pmatrix} \right\|
\]

where \(zI_N\) is the identity matrix of size \(N\), \(e_n\) is the \(n\)-th standard basis vector in \(\mathbb{C}^N\), and \(\epsilon_n = (\epsilon_{n-k+1}, \ldots, \epsilon_{n-k+1} e_{1} \ldots e_{n-k})\).
Thus, we choose $r_H$ to be a small number satisfying
\[ r_H \leq \frac{s_0}{4(|z| + 1 + C_0)}, \]
then by (3.8), we can choose $s_0 < c$ to obtain
\[ \mathbb{P} \left( \exists w' \in B(w, 2r_H) : \|Hw'\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \mathbb{P} \left( \|Hw\| \leq s_0, \|X\| \leq C_0 \right) \leq \exp(-cn), \]

Note that for small $\theta > 0$ that will be determined in the sequel, and $I \subseteq [N + n - k]$ with $|I| = \theta(N + n - k)$, by Lemma 2.5, the set of unit vector in $\mathbb{S}^{N+n-k-1}$ supported in $I$ has a $r_H$-net of cardinality bounded by $(3/r_H)^{2|I|}$. Thus,
\[ \mathbb{P} \left( \exists w' \in \text{Comp}(\theta, r_H) : \|Hw'\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \sum_{I \subseteq [N+n-k], |I| = \theta(N+n-k)} \mathbb{P} \left( \exists w' \in \mathbb{S}^{2n-1} \cap B(\mathbb{S}^{2n-1}_I, r_H) : \|Hw'\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \left( \frac{N + n - k}{\theta(N + n - k)} \right) \left( \frac{3}{r_H} \right)^{2\theta(N+n-k)} \mathbb{P} \left( \exists w' \in B(w, 2r_H) : \|Hw'\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \left( \frac{N + n - k}{\theta(N + n - k)} \right) \left( \frac{3}{r_H} \right)^{2\theta(N+n-k)} \exp(-cn). \]

By the Stirling formula,
\[ \left( \frac{N + n - k}{\theta(N + n - k)} \right) = \frac{(N + n - k) \cdots (N + n - k - \theta(N + n - k) + 1)}{(\theta(N + n - k))!} \leq \frac{(N + n - k)^{\theta(N+n-k)}}{(\theta(N + n - k))!} \sim \frac{(N + n - k)^{\theta(N+n-k)}e^{\theta(N+n-k)}}{\sqrt{2\pi \theta(N + n - k)(\theta(N + n - k))^{\theta(N+n-k)}}} = \frac{e^{\theta(N+n-k)}}{\sqrt{2\pi \theta(N + n - k)}}. \]

Hence, when $n$ large,
\[ \mathbb{P} \left( \exists w' \in \text{Comp}(\theta, r_H) : \|Hw'\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \left( \frac{e^{\theta(N+n-k)}}{\theta} \right) \left( \frac{3}{r_H} \right)^{2\theta(N+n-k)} \exp(-cn) = \exp \left( -cn + \theta(N + n - k) \ln \left( \frac{9e}{\theta r_H^2} \right) \right). \]
Since $\theta \ln((9e)/(\theta r_H^2))$ tends to zero as $\theta$ tends to zero, we can choose $\theta = \theta_0$, where $\theta_0$ is sufficiently small, such that

$$\mathbb{P} \left( \exists w' \in \text{Comp}(\theta_0, r_H) : \|H w\| \leq s_0/2, \|X\| \leq C_0 \right) \leq \exp(-cn),$$

which lead to (3.4).

The proof of (3.4) for the case $2k + 1 > n$ is similar and is omitted.

3.2. Estimate (3.5) for incompressible vectors for the case $2k + 1 \leq n$

We now establish the estimate (3.5) for the case $2k + 1 \leq n$ with $\theta = \theta_0$ and $\rho = r_H$. By Lemma 2.3, it is enough to prove

$$\mathbb{P} \left( \text{dist}(H_{[N+n-k],l}, H_{-l}) \leq n^{-13/11}, \|X\| \leq C_0 \right) \leq Cn^{-1/22}, \forall l \in [N + n - k]. \quad (3.10)$$

By Lemma 2.4, we have

$$\text{dist}(H_{[N+n-k],l}, H_{-l}) = \frac{\text{Num}}{\text{Den}}, \quad (3.11)$$

where

$$\text{Num} = \left| Hll - H_{l,[N+n-k]\setminus\{l\}} (H_{[N+n-k]\setminus\{l\},[N+n-k]\setminus\{l\}})^{-1} H_{[N+n-k]\setminus\{l\},l} \right|, \quad (3.12)$$

and

$$\text{Den} = \sqrt{1 + \left\| H_{l,[N+n-k]\setminus\{l\}} (H_{[N+n-k]\setminus\{l\},[N+n-k]\setminus\{l\}})^{-1} \right\|^2}. \quad (3.13)$$

Next, we compute (3.11). For the case $N + k + 1 \leq l \leq N + n - k$, Num in (3.12) is a quadratic form of the vector $X_{l-N}$. The computations are developed in Section 3.2.1. For the case $N + 1 \leq l \leq N + k$, Num is a bilinear form of the vectors $X_{l-N}$ and $X_{l-N+n-k}$. We develop the computation in Section 3.2.2, following the idea in Section 3.2.1 with necessary modification. The Num for the last case $1 \leq l \leq N$ is a mixture of the previous two cases, and the computations are developed in Section 3.2.3 using the same idea introduced in Section 3.2.1.

3.2.1. Case of $N + k + 1 \leq l \leq N + n - k$

We estimate (3.11) for the case $l = N + n - k$ first. Recall the definition of $H$, we have

$$H_{N+n-k,N+n-k} = 0, \quad H_{N+n-k,[N+n-k-1]} = \left( X_{n-k}^*, \left( e_{k}^{(n-k-1)} \right)^T \right),$$

$$H_{[N+n-k-1],N+n-k} = \left( \begin{array}{c} X_{n-k} \vspace{1em} e_{n-2k}^{(n-k-1)} \end{array} \right),$$

$$H_{[N+n-k-1],[N+n-k-1]} = \begin{pmatrix} zI_N & Y^{(1)} \\ Y^{(2)} & B \end{pmatrix}$$

$$= \begin{pmatrix} zI_N & (X_{n-k+1}, \ldots, X_n, X_{k+1}, \ldots, X_{n-k-1}) \\ (X_1, \ldots, X_{n-k-1})^* & \left( e_{n-2k+1}^{(n-k-1)}, \ldots, e_{n-k-1}^{(n-k-1)}, 0, e_1^{(n-k-1)}, \ldots, e_{n-2k-1}^{(n-k-1)} \right) \end{pmatrix}. \quad (3.14)$$
We first show that \( H_{[N+n-k-1],[N+n-k-1]} \) is invertible almost surely. Apply the row operation to the determinant, we can see that
\[
det \left( H_{[N+n-k-1],[N+n-k-1]} \right) = \det \left( B - z^{-1}Y^{(2)}Y^{(1)} \right).
\]
By a similar argument above, we can show that the determinant is a non-zero polynomial of the entries of \( X \). Since the entries of \( X \) have density, the determinant vanishes with probability zero.

Denote
\[
\left( H_{[N+n-k-1],[N+n-k-1]} \right)^{-1} = \begin{pmatrix} D & E \\ F & G \end{pmatrix},
\]
where \( D \in \mathbb{C}^{N \times N}, G \in \mathbb{C}^{(n-k-1) \times (n-k-1)} \).

Next, we compute the Num given by (3.12) and Den given by (3.13) individually.

**Step (a).** We first consider Num. This is a quadratic form of the vector \( X_{n-k} \) and can be estimated via small ball probability using Lemma 2.8 and Lemma 2.7.

Let \( \xi = \{ \xi_1, \ldots, \xi_N \} \) be i.i.d. Bernoulli random variables with \( \mathbb{P}(\xi_1 = 1) = p \), where \( p \in (0, 1) \) will be determined in the sequel. Moreover, we choose these variables to be independent of everything else. Set \( I = \{ i \in [N] : \xi_i = 1 \} \). Choose three random vectors \( x, x', x'' \in \mathbb{C}^N \), such that their entries are independent each other and independent of everything else, and that \( x, x', x'' \overset{d}{=} X_{n-k} \). Set
\[
u = (x)_I, \ \upsilon = (x')_I, \ \omega = (x'')_I.
\]
Denote \( \tilde{X} = (X_1, \ldots, X_{n-k-1}, X_{n-k+1}, \ldots, X_n) \), then by (3.12), (3.14), Cauchy-Schwarz inequality and Lemma 2.8, we have
\[
\mathbb{P}(\text{Num} \leq t, \|X\| \leq C_0)^2 \\
\leq \mathbb{P}(\text{Num} \leq t, \|\tilde{X}\| \leq C_0)^2 \\
= \mathbb{P}(\|\tilde{X}\| \leq C_0, \\
\left| X_{n-k}^*DX_{n-k} + \left( e_k^{(n-k-1)} \right)^T F X_{n-k} + X_{n-k}^*Ee_{n-k}^{(n-k-1)} + \left( e_k^{(n-k-1)} \right)^T Ge_{n-k}^{(n-k-1)} \right| \leq t \right)^2 \\
\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{X_{n-k}} \left[ 1 \left| X_{n-k}^*DX_{n-k} + \left( e_k^{(n-k-1)} \right)^T F X_{n-k} + X_{n-k}^*Ee_{n-k}^{(n-k-1)} + \left( e_k^{(n-k-1)} \right)^T Ge_{n-k}^{(n-k-1)} \right| \leq t \right] \right]^2 \\
\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ S_u \left( (v-w)^*D_{I_e}I_u + u^*D_{I_e}(v-w), 2t \right) \right] \left( 1 \|\tilde{X}\| \leq C_0 \right) \right] \\
= \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ S_u \left( (\Pi_{I_e}(x' - x''))^*D (\Pi_{I_e}x) + (\Pi_{I_e}x)^*D (\Pi_{I_e}(x' - x'')) , 2t \right) \right] \left( 1 \|\tilde{X}\| \leq C_0 \right) \right]. \quad (3.15)
\]

Denote
\[
y = \frac{D (\Pi_{I_e}(x' - x''))}{\|D (\Pi_{I_e}(x' - x''))\|}, \ \alpha = \frac{\sqrt{n} \|D (\Pi_{I_e}(x' - x''))\|}{\|D\|_{HS}},
\]
\[
\hat{\gamma}^* = \frac{\langle \Pi E(x' - x'') \rangle^* D}{\|\langle \Pi E(x' - x'') \rangle^* D \|} \, , \quad \hat{\alpha} = \sqrt{\frac{n}{D}} \|\langle \Pi E(x' - x'') \rangle^* D \|_{HS},
\]

Here, we use the convention that \( y = \hat{y} = 0 \) and \( \alpha = \hat{\alpha} = 0 \) if \( I = [N] \). Let \( W_i = \hat{\alpha}^{-1} \overline{y}_i x_i + \alpha y_i \overline{x}_i \) for \( i \in I \). When conditioning on \( v, w \) and \( \overline{X} \), \( \{W_i : i \in I\} \) are independent. Besides, we have

\[
\langle \Pi E(x' - x'') \rangle^* D (\Pi E x) + (\Pi E x)^* D \langle \Pi E(x' - x'') \rangle = \frac{D}{\sqrt{n}} \langle \hat{\alpha} \hat{y}^* (\Pi E x) + (\Pi E x)^* y \alpha \rangle
\]

\[
= \frac{D}{\sqrt{n}} \sum_{i \in I} W_i. \tag{3.16}
\]

Hence, by (3.15), (3.16), Definition 2.1 and Lemma 2.6,

\[
\mathbb{P} \left( \text{Num} \leq t, \|X\| \leq C_0 \right)^2 \leq \mathbb{E}_{\overline{X}} \left[ \mathbb{E}_{v, w} \left[ S_u \left( \|D\|_{HS} \sum_{i \in I} W_i, 2t \right) \mathbb{I}_{\|\overline{X}\| \leq C_0} \right] \right]
\]

\[
= \mathbb{E}_{\overline{X}} \left[ \mathbb{E}_{v, w} \left[ S_u \left( \sum_{i \in I} W_i, \frac{2\sqrt{n} t}{\|D\|_{HS}} \|\overline{X}\| \leq C_0 \right) \right] \right]
\]

\[
\leq \mathbb{E}_{\overline{X}} \left[ \mathbb{E}_{v, w} \left[ S_u \left( \sum_{i \in I \cap J} W_i, \frac{2\sqrt{n} t}{\|D\|_{HS}} \|\overline{X}\| \leq C_0 \right) \right] \right], \tag{3.17}
\]

where the index set \( J \) is given by

\[
J = \left\{ j \in [N] : \frac{r_H}{\sqrt{\theta_0 N}} \leq |y_j| \leq \frac{2}{\sqrt{\theta_0 N}}, |\overline{y}_j| \leq \frac{2}{\sqrt{\theta_0 N}} \right\}. \tag{3.18}
\]

Note that \( x_i \) has the same distribution as \( X_{11}^{(n)} \), by conditions (C1), (C2) and Cauchy-Schwarz inequality, we have

\[
\sum_{i \in I \cap J} \mathbb{E}_u \left[ |W_i|^2 \right] = \sum_{i \in I \cap J} \left( \hat{\alpha}^2 |\overline{y}_i|^2 + \alpha^2 |y_i|^2 \right) \mathbb{E} \left[ |x_i|^2 \right] + 2\alpha \hat{\alpha} \Re \left( \mathbb{E} \left[ x_i^2 \overline{y}_i \right] \right)
\]

\[
= \sum_{i \in I \cap J} \left( \hat{\alpha}^2 |\overline{y}_i|^2 + \alpha^2 |y_i|^2 \right) \mathbb{E} \left[ |X^{(n)}_{11}|^2 \right] + 2\alpha \hat{\alpha} \Re \left( \mathbb{E} \left[ \left( X^{(n)}_{11} \right)^2 \overline{y}_i \right] \right)
\]

\[
\geq \sum_{i \in I \cap J} \left( \hat{\alpha}^2 |\overline{y}_i|^2 + \alpha^2 |y_i|^2 \right) \mathbb{E} \left[ |X^{(n)}_{11}|^2 \right] - 2\alpha \hat{\alpha} \mathbb{E} \left[ \left( X^{(n)}_{11} \right)^2 |\overline{y}_i| |y_i| \right]
\]

\[
\geq \sum_{i \in I \cap J} \left( \hat{\alpha}^2 |\overline{y}_i|^2 + \alpha^2 |y_i|^2 \right) \left( \frac{1}{n} - \mathbb{E} \left[ \left( X^{(n)}_{11} \right)^2 \right] \right)
\]

\[
\geq \frac{c_0}{nN} r_H^2 \alpha^2 |I \cap J|, \tag{3.19}
\]

and

\[
\sum_{i \in I \cap J} \mathbb{E}_u \left[ |W_i|^3 \right] \leq 4 \sum_{i \in I \cap J} \mathbb{E}_u \left[ |\hat{\alpha} \overline{y}_i u_i|^3 + |\alpha y_i \overline{u}_i|^3 \right]
\]

\[
\leq \frac{32}{(\theta_0 N)^{3/2}} \sum_{i \in I \cap J} \mathbb{E}_u \left[ \hat{\alpha}^3 |u_i|^3 + \alpha^3 |\overline{u}_i|^3 \right].
\]
Lemma 3.1 (established later): for any deterministic vector $\mathbf{x}$, then by Hoeffding concentration inequality (Lemma B.1),

\[
\mathbb{E} \left[ \left| X_{11}^{(n)} \right|^3 \right] |I \cap J| \leq \frac{32}{(\theta_0 N)^{3/2}} (\alpha^3 + \tilde{\alpha}^3) \left( \mathbb{E} \left[ \left| X_{11}^{(n)} \right|^4 \right] \right)^{3/4} |I \cap J| \leq \frac{C}{(\theta_0 N n)^{3/2}} (\alpha^3 + \tilde{\alpha}^3) |I \cap J|.
\]

(3.20)

Here, $C$ is a large constant. Hence, by (3.19), (3.20) and Lemma 2.7,

\[
\mathcal{S}_u \left( \sum_{i \in I \cap J} W_i, \frac{2\sqrt{nt}}{\|D\|_H} \right) \leq \frac{c}{\sqrt{\sum_{i \in I \cap J} \mathbb{E}_u[|W_i|^2]}} \frac{2\sqrt{nt}}{\|D\|_H} + \frac{c}{\sqrt{\sum_{i \in I \cap J} \mathbb{E}_u[|W_i|^2]}} \frac{3\sqrt{nt}}{(\sum_{i \in I \cap J} \mathbb{E}_u[|W_i|^2])^{3/2}} \leq \frac{C n \sqrt{N t}}{r_H \alpha \sqrt{|I \cap J|}} \|D\|_H + \frac{C (\alpha^3 + \tilde{\alpha}^3)}{\theta_0^3 r_H^2 \alpha^3 |I \cap J|}.
\]

(3.21)

Next, we estimate the lower bound of $|I \cap J|$. Take $p = 1 - \theta_0/4$ and set the event

\[
\mathcal{E}_I = \{ |I| > \sqrt{N (1 - \theta_0/3)} \} = \left\{ \sum_{i=1}^N \xi_i > \sqrt{N (1 - \theta_0/3)} \right\},
\]

then by Hoeffding concentration inequality (Lemma B.1),

\[
\mathbb{P} \left( \mathcal{E}_I^c \right) = \mathbb{P} \left( \sum_{i=1}^N \xi_i \leq \sqrt{N (1 - \theta_0/3)} \right) = \mathbb{P} \left( \sum_{i=1}^N (\xi_i - \mathbb{E}[\xi_i]) \leq -N \theta_0/12 \right) \leq \exp(-cn \theta_0^2).
\]

(3.22)

Now we estimate $|J|$. To do this, we need the following estimation on the matrix $D$ from Lemma 3.1 (established later): for any deterministic vector $d \in \mathbb{C}^N \setminus \{0\}$,

\[
\mathbb{P} \left( \frac{Dd}{\|Dd\|} \in \text{Comp}(\theta_0, r_H), \|\tilde{X}\| \leq C_0 \right) \leq \exp(-cn N).
\]

Since, the entries of $x'$ and $x''$ are independent and have the same distribution as $X_{11}^{(n)}$, which has continuous density. If $I^c \neq \emptyset$, then $\Pi_{I^c} (x' - x'') = 0$ with probability zero. Denote

\[
\mathcal{E}_{y, \text{Incomp}} = \{ y \in \text{Incomp}(\theta_0, r_H) \},
\]

then by Lemma 3.1,

\[
\mathbb{P} \left( \mathcal{E}_{y, \text{Incomp}}^c \cap \left\{ \|\tilde{X}\| \leq C_0 \right\} \right) \leq \mathbb{P} (I = [N]) + \mathbb{P} \left( \frac{D (\Pi_{I^c} (x' - x''))}{\|D (\Pi_{I^c} (x' - x''))\|} \in \text{Comp}(\theta_0, r_H), \Pi_{I^c} (x' - x'') \neq 0 \right) = \mathbb{P} (I = [N]) + \mathbb{E}_{x', x''} \left[ \mathbb{P} \left( \frac{D (\Pi_{I^c} (x' - x''))}{\|D (\Pi_{I^c} (x' - x''))\|} \in \text{Comp}(\theta_0, r_H) \left| x', x'' \right\} 1_{\Pi_{I^c} (x' - x'') \neq 0} \right] \leq \mathbb{P} (I = [N]) + \exp(-cn N)
\]
\[= p^N + \exp(-cN) \leq \exp(-cN). \quad (3.23)\]

On \( \mathcal{E}_{y, \text{Incomp}} \), by Lemma 2.2, \(|J| \geq \theta_0 N/2 \). Thus, by (3.22), on the event \( \mathcal{E}_{y, \text{Incomp}} \cap \mathcal{E}_I \),

\[|I \cap J| \geq |I| + |J| - N > N(1 - \theta_0/3) + \theta_0 N/2 - N = \theta_0 N/6. \quad (3.24)\]

Next, we estimate \((\alpha^3 + \tilde{\alpha}^3)/\alpha^3\) and \(\alpha^{-1}\) in (3.21). Denote \( \mathcal{E}_\alpha = \{ \gamma \leq \alpha \leq \beta^{-1}, \tilde{\alpha} \leq \beta^{-1} \} \), where \( \beta \) and \( \gamma \) are \( o(1) \). Then on the event \( \mathcal{E}_\alpha \), we have

\[\frac{\alpha^3 + \tilde{\alpha}^3}{\alpha^3} \leq 2\beta^{-3}\gamma^{-3}, \quad \alpha^{-1} \leq \gamma^{-1}. \quad (3.25)\]

Next, we show that the event \( \mathcal{E}_\alpha \) has high probability. Indeed, we have

\[\mathbb{P}\left( \mathcal{E}_\alpha^c, \|\tilde{X}\| \leq C_0 \right) \leq \mathbb{P}\left( \alpha < \gamma, \|\tilde{X}\| \leq C_0 \right) + \mathbb{P}(\alpha > \beta^{-1}) + \mathbb{P}(\tilde{\alpha} > \beta^{-1}). \quad (3.26)\]

Recall that \((\Pi_i c(x' - x'')) = (1 - \xi_i) (x' - x'')_i \), so the entries of \( \Pi_i c(x' - x'') \) are i.i.d. with mean zero and variance \( 2(1 - p)/n \). Thus, by Markov inequality, the second term of the right hand side of (3.26) is

\[\mathbb{P}(\alpha > \beta^{-1}) = \mathbb{E}_{\tilde{X}} \mathbb{P}\left( \alpha > \beta^{-1} | \tilde{X} \right) = \mathbb{E}_{\tilde{X}} \mathbb{P}\left( \|D (\Pi_i c(x' - x''))\| > \frac{\|D\|_{HS}}{\sqrt{n} \beta} \|\tilde{X}\| \right) \leq n \beta^2 \mathbb{E}_{\tilde{X}} \mathbb{E}\left[ \|D (\Pi_i c(x' - x''))\|^2 \|\tilde{X}\| \right] = n \beta^2 \mathbb{E}_{\tilde{X}} \mathbb{E}\left[ \frac{\|D\|^2_{HS}}{n} \sum_{i_1, i_2, i_3 = 1}^n (\Pi_i c(x' - x''))_{i_1} D_{i_1 i_2}^{*} D_{i_2 i_3} (\Pi_i c(x' - x''))_{i_3} \|\tilde{X}\| \right] = 2(1 - p) \beta^2 \mathbb{E}_{\tilde{X}} \mathbb{E}\left[ \frac{\|D\|^2_{HS}}{n} \sum_{i_1, i_2 = 1}^n D_{i_1 i_2}^{*} D_{i_2 i_1} \right] = 2(1 - p) \beta^2. \quad (3.27)\]

By the same argument, the third term of the right hand side of (3.26) is also bounded by \( 2(1 - p) \beta^2 \). Next, we deal with the first term of (3.26). We denote

\[
\left( \tilde{u}^{(i)} \right)^* = \frac{\left( e_i^{(N)} \right)^* D}{\left\| \left( e_i^{(N)} \right)^* D \right\|}, \quad i \in [N].
\]

Then we have

\[
\|D (\Pi_i c(x' - x''))\|^2 = \sum_{i = 1}^N \left\| \left( e_i^{(N)} \right)^* D (\Pi_i c(x' - x'')) \right\|^2 = \sum_{i = 1}^N \left\| \left( e_i^{(N)} \right)^* D \right\|^2 \left( \tilde{u}^{(i)} \right)^* (\Pi_i c(x' - x''))^2. \quad (3.28)
\]
Denote
\[ I_i = \left\{ j \in [N] : \frac{r_H}{\sqrt{N}} \leq |\tilde{u}^{(i)}_j| \leq \frac{2}{\sqrt{\theta_0 N}} \right\}. \]

Note that
\[ \sum_{i=1}^{n} \left\| (e_i^{(N)})^* D \right\|^2 = \|D\|^2_{HS}, \]

by Lemma C.1, Lemma 2.6 and Lemma 2.7, we have
\[
P_{v,w,\xi} \left( \sum_{i=1}^{N} \left\| (e_i^{(N)})^* D \right\|^2 \left| (\tilde{u}^{(i)})^* (\Pi_F (x' - x'')) \right|^2 \leq \frac{\gamma^2}{n} \|D\|^2_{HS} \right)
\leq 2 \sum_{i=1}^{N} \left\| (e_i^{(N)})^* D \right\|^2 \|D\|^2_{HS} \left( \frac{C \sqrt{N} \gamma}{\sqrt{1 - pr_H \sqrt{|I_i|}}} + \frac{C}{\sqrt{1 - p\theta_0^{3/2} r_H^3 \sqrt{|I_i|}}} \right)
\leq 2 \left( \frac{C \sqrt{N} \gamma}{\sqrt{1 - pr_H}} + \frac{C}{\sqrt{1 - p\theta_0^{3/2} r_H^3 \sqrt{|I_i|}}} \right) \max_{i \in [N]} \frac{1}{\sqrt{|I_i|}}.
\]

Let
\[ E_{u,\text{Incomp}} = \left\{ \tilde{u}^{(i)} \in \text{Incomp}(\theta_0, r_H), \forall i \in [N] \right\}. \]

Then by Lemma 2.2, on \( E_{u,\text{Incomp}} \), \( |I_i| \geq \theta_0 N/2 \) for all \( i \in [N] \). Moreover, by Remark 3.1, \( P \left( \bar{E}_{u,\text{Incomp}}, \|\bar{X}\| \leq C_0 \right) \leq N \exp(-cN) \). Thus, by (3.28) and (3.29), the first term of (3.26) is
\[
P \left( \alpha < \gamma, \|\bar{X}\| \leq C_0 \right)
= P \left( \|D (\Pi_F (x' - x''))\| < \frac{\gamma}{\sqrt{n}} \|D\|_{HS}, \|\bar{X}\| \leq C_0 \right)
= P \left( \sum_{i=1}^{N} \left\| (e_i^{(N)})^* D \right\|^2 \left| (\tilde{u}^{(i)})^* (\Pi_F (x' - x'')) \right|^2 < \frac{\gamma^2}{n} \|D\|^2_{HS}, \|\bar{X}\| \leq C_0 \right)
= \mathbb{E}_{\bar{X}} \left[ P \left( \sum_{i=1}^{N} \left\| (e_i^{(N)})^* D \right\|^2 \left| (\tilde{u}^{(i)})^* (\Pi_F (x' - x'')) \right|^2 < \frac{\gamma^2}{n} \|D\|^2_{HS} \right| \|\bar{X}\| \leq C_0 \right].
\begin{align*}
\mathbb{E}_{\tilde{X}} \left[ \mathbb{P} \left( \sum_{i=1}^{N} \left\| \left( e_{i}^{(N)} \right)^{*} D \right\|^{2} \left( \tilde{u}^{(i)} \right)^{*} \left( \Pi_{B}(x' - x'') \right) \right\| ^{2} < \frac{\gamma^{2}}{n} \left\| D \right\|_{HS}^{2} \left\| \tilde{X} \right\| \right) \mathbb{1}_{\varepsilon_{u, \text{Incomp}}} \mathbb{1}_{\left\| \tilde{X} \right\| \leq C_{0}} \right] \\
+ \mathbb{E}_{\tilde{X}} \left[ \mathbb{P} \left( \sum_{i=1}^{N} \left\| \left( e_{i}^{(N)} \right)^{*} D \right\|^{2} \left( \tilde{u}^{(i)} \right)^{*} \left( \Pi_{B}(x' - x'') \right) \right\| ^{2} < \frac{\gamma^{2}}{n} \left\| D \right\|_{HS}^{2} \left\| \tilde{X} \right\| \right) \mathbb{1}_{\varepsilon_{u, \text{Incomp}}} \mathbb{1}_{\left\| \tilde{X} \right\| \leq C_{0}} \right]
\leq 2 \left( \frac{C \sqrt{N} \gamma}{\sqrt{1 - pr_{H}}} + \frac{C}{\sqrt{1 - pr_{H}^{3/2} r_{H}^{0}}} \right) \frac{\sqrt{\gamma}}{\sqrt{\theta_{0} N}} + \mathbb{P} \left( \varepsilon_{u, \text{Incomp}}, \left\| \tilde{X} \right\| \leq C_{0} \right)
\leq \frac{C \gamma}{\sqrt{(1 - p) \theta_{0} r_{H}}} + \frac{C}{\sqrt{(1 - p) \theta_{0}^{2} r_{H}^{3} \sqrt{N}}} + N \exp(-cN).
\end{align*}

Therefore, by (3.26), (3.27) and (3.30),
\begin{align*}
\mathbb{P} \left( \varepsilon_{\alpha}, \left\| \tilde{X} \right\| \leq C_{0} \right) \leq 4(1 - p) \beta^{2} + \frac{C \gamma}{\sqrt{(1 - p) \theta_{0} r_{H}}} + \frac{C}{\sqrt{(1 - p) \theta_{0}^{2} r_{H}^{3} \sqrt{N}}} + N \exp(-cN).
\end{align*}

Lastly, by (3.17), (3.21), (3.22), (3.23), (3.24), (3.25) and (3.31),
\begin{align*}
\mathbb{P} \left( \text{Num} \leq s \left\| D \right\|_{HS}, \left\| X \right\| \leq C_{0} \right)^{2}
\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ S_{u} \left( \sum_{i \in I \cap J} W_{i}, 2\sqrt{n}s \right)^{1}_{\left\| \tilde{X} \right\| \leq C_{0}} \right] \right]
\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ S_{u} \left( \sum_{i \in I \cap J} W_{i}, 2\sqrt{n}s \right)^{1}_{\left\| \tilde{X} \right\| \leq C_{0}} \right] \right] + \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ 1_{\left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\}} \cup \varepsilon_{u, \text{Incomp}} \cup \varepsilon_{I} \cup \varepsilon_{\alpha} \right] \right]
\leq \frac{C_{n s}}{r_{H} \gamma \sqrt{\theta_{0}}} + \frac{C_{\beta^{3} \gamma^{3}}}{\theta_{0}^{2} r_{H} \sqrt{N}} + \mathbb{P} \left( \left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\} \cup \varepsilon_{u, \text{Incomp}} \cup \varepsilon_{I} \cup \varepsilon_{\alpha} \right)
\leq \frac{C_{n s}}{r_{H} \gamma \sqrt{\theta_{0}}} + \frac{C_{\beta^{3} \gamma^{3}}}{\theta_{0}^{2} r_{H} \sqrt{N}} + \mathbb{P} \left( \left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\} \cup \varepsilon_{y, \text{Incomp}} \right) + \mathbb{P} \left( \varepsilon_{I} \right) + \mathbb{P} \left( \left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\} \cup \varepsilon_{\alpha} \right)
\leq \frac{C \left( \theta_{0}, r_{H} \right)}{\gamma} \frac{\sqrt{n s}}{\beta^{3} \gamma^{3} \sqrt{N}} + \frac{C \left( \theta_{0}, r_{H} \right)}{\beta^{3} \gamma^{3} \sqrt{N}} + \mathbb{P} \left( \left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\} \cup \varepsilon_{\alpha} \right)
+ \mathbb{P} \left( \left\{ \left\| \tilde{X} \right\| \leq C_{0} \right\} \cup \varepsilon_{\alpha} \right)
\leq \frac{C \left( \theta_{0}, r_{H} \right)}{\sqrt{N}} \left( \frac{1}{\gamma} + \frac{1}{\beta^{3} \gamma^{3} \sqrt{N}} + \beta^{2} + \gamma + \frac{1}{\sqrt{N}} \right).
\end{align*}

Here, \( C \left( \theta_{0}, r_{H} \right) \) is a large positive constant that may depend on \( \theta_{0}, r_{H} \) (and \( z \)). Then we may choose \( \beta = n^{-1/22} \) and \( \gamma = n^{-1/11} \) to obtain
\begin{align*}
\mathbb{P} \left( \text{Num} \leq s \left\| D \right\|_{HS}, \left\| X \right\| \leq C_{0} \right)^{2} \leq C \left( \theta_{0}, r_{H} \right) \left( n^{12/11} s + n^{-1/11} \right).
\end{align*}
Step (b). We compute $\text{Den}$ given by (3.13). In order to upper bound $\text{Den}$ by $\|D\|_{HS}$, we aim to upper bound 1, $\|F\|$, $\|E\|$ and $\|G\|$ by $\|D\|_{HS}$, up to some positive constants.

Recall the definition of $H$ in (3.14) and the blocking of $\left(H_{[N+n−k−1],[N+n−k−1]}\right)^{−1}$, we have the following identity

$$
\begin{pmatrix}
zI_N & Y^{(1)} \\
Y^{(2)} & B \\
\end{pmatrix}
\begin{pmatrix}
D & E \\
F & G \\
\end{pmatrix}
= 
\begin{pmatrix}
zI_N & Y^{(1)} \\
Y^{(2)} & B \\
\end{pmatrix}
= I_{N+n−k−1}.
\tag{3.33}
$$

We first control $\|F\|$ with high probability. For any $u \in \mathbb{S}^{N-1}$, denote $v = Du$ and $w = Fu$ then we have the identity

$$
\begin{pmatrix}
u \\
0 \\
\end{pmatrix} = 
\begin{pmatrix}
zI_N & Y^{(1)} \\
Y^{(2)} & B \\
\end{pmatrix}
\begin{pmatrix}
v \\
w \\
\end{pmatrix}.
\tag{3.34}
$$

Hence, on the event $\{\|X\| \leq C_0\}$, we have

$$
\|B^*Bw\| = \|B^*Y^{(2)}v\| \leq C_0 \|v\|.
\tag{3.35}
$$

Denote the event $\mathcal{E}_{X_n} = \{\|X_n\| \geq c\}$ for a small constant $c$. Then by Lemma B.2,

$$
\mathbb{P}(\mathcal{E}_{X_n}^c) \leq \exp(-cN).
\tag{3.36}
$$

As explained in the proof of Lemma 3.1, by (3.42), (3.35) and (3.34), on the event $\mathcal{E}_{X_n}$, we have

$$
\|Fu\|^2 = \|w\|^2 = \|B^*Bw\|^2 + |w_k|^2 \\
\leq C_0^2 \|v\|^2 + \frac{|w_k|^2}{\epsilon^2} \|X_n\|^2 \\
\leq C_0^2 \|v\|^2 + \frac{1}{\epsilon^2} \|Y^{(1)}w - Y^{(1)}B^*Bw\|^2 \\
\leq C_0^2 \|v\|^2 + \frac{1}{\epsilon^2} \|u - zv - Y^{(1)}B^*Bw\|^2 \\
\leq C \left(1 + \|v\|^2\right) \\
= C \left(1 + \|Du\|^2\right).
$$

Take the supremum over $u$, we obtain $\|F\| \leq C(z)(1 + \|D\|)$ on the event $\mathcal{E}_{X_n}$.

Next, on the event $\{\|X\| \leq C_0\}$, we have $\|D\| \geq c$ for a small constant $c$. Suppose not, then by (3.33), $\|BF\| = \|Y^{(2)}D\| \leq C_0c$. An argument that similar to (3.42), we may obtain from (3.33) that $I_N = zD + Y^{(1)}F = zD + Y^{(1)}B^*BF + X_nF_{k,[N]}$. Thus, $\|I_N - X_nF_{k,[N]}\| \leq |z|c + C_0^2c$. Then we can find an unit eigenvector $v'$ of the rank 1 matrix $X_nF_{k,[N]}$ associate to the eigenvalue 0. Then $\|I_N - X_nF_{k,[N]}\| \geq \|(I_N - X_nF_{k,[N]}) v'\| = \|v'\| = 1$, which leads to a contradiction if we choose $c$ to be small enough. Thus, on the event $\mathcal{E}_{X_n} \cap \{\|X\| \leq C_0\}$, we have $\|F\| \leq C\|D\|$. The control on $\|E\|$ and $\|G\|$ are similar, which are sketched below. By the identity

$$
(u^T, 0) = (u^T D, u^T E) 
\begin{pmatrix}
zI_N & Y^{(1)} \\
Y^{(2)} & B \\
\end{pmatrix},
\quad u \in \mathbb{S}^{N-1},
$$
and (3.33), we have
\[
\|u^\top E\|^2 = \|u^\top EBB^*\|^2 + |(u^\top E)_{n-2k}|^2 \\
\leq \|u^\top EBB^*\|^2 + \frac{1}{c^2} |(u^\top E)_{n-2k}|^2 \|X_{n-2k}\|^2 \\
= \|u^\top EBB^*\|^2 + \frac{1}{c^2} \|u^\top E(I_{n-k-1} - BB^*)Y^{(2)}\|^2 \\
= \|u^\top EBB^*\|^2 + \frac{1}{c^2} \|u^\top EY^{(2)} - u^\top EBB^*Y^{(2)}\|^2 \\
= \|u^\top DY^{(1)}B^*\|^2 + \frac{1}{c^2} \|u^\top - Zu^\top D + u^\top DY^{(1)}B^*Y^{(2)}\|^2 \\
\leq C(1 + \|D\|^2) \leq C\|D\|^2. \tag{3.37}
\]
on the event \(\{\|X\| \leq C_0\} \cap \mathcal{E}_{X_{n-2k}}\), where \(\mathcal{E}_{X_{n-2k}} = \{\|X_{n-2k}\| \leq c\}\) for a small constant \(c\). Note that by Lemma B.2, we have also
\[
\mathbb{P}\left(\mathcal{E}^c_{X_{n-2k}}\right) \leq \exp(-cN). \tag{3.38}
\]
By the identity
\[
\begin{pmatrix} 0 \\ u \end{pmatrix} = \begin{pmatrix} zI_N \\ Y^{(2)} \end{pmatrix} \begin{pmatrix} Eu \\ Gu \end{pmatrix}, \quad u \in \mathbb{S}^{n-k-1},
\]
and (3.33), we have
\[
\|Gu\|^2 = \|B^*BGu\|^2 + |(Gu)_k|^2 \\
\leq \|B^*BGu\|^2 + \frac{|(Gu)_k|^2}{c^2} \|X_n\|^2 \\
\leq \|B^*BGu\|^2 + \frac{1}{c^2} \|Y^{(1)}Gu - Y^{(1)}B^*BGu\|^2 \\
\leq \|B^*(I_{n-k-1} - Y^{(2)}E)u\|^2 + \frac{1}{c^2} \|-zEu - Y^{(1)}B^*(I_{n-k-1} - Y^{(2)}E)u\|^2 \\
\leq C(1 + \|E\|^2) \leq C(\|D\|^2 + \|E\|^2), \tag{3.39}
\]
on the event \(\{\|X\| \leq C_0\} \cap \mathcal{E}_X\). Take supremum over \(u\) in (3.37) to obtain \(\|E\| \leq C\|D\|\) on the event \(\{\|X\| \leq C_0\} \cap \mathcal{E}_{X_{n-2k}}\). Then take supremum over \(u\) in (3.39) to obtain \(\|G\| \leq C\|D\|\) on the event \(\{\|X\| \leq C_0\} \cap \mathcal{E}_{X_{n-2k}} \cap \mathcal{E}_X\).

Hence, on the event \(\mathcal{E}_{X_{n-2k}} \cap \mathcal{E}_X \cap \{\|X\| \leq C_0\}\), we have
\[
\text{Den}^2 = 1 + \left\| \left( X_{n-k}^* \left( e_{k}^{(n-k-1)} \right)^\top \right) \begin{pmatrix} D & E \\ F & G \end{pmatrix} \right\|^2 \\
\leq 1 + (1 + C_0^2) \left( \|D\|^2 + \|E\|^2 + \|F\|^2 + \|G\|^2 \right) \\
\leq C\|D\|^2 \\
\leq C\|D\|^2_{HS}, \tag{3.40}
\]
where we use the norm relationship \(\|D\| \leq \|D\|_{HS}\).

Therefore, by (3.11), (3.40), (3.36), (3.38) and (3.32),
\[
\mathbb{P}\left(\text{dist}(H_{[N+n-k],N+n-k}, H_{-(N+n-k)}) \leq s, \|X\| \leq C_0\right)
The proof of (3.10) for the case \( l = N + n - k \) is finished when we choose \( s = n^{-13/11} \). For the case \( N + k + 1 \leq l < N + n - k \) the proof are similar.

The following Lemma has been used previously.

**Lemma 3.1.** For any deterministic vector \( d \in \mathbb{C}^N \setminus \{0\} \),

\[
\mathbb{P} \left( \frac{Dd}{\|Dd\|} \in \text{Comp}(\theta_0, r_H), \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN).
\]

**Proof.** (of Lemma 3.1) We first show that for deterministic vector \( v \in \mathbb{S}^{N-1} \) supported on a deterministic index set \( I \) with \( |I| = \theta_0 N \), for small \( t \),

\[
\mathbb{P} \left( \inf_{w \in \mathbb{C}^{n-k-1}} \text{dist} \left( H_{[N+n-k-1],[N+n-k-1]} \left( \begin{array}{c} v \\ w \end{array} \right), \text{Span} \left\{ \left( \begin{array}{c} d \\ 0 \end{array} \right) \right\} \right) \leq t, \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN).
\]

(3.41)

Recall the definition of \( B \), we can see that \( B^*B \) is a diagonal matrix, whose \( k \)-th diagonal entry is zero and other diagonal entries are 1. Thus,

\[
Y^{(1)}w = \sum_{i=1}^{n-k-1} \left( Y^{(1)} \right)_{[N],i} w_i \\
= \left( Y^{(1)} \right)_{[N],k} w_k + \sum_{i \in [n-k-1] \setminus \{k\}} \left( Y^{(1)} \right)_{[N],i} w_i \\
= X_n w_k + Y^{(1)} B^* B w.
\]

(3.42)

Recall the blocking of \( H_{[N+n-k-1],[N+n-k-1]} \), we have

\[
\mathbb{P} \left( \inf_{w \in \mathbb{C}^{n-k-1}} \text{dist} \left( H_{[N+n-k-1],[N+n-k-1]} \left( \begin{array}{c} v \\ w \end{array} \right), \text{Span} \left\{ \left( \begin{array}{c} d \\ 0 \end{array} \right) \right\} \right) \leq t, \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN).
\]

\[
\mathbb{P} \left( \exists w \in \mathbb{C}^{n-k-1}, \exists a \in \mathbb{C} : \|zv + Y^{(1)} w + ad\| \leq t, \|Y^{(2)} v + Bw\| \leq t, \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN).
\]

(3.43)

For a large enough constant \( C'(t) \) that depends on \( t \), by Lemma B.2, we have

\[
\mathbb{P} \left( \exists w_k, a \in \mathbb{C} : \|zv + X_n w_k - Y^{(1)} B^* Y^{(2)} v + ad\| \leq t(1 + C_0), |w_k| \geq C'(t), \|\bar{X}\| \leq C_0 \right)
\]
\[ \begin{align*}
&\leq \mathbb{P}\left( \exists w_k, a \in \mathbb{C} : \|zv + X_n w_k + ad\| \leq C_0^2 + t(1 + C_0), |w_k| \geq C'(t), \|\bar{X}\| \leq C_0 \right) \\
&\leq \mathbb{P}\left( \exists w_k, a \in \mathbb{C} : \left\| \frac{z}{w_k} v + X_n + \frac{a}{w_k} d \right\| \leq \frac{C_0^2 + t(1 + C_0)}{C'(t)}, |w_k| \geq C'(t), \|\bar{X}\| \leq C_0 \right) \\
&\leq \mathbb{P}\left( \text{dist } (X_n, \text{Span } \{v, d\}) \leq \frac{C_0^2 + t(1 + C_0)}{C'(t)} \right) \\
&\leq \exp(-cN). \tag{3.44}
\end{align*} \]

We divide the interval \([-C'(t), C'(t)]\) by \(-C'(t) = a_1 < \ldots < a_{M(t)} = C'(t)\), such that \(|a_{i+1} - a_i| \leq 3C'(t)/M(t)\), where \(M(t)\) is a large enough constant. Then we have

\[ \begin{align*}
&\sum_{i=1}^{M(t)-1} \mathbb{P}\left( \exists w_k, a \in \mathbb{C} : \left\| zv + X_n w_k - Y^{(1)} B^* Y^{(2)} v + ad \right\| \leq t(1 + C_0), |w_k| < C'(t), \|\bar{X}\| \leq C_0 \right) \\
&\leq \sum_{i=1}^{M(t)-1} \mathbb{P}\left( \exists a \in \mathbb{C} : \left\| zv + X_n a_i - Y^{(1)} B^* Y^{(2)} v + ad \right\| \leq t(1 + C_0) + \frac{3C_0 C'(t)}{M(t)}, \|\bar{X}\| \leq C_0 \right) \\
&\leq \sum_{i=1}^{M(t)-1} \mathbb{P}\left( \exists a \in \mathbb{C} : \left\| zv + Y^{(1)} \left( a_i e_k - B^* Y^{(2)} v \right) + ad \right\| \leq t(1 + C_0) + \frac{3C_0 C'(t)}{M(t)} \right). \tag{3.45}
\end{align*} \]

We denote \(u_i = a_i e_k - B^* Y^{(2)} v\) then we have

\[ u_i = a_i e_k - \begin{pmatrix}
0_{(k-1)\times (n-2k-1)} & 0_{(k-1)\times 1} & I_{k-1} \\
0_{1\times (n-2k-1)} & 0_{1\times (k-1)} & 0_{1\times (n-2k-1)\times (k-1)} \\
I_{n-2k-1} & 0_{(n-2k-1)\times 1} & 0_{(n-2k-1)\times (k-1)} \\
\end{pmatrix} \begin{pmatrix}
X_1^* \\
\vdots \\
X_{n-k-1}^*
\end{pmatrix} v \]

Note that \(v\) is supported on \(I\), we can see that \(u_i\) is independent of \(X_{p_i[n]}\), which implies that \(u_i\) is independent of \(Y^{(1)}_{p_i[n-k-1]}\). Moreover, by Lemma B.2, we have

\[ \mathbb{P}\left( \|u_i\|^2 < c \right) \leq \mathbb{P}\left( \sum_{i \in [n-k-1]\setminus \{n-2k\}} (X_i^* v)^2 < c \right) \]

\[ = \mathbb{P}\left( \| (X_1, \ldots, X_{n-2k-1}, X_{n-2k+1}, \ldots, X_{n-k-1})^* v \|^2 < c \right) \]

\[ \leq \exp(-cN). \tag{3.46} \]

Thus, by (3.45), (3.46) and Lemma B.2,

\[ \mathbb{P}\left( \exists w_k, a \in \mathbb{C} : \left\| zv + X_n w_k - Y^{(1)} B^* Y^{(2)} v + ad \right\| \leq t(1 + C_0), |w_k| < C'(t), \|\bar{X}\| \leq C_0 \right) \]

\[ \leq \exp(-cN). \]
Similarly, by (3.42) and Lemma B.2, we have

\[
\mathbb{P}\left( \exists w \in \mathbb{C}^{n-k-1}, \exists r' \geq 0 : H_{[N+n-k-1],[N+n-k-1]} \begin{pmatrix} 0 \\ w \end{pmatrix} = \begin{pmatrix} r'd \\ 0 \end{pmatrix} : \left\| X \right\| \leq C_0 \right) = \mathbb{P}\left( \exists w \in \mathbb{C}^{n-k-1}, \exists r' \geq 0 : Y^{(1)}w = r'd, Bw = 0, \left\| X \right\| \leq C_0 \right) = \mathbb{P}\left( \exists w_k \in \mathbb{C}, \exists r' \geq 0 : X_nw_k = r'd, Bw = 0, \left\| X \right\| \leq C_0 \right) = \mathbb{P}\left( \exists w_k \in \mathbb{C} \setminus \{0\}, \exists r' \geq 0 : X_nw_k = \begin{pmatrix} r' \\ w_k \end{pmatrix} \right) = \mathbb{P}(X_n \in \text{Span}\{d\}) \leq \exp(-cN). \tag{3.48}
\]

By Lemma 2.5, the set of compressible vectors \( \text{Comp}(\theta_0, r_H) \) lies in a \( r_H \)-neighbourhood of \( \mathbb{S}_I^{N-1} \) for some \( I \subset [N] \) with \( |I| = \theta_0N \), and the set \( \mathbb{S}_I^{N-1} \) has a \( r_H \)-net of cardinal number bounded by \( (3/r_H)^{2\theta_0N} \). Moreover, by (3.9) and Lemma A.3, we have \( \left\| H_{[N+n-k-1],[N+n-k-1]} \right\| \leq \left\| H \right\| \leq |z| + 1 + C_0 \). Thus, by (3.41) and (3.48),

\[
\mathbb{P}\left( \frac{Dd}{\left\| Dd \right\|} \in \text{Comp}(\theta_0, r_H), \left\| X \right\| \leq C_0 \right) = \mathbb{P}\left( \exists w \in \text{Comp}(\theta_0, r_H), r \geq 0, w \in \mathbb{C}^{n-k-1} : H_{[N+n-k-1],[N+n-k-1]} \begin{pmatrix} rv \\ w \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix} : \left\| X \right\| \leq C_0 \right) = \mathbb{P}\left( \exists w \in \mathbb{C}^{n-k-1} : H_{[N+n-k-1],[N+n-k-1]} \begin{pmatrix} 0 \\ w \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix} : \left\| X \right\| \leq C_0 \right).
\]
+ \mathbb{P} \left( \exists v \in \text{Comp}(\theta_0, r_H), r > 0, w \in \mathbb{C}^{n-k-1}: H_{[N+n-k],[N+n-k-1]} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} d/r \\ 0 \end{pmatrix}, \| \tilde{X} \| \leq C_0 \right) \\
\leq \exp(-cN) + \binom{N}{\theta_0, N} \frac{3}{r_H}^{26_0N} \times \mathbb{P} \left( \exists v \in B(v_0, 2r_H), r > 0, w \in \mathbb{C}^{n-k-1}: H_{[N+n-k-1],[N+n-k-1]} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} d/r \\ 0 \end{pmatrix}, \| \tilde{X} \| \leq C_0 \right) \\
\leq \exp(-cN) + \binom{N}{\theta_0, N} \frac{3}{r_H}^{26_0N} \times \mathbb{P} \left( \exists w \in \mathbb{C}^{n-k-1}: \text{dist} \left( H_{[N+n-k-1],[N+n-k-1]} \begin{pmatrix} v_0 \\ w \end{pmatrix}, \text{Span} \left\{ \begin{pmatrix} d \\ 0 \end{pmatrix} \right\} \right) \leq 2r_H(|z| + 1 + C_0), \| \tilde{X} \| \leq C_0 \right) \\
\leq \exp(-cN) + \binom{N}{\theta_0, N} \frac{3}{r_H}^{26_0N} \exp(-cN) \\
\leq \exp(-cN).

In the last inequality, we just use the Stirling formula as what we do at the end of Step 1. We may replace the \( \theta_0 \) by a smaller one if necessary. Hence, the proof of Lemma 3.1 is finished. \( \square \)

**Remark 3.1.** By a similar argument, one can show that

\[
\mathbb{P} \left( \frac{d^*D}{\|d^*D\|} \in \text{Comp}(\theta_0, r_H), \| \tilde{X} \| \leq C_0 \right) \leq \exp(-cN).
\]

**3.2.2 Case of \( N + 1 \leq l \leq N + k \)**

The estimation is similar to the previous case of \( N + k + 1 \leq l \leq N + n - k \) and the proof is sketched as follows. Without loss of generality, we only estimate (3.11) for the case \( l = N + 1 \). First of all, we have

\[
H_{N+1,N+1} = 0, \ H_{N+1,[N+n-k]\{N+1\}} = \left( X_1^*, \left( e_k^{(n-k-1)} \right)^T \right), \\
H_{[N+n-k]\{N+1\},N+1} = \begin{pmatrix} X_{n-k+1} \\ e_k^{(n-k-1)} \end{pmatrix}, \\
H_{[N+n-k]\{N+1\},[N+n-k]\{N+1\}} = \begin{pmatrix} zI_N \\ (X_2, \ldots, X_{n-k})^{*} \end{pmatrix} \begin{pmatrix} \left( e_k^{(n-k-1)} \right)^T, \ldots, e_k^{(n-k-1)} \end{pmatrix}.
\]

By showing that the determinant of \( H_{[N+n-k]\{N+1\},[N+n-k]\{N+1\}} \) is a non-zero polynomial of the entries of \( X \), one can deduce the invertibility of \( H_{[N+n-k]\{N+1\},[N+n-k]\{N+1\}} \). Next, we denote

\[
\left( H_{[N+n-k]\{N+1\},[N+n-k]\{N+1\}} \right)^{-1} = \begin{pmatrix} D \\ E \\ F \\ G \end{pmatrix},
\]

where \( D \in \mathbb{C}^{N \times N} \) and \( G \in \mathbb{C}^{(n-k-1) \times (n-k-1)} \). Denote a random vector \( Y = (X_1^*, X_{n-k+1}^*)^* \), then by (3.12),

\[
\text{Num} = \left| X_1^*DX_{n-k+1} + X_1^*e_k^{(n-k+1)} + \left( e_k^{(n-k-1)} \right)^T FX_{n-k+1} + \left( e_k^{(n-k-1)} \right)^T Ge_k^{(n-k-1)} \right|
\]
Similar to Step (a) in Section 3.2.1, we introduce an independent family for some deterministic vector $\tilde{x}$:

$$
\begin{align*}
&\left| Y^* \begin{pmatrix} 0_{N \times N} & D \\ 0_{N \times N} & 0_{N \times N} \end{pmatrix} Y + Y^* \begin{pmatrix} E_{n-2k} & (n-k+1) \end{pmatrix}^T F \right| \\
&= \left| Y^* \bar{D} Y + Y^* \tilde{b}^{(1)} + \left( \tilde{b}^{(2)} \right)^* Y + \tilde{a} \right|
\end{align*}
$$

where

$$
\bar{D} = \begin{pmatrix} 0_{N \times N} & D \\ 0_{N \times N} & 0_{N \times N} \end{pmatrix}, \quad \tilde{b}^{(1)} = \begin{pmatrix} E_{n-2k} \\ 0_{N \times 1} \end{pmatrix},
$$

$$
\left( \tilde{b}^{(2)} \right)^* = \begin{pmatrix} 0_{1 \times N} & (n-k-1) \end{pmatrix}^T F, \quad \tilde{a} = \begin{pmatrix} (n-k+1) \end{pmatrix}^T G_{n-2k}.
$$

Similar to Step (a) in Section 3.2.1, we introduce an independent family $\xi = \{\xi_1, \ldots, \xi_N\}$ of Bernoulli random variables and denote $I = \{i \in [N] : \xi_i = 1\}$, $\tilde{I} = I \cup (I + N) = I \cup \{i \in [2N] \setminus [N] : i - N \in I\}$. Choose independent random vectors $\bar{x}, \bar{x'}, \bar{x''} \overset{d}{=} Y$ and set $\bar{u} = (\bar{x})_I$, $\bar{v} = (\bar{x'})_I$, and $\bar{w} = (\bar{x''})_I$. Denote $\bar{X} = (X_2, \ldots, X_{n-k}, X_{n-k+2}, \ldots, X_n)$. Then by Cauchy-Schwarz inequality and Lemma 2.8,

$$
P(\text{Num} \leq t, \|X\| \leq C_0)^2 \leq P \left( \|Y^* \bar{D} Y + Y^* \tilde{b}^{(1)} + \left( \tilde{b}^{(2)} \right)^* Y + \tilde{a} \| \leq t, \|\bar{X}\| \leq C_0 \right)^2
$$

$$
\leq E_{\tilde{X}} \left[ E_Y \left[ 1_{\|Y^* \bar{D} Y + Y^* \tilde{b}^{(1)} + \left( \tilde{b}^{(2)} \right)^* Y + \tilde{a} \| \leq t} \right] 1_{\|\bar{X}\| \leq C_0} \right]^2
$$

$$
\leq E_{\tilde{X}} \left[ E_Y \left[ 1_{\|Y^* \bar{D} Y + Y^* \tilde{b}^{(1)} + \left( \tilde{b}^{(2)} \right)^* Y + \tilde{a} \| \leq t} \right] 1_{\|\bar{X}\| \leq C_0} \right]^2
$$

$$
\leq E_{\tilde{X}} \left[ E_{\bar{x}, \bar{v}, \bar{w}} \left[ S_{\tilde{a}} \left( (\bar{v} - \bar{w})^* \bar{D} \bar{v} + \bar{w}^* \bar{D} \bar{w} + 2t \right) \right] 1_{\|\bar{X}\| \leq C_0} \right]
$$

$$
= E_{\tilde{X}} \left[ E_{\bar{x}, \bar{v}, \bar{w}} \left[ S_{\tilde{a}} \left( (\bar{y} \bar{z} - \bar{z}'')^* \bar{D} \bar{y} + \bar{y}' \bar{D} \bar{y} + 2t \right) \right] 1_{\|\bar{X}\| \leq C_0} \right].
$$

Next, we need the following lemma, which is analogous to Lemma 3.1.

**Lemma 3.2.** For any deterministic vector $d \in \mathbb{C}^{2N}$ such that $\Pi_{[2N] \setminus [N]} d \neq 0$,

$$
P \left( \frac{\bar{D} d}{\|\bar{D} d\|} \in \text{Comp}(\theta_0, r_H), \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN).
$$

**Proof.** Replacing $\theta_0$ by a smaller one if necessary, one can follow the proof of Lemma 3.1 to show that for small $\theta_0$,

$$
P \left( \frac{\bar{D} \tilde{d}}{\|\bar{D} \tilde{d}\|} \in \text{Comp}(\theta_0/2, r_H), \|\bar{X}\| \leq C_0 \right) \leq \exp(-cN),
$$

for some deterministic vector $\tilde{d} \in \mathbb{C}^N \setminus \{0\}$.

Now for any deterministic vector $d \in \mathbb{C}^{2N}$, we write $d = (d_1^*, d_2^*) \in \mathbb{C}^{2N}$, where $d_1, d_2 \in \mathbb{C}^N$. Note that

$$
\bar{D} d = \begin{pmatrix} D d_2 \\ 0_{N \times 1} \end{pmatrix}.
$$
By the definition of compressible vectors, we can see that
\[
\left\{ \frac{\tilde{D}d}{\| \tilde{D}d \|} \in \text{Comp}(\theta_0/2, r_H) \right\} \subseteq \left\{ \frac{Dd_2}{\| Dd_2 \|} \in \text{Comp}(\theta_0, r_H) \right\}.
\]
Indeed, on the event \{ \tilde{D}d/\| \tilde{D}d \| \in \text{Comp}(\theta_0/2, r_H) \}, one can find a unit vector \( v \) and \( J \in [2N] \) with \( |J| = \theta_0 N \), such that \( v_{[2N]\setminus J} = 0 \) and \( \tilde{D}d/\| \tilde{D}d \| - v \leq r_H \). If \( J \cap [N] = \emptyset \), then \( v_{[N]} = 0 \), which leads to
\[
\sqrt{2} = \sqrt{1 + \| v \|^2} = \frac{\| \tilde{D}d \| - v}{\| \tilde{D}d \|} \leq r_H.
\]
This is a contradiction since \( r_H \) is chosen to be small enough. Therefore, by triangle inequality, we have
\[
\left\| \frac{Dd_2}{\| Dd_2 \|} - \frac{v_{[N]}}{\| v_{[N]} \|} \right\| \leq \frac{Dd_2}{\| Dd_2 \|} - v_{[N]} + 1 - \| v_{[N]} \|
\]
\[
= \sqrt{\| \tilde{D}d \| - v \|^2 - \| v_{[2N]\setminus[N]} \|^2 + 1 - \| v_{[N]} \|^2}
\]
\[
\leq \sqrt{\frac{r_H^2 - \| v_{[2N]\setminus[N]} \|^2 + 1 - \| v_{[N]} \|^2}{r_H^2 - 1 + \| v_{[N]} \|^2 + 1 - \| v_{[N]} \|^2}}.
\]
The right-hand side is increasing with respect to \( \| v_{[N]} \| \), and obtains the maximum value \( r_H \) when \( \| v_{[N]} \| = 1 \). Hence, we complete the proof. \( \square \)

We can define \( y, \tilde{y}, \alpha, \tilde{\alpha}, W_i \) as we do in Section 3.2.1, where \( D, I, x' \) and \( x'' \) should be replaced by \( \tilde{D}, \tilde{I}, \tilde{x}' \) and \( \tilde{x}'' \), respectively. Let \( J \) be the index set given by (3.18) with \( N \) replaced by \( 2N \). Then we can obtain the corresponding upper bound (3.21) with \( I, J \) and \( D \) replaced by \( \tilde{I}, \tilde{J} \) and \( \tilde{D} \), respectively. If \( I^R \neq \emptyset \), then \( \Pi_{p^c+N}(\tilde{x}' - \tilde{x}'') = 0 \) with probability zero since the entries of \( Y \) have continuous density. Then by the Lemma 3.2, one can still obtain that the probability of \( y \in \text{Incomp}(\theta_0, r_H) \) is at most \( \exp(-cN) \). Besides,
\[
\mathbb{P}\left( I^R = \emptyset \right) = p^N = \exp(-cN).
\]
Thus, the probability of \( y \in \text{Incomp}(\theta_0, r_H) \) is at least \( 1 - \exp(-cN) \). On the event \( \{ y \in \text{Incomp}(\theta_0, r_H) \} \cap E_I \), we still have \( |I \cap J| \geq \theta_0 N/3 \) for a small constant \( c \) that only depends on \( \theta_0 \). Then the computation of (3.31) is still valid with \( D \) replaced by \( \tilde{D} \), the unit vector \( e_i^{(N)} \) replaced by \( e_i^{(2N)} \) and the range of the index \( j \) should be \( [2N] \). Thus, one may derive (3.32).

The computation of Den given by (3.13) as well as the estimation
\[
\mathbb{P}\left( \frac{\text{Num}}{\text{Den}} \leq s, |X| \leq C_0 \right) \leq C(\theta_0, r_H)\sqrt{n^{12/11}s + n^{-1/11}} + \exp(-cN)
\]
can be deduced step by step as in Section 3.2.1. By choosing \( s = n^{-13/11} \), we obtain (3.10) for the case \( l = N + 1 \).
3.2.3. Case of $1 \leq l \leq N$

The estimation is similar to the previous case of $N + k + 1 \leq l \leq N + n - k$ and the proof is sketched as follows. Without loss of generality, we only estimate (3.11) for the case $l = N$. First of all, we have

$$H_{N,N} = z, \quad H_{N,[N+n-k]\{N\}} = \left(0_{(N-1)\times 1}, (X_{N,[n]\{n-k\}, X_{N,[n-k]\{k\}})\right),$$

$$H_{[N+n-k]\{N\},N} = \left(0_{(N-1)\times 1}, (X_{N,[n-k]\{k\}})^*\right),$$

$$H_{[N+n-k]\{N\},[N+n-k]\{N\}} = \left((X_{N-1},[n-k])^*, (X_{N-1},[n-k])^\dagger, (X_{n-k}^\dagger, e_1, \ldots, e_{n-k}^\dagger)\right).$$

By showing that the determinant of $H_{[N+n-k]\{N\},[N+n-k]\{N\}}$ is a non-zero polynomial of the entries of $X$, one can deduce the invertibility of $H_{[N+n-k]\{N\},[N+n-k]\{N\}}$. Next, we denote

$$(H_{[N+n-k]\{N\},[N+n-k]\{N\}})^{-1} = \begin{pmatrix} D & E \\ F & G \end{pmatrix},$$

where $D \in \mathbb{C}^{(N-1)\times (N-1)}$ and $G \in \mathbb{C}^{(n-k)\times (n-k)}$. Denote a row random vector

$$Y = (Y_1, Y_2, Y_3) = (X_{N,[k]}, X_{N,[n]\{n-k\}, X_{N,[n-k]\{k\}}),$$

then by (3.12),

$$\text{Num} = \left| z - (0_{1\times (N-1)}, Y_2, Y_3) \begin{pmatrix} D & E \\ F & G \end{pmatrix} (0_{1\times (N-1)}, Y_1, Y_3)^* \right|$$

$$= \left| z - (Y_2, Y_3) G (Y_1, Y_3)^* \right|$$

$$= \left| z - Y \begin{pmatrix} 0_{k \times k} & 0_{k \times k} & 0_{k \times (n-2k)} \\ G_{[k],[k]} & 0_{k \times k} & G_{[k],[n-2k]} \\ 0_{(n-2k) \times k} & G_{[n-2k],[n-2k]} & 0_{(n-2k) \times k} \end{pmatrix} Y^* \right|.$$

We denote the matrix above as $\tilde{G}$ then $\text{Num} = \left| z - Y \tilde{G} Y^* \right|$. Similar to Step (a) in Section 3.2.1, we introduce an independent family $\xi = \{\xi_1, \ldots, \xi_n\}$ of Bernoulli random variables and denote $I = \{i \in [n]: \xi_i = 1\}$. Choose independent random vectors $x, x', x'' \equiv Y$ and set $u = (x)_I, v = (x')_I$ and $w = (x'')_I$. Denote $\tilde{X} = X_{[N-1],[n]}$. Then by Cauchy-Schwarz inequality and Lemma 2.8,

$$\mathbb{P}(\text{Num} \leq t, \|X\| \leq C_0)^2$$

$$\leq \mathbb{P}\left(\left| z - Y \tilde{G} Y^* \right| \leq t, \|\tilde{X}\| \leq C_0\right)^2$$

$$= \left(\mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{X_{N,[n]}} \left[ 1_{\|z - Y \tilde{G} Y^*\| \leq t} \right] \|\tilde{X}\| \leq C_0 \right] \right)^2$$

$$\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_Y \left[ 1_{\|z - Y \tilde{G} Y^*\| \leq t} \right] \|\tilde{X}\| \leq C_0 \right]$$

$$\leq \mathbb{E}_{\tilde{X}} \left[ \mathbb{E}_{v,w} \left[ S_u \left( (v - w)^* \tilde{G}_I u + u^* \tilde{G}_I^* (v - w), 2t \right) \right] \right] \|\tilde{X}\| \leq C_0 \right]$$
Then we can define $y, \tilde{y}, \alpha, \tilde{\alpha}, W_i$ as we do in Section 3.2.1, where $D$ should be replaced by $\tilde{G}$. One can compute the corresponding upper bound (3.21), where $J$ should be given in (3.18) with $N$ replaced by $n$.

Besides, we have the following lemma that is analogous to Lemma 3.1.

**Lemma 3.3.** For any deterministic vector $d \in \mathbb{C}^n$ such that $\Pi_{[k] \cup ([n] \setminus [2k])} d \neq 0$,

$$\mathbb{P} \left( \frac{\tilde{G}d}{\| \tilde{G}d \|} \in \text{Comp}(\theta_0, r_H), \| \tilde{X} \| \leq C_0 \right) \leq \exp(-cN).$$

**Proof.** Again one can follow the proof of Lemma 3.1 to show that for small $\theta'_0$,

$$\mathbb{P} \left( \frac{Gd}{\| Gd \|} \in \text{Comp}(\theta'_0, r_H), \| \tilde{X} \| \leq C_0 \right) \leq \exp(-cN),$$

for some deterministic vector $\tilde{d} \in \mathbb{C}^{n-k} \setminus \{0\}$.

Now for any deterministic vector $d \in \mathbb{C}^n$, we write $d = (d_1^*, d_2^*, d_3^*)^* \in \mathbb{C}^n$, where $d_1, d_2 \in \mathbb{C}^k$ and $d_3 \in \mathbb{C}^{n-2k}$. Note that

$$\tilde{G}d = \begin{pmatrix} 0_{k \times 1} \\ d_1 \\ d_3 \end{pmatrix}. $$

By the definition of compressible vectors, we can see that

$$\left\{ \tilde{G}d \in \text{Comp}(\theta_0, r_H) \right\} \subseteq \left\{ G(d_1^*, d_3^*)^* \in \text{Comp}(\theta'_0, r_H) \right\},$$

if $\theta_0 n \leq \theta'_0 (n-k)$. Recalling the ratio (1.3), we may replace $\theta_0$ by a smaller one if necessary. The proof is complete. $\square$

On the event $\{ I^c \not\subseteq [2k] \setminus [k] \}$, $\Pi_{[k] \cup ([n] \setminus [2k])}(x' - x'') = 0$ with probability zero since the entries of $Y$ have continuous density. Then by the Lemma 3.3, one can still obtain that the probability of $y \in \text{Comp}(\theta_0, r_H)$ is at most $\exp(-cN)$. Besides,

$$\mathbb{P} \left( r^c \subseteq [2k] \setminus [k] \right) = \mathbb{P}(\xi_i = 1, \forall i \in [k] \cup ([n] \setminus [2k])) = p^{n-k} = \exp(-cN).$$

Thus, the probability of $y \in \text{Incomp}(\theta_0, r_H)$ is at least $1 - \exp(-cN)$. Similarly, one can show that the probability of the event $\mathcal{E}'_j = \{|I| > n(1-\theta_0/3)\}$ is at least $1 - \exp(-cN)$. On the event $\{ y \in \text{Incomp}(\theta_0, r_H) \} \cap \mathcal{E}'_j$, we still have $|I \cap J| \geq \theta_0 n/6$.

Note that

$$\left\{ e_i^{(n)} \right\}^* \tilde{G} = \begin{cases} 0, & i \in [2k] \setminus [k], \\
(e_i^{(n-k)})^* G, & i \in [k] \cup ([n] \setminus [2k]).
\end{cases} $$
Thus, (3.10) follows from the estimations on Num and Den.

Without loss of generality, we only estimate (3.11) for the case \(N \in \mathbb{N}\). Then the computation of (3.31) is still valid with \(D\) replaced by \(\tilde{G}\) and the index \(i\) should be in the index set \([k] \cup ([n] \setminus [2k])\). Thus, one may derive (3.32) with \(D\) replaced by \(G\).

Next, we compute Den given by (3.13). Recall the definition of \(H\) in (3.14) and the blocking of \((H_{[N+n-k]}[N],[N+n-k])^{-1}\), we have the following identity

\[
I_{N+n-k-1} = \begin{pmatrix}
H_{[N-n-k]}[N],[N-1] & H_{[N-n-k]}[N],[N+n-k] \\
D & F
\end{pmatrix} \begin{pmatrix}
zI_{N-1} & H_{[N-n-k]}[N],[N] \\
F & G
\end{pmatrix}.
\]

Note that the submatrix \(H_{[N-n-k]}[N],[N+n-k][N]\) is a permutation matrix, \(H_{[N-n-k]}[N],[N+n-k][N]\) is a submatrix of \(X\) up to a permutation, \(H_{[N-n-k]}[N],[N-1]\) is a submatrix of \(X^*\). Thus, on the event \(\{\|X\| \leq C_0\}\), we have

\[
\|F\| = \frac{1}{|z|} \|FzI_{N-1}\| = \frac{1}{|z|} \|GH_{[N-n-k]}[N],[N-1]\| \leq C_0 \|G\|.
\]

Moreover, on the event \(\{\|X\| \leq C_0\}\), suppose that \(\|G\| \leq c'\) for a small constant \(c'\) that may depend on \(|z|\) and will be determined later, then \(\|F\| \leq C_0 c'/|z|\). Thus,

\[
1 = \|I_{n-k}\| = \|FH_{[N-n-k]}[N]+GH_{[N-n-k]}[N],[N+n-k][N]\| \leq C_0 \|F\| + \|G\| \leq \frac{C_0^2 c'}{|z|} + c'.
\]

We may choose \(c' < |z|/ (|z| + C_0^2)\) to reach a contradiction. Thus, \(\|G\| > c'\) on the event \(\{\|X\| \leq C_0\}\).

Therefore, on the event \(\{\|X\| \leq C_0\}\),

\[
\text{Den}^2 = 1 + \| (X_{[N],[n-k]}, X_{[N],[n-k]}^*) F \|^2 + \| (X_{[N],[n-k]}, X_{[N],[n-k]}^*) G \|^2
\]

\[
\leq 1 + C_0^2 \|F\|^2 + C_0^2 \|G\|^2
\]

\[
\leq C \|G\|^2.
\]

Thus, (3.10) follows from the estimations on Num and Den.

### 3.3. Estimate (3.5) for incompressible vectors for the case \(2k + 1 > n\)

We now establish estimation (3.5) for the case \(2k + 1 > n\) with \(\theta = \theta_0\) and \(\rho = r_H\). The proof is similar to that in Section 3.2 and is sketched as follows.

It is enough to prove (3.10). Moreover, (3.11), (3.12) and (3.13) are still valid.

#### 3.3.1. Case of \(N + 1 \leq l \leq N + n - k\)

Without loss of generality, we only estimate (3.11) for the case \(l = N + n - k\). Recall the definition of \(H\), we have

\[
H_{N+n-k,N+n-k} = 1, \quad H_{N+n-k,[N+n-k-1]} = (X_{n-k}^*, 0_{1 \times (n-k-1)}).
\]
\[
H_{[N+n-k-1],N+n-k} = \begin{pmatrix}
X_n \\
0_{(n-k-1)\times 1}
\end{pmatrix},
\]

\[
H_{[N+n-k-1],[N+n-k-1]} = \begin{pmatrix}
zI_N \\
(X_1, \ldots, X_{n-k})^* \\
I_{n-k-1}
\end{pmatrix}.
\]

One can follow the argument in Section 3.2.2 to obtain the existence of

\[
(H_{[N+n-k-1],[N+n-k-1]})^{-1} = \begin{pmatrix} D & E \\ F & G \end{pmatrix},
\]

and the estimation (3.32) of Num. For the estimation of Den, one can follow the argument in Section 3.2.3 to obtain

\[
\text{Den}^2 \leq C\|D\|^2,
\]
on the event \(\{\|X\| \leq C_0\}\). The estimation (3.10) follows from the estimations on Num and Den.

### 3.3.2. Case of \(1 \leq l \leq N\)

Without loss of generality, we only estimate (3.11) for the case \(l = N\). Recall the definition of \(H\), we have

\[
H_{N,N} = z, \quad H_{N,[N+n-k]\{N\}} = \begin{pmatrix}
0_{1\times(N-1)}, X_{N,[n]\{k\}} \\
X_{N,[n]\{k\}}^*
\end{pmatrix}, \quad H_{[N+n-k-1],[N+n-k-1]} = \begin{pmatrix}
zI_{N-1} \\
(X_{[N-1],[n-k]}^*) \\
I_{n-k}
\end{pmatrix}.
\]

One can follow the argument in Section 3.2.2 to obtain the existence of

\[
(H_{[N+n-k-1],[N+n-k-1]\{N\}})^{-1} = \begin{pmatrix} D & E \\ F & G \end{pmatrix},
\]

and the estimation (3.32) of Num with \(D\) replaced by \(G\). For the estimation of Den, one can follow the argument in Section 3.2.3 to obtain

\[
\text{Den}^2 \leq C\|G\|^2,
\]
on the event \(\{\|X\| \leq C_0\}\). The estimation (3.10) follows from the estimations on Num and Den.

### 4. Limiting eigenvalue empirical distribution

Though the small rank perturbation for Hermitian matrices fails in general, with the estimation on least singular value in Section 3, it turns out that the limit of \(\mu_{Y(n)}\) when \(k/n = o(\ln^{-1} n)\) is the same as the limit of \(\mu_{Z(n)}\). The detail argument is developed in Section 4.1. In addition, we establish the limit of \(\mu_{Y(n)}\) when \(k \geq n/2\) in Section 4.2. We also provide in Section 4.3 a discussion on the case \(\gamma_1 \in (0, 1/2)\).
4.1. The case $k/n = o(\ln^{-1} n)$

Let

$$g(x) = \frac{x(1 - \gamma_0 + 2x)^2}{1 + x}, \quad x \in [0 \lor (\gamma_0 - 1), \gamma_0].$$

then Bose and Hachem [2020] showed that $g$ is increasing and invertible on its domain. Moreover, under the conditions (C1) and (C2), Bose and Hachem [2020] also showed that $\mu_{Z(n)}$ of the matrix in (1.4) converges to a deterministic rotation invariant probability measure $\mu^{(\gamma_0)}$ in probability. Moreover, the distribution function of the radial component of $\mu^{(\gamma_0)}$ is

$$\mu^{(\gamma_0)}(B(0, r)) = \begin{cases} \gamma_0^{-1} g^{-1}(r^2), & 0 \leq r \leq \gamma_0^{1/2}(\gamma_0 + 1)^{1/2}, \\ 1, & r > \gamma_0^{1/2}(\gamma_0 + 1)^{1/2}, \end{cases}$$

if $\gamma_0 \leq 1$, and

$$\mu^{(\gamma_0)}(B(0, r)) = \begin{cases} 1 - \gamma_0^{-1}, & 0 \leq r \leq (\gamma_0 - 1)^{3/2}\gamma_0^{-1/2}, \\ \gamma_0^{-1} g^{-1}(r^2), & (\gamma_0 - 1)^{3/2}\gamma_0^{-1/2} \leq r \leq \gamma_0^{1/2}(\gamma_0 + 1)^{1/2}, \\ 1, & r > \gamma_0^{1/2}(\gamma_0 + 1)^{1/2}, \end{cases}$$

if $\gamma_0 > 1$. Recall the matrix of interest $Y^{(n)}$ in (1.1).

**Theorem 4.1.** Let $k = 1$ and let $N$ satisfy (1.3). Assume that the conditions (C1) and (C2) hold. Then $\mu_{Y^{(n)}}$ converges weakly to $\mu^{(\gamma_0)}$ in probability.

**Proof.** Note that $A^{(n)} - J^{(n)}$ is a rank one matrix, so is $Y^{(n)} - Z^{(n)}$. Thus, by Lemma A.2, we have

$$s_i\left(Y^{(n)} - zI_N\right) \geq s_{i+1}\left(Z^{(n)} - zI_N\right), \quad \forall i \in [N - 1].$$

Thus, for small $\delta \in (0, 1)$, on the event $\{s_N\left(Y^{(n)} - zI_N\right) < \delta\}$,

$$\left|\int_{0}^{\delta} \ln(\lambda) d\nu_{Y^{(n)} - zI_N}(\lambda)\right| \leq \frac{1}{N} \sum_{s_i(Y^{(n)} - zI_N) < \delta, i < N} \left|\ln\left(s_i\left(Y^{(n)} - zI_N\right)\right)\right|$$

$$= \frac{1}{N} \left|\ln\left(s_N\left(Y^{(n)} - zI_N\right)\right)\right| + \frac{1}{N} \sum_{s_i(Y^{(n)} - zI_N) < \delta, i < N} \left|\ln\left(s_i\left(Y^{(n)} - zI_N\right)\right)\right|$$

$$\leq \frac{1}{N} \left|\ln\left(s_N\left(Y^{(n)} - zI_N\right)\right)\right| + \frac{1}{N} \sum_{s_i(Y^{(n)} - zI_N) < \delta, i < N} \left|\ln\left(s_{i+1}\left(Z^{(n)} - zI_N\right)\right)\right|$$

$$\leq \frac{1}{N} \left|\ln\left(s_N\left(Y^{(n)} - zI_N\right)\right)\right| + \frac{1}{N} \sum_{s_i(Z^{(n)} - zI_N) < \delta} \left|\ln\left(s_i\left(Z^{(n)} - zI_N\right)\right)\right|$$
\[
\frac{1}{N} \ln \left( \sigma_N \left( Y^{(n)} - zI_N \right) \right) + \left| \int_0^\delta \ln(\lambda) d\nu_{(Z^{(n)}-zI_N)}(\lambda) \right|.
\]

Thus, for any \( \epsilon > 0 \), by Theorem 3.1, we have
\[
P \left( \left| \int_0^\delta \ln(\lambda) d\nu_{(Y^{(n)}-zI_N)}(\lambda) \right| > \epsilon \right)
\]
\[
= P \left( \left| \int_0^\delta \ln(\lambda) d\nu_{(Y^{(n)}-zI_N)}(\lambda) \right| > \epsilon, \sigma_N \left( Y^{(n)} - zI_N \right) < \delta \right)
\]
\[
\leq P \left( \frac{1}{N} \left| \ln \left( \sigma_N \left( Y^{(n)} - zI_N \right) \right) \right| + \left| \int_0^\delta \ln(\lambda) d\nu_{(Z^{(n)}-zI_N)}(\lambda) \right| > \epsilon, \sigma_N \left( Y^{(n)} - zI_N \right) < \delta \right)
\]
\[
\leq P \left( \frac{1}{N} \left| \ln \left( \sigma_N \left( Y^{(n)} - zI_N \right) \right) \right| + \left| \int_0^\delta \ln(\lambda) d\nu_{(Z^{(n)}-zI_N)}(\lambda) \right| > \epsilon, n^{-13/11} < \sigma_N \left( Y^{(n)} - zI_N \right) < \delta \right) + Cn^{-1/22} + P \left( \|X\| > C_0 \right)
\]
\[
\leq P \left( \frac{13 \ln n}{11N} + \left| \int_0^\delta \ln(\lambda) d\nu_{(Z^{(n)}-zI_N)}(\lambda) \right| > \epsilon \right) + Cn^{-1/22} + P \left( \|X\| > C_0 \right).
\]

In [Bose and Hachem, 2020, (35)], for all \( z \in \mathbb{C} \setminus \{0\} \), for all \( \epsilon > 0 \),
\[
\lim_{\delta \to 0^+} \lim_{n \to \infty} P \left( \left| \int_0^\delta \ln(\lambda) d\nu_{(Z^{(n)}-zI_N)}(\lambda) \right| > \epsilon \right) = 0.
\]

Thus, when choosing \( C_0 \) large, we have
\[
\lim_{\delta \to 0^+} \lim_{n \to \infty} P \left( \left| \int_0^\delta \ln(\lambda) d\nu_{(Y^{(n)}-zI_N)}(\lambda) \right| > \epsilon \right) = 0,
\]
where we obtain the uniform integrability of the logarithm with respect to the \( \nu_{(Y^{(n)}-zI_N)} \) in probability, for all \( z \in \mathbb{C} \setminus \{0\} \).

Next, we consider the Hermitian matrix
\[
\Sigma_n(z) = \begin{pmatrix} 0 & Y^{(n)} - zI_N \\ (Y^{(n)} - zI_N)^* & 0 \end{pmatrix}.
\]

and
\[
\Sigma_J(z) = \begin{pmatrix} 0 & Z^{(n)} - zI_N \\ (Z^{(n)} - zI_N)^* & 0 \end{pmatrix}.
\]

Then by Bose and Hachem [2020], there exists a probability measure \( \nu_{z} \), such that \( \nu_{(Z^{(n)}-zI_N)} \) converges weakly to \( \nu_{z} \) almost surely. Note that the set of eigenvalues of \( \Sigma_J \) is
\[
\{ \lambda_i(\Sigma_J) : i \in [2N] \} = \left\{ \pm s_i \left( Z^{(n)} - zI_N \right) : i \in [N] \right\},
\]
we have
\[
\mu_{\Sigma_J}(x) = \frac{\nu_{(Z^{(n)}-zI_N)}(x) + \nu_{(Z^{(n)}-zI_N)}(-x)}{2}.
\]
Thus, if we denote by $\tilde{\nu}_z$ the symmetrization of $\nu_z$, which is the probability measure defined by $\tilde{\nu}_z(E) = \nu_z(E) + \nu_z(-E)$ for all Borel set $E$, then $\mu_{\Sigma_J(z)}$ converges weakly to $\tilde{\nu}_z$ almost surely for almost all $z \in \mathbb{C}$. Moreover, since $A^{(n)} - J^{(n)}$ has rank one, $\Sigma_A(z) - \Sigma_J(z)$ is a rank two matrix. By the Stability of ESD laws with respect to small rank perturbations ([Tao, 2012, Exercise 2.4.4]), we can deduce the weakly convergence of $\mu_{\Sigma_A(z)}$ towards $\tilde{\nu}_z$ almost surely for almost all $z \in \mathbb{C}$. Hence, we obtain the weakly convergence of $\nu_{(Y^{(n)} - zI_N)}$ towards $\nu_z$.

Therefore, by Lemma 2.10, $\mu_{Y^{(n)}}$ converges weakly to some probability measure $\mu'$ in probability for almost all $z \in \mathbb{C}$, and the limit measure $\mu'$ is satisfies

$$\mathcal{L}_{\mu'}(z) = -\int_0^\infty \ln(\lambda) d\nu_z(\lambda).$$

Since the singular value empirical distributions of $Y^{(n)} - zI_N$ and $Z^{(n)} - zI_N$ converge to the same limit $\nu_z$, we have $\mathcal{L}_{\mu'}(z) = \mathcal{L}_{\mu^{(\gamma_0)}}(z)$ for almost all $z \in \mathbb{C}$. Then by Lemma 2.9, we have $\mu' = \mu^{(\gamma_0)}$.

**Theorem 4.2.** Let $N, k$ satisfy (1.3) such that $k/n = o(\ln^{-1} n)$. Assume that the conditions (C1) and (C2) hold. then $\mu_{Y^{(n)}}$ converges weakly to $\mu^{(\gamma_0)}$ in probability.

**Proof.** We consider the case $k = 2$ first. The proof is similar to the proof of Theorem 4.1, which is sketched below. Recall that

$$Y^{(n)} = \sum_{i=1}^{n-2} X_{i+2} X_i^*. $$

Let

$$\tilde{Y}^{(n)} = Y^{(n)} + X_2X_{n_0}^* = (X_1, X_3, \ldots, X_{n_0}, X_2, X_4, \ldots, X_{n_e}) \begin{pmatrix} 0_{1 \times (n-1)} & 0 \\ I_{n-1} & 0_{(n-1) \times 1} \end{pmatrix} \begin{pmatrix} X_1^* \\ \vdots \\ X_{n_0}^* \\ X_2^* \\ \vdots \\ X_{n_e}^* \end{pmatrix},$$

where $n_0 = 2[(n-1)/2] + 1$ is the largest odd number that does not exceed $n$, and $n_e = 2[n/2]$ is the largest even number that does not exceed $n$.

Since $(Y^{(n)} - zI_N) - (\tilde{Y}^{(n)} - zI_N) = X_2X_{n_0}^*$ is a rank one matrix, then by Lemma A.2 and the argument in the beginning of the proof of Theorem 4.1, we can obtain

$$\left| \int_0^\delta \ln(\lambda) d\nu_{Y^{(n)} - zI_N}(\lambda) \right| \leq \frac{1}{N} \ln \left| s_N \left( Y^{(n)} - zI_N \right) \right| + \left| \int_0^\delta \ln(\lambda) d\nu_{\tilde{Y}^{(n)} - zI_N}(\lambda) \right|.$$ 

Note that $(X_1, X_3, \ldots, X_{n_0}, X_2, X_4, \ldots, X_{n_e}) \overset{d}{=} X$, the logarithm function is uniform integrable near zero with respect to $\tilde{Y}^{(n)} - zI_N$. Hence, by Theorem 3.1, one can obtain the uniform integrability of the logarithm function with respect to $Y^{(n)} - zI_N$ in probability for all $z \in \mathbb{C} \setminus \{0\}$ by using a similar argument to the proof of Theorem 4.1.
Next, we denote
\[
\Sigma_{Y(n)}(z) = \begin{pmatrix} 0 & Y(n) - zI_N \\ (Y(n) - zI_N)^* & 0 \end{pmatrix}, \quad \Sigma_{\tilde{Y}(n)}(z) = \begin{pmatrix} 0 & \tilde{Y}(n) - zI_N \\ (\tilde{Y}(n) - zI_N)^* & 0 \end{pmatrix}.
\]

Then by the proof of Theorem 4.1, \(\mu_{\Sigma_{Y(n)}(z)}\) converges weakly to \(\tilde{\nu}_z\) almost surely for almost all \(z \in \mathbb{C}\). Since \(\Sigma_{Y(n)}(z) - \Sigma_{\tilde{Y}(n)}(z)\) is a rank two matrix, by the Stability of ESD laws with respect to small rank perturbations ([Tao, 2012, Exercise 2.4.4]), we can deduce the weak convergence of \(\nu\) towards \(\tilde{\nu}_z\) almost surely for almost all \(z \in \mathbb{C}\). Then we obtain the weakly convergence of \(\nu_{Y(n)} - zI_N\) towards \(\tilde{\nu}_z\) almost surely for almost all \(z \in \mathbb{C}\).

Therefore, by Lemma 2.10, \(\mu_Y\) converges weakly to a probability measure in probability for almost all \(z \in \mathbb{C}\), whose logarithmic potential is the same as \(\mu_{(\gamma_0)}\). Then the theorem follows from Lemma 2.9.

The general case \(k/n = o(\ln^{-1} n)\) is similar. We rearrange the columns \(X_1, \ldots, X_n\) according to \(l = l(i)\), the remainder of division of \(i\) by \(k\). More precisely, the rearranged column vectors are
\[
(\tilde{X}_1, \ldots, \tilde{X}_n) = \left( \begin{array}{c} X_{1}, X_{1+k}, X_{1+2k}, \ldots, X_{l=1} \ \cdots \ X_{2}, X_{2+k}, X_{2+2k}, \ldots, X_{l=2} \ \cdots \ X_{k}, X_{2k}, X_{3k} \ldots \end{array} \right).
\]

Clearly, \((\tilde{X}_1, \ldots, \tilde{X}_n) \overset{d}{=} X\). Consider
\[
\tilde{Y}(n) = \sum_{i=1}^{n-1} \tilde{X}_{i+1} \tilde{X}_i^* - \sum_{i=1}^{n-k} X_{i+k} X_i^* + \sum_{l=1}^{k-1} X_{l+1} X_{a_l}^* = Y(n) + \sum_{l=1}^{k-1} X_{l+1} X_{a_l}^*.
\]

where \(a_l\) is the largest integer that does not exceed \(n\) and has remainder \(l\) modulo \(k\), that is, \(X_{a_l}\) is the last term of the \(l\)th block in the rearrangement above. Similar to the proof of Theorem 4.1, we can obtain
\[
\left| \int_0^\delta \ln(\lambda) d\nu_{Y(n)} - zI_N (\lambda) \right| \leq \frac{k-1}{N} \left| \ln \left( s_N \left( Y(n) - zI_N \right) \right) \right| + \left| \int_0^\delta \ln(\lambda) d\nu_{\tilde{Y}(n)} - zI_N (\lambda) \right|.
\]

Note that the rank of \(\tilde{Y}(n) - Y(n)\) is at most \(k - 1\). Moreover, we can write
\[
\tilde{Y}(n) = \begin{pmatrix} \tilde{X}_1, \ldots, \tilde{X}_n \end{pmatrix} \begin{pmatrix} 0_{1 \times (n-1)} & 0 \\ I_{n-1} & 0_{(n-1) \times 1} \end{pmatrix} \begin{pmatrix} \tilde{X}_1^* \\ \vdots \\ \tilde{X}_n^* \end{pmatrix}.
\]

The uniform integrability of the logarithm function with respect to \(Y(n) - zI_N\) can then be deduced similarly to the previous case of \(k = 2\).

Since the matrix \(\Sigma_{Y(n)}(z) - \Sigma_{\tilde{Y}(n)}(z)\) has rank \(2k\), which is \(o(n)\), we can also deduce the weak convergence in probability of \(\mu_{Y(n)}\) towards \(\mu_{(\gamma_0)}\), again similar to the case \(k = 2\).

\[\Box\]

**4.2. The case \(k \geq n/2\)**

**Theorem 4.3.** Assume that the conditions (C1) and (C2) hold. Let \(N, k\) satisfy (1.3) such that \(k \geq n/2\), then there exists a probability measure \(\mu_{(\gamma_0, \gamma_1)}\), such that \(\mu_{Y(n)}\) converges weakly to \(\mu_{(\gamma_0, \gamma_1)}\) in probability.
Proof. We apply the logarithmic potential technique (Lemma 2.10) to obtain the convergence of \( \{ \nu_{Y^{(n)}} : n \in \mathbb{N}_+ \} \). We divide the proof into two steps. In Step 1, we prove the uniform integrability of the logarithm function for the family \( \{ \nu_{Y^{(n)}-zI_N} : N \in \mathbb{N}_+ \} \) in probability. Then we prove the almost sure convergence of the singular value empirical measure \( \{ \nu_{Y^{(n)}-zI_N} : N \in \mathbb{N}_+ \} \) in Step 2.

**Step 1.** We still denote

\[
\Sigma_{Y^{(n)}}(z) = \begin{pmatrix} Y^{(n)} - zI_N \end{pmatrix}^* \begin{pmatrix} Y^{(n)} - zI_N \end{pmatrix}.
\]

For \( \eta \in \mathbb{C}_+ = \{ w \in \mathbb{C} : \text{Re} w > 0 \} \), we denote the resolvent \( G(z, \eta) = (\Sigma_{Y^{(n)}}(z) - \eta I_{2N})^{-1} \), then by Lemma A.1, we have

\[
G(z, \eta) = \begin{pmatrix} -\eta I_N & Y^{(n)} - zI_N \end{pmatrix} \begin{pmatrix} Y^{(n)} - zI_N \end{pmatrix}^{-1} = \begin{pmatrix} G_{11}(z, \eta), & G_{12}(z, \eta) \end{pmatrix}, \tag{4.1}
\]

where

\[
G_{11}(z, \eta) = \eta \left( (Y^{(n)} - zI_N)(Y^{(n)} - zI_N)^* - \eta^2 I_N \right)^{-1}, \tag{4.2}
\]

\[
G_{12}(z, \eta) = \left( (Y^{(n)} - zI_N)(Y^{(n)} - zI_N)^* - \eta^2 I_N \right)^{-1}(Y^{(n)} - zI_N), \tag{4.3}
\]

\[
G_{21}(z, \eta) = (Y^{(n)} - zI_N)^* \left( (Y^{(n)} - zI_N)(Y^{(n)} - zI_N)^* - \eta^2 I_N \right)^{-1}(Y^{(n)} - zI_N), \tag{4.4}
\]

\[
G_{22}(z, \eta) = \eta \left( (Y^{(n)} - zI_N)^* (Y^{(n)} - zI_N) - \eta^2 I_N \right)^{-1}. \tag{4.5}
\]

We first establish the following so-called Wegner estimation

\[
-\frac{1}{n} \begin{pmatrix} \text{Tr} \left( G(z, ut) \right) \end{pmatrix} \leq C \left( 1 + t^{-\alpha_n - \beta} \right), \forall z \neq 0, \forall t \in (0, 1/2), \tag{4.6}
\]

for some positive constants \( C, \alpha, \beta \).

We follow the idea in Bose and Hachem [2020] to prove (4.6). By a standard concentration argument (see [Bose and Hachem, 2020, Proposition 26]) one can assume that the entries of \( X^{(n)} \) are complex Gaussian. Note that for \( k, l, i \in [N] \) and \( j \in [n] \), we have

\[
\frac{\partial (G_{11})_{kl}}{\partial X_{ij}} = - \left( G_{12}X \left( A^{(n)} \right)^* \right)_{kj} (G_{11})_{il} - (G_{11}XA^{(n)})_{kj} (G_{21})_{il}, \tag{4.7}
\]

\[
\frac{\partial (G_{12})_{kl}}{\partial X_{ij}} = - \left( G_{12}X \left( A^{(n)} \right) \right)_{kj} (G_{22})_{il} - (G_{12}XA^{(n)})_{kj} (G_{12})_{il}. \tag{4.8}
\]

By (4.7), (4.8) and the Integration by Parts formula for Gaussian variables, for \( i, j \in [n] \), we have

\[
\begin{align*}
\mathbb{E} [X_i^*G_{11}X_j] &= 1_{\{i=j\}} \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{11} \right] - \mathbb{E} \left[ (X^*G_{12}X \left( A^{(n)} \right)^* \right]_{ij} \cdot \frac{1}{n} \text{Tr} G_{11} \\
&\quad - \mathbb{E} \left[ (X^*G_{11}XA^{(n)})_{ij} \right] \cdot \frac{1}{n} \text{Tr} G_{21} , \tag{4.9}
\end{align*}
\]

\[
\begin{align*}
\mathbb{E} [X_i^*G_{12}X_j] &= 1_{\{i=j\}} \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{12} \right] - \mathbb{E} \left[ (X^*G_{11}X \left( A^{(n)} \right) \right]_{ij} \cdot \frac{1}{n} \text{Tr} G_{21} \\
&\quad - \mathbb{E} \left[ (X^*G_{11}XA^{(n)})_{ij} \right] \cdot \frac{1}{n} \text{Tr} G_{22}.
\end{align*}
\]
By Lemma C.2 and (4.7), we have

\[
\Var \left( \frac{1}{n} \Tr G_{11} \right) = \frac{1}{n^2} \Var (\Tr G_{11}) \\
\leq \frac{1}{2n^3} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( \frac{\partial \Tr G_{11}}{\partial X_{ij}} \right)^2 \right] + \frac{1}{n^3} \mathbb{E} \left[ \left( \sum_{k \in [N]} k \frac{\partial (G_{11})_{kk}}{\partial X_{ij}} \right)^2 \right] \\
= \frac{1}{n^3} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( \sum_{k \in [N]} k \frac{\partial (G_{11})_{kk}}{\partial X_{ij}} \right)^2 \right] \\
= \frac{2}{n^3} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( (G_{11}G_{12}X(A^{(n)})^*)_{ij} + (G_{21}G_{11}X(A^{(n)})^*)_{ij} \right)^2 \right] \\
\leq \frac{4}{n^2} \mathbb{E} \left[ \left\| G_{11}G_{12}X(A^{(n)})^* \right\|_{HS}^2 + \left\| G_{21}G_{11}X(A^{(n)})^* \right\|_{HS}^2 \right] \\
\leq \frac{4}{n^2} \mathbb{E} \left[ \left\| G_{11}G_{12}X(A^{(n)})^* \right\|^2 + \left\| G_{21}G_{11}X(A^{(n)})^* \right\|^2 \right].
\]

Here, we use the fact that \( \| M \|_{HS} \leq \sqrt{n} \| M \| \) for \( M \in \mathbb{C}^{n \times n} \). Note that \( \| A^{(n)} \| = 1, \| X \| \to 1 + \sqrt{\gamma_0}, \left\| \left( (Y^{(n)} - zI_N) (Y^{(n)} - zI_N)^* - \eta^2 I_N \right)^{-1} \right\| \leq (3\eta)^{-2} \), we have

\[
\Var \left( \frac{1}{n} \Tr G_{11} \right) \leq \frac{C'\| \eta \|^2}{n^2 (3\eta)^8}.
\]

Here, \( C' \) is a positive constant that depends only on \( \gamma_0 \) and \( z \) and may vary in different places. By a similar argument, one can obtain

\[
\Var \left( \frac{1}{n} \Tr G_{ij} \right) \leq \frac{C'(|\eta|^4 + 1)}{n^2 (3\eta)^8}, \quad i, j = 1, 2.
\]

By Lemma C.2 and (4.7), for \( k, l \in [N] \), we have

\[
\Var \left( X_k^* G_{11} X_l \right) \\
\leq \frac{1}{2n} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( \frac{\partial (X_k^* G_{11} X_l)}{\partial X_{ij}} \right)^2 \right] + \frac{1}{n} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( \frac{\partial (X_k^* G_{11} X_l)}{\partial X_{ij}} \right)^2 \right] \\
= \frac{1}{n} \mathbb{E} \left[ \sum_{i \in [N], j \in [n]} \left( \frac{\partial (X_k^* G_{11} X_l)}{\partial X_{ij}} \right)^2 \right].
\]
We denote

\[ A_{ij} = A^{(n)}_{ij}, \]

then \( \text{Var}(\xi_{ij} A_{ij}) \leq \frac{C}{n} \left( 1 + \frac{\eta^2}{n} \right) \). Hence, substitute (4.11) and (4.12) to (4.9) and (4.10), we have

\[
\mathbb{E} [X_{ij}^* G_{ij} X_j] = 1_{(i=j)} \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] - \mathbb{E} \left[ X_{ij}^* G_{ij} X_j \right] X_{ij}^* \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] \]

\[
\mathbb{E} [X_{ij}^* G_{ij} X_j] = 1_{(i=j)} \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] - \mathbb{E} \left[ X_{ij}^* G_{ij} X_j \right] X_{ij}^* \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] \]

\[
\text{Var}(X_{ij}^* G_{ij} X_j) \leq \frac{C'}{n} \left( \eta^2 \right) \left( 1 + \frac{1}{(3\eta)^4} \right) \]

By a similar argument, one can obtain

\[
\text{Var}(X_{ij}^* G_{ij} X_j) \leq \frac{C' \left( 1 + \frac{|\eta|^4}{n} \right) \left( 1 + \frac{(3\eta)^4}{n} \right)}{n \left( 3\eta \right)^8}, \quad i, j = 1, 2. \tag{4.12}
\]

We denote

\[
\Psi(\eta) = \frac{(1 + |\eta|^4) \left( 1 + (3\eta)^2 \right)}{(3\eta)^8}, \tag{4.13}
\]

then \( \text{Var} \left( \frac{1}{n} \text{Tr} G_{ij} \right) \text{Var}(X_{ij}^* G_{ij} X_j) \leq \frac{C'}{n} \Psi(\eta)^2 \). Hence, substitute (4.11) and (4.12) to (4.9) and (4.10), we have

\[
\mathbb{E} [X_{ij}^* G_{ij} X_j] = 1_{(i=j)} \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] - \mathbb{E} \left[ X_{ij}^* G_{ij} X_j \right] X_{ij}^* \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] \]

\[
\text{Var}(X_{ij}^* G_{ij} X_j) \leq \frac{C'}{n} \left( \eta^2 \right) \left( 1 + \frac{1}{(3\eta)^4} \right) \]

Here, we use the notation \( O(\Psi(\eta)/n^{3/2}) \) to represent a number whose absolute value is bounded by \( C\Psi(\eta)/n^{3/2} \) for some large constant \( C \) that depends on \( \gamma, \gamma_0, \gamma_1 \). The constant may vary in different place. We denote \( A_{ij} = \mathbb{E} [X^* G_{ij} X] \) and \( g_{ij} = \mathbb{E} \left[ \frac{1}{n} \text{Tr} G_{ij} \right] \) for \( i, j = 1, 2 \). Noting that \( g_{11} = g_{22} \), we can write (4.14) and (4.15) as

\[
A_{11} \left( I_n + g_{21} A^{(n)} \right) + g_{11} A_{12} \left( A^{(n)} \right)^* = g_{11} I_n + O \left( \Psi(\eta) n^{3/2} \right) E_n, \tag{4.16}
\]

\[
g_{11} A_{11} A^{(n)} + A_{12} \left( I_n + g_{12} \left( A^{(n)} \right)^* \right) = g_{12} I_n + O \left( \Psi(\eta) n^{3/2} \right) E_n. \tag{4.17}
\]
Here, $E_n$ is a $n \times n$ matrix with all the entries equal to 1. $O(\Psi(\eta)/n^{3/2})E_n$ is a matrix whose entries are bounded by $O(\Psi(\eta)/n^{3/2})$. By (4.16) and (4.17), we have

$$A_{11} \left( \left( I_n + g_{21} A^{(n)} \right) \left( I_n + g_{12} \left( A^{(n)} \right)^* \right) - g_{11}^2 A^{(n)} \left( A^{(n)} \right)^* \right) = g_{11} I_n + O \left( \frac{\Psi'(\eta)}{n^{3/2}} \right) E_n,$$

(4.18)

$$A_{12} \left( \left( I_n + g_{12} \left( A^{(n)} \right)^* \right) \left( I_n + g_{21} A^{(n)} \right) - g_{11}^2 \left( A^{(n)} \right)^* A^{(n)} \right) = g_{12} I_n + (g_{12}g_{21} - g_{11}^2) A^{(n)} + O \left( \frac{\Psi'(\eta)}{n^{3/2}} \right) E_n,$$

(4.19)

where

$$\Psi'(\eta) = \Psi(\eta) \left( 1 + \frac{1}{(3\eta)^2} + \frac{\eta}{(3\eta)^2} \right).$$

To compute $A_{11}$ from (4.18), we can write

$$\left( I_n + g_{21} A^{(n)} \right) \left( I_n + g_{12} \left( A^{(n)} \right)^* \right) - g_{11}^2 A^{(n)} \left( A^{(n)} \right)^* = I_k \left( \begin{array}{c} g_{12} I_{n-k} \\ 0_{(2k-n) \times (n-k)} \end{array} \right) \left( \begin{array}{c} 0_{(n-k) \times (2k-n)} \\ I_{2k-n} \end{array} \right) \left( \begin{array}{c} 0_{(2k-n) \times (n-k)} \\ 1 - g_{11}^2 I_{n-k} \end{array} \right).$$

By Lemma A.1, if $g_{11}^2 \neq 1$, we have

$$\left( \left( I_n + g_{21} A^{(n)} \right) \left( I_n + g_{12} \left( A^{(n)} \right)^* \right) - g_{11}^2 A^{(n)} \left( A^{(n)} \right)^* \right)^{-1} = \left( \begin{array}{cc} 1 - g_{11}^2 + g_{12}g_{21} I_{n-k} & 0_{(n-k) \times (2k-n)} \\ -g_{21} & I_{2k-n} \end{array} \right) \left( \begin{array}{cc} 1 - g_{11}^2 I_{n-k} & 0_{(2k-n) \times (n-k)} \\ 0_{(2k-n) \times (n-k)} & 1 - g_{11}^2 I_{n-k} \end{array} \right).$$

Hence, substitute it to (4.18), we have

$$A_{11} = \left( g_{11} I_n + O \left( \frac{\Psi'(\eta)}{n^{3/2}} \right) E_n \right) \left( \begin{array}{cc} 1 - g_{11}^2 + g_{12}g_{21} I_{n-k} & 0_{(n-k) \times (2k-n)} \\ -g_{21} & I_{2k-n} \end{array} \right) \left( \begin{array}{cc} 1 - g_{11}^2 I_{n-k} & 0_{(2k-n) \times (n-k)} \\ 0_{(2k-n) \times (n-k)} & 1 - g_{11}^2 I_{n-k} \end{array} \right).$$

Similarly, we have

$$A_{12} = \left( g_{12} I_n + (g_{12}g_{21} - g_{11}^2) A^{(n)} + O \left( \frac{\Psi'(\eta)}{n^{3/2}} \right) E_n \right) \left( \begin{array}{cc} 1 - g_{11}^2 I_{n-k} & 0_{(n-k) \times (2k-n)} \\ -g_{21} & I_{2k-n} \end{array} \right) \left( \begin{array}{cc} 1 - g_{11}^2 I_{n-k} & 0_{(2k-n) \times (n-k)} \\ 0_{(2k-n) \times (n-k)} & 1 - g_{11}^2 I_{n-k} \end{array} \right).$$

(4.21)
In addition, by (4.1), we have the following identities.

\[-\eta G_{11} + G_{12} \left(Y^{(n)}\right)^* - zg_{12} = I_N,\]

\[G_{11} Y^{(n)} - zg_{11} - \eta G_{12} = 0.\]

By taking the expectation of trace, we have

\[-\eta g_{11} - zg_{12} + \frac{1}{n} \text{Tr} \left( A_{12} \left( A^{(n)} \right)^* \right) = \frac{N}{n}\]  \hspace{1cm} (4.22)

\[-zg_{11} - \eta g_{12} + \frac{1}{n} \text{Tr} \left( A_{11} A^{(n)} \right) = 0.\]  \hspace{1cm} (4.23)

To prove (4.6), we choose \(\eta = it\) where \(t \in (0, 1/2).\) By (4.2), (4.3), (4.4) and (4.5), we can see that \(g_{11}(z, it) = g_{22}(z, it)\) is pure imaginary so we can write \(g_{11}(z, it) = is(z, t)\) with \(s(z, t) > 0.\) Furthermore, we have \(g_{12}(z, it) = g_{21}(z, it)\) and \(|g_{12}| \leq C t^{-2}.\) By some computation, one can easily see that \(\Psi(it) \leq 4t^{-8}\) and \(\Psi'(it) \leq 12t^{-10}.\) Hence, we can simplify (4.20) as

\[A_{11} = ts \begin{pmatrix}
\frac{1 + s^2 + |g_{12}|^2}{1 + s^2} I_{n-k} & 0_{(n-k) \times (2k-n)} \\
0_{(2k-n) \times (n-k)} & I_{2k-n}
\end{pmatrix} \begin{pmatrix}
\frac{-g_{12}}{1 + s^2} I_{n-k} & 0_{(2k-n) \times (n-k)} \\
0_{(n-k) \times (2k-n)} & \frac{1}{1 + s^2} I_{n-k}
\end{pmatrix} + O \left( \frac{1}{n^{3/2} t^{14}} \right) E_n.
\]

Thus,

\[
\frac{1}{n} \text{Tr} \left( A_{11} A^{(n)} \right) = \frac{n - k}{n} \left( -ts g_{12} \right) + O \left( \frac{1}{n^{3/2} t^{14}} \right).
\]

Together with (4.23), we have

\[g_{12} = \left( t + \frac{n - k}{n} \cdot \frac{s}{1 + s^2} \right)^{-1} \left( -zs + O \left( \frac{1}{n^{3/2} t^{14}} \right) \right).\]  \hspace{1cm} (4.24)

Similarly, we can simplify (4.21) as

\[A_{12} = \left( \frac{g_{12}}{1 + s^2} I_{n-k} \right) \begin{pmatrix}
0_{(2k-n) \times (n-k)} & \frac{-g_{12}}{1 + s^2} I_{n-k} \\
\frac{g_{12} I_{2k-n}}{s^2} & 0_{(n-k) \times (2k-n)}
\end{pmatrix} + O \left( \frac{1}{n^{3/2} t^{14}} \right) E_n.
\]

Hence,

\[
\frac{1}{n} \text{Tr} \left( A_{12} \left( A^{(n)} \right)^* \right) = \frac{n - k}{n} \frac{s^2}{1 + s^2} + O \left( \frac{1}{n^{3/2} t^{14}} \right).
\]

Together with (4.24) and (4.22), we have

\[
\left( t + \frac{n - k}{n} \cdot \frac{s}{1 + s^2} \right) \left( ts + \frac{n - k}{n} \frac{s^2}{1 + s^2} - \frac{N}{n} \right) + s|z|^2 = O \left( \frac{1}{n^{3/2} t^{14}} \right).
\]
Thus,
\[
    s = \frac{N}{n} \left( \frac{t + n - k \cdot \frac{s}{n}}{1 + s^2} \right) - \frac{n - k \cdot \frac{2ts^2}{1 + s^2}}{n} - \frac{(n - k \cdot \frac{s^3}{(1 + s^2)^2} + O \left( \frac{1}{n^{3/2}t^4} \right)}{z}^2 + t^2
\]
\[
    \leq \frac{N}{n} \left( \frac{t + n - k \cdot \frac{s}{n}}{1 + s^2} \right) + O \left( \frac{1}{n^{3/2}t^4} \right),
\]
which establishes (4.6) for \( \alpha = 16 \) and \( \beta = 3/2 \).

Therefore, by [Guionnet et al., 2011, Lemma 15] and (4.6), we have
\[
    \mathbb{E} \mu_{\Sigma_{\gamma(n)}}((-t, t)) \leq C \left( 1 + t^{-\alpha}n^{-\beta} \right).
\]
Note that \( 2\mu_{\Sigma_{\gamma(n)}}(\cdot) = \nu_{\gamma(n) - zI_n}(\cdot) + \nu_{\gamma(n) - zI_n}(\cdot) \), by a standard argument (see [Guionnet et al., 2011, Proposition 14]), one can deduce the following uniform integrability of the logarithm function from Theorem 3.1 and (4.25)
\[
    \lim_{K \to +\infty} \sup_{n \in \mathbb{N}} \mathbb{P} \left( \left| \int_{|\ln \lambda| > K} \ln \lambda |\nu_{\gamma(n) - zI_n}(d\lambda)\right| > \epsilon \right) = 0, \forall \epsilon > 0.
\]

**Step 2.** By a standard concentration argument (see [Bose and Hachem, 2020, Proposition 26]) one can assume that the entries of \( X^{(n)} \) are complex Gaussian. For test function \( f \in C_b(\mathbb{R}) \), for \( z \in \mathbb{C} \), we have
\[
    \int f(x) d\nu_{\gamma(n) - zI_n}(x) = \int f(\sqrt{x}) d\mu_{\gamma(n) - zI_n}(\gamma(n) - zI_n)(x)
\]
\[
    = \int f(\sqrt{x}) d\mu_{\gamma(n)}(\gamma(n))^{*} - z(\gamma(n))^{*} - \mathbb{E}(\gamma(n))^{*} + |z|^2 I_n(x)
\]
\[
    = \int f(\sqrt{x + |z|^2}) d\mu_{\gamma(n)}(\gamma(n))^{*} - z(\gamma(n))^{*} - \mathbb{E}(\gamma(n))(x).
\]

By [Horn and Johnson, 2013, Theorem 1.3.22], we have
\[
    \mu_{\gamma(n)}(\gamma(n))^{*} - z(\gamma(n))^{*} - \mathbb{E}(\gamma(n))
\]
\[
    = \frac{n}{N} \mu_{\gamma(n)}(\gamma(n))^{*} - z(\gamma(n))^{*} - \mathbb{E}(\gamma(n))^{*} - z\mathbb{E}(\gamma(n))^{*} - zA^{(n)}(X^{(n)})^{*} + \frac{N - n}{N} \delta_0.
\]

Denote \( \gamma_0^+ = [\gamma_0] + 1 \) be the smallest integer that is strictly greater than \( \gamma_0 \). Without loss of generality, we can assume that \( N < \gamma_0^+ n \). Let \( \tilde{X}^{(n)} \) be a \((\gamma_0^+ n) \times n\) matrices whose entries are complex Gaussian with mean zero and variance \( 1/n \), such that \( \tilde{X}^{(n)}_{[N],[n]} = X^{(n)} \). Then Wishart matrix \( (X^{(n)})^{*} X^{(n)} \) can be written as
\[
    (X^{(n)})^{*} X^{(n)} = (\tilde{X}^{(n)})^{*} \begin{pmatrix} I_N & 0_{N \times (\gamma_0^+ n - N)} \\ 0_{(\gamma_0^+ n - N) \times N} & 0_{(\gamma_0^+ n - N) \times (\gamma_0^+ n - N)} \end{pmatrix} \tilde{X}^{(n)}.
\]
Under the condition (1.3), for all polynomial $P$ in 2 non-commutative indeterminates, one has
\[
\frac{1}{n} \text{Tr} \left[ P \left( (A^{(n)}), (A^{(n)})^* \right) \right] \rightarrow \tau \left[ P(a, a^*) \right]
\]
for some non-commutative element $a$ in a $C^*$-probability space $(A, *, \tau, \| \cdot \|)$ with a faithful trace $\tau$. Besides, the Hermitian matrix
\[
\begin{pmatrix}
I_N & 0_{N \times (\gamma_0^+ n - N)} \\
0_{(\gamma_0^+ n - N) \times N} & 0_{(\gamma_0^+ n - N) \times (\gamma_0^+ n - N)}
\end{pmatrix}
\]
converges to the law
\[
\frac{\gamma_0}{\gamma_0} \delta_1 + \frac{\gamma_0^- - \gamma_0}{\gamma_0} \delta_0
\]
in the $C^*$-probability space of random matrices. Therefore, by [Male, 2012, Corollary 2.2, Theorem 1.6], there exists a non-commutative random variable $x$ satisfying that $x$ and $a$ are free, such that for any polynomial $P$ in 3 non-commutative indeterminates,
\[
\frac{1}{n} \text{Tr} \left[ P \left( (X^{(n)})^* X^{(n)}, A^{(n)}, (A^{(n)})^* \right) \right] \rightarrow \tau \left[ P(x, a, a^*) \right], \ n \rightarrow \infty,
\]
(4.28)
almost surely. Consequently, the eigenvalue empirical measure of $P \left( (X^{(n)})^* X^{(n)}, A^{(n)}, (A^{(n)})^* \right)$ converges almost surely. In particular, for $z \in \mathbb{C}$, by choosing the polynomial $P(x, v, w) = vxwx - zwx - \overline{z}vx$, we obtain the almost sure convergence of
\[
\mu_{\gamma_0} (X^{(n)})^* X^{(n)} (A^{(n)})^* X^{(n)} - z (A^{(n)})^* X^{(n)} - \overline{z} A^{(n)} (X^{(n)})^* X^{(n)}.
\]
Together with (1.3), (4.27) and (4.26), one can easily obtain the almost sure convergence of
\[
\nu_{\gamma_0} (n) - z I_N : N \in \mathbb{N}_+ \}.
\]
The proof is concluded by Step 1, Step 2 and Lemma 2.10.

Finally, we observe that $Y^{(n)}$ is a product of two independent rectangular matrices. It turns out that the limiting distribution $\mu^{(\tau_0, \tau_1)}$ has an explicit form if moreover, the matrix entries are Gaussian, see Qi and Zhao [2021], Akemann and Ipsen [2015], Zeng [2017] for examples. Since the Gaussian case is a special case covered in Theorem 4.3, this explicit form remains in the general case. Now we determine $\mu^{(\tau_0, \tau_1)}$ explicitly.

**Corollary 4.1.** Assume that the conditions in Theorem 4.3 are satisfied. Then the limit measure $\mu^{(\tau_0, \tau_1)}$ can be characterized as follows.

- **If** $\gamma_0 + \gamma_1 \leq 1$, then
  \[
  \mu^{(\tau_0, \tau_1)} (dz) = \frac{1}{\pi \gamma_0 \sqrt{(1 - \gamma_1 - \gamma_0)^2 + 4|\gamma|^2}} 1\left\{ |\gamma| \leq \sqrt{\gamma_0 (1 - \gamma_1)} \right\} \, dz.
  \]
  (4.29)

- **If** $\gamma_0 + \gamma_1 \geq 1$, then
  \[
  \mu^{(\tau_0, \tau_1)} (dz) = \frac{1}{\pi \gamma_0 \sqrt{(\gamma_1 + \gamma_0 - 1)^2 + 4|\gamma|^2}} 1\left\{ |\gamma| \leq \sqrt{\gamma_0 (1 - \gamma_1)} \right\} \, dz + \frac{\gamma_1 + \gamma_0 - 1}{\gamma_0} \delta_0 (dz)
  \]
  (4.30)
Proof. From the proof of Theorem 4.3, Lemma 2.9 and Lemma 2.10, it is enough to determine the limit measure \( \mu^{(\gamma_0, \gamma_1)} \) for the complex Gaussian case. We can write

\[
Y^{(n)} = (X_{k+1}, \ldots, X_n) \begin{pmatrix} X_1^* \\ \vdots \\ X_{n-k}^* \end{pmatrix},
\]

since \( k \geq n/2 \), the two matrices \((X_{k+1}, \ldots, X_n)\) and \((X_1, \ldots, X_{n-k})\) are independent.

Consider the case \( N \leq n - k \) first. By [Zeng, 2017, Corollary 1.2], the empirical spectral distribution of the matrix

\[
\frac{n}{\sqrt{N(n-k)}} Y^{(n)}
\]

converges weakly to a probability measure with density function

\[
\rho(z) = \frac{1}{\pi} \left( \frac{1 - \gamma_0}{1 - \gamma_1} \right)^2 + \frac{4\gamma_0}{1 - \gamma_1} |z|^2 \right)^{-1/2} 1_{\{|z| \leq 1\}},
\]

almost surely. Note that \( n/\sqrt{N(n-k)} \to 1/\sqrt{\gamma_0(1 - \gamma_1)} \), the empirical spectral distribution \( \mu_{Y^{(n)}} \) converges weakly to \( \mu^{(\gamma_0, \gamma_1)} \), whose density function is given in (4.29), almost surely.

Next, we consider the case \( N > n - k \). We set the \((n-k) \times (n-k)\) matrix by

\[
\hat{Y}^{(n)} = \begin{pmatrix} X_1^* \\ \vdots \\ X_{n-k}^* \end{pmatrix} (X_{k+1}, \ldots, X_n).
\]

By [Zeng, 2017, Corollary 1.2], the empirical spectral distribution of the matrix

\[
\frac{n}{\sqrt{N(n-k)}} \hat{Y}^{(n)}
\]

converges weakly to a probability measure with density function

\[
\rho(z) = \frac{1}{\pi} \left( \frac{1 - \gamma_1}{\gamma_0} \right)^2 + \frac{4(1 - \gamma_1)}{\gamma_0} |z|^2 \right)^{-1/2} 1_{\{|z| \leq 1\}}, \tag{4.31}
\]

almost surely. Moreover, by [Horn and Johnson, 2013, Theorem 1.3.22], \( Y^{(n)} \) and \( \hat{Y}^{(n)} \) have the same nonzero eigenvalues, which implies that

\[
\mu_{Y^{(n)}} = \frac{(n-k)}{N} \mu_{\hat{Y}^{(n)}} + \frac{N - (n-k)}{N} \delta_0, \tag{4.32}
\]

Under the ratio assumption (1.3), one can derive (4.30) from (4.31) and (4.32).

4.3. The case \( \epsilon n < k < n/2 \) for some \( \epsilon > 0 \)

The previous results do not cover the case where \( \ln^{-1} n \ll k/n < 1/2 \). Consider the case where \( \epsilon < k/n < 1/2 \) for some \( \epsilon > 0 \) and large \( k, n \). Roughly speaking, we would implement
the same calculation done for the case $k \geq n/2$ in Section 4.2. A key step is to establish the corresponding Wegner estimate (4.6). We conjecture that such estimate is indeed possible, and the eigenvalue distribution still have a limit (as for the case $k \geq n/2$). We explain this fact with the case $n/3 \leq k < n/2$.

The argument for the uniform integrability in probability of the logarithm function in Theorem 4.3 still works. More precisely, one can define the matrix $G(z, \eta)$ and write it as a $2 \times 2$ blocking matrix as in (4.1). The estimates of the variance of $\text{Tr}G_{ij}$ and $X_p^*G_{ij}X_q$ in (4.11) and (4.12) only use the fact $\|A^{(n)}\| = 1$, and hence are still valid. Thus, the equations (4.16) and (4.17) still hold without any modification. One can follow the same idea to solve $A_{11}$ and $A_{12}$ from the system of equations (4.16) and (4.17), noting that the inverse of the two matrices

\[
\begin{align*}
\left( I_n + g_{21}A^{(n)} \right) &\left( I_n + g_{12} \left( A^{(n)} \right)^* \right) - g_{11}^2 A^{(n)} \left( A^{(n)} \right)^*, \\
\left( I_n + g_{12} \left( A^{(n)} \right)^* \right) &\left( I_n + g_{21}A^{(n)} \right) - g_{11}^2 \left( A^{(n)} \right)^* A^{(n)}
\end{align*}
\]

(4.33)

are the matrices with non-vanishing $(i, j)$ entries if $|i - j| \in \{0, \pm k, \pm 2k\}$. Plugging $A_{11}$ and $A_{12}$ into the identities (4.22) and (4.23), one may establish the required Wegner estimation (4.6), which leads to the uniform integrability of the logarithm function. Under the ratio condition (1.3), the argument of the convergence of $\nu_{Y^{(n)} - zI_N}$ in Theorem 4.3 still works. Therefore, by Lemma 2.10, there exists a probability measure $\mu^{(\gamma_0, \gamma_1)}$, such that $\mu_{Y^{(n)}}$ converges weakly to $\mu^{(\gamma_0, \gamma_1)}$ in probability.

In general, when $n/(K + 1) \leq k < n/K$ for some fixed positive integer $K$, the argument above is still valid, where the inverse of the two matrices (4.33) are the matrices with non-vanishing $(i, j)$ entries if $|i - j| \in \{0, \pm k, \pm 2k, \ldots, \pm Kk\}$. However, the exact computations for the Wegner estimate seem much more complicated for general $K$ then the case $K = 1$ considered in Section 4.2.

Appendix A: Matrices

The following linear algebraic lemmas could be found in Bai and Silverstein [2010] and Tao [2012].

**Lemma A.1.** For $A \in \mathbb{C}^{p \times p}$, $D \in \mathbb{C}^{q \times q}$, $B \in \mathbb{C}^{p \times q}$, $C \in \mathbb{C}^{q \times p}$, if $D$ and $A - BD^{-1}C$ are invertible, then

\[
\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}.
\]

**Lemma A.2.** For $A, B \in \mathbb{C}^{n \times p}$, we have

\[
s_{i+j-1}(A + B) \leq s_i(A) + s_j(B), \quad \forall i \in \mathbb{N}_+.
\]

**Lemma A.3.** [Bai and Silverstein, 2010, Exercise 1.3.22] For $A \in \mathbb{C}^{m \times n}$, let $C \in \mathbb{C}^{p \times q}$ be a submatrix of $A$, then singular values of $A$ and $C$ satisfies

\[
s_i(C) \leq s_i(A), \quad \forall i \in \mathbb{N}_+.
\]
Appendix B: Concentration inequalities

The following Hoeffding’s inequality could be found in Tao [2012].

**Lemma B.1.** [Tao, 2012, Exercise 2.1.4] Let $X_1, \ldots, X_n$ be independent real random variables, with $X_i$ taking values in an interval $[0, 1]$, and let $S_n = X_1 + \cdots + X_n$. Then

$$
\mathbb{P} \left( |S_n - \mathbb{E}[S_n]| \geq \sqrt{n} \lambda \right) \leq C \exp \left( -c \lambda^2 \right),
$$

for some absolute constants $c, C > 0$.

The following lemma could be found in Tao et al. [2010], see also [Bordenave and Chafaï, 2012, Lemma 4.13].

**Lemma B.2.** [Bose and Hachem, 2020, Lemma 4] Let $X$ be a $N \times n$ random matrix satisfying condition (C1). Let $u \in S^{n-1}$ be a deterministic vector and $W$ be a deterministic $d$-dimensional vector subspace of $\mathbb{C}^N$, where $d$ does not depend on $N$ and $n$. Then for large $n, N$,

$$
\mathbb{P} (\text{dist} (Xu, W) \leq c) \leq \exp(-cN).
$$

Appendix C: Other lemmas

The following lemma could be found in Vershynin [2014].

**Lemma C.1.** [Vershynin, 2014, Lemma 8.3] Let $Z_1, \ldots, Z_n$ be arbitrary non-negative random variables (not necessarily independent), and $p_1, \ldots, p_n$ be non-negative numbers such that their sum equals to 1. Then for every $t \in \mathbb{R}$,

$$
\mathbb{P} \left( \sum_{i=1}^{n} p_i Z_i \leq t \right) < 2 \sum_{i=1}^{n} p_i \mathbb{P} (Z_i \leq 2t).
$$

The following lemma is known as Poincaré-Nash inequality and could be found in Pastur [2005].

**Lemma C.2.** [Pastur, 2005, Proposition 2.4] Let $\xi = (\xi_1, \ldots, \xi_d)^\top$ be a real centered Gaussian random vector with covariance matrix $\Sigma$. Let $\Phi_1, \Phi_2 : \mathbb{R}^d \to \mathbb{C}$ be two functions with bounded partial derivatives, then

$$
\text{Cov} (\Phi_1, \Phi_2) = \mathbb{E} [\Phi_1 \Phi_2] - \mathbb{E} [\Phi_1] \mathbb{E} [\Phi_2] \leq \sqrt{\mathbb{E} [(\nabla \Phi_1(\xi))^\top \Sigma (\nabla \Phi_1(\xi))] \mathbb{E} [(\nabla \Phi_2(\xi))^\top \Sigma (\nabla \Phi_2(\xi))]}. 
$$

Acknowledgments  The authors would like to thank Walid Hachem for pointing to us the fact that the least singular value estimate in Section 3 does not require the existence of density for the matrix entries using an approximation argument (see the beginning of the proof of Theorem 3.1).

We are grateful to the Editor and a referee for numerous valuable and detailed comments that have helped improve the paper. In particular, the computation of the explicit form of the LSD $\mu^{(\gamma_0, \gamma_1)}$ in Corollary 4.1 is suggested by the referee.
References


