A Local Limit Theorem and Delocalization of Eigenvectors for Polynomials in Two Matrices

Ching-Wei Ho
Department of Mathematics
Indiana University
Bloomington, IN 47405
cwho@iu.edu

May 1, 2020

Abstract

We propose a boundary regularity condition for the $M_n(\mathbb{C})$-valued subordination functions in free probability to prove a local limit theorem and delocalization of eigenvectors for self-adjoint polynomials in two random matrices. We prove this through estimating the pair of $M_n(\mathbb{C})$-valued approximate subordination functions for the sum of two $M_n(\mathbb{C})$-valued random matrices $\gamma_1 \otimes C_N + \gamma_2 \otimes U_N^* D_N U_N$, where $C_N, D_N$ are deterministic diagonal matrices, and $U_N$ is Haar unitary.

Contents

1 Introduction 1

2 Preliminary 3
2.1 Free Probability 10
2.1.1 Strong Convergence 14
2.2 Operator-valued Free Probability 15
2.3 Linearization 16
2.4 Newton’s Method 17

3 The Unitarily Invariant Model 18
3.1 The Sum of Two $M_n(\mathbb{C})$-valued Matrices 18
3.2 Main Results 23
3.2.1 A Local Limit Theorem for Polynomials 23
3.2.2 Delocalization of Eigenvectors 23

4 Acknowledgement 25

1 Introduction

Let $\{C_N\}_{N \in \mathbb{N}}$ and $\{D_N\}_{N \in \mathbb{N}}$ be deterministic self-adjoint diagonal $N \times N$ matrices. Suppose that $\{U_N\}_{N \in \mathbb{N}}$ is a sequence of $N \times N$ random unitary matrices distributed uniformly in $U(N)$. We consider random matrices $c_N = C_N$ and $d_N = U_N^* D_N U_N$. The law of $d_N$ is independent of unitary conjugation. Suppose that there are free random variables $c, d$ such that $c_N \to c$ and $d_N \to d$ in distribution.
Suppose that $P$ is a self-adjoint polynomial in two noncommuting indeterminates. Our purpose is to study the density of the eigenvalues of the random matrix $X_N = P(e_N, d_N)$. In other words, given $x \in \mathbb{R}$ and a sequence $(\eta_N)_{N \in \mathbb{N}}$ that converge to 0, we consider the ratio

$$M_{\eta_N}(x) \quad 2N\eta_N$$

where $M_{\eta_N}(x)$ denotes the (random) number of eigenvalues of $X_N$ in the interval $[x - \eta_N, x + \eta_N]$. We provide technical conditions under which the limit

$$\lim_{N \to \infty} \frac{M_{\eta_N}(x)}{2N\eta_N} = \rho(x)$$

exists in probability, and the convergence is uniform for all $x$ in some closed interval $I$. The sequence $\eta_N$ is of the form $\eta_N = (cN^{-1/2} \log N)^{\alpha}$ where $c$ is a large constant and $\alpha \in (0, 1)$.

Under the same technical conditions, we prove a delocalization result for the eigenvectors of $X_N$ corresponding to eigenvalues in the interval $I$. More precisely, the eigenvectors $v_1^{(N)}, \ldots, v_{k_N}^{(N)}$ with the corresponding eigenvalue in $I$ satisfy

$$\mathbb{P} \left( \max_{j=1, \ldots, k_N} \max_{i=1, \ldots, N} |v_j^{(N)}(i)|^2 > N^{-\frac{1}{12}} \log N \right) \leq \exp(-N^{-\delta})$$

for all sufficiently large $N$, for some $\delta > 0$. Here $v_a^{(N)}(i)$ denotes the $i$-coordinate of the vector $v_a^{(N)}$.

The technical conditions we require concern regularity of a certain matrix-valued subordination function whose existence is proved in free probability theory. Our methods extend work done by Kargin [21] for the random matrix $c_N + d_N$. If $p(c_N, d_N) = c_N + d_N$ and $I$ is contained in the bulk of the eigenvalues of $X_N$ then the technical conditions are equivalent to the ones proposed in [21] and are known to be true. For general self-adjoint $p$, the technical conditions are believed to be true when the closed interval $I$ considered is inside the interior of the bulk of the $X_N$.

There have been further developments in local limit theorems regarding the model $c_N + d_N$. Bao, Erdős and Schnelli [4,5] improved the length of the interval to the optimal scale of $N^{-1+\epsilon}$. Discussions of the local law at the edge can be found in [6]. Erdős, Krüger and Nemish [17] studied the local laws for polynomials of Wigner matrices.

It is well-known that the eigenvalue distributions of general Wigner matrices converge weakly to the semicircle law; this is a macroscopic phenomenon of the asymptotic behavior of the eigenvalue distributions. On a microscopic scale, Erdős, Schlein, and Yau [18] proved a local limit theorem for general Wigner matrices, for which the number of eigenvalues in an interval of length $N^{-\frac{2}{3}} \log N$ around a point $x \in \mathbb{R}$ is concerned. More precisely, they prove that if $\eta^* \geq C N^{-2/3} \log N$ and $\kappa > 0$, then

$$\mathbb{P} \left\{ \sup_{|x| \leq 2 - \kappa} \frac{|M_{\eta^*}(x)|}{2N\eta^*} - \rho_{sc}(x) \geq \epsilon \right\} \to 0$$

for any $\epsilon > 0$ and sufficiently large $N$, where $\rho_{sc}$ is the density of the semicircle law and $M_{\eta^*}(x)$ is the number of eigenvalues of the $N \times N$ Wigner matrices in the interval $[x - \eta^*, x + \eta^*]$. They indeed showed that the decay of the probability is exponential. Note that the supremum inside the probability is taken over $[-2 + \kappa, 2 - \kappa]$, which does not include the boundary $\{-2, 2\}$ of the support.

In addition to the eigenvalue distribution in the macroscopic and microscopic scales, one can investigate the eigenvector behaviors of the random matrix. Erdős, Schlein, and Yau [18] proved that no eigenvector is strongly localized for a general Wigner matrix. More precisely, they proved that, given any small enough $\eta > 0$ and any
Voiculescu [29] discovered that free probability can be used to study the limit of empirical eigenvalue distributions of random matrices. Convergence of moments and Cauchy transforms have been the main tools to study the large-$N$ limit behavior of Hermitian random matrices. Collins and Male [15] proved that if $c_N$ and $d_N$ are random matrices with strong free limits $c$ and $d$ (in the sense that no eigenvalue of $c_N$ (resp. $d_N$) string from the spectrum of $c$ (resp. $d$)), one of them has law independent of unitary conjugation, then, given any polynomial $P$, the sequence $P(c_N, d_N)$ converges strongly in distribution to the $P(c, d)$. In particular, the empirical eigenvalue distributions of random matrices $c_N + d_N$ converges strongly in distribution to the free additive convolution $\mu_c \boxplus \mu_d$ where $\mu_a$ denotes the law of the random variable $a$.

In order to understand properties of $P(c_N, d_N)$ for a general self-adjoint noncommutative polynomial $P$, we use the linearization technique to transform the nonlinear polynomial into a linear polynomial with matrix coefficients. We isolate some regularity properties of the subordination functions that allows us to estimate local behaviors of the empirical eigenvalue distributions of $P(c_N, d_N)$. The idea of the regularity assumption comes from the very smooth plots of the limiting distribution of $P(c, d)$, where $c, d$ are freely independent semicircular random variables, drawn in Belinschi, Mai and Speicher’s paper [21]. They plotted the graphs ([21] Theorem 4.1) by first taking a linearization of the underlying polynomial, computing the (approximate) subordination functions as the $M_n(\mathbb{C})$-valued Denjoy-Wolff point by iterations of a function at $\beta = ze_{1,1} - \gamma_0 + i\eta$, where $\eta > 0$ and $3\varepsilon > 0$ are small. This work suggests that the boundary of the $M_n(\mathbb{C})$-valued subordination functions from the linearization is highly regular.

This paper is organized as follows. In Section 2, we include some background on free probability, operator-valued probability, and linearization. The statements of the existence of the subordination functions and the procedure of linearization to a polynomial in noncommuting variables can be found here. We also include the statement of Newton’s method for functions defined on a Banach space. In Section 3, we estimate the approximate subordination functions of the $M_n(\mathbb{C})$-valued unitarily invariant random matrix model; the estimates are done through Newton’s method. We also show how to make use of the estimates to prove the local limit theorem and delocalization of eigenvectors for polynomials.

2 Preliminary

2.1 Free Probability

Definition 2.1. 1. A $W^*$-probability space is a pair $(\mathcal{A}, \tau)$ where $\mathcal{A}$ is a von Neumann algebra and $\tau$ is a normal, faithful tracial state on $\mathcal{A}$. The elements in $\mathcal{A}$ are called (noncommutative) random variables.

2. The $*$-subalgebras $\mathcal{A}_1, \cdots, \mathcal{A}_n \subseteq \mathcal{A}$ are said to be free or freely independent if, given any $i_1, i_2, \cdots, i_m \in \{1, \cdots, n\}$ with $i_k \neq i_{k+1}$, $a_{i_j} \in \mathcal{A}_{i_j}$ are centered, then we also have $\tau(a_1 a_2 \cdots a_m) = 0$. The random variables $a_1, \cdots, a_m$ are free or freely independent if the $*$-subalgebras they generate are free.

3. For a self-adjoint element $a \in \mathcal{A}$, the law or distribution $\mu$ of $a$ is a probability measure on $\mathbb{R}$ such that whenever $f$ is a bounded continuous function, we have

$$\int_{\mathbb{R}} f \, d\mu = \tau(f(a)).$$

Recall that the Cauchy transform of the law $\mu$ of $a$ on the real line is given by

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z - t} \mu(dt) = \tau((z - a)^{-1})$$
for \( z \notin \text{supp} \mu \). The Cauchy transform can be defined for any finite Borel measure; however, we only use it for probability measures. The transform \( G_\mu \) maps the upper half plane \( \mathbb{H}^+ (\mathbb{C}) \) into the lower half plane \( \mathbb{H}^- (\mathbb{C}) \), and \( \lim_{y \uparrow +\infty} iy G_\mu (iy) = \mu (\mathbb{R}) \). More results of Cauchy transform can be found in [1].

The measure \( \mu \) can be recovered from \( G_\mu \) using the Stieltjes inversion formula, that expresses \( \mu \) as a weak limit

\[
\mu (dx) = \lim_{y \downarrow 0} -\frac{1}{\pi} \Im G_\mu (x + iy) \, dx
\]

The absolutely continuous part of \( \mu \) relative to Lebesgue measure is given by

\[
\frac{d\mu}{dt} (x) = \lim_{y \downarrow 0} -\frac{1}{\pi} \Im G_\mu (x + iy)
\]

and almost everywhere relative to the singular part of \( \mu \),

\[
\lim_{y \downarrow 0} -\frac{1}{\pi} \Im G_\mu (x + iy) = +\infty.
\]

### 2.1.1 Strong Convergence

**Definition 2.2.** Let \( \mathbf{a} = (a_1, \ldots, a_k) \) be a \( k \)-tuple of random variables in the \( W^* \)-probability space \((\mathcal{A}, \tau)\). The joint distribution is the linear form \( P \mapsto \tau [P (\mathbf{a}, \mathbf{a}^*)] \) for all noncommutative polynomials with \( 2k \) noncommuting indeterminates.

1. We say that the \( k \)-tuples \( \mathbf{a}_N = (a_1^{(N)}, \ldots, a_k^{(N)}) \), \( N \in \mathbb{N} \), in \( W^* \)-probability spaces \((\mathcal{A}_N, \tau_N)\) converge in distribution to \( \mathbf{a} \) if

\[
\tau_N [P (\mathbf{a}_N, \mathbf{a}_N^*)] \to \tau [P (\mathbf{a}, \mathbf{a}^*)]
\]

for every noncommutative polynomial \( P \) in \( 2k \) noncommuting indeterminates.

2. We say that the \( k \)-tuples \( \mathbf{a}_N \) converge strongly in distribution if, in addition to the convergence in distribution, we also have

\[
\| P (\mathbf{a}_N, \mathbf{a}_N^*) \| \to \| P (\mathbf{a}, \mathbf{a}^*) \|
\]

for all noncommutative polynomials \( P \) in \( 2k \) noncommuting indeterminates.

Haagerup and Thorbjørnsen [19] proved the strong asymptotic freeness of independent GUE matrices, which are \( N \times N \) Hermitian matrices with independent complex Gaussian entries with variance \( \frac{1}{N} \) in the upper triangular part. Thus, the theorem proved in [19] is an example of strong convergence.

**Example 2.3 ([19]).** For any integer \( N \geq 1 \), let \( X_1^{(N)}, X_2^{(N)}, \ldots, X_p^{(N)} \) be \( N \times N \) independent GUE matrices and let \( (x_1, \ldots, x_p) \) be a semicircular system in a \( W^* \)-probability space with faithful state. Then, almost surely, for all polynomials \( P \) in \( p \) noncommuting indeterminates, one has

\[
\| P (X_1^{(N)}, \ldots, X_p^{(N)}) \| \to \| P (x_1, \ldots, x_p) \|
\]
as \( N \to \infty \), where \( \| \cdot \| \) are the matrix operator norm on the left hand side and the \( C^* \)-algebra norm on the right hand side.

The above example has been generalized to other matrix models; see [2] [14] [23] [24].

By Haar unitary matrices we mean random matrices distributed according to the Haar measure on the unitary group. A non-commutative random variable \( u \in \mathcal{A} \) is called a Haar unitary if it is unitary and \( \tau [u^n] = \delta_{n0} \). Collins and Male [15] proved a strong limit in distribution of independent Haar unitary matrices and (possibly random) matrices that are independent of the Haar unitary.
**Proposition 2.4** (Proposition 2.1, [15]). Let \( x_N = (x_1^{(N)}, \ldots, x_p^{(N)}) \) and \( x = (x_1, \ldots, x_p) \) be \( p \)-tuples of variables in \( C^* \)-probability spaces \((\mathcal{A}_N, \tau_N)\) and \((\mathcal{A}, \tau)\) with faithful states. Then, the followings are equivalent:

1. \( x_N \) converges strongly in distribution to \( x \).

2. for any continuous map \( f_i, g_i : \mathbb{R} \to \mathbb{C}, i = 1, \ldots, p \), the family of variables \((f_i(\mathbb{R}x_i^{(N)}), g_i(\Im x_i^{(N)}))\) converges strongly in distribution to \((f_i(\mathbb{R}x_i), g_i(\Im x_i))\).

3. for any self-adjoint variable \( h_N = P(x_N) \), where \( P \) is a fixed non-commutative polynomials, \( \mu_{h_N} \) converges weakly to \( \mu_h \), where \( h = P(x) \). Moreover, the support of \( \mu_{h_N} \) converges to the support of \( \mu_h \) in the Hausdorff distance.

In particular, the strong convergence in distribution of a single self-adjoint variable is its convergence in distribution together with the Hausdorff convergence of its spectrum.

The random matrices that we consider are \( c_N = C_N \) and \( d_N = U_N D_N U_N^* \) where \( C_N, D_N \) are self-adjoint deterministic diagonal matrices and \( U_N \) is Haar distributed on the unitary group \( U(N) \). The more complicated model in which \( C_N \) and \( D_N \) are random diagonal matrices follows from applying Fubini’s theorem to the laws of the eigenvalue distributions of \( c_N \) and \( D_N \). This model was also considered in, for example, [9, 10, 21].

**2.2 Operator-valued Free Probability**

Voiculescu [31] introduced the concept of \( W^* \)-operator-valued noncommutative probability space. This is a triple \((\mathcal{A}, \phi, B)\) where \( \mathcal{A} \) is a von Neumann algebra, \( B \subseteq \mathcal{A} \) is a von Neumann subalgebra of \( \mathcal{A} \), and \( \phi : \mathcal{A} \to B \) is a unit-preserving conditional expectation. For us, the relevant spaces are \((M_n(\mathbb{C}) \otimes \mathcal{A}, \id \otimes \tau, M_n(\mathbb{C}))\) and \((M_n(\mathbb{C}) \otimes (L^\infty \otimes M_n(\mathbb{C})), \id \otimes \tr_n, M_n(\mathbb{C}))\), where \((\mathcal{A}, \tau)\) is a \( W^* \)-probability space and \( \tr_n \) is the normalized trace on matrices.

When \( x \in \mathcal{A} \), we denote by \( B\langle x \rangle \) the algebra generated by \( B \) and \( x \). Freeness in the operator-valued case is defined as follows.

**Definition 2.5.** Two algebras \( \mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{A} \) containing \( B \) are said to be free over \( B \) if

\[
\phi[x_1 x_2 \cdots x_n] = 0
\]

whenever \( n \in \mathbb{N}, x_j \in \mathcal{A}_j, \phi[x_j] = 0 \) and \( i_j \neq i_{j+1}, \) for \( j = 1, \ldots, n-1 \). Two random variables \( x, y \in \mathcal{A} \) are said to be free over \( B \) if \( B\langle x \rangle \) and \( B\langle y \rangle \) are free over \( B \).

An operator-valued (or \( B \)-valued) random variable \( x \in \mathcal{A} \), the distribution \( \mu_x \) is the set of multilinear maps \( m_n^x : B^{n-1} \to B \) given by

\[
m_n^x(b_1, \ldots, b_{n-1}) = \phi[xb_1xb_2\cdots xb_{n-1}x], \quad b_1, \ldots, b_{n-1} \in B,
\]

for all \( n \in \mathbb{N} \). When \( n = 0, m_0^x \) is the constant \( 1 \in \mathcal{B} \); when \( n = 1, m_1^x = \phi[x] \).

If \( x, y \) are free over \( B \), then the distribution \( \mu_{x+y} \) of \( x + y \) depends only on the distributions \( \mu_x \) and \( \mu_y \). We denote this distribution by \( \mu_x \boxplus \mu_y \).

We use the \( \mathbb{H}^+(B) \) to denote the upper half plane of \( B \)

\[
\mathbb{H}^+(B) = \left\{ b \in B : \Im b = \frac{b^* - b}{2i} \in B \text{ is positive} \right\}.
\]

Similarly, \( \mathbb{H}^-(B) = \left\{ b \in B : \Im b = \frac{b - b^*}{2i} \in B \text{ is negative} \right\} \) denotes the lower half plane of \( B \). The operator-valued Cauchy transform of a variable \( x \in \mathcal{A} \) is defined by

\[
G_x(b) = \phi[(b - x)^{-1}]
\]
for those \( b \in B \) such that \( b - x \) is invertible in \( \mathcal{A} \). If \( b \in \mathbb{H}^+(B) \), then \( b - x \) is invertible and \( G_x(b) \in \mathbb{H}^-(B) \). The operator-valued Cauchy transform and its fully matricial extension is a very powerful tool to study operator-valued distributions. See [7, 31, 32] for more details.

Suppose \((\mathcal{A}, \phi, B)\) is a \( W^*\)-operator-valued non-commutative probability space and \( x, y \in \mathcal{A} \) are freely independent over \( B \). There are two powerful tools to study the distribution \( \mu_x \oplus \mu_y \) of \( x + y \). The first one is the \( \mathcal{R} \)-transform, introduced by Voiculescu [28, 31]. It has been well-studied also in, for example, [16, 27]. The second one is the subordination function. Its existence was proved in the scalar case in [13, 30], and in the operator-valued case in [11, 32]. The theorem is as follows.

**Theorem 2.6.** Assume that \((\mathcal{A}, \phi, B)\) is a \( W^*\)-operator-valued noncommutative probability space and \( x, y \in \mathcal{A} \) are two self-adjoint operator-valued random variables that are free over \( B \). Then there exists a unique pair of Fréchet analytic maps \( \omega_1, \omega_2 : \mathbb{H}^+(B) \to \mathbb{H}^+(B) \) such that

1. \( \exists \omega_j(b) \geq \exists b \) for all \( b \in \mathbb{H}^+(B) \), \( j = 1, 2 \);
2. \( G_x(\omega_1(b)) = G_y(\omega_2(b)) = (\omega_1(b) + \omega_2(b) - b)^{-1} \) for all \( b \in \mathbb{H}^+(B) \).
3. \( G_x(\omega_1(b)) = G_y(\omega_2(b)) = G_{x+y}(b) \) for all \( b \in \mathbb{H}^+(B) \).

### 2.3 Linearization

As in [2, 11], linearization is used to reduce a problem about a polynomial in several variables to an addition of matrices with the variables in their entries. A *linearization* of \( P \in \mathbb{C}\langle X_1, \ldots, X_k \rangle \) is a linear polynomial \( L \) with matrix coefficients \( \gamma_0, \ldots, \gamma_k \),

\[
L = \gamma_0 \otimes 1 + \gamma_1 \otimes X_1 + \ldots + \gamma_k \otimes X_k
\]

satisfying the property that given \( z \in \mathbb{C} \) and \( a_1, \ldots, a_k \) in a \( W^*\)-probability space \( \mathcal{A} \), \( z - P(a_1, \ldots, a_k) \) is invertible in \( \mathcal{A} \) if and only if \( ze_{1,1} \otimes 1 - L(a_1, \ldots, a_k) \) is invertible in \( M_n(\mathcal{A}) = M_n(\mathbb{C}) \otimes \mathcal{A} \), where \( e_{1,1} \) is the matrix element with 1 in the \((1,1)\)-entry and 0 elsewhere. Indeed it can be defined in a more general way by choosing any \( \alpha \in M_n(\mathbb{C}) \), with a different \( L \), \( z - P(a_1, \ldots, a_k) \) is invertible in \( \mathcal{A} \) if and only if \( z\alpha \otimes 1 - L(a_1, \ldots, a_k) \) is invertible in \( M_n(\mathcal{A}) \). Every polynomial possesses a linearization in this sense [25].

Now we describe the linearization process from [2, 22]. The same procedure is also used in [9] to study outliers of a polynomial in unitarily invariant random matrices.

Given a polynomial \( P \in \mathbb{C}\langle X_1, \ldots, X_k \rangle \), the linear polynomial \( L \in M_n(\mathbb{C}\langle X_1, \ldots, X_k \rangle) \) is of the form

\[
L = \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix},
\]

where \( u \in M_{1,(n-1)}(\mathbb{C}\langle X_1, \ldots, X_k \rangle) \), \( v \in M_{(n-1),1}(\mathbb{C}\langle X_1, \ldots, X_k \rangle) \) \( Q \in M_{n-1}(\mathbb{C}\langle X_1, \ldots, X_k \rangle) \). The matrix \( Q \) has to be invertible. The polynomial entries of \( Q \), \( u \) and \( v \) all have degree less than or equal 1. Moreover, \( uQ^{-1}v = -P \). If \( P \) is a self-adjoint polynomial, the coefficients \( \gamma_i \) of \( L \) can be chosen to be self-adjoint matrices.

If \( P \) is a monomial of degree 0 or 1, we set \( n = 1 \) and \( L = P \). If \( P = cX_{i_1}X_{i_2} \ldots X_{i_l} \), where \( l \geq 2 \) and \( i_1, \ldots, i_l \in \{1, \ldots, k\} \), then the matrix is of size \((l + 1) \times (l + 1)\) and

\[
L = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 & c \\
0 & 0 & \cdots & 0 & X_{i_1} & -1 \\
0 & 0 & \cdots & X_{i_2} & -1 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & X_{i_{l-1}} & \cdots & 0 & 0 & 0 \\
X_{i_l} & -1 & \cdots & 0 & 0 & 0
\end{pmatrix}
\]
The lower right \( l \times l \) matrix has an inverse of degree \( l - 1 \) in \( M_{l-1}(\mathbb{C}(X_1, \ldots, X_k)) \) \([22]\). Suppose now that \( P = P_1 + P_2 \), where \( P_1 \in \mathbb{C}(X_1, \ldots, X_k) \) with linearization

\[
L_i = \begin{pmatrix} 0 & u_i \\ v_i & Q_i \end{pmatrix} \in M_{n_i}(\mathbb{C}(X_1, \ldots, X_k)).
\]

Then

\[
L = \begin{pmatrix} 0 & u_1 & u_2 \\ v_1 & Q_1 & 0 \\ v_2 & 0 & Q_2 \end{pmatrix} = \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix} \in M_{n_1+n_2-1}(\mathbb{C}(X_1, \ldots, X_k))
\]

is a linearization of \( P \). Hence, every \( P \) possesses a linearization.

If \( P \) is a self-adjoint polynomial, then \( P = P_0 + P_0^* \) for some polynomial \( P_0 \). Suppose

\[
L_0 = \begin{pmatrix} 0 & u_0 \\ v_0 & Q_0 \end{pmatrix}
\]

is a linearization of \( P_0 \), then

\[
L_i = \begin{pmatrix} 0 & u_i \\ v_i & Q_i \end{pmatrix} \in M_{n_i}(\mathbb{C}(X_1, \ldots, X_k)).
\]

Then

\[
L = \begin{pmatrix} 0 & u_0 & u_0^* \\ u_0^* & 0 & Q_0^* \\ v_0 & Q_0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & u^* \\ u & Q \end{pmatrix}
\]

is a self-adjoint linear polynomial for \( P \). The constant term of \( Q^{-1} \) has spectrum contained in \( \{1, -1\} \) \([22]\).

More properties of this linearization process were proved in \([9\), Section 4\]. The following lemma, which gives an estimate of norm between the polynomial \( P \) and the linearization \( L \), is of particular interest to us.

**Lemma 2.7** (Lemma 4.3 \([9]\)). Suppose that \( P = P^* \in \mathbb{C}(X_1, \ldots, X_k) \), and let \( L \) be a linearization of \( P \) constructed with the properties above. There exist polynomials \( T_1, T_2 \) with nonnegative coefficients with the following property: given arbitrary selfadjoint elements \( S_1, \ldots, S_k \) in a unital \( C^*-\)algebra \( \mathcal{A} \), and given \( z_0 \in \mathbb{C} \) such that \( z_0 - P(S) \) is invertible, we have

\[
\| (z_0 e_{1,1} - L(S))^{-1} \| \leq T_1(\| S_1 \|, \ldots, \| S_k \|) \| z_0 - P(S)^{-1} \| + T_2(\| S_1 \|, \ldots, \| S_k \|).
\]

In particular, given constants \( C, \delta > 0 \), there exists \( \varepsilon > 0 \) such that \( \text{dist}(z_0, \sigma(P(S))) \geq \delta \) and \( \| S_1 \| + \ldots + \| S_k \| \leq C \) imply \( \text{dist}(0, \sigma(z_0 e_{1,1} - L(S))) \geq \varepsilon \).

### 2.4 Newton’s Method

Newton’s Method provides a way to locate a zero from an initial guess; it also gives an estimate about the distance between the zero and the initial guess. It has been generalized into different setups to fit the needs of different applications. The one that we need is to locate a zero of a nonlinear operator on a Banach space.

Let \( F \) be a nonlinear operator on a domain \( G \) in a Banach space \( X \). To use Newton’s method to locate the solution \( F(x) = 0 \), we first make an initial guess \( y_0 \). If \( F \) is twice differentiable, and \( F' \) is invertible in a neighborhood of \( y_0 \), then the iterations

\[
y_{n+1} = y_n - [F'(y_n)]^{-1}F(y_n)
\]

converge as \( n \to \infty \) to the solution of \( F(x) = 0 \) under the sufficient conditions of the following theorem due to Kantorovich \([20]\).
Theorem 2.8 (Kantorovich). Assume, in the framework described above, there are constants $C_0$, $\delta_0$ and $A$ such that the following conditions hold:

1. The operator $F'(y_0)$ is invertible whose inverse has norm $\|F'(y_0)^{-1}\| \leq C_0$;
2. $\|F'(y_0)^{-1}F(y_0)\| \leq \delta_0$;
3. the second derivative $F''(y)$, $y \in G$ is bounded in $G$ by $A$, namely $\|F''(y)\| \leq A$;
4. the constants $C_0$, $\delta_0$, $A$ satisfy the inequality $h_0 = C_0\delta_0A \leq \frac{1}{2}$.
5. $B\left(y_0, \frac{1-\sqrt{1-2h_0}}{h_0}\delta_0\right) \subseteq G$.

Then the equation $F(w) = 0$ has a unique solution $w^*$ in the ball $B\left(y_0, \frac{1-\sqrt{1-2h_0}}{h_0}\delta_0\right)$ centered at our initial approximation $w_0$ and the iterations of the Newton’s method $w_n \rightarrow w^*$ as $n \rightarrow \infty$.

3 The Unitarily Invariant Model

3.1 The Sum of Two $M_n(\mathbb{C})$-valued Matrices

Fix two sequences $\{C_N\}_{N \in \mathbb{N}}$ and $\{D_N\}_{N \in \mathbb{N}}$ of deterministic diagonal matrices $C_N, D_N \in M_N(\mathbb{R})$. Suppose that both sequences have a strong limiting (deterministic) empirical eigenvalue distribution. Let $\{U_N\}_{N \in \mathbb{N}}$ be a sequence of $N \times N$ unitary matrices distributed uniformly on $U(N)$.

We view $d_N = U_N D_N U_N^*$ and $c_N = C_N$ as random matrices in the space $M_N(\mathbb{C}) \otimes L^\infty$. Let $(\mathcal{A}, \tau)$ be a $W^*$-probability space and let $c, d \in \mathcal{A}$ be freely independent self-adjoint variables such that $(c_N, d_N) \rightarrow (c, d)$ strongly in distribution. The $M_n(\mathbb{C})$-valued probability spaces $(M_n(\mathbb{C}) \otimes (M_N(\mathbb{C}) \otimes L^\infty), M_n(\mathbb{E}_N), M_n(\mathbb{C}))$, where $\mathbb{E}$ is the expectation, fit in the framework of operator-valued noncommutative probability space introduced in Section 2.2. We look at the “limit” $M_n(\mathbb{C})$-valued probability space $(M_n(\mathcal{A}), M_n(\tau), M_n(\mathbb{C}))$.

We focus on the $M_n(\mathbb{C})$-valued unitarily invariant matrix model

$$H_N = \gamma_1 \otimes c_N + \gamma_2 \otimes d_N \in M_n(\mathcal{A})$$

where $\gamma_1, \gamma_2$ are Hermitian matrices in $M_n(\mathbb{C})$. (Later we apply the results in this section to a linearization of a noncommutative polynomial $P(c_N, d_N)$.)

Notation 3.1. We use the following notations throughout the rest of this paper:

1. $\mathbb{E}$ is the expectation of a random variable (in a probability space). We also use $\mathbb{E}$ to mean taking expectation entrywise in a matrix; $\mathbb{E}$ is completely positive.
2. $H_N = \gamma_1 \otimes c_N + \gamma_2 \otimes d_N$; and $H = \gamma_1 \otimes c + \gamma_2 \otimes d$.
3. $R_N(\beta) = (\beta \otimes I_N - H_N)^{-1}$ is the resolvent of $H_N$, where $\beta \in \mathbb{H}^+(M_n(\mathbb{C}))$.
4. The (random $M_n(\mathbb{C})$-valued) Cauchy transform $G_{H_N}(\beta) = M_n(\mathbb{E}_N)(R_N(\beta))$. We also write the (random) Cauchy transform $G_{c_N}(\beta) = M_n(\mathbb{E}_N)((\beta \otimes I_N - \gamma_1 \otimes c_N)^{-1})$ and a similar definition for $d_N$.
5. We write $G_H(\beta) = M_n(\tau)((\beta \otimes I_N - H)^{-1})$, $\beta \in \mathbb{H}^+(M_n(\mathbb{C}))$, and $R(\beta) = (\beta \otimes I_N - H)^{-1}$. We sometimes write $\beta$ in place of $\beta \otimes I_N$ when context is clear.
6. We define approximate subordination functions, for $\beta \in \mathbb{H}^+(M_n(\mathbb{C}))$ by
\[ \begin{align*}
\omega_1^{(N)}(\beta) &= \beta - (\mathbb{E} G_{H_N}(\beta))^{-1}(\mathbb{E} f_1^{(N)}(\beta)), \\
\omega_2^{(N)}(\beta) &= \beta - (\mathbb{E} G_{H_N}(\beta))^{-1}(\mathbb{E} f_2^{(N)}(\beta))
\end{align*} \]

where \( f_1^{(N)}(\beta) = M_n(\text{tr}_N)(R_N(\beta) \cdot (\gamma_2 \otimes d_N)) \) and \( f_2^{(N)}(\beta) = M_n(\text{tr}_N)(R_N(\beta) \cdot (\gamma_1 \otimes e_N)). \)

These are analogs of definitions from [21]. We observe that

\[ f_1^{(N)}(\beta) + f_2^{(N)}(\beta) = M_n(\text{tr}_N)((H_N - \beta)R_N(\beta)) + M_n(\text{tr}_N)(\beta R_N(\beta)) = -I_N + \beta G_{H_N}(\beta) \]

and

\[ \omega_1^{(N)}(\beta) + \omega_2^{(N)}(\beta) = \beta + (\mathbb{E} G_{H_N}(\beta))^{-1}. \]

The approximate subordination functions satisfy the same equation as the subordination functions of the free random variables in the limit, as \( N \to \infty. \) We will show that these functions are “almost” self-map of the upper half plane \( \mathbb{H}^+(M_n(\mathbb{C})) \) (See Proposition 3.6) and that they converge to the subordination functions of the free random variables (See equation (3.7) and [9, Proposition 8.8]). An explicit estimate of these approximate subordination functions is given in Proposition 3.11. We use them when we approximate the Cauchy transform \( G_{H_N}(\beta) \) when \( \beta \) is very close to the boundary \( \partial \mathbb{H}^+(M_n(\mathbb{C})) \) of the upper half plane \( \mathbb{H}^+(M_n(\mathbb{C})). \)

From now on, we simply write \( \text{tr} \) instead of \( \text{tr}_N \) for the normalized trace for \( N \times N \) matrices.

**Lemma 3.2.** We have

\[ \mathbb{E}(G_{H_N}(\beta) \otimes I_N \otimes \gamma_2 \otimes d_N \cdot R_N(\beta)) = \mathbb{E}[f_1^{(N)}(\beta)R_N(\beta)]. \]

**Proof.** By [9] Proof of Lemma 8.1, for any \( Z \in M_N(\mathbb{C}), \)

\[ \mathbb{E}(R_N(\beta)[I_n \otimes Z, \gamma_2 \otimes d_N]R_N(\beta)) = 0 \quad (3.1) \]

where the \([\cdot, \cdot]\) is the commutator.

The matrix \( R_N(\beta) \in M_n(\mathbb{C}) \otimes L^\infty \) is of the form \( \sum_{j_1,j_2=1}^n e_{j_1,j_2} \otimes A_{j_1,j_2}, \) where \( e_{j_1,j_2} \) is the \( n \times n \) matrix with 1 in the \((j_1,j_2)\)-entry and 0 elsewhere. We take \( Z = e_{ab}. \) Since \((Zd_N)_{xy} = \delta_{xa}(d_N)_{by}\) and \((d_NZ)_{xy} = \delta_{yb}(d_N)_{xa},\) the \((u, v)\)-entries of \( A_j(Zd_N)A_k \) and \( A_j(d_NZ)A_k \) are \((A_j)_{ua}(d_NA_k)_{vb}\) and \((A_jd_N)_{ua}(A_k)_{vb}\) respectively. Expanding the matrix multiplication gives

\[ R_N(\beta) \cdot \gamma_2 \otimes (Zd_N) \cdot R_N(\beta) = \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \otimes A_{j_1,j_2}(Zd_N)A_{k_1,k_2} \]

and similarly

\[ R_N(\beta) \cdot \gamma_2 \otimes (d_NZ) \cdot R_N(\beta) = \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \otimes A_{j_1,j_2}(d_NZ)A_{k_1,k_2}. \]

By (3.1), the expectation of above two quantities are equal, so

\[ \mathbb{E} \left[ \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \cdot (A_{j_1,j_2})_{ua}(d_NA_{k_1,k_2})_{vb} \right] = \mathbb{E} \left[ \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \cdot (A_{j_1,j_2}d_N)_{ua}(A_{k_1,k_2})_{vb} \right]. \]

We first take \( u = a, \) then sum over all \( a = 1, 2, \ldots, N \) and divide by \( N; \) we have

\[ \mathbb{E} \left[ \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \cdot \text{tr}(A_{j_1,j_2})(d_NA_{k_1,k_2})_{vb} \right] = \mathbb{E} \left[ \sum_{j_1,j_2,k_1,k_2} e_{j_1,j_2} \gamma_2 e_{k_1,k_2} \cdot \text{tr}(A_{j_1,j_2}d_N)(A_{k_1,k_2})_{vb} \right]. \]
Since this is true for all \( b, v \), we see that

\[
\mathbb{E} \left[ \sum_{j_1, j_2, k_1, k_2} e_{j_1, j_2} \gamma_2 e_{k_1, k_2} \otimes \text{tr}(A_{j_1, j_2})(d_N A_{k_1, k_2}) \right] = \mathbb{E} \left[ \sum_{j_1, j_2, k_1, k_2} e_{j_1, j_2} \gamma_2 e_{k_1, k_2} \otimes \text{tr}(A_{j_1, j_2}d_N A_{k_1, k_2}) \right].
\]

Thus

\[
\mathbb{E}(G_{H_N}(\beta) \gamma_2 \otimes d_N R_N(\beta)) = \mathbb{E} \left[ \sum_{j_1, j_2} e_{j_1, j_2} \otimes (\text{tr}(A_{j_1, j_2})I_N) \cdot \gamma_2 \otimes d_N \cdot \sum_{k_1, k_2} e_{k_1, k_2} \otimes A_{k_1, k_2} \right]
\]

\[
= \mathbb{E} \left[ \sum_{j_1, j_2, k_1, k_2} e_{j_1, j_2} \gamma_2 e_{k_1, k_2} \otimes \text{tr}(A_{j_1, j_2})(d_N A_{k_1, k_2}) \right]
\]

\[
= \mathbb{E} \left[ \sum_{j_1, j_2, k_1, k_2} e_{j_1, j_2} \gamma_2 e_{k_1, k_2} \otimes \text{tr}(A_{j_1, j_2}d_N A_{k_1, k_2}) \right]
\]

\[
= \mathbb{E} \left[ \left( \sum_{j_1, j_2} e_{j_1, j_2} \gamma_2 \otimes \text{tr}(A_{j_1, j_2}d_N)I_N \right) \left( \sum_{k_1, k_2} e_{k_1, k_2} \otimes A_{k_1, k_2} \right) \right]
\]

\[
= \mathbb{E} [M_n(\text{tr})(R_N(\beta) \cdot \gamma_2 \otimes d_N) \otimes I_N \cdot R_N(\beta)]
\]

\[
= \mathbb{E}[f_1^{(N)}(\beta)R_N(\beta)],
\]

and the lemma follows. \( \square \)

The subordination function \( \omega_1 \) is given by

\[
\omega_1(\beta) \otimes 1 = M_n(\tau)[R(\beta)]^{-1} + \gamma_1 \otimes c.
\]

To imitate the behavior of \( \omega_1 \) in the case of finite dimensional matrices, we try to look at \( \mathbb{E}[R_N(\beta)]^{-1} + \gamma_1 \otimes c_N \).
However, this function is no longer of the form \( M_n(\mathbb{C}) \otimes I_N \) because \( c_N \) and \( d_N \) are not free.

We next estimate how far \( \omega_1^{(N)} \) is from \( \mathbb{E}[R_N(\beta)]^{-1} + \gamma_1 \otimes c_N \). The error term \( \delta_1(\beta) \) is defined to be

\[
\delta_1(\beta) = \mathbb{E}[G_{H_N}(\beta)]^{-1} \mathbb{E} \Delta_1(\beta)
\]

(3.3)

where

\[
\Delta_1(\beta) = \mathbb{E}[(f_1^{(N)} - \mathbb{E}f_1^{(N)}) R_N(\beta)] - \mathbb{E}[(G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta))(\beta - \gamma_1 \otimes c_N) R_N(\beta)].
\]

(3.4)

The definition of \( \Delta_1 \) allows us to approximate by applying concentration inequality to \( f_1^{(N)} - \mathbb{E}f_1^{(N)} \) and \( G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta) \). We also define \( \delta_2 \) and \( \Delta_2 \) analogously, by replacing \( \gamma_1 \otimes c_N \) by \( \gamma_2 \otimes d_N \) and \( f_1^{(N)} \) by \( f_2^{(N)} \). We use the same notation as above.

**Proposition 3.3.** We have

\[
(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N) \mathbb{E}[R_N(\beta)] = 1 + \delta_1(\beta).
\]

(3.5)
Remark 3.4. We will prove that under extra assumptions in Proposition 3.6, $\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N$ is indeed invertible, and $\omega_1^{(N)}(\beta)$ is in $\mathbb{H}^+(M_n(\mathbb{C}))$.

Proof. By the identity $(\beta - \gamma_1 \otimes c_N)R_N(\beta) = 1 + (\gamma_2 \otimes d_N)R_N(\beta)$, multiplying $G_H(\beta) \otimes I_N$ and taking expectation give us

$$
E[G_H(\beta) \otimes I_N \cdot (\beta - \gamma_1 \otimes c_N) \cdot R_N(\beta)] = E[G_H(\beta) \otimes I_N] + E[G_H(\beta) \otimes I_N \cdot \gamma_2 \otimes d_N \cdot R_N(\beta)]
$$

where the last equality follows from Lemma 3.2. Now the left hand side is

$$
E[(G_H(\beta) - E[G_H(\beta)])(\beta - \gamma_1 \otimes c_N)R_N(\beta)] + E[G_H(\beta)]E[(\beta - \gamma_1 \otimes c_N)R_N(\beta)]
$$

while the right hand side is

$$
E[G_H(\beta)] + E[(f_1^{(N)}(\beta) - E[f_1^{(N)}(\beta)])R_N(\beta)] + E[f_1^{(N)}(\beta)]E[R_N(\beta)].
$$

Rearranging the terms gives

$$
E[G_H(\beta)]E[(\beta - \gamma_1 \otimes c_N - (EG_H(\beta))^{-1}(E[f_d_N(\beta)]) \cdot R_N(\beta)]
$$

$$
= E[G_H(\beta)] + E[(f_1^{(N)}(\beta) - E[f_1^{(N)}(\beta)]) \cdot R_N(\beta)] - E[(G_H(\beta) - EG_H(\beta)) \cdot (\beta - \gamma_1 \otimes c_N) \cdot R_N(\beta)].
$$

Recall that $\omega_1^{(N)}(\beta) = \beta - (EG_H(\beta))^{-1}(E[f_1^{(N)}(\beta)]$; it follows that

$$
(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)E[R_N(\beta)] = 1 + \delta_1(\beta)
$$

where $\Delta_1(\beta)$ is defined in 3.4. \hfill \Box

Assume that $\eta \in (0, 1)$ and $z \in K$ for some bounded set $K \subseteq \mathbb{H}^+(\mathbb{C})$. Using linearization technique, we are particularly interested in $\beta = ze_{1,1} - \gamma_0 + i\eta$. Its real part is bounded uniformly for all $z \in K$ and $\|z\beta^{-1}\| = \frac{1}{\eta}$ since $z \in \mathbb{H}^+(\mathbb{C})$. We look more generally at $\{\beta \in \mathbb{H}^+(M_n(\mathbb{C})) : \|\beta\| \leq K, \|z\beta^{-1}\| = \frac{1}{\eta}\}$ where $\eta \in (0, 1)$, and $K > 0$.

Lemma 3.5. Suppose that $\beta \in \mathbb{H}^+(M_n(\mathbb{C}))$, $\|\beta\| \leq K$, and $\|z\beta^{-1}\| = \frac{1}{\eta} > 1$. Then

$$
\|E[\omega_1^{(N)}(\beta)]\| \leq C \left( \frac{1}{N^2 \eta^3} \right)
$$

where $C$ depends only on $K$, $\gamma_1$, $\gamma_2$, $\|c\|$, $\|d\|$.

Proof. By [7] Equation 13,

$$
\|R_N(\beta)\| \leq \|z\beta^{-1}\| = \frac{1}{\eta}.
$$

We estimate $\Delta$ term-by-term. For the second term $E[(G_H(\beta) - EG_H(\beta)) \cdot (\beta - \gamma_1 \otimes c_N) \cdot R_N(\beta)]$, we have

$$
\|E[(G_H(\beta) - EG_H(\beta)) \cdot (\beta - \gamma_1 \otimes c_N) \cdot R_N(\beta)]\| \leq \|E[(G_H(\beta) - EG_H(\beta)) \cdot (\beta - \gamma_1 \otimes c_N)]\| \cdot \|R_N(\beta)\|
$$

$$
\leq \frac{C(K, \gamma_1, \gamma_2, c_N, d_N)}{\eta} \|G_H(\beta) - EG_H(\beta)\|.
$$
By concentration inequality, and by the fact that \( G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta) = M_n(\text{tr})(R_N(\beta) - \mathbb{E}(R_N(\beta))) \) has rank bounded by \( n \) which is independent of \( N \), we obtain, as in [9] Proof of Proposition 8.4,

\[
\mathbb{P}(\|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| > \varepsilon) \leq 2 \exp\left(\frac{-N\varepsilon^2}{8r^4\|\gamma\|^2}\right)
\]

for any \( \varepsilon > 0 \) and \( \alpha \in (0, \frac{1}{2}) \), where \( r = \sup_N \|R_N(\beta)\| \) (The norm is taken over \( M_n(\mathbb{C}) \otimes L^\infty \), which is a positive constant, not a random norm). Since \( r = O(1/\eta) \), we have

\[
\mathbb{P}(\|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)(\beta - \gamma_1 \otimes c_N)R_N(\beta)\| > \varepsilon) \leq 2 \exp\left(\frac{-\varepsilon^2CN\eta^6}{\|\gamma\|^2}\right).
\]

For the first term \( (f^{(N)}_1 - \mathbb{E}f^{(N)}_1)R_N(\beta) \), since \( f_{d_N}(\beta) = -I_n + M_n(\text{tr})(\beta - \gamma_1 \otimes c_N)R_N(\beta) \), we have

\[
(f_{d_N} - \mathbb{E}f_{d_N}) = M_n(\text{tr})(\beta - \gamma_1 \otimes c_N)(R_N(\beta) - \mathbb{E}R_N(\beta)) = R_N(\beta) - \mathbb{E}R_N(\beta) - M_n(\text{tr})(\gamma_1 \otimes c_N)(R_N(\beta) - \mathbb{E}R_N(\beta)).
\]

The terms \( M_n(\text{tr})[R_N(\beta) - \mathbb{E}R_N(\beta)] \) and \( M_n(\text{tr})[(\gamma_1 \otimes c_N)(R_N(\beta) - \mathbb{E}R_N(\beta))] \) are linear transformations of \( R_N(\beta) - \mathbb{E}R_N(\beta) \) with at most rank \( n \). Hence,

\[
\mathbb{P}\left(\|f^{(N)}_1 - \mathbb{E}f^{(N)}_1\| > \varepsilon\right) \leq 2 \exp\left(\frac{-\varepsilon^2CN\eta^6}{\|\gamma\|^2}\right).
\]

Using triangle inequality, we have

\[
\mathbb{P}(\|\Delta_1\| > \varepsilon) \leq 4 \exp\left(\frac{-CN\eta^6\varepsilon^2}{\|\gamma\|^2}\right).
\]

Thus

\[
\|\mathbb{E}\Delta_1(\beta)\| \leq \mathbb{E}\|\Delta_1(\beta)\| \leq \int_0^\infty 4 \exp\left(\frac{-CN\eta^6t^2}{\|\gamma\|^2}\right) dt = O\left(\frac{1}{N^\frac{1}{2}\eta^\theta}\right),
\]

the lemma is proved.

We now would like to consider the map \( \beta \mapsto \mathbb{E}[G_{H_N}(\beta)]^{-1} \) for \( \beta \in \mathbb{H}^+(M_n(\mathbb{C})) \). Analogue to [10] where the completely positive map \( \mathbb{E} \) is considered, we consider

\[
\tilde{F}(\varepsilon) = ((M_n(\text{tr})\mathbb{E})[(\mathbb{H}\beta - \varepsilon(H_N - \Re \beta)^{-1})]^{-1} = z \mathbb{H}\beta + \Re \beta - M_n(\text{tr})[\mathbb{E}(H_N)]
\]

\[
-\frac{1}{z} (M_n(\text{tr})\mathbb{E})[H_N - (M_n(\text{tr})\mathbb{E})(H_N)](\mathbb{H}\beta^{-1})(H_N - (M_n(\text{tr})\mathbb{E})(H_N)) + O\left(\frac{1}{z}\right).
\]

Hence,

\[
F(z) = ((M_n(\text{tr})\mathbb{E})[(\Re \beta + z\mathbb{H}\beta - H_N)^{-1}]^{-1} = A + Bz - \int_\Re \frac{\rho(dt)}{z - t}
\]

where \( A = \Re \beta - (M_n(\text{tr})\mathbb{E})(H_N) \), \( B = \mathbb{H}\beta \) and

\[
\rho(\Re) = \lim_{z \to \infty} z(A + Bz - F(z)) = (M_n(\text{tr})\mathbb{E})[(H_N - (M_n(\text{tr})\mathbb{E})(H_N))(\mathbb{H}\beta^{-1})(H_N - (M_n(\text{tr})\mathbb{E})(H_N))].
\]

It follows that

\[
\|(\mathbb{E}G_{H_N}(\beta))^{-1}\| \leq \|\beta\| + \|M_n(\text{tr})\mathbb{E}(H_N)\| + \|\rho(\Re)\| = O\left(\frac{1}{\eta}\right)
\]

12
where constant of the $O$ in last equality is for $\Im \beta \geq \eta$, with $\eta \in (0, 1)$ and depends on $\|\beta\| \leq K$.

Hence, if $\|\beta\| \leq K$ and $\Im \beta \geq \eta \in (0, 1)$, Lemma 3.3 shows that

$$\|((\mathbb{E}G_{H_N}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))\| = O \left( \frac{1}{N^{\frac{1}{4}\eta^4}} \right).$$  (3.6)

The preceding observation shows that $\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N$ is invertible as long as $\|((\mathbb{E}G_{H_N}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))\|$ is small, because the invertible matrices form an open set. When $\|((\mathbb{E}G_{H}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))\|$ is small, we can write (3.3) as

$$\mathbb{E}[R_N(\beta)] = (\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1} \left( I + \mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}\Delta_1(\beta) \right).$$

Hence $\mathbb{E}[R_N(\beta)]^{-1} = \left( I + \mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}\Delta_1(\beta) \right)^{-1} (\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)$ which shows

$$\omega_1^{(N)}(\beta) = \mathbb{E}[R_N(\beta)]^{-1} + \gamma_1 \otimes c_N - \left[ (I + \mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}\Delta(\beta))^{-1} - I \right] (\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N).$$  (3.7)

The above formula shows the difference between this definition of $\omega$ and the choice from [9]. This formula will be used as a measure of how far $\omega$ is from being a self-map on the upper half plane of $M_n(\mathbb{C})$. Then, we estimate the difference of $\omega_1^{(N)}$ from $\mathbb{E}[R_N(\beta)]^{-1} + \gamma_1 \otimes c_N$. The advantages of our choice are that the $\omega$’s are already in $M_n(\mathbb{C})$ which allows us to estimate the Cauchy transforms $G_{c_N}(\omega_1^{(N)}(\beta))$ and $G_{d_N}(\omega_2^{(N)}(\beta))$; that the functions $\omega_1^{(N)}$ and $\omega_2^{(N)}$ satisfy the same equation as the subordination functions also allows us to apply Newton’s method.

The following proposition gives an estimate of the imaginary part of $\omega_1^{(N)}(\beta)$, in addition to the invertibility of $\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N$.

**Proposition 3.6.** Suppose $\beta \in \mathbb{H}^+(\mathbb{C})$, $\|\beta\| \leq K$ and $\|\Im(\beta)^{-1}\| = \frac{1}{\eta}$ $\eta \in (0, 1)$. If $N^{\frac{1}{2}}\eta^4$ is large enough, then

$$\Im(\omega_1^{(N)}(\beta)) \geq \Im \beta - \frac{c}{N^{\frac{1}{2}}\eta^6}$$

for some positive constant $c$. In particular, if $N^{\frac{1}{2}}\eta^6$ is large enough, $\omega_1^{(N)}(\beta) \in \mathbb{H}^+(M_n(\mathbb{C}))$.

**Proof.** Suppose $N^{\frac{1}{2}}\eta^4$ is large enough such that $\|((\mathbb{E}G_{H}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))\| < \frac{1}{2}$. By power series expansion, we know that if $\|X\| \leq \varepsilon < \frac{1}{2}$, then $\|(I + X)^{-1} - I\| \leq 2\varepsilon$. We have

$$\|(I + (\mathbb{E}G_{H}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))^{-1} - I\| = O \left( \frac{1}{N^{\frac{1}{2}}\eta^4} \right).$$

Recall that $f_1^{(N)}(\beta) = M_n(tr)(R_N(\beta) \cdot \gamma_2 \otimes d_N)$. We have the estimate

$$\|\mathbb{E} f_1^{(N)}(\beta)\| \leq \|\mathbb{E} f_1^{(N)}(\beta)\| \leq \gamma_2 \otimes d_N \cdot \mathbb{E}\|R_N(\beta)\| = O \left( \frac{1}{\eta} \right).$$

Hence if $\beta \in \mathbb{H}^+(M_n(\mathbb{C}))$, $\|\beta\| \leq K$ and $\|\Im(\beta)^{-1}\| = \frac{1}{\eta}$, $\eta \in (0, 1)$, then

$$\|\omega_1^{(N)}(\beta)\| = \|\beta - (\mathbb{E}G_{H_N}(\beta))^{-1}\mathbb{E} f_1^{(N)}(\beta)\| = O \left( \frac{1}{\eta^2} \right).$$

13
It follows that
\[
\|(I + (\mathbb{E}G_{H_N}(\beta))^{-1}\mathbb{E}\Delta_1(\beta))^{-1} - I)(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)\| = O\left(\frac{1}{N^{\frac{1}{2}}\eta^6}\right).
\]

Since \(\mathfrak{S}(\mathbb{E}R_N(\beta))^{-1} \geq \mathfrak{S}(\beta)\), we have, by looking at the expression (3.7),
\[
\mathfrak{S}(\omega_1^{(N)}(\beta)) \geq \mathfrak{S}\beta - \frac{c}{N^{\frac{1}{2}}\eta^6},
\]
for some positive constant \(c\).

If \(N^{\frac{1}{2}}\eta^6\) large enough, \(\omega_1^{(N)}(\beta) \in \mathbb{H}^+(M_n(\mathbb{C}))\) because \(\mathfrak{S}\beta \geq \epsilon > 0\) for some \(\epsilon\). \(\square\)

**Theorem 3.7.** Suppose \(\|\beta\| \leq K, \beta \in \mathbb{H}^+(\mathbb{C}), \|\mathfrak{S}\beta^{-1}\| = \frac{1}{\eta}, \eta \in (0, 1)\). If \(N^{\frac{1}{2}}\eta^4\) is large enough, then the following estimates hold:

1. \(\|\mathbb{E}R_N(\beta) - (\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1}\| = O\left(\frac{1}{N^{\frac{1}{2}}\eta^6}\right)\).
2. \(\|\mathbb{E}G_H(\beta) - M_n(\operatorname{tr})(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1}\| = O\left(\frac{1}{N^{\frac{1}{2}}\eta^6}\right)\).

**Proof.** Equation (3.5) says
\[
(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1} = \mathbb{E}[R_N(\beta)] (I + \mathbb{E}[G_{H}(\beta)]^{-1}\mathbb{E}[-\Delta])^{-1}.
\]

By Equation (3.6), \(\|\mathbb{E}G_N(\beta)^{-1}\mathbb{E}\Delta_1(\beta)\| = O\left(\frac{1}{N^{\frac{1}{2}}\eta^4}\right)\). If \(\|\mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}[\Delta_1(\beta)]\| < \frac{1}{2}\), then using power series expansion to \((I + \mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}[\Delta_1(\beta)])^{-1}\), we have
\[
\|(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1}\| = O\left(\frac{1}{\eta}\right).
\]

It follows that the error term \(\delta_1(\beta)\) from proposition 3.3 satisfies
\[
\|\mathbb{E}R_N(\beta) - (\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1}\| \leq \|(\omega_1^{(N)}(\beta) - \gamma_1 \otimes c_N)^{-1}\| \|\mathbb{E}[G_{H_N}(\beta)]^{-1}\mathbb{E}\Delta_1(\beta)\| = O\left(\frac{1}{N^{\frac{1}{2}}\eta^5}\right),
\]
which is the first assertion. The second assertion follows directly from \(M_n(\operatorname{tr})\) is continuous. \(\square\)

We now extend the notion of smoothness in [21] to \(M_n(\mathbb{C})\)-valued case. The following reduces to the smoothness condition in [21] when \(n = 1\).

**Definition 3.8.** Suppose that \(c, d\) are freely independent random variables, with corresponding distributions \(\mu_c\) and \(\mu_d\), and \(\gamma_1, \gamma_2 \in M_n(\mathbb{C})\) are self-adjoint matrices. Let
\[
H = \gamma_1 \otimes c + \gamma_2 \otimes d \in M_n(\mathfrak{A})
\]
be a sum of \(M_n(\mathbb{C})\)-valued free variables, equipped with the conditional expectation \(M_n(\tau)\).

We say that the triple \((\mu_c, \mu_d, H)\) is regular on \((S, \mathbb{V})\), where \(S \subseteq \partial \mathbb{H}^+(M_n(\mathbb{C}))\), \(\mathbb{V}\) is a subspace of Hermitian matrices, if the restrictions of \(\omega_1\) and \(\omega_2\) to \(\{x + iy \in \mathbb{H}^+(M_n(\mathbb{C})) : x \in S, y \in \mathbb{V}\}\) extend continuously to \(\partial\{x + iy \in \mathbb{H}^+(M_n(\mathbb{C})) : x \in S, y \in \mathbb{V}\}\) and if the triple \((\mu_c, \mu_d, H)\) satisfies the following two conditions:
A. For all $\bar{\beta} \in \bar{S}$, $\omega_1(\bar{\beta}) - t\gamma_1$ is invertible for every $t \in \text{supp } \mu_c$ and $\omega_2(\bar{\beta}) - t\gamma_2$ is invertible for every $t \in \text{supp } \mu_d$.

B. For each $(w_1, w_2) := (\omega_1(\bar{\beta}), \omega_2(\bar{\beta}))$, for $\bar{\beta} \in \bar{S}$, the map from $M_n(\mathbb{C}) \times M_n(\mathbb{C})$ to $M_n(\mathbb{C}) \times M_n(\mathbb{C})$ given by

$$(b_1, b_2) \mapsto \left(\begin{array}{c}
-(w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1} + \int_{\mathbb{R}}(w_1 - t\gamma_1)^{-1}b_1(w_1 - t\gamma_1)^{-1}\mu_c(dt) \\
-(w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1} + \int_{\mathbb{R}}(w_2 - t\gamma_2)^{-1}b_2(w_2 - t\gamma_2)^{-1}\mu_d(dt)
\end{array} \right)$$

is invertible.

**Remark 3.9.** We will take $S = \{z \cdot e_{1,1} - \gamma_0 : z = x + iy, x \in I, y \in (0, \eta]\}$ for a fixed interval and $V = \mathbb{R}$ in Section 3.2 in which we apply the linearization technique to polynomials.

Given two probability measures $\mu$ and $\nu$, we use the Lévy distance

$$d_L(\mu, \nu) = \inf\{s \geq 0 : F_\nu(x-s) - s \leq F_\mu(x) \leq F_\nu(x+s) + s \text{ for all } x \in \mathbb{R}\}$$

to measure the distance between $\mu$ and $\nu$, where $F_\mu$ and $F_\nu$ are the cumulative distribution functions of $\mu$ and $\nu$ respectively. A sequence $\mu_n$ of probability measures converges weakly to $\mu$, if and only if $d_L(\mu_n, \mu) \to 0$ (See Theorem III.1.2 on page 314 and Exercise III.1.4 on page 316 in [26]).

We set $LD(c_N, d_N) = \max\{d_L(\mu_{c,N}, \mu_c), d_L(\mu_{d,N}, \mu_d)\}$. Note that even though $d_N$ is a random matrix, the law $\mu_{d,N}$ is deterministic, since $d_N$ is a unitary conjugate of the deterministic self-adjoint diagonal matrix $D_N$. In our random matrix model, $(c_N, d_N) \to (c, d)$ strongly in distribution, so, in particular, $LD(c_N, d_N) \to 0$. Denote by $\varepsilon^{(N)}_1(\beta) = (\omega_1(\beta) - \gamma_1 \otimes c)^{-1}\delta_1(\beta)$, $\varepsilon^{(N)}_2(\beta) = (\omega_2(\beta) - \gamma_1 \otimes d)^{-1}\delta_2(\beta)$ where $\delta_1$ is as in (3.3) and $\delta_2$ is defined analogously, and

$$\varepsilon^{(N)}(\beta) = \|\varepsilon^{(N)}_1(\beta)\| + \|\varepsilon^{(N)}_2(\beta)\|.$$ 

For convenience, we introduce the notation $\mathcal{R}_\eta := \{x + iy : x \in S, y \in V, 0 < y \leq \eta\}$ where $\eta > 0$. Although $\mathcal{R}_\eta$ depends on $V$ too, there is no ambiguity because $V$ does not change throughout the proof, or in the application to linearization.

When $X$ and $Y$ are Banach spaces, we denote by $B(X, Y)$ the space of all bounded linear operators from $X$ to $Y$.

To prove the next proposition, we need the following version of [21] Lemma 2.1 to use Lévy distance to approximate integrals. This version of the lemma only considers integration on a compact interval, instead of on the real line. This small change is due to the fact that we will integrate a continuous but possibly unbounded function on $\mathbb{R}$ against a compactly supported measure; integrating over a compact interval poses no question of integrability.

Since $(c_N, d_N) \to (c, d)$ strongly in distribution, supp $\mu_{c,N} \to$ supp $\mu_c$ and supp $\mu_{d,N} \to$ supp $\mu_d$ in the Hausdorff metric. Consequently, there is an $M > 0$ such that for large enough $N$, $\mu_{c,N}, \mu_{c,N}$ and $\mu_{d,N}$ are all supported in $[-M, M]$.

**Lemma 3.10.** Suppose that $\mu$ and $\nu$ are probability measures supported in $[-M, M]$. Suppose also that $d_L(\mu, \nu) = l$. Assume that $h$ is a $C^1$ real-valued function such that $\int_{-M}^M |h(u)| du < \infty$ and $\int_{-M}^M |h'(u)| du < \infty$. Then

$$\int_{-M}^M |h(u)[F_\nu(\eta u) - F_\mu(\eta u)]| du \leq cl \max\{1, \eta^{-1}\}$$

where $c > 0$ can be chosen to be

$$c = \max\left\{ \sup_{\xi \in [-M, M]} 2|h(\xi)|, \int_{-M}^M |h'(\xi)| d\xi \right\}.$$
We use Newton’s method to estimate how far $\omega_1^{(N)}$ and $\omega_2^{(N)}$ are from the subordination functions $\omega_1$ and $\omega_2$ in the limit.

**Proposition 3.11.** Assume that $(\mu_c, \mu_d, H)$ is regular on $(S, V)$, and that $\bar{S}$ is compact. Then for some positive $\varepsilon, \bar{\eta}, \bar{\varepsilon}$, if $\varepsilon := \varepsilon^{(N)}(\beta) \leq \varepsilon, \bar{\varepsilon} := L\varepsilon, d\varepsilon \leq \bar{\varepsilon}$, and $\beta \in \bar{R}_\eta$, we have

$$\max\{\|\omega_1^{(N)}(\beta) - \omega_1(\beta)\|, \|\omega_2^{(N)}(\beta) - \omega_2(\beta)\|\} = O(\varepsilon + \bar{\varepsilon})$$

where the constant in the $O$ depends only on $c, d, \gamma_1, \gamma_2$.

**Proof.** Consider the function $F : M_n(C) \times M_n(C) \to M_n(C) \times M_n(C)$ defined by

$$F(w_1, w_2) = \left( (w_1 + w_2 - \beta)^{-1} - G_{c_N}(w_1) - \varepsilon_1^{(N)}(\beta), (w_1 + w_2 - \beta)^{-1} - G_{d_N}(w_2) - \varepsilon_2^{(N)}(\beta) \right).$$

The function $F$ depends on $\beta$ but we will find bounds that are independent of $\beta \in \bar{R}_\eta$ for some $\bar{\eta}$. We make the initial guess $w_0(\beta) = (\omega_1(\beta), \omega_2(\beta))$ in Newton’s method. To simply our notations, we also use the notation $\omega_1, \omega_2$ for the continuous extensions of these functions to $\{x + iy \in \mathbb{H}^+(M_n(C)) : x \in S, y \in V\}$.

Since $(\mu_c, \mu_d, H)$ is regular on $(S, V)$ in the sense of Definition 3.8, $\omega_1$ and $\omega_2$ are continuous on $\bar{R}_1$, and $\omega_1(\beta) - t\gamma_1$ is invertible for all $\beta \in \bar{R}_1$ and $t \in \text{supp} \mu_c$. Because $(c_N, d_N) \to (c, d)$ strongly in distribution, $\text{supp} \mu_c \to \text{supp} \mu_c$ in the Hausdorff metric. Since invertible matrices form an open set, $\omega_1(\beta) - t\gamma_1$ is invertible for all $\beta \in S$ and $t \in \text{supp} \mu_{c_N}$ for all large enough $N$. Similarly, $\omega_2(\beta) - t\gamma_2$ is invertible for all $\beta \in S$ and $t \in \text{supp} \mu_{d_N}$ for all large enough $N$. Because $M_n(C)$ is finite dimensional, $\bar{R}_1$ is compact.

Since $\text{supp} \mu_c$ and $\bar{R}_1$ are compact, we have $\|\omega_1(\beta) - t\gamma_1\| \leq C$ for some $C$ independent of $\beta \in \bar{R}_1$ and $t \in \text{supp} \mu_c$. Choosing a bigger $C$ if necessary, $\|\omega_1(\beta) - t\gamma_1\| \leq C$ for some $C$ independent of $\beta \in \bar{R}_1$ and $t \in \text{supp} \mu_{c_N}$ for all large enough $N$. Similarly, $\|\omega_2(\beta) - t\gamma_2\| \leq C$ for some (possibly still larger) $C$ independent of $\beta \in \bar{R}_1$, $t \in \text{supp} \mu_d$ and $t \in \text{supp} \mu_{d_N}$ for large enough $N$. In particular, the entries satisfy

$$\|(\omega_1(\beta) - t\gamma_1)^{-1})_{ij}\| \leq C \text{ and } \|(\omega_2(\beta) - t\gamma_2)^{-1})_{ij}\| \leq C. \tag{3.8}$$

The determinant of $\omega_1(\beta) - t\gamma_2$ is a polynomial in $t$ with degree at most $n$, and each entry of the matrix cofactor of $\omega_1(\beta) - t\gamma_2$ is a polynomial in $t$ with degree at most $n - 1$. Each entry of $(\omega_1(\beta) - t\gamma_2)^{-1}$ can then be expressed as a rational function

$$\frac{P_\beta(t)}{Q_\beta(t)}$$

for some polynomials $P_\beta$ and $Q_\beta$ of degrees at most $n$ and $n - 1$ respectively, and whose coefficients are continuous functions of $\beta$. When $N$ is large enough, by (3.8), $Q_\beta(t)$ are bounded away from 0 for all $t \in \text{supp} \mu_{c_N}$ and $\beta \in \bar{S}$ when $j = 1$, and $t \in \text{supp} \mu_{d_N}$ and $\beta \in \bar{S}$ when $j = 2$. Suppose, for large enough $N$, $\mu_c, \mu_d, \mu_{c_N}$ and $\mu_{d_N}$ are all supported in $[-M, M]$. So, by integration by parts and using the fact that $F_{\mu_c}(M) = 1$, each entry of $\int_{-M}^{M} (\omega_1(\beta) - t\gamma_1)^{-1} \mu_c(dt)$ is of the form

$$\frac{P_\beta(M)}{Q_\beta(M)} - \int_{-M}^{M} \frac{Q_\beta(\lambda)P'_\beta(\lambda) - P_\beta(\lambda)Q'_\beta(\lambda)}{Q_\beta(\lambda)^2} F_{\mu_c}(\lambda) \, d\lambda$$

It follows that

$$G_c(\omega_1(\beta)) - G_{c_N}(\omega_1(\beta)) = \int_{-M}^{M} \frac{Q_\beta(\lambda)P'_\beta(\lambda) - P_\beta(\lambda)Q'_\beta(\lambda)}{Q_\beta(\lambda)^2} [F_{\mu_{c_N}}(\lambda) - F_{\mu_c}(\lambda)] \, d\lambda.$$

By Lemma 3.10 and the fact that $Q_\beta(t)$ are bounded away from 0 for all $t \in \text{supp} \mu_{c_N}$, $t \in \text{supp} \mu_c$ and $\beta \in \bar{S}$, there is a constant $c$ independent of $\beta$ such that

$$\|G_c(\omega_1(\beta)) - G_{c_N}(\omega_1(\beta))\| \leq cl.$$
The similar estimate holds for $d_N$ and $d$ by symmetry. So we have
\[
\left\| \begin{bmatrix} G_c(\omega_1(\beta)) - G_{cN}(\omega_1(\beta)) \\ G_d(\omega_2(\beta)) - G_{dN}(\omega_2(\beta)) \end{bmatrix} \right\| \leq M \eta
\]
for some constant $M$.

Now, by definition of $\epsilon^{(N)}$, $\|\epsilon_1^{(N)}(\beta), \epsilon_2^{(N)}(\beta)\| \leq \epsilon^{(N)}(\beta)$ and we have
\[
\| F(w_0) \| = \left\| \begin{bmatrix} (G_c(\omega_1(\beta)) - G_{cN}(\omega_1(\beta))) - \epsilon_1^{(N)}(\beta) \\ (G_d(\omega_2(\beta)) - G_{dN}(\omega_2(\beta))) - \epsilon_2^{(N)}(\beta) \end{bmatrix} \right\| = O(1 + \epsilon)
\]
where the constant term in the $O$-term is uniform for all $\beta \in \overline{R}_1$.

In order to apply Newton’s method, we need to estimate the norm of the inverse of the derivative. It is easy to compute that the derivative $F'_{w_1, w_2} \in B(M_n(\mathbb{C})^2, M_n(\mathbb{C})^2)$ is given by
\[
F'_{w_1, w_2}(b_1, b_2) = \left( \begin{array}{c}
-(w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1} + \int_{\mathbb{R}}(w_1 - t\gamma_1)^{-1}b_1(w_1 - t\gamma_1)^{-1} \mu_c(dt) \\
-(w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1} + \int_{\mathbb{R}}(w_2 - t\gamma_2)^{-1}b_2(w_2 - t\gamma_2)^{-1} \mu_d(dt)
\end{array} \right)
\]
which is invertible when $(w_1, w_2) = w_0(\beta)$, for $\beta \in \bar{S}$ if $\bar{l}$ is chosen to be small enough so that $d_L(\mu_{cN}, \mu_c)$ and $d_L(\mu_d, \mu_d)$ small, by the assumption (B) that $(\mu_c, \mu_d, H)$ is regular, and that invertible operators in the space $B(M_n(\mathbb{C})^2, M_n(\mathbb{C})^2)$ form an open set. It is evident that $F'$ is continuous, as a function of $(w_1, w_2)$, on $M_n(\mathbb{C}) \times M_n(\mathbb{C})$. Since invertible operators on $M_n(\mathbb{C})^2 \to M_n(\mathbb{C})^2$ form an open set and $\omega_1, \omega_2$ are continuous, there is an $\bar{n} > 0$ such that $F'_{\omega_1(\beta), \omega_2(\beta)}$ is invertible for $\beta \in \overline{R}_{\bar{n}}$. Since $M_n(\mathbb{C})$ and $V$ are finite dimensional, $\overline{R}_{\bar{n}}$ is compact. It follows that $\|F'_{\omega_1(\beta), \omega_2(\beta)}\|^{-1}$ is bounded uniformly on $\overline{R}_{\bar{n}}$.

We now estimate the norm of the second derivative $F'' : M_n(\mathbb{C})^2 \to B(M_n(\mathbb{C})^2, B(M_n(\mathbb{C})^2, M_n(\mathbb{C})^2))$. The first entry of $F''_{w_1, w_2}(h_1, h_2)$ is given by
\[
(F''_{w_1, w_2}(h_1, h_2))_1(b_1, b_2) = (w_1 + w_2 - \beta)^{-1}(h_1 + h_2)(w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1}
\]
\[
+ (w_1 + w_2 - \beta)^{-1}(b_1 + b_2)(w_1 + w_2 - \beta)^{-1}(h_1 + h_2)(w_1 + w_2 - \beta)^{-1}
\]
\[
- \int_{\mathbb{R}}[(w_1 - t\gamma_1)^{-1}h_1(w_1 - t\gamma_1)^{-1}b_1(w_1 - t\gamma_1)^{-1}
\]
\[
+ (w_1 - t\gamma_1)^{-1}b_1(w_1 - t\gamma_1)^{-1}h_1(w_1 - t\gamma_1)^{-1}] \mu_c(dt).
\]
Then the norm of $\|F''_{\omega_1(\beta), \omega_2(\beta)}(h_1, h_2)\|_1$ is bounded by
\[
\|F''_{\omega_1(\beta), \omega_2(\beta)}(h_1, h_2)\|_1 \leq 2\|\omega_1(\beta) + \omega_2(\beta) - \beta\|_1^3 + \sup_{\beta, t} \|\omega_1(\beta) - t\gamma_1\|_1^3 \|\omega_1(\beta)\| \|h_1, h_2\|
\]
which shows the norm in the space $L(M_n(\mathbb{C})^2, L(M_n(\mathbb{C})^2, M_n(\mathbb{C})^2))$ of
\[
\|F''_{\omega_1(\beta), \omega_2(\beta)}\|_1 \leq 2 \left(\|\omega_1(\beta) + \omega_2(\beta) - \beta\|_1^3 + \sup_{\beta, t} \|\omega_1(\beta) - t\gamma_1\|_1^3\right) \leq 4C^3
\]
by the facts that $\sup_{\beta, t} \|\omega_1(\beta) - t\gamma_1\|_1 \leq C$ and $\|\omega_1(\beta) + \omega_2(\beta) - \beta\|_1 \leq \|G_c(\omega_1(\beta))\| \leq C$. A similar computation shows $\|F''_{w_1, w_2}\|_2 \leq 4C^3$. Since $\omega_1$ and $\omega_2$ are continuous on $\overline{R}_1$, $\|F''_{\omega_1(\beta), \omega_2(\beta)}\|$ is bounded uniformly, say by $A$, in $\overline{R}_{\bar{n}}$. 

17
We take \( C_0 = \sup_{\beta \in \overline{\mathcal{R}_\eta}} \| F'(w_0(\beta)) \|. \) Since \( \| [F' (w_0(\beta))^{-1}] F( w_0(\beta)) \| \leq C_0 \| F( w_0(\beta)) \|, \) we take \( \delta_0 = C_0 \sup_{\beta} \| F( w_0(\beta)) \| \) which is \( O(\varepsilon + l) \), with the constant in the \( O \)-term uniform in \( \beta \in \overline{\mathcal{R}_\eta} \). We can take small enough \( l \) and \( \varepsilon \) such that if \( l \leq \bar{l} \) and \( \varepsilon(\beta) \leq \bar{\epsilon} \), then \( h_0 = C_0 \delta_0 A < \frac{1}{2} \) so that Newton’s method guarantees the unique solution is in a neighborhood of \( w_0 \). Newton’s method indeed guarantees the solution \((w_1, w_2)\) satisfying

\[
\| w_1 - \omega_1(\beta) \| = O(\varepsilon + l) \quad \text{and} \quad \| w_2 - \omega_2(\beta) \| = O(\varepsilon + l).
\]

The functions \( \omega_1^{(N)}(\beta) \) and \( \omega_2^{(N)}(\beta) \) is the zero of \( F, F(w_1, w_2) = 0 \). Also, (3.5) shows that \( \omega_1^{(N)}(\beta) \to \omega_1(\beta) \) and \( \omega_2^{(N)}(\beta) \to \omega_2(\beta) \) as \( N \to \infty \). If follows that for small \( \bar{l} \) and \( \bar{\varepsilon} \), the solution of \( F(w_1, w_2) = 0 \) found by Newton’s method is the pair \( (\omega_1^{(N)}(\beta), \omega_2^{(N)}(\beta)) \) which proves for all \( \beta \in \overline{\mathcal{R}_\eta} \) (but not \( \mathcal{R}_\eta \)).

\[
\| \omega_1^{(N)}(\beta) - \omega_1(\beta) \| = O(\varepsilon + l) \quad \text{and} \quad \| \omega_2^{(N)}(\beta) - \omega_2(\beta) \| = O(\varepsilon + l).
\]

\( \square \)

### 3.2 Main Results

#### 3.2.1 A Local Limit Theorem for Polynomials

Now we move on to estimating the (scalar) Cauchy transform of a polynomial \( P(c_N, d_N) \) in \( c_N \) and \( d_N \). We recall that \( (c_N, d_N) \to (c, d) \) strongly in distribution. The polynomial \( P \) possesses a linearization as in Section 2.3 which gives a linear polynomial with coefficients in \( M_n(\mathbb{C}) \)

\[
L_N = \gamma_0 + \gamma_1 \otimes c_N + \gamma_2 \otimes d_N.
\]

We apply the results from the preceding section to \( H_N = \gamma_1 \otimes c_N + \gamma_2 \otimes d_N \) and its large-\( N \) limit \( H = \gamma_1 \otimes c + \gamma_2 \otimes d \).

**Lemma 3.12.** Assume that \((\mu_c, \mu_d, H)\) is regular on \( (S, V) \), and that \( \bar{S} \) is compact. Then for some sufficiently small \( \varepsilon = \varepsilon(\beta), l = LD(c_N, d_N) \), and \( \eta > 0 \), we have

\[
\| \mathbb{E} G_{H_N}(\beta) - G_H(\beta) \| = O(\varepsilon + l)
\]

where \( \Re \beta \in S, \Im \beta \in V, \Im \beta > 0, \| (\Im \beta)^{-1} \| = \frac{1}{n} \).

**Proof.** Because \( \mathbb{E} G_{H_N}(\beta) = (\omega_1^{(N)}(\beta) + \omega_2^{(N)}(\beta) - \beta)^{-1} \) and \( G_H(\beta) = (\omega_1(\beta) + \omega_2(\beta) - \beta)^{-1} \), we see that

\[
\mathbb{E} [ M_n( \text{tr} (R_N(\beta)) ) - M_n( \text{tr} (R(\beta)) ) = (\omega_1^{(N)}(\beta) + \omega_2^{(N)}(\beta) - \beta)^{-1} \left( (\omega_1^{(N)}(\beta) + (\omega_2 - \omega_2^{(N)}(\beta)))(\omega_1(\beta) + \omega_2(\beta) - \beta)^{-1} \right).
\]

By the subordination relation

\[
(\omega_1(\beta) + \omega_2(\beta) - \beta)^{-1} = \int_{\mathbb{R}} (\omega_2(\beta) - t\gamma_2)^{-1} \mu_d(dt),
\]

\((\mu_c, \mu_d, H)\) being regular implies that \( (\omega_1(\beta) + \omega_2(\beta) - \beta)^{-1} \) is uniformly bounded for small enough \( \eta > 0 \), \( \Re \beta \in S, \Im \beta \in V, \Im \beta > 0, \| (\Im \beta)^{-1} \| = \frac{1}{n} \). Proposition 3.11 shows that \( \omega_1^{(N)} \) and \( \omega_2^{(N)} \) are close to \( \omega_1 \) and \( \omega_2 \) respectively, so when \( N \) large, \( (\omega_1^{(N)}(\beta) + \omega_2^{(N)}(\beta) - \beta)^{-1} \) is also bounded. The norm of the middle term \( \| (\omega_1 - \omega_1^{(N)})(\beta) + (\omega_d - \omega_d^{(N)}(\beta)) \| = O(\varepsilon + l) \) by the conclusion of Proposition 3.11 provided that \( n, \varepsilon, l \) are small enough. \( \square \)
Lemma 3.13. For some positive constants $c_1$ and $c''$ which may depend on $|I|$ and $\kappa$ and for all $\delta > 0$,

$$
\mathbb{P} \left( \sup_{I,\kappa, \cdot e_{1,1} - \gamma_0 + i\eta} \|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| > \delta \right) \leq \exp \left( \frac{-c''N\eta^4\delta^2}{\eta_1\delta^2} \right)
$$

where the supremum is taken over $\beta \in \Pi_{I,\kappa, \cdot e_{1,1} - \gamma_0 + i\eta}$, provided that $N \geq \frac{-c_1\log(\eta^4\delta^2)}{\eta^4\delta^2}$ for some $c_1$ large enough.

Proof. It is easy to compute $\|G'_{H_N}(\beta)\| \leq \|(3\beta)^{-1}\|^2$ and $\|\mathbb{E}G'_{H_N}(\beta)\| \leq \|(3\beta)^{-1}\|^2$.

We create an $\frac{\eta^4}{\delta^2}$-net on $\Pi_{I,\kappa}$, which requires us $O(|I| \cdot \kappa)$ number of points. If $\|G_{H_N}(\beta_0) - \mathbb{E}G_{H_N}(\beta_0)\| \leq \delta/2$ for every point $\beta_0$ on the net, then given any $\beta = z e_{1,1} - \gamma_0 + i\eta$, choose $\beta_0$ such that $\|\beta - \beta_0\| \leq \frac{1}{\sqrt{2}} \frac{\eta^4}{\delta^2} < \frac{\eta^4}{\delta^2}$. Then

$$
\|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| \leq \|G_{H_N}(\beta) - G_{H_N}(\beta_0)\| + \|G_{H_N}(\beta_0) - \mathbb{E}G_{H_N}(\beta_0)\| + \|\mathbb{E}G_{H_N}(\beta_0) - \mathbb{E}G_{H_N}(\beta)\|
$$

$$
\leq \|(3\beta_0)^{-1}\|^2\|\beta - \beta_0\| + \frac{\delta}{2} + \|(3\beta_0)^{-1}\|^2\|\beta - \beta_0\| \leq \delta.
$$

Recall, by the concentration inequality, for all $\alpha \in (0, \frac{1}{2})$, for every $\beta \in \mathbb{H}^+(\mathbb{M}_n(\mathbb{C}))$, $\|\beta^{-1}\| = \frac{1}{\eta}$, we have

$$
\mathbb{P} \left( \|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| > \delta \right) \leq 2 \exp \left( \frac{-cN\eta^4\delta^2}{\eta_2\delta^2\|d_N\|^2} \right).
$$

Since there are $O(|I| \cdot \kappa)$ number of points on the net,

$$
\mathbb{P} \left( \sup_{I,\kappa, \cdot e_{1,1} - \gamma_0 + i\eta} \|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| > \delta \right) \leq \frac{C|I| \cdot \kappa}{\eta^4\delta^2} \exp \left( \frac{-cN\eta^4\delta^2}{\eta_2\delta^2\|d_N\|^2} \right)
$$

$$
\leq \exp \left( \frac{-cN\eta^4\delta^2}{\eta_2\delta^2\|d_N\|^2} + \log \left( \frac{C|I| \cdot \kappa}{\eta^4\delta^2} \right) \right)
$$

$$
\leq \exp \left( \frac{-c''N\eta^4\delta^2}{\eta_2\delta^2\|d_N\|^2} \right)
$$

if $N \geq \frac{-c_1\log(\eta^4\delta^2)}{\eta_2\delta^2}$ for some $c_1$ large enough.

Let $\delta^2 = \frac{\log N}{N\eta^2}$. Then $N \geq \frac{-c_1\log(\eta^4\delta^2)}{\eta^4\delta^2} = \frac{-c_1\log((\log N)/N)}{(\log N)/N}$ when $N$ large. If $\eta = N^{-1/12}\sqrt{\log N}$, then Lemma 3.13 shows that, as long as $N$ is large enough,

$$
\mathbb{P} \left( \sup_{I,\kappa, \cdot e_{1,1} - \gamma_0 + i\eta} \|G_{H_N}(\beta) - \mathbb{E}G_{H_N}(\beta)\| > \frac{1}{N^{1/2}\sqrt{\log N}} \right) \leq \exp (-c \log N)
$$

for some constant $c > 0$.

Given any $\varepsilon > 0$, by (3.6) and Proposition 3.6 $\varepsilon(N)(\beta) < \varepsilon$ when $N$ large. Lemma 3.12 shows that

$$
\|\mathbb{E}G_{H_N}(\beta) - G_H(\beta)\| \leq \varepsilon + Cl.
$$

We have proved the following.
Proposition 3.14. Assume that the triple \((\mu_c, \mu_d, H)\) is regular in \((\Pi_{I,K} \cdot e_{1,1} - \gamma_0, \mathbb{R} \cdot I_n)\). Assume also that 
\[ l := LD(c_N, d_N) \leq \tilde{l} \text{ where } \tilde{l} \text{ is chosen as in Proposition 3.17} \]
Let \( \eta = N^{-1/12} \sqrt{\log N} \). Then for some positive constants \( C \) and \( c, \) and for every \( \varepsilon > 0, \)
\[
\mathbb{P} \left( \sup_{\Pi_{I,K} \cdot e_{1,1} - \gamma_0 + i\eta} \| G_{H_N}(\beta) - G_H(\beta) \| > \varepsilon + Cl \right) \leq \exp \left( -c \log N \right)
\]
for all sufficiently large \( N. \)

We denote the Hilbert-Schmidt norm by \( \| \cdot \|_{HS} \). The function \( X \mapsto \| X \|_{HS} \) is clearly Lipschitz, with Lipschitz constant 1, in the Hilbert-Schmidt norm. By [3 Corollary 4.4.28] and [9 Lemma 8.3], for each \( \beta \) with \( \| \Im \beta \| = \eta, \)
\[
\mathbb{E} \left( \| R_N(\beta) \|_{HS} - \mathbb{E}[\| R_N(\beta) \|_{HS}] \geq \varepsilon \right) \leq 2 \exp \left( \frac{-\varepsilon^2 N \eta^4}{16 \| \gamma_2 \|^2 \| D_N \|^2} \right), \tag{3.9}
\]
Lemma 3.15. Let \( \eta = N^{-1/16} (\log N)^{\delta}, \delta > 0 \) and \( \beta \) is of the form \( z e_{1,1} - \gamma_0 + i\eta, \) where \( z \in \Pi_{I,K}. \) If \((\mu_c, \mu_d, H)\) is regular in \((\Pi_{I,K} \cdot e_{1,1} - \gamma_0, \mathbb{R} \cdot I_n)\), then the (random) matrix norm \( \| R_N(\beta) \| \) of \( R_N(\beta) \) is bounded uniformly with large probability. That is, there exists a \( C > 0 \) such that
\[
\mathbb{P} \left( \sup_{\Pi_{I,K} \cdot e_{1,1} - \gamma_0 + i\eta} \| R_N(\beta) \|_{HS} > C \right) \leq \exp \left( \frac{-c'' N \eta^4 \varepsilon^2}{\| \gamma_2 \|^2 \| d_N \|^2} \right)
\]
for all large enough \( N. \)

Proof. Recall that, by Proposition 3.11 and Theorem 3.7 (2) that \( \varepsilon^{(N)}(\beta) = O \left( \frac{1}{N^2 \eta^5} \right), \)
\[
\| \omega_d(\beta) - \omega_{d_N}(\beta) \| = O \left( \frac{1}{N^{2/5} \eta^5} + LD(c_N, d_N) \right) \tag{3.10}
\]
when \( N \) is large enough, so that \( LD(c_N, d_N) \) is small enough. We also recall that
\[
LD(c_N, d_N) = \max \{ d_L(\mu_c, \mu_d), d_L(\mu_{d_N}, \mu_d) \}.
\]
If \( N^{2/5} \eta^4 \) is large enough, by Theorem 3.7 \( \| \mathbb{E} R_N(\beta) - (\omega_1^{(N)}(\beta) - \gamma_1 \circ c_N)^{-1} \| = O \left( \frac{1}{N^{2/5} \eta^5} \right). \) We notice that
\[
(\omega_1^{(N)}(\beta) - \gamma_1 \circ c_N)^{-1} = \int_{\mathbb{R}} (\omega_1^{(N)}(\beta) - t\gamma_1)^{-1} E_N(dt)
\]
where \( E_N \) is the (discrete) spectral measure for \( c_N. \) The support of \( E_N \) converges to \( \text{supp } \mu_c \) in Hausdorff distance by strong convergence in distribution. Using equation (3.10) and Proposition 3.11 when \( N \) is large enough, regularity on \((\Pi_{I,K} \cdot e_{1,1} - \gamma_0, \mathbb{R} \cdot I_n)\) implies that \((\omega_1^{(N)}(\beta) - \gamma_1 \circ c_N)^{-1} \) is bounded uniformly. We conclude that \( \| \mathbb{E} R_N(\beta) \| < C \) for some \( C > 0, \) for all \( \beta = ze_{1,1} - \gamma_0 + i\eta \) where \( z \in \Pi_{I,K}. \)

By (3.9), for a particular \( \beta_0, \) with probability at least \( 1 - 2 \exp \left( \frac{-N^{2/5} \eta^4}{16 \| \gamma_2 \|^2 \| d_N \|^2} \right), \) we have
\[
\| R_N(\beta_0) \|_{HS} \leq \varepsilon + \mathbb{E}[\| R_N(\beta_0) \|_{HS}].
\]
We create an \( \frac{N^{2/5} \eta^4}{4} \)-net on \( \Pi_{I,K}, \) which requires \( O(|I| \cdot \kappa) \) number of points. Given any \( \beta, \) there exists \( \beta_0 \) such that \( \| \beta - \beta_0 \| \leq \frac{N^{2/5} \eta^4}{4} \) and \( \| R_N(\beta_0) \| \leq \frac{\varepsilon}{2} + \mathbb{E}[\| R_N(\beta_0) \|]. \) Then using the inequality \( \| AB \|_{HS} \leq \| A \| \| B \|_{HS} \) we have the estimates
\[
\| R_N(\beta) \|_{HS} - \| R_N(\beta_0) \|_{HS} \leq \| R_N(\beta) - R_N(\beta_0) \|_{HS} \leq \| \beta - \beta_0 \|_{HS} \| (\Im \beta)^{-1} \| \| (\Im \beta_0)^{-1} \| \leq \frac{\varepsilon}{4}.
\]
Similarly,
\[
\|\mathbb{E}R_N(\beta)\|_{HS} - \|\mathbb{E}R_N(\beta_0)\|_{HS} \leq \|\mathbb{E}R_N(\beta) - \mathbb{E}R_N(\beta_0)\|_{HS} \leq \|\beta - \beta_0\|_{HS}(\|\beta_3\| - \|\beta_0\|) \leq \frac{\varepsilon}{4}.
\]
It follows that, when \( N \) is sufficiently large,
\[
P \left( \sup_{\Pi_{I,\varepsilon} \in J_{1, \gamma_0}} \| R_N(\beta) \|_{HS} > \varepsilon + \varepsilon \| R_N(\beta) \| \right) \leq C' \| I \cdot \kappa \| \exp \left( \frac{-c \eta \beta^2}{\| \gamma_2 \|^2 \| d_N \|^2} \right) \leq \exp \left( \frac{-c'' \eta \beta^2}{\| \gamma_2 \|^2 \| d_N \|^2} \right)
\]
for some constants \( C', c, c'' > 0 \). Since \( \| \mathbb{E}R_N(\beta) \| < C \), we conclude the proof. \( \square \)

We use the regularity at the boundary of \( M_n(tr)[(\varepsilon e_{1,1} - L_N + i\eta)^{-1}] \) to get estimates of \( M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] \); the latter is what we actually want to estimate because the \( (1,1) \)-entry of \( M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] \) will give us the Cauchy transform of \( P(c_N, d_N) \).

We have
\[
\| M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] - M_n(\tau)[(\varepsilon e_{1,1} - L)^{-1}] \|_{HS} \\
\leq \| M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] - M_n(tr)[(\varepsilon e_{1,1} - L_N + i\eta)^{-1}] \|_{HS} + \| M_n(\tau)[(\varepsilon e_{1,1} - L + i\eta)^{-1}] \|_{HS} \\
\leq \eta \| (\varepsilon e_{1,1} - L_N)^{-1} \| \| (\varepsilon e_{1,1} - L_N + i\eta)^{-1} \|_{HS} + \| M_n(\tau)[(\varepsilon e_{1,1} - L + i\eta)^{-1}] \|_{HS} \\
\leq M \eta \| (\varepsilon e_{1,1} - L_N)^{-1} \| \| (\varepsilon e_{1,1} - L + i\eta)^{-1} \|_{HS} + \| M_n(\tau)[(\varepsilon e_{1,1} - L + i\eta)^{-1}] \|_{HS}.
\]
We take \( z = x + i\eta^\alpha \), where \( x \in I \) and \( 0 < \alpha < 1 \). The quantities \( \| (\varepsilon e_{1,1} - L_N + i\eta)^{-1} \| \), \( \| (\varepsilon e_{1,1} - L + i\eta)^{-1} \| \) are bounded with large probability, by Lemma 3.15. Since \( (\mu_c, \mu_d, H) \) is regular on \( \Pi_{I,\varepsilon} \), and by Lemma 2.7,
\[
\| M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] - M_n(\tau)[(\varepsilon e_{1,1} - L)^{-1}] \|_{HS} \\
\leq M \eta^{1-\alpha} + \| M_n(tr)[(\varepsilon e_{1,1} - L_N + i\eta)^{-1}] - M_n(\tau)[(\varepsilon e_{1,1} - L + i\eta)^{-1}] \|_{HS} \leq M \eta^{1-\alpha} + \sqrt{\eta} \| M_n(tr)[(\varepsilon e_{1,1} - L_N + i\eta)^{-1}] - M_n(\tau)[(\varepsilon e_{1,1} - L + i\eta)^{-1}] \|_{HS}
\]
for some constant \( M > 0 \). We note that we used the fact that \( \| A \|_{HS}^2 = n \cdot \text{tr}(A^*A) \leq n \| A^*A \| \leq n \| A \|^2 \) for any matrix \( A \). The preceding observation leads to the following result, by Proposition 3.14.

**Proposition 3.16.** Assume that the triple \( (\mu_c, \mu_d, H) \) is regular on \( (\Pi_{I,\kappa} \cdot \varepsilon_{1,1} - \gamma_0, \mathbb{R} \cdot I_n) \) for some \( \kappa > 0 \). Assume also that \( \lambda := LD(c_N, d_N) \leq \tilde{l} \) where \( \tilde{l} \) is chosen such that the conclusion of Proposition 3.11 holds. Let \( \eta = N^{-1/12} \sqrt{\log N} \). Then for some positive constants \( M, C, c, \) and for every \( 0 < \alpha < 1, \varepsilon > 0, \)
\[
P \left( \sup_{\Pi_{I,\varepsilon} \in J_{1, \gamma_0}} \| M_n(tr)[(\varepsilon e_{1,1} - L_N)^{-1}] - M_n(\tau)[(\varepsilon e_{1,1} - L)^{-1}] \|_{HS} > M \eta^{1-\alpha} + \varepsilon + C\tilde{l} \right) \leq \exp(-c \log N)
\]
for all sufficiently large \( N \).

Being regular on \( (\Pi_{I,\kappa} \cdot \varepsilon_{1,1} - \gamma_0, \mathbb{R} \cdot I_n) \) ensures that \( \mu_{P(c,d)} \) is absolutely continuous on the interval \( I \); this follows from two results. In [8] Lemma 1.10 Belinschi showed that if we denote \( \mu^{sc} \) as the singular continuous part of \( \mu \), then for \( \mu^{sc} \)-almost all \( x \in \mathbb{R} \), the nontangential limit of the imaginary part of the Cauchy transform is infinite. Bercovici and Voiculescu [12] proved that \( (z - x)G_\mu(x) \to \mu(\{x\}) \) as \( z \to x \) nontangentially.

The main theorem of this section is a local limit theorem. The proof is almost identical to the ones given by [18] Corollary 4.2 and [21] Theorem 1.4.
\textbf{Theorem 3.17.} Suppose that \( \mu_{cN} \to \mu_c \) and \( \mu_{dN} \to \mu_d \) weakly. Using the same notation as above: \( H \) is the nonconstant part of the linearization \( L \) of \( P \). Suppose \( (\mu_c, \mu_d, H) \) is regular on \((\Pi_{I, \infty} \cdot e_{1,1} - \gamma_0, \mathbb{R} \cdot I_n)\). Let \( \rho \) be the density of \( P(c, d) \), that is the absolutely continuous part of \( \mu_{P(c,d)} \). Assume that \( \eta_N^* = N^{-\frac{1}{2\alpha}} \log N \) and fix any \( \alpha \in (0, 1) \). Then for every \( x \in I \),

\[
\frac{M(\eta_N^*)^\alpha(x)}{2(\eta_N^*)^\alpha N} \to \rho(x)
\]

in probability, where \( M(\eta_N^*)^\alpha(x) \) is the number of eigenvalues in the interval centered at \( x \) of length \( 2(\eta_N^*)^\alpha \).

\textbf{Proof.} In this proof, we will drop the subscript of \( \eta_N^* \), and write \( \eta^* \) instead. Fix any \( \alpha \in (0, 1) \). Let \( \eta = cN^{-1/12} \) and \( c \) is sufficiently large, and let \( \eta^* = B^{\frac{1}{2\alpha}} \eta \). We write \( I^* = [x - (\eta^*)^\alpha, x + (\eta^*)^\alpha] \). Let

\[
R(\lambda) := \frac{1}{\pi} \int_{x-(\eta^*)^\alpha}^{x+(\eta^*)^\alpha} \frac{\eta^\alpha}{(t-\lambda)^2 + (\eta^\alpha)^2} \, dt
= \frac{1}{\pi} \left( \arctan \left( \frac{x-\lambda}{\eta^\alpha} + B \right) - \arctan \left( \frac{x-\lambda}{\eta^\alpha} - B \right) \right).
\]

Then \( R = 1_{I^*} + T_1 + T_2 + T_3 \) where \( 1_{I^*} \) is the indicator function of \( I^* \) and functions \( T_1, T_2 \) and \( T_3 \) satisfy the following properties:

\[
|T_1| \leq \frac{c}{\sqrt{B}} \quad \text{supp}(T_1) \subseteq I_1 = [x - 2(\eta^*)^\alpha, x + 2(\eta^*)^\alpha];
|T_2| \leq \frac{c}{\sqrt{B}} \quad \text{supp}(T_2) \subseteq J_1 \cup J_2;
|T_3| \leq \frac{C_{\eta^\alpha}(\eta^*)^\alpha}{(x-x^2+(\eta^*)^\alpha)} \quad \text{supp}(T_3) \subseteq I_1.
\]

The \( J_1 \) and \( J_2 \) above are the intervals of length \( \sqrt{B} \eta^\alpha \) with midpoints \( x - (\eta^*)^\alpha \) and \( x + (\eta^*)^\alpha \) respectively. Note that

\[
\frac{M(\eta^*)^\alpha(x)}{2(\eta^*)^\alpha N} = \frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} 1_{I^*}(\lambda) \mu_{PN}(d\lambda)
= \frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} R(\lambda) \mu_{PN}(d\lambda) - \frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} (T_1 + T_2 + T_3) \mu_{PN}(d\lambda).
\]

The second integral can be estimated as

\[
\frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} (T_1 + T_2 + T_3) \mu_{PN}(d\lambda) \leq \frac{c}{\sqrt{B}} \frac{M_{I_1}(x)}{2(\eta^*)^\alpha N} + \frac{M_{J_1}(x) + M_{J_2}(x)}{2(\eta^*)^\alpha N} + \frac{C_{\eta^\alpha}}{(\eta^*)^\alpha} \rho_{(\eta^*)^\alpha}(x)
\]

where \( M_I \) denotes the number of eigenvalues of \( H_N \) in interval \( I \), and

\[
\rho_{(\eta^*)^\alpha}(x) := \frac{1}{\pi} \Im G_{PN}(x + i(\eta^*)^\alpha) = \frac{1}{\pi} \int_{\mathbb{R}} (\eta^*)^\alpha \frac{1}{(x-\lambda)^2 + (\eta^*)^{2\alpha}} \mu_{PN}(d\lambda).
\]

Hence, by using the inequality

\[
\frac{1}{\pi} \int_{\mathbb{R}} \frac{(\eta^*)^\alpha}{(x-\lambda)^2 + (\eta^*)^{2\alpha}} \mu_{PN}(d\lambda) \geq \frac{1}{\pi} \int_{x-(\eta^*)^\alpha}^{x+(\eta^*)^\alpha} \frac{(\eta^*)^\alpha}{(x-\lambda)^2 + (\eta^*)^{2\alpha}} \mu_{PN}(d\lambda) \geq \frac{1}{2\pi} \frac{M(\eta^*)^\alpha(x)}{N},
\]

we have

\[
\frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} |T_1 + T_2 + T_3| \mu_{PN}(d\lambda) \leq \frac{c}{\sqrt{B}} \rho_{2(\eta^*)^\alpha}(x) + \rho_{\sqrt{B}(\eta^*)^\alpha}(x - (\eta^*)^\alpha) + \rho_{\sqrt{B}(\eta^*)^\alpha}(x + (\eta^*)^\alpha) + \rho_{(\eta^*)^\alpha}(x).
\]
By the linearization property, $\mathbb{E} G_{P_N}(x + i(\eta^*)^\alpha) = M_N(\text{Etr})[(x + i(\eta^*)^\alpha)e_{1,1} - L(c_N, d_N))^{-1}]_{1,1}$. So, putting $z = x + i(\eta^*)^\alpha$ we look at

$$\left\{ M_N(\text{Etr})[(ze_{1,1} - L(c_N, d_N))^{-1}] - M_N(\text{Etr})[(ze_{1,1} + i\eta - L(c_N, d_N))^{-1}] \right\}$$

and

$$= M_N(\text{Etr}) \left( i(ze_{1,1} - L(c_N, d_N))^{-1}(ze_{1,1} - L(c_N, d_N) + i\eta)^{-1} \right) + M_N(\text{Etr})[(ze_{1,1} + i\eta - L(c_N, d_N))^{-1}]

= M_N(\text{Etr}) \left( i(ze_{1,1} - L(c_N, d_N))^{-1}(ze_{1,1} - \gamma_0 + i\eta - H_N)^{-1} \right) + G_{H_N}(ze_{1,1} - \gamma_0 + i\eta)

The first term is bounded by $\eta^{-1-\alpha}\|ze_{1,1} - \gamma_0 + i\eta - H_N)^{-1}\|$, which is bounded (and indeed goes to 0 as $\eta \to 0$) by the assumption that $(\mu_c, \mu_d, H)$ is regular and Proposition 3.11. The argument here is similar to the proof Lemma 3.12. The regularity assumption implies that

$$G_{H_N}(ze_{1,1} - \gamma_0 + i\eta)

=[G_{H_N}(ze_{1,1} - \gamma_0 + i\eta) - G_{c_N}(\omega_1^{(N)}(ze_{1,1} - \gamma_0 + i\eta))] + G_{c_N}(\omega_1^{(N)}(ze_{1,1} - \gamma_0 + i\eta))

=O(N^{-\frac{3}{2}}\eta^{-5}) + G_{c_N}(\omega_1^{(N)}(ze_{1,1} - \gamma_0 + i\eta))

=o(1) + G_{c_N}(\omega_1^{(N)}(ze_{1,1} - \gamma_0 + i\eta))$$

is bounded, so is $\mathbb{E} G_{P_N}(x + i(\eta^*)^\alpha)$. Whence, $\frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} R(\lambda)\mu_{P_N}(d\lambda) = O(B^{-\frac{1}{2}})$.

Now the main term $\frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} R(\lambda)\mu_{P_N}(d\lambda) = \frac{1}{\pi} \int_{\mathbb{R}} \Im G_P(t + i(\eta^*)^\alpha) dt$

and $\frac{1}{2(\eta^*)^\alpha} \int_{\mathbb{R}} \frac{1}{\pi} \Im \left[ G_{P_N}(t + i(\eta^*)^\alpha) - G_P(t + i(\eta^*)^\alpha) \right] dt.$

The first part converges to $\rho(x)$ because of the assumption that $(\mu_c, \mu_d, H)$ is regular at $x$. For the second term, it goes to zero in probability by Proposition 3.16 because we assume both $\|\mu_{c_N} - \mu_c\| \to 0$ and $\|\mu_{d_N} - \mu_d\| \to 0$ as $N \to \infty$. \hfill $\square$

### 3.2.2 Delocalization of Eigenvectors

Delocalization of eigenvectors was proved in [18] for the general Wigner random matrices. Kargin [21] considered delocalization of different order.

**Definition 3.18 (21).** We say that the eigenvectors $v_1^{(N)}, \ldots, v_k^{(N)}$ with corresponding eigenvalues in $I$ of a sequence of matrices are delocalized at length $N^{-\theta}$ in the interval $I$ if there exists $\delta > 0$ such that

$$\mathbb{P} \left( \max_{j=1,\ldots,k_N} \max_{i=1,\ldots,N} |v_j^{(N)}(i)| > N^{-\theta} \log N \right) \leq \exp(-N^{-\delta})$$

for all sufficiently large $N$.

Kargin proved the delocalization of eigenvectors for a sum of random matrices of the form $C_N + U_N D_N U_N$, using smoothness assumption. The next theorem shows that if we have the regularity condition for the $M_n(\mathbb{C})$-valued matrices arised from linearization of polynomials, we can prove the delocalization of eigenvectors for polynomials $P(c_N, d_N)$ of two random matrices.
Theorem 3.19. Suppose that \((c_N, d_N) \to (c, d)\) strongly in distribution. Denote \(H\) as the nonconstant part of the linearization \(L\) of \(P\). Suppose also that \((\mu_c, \rho_d, H)\) is regular on \((\Pi_{I, \kappa} \cdot e_{1,1} - \gamma_0, \mathbb{R} \cdot I_n)\), where \(I\) is a compact interval on \(\mathbb{R}\) and \(\kappa > 0\). Then, given any \(\alpha \in (0, 1)\), for all sufficiently large \(N\), the eigenvectors of \(P(c_N, d_N)\) are delocalized at length \(N^{-\frac{\alpha}{24}}\) in \(I\).

Proof. It suffices to show that, given any \(v_m^{(N)}\) of \(P(c_N, d_N)\),
\[
\mathbb{P}(|v_m^{(N)}(i)|^2 > M'N^{-\frac{\alpha}{24}} \log N) \leq 2 \exp(-K'N\varepsilon^2)
\]
for some positive constant \(K'\) because \(k_N \leq N\); we are taking the intersection of at most \(N^2\) events of the form \(|v_m(k)|^2 > M'N^{-\frac{\alpha}{24}} \log N\) in the probability in the statement of delocalization of eigenvectors.

Let \(\eta = N^{-\frac{\alpha}{24}}(\log N)^{1/\alpha}\). We write \(\beta = ze_{1,1} - \gamma_0 + i\eta\) for \(z \in \Pi_{I, \kappa}\). By Proposition 3.6, \(\omega_1^{(N)}(\beta) \in \mathbb{H}^+(M_n(\mathbb{C}))\) when \(N\) is large.

Again, as in [9, Proof of Proposition 8.4], the concentration inequality implies there is a constant \(K > 0\) such that for all \(i = 1, 2, \ldots, n\),
\[
\mathbb{P}(|(R_N(\beta) - \mathbb{E}R_N(\beta))_{i,i}| > \varepsilon) \leq 2 \exp(-KN\varepsilon^2)
\]
for all \(\varepsilon > 0\). The \((1,1)\)-block of \(R_N(ze_{1,1} - \gamma_0)\) is the resolvent \((z - P(c_N, d_N))^{-1}\); the \((i,i)\)-entry of \(R_N(ze_{1,1} - \gamma_0)\) is the \((i,i)\)-entry of the resolvent \((z - P(c_N, d_N))^{-1}\).

Now, let \(z = x + i\eta^\alpha\) where \(x \in I\) and \(\alpha \in (0, 1)\). We have
\[
(z_{e_{1,1}} - L)^{-1} = (ze_{1,1} - L)^{-1} - (ze_{1,1} + i\eta - L)^{-1} + [R(\beta) - \mathbb{E}R(\beta)] + \mathbb{E}[R(\beta)]
\]
\[
= i\eta(z_{e_{1,1}} - \gamma_0 - \gamma_1 \otimes c_N - \gamma_2 \otimes d_N)^{-1} (ze_{1,1} - \gamma_0 + i\eta - \gamma_1 \otimes c_N - \gamma_2 \otimes d_N)^{-1}
\]
\[
+ [R_N(\beta) - \mathbb{E}R_N(\beta)] + \mathbb{E}R_N(\beta).
\]
It follows that, by applying Lemma 2.7 to the first term, and by Lemma 3.13 there is an \(M > 0\) such that, with probability at least \(2 \exp(-K'N\varepsilon^2)\) for some \(K' > 0\),
\[
||((z - P(c_N, d_N))^{-1})_{i,i}| \leq M\eta^{-\alpha} + |[R_N(\beta) - \mathbb{E}R_N(\beta)]_{k,k}| + C \leq M\eta^{-\alpha} + \varepsilon + C. \tag{3.14}
\]

Let \(v_m^{(N)}, m = 1, \ldots, N\), be the eigenvectors of \(P(c_N, d_N)\), with corresponding eigenvalues \(\lambda_m\). Since
\[
(z - P(c_N, d_N))^{-1} = \sum_{m=1}^N \frac{v_m v_m^*}{z - \lambda_m},
\]
and so
\[
-\Im((z - P(c_N, d_N))^{-1})_{i,i} = \sum_{m=1}^N \eta^\alpha |v_m(i)|^2 \frac{\eta^\alpha}{(z - \lambda_m)^2 + \eta^{2\alpha}}
\]
where \(v_m(i)\) means the \(i\)th-component of \(v_m\). Taking \(x = \lambda_m\) for a particular \(\lambda_m \in I\) and using the inequality (3.14) give
\[
|v_m(i)|^2 \leq -\eta^\alpha \Im((z - P(c_N, d_N))^{-1})_{k,k}
\]
\[
\leq M\eta + \eta^\alpha |[R_N(\beta) - \mathbb{E}R_N(\beta)]_{k,k}| + C\eta^\alpha
\]
\[
\leq (M + \varepsilon + C)\eta^\alpha = (M + \varepsilon + C)N^{-\frac{\alpha}{24}} \log N
\]
since \(\eta < 1\) when \(N\) large. It follows that
\[
\mathbb{P}(|v_m(i)|^2 > M'N^{-\frac{\alpha}{24}} \log N) \leq 2 \exp(-K'N\varepsilon^2).
\]
This concludes the proof. \(\square\)
4 Acknowledgement

The author would like to thank Hari Bercovici for useful conversations and discussions; the author would also like to thank his careful reading and valuable advice that improves the manuscript. The author learned the use of linearization from conversations with Tobias Mai and Roland Speicher, when he visited Germany, organized and funded by Todd Kemp and Roland Speicher.

The author would also like to thank the anonymous referee providing useful comments and suggestions to improve the manuscript.

References

[1] Akhiezer, N. I. The classical moment problem and some related questions in analysis. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.

[2] Anderson, G. W. Convergence of the largest singular value of a polynomial in independent Wigner matrices. Ann. Probab. 41, 3B (2013), 2103–2181.

[3] Anderson, G. W., Guionnet, A., and Zeitouni, O. An introduction to random matrices, vol. 118 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010.

[4] Bao, Z., Erdös, L., and Schnelli, K. Convergence rate for spectral distribution of addition of random matrices. Advances in Mathematics 319 (2017), 251 – 291.

[5] Bao, Z., Erdös, L., and Schnelli, K. Local law of addition of random matrices on optimal scale. Communications in Mathematical Physics 349, 3 (Feb 2017), 947–990.

[6] Bao, Z., Erdös, L., and Schnelli, K. Spectral rigidity for addition of random matrices at the regular edge. Arxiv:1708.01597. (2017).

[7] Belinschi, S., Popa, M., and Vinnikov, V. Infinite divisibility and a non-commutative boolean-to-free bercovicipata bijection. Journal of Functional Analysis 262, 1 (2012), 94 – 123.

[8] Belinschi, S. T. A note on regularity for free convolutions. Annales de l’Institut Henri Poincare (B) Probability and Statistics 42, 5 (2006), 635 – 648.

[9] Belinschi, S. T., Bercovici, H., and Capitaine, M. Outlying eigenvalues of a polynomial in large random matrices. arXiv:1703.08102 (2018).

[10] Belinschi, S. T., Bercovici, H., Capitaine, M., and Yvier, M. Outliers in the spectrum of large deformed unitarily invariant models. Ann. Probab. 45, 6A (11 2017), 3571–3625.

[11] Belinschi, S. T., Mai, T., and Speicher, R. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. Journal für die reine und angewandte Mathematik (Crelles Journal) (03 2013).

[12] Bercovici, H., and Voiculescu, D. Regularity questions for free convolution. In Nonselfadjoint operator algebras, operator theory, and related topics, vol. 104 of Oper. Theory Adv. Appl. Birkhäuser, Basel, 1998, pp. 37–47.

[13] Biane, P. Processes with free increments. Mathematische Zeitschrift 227, 1 (Jan 1998), 143–174.
[14] Capitaine, M., and Donati-Martin, C. Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* 56, 2 (2007), 767–803.

[15] Collins, B., and Male, C. The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér.* (4) 47, 1 (2014), 147–163.

[16] Dykema, K. J. On the s-transform over a Banach algebra. *Journal of Functional Analysis* 231, 1 (2006), 90–110.

[17] Erdős, L., Krüger, T., and Nemish, Y. Local laws for polynomials of Wigner matrices. *Journal of Functional Analysis* 278, 12 (2020), 108507.

[18] Erdős, L., Schlein, B., and Yau, H.-T. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* 37, 3 (05 2009), 815–852.

[19] Haagerup, U., and Thorbjørnsen, S. A new application of random matrices: Ext(C∗ red(F2)) is not a group. *Ann. of Math.* (2) 162, 2 (2005), 711–775.

[20] Kantorovič, L. V. Functional analysis and applied mathematics. *Uspehi Matem. Nauk (N.S.)* 3, 6(28) (1948), 89–185.

[21] Kargin, V. Subordination for the sum of two random matrices. *Ann. Probab.* 43, 4 (07 2015), 2119–2150.

[22] Mai, T. *On the analytic theory of non-commutative distributions in free probability.* PhD thesis, 01 2017.

[23] Male, C. The norm of polynomials in large random and deterministic matrices. *Probab. Theory Related Fields* 154, 3-4 (2012), 477–532. With an appendix by Dimitri Shlyakhtenko.

[24] Schultz, H. Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* 131, 2 (2005), 261–309.

[25] Schützenberger, M. P. On the definition of a family of automata. *Information and Control* 4 (1961), 245–270.

[26] Shiryaev, A. N. *Probability*, second ed., vol. 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[27] Speicher, R. Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. *Mem. Amer. Math. Soc.* 132, 627 (1998), x+88.

[28] Voiculescu, D. Addition of certain non-commuting random variables. *Journal of Functional Analysis* 66, 3 (1986), 323 – 346.

[29] Voiculescu, D. Limit laws for random matrices and free products. *Inventiones mathematicae* 104, 1 (Dec 1991), 201–220.

[30] Voiculescu, D. The analogues of entropy and of Fisher’s information measure in free probability theory. I. *Comm. Math. Phys.* 155, 1 (1993), 71–92.

[31] Voiculescu, D. Operations on certain non-commutative operator-valued random variables. *Astérisque*, 232 (1995), 243–275. Recent advances in operator algebras (Orléans, 1992).

[32] Voiculescu, D. The coalgebra of the free difference quotient and free probability. *Internat. Math. Res. Notices*, 2 (2000), 79–106.