Convergence Rate of Empirical Spectral Distribution of Random Matrices From Linear Codes
Chin Hei Chan, Vahid Tarokh, and Maosheng Xiong

Abstract—It is known that the empirical spectral distribution of random matrices obtained from linear codes of increasing length converges to the well-known Marchenko-Pastur law, if the Hamming distance of the dual codes is at least 5. In this paper, we prove that the convergence rate in probability is at least of the order $n^{-1/4}$ where $n$ is the length of the code.

Index Terms—Group randomness, linear code, dual distance, empirical spectral measure, random matrix theory, Marchenko-Pastur law.

I. INTRODUCTION

Random matrix theory is the study of matrices whose entries are random variables. Of particular interest is the study of eigenvalue statistics of random matrices such as the empirical spectral measure. It has been broadly investigated in a wide variety of areas, including statistics [24], number theory [17], economics [18], theoretical physics [23] and communication theory [22].

Most of the matrix models in the literature are random matrices with independent entries. In a recent series of work (initiated in [2] and developed further in [1], [3], [25]), the authors considered a class of sample-covariance type matrices formed randomly from linear codes over a finite field, and proved that if the Hamming distance of the dual codes is at least 5, then as the length of the codes goes to infinity, the empirical spectral distribution of the random matrices obtained in this way converges to the well-known Marchenko-Pastur (MP) law. Since truly random matrices (i.e., random matrices with i.i.d. entries) of large size satisfy this Marchenko-Pastur (MP) law. Since truly random matrices (i.e., random matrices with i.i.d. entries) of large size satisfy this property, this can be interpreted as that sequences from linear codes of dual distance at least 5 behave like random among themselves. This is a new pseudo-random test for sequences and is called a “group randomness” property [1]. It may have many potential applications.

How fast does the empirical spectral distribution converge to the MP law? This question is interesting in its own rights, and important in applications as one may wish to use linear codes of proper length to generate pseudo-random matrices. Along with proving the convergence in expectation, the authors in [25] obtained a convergence rate of the order $\frac{\log \log n}{\log n}$ where $n$ is the length of the code. This is quite unsatisfactory, as the numerical data showed clearly that the convergence is rather fast with respect to $n$. In this paper, we prove that the convergence rate is indeed at least of the order $n^{-\frac{1}{4}}$ in probability. This substantially improves the previous result.

To introduce our main result, we need some notation.

Let $\mathcal{C}$ be a linear code of length $n$ and dimension $k$ over the finite field $\mathbb{F}_q$ of order $q$, where $q$ is a prime power. $\mathcal{C}$ is called an $[n, k, d]_q$ linear code for short. The most interesting case is the binary linear codes, corresponding to $q = 2$. The dual code $\mathcal{C}^\perp$ consists of the $n$-tuples in $\mathbb{F}_q^n$ which are orthogonal to all codewords of $\mathcal{C}$ under the standard inner product. Clearly, $\mathcal{C}^\perp$ is also a linear code. Denote by $d^\perp$ the Hamming distance of $\mathcal{C}^\perp$. It is called the dual distance of $\mathcal{C}$.

Let $\psi : \mathbb{F}_n \rightarrow \mathbb{C}^\times$ be the standard additive character. To be more precise, if $\mathbb{F}_q$ has characteristic $l$, which is a prime number, then $\psi$ is given by $\beta \mapsto \exp\left(2\pi i Tr_{q/l}(\beta)/l\right)$, where $Tr_{q/l}$ is the absolute trace mapping from $\mathbb{F}_q$ to $\mathbb{F}_l$. In particular, if $q = l = 2$, then the map $\psi : \mathbb{F}_2 := \{0, 1\} \rightarrow \{-1, 1\}$ is defined as $\beta \mapsto (-1)^\beta$. We extend $\psi$ component-wise to $\mathbb{F}_q^n$ and obtain the map $\psi : \mathbb{F}_q^n \rightarrow (\mathbb{C}^\times)^n$. Denote $\mathcal{D} := \psi(\mathcal{C})$.

Denote by $\Phi_n$ a $p \times n$ matrix whose rows are chosen from $\mathcal{D}$ uniformly and independently. This makes the set $\mathcal{D}^p$ a probability space with the uniform probability.

Let $G_n$ be the Gram matrix of $X = \frac{1}{\sqrt{n}} \Phi_n$, that is,

$$G_n = XX^* = \frac{1}{n} \Phi_n \Phi_n^*,$$

where $A^*$ means the conjugate transpose of the matrix $A$. Let $\mu_n$ be the empirical spectral measure of $G_n$, that is,

$$\mu_n = \frac{1}{p} \sum_{j=1}^{p} \delta_{\lambda_j},$$

where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ are the eigenvalues of $G_n$ and $\delta_\lambda$ is the Dirac measure at the point $\lambda$. Note that $\mu_n$ is a random measure, that is, for any interval $I \subset \mathbb{R}$, the value $\mu_n(I)$ is a random variable with respect to the probability space $\mathcal{D}^p$. Our main result is as follows.

Theorem 1: Assume that $y := p/n \in (0, 1)$ is fixed. If $d^\perp \geq 5$, then

$$|\mu_n(I) - \varphi_{\text{MP}, y}(I)| \leq n^{-\frac{1}{4}}$$
uniformly for all intervals $I \subset \mathbb{R}$. Here $q_{\text{MP,y}}$ is the empirical spectral measure of the Marchenko-Pastur law whose density function is given by

$$d q_{\text{MP,y}}(x) = \frac{1}{2\pi y} \sqrt{(b - x)(x - a)} I_{[a,b]} dx,$$

where the constant $a$ and $b$ are defined as

$$a = (1 - \sqrt{y})^2, b = (1 + \sqrt{y})^2,$$

and $I_{[a,b]}$ is the indicator function of the interval $[a, b]$.

**Remark 1:** The symbol $\prec$ in (3) is a standard notation for “stochastic domination” in probability theory (see [8] for details). Here it means that for any $\varepsilon > 0$ and any $D > 0$, there is a quantity $n_0(\varepsilon, D)$, such that whenever $n \geq n_0(\varepsilon, D)$, we have

$$\sup_{\mathbb{R}} P \left[ |\mu_n(I) - q_{\text{MP,y}}(I)| > n^{-\frac{1}{2} + \varepsilon} \right] \leq n^{-D},$$

where $P$ is the probability with respect to $D^p$ and the supremum can also be taken over all linear codes $C$ of length $n$ over $\mathbb{F}_q$ with $d^\perp \geq 5$.

**Remark 2:** [Theorem 1] is reminiscent of a well-known result of Sidelnikov ([16], [21]) which states that for any $[n, k, d]$ binary linear code $C$ with dual distance $d^\perp \geq 3$, one has

$$|A(z) - \Phi(z)| \leq \frac{9}{\sqrt{d^\perp}}.$$

Here $A(z)$ is the normalized cumulative weight distribution of $C$ and

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

Hence the “randomness” of the weight distribution of $C$ is ensured if $d^\perp$ is sufficiently large. With this respect Theorem 1 is a little surprising since the condition $d^\perp \geq 5$ already ensures a fast convergence rate to the MP law (see Equation (3)). We emphasize that the condition $d^\perp \geq 5$ is also optimal: the work [1] showed that the empirical spectral distribution of random matrices based on binary Simplex (shortened first-order Reed-Muller) codes with $d^\perp = 3$ does not converge to the MP law; a similar calculation shows that the empirical spectral distribution of random matrices based on binary first-order Reed-Muller codes with $d^\perp = 4$ does not converge to the MP law either.

**Remark 3:** It seems quite possible to extend our results to nonlinear codes, where the dual distance $d^\perp$ is defined in [16, Chapter 5]. In this paper, however, we focus only on linear codes.

**Remark 4:** For application purposes, from Theorem 1, binary linear codes of dual distance 5 with large length and small dimension are desirable as they can be used to generate random matrices efficiently. Here we mention two constructions of binary linear codes with parameters $[2^m - 1, 2m]$ and dual distance 5. The first family is the dual of primitive double-error correcting BCH codes ([13]). The second family of such codes, which includes the well-known Gold codes, can be constructed as follows: Let $f : \mathbb{F}_{2^n} \rightarrow \mathbb{F}_{2^n}$ be a function such that $f(0) = 0$. Let $n = 2^m - 1$ and $\alpha$ be a primitive element of $\mathbb{F}_{2^n}$. Define a matrix

$$H_f := \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \end{bmatrix}.$$

Given a basis of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2 := \{0, 1\}$, each element of $\mathbb{F}_{2^n}$ can be identified as an $m \times 1$ column vector in $\mathbb{F}_2$, hence the $H_f$ above can be considered as a binary matrix of size $2m \times n$. Denote by $C_f$ the binary linear code obtained from $H_f$ as a generator matrix. Note that $C_f$ has length $2^m - 1$ and dimension $2m$. It is known that the dual distance of $C_f$ is 5 if and only if $f$ is an almost perfect nonlinear (APN) function [10], [20]. Since there are many APNs when $m$ is odd, this provides a general construction of binary linear codes of dual distance 5 which may be of interest for applications.

**Remark 5:** Binary linear codes can be used to construct deterministic sensing matrices which satisfy the important “statistical restricted isometry property” ( [7], [11]). From an $[n, k, d]$ binary linear code $C$, letting $N = 2^k$ and using the same notation as in Theorem 1 for easy comparison, one obtains an $N \times n$ matrix $\Phi_C$ whose rows consist of all the codewords of $C$ under the map $\psi$ (so that $\Phi_C$ is a matrix of entries $\pm 1$). The probability space is choosing $p$ distinct rows uniformly at random from $\Phi_C$ to form a submatrix $\Phi_p$. Then the sensing matrix $\frac{1}{\sqrt{n}} \Phi_p$ is said to have the $(p, \delta, \varepsilon)$-StRIP if

$$P \left( \left\| \frac{1}{\sqrt{n}} \Phi_p \Phi_p^* - I_p \right\|_2 > \delta \right) < \varepsilon.$$

Equation (6) states that the event that all the eigenvalues of the matrix $\frac{1}{\sqrt{n}} \Phi_p \Phi_p^*$ lie in the interval $[1 - \delta, 1 + \delta]$ has probability $1 - \varepsilon$. Since the probability space in (6) is essentially the same as that in (3) of Theorem 1, it can be seen that Theorem 1 provides a strong and much more precise description about how the eigenvalues of $\frac{1}{\sqrt{n}} \Phi_p \Phi_p^*$ are distributed along the real line, but Theorem 1 falls short of proving (6). It seems possible to prove (6) by considering a slightly different normalization of the random matrix from linear codes as done in ([12]). We might come back to this question in the future.

For truly random matrices with i.i.d. entries, finding the rate of convergence has been a long-standing question, starting from [4], [5], [15] in early 1990s. Great progress has been made in the last 10 years, culminating in achieving the optimal rate of convergence $n^{-1}$ where $n$ is the size of the matrix (see [8], [14], [19]). The major technique is the use of the Stieltjes transform. In this paper we also use this technique.

The convergence rate problem for the empirical spectral distribution of large sample covariance random matrices has been studied for example in [5], [9], and in particular in [9] an optimal rate of convergence of order $n^{-1}$ was obtained under quite general conditions. However, despite our best effort, none of the techniques in [5] and [9] can be easily applied directly to our setting. Instead we use a combination of ideas from [5] and [9]. Moreover, it is not clear to us what the best rate of convergence is under general linear codes in terms of dual distance. It might be interesting to find out a general sufficient condition on the dual distance that guarantees the optimal rate of convergence $n^{-1}$. We hope to stress this problem in the future.
The paper is now organized as follows. In Section II, Preliminaries we introduce the main tool, the Stieltjes transform and related formulas and lemmas which will play important roles in the Proof of Theorem 1. In Section III, we show how Theorem 1 can be derived directly from a major statement in terms of the Stieltjes transform (Theorem 4). While the argument is standard, it is quite technical and non-trivial. To streamline the idea of the proof, we put some of the arguments in Section IV-C Appendix. In Section IV, we give a detailed proof of Theorem 4.

II. Preliminaries

A. Stieltjes Transform

In this section we recall some basic knowledge of Stieltjes transform. Interested readers may refer to [6, Chapter B.2] for more details.

Let $F$ be an arbitrary real function with bounded variation, and $\mu$ be the corresponding (signed) measure. The Stieltjes transform of $F$ (or $\mu$) is defined by

$$s(z) := \int_{-\infty}^{\infty} \frac{dF(x)}{x-z} = \int_{\mathbb{R}} \frac{\mu(dx)}{x-z},$$

where $z$ is a complex variable outside the support of $F$ (or $\mu$). In particular, $s(z)$ is well-defined for all $z \in \mathbb{C}^+ := \{ z \in \mathbb{C} : \Re z > 0 \}$, the upper half complex plane. Here $\Re z$ is the imaginary part of $z$.

It can be verified that $s(z) \in \mathbb{C}^+$ for all $z \in \mathbb{C}^+$. The complex variable $z$ is commonly written as $z = E + i\eta$ for $E, \eta \in \mathbb{R}$.

The Stieltjes transform is useful because a function of bounded variation (or signed measures) which will play important roles in the Proof of Theorem 1. In Section III, we show how the convergence of Stieltjes transform is both necessary and sufficient for the convergence of the underlying distribution (see [6, Theorem B.9]).

B. Resolvent Identities and Formulas for Green Function Entries

Let $X = (X_{jk})$ be a $p \times n$ matrix. Denote by $G$ the Green function of $XX^*$, that is,

$$G := G(z) = (XX^* - zI)^{-1},$$

where $z \in \mathbb{C}^+$ and $I$ is the identity matrix.

Given a subset $T \subset [1 \ldots p] := \{1, 2, \ldots, p\}$, let $X^{(T)}$ be the $p \times n$ matrix whose $(j,k)$-th entry is defined by $(X^{(T)})_{jk} := X_{jk}$. In addition, let $G^{(T)}$ be the Green function of $X^{(T)}X^{(T)^*}$. We write $R$ and $R^{(T)}$ as the Green functions of $X^*X$ and $X^{(T)*}X^{(T)}$ respectively. Then for $\ell \in [1 \ldots p] \setminus T$, we have [9, (3.8)]

$$\frac{1}{G^{(T)}_{\ell \ell}} = -z - \sum_{j,k} X_{jk} R^{(T)}_{jk} X_{k\ell},$$

where the indices $j,k$ vary in $[1 \ldots n]$, and $R^{(T)}_{jk}$ is the $(j,k)$-th entry of the matrix $R^{(T)}(T \cup \{\ell\})$.

The two Green functions $G^{(T)}$ and $R^{(T)}$ are related by the following identity ( [9, Lemma 3.9]):

$$\text{Tr}G^{(T)} - \text{Tr}R^{(T)} = \frac{n - (p - |T|)}{z}. \quad (8)$$

Here $|T|$ is the cardinality of the set $T$, and $\text{Tr}A$ is the trace of the matrix $A$.

Recall that we denote $\eta = \Re z$. Then we have the following eigenvalue interlacing property ( [9, Lemma 3.10])

$$|\text{Tr}G^{(T)} - \text{Tr}G| \leq C\eta^{-1}, \quad (9)$$

where $C$ is a constant depending on the set $T$ only, and also the Wald’s identity (see [9, (3.14)]) or [8, (3.6)])

$$\sum_k |R^{(T)}_{jk}|^2 = \eta^{-1} \Im R^{(T)}_{jj}. \quad (10)$$

C. Stieltjes Transform of the Marchenko-Pastur Law

The Stieltjes transform $s_{\text{MP}, y}$ of the Marchenko-Pastur distribution given in (4) can be computed as (see [5])

$$s_{\text{MP}, y}(z) = -\frac{y + z - 1 - \sqrt{(y + z - 1)^2 - 4yz}}{2yz}, \quad (11)$$

It is well-known that $s_{\text{MP}, y}(z)$ is the unique function that satisfies the equation of $u(z)$ in

$$u(z) = \frac{1}{1 - y - z - yzu(z)} \quad (12)$$

such that $\Im u(z) > 0$ whenever $\eta := \Re z > 0$.

If a function $f : \mathbb{C}^+ \to \mathbb{C}^+$ satisfies Equation 12 with a small perturbation, we then expect that $f(z)$ should be quite close to $s_{\text{MP}, y}(z)$ as well. This is quantified by the following result. First, we define

$$\kappa := \min\{|E-a|, |E-b|\} \quad (13)$$

where $a$ and $b$ are constants given in (5) and for a fixed constant $\tau > 0$, we define

$$S_{\tau} := \left\{ z = E + i\eta : \kappa \leq \tau^{-1}, n^{-1/4}\tau \leq \eta \leq \tau^{-1} \right\}. \quad (14)$$

**Lemma 2:** [9, Lemma 4.5] Suppose the function $\delta : S_{\tau} \to (0, \infty)$ satisfies:

1. $n^{-2} \leq \delta(z) \leq \varepsilon$ for some fixed constant $\varepsilon > 0$ for all $z \in S_{\tau}$;
2. $\delta$ is Lipschitz continuous with Lipschitz constant $n$;
3. for each fixed $E$, the function $\eta \mapsto \delta(E + i\eta)$ is nonincreasing for $\eta > 0$.

Suppose $u : S_{\tau} \to \mathbb{C}$ is the Stieltjes transform of a probability measure satisfying

$$u(z) = \frac{1}{1 - y - z - yzu(z) + \Delta(z)} \quad (15)$$

for some $\Delta(z)$.

Fix $z \in S_{\tau}$ and define

$$L(z) := \left\{ w \in S_{\tau} : \Re w = \Re z, \Im w \in [\Re z, 1] \cap (n^{-5}N) \right\}.$$
where $\Re z$ is the real part of $z \in \mathbb{C}$. Suppose that
\[ |\Delta(w)| \leq \delta(w), \quad \forall w \in L(z) \cup \{z\}. \tag{16} \]
Then we have
\[ |u(z) - s_{\text{MP}, y}(z)| \leq \frac{C\delta(z)}{\sqrt{\kappa + \eta + \delta(z)}}, \]
where $\kappa$ is the $z$-dependent variable defined as in (13).

### D. Convergence of Stieltjes Transform in Probability

The following result is useful to bound the convergence rate of a random Stieltjes transform in probability.

**Lemma 3:** Let $\mathcal{M}$ be a $p \times n$ random matrix with independent rows, $S = \mathcal{M}\mathcal{M}^*$, and $m(z)$ be the Stieltjes transform of the empirical spectral distribution of $S$. Then
\[ \mathbb{P}(\{m(z) - Em(z)\} > r) \leq 2 \exp\left( -\frac{pn^2r^2}{8} \right). \]

**Proof of Lemma 3:** Note that the $(j, k)$-th entry of $S$ is simply the inner product of the $j$-th and $k$-th rows of $\mathcal{M}$. Hence varying one row of $\mathcal{M}$ only gives an additive perturbation of $S$ of rank at most two. Applying the resolvent identity [8, (2.3)], we see that the Green function is also only affected by an additive perturbation by a matrix of rank of at most two and operator norm at most $2\eta^{-1}$. Then the desired result follows directly by applying the McDiarmid’s Lemma [8, Lemma F.3].

For the purpose of this paper, we define an $n$-dependent event $\Xi$ to hold with high probability if for any $D > 0$, there is a quantity $n_0 = n_0(D) > 0$ such that $\mathbb{P}(\Xi) \geq 1 - n^{-D}$ for any $n > n_0$.

### III. PROOF OF THEOREM 1

From this section onwards, let $\mathcal{C}$ be a linear code of length $n$ over $\mathbb{F}_q$ with dual distance $d^2 \geq 5$. Let $\psi: \mathbb{F}_q \to \mathbb{C}^\times$ be the standard additive character, extended to $\mathbb{F}_q^n$ component-wisely. Write $D = \psi(\mathcal{C})$.

Let $\Phi_n$ be a $p \times n$ random matrix whose rows are picked from $D$ uniformly and independently. This makes $\mathcal{D}_n$ a probability space. Let $y := p/n \in (0, 1)$ be fixed. Write $X = n^{-1/2}\Phi_n$ and $G_n = XX^*$ the Gram matrix of $X$. Furthermore, let $\mu_n$ be the empirical spectral measure of $G_n$ given by (2).

Denote $s_{\mathcal{G}_n}(z)$ to be the Stieltjes transform of $\mu_n$, which is given by
\[ s_{\mathcal{G}_n}(z) = \frac{1}{p} \sum_{j=1}^p \frac{1}{\lambda_j - z} = \frac{1}{p} \text{Tr}G, \]
where $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of the matrix $G_n$, and $G$ is the Green function of $G_n$, that is, $G := G(z) = (G_n - zI)^{-1}$. Note that in this setting this Stieltjes transform $s_{\mathcal{G}_n}(z)$ is itself a random variable.

Denote
\[ s_n(z) := \mathbb{E}s_{\mathcal{G}_n}(z) = \frac{1}{p} \mathbb{E}\text{Tr}G. \tag{17} \]
Here $\mathbb{E}$ is the expectation with respect to the probability space $\mathcal{D}_n$.

### A. An Equation for $s_n(z)$

In the following result, we write $s_n(z)$ defined in (17) in the form of the equation (12) with a small perturbation.

**Theorem 4:** For any $z \in \mathbb{S}_\tau$,
\[ s_n(z) = \frac{1}{1 - y - z - yzs_n(z) + \Delta(z)} \]
where $\Delta(z) = O(n^{-1}y^{-3})$.

We remark that Theorem 4 is a major technical result regarding the expected Stieltjes transform $s_n(z)$, from which Theorem 1 can be derived directly without reference to linear codes at all. The proof of Theorem 4 is, however, quite complicated and is directly related to properties of linear codes. To streamline the idea of the proof, here we assume Theorem 4 and sketch a proof of Theorem 1. The proof of Theorem 4 is postponed to Section IV.

### B. Proof of Theorem 1

Assuming Theorem 4, we can first estimate the term $|s_{\mathcal{G}_n}(z) - s_{\text{MP}, y}(z)|$, following ideas from [8] and [9].

**Theorem 5:** Assume that Theorem 4 holds. Then for any fixed $z \in \mathbb{S}_\tau$, we have
\[ |s_{\mathcal{G}_n}(z) - s_{\text{MP}, y}(z)| \leq n^\tau(n^{-1/4} + n^{-1}\eta^{-7/2}) \]
with high probability.

**Proof of Theorem 5:** We can check that all the conditions of Lemma 2 are satisfied: first by Theorem 4 we see that (15) holds for $u(z) = s_n(z)$; in addition, (16) holds for $\delta(z) = \delta(E + iy) = Cn^{-1}\eta^{-3}$, and this function is independent of $E$, nonincreasing in $\eta > 0$ and Lipschitz continuous with Lipschitz constant $Cn^{-4}\tau < n$. Hence by Lemma 2, we have
\[ |s_n(z) - s_{\text{MP}, y}(z)| \leq \frac{C\delta(z)}{\sqrt{\kappa + \eta + \delta(z)}}. \]
Note that in $\mathbb{S}_\tau$ we have $\delta(z) = O(n^{-1/4 - 3\tau}) = o(\eta)$. Therefore we have
\[ |s_n(z) - s_{\text{MP}, y}(z)| = O(n^{-1}\eta^{-7/2}) \tag{18} \]
for all $z \in \mathbb{S}_\tau$.

Now Lemma 3 implies that
\[ \mathbb{P}\left(|s_{\mathcal{G}_n}(z) - s_n(z)| > n^{-1/4}\right) \leq 2 \exp\left(-\frac{yn(n^{-1/4})^4}{8}\right) = 2 \exp\left(-\frac{yn^{4\tau}}{8}\right) \leq n^{-D} \]
on $\mathbb{S}_\tau$, for any $D > 0$ and large enough $n$. Combining this with (18) completes the proof of Theorem 5.

Finally, armed with Theorem 5, we can derive Theorem 1 from a standard application of the Helffer-Sjöstrand formula in random matrix theory. The argument is essentially complex analysis. Interested readers may refer to Section IV-C Appendix for details.

### IV. PROOF OF THEOREM 4

Now we give a detailed proof of Theorem 4, in which the condition that $d^\perp \geq 5$ becomes essential.
A. Linear Codes With $d^± ≥ 5$

Recall the notation from the beginning of Section III. Let $C$ be a linear code of length $n$ over $\mathbb{F}_q$. First, is a simple orthogonality result regarding $C$.

**Lemma 6:** Let $a \in \mathbb{F}_q^n$. Then

$$\frac{1}{\# C} \sum_{c \in C} \psi(a \cdot c) = \begin{cases} 1 & (a \in C^⊥), \\ 0 & (a \notin C^⊥). \end{cases}$$

Here $a \cdot c$ is the usual inner product between the vectors $a$ and $c$.

As in Section III, let $\Phi_n$ be a $p \times n$ random matrix whose rows are picked from $D = \psi(C)$ uniformly and independently and let $X = n^{-1/2} \Phi_n$. Denote by $X_{jk}$ the $(j,k)$-th entry of $X$.

**Corollary 7:** Assume $d^± ≥ 5$. Then for any $\ell \in [1..p]$, 

(a) $E(X_{jk} X_{tk}) = 0$ if $j \neq k$;

(b) $E(X_{jk} X_{tk} X_{tk} X_{tk}) = 0$ if the indices $j,t,k,s$ do not come in pairs. If the indices come in pairs, then $E(X_{jk} X_{tk} X_{tk} X_{tk}) \leq n^{-2}$.

Here $E$ is the expectation with respect to the probability space $D^p$.

**Proof of Corollary 7:** For simplicity, denote by $e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{F}_q^n$ the vector with a 1 at the $i$-th entry and 0 at all other places.

(a) It is easy to see that

$$E(X_{jk} X_{tk}) = n^{-1}(\# C)^{-1} \sum_{c \in C} \psi(c_j - c_k) = n^{-1}(\# C)^{-1} \sum_{c \in C} \psi((e_j - e_k) \cdot c).$$

As $d^± ≥ 5$ and $j \neq k$, so $0 \neq e_j - e_k \notin C^⊥$, and the desired result follows directly from Lemma 6.

(b) Again we can check that

$$E(X_{jk} X_{tk} X_{tk} X_{tk}) = n^{-2}(\# C)^{-1} \sum_{c \in C} \psi((e_j + e_t - e_k - e_s) \cdot c).$$

If the indices $j,t,k,s$ do not come in pairs, since $d^± ≥ 5$, we have $0 \neq e_j + e_t - e_k - e_s \notin C^⊥$, and the result is zero by Lemma 6; if the indices $j,t,k,s$ do come in pairs, noting that $|X_{jk}| = n^{-1/2}$, we also obtain the desired estimate. This completes the proof of Corollary 7. \hfill \Box

B. Resolvent Identities

We start with the resolvent identity (7) for $T = \emptyset$. The sum on the right of (7) can be written as

$$\frac{1}{G_{\ell\ell}} = -z - \frac{z}{n} \text{TrR}^{(\ell)} - Z_{\ell\ell} = -z - \frac{z}{n} \left( \text{TrG}^{(\ell)} - \frac{n - p + 1}{z} \right) - Z_{\ell\ell} = 1 - y - z - y z s_n(z) + Y_{\ell},$$

where

$$Y_{\ell} = y z s_n(z) - \frac{z}{n} \text{TrG}^{(\ell)} + \frac{1}{n} - Z_{\ell\ell}.$$  \hspace{1cm} (20)

Using (7) and (8) we have

$$\frac{1}{G_{\ell\ell}} = -z - \frac{z}{n} \text{TrR}^{(\ell)} - Z_{\ell\ell} = -z - \frac{z}{n} \left( \text{TrG}^{(\ell)} - \frac{n - p + 1}{z} \right) - Z_{\ell\ell} = 1 - y - z - y z s_n(z) + Y_{\ell}.$$  \hspace{1cm} (21)

C. Estimates of $Z_{\ell\ell}$ and $Y_{\ell}$

We now give estimates on the $(z$-dependent) random variable $Z_{\ell\ell}$. First, given $T \subset [1..p]$, we denote $E(T)(\cdot) := E(T)(\cdot X_T)$.

**Lemma 8:** For any $\ell \in [1..p]$, we have

(a) $E(\ell) Z_{\ell} = E Z_{\ell} = 0$;

(b) $E|Z_{\ell}|^2 = O(n^{-1} \eta^{-2})$.

**Proof of Lemma 8:** (a) From the definition of $Z_{\ell}$ in (19), we have

$$E(\ell) Z_{\ell} = z \sum_{j \neq k} R_{jk}^{(\ell)} E(X_{jk} X_{tk}) = 0,$$  \hspace{1cm} (20)

where the first equality follows from the fact that rows of $X$ are independent, and second equality follows from statement (a) of Corollary 7. The proof of the result on $EZ_{\ell}$ is similar by replacing $R_{jk}^{(\ell)}$ with $E R_{jk}^{(\ell)}$.

(b) Expanding $|Z_{\ell}|^2$ and taking expectation $E$ inside, noting that the rows of $X$ are independent, we have

$$E|Z_{\ell}|^2 = |z|^2 E \left| \sum_{j \neq k} X_{jk} R_{jk}^{(\ell)} X_{tk} \right|^2 = |z|^2 \sum_{j \neq k, s \neq t} \left( R_{jk}^{(\ell)} R_{st}^{(\ell)} \right) E(X_{jk} X_{tk} X_{tk} X_{ts}).$$

Since $d^± ≥ 5$, by using statement (b) of Corollary 7 and Wald’s identity (10), together with the trivial bound $|R_{jj}^{(\ell)}| \leq \eta^{-1}$, we obtain

$$E|Z_{\ell}|^2 = C |z|^2 n^2 \sum_{j,k} E|R_{jk}^{(\ell)}|^2 \leq C n^{-2}.$$  \hspace{1cm} (21)

Here $C$ is a generic constant which may be different in each occurrence.

The above estimations lead to the following estimations about $Y_{\ell}$.

**Lemma 9:** For any $\ell \in [1..p]$, we have

(a) $E Y_{\ell} = O(n^{-1} \eta^{-1})$;

(b) $E|Y_{\ell}|^2 = O(n^{-1} \eta^{-2})$. 


Proof of Lemma 9: (a) By (21) we get
\[
\mathbb{E}Y_\ell = \frac{z}{n} \mathbb{E}(\text{Tr}G - \text{Tr}G^t) + \frac{1}{n} \mathbb{E}Z_\ell
\]
where the second equality follows from (a) of Lemma 8. Using (9) we easily obtain
\[
|\mathbb{E}Y_\ell| \leq \frac{C|z|}{n\eta} \leq \frac{C}{n\eta}
\]
(b) We split \(\mathbb{E}|Y_\ell|^2\) as
\[
\mathbb{E}|Y_\ell|^2 = \mathbb{E}|Y_\ell - \mathbb{E}Y_\ell|^2 + |\mathbb{E}Y_\ell|^2 = V_1 + V_2 + |\mathbb{E}Y_\ell|^2,
\]
where
\[
V_1 = \mathbb{E}|Y_\ell - \mathbb{E}(Y_\ell)|^2, \quad V_2 = \mathbb{E}|\mathbb{E}(Y_\ell)|^2 - \mathbb{E}Y_\ell.
\]
We first estimate \(V_1\). By the definition of \(Y_\ell\) in (21) and applying (a) of Lemma 8, we see that
\[
Y_\ell - \mathbb{E}(Y_\ell) = -Z_\ell + \mathbb{E}(Z_\ell) = -Z_\ell.
\]
Then by (b) of Lemma 8 we obtain
\[
V_1 = \mathbb{E}|Z_\ell|^2 = O(n^{-1}\eta^{-2}). \quad (23)
\]
Next we estimate \(V_2\). Again by (21) and Lemma 8, we have
\[
\mathbb{E}Z_\ell - \mathbb{E}Y_\ell = -\frac{z}{n} \left( \mathbb{E}\text{Tr}G^t - \mathbb{E}\text{Tr}G^t \right) - (\mathbb{E}Z_\ell - \mathbb{E}Y_\ell) = -\frac{z}{n} \left( \mathbb{E}\text{Tr}G^t - \mathbb{E}\text{Tr}G^t \right).
\]
Hence
\[
V_2 = \frac{|z|^2}{n^2} (\mathbb{E}|\text{Tr}G^t - \mathbb{E}\text{Tr}G|^2)
\]
\[
= \frac{|z|^2}{n^2} \sum_{m \neq \ell} \mathbb{E}|\text{Tr}G^t - \mathbb{E}\text{Tr}G|^2, \quad (24)
\]
where \(T_0 := 0\) and \(T_m := [1 \ldots m]\) for \(m \in [1 \ldots p]\). The second equality follows from applying successively the law of total variance to the rows of \(\Phi_n\).

For \(m \neq \ell\), denote \(\gamma_m := \mathbb{E}T_m T_m^{-1} \mathbb{E}\text{Tr}G^t - \mathbb{E}T_m \mathbb{E}\text{Tr}G^t\) and \(\sigma_m := \mathbb{E}T_m^{-1} \mathbb{E}\text{Tr}G^t - \mathbb{E}T_m \mathbb{E}\text{Tr}G^t\). It is easy to check that
\[
\gamma_m = \mathbb{E}T_m^{-1} \sigma_m - \mathbb{E}T_m \sigma_m.
\]
Thus by (9) we have \(|\gamma_m| \leq C\eta^{-1}\).

Putting this into (24) yields
\[
V_2 \leq \frac{C|z|^2}{n^2\eta} \leq \frac{C}{n^2\eta^2}.
\]
Plugging the estimates of \(\mathbb{E}Y_\ell\) in statement (a), \(V_1\) in (23) and \(V_2\) above into the equation (22), we obtain the desired estimate of \(|\mathbb{E}|Y_\ell|^2\). This finishes the proof of Lemma 9.

We can now complete the proof of Theorem 4.

Proof of Theorem 4: Taking reciprocal and then expectation on both sides of (20), we get
\[
\mathbb{E}G_{\ell\ell} = \mathbb{E}\frac{1}{\alpha_n + Y_\ell} = \frac{1}{\alpha_n} + A_\ell = \frac{1}{\alpha_n + \Delta_\ell},
\]
where
\[
A_\ell = \mathbb{E}\frac{1}{\alpha_n + Y_\ell} - \frac{1}{\alpha_n} = -\frac{1}{\alpha_n^2} \mathbb{E}Y_\ell + \frac{1}{\alpha_n^2} \mathbb{E}Y_\ell^2 + \frac{Y_\ell}{\alpha_n + Y_\ell},
\]
and
\[
\Delta_\ell = \left( \frac{1}{\alpha_n} + A_\ell \right)^{-1} - \frac{1}{\alpha_n} = -\frac{\alpha_n^2 A_\ell}{1 + \alpha_n A_\ell}.
\]
Multiplying \(\alpha_n^2\) on both sides of (26) and using the estimate \(|\alpha_n + Y_\ell| \geq \eta\), we obtain
\[
|\alpha_n^2 A_\ell| = \left| -\mathbb{E}Y_\ell + \frac{\mathbb{E}Y_\ell^2}{\alpha_n + Y_\ell} \right| \leq |\mathbb{E}Y_\ell| + \frac{1}{\eta} |\mathbb{E}|Y_\ell|^2. \quad (28)
\]
Putting the results of Lemma 9 in (28), we get
\[
|\alpha_n^2 A_\ell| \leq \frac{C}{n\eta^3}.
\]
Using the fact that \(|\alpha_n|^{-1} \leq \eta^{-1}\), we have \(|\alpha_n A_\ell| \leq Cn^{-1}\eta^{-4}\), so that \(|1 + \alpha_n A_\ell| \geq C\).

Substituting all these into (27) yields
\[
|\Delta_\ell| \leq \frac{C}{n\eta^3}
\]
for all \(z \in S_r\).

Then the theorem follows directly from summing both sides of (25) for all \(\ell \in [1 \ldots p]\) and then dividing both sides by \(p\). \(\Box\)

APPENDIX

In this section, we use Helffer-Sjöstrand formula to prove Theorem 1 from Theorem 5. This is a standard procedure well-known in random matrix theory. We follow the idea based on [8, Appendix C].

First we define the signed measure \(\hat{\mu}_n\) and its Stieltjes transform \(\hat{s}_n\) by
\[
\hat{\mu}_n := \mu_n - \delta_{\text{MP},y}, \hat{s}_n(z) = \int \frac{\hat{\mu}_n(dx)}{x - z} = s_n(z) - s_{\text{MP},y}(z).
\]
Now fix \(\varepsilon \in (0, 1/4)\) and define \(\tilde{\eta} := \eta^{-1/4 + \varepsilon/2}\). For any interval \(I \subset [a - 1, b + 1]\), where \(a\) and \(b\) are constants defined in (5), we choose a smoothed indicator function \(f = f_{\tilde{\eta}} \in C^\infty_c(\mathbb{R}; [0, 1])\) satisfying \(f(u) = 1\) for \(u \in I\), \(f(u) = 0\) for \(\text{dist}(E, I) \geq \tilde{\eta}\), \(\|f''\|_\infty \leq C\tilde{\eta}^{-1}\), and \(\|f''\|_\infty \leq C\tilde{\eta}^{-2}\). These imply that the supports of \(f'\) and \(f''\) have Lebesgue measure bounded by \(2\tilde{\eta}\). In addition, choose a smooth even cutoff function \(\chi \in C^\infty_c(\mathbb{R}; [0, 1])\) with \(\chi(v) = 1\) for \(|v| \leq 1\), \(\chi(v) = 0\) for \(|v| \geq 2\) and \(|\chi'|_\infty \leq 2\). Throughout this section, \(C\) represents a positive constant whose value may vary in each appearance.

Then by the Helffer-Sjöstrand formula, we get
\[
\int f(x)\hat{\mu}_n(dx) = \frac{1}{2\pi} \int \int (\partial_u + i\partial_v)[f(u) + ivf'(u)]\chi(v)]\hat{s}_n(u + iv)dvdu.
\]
As LHS is real, we can write as

\[
\int f(\lambda)\hat{\mu}_n(d\lambda) = -\frac{1}{2\pi} \int_{|v|\leq \eta} f''(u)\chi(v) v^3 \hat{s}_n(u + iv)dvdu
\]

(29)

\[
-\frac{1}{2\pi} \int_{|v|> \eta} f''(u)\chi(v) v^3 \hat{s}_n(u + iv)dvdu
\]

(30)

\[
\frac{1}{2\pi} \int \int (f(u) + ivf'(u))\chi^2(v)\hat{s}_n(u + iv)dvdu
\]

(31)

First, by the trivial identity \( \hat{s}_n(u - iv) = \hat{s}_n(u + iv) \) and the fact that \( \hat{s}_n \) is Lipschitz continuous on the compact set \( S_{\varepsilon/2} \), we can easily extend Theorem 5 as follows:

**Lemma 10:** For any fixed \( \varepsilon > 0 \), we have, with high probability,

\[
|\hat{s}_n(u + iv)| \leq n^{\varepsilon/2}(n^{-1/4} + n^{-1}|v|^{-7/2}),
\]

for all \( u, v \in \mathbb{R} \) such that \( \min\{|u-a|,|u-b|\} \leq 2\varepsilon^{-1} \) and \( |v| \in [\eta, 2\varepsilon^{-1}] \).

We may now estimate the terms appearing in (29)-(31). First, for the term in (31), by using the fact that \( \chi \) is even with support in \([-2, 2] \setminus (-1, 1) \), we have

\[
\left| \int \int (f(u) + ivf'(u))\chi^2(v)\hat{s}_n(u + iv)dvdu \right| \leq C\eta
\]

(32)

with high probability.

We next estimate the term in (29). Since \( v \) is small, we cannot apply Lemma 10 directly. However it can be proved that for all \( u \), the function \( v \mapsto v^3(s_{\varepsilon}(u + iv)) \) is nondecreasing for \( v > 0 \). This implies, for \( v \in (0, \eta) \),

\[
v^3\hat{s}_n(u + iv) \leq v^3s_{\varepsilon}(u + iv) \leq \hat{\eta}_\varepsilon^3s_{\varepsilon}(u + i\hat{\eta}_\varepsilon) \leq \hat{\eta}_\varepsilon^{\varepsilon/2}(n^{-1/4} + n^{-1}\eta^{-7/2}) + C\]

\[
\leq C\eta
\]

with high probability.

Hence we have

\[
\left| \int \int_{|v|\leq \eta} f''(u)\chi(v) v^3 (\hat{s}_n(u + iv))dvdu \right| \leq \hat{\eta}_\varepsilon^{-1} \int_{|v|\leq \eta} C\eta dv \leq C\eta.
\]

(33)

For the term in (30), we have

\[
\int \int_{|v|> \eta} f''(u)\chi(v) v^3 (\hat{s}_n(u + iv))dvdu
\]

\[
= \left[ \int_{|v|> \eta} f'(u)\chi(v) v (\hat{s}_n(u + iv))dv \right]_{u=\sup supp f''} - \left[ \int_{|v|> \eta} f'(u)\chi(v) v \frac{\partial}{\partial u} (\hat{s}_n(u + iv))dv \right]_{u=\inf supp f''}
\]

with high probability. Hence

\[
|\hat{\mu}_n(I)| \leq Cn^{-1/4+\varepsilon/2} \leq n^{-1/4+\varepsilon}
\]

with high probability. As \( \varepsilon \in (0, 1/4) \) is arbitrary, we conclude that \( \hat{\mu}_n(I) = O_{\varepsilon}(n^{-1/4}) \) for any \( I \subset [a - 1, b + 1] \).

Then for a general interval \( I \subset \mathbb{R} \), we first note that we have proved that \( \hat{\mu}_n([a, b]) = O_{-\varepsilon}(n^{-1/4}) \). As \( \mu_n \) is a

The first term is zero. As for the second, by Cauchy-Riemann equation, we have

\[
\int \int_{|v|> \eta} f''(u)\chi(v) v\Im(\hat{\mu}_n(u + iv))dvdu
\]

\[
= \left[ \int \supp f'' \int_{|v|> \eta} f'(u)\chi(v) v \frac{\partial}{\partial v} (\hat{s}_n(u + iv))dv \right]_\eta
\]

\[
= 2 \left[ \int \supp f'' \int_{|v|> \eta} f'(u)\chi(v) v (\hat{s}_n(u + iv))dv \right]_\eta
\]

\[
- 2 \int \supp f'' \int_{|v|> \eta} f'(u)(\chi(v) v + \chi(v)) (\hat{s}_n(u + iv))dvdu
\]

For the first term, we get

\[
\left| \left[ \int \supp f'' \int_{|v|> \eta} f'(u)\chi(v) v (\hat{s}_n(u + iv))dv \right]_\eta \right| \leq C\eta
\]

(34)

For the second term, we get

\[
\left| \int \supp f'' \int_{|v|> \eta} f'(u)(\chi(v) v + \chi(v)) (\hat{s}_n(u + iv))dvdu \right| \leq C\eta
\]

(35)

Putting all together, we get

\[
\left| \int f(\lambda)\hat{\mu}_n(d\lambda) \right| \leq C\eta
\]

with high probability.
probability measure and \( q_{\text{MP}}(a, b) = 1 \), we deduce that 
\[ \mu_n([a, b]) = O_n^{-1/4}. \]
Therefore we have 
\[ \mu_n(I) = \mu_n(I \cap [a, b]) + \mu_n(I \setminus [a, b]) \]
\[ = \mu_n(I \cap [a, b]) + O_n^{-1/4} \]
\[ = q_{\text{MP}}(I) + O_n^{-1/4} \]
where in the last step we use that \( \mu_n(I) = O_n^{-1/4} \) for \( I \subset [a, b] \). From the calculation it is easy to see that the above estimate holds simultaneously for all \( I \) (i.e. the constant absorbed by \( O_n \) is independent of \( I \)).

This completes the proof of Theorem 1.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for many invaluable suggestions which improve the presentation of the paper. V. Tarokh Vahid Tarokh would like to thank David Forney, Rob Calderbank, and Neil Sloane for stimulating discussions. M. Xiong would like to thank Zhigang Bao for comments and suggestions.

REFERENCES

[1] B. Babadi, S. S. Ghassemzadeh, and V. Tarokh, “Group randomness properties of pseudo-noise and gold sequences,” in Proc. 12th Can. Workshop Inf. Theory, Kelowna, BC, Canada, May 2011, pp. 42–46.
[2] B. Babadi and V. Tarokh, “Random frames from binary linear block codes,” in Proc. 44th Ann. Conf. Inf. Sci. Syst. (CISS), Princeton, NJ, USA, Mar. 2010, pp. 1–3.
[3] B. Babadi and V. Tarokh, “Spectral distribution of random matrices from binary linear block codes,” IEEE Trans. Inf. Theory, vol. 57, no. 6, pp. 3955–3962, Jun. 2011.
[4] Z. D. Bai, “Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices,” Ann. Probab., vol. 21, no. 2, pp. 625–648, Apr. 1993.
[5] Z. D. Bai, “Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices,” Ann. Probab., vol. 21, no. 2, pp. 649–672, Apr. 1993.
[6] Z. Bai and J. W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2010.
[7] A. Barg, A. Mazumdar, and R. Wang, “Restricted isometry property of random subdictionaries,” IEEE Trans. Inf. Theory, vol. 61, no. 8, pp. 4440–4450, Aug. 2015.
[8] F. Benaych-Georges and A. Knowles, “Local semicircle law for Wigner matrices,” in Advanced Topics in Random Matrices. Paris, France: Panoramas et Synthèses 53, Sociétè Mathématique de France, 2018, ch. 1, pp. 1–90.
[9] B. Alex, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, “Isotropic local laws for sample covariance and generalized wigner matrices,” Electron. J. Probab., vol. 19, no. 33, pp. 1–53, 2014.
[10] C. Blondeau and K. Nyberg, “Perfect nonlinear functions and cryptography,” Finite Fields Appl., vol. 32, pp. 120–147, Mar. 2015.
[11] R. Calderbank, S. Howard, and S. Jafarpour, “Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property,” IEEE J. Sel. Topics Signal Process., vol. 4, no. 2, pp. 358–374, Apr. 2010.

Chin Hei Chan received the Ph.D. degree in mathematics from The Hong Kong University of Science and Technology, Hong Kong, in 2019. He is currently a Post-Doctoral Fellow with The Hong Kong University of Science and Technology. His research interests include algebraic coding theory and number theory.

Vahid Tarokh received the Ph.D. degree in electrical engineering from the University of Waterloo, Ontario, Canada, in 1995. He worked with AT&T Labs-Research until 2000 and the Massachusetts Institute of Technology (MIT) as an Associate Professor. In 2002, he joined Harvard University as a Hammond Vinton Hayes Senior Fellow of electrical engineering and a Perkins Professor of applied mathematics. In January 2018, he joined Duke University as the Rhodes Family Professor of electrical and computer engineering, computer science, and mathematics, and a Bass Connections Endowed Professor. Since January 2019, he has been with Duke University as a Microsoft Data Science Investigator.

Maosheng Xiong received the Ph.D. degree in mathematics from the University of Illinois at Urbana–Champaign in 2007. He held a post-doctoral position at Pennsylvania State University from August 2007 to June 2010. In 2010, he joined the Department of Mathematics, The Hong Kong University of Science and Technology, where he is currently an Associate Professor. His area of research interests is in algebraic coding theory and number theory.