On a question of Babadi and Tarokh

Jing Xia[1] , Maosheng Xiong[2]

Abstract

In a recent remarkable paper [3], Babadi and Tarokh proved the "randomness" of sequences arising from binary linear block codes in the sense of spectral distribution, provided that their dual distances are sufficiently large. However, numerical experiments conducted by the authors revealed that Gold sequences which have dual distance 5 also satisfy such randomness property. Hence the interesting question was raised as to whether or not the stringent requirement of large dual distances can be relaxed in the theorem in order to explain the randomness of Gold sequences. This paper improves their result on several fronts and provides an affirmative answer to this question.

Index Terms

Asymptotic spectral distribution, coding theory, Marchenko-Pastur law, random matrix theory, randomness of sequences.

I. Introduction

The elegant theory of random matrices, and in particular properties of their spectral distribution, have been studied for a long time but remain a prominent and active research area due to its wide and important applications in many diverse disciplines such as mathematical statistics, theoretical physics, number theory, and more recently in economics [10] and communication theory [12]. Most of the random models considered so far are matrices whose entries have i.i.d. structures. In a remarkable paper, Babadi and Tarokh [3] considered matrices formed by choosing randomly codewords from some linear block codes with large dual distance and proved that these matrices behave like random matrices with i.i.d. entries, as long as the empirical spectral distribution is concerned. To describe their beautiful result, we need some notation.

Let $\mathcal{C}$ be an $[n, k, d]$ binary linear block code of length n, dimension k and minimum Hamming distance d over $\mathrm{GF}(2)$. The dual code of $\mathcal{C}$, denoted by $\mathcal{C}^\perp$, is an $[n, n-k, d^\perp]$ binary linear block

1. Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA, USA

2. Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

code over $\mathrm{GF}(2)$ such that all the codewords of $\mathcal{C}^{\perp}$ are orthogonal to those of $\mathcal{C}$ with the inner product defined over $\mathrm{GF}(2)^n$. Let $\epsilon : \mathrm{GF}(2)^n \to \{-1, 1\}^n$ be the component-wise mapping $\epsilon(v_i) := (-1)^{v_i}$, for $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathrm{GF}(2)^n$. For $p < n$, let $\mathbf{\Phi}_{\mathcal{C}}$ be a $p \times n$ random matrix whose rows are obtained by mapping a uniformly drawn set of size p of the codewords of $\mathcal{C}$ under ϵ. The *Gram matrix* of $\mathbf{\Phi}_{\mathcal{C}}$ is defined as $\mathcal{G}_{\mathcal{C}} := \mathbf{\Phi}_{\mathcal{C}}\mathbf{\Phi}_{\mathcal{C}}^T$, where $\mathbf{\Phi}_{\mathcal{C}}^T$ is the transpose of $\mathbf{\Phi}_{\mathcal{C}}$. Let $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the set of eigenvalues of an $n \times n$ matrix $\mathbf{A}$. The *spectral measure* of $\mathbf{A}$ is defined by

$$\mu_{\mathbf{A}} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

where δ_z is the Dirac measure. The empirical spectral distribution of $\mathbf{A}$ is defined as

$$M_{\mathbf{A}}(z) := \int_{-\infty}^{z} \mu_{\mathbf{A}}(\mathrm{d}z).$$

Babadi and Tarokh proved the following result ([3, Theorem 2.1]):

Consider a sequence of $[n, k_n, d_n]$ binary linear block codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$. Let $\mathbf{\Phi}_{\mathcal{C}_n}$ be a $p \times n$ random matrix based on $\mathcal{C}_n$, let $\mathcal{G}_{\mathcal{C}_n}$ denote the Gram matrix of the matrix $\frac{1}{\sqrt{n}}\mathbf{\Phi}_{\mathcal{C}_n}$, and let $M_{\mathcal{C}_n}(z)$ denote the empirical spectral distribution of $\mathcal{G}_{\mathcal{C}_n}$. Finally, let r_n be the greatest even integer less than or equal to $[(d_n^{\perp} - 1)/2]$, and let $r := \liminf_n r_n$. Then, as $n \to \infty$ with $y := p/n \in (0, 1)$ fixed, we have

$$\limsup_n |M_{\mathcal{C}_n}(z) - M_{\mathrm{MP}}(z)| \leq c(y, r)\left(r^{-1} + r^{-2}\right)$$

almost surely for all z, where $c(y, r)$ is a bounded function of r (which can be given explicitly), and $M_{\mathrm{MP}}(z)$ is the distribution corresponding to the Marchenko-Pastur measure μ_{MP} whose density is given by

$$\frac{\mathrm{d}\mu_{\mathrm{MP}}}{\mathrm{d}z} := \frac{1}{2\pi z y}\sqrt{(b - z)(z - a)}\, 1_{(a \leq z \leq b)} \ ,$$

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

It is well-known that as the dimensions grow to infinity, the empirical spectral distribution of the Gram matrix of real i.i.d. random matrices follows the Marchenko-Pastur law [8]. With this respect, the above result indicates that the matrix $\frac{1}{\sqrt{n}}\mathbf{\Phi}_{\mathcal{C}}$ based on the binary linear block code $\mathcal{C}$ is very close to random i.i.d. generated matrices as $n \to \infty$, if the dual distance of the code $\mathcal{C}$ is large enough. Numerical experiments conducted by the authors [3] on some low-rate BCH codes confirmed the significant similarity of the empirical distribution to the Marchenko-Pastur law for dimensions (and consequently, dual distances) as

small as $n = 63$.

However, there is an interesting phenomenon: the authors [2] also conducted some numerical experiments on Gold sequences and found convincing similarity of the empirical distributions to the Marchenko-Pastur law as well. This is a little surprising because Gold sequences arise from Gold codes [6] whose dual distances are always 5, which is relatively small. In a more recent interesting paper [4], investigating much further on the topic, the authors proved decisively the "randomness" of products of matrices arising from different binary linear block codes under large dual distances. At the end of the paper [4] Babadi and Tarokh also conducted numerical experiments and found numerical evidence of randomness on some Gold sequences. Hence they raised the natural question as to relaxing the stringent requirement of large dual distances in the results in order to explain the mysterious randomness of Gold sequences.

The purpose of this paper is to provide an affirmative answer to this questions. While binary linear block codes are most useful in practice, it is worthwhile to consider, at least in theory, linear block codes over a general finite field $\mathrm{GF}(q)$ where q is a prime power, especially when it does not require any substantial effort. For this purpose, denote by $\psi : \mathrm{GF}(q) \to \mathbb{C}^*$ the standard additive character given by

$$\psi(z) = \exp\left(\frac{2\pi\sqrt{-1}\,\mathrm{Tr}_{q/l}(z)}{l}\right),$$

here l is any prime number and q is a power of l, and $\mathrm{Tr}_{q/l}$ denotes the trace mapping from $\mathrm{GF}(q)$ to $\mathrm{GF}(l)$. When $q = l = 2$, then $\psi(z) = (-1)^z$ for $z \in \mathrm{GF}(2)$ which was considered before. It is known that $\psi(z)$ is a complex p-th root of unity.

Let $\mathcal{C}$ be an $[n, k, d]$ linear block code of length n, dimension k and minimum Hamming distance d over $\mathrm{GF}(q)$. The dual code of $\mathcal{C}$, denoted by $\mathcal{C}^\perp$, is an $[n, n-k, d^\perp]$ linear block code over $\mathrm{GF}(q)$ such that all the codewords of $\mathcal{C}^\perp$ are orthogonal to those of $\mathcal{C}$ with the natural inner product defined over $\mathrm{GF}(q)^n$. Let $\epsilon : \mathrm{GF}(q)^n \to (\mathbb{C}^*)^n$ be the component-wise mapping $\epsilon(v_i) := \psi(v_i)$, for $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathrm{GF}(q)^n$. For $p < n$, let $\mathbf{\Phi}_\mathcal{C}$ be a $p \times n$ random matrix whose rows are obtained by mapping a uniformly drawn set of size p of the codewords of $\mathcal{C}$ under ϵ. The *Gram matrix* of the $p \times n$ matrix $\mathbf{\Phi}_\mathcal{C}$ is defined as $\mathcal{G}_\mathcal{C} := \mathbf{\Phi}_\mathcal{C}\mathbf{\Phi}_\mathcal{C}^*$, where $\mathbf{\Phi}_\mathcal{C}^*$ is the conjugate transpose of $\mathbf{\Phi}_\mathcal{C}$. We prove

Theorem 1. *Let $\mathcal{C}$ be an $[n, k, d]$ linear block code over $\mathrm{GF}(q)$. Let $\mathbf{\Phi}_\mathcal{C}$ be a $p \times n$ random matrix based on $\mathcal{C}$, let $\mathcal{G}_\mathcal{C}$ denote the Gram matrix of $\frac{1}{\sqrt{n}}\mathbf{\Phi}_\mathcal{C}$, and let $M_\mathcal{C}(z)$ denote the empirical spectral distribution*

of $\mathcal{G}_{\mathcal{C}}$. Suppose n is sufficiently large. Then if $d^{\perp} \geq 5$ and for any $y := p/n \in (0,1)$, we have

$$\sup_{z\in\mathbb{R}} |M_{\mathcal{C}}(z) - M_{\mathrm{MP}}(z)| \leq \frac{800}{\sqrt{y}(1-y)} \frac{\log\log n}{\log n} \ . \tag{1}$$

A. Discussion of the Main Theorem

Theorem 1 might look a little surprising, compared with the celebrated result by Sidel'nikov [11]: for any $[n,k,d]$ binary linear block code $\mathcal{C}$ with $d^{\perp} \geq 3$, we have

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

as $n \to \infty$, where $A(z)$ is the cumulative weight distribution function of the code $\mathcal{C}$ and

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

Hence the "randomness" of the weight distribution of $\mathcal{C}$ is ensured if $d^{\perp}$ is sufficiently large. In Theorem 1, however, we only require $d^{\perp} \geq 5$.

Gold codes have three distinct non-zero weights which are known [6]. By applying the MacWilliams identity [7] and by using **Mathematica**, it can be readily verified that the dual distance of Gold codes is always 5, hence Theorem 1 is applicable and confirms that Gold sequences behave like random i.i.d. sequences, in the sense of the spectral distribution.

The condition $d^{\perp} \geq 5$ in Theorem 1 can be slightly improved by assuming that the number of weight 4 codewords in $C^{\perp}$ is relatively small (see Theorem 2 in Section II), and the inequality (1) of same kind still holds true, if 800 replaced by a larger constant on the right hand side of (1). On the other hand, however, if $d^{\perp} = 3$, then Theorem 1 may not be true: Babadi, Ghassemzadeh and Tarokh ([2, Theorem 3.1]) proved that shortened first-order Reed-Muller (Simplex) codes which have dual distance 3 have substantially different behavior in the sense of the spectral distribution.

The proof of Theorem 1 follows essentially the strategy used by Babadi and Tarokh in [3], but here in the paper some essence of number theory plays more prominent roles in the study. This might become more apparent in Section II when we study the l-moment of the spectral measure. We shall prove Theorem 2, which improves [3, Lemma 3.3] substantially. Equipped with Theorem 2, in Section III we will prove Theorem 1 directly. In the proof of Theorem 2, however, some very complicated issues of combinatorial nature arise which need to be taken care of. To streamline the ideas of the paper, we treat those issues in Section IV.

II. Estimate of the l-th moment

In this section we study the l-th moment of the spectral distribution, similar to [3, Lemma 3.3]. We use slightly different notation, which might be more suited for the problem.

As in Introduction, let $\mathcal{C}$ be an $[n,k,d]$ linear block code over $\mathrm{GF}(q)$, and let $\epsilon : \mathrm{GF}(q)^n \to (\mathbb{C}^*)^n$ be the component-wise mapping. Define $\mathcal{D} = \epsilon(\mathcal{C})$. Let $N := q^k$ be the cardinality of $\mathcal{D}$ (and $\mathcal{C}$). Let $p < n$. In order to choose randomly p elements from $\mathcal{D}$, we define Ω_p to be the set of all maps $s : [1,p] \to \mathcal{D}$ endowed with the uniform probability, here $[1,p]$ denotes the set of integers from 1 to p. Hence Ω_p is a probability space with cardinality $|\Omega_p| = N^p$. For each $s \in \Omega_p$, the $p \times n$ matrix $\mathbf{\Phi}_s$ corresponding to s is given by

$$\mathbf{\Phi}_s^T = \left[s(1)^T, s(2)^T, \ldots, s(p)^T \right]_{n \times p} \ ,$$

here we have written $s(i) \in \mathcal{D}$ as $1 \times n$-row vectors. For any $\mathbf{u} = (u_1, \ldots, u_n), \mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n$, the (Hermitian) inner product is

$$\langle \mathbf{u}, \mathbf{v} \rangle := u_1 \bar{v}_1 + \cdots + u_n \bar{v}_n.$$

Let $\mathcal{G}(s)$ be the Gram matrix of $\frac{1}{\sqrt{n}} \mathbf{\Phi}_s$. This is a $p \times p$ Hermitian matrix with the (ij)-th entry given by $\langle s(i), s(j) \rangle / n$. Let $\lambda_1(s), \lambda_2(s), \ldots, \lambda_p(s) \in \mathbb{R}$ be the eigenvalues of $\mathcal{G}(s)$. For any positive integer l, define

$$A_l(s) := \frac{1}{p} \sum_{i=1}^{p} \lambda_i(s)^l = \frac{1}{p} \ \mathrm{Tr} \left(\mathcal{G}(s)^l \right).$$

The purpose of this section is to compute $\mathbb{E}\left(A_l(s), \Omega_p\right)$, the l-th moment of the spectral measure. We prove a general result:

Theorem 2. *Let $y := p/n \in (0,1)$. Let A be the number of weight 4 codewords in $\mathcal{C}^\perp$. Then for any $2 \le l < \sqrt{p}$, we have*

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \sum_{i=0}^{l-1} \frac{y^i}{i+1} \binom{l}{i} \binom{l-1}{i} + E_l, \tag{2}$$

where E_l is bounded by

$$|E_l| \le \left(4 + 2\sqrt{\frac{2A}{q-1} + \frac{1}{4}} \right) \frac{l^{l+1}}{n} \ ,$$

The rest of this section is devoted to a proof of Theorem 2.

A. Problem setting up

We say that $\gamma : [0, l] \to [1, p]$ is a closed path if γ is a map with $\gamma(0) = \gamma(l)$. Denote by $\Pi_{l,p}$ the set of all closed paths from $[0, l]$ to $[1, p]$. For each $\gamma \in \Pi_{l,p}$ and $s \in \Omega_p$, define

$$\omega_\gamma(s) := \langle s \circ \gamma(0), s \circ \gamma(1)\rangle\langle s \circ \gamma(1), s \circ \gamma(2)\rangle \cdots \langle s \circ \gamma(l-1), s \circ \gamma(l)\rangle.$$

Expanding $\mathrm{Tr}\left(\mathcal{G}(s)^l\right)$, it is easy to see that

$$A_l(s) = \frac{1}{pn^l} \sum_{\gamma \in \Pi_{l,p}} \omega_\gamma(s).$$

Hence

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \frac{1}{pn^l} \sum_{\gamma \in \Pi_{l,p}} \mathbb{E}\left(\omega_\gamma(s), \Omega_p\right).$$

Let Σ_p be the group of permutations of the set $[1, p]$. Then Σ_p acts on $\Pi_{k,p}$, since $\sigma \circ \gamma \in \Pi_{l,p}$ whenever $\gamma \in \Pi_{l,p}$ and $\sigma \in \Sigma_p$. Let $[\gamma]$ be the equivalent class of γ, that is,

$$[\gamma] = \{\sigma \circ \gamma : \sigma \in \Sigma_p\}.$$

We may write

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \frac{1}{pn^l} \sum_{\gamma \in \Pi_{l,p}/\Sigma_p} \sum_{\tau \in [\gamma]} \mathbb{E}\left(\omega_\tau(s), \Omega_p\right).$$

For any fixed $\sigma \in \Sigma_p$, as s runs over Ω_p, clearly $s \circ \sigma$ also runs over Ω_p, hence

$$\mathbb{E}\left(\omega_{\sigma\circ\gamma}(s), \Omega_p\right) = \mathbb{E}\left(\omega_\gamma(s \circ \sigma), \Omega_p\right) = \mathbb{E}\left(\omega_\gamma(s), \Omega_p\right).$$

Moreover, let

$$V_\gamma := \gamma\left([0, l]\right) \subset [1, p], \quad v_\gamma := \#V_\gamma,$$

and define the probability space

$$\Omega(V_\gamma) := \{s : V_\gamma \to \mathcal{D}\}$$

assigned with the uniform probability. It is clear that $\#[\gamma] = \frac{p!}{(p-v_\gamma)!}, \#\Omega(V_\gamma) = N^{v_\gamma}$ and

$$\mathbb{E}\left(\omega_\gamma(s), \Omega_p\right) = \mathbb{E}\left(\omega_\gamma(s), \Omega(V_\gamma)\right).$$

Summarizing the above considerations, we have

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \frac{1}{pn^l} \sum_{\gamma \in \Pi_{l,p}/\Sigma_p} \frac{p!}{(p-v_\gamma)!}\, \mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right). \tag{3}$$

B. Study of $\mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right)$

Up to this point everything is essentially the same as in the proof of [3, Lemma 3.3]. The main innovation of the paper is to use number theory to treat $\mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right)$ in a more careful way.

Let $H = (h_{ij})_{n\times k}$ be a generating matrix of $\mathcal{C}$, that is, each codeword of $\mathcal{C}$ is given by

$$c(\mathbf{x}) := H[x_1, \ldots, x_k]^T, \tag{4}$$

for some unique $\mathbf{x} = (x_1, \ldots, x_k) \in \mathrm{GF}(q)^k$. Hence each $s(i) \in \mathcal{D}$ corresponds to a unique vector, which we may record as $(s(i)_1, \ldots, s(i)_k) \in \mathrm{GF}(q)^k$. From (4), the t-th entry of $s(i)$ is given by

$$s(i)[t] = \psi\left(\sum_{j=1}^{k} h_{tj} s(i)_j\right),$$

where $\psi : \mathrm{GF}(q) \to \mathbb{C}^*$ is the standard additive character. So

$$\langle s\circ\gamma(u), s\circ\gamma(u+1)\rangle = \sum_{t=1}^{n} \psi\left(\sum_{j=1}^{k} h_{tj} s\circ\gamma(u)_j - \sum_{j=1}^{k} h_{tj} s\circ\gamma(u+1)_j\right),$$

and hence

$$\begin{aligned}\omega_\gamma(s) \;=\; & \sum_{1\le t_0,t_1,\ldots,t_{l-1}\le n} \psi\left(\sum_{j=1}^{k} h_{t_0 j}\left\{s\circ\gamma(0)_j - s\circ\gamma(1)_j\right\}\right) \\ & \times\psi\left(\sum_{j=1}^{k} h_{t_1 j}\left\{s\circ\gamma(1)_j - s\circ\gamma(2)_j\right\}\right)\cdots\psi\left(\sum_{j=1}^{k} h_{t_{l-1} j}\left\{s\circ\gamma(l-1)_j - s\circ\gamma(0)_j\right\}\right).\end{aligned}$$

Now suppose

$$V_\gamma = \{z_a : 1 \le a \le v_\gamma\} \subset [1, p],$$

and for each a, let $I_a := \gamma^{-1}(z_a)$. For each $u \in I_a$, we have $\gamma(u) = z_a$ and clearly $[0, l-1] = \bigcup_a I_a$ is a partition. We may collect the term $s(z_a)$ together on the right hand side of $\omega_\gamma(s)$ above and rewrite it as

$$\omega_\gamma(s) = \sum_{1\le t_0,t_1,\ldots,t_{l-1}\le n} \prod_{a=1}^{v_\gamma}\prod_{j=1}^{k} \psi\left(s(z_a)_j \sum_{u\in I_a}\{h_{t_u j} - h_{t_{u-1} j}\}\right).$$

Here when $u = 0$, we interpret $t_{0-1} := t_{l-1}$ (we will use this convent multiple times in the paper).

Therefore

$$\mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right) = \frac{1}{N^{v_\gamma}} \sum_{\substack{s(z_a)_j \in \mathrm{GF}(q) \\ 1 \le a \le v_\gamma \\ 1 \le j \le k}} \omega_\gamma(s).$$

The orthogonality property

$$\sum_{z \in \mathrm{GF}(q)} \psi(zx) = \begin{cases} 0 & : \quad \text{if } x \in \mathrm{GF}(q) \setminus \{0\}; \\ q & : \quad \text{if } x = 0, \end{cases}$$

implies that if for some a and for some j we have

$$\sum_{u \in I_a} \left(h_{t_u j} - h_{t_{u-1} j}\right) \neq 0\,,$$

then their contribution to $\mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right)$ is zero. So we conclude that the quantity $\mathbb{E}\left(\omega_\gamma(s), \Omega\left(V_\gamma\right)\right)$ is the same as W_γ, which is the number of solutions $(t_0, t_1, \ldots, t_{l-1})$ such that $1 \le t_0, t_1, \ldots, t_{l-1} \le n$ and

$$\sum_{u \in I_a} \left(\mathbf{h}_{t_u} - \mathbf{h}_{t_{u-1}}\right) = \mathbf{0}, \quad \forall\ 1 \le a \le v_\gamma,$$

here $\mathbf{h}_i$ denotes the i-th row of the matrix H, and

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \frac{1}{pn^l} \sum_{\gamma \in \Pi_{l,p}/\Sigma_p} \frac{p!}{(p - v_\gamma)!}\, W_\gamma. \tag{5}$$

C. Proof of Theorem 2

The combinatorial nature of solving W_γ, while elementary, presents some technical challenge. To streamline the idea of the proof, and for the sake of clarity, we leave the analysis of W_γ to Section IV. Here instead we quote the main results to continue our proof.

In Section IV we prove that there is a subset $\Gamma \subset \Pi_{l,p}/\Sigma_p$ with the following property:

$$\begin{cases} W_\gamma = n^{l - v_\gamma + 1} & : \quad \text{if } \gamma \in \Gamma; \\ 0 \le W_\gamma \le C_A\, n^{l - v_\gamma} & : \quad \text{if } \gamma \notin \Gamma, \end{cases}$$

where C_A is given in (19). Using this we find that

$$\mathbb{E}\left(A_l(s), \Omega_p\right) = \frac{n}{p} \sum_{\substack{\gamma \in \Pi_{l,p}/\Sigma_p \\ \gamma \in \Gamma}} \frac{p!}{(p - v_\gamma)!\, n^{v_\gamma}} + E_1, \tag{6}$$

where E_1 is bounded by

$$|E_1| \leq \frac{C_A}{p} \sum_{\gamma \in \Pi_{l,p}/\Sigma_p} \frac{p!}{(p-v_\gamma)!\, n^{v_\gamma}} \leq \frac{C_A}{p} \sum_{v=1}^{l} \left(\frac{p}{n}\right)^v \sum_{\substack{\gamma \in \Pi_{l,p}/\Sigma_p \\ v_\gamma = v}} 1 \ .$$

It is easy to see that

$$\sum_{\substack{\gamma \in \Pi_{l,p}/\Sigma_p \\ v_\gamma = v}} 1 < v^l \leq l^l,$$

and hence

$$|E_1| \leq C_A\, l^{l+1}/n.$$

On the other hand, it is also proved in Section IV that

$$\sum_{\substack{\gamma \in \Gamma \subset \Pi_{l,p}/\Sigma_p \\ v_\gamma = v}} 1 = \frac{1}{v} \binom{l}{v-1} \binom{l-1}{v-1}.$$

Suppose $2 \leq l < \sqrt{p}$. For $v \geq 2$, using

$$p^v \geq \frac{p!}{(p-v)!} > p^v \left(1 - v/p\right)^{v-1} \geq p^v \left(1 - v(v-1)/p\right),$$

in (6), we can finally obtain, after some simplifying, the desired result (2). This completes the proof of Theorem 2. □

III. Proof of Theorem 1

Given Theorem 2, the proof of Theorem 1 follows essentially arguments in [3], though some of our analysis is more precise.

A. Some lemmas

Fix $y \in (0,1)$, let $\mathbf{x}$ be a Marchenko-Pastur random variable whose density function is given by

$$\frac{\mathrm{d}\mu_{\mathrm{MP}}}{\mathrm{d}z} := \frac{1}{2\pi z y} \sqrt{(b-z)(z-a)}\, 1_{(a \leq z \leq b)} \ ,$$

here $a = (1-\sqrt{y})^2$ *and* $b = (1+\sqrt{y})^2$. It is known that the l-th moment of $\mathbf{x}$ is given by

$$m_{\mathrm{MP}}^{(l)} = \mathbb{E}(\mathbf{x}^l) = \sum_{i=0}^{l-1} \frac{y^i}{i+1} \binom{l}{i} \binom{l-1}{i}. \tag{7}$$

Define

$$b_{\mathrm{MP}}^{(l)} := \mathbb{E}\left((\mathbf{x}-1)^l\right).$$

Clearly $b_{\mathrm{MP}}^{(0)} = 1, b_{\mathrm{MP}}^{(1)} = 0$. We first prove

Lemma 3. *For any $l \geq 2$ we have*

$$\left|b_{\mathrm{MP}}^{(l)}\right| < \frac{l^3(8e^2)^l y}{8\pi}. \tag{8}$$

Proof. Expanding $\mathbb{E}\left((\mathbf{x}-1)^l\right)$ and using (7) we have

$$b_{\mathrm{MP}}^{(l)} = \sum_{i=1}^{l-1} \frac{y^i}{i+1} \sum_{t=i+1}^{l} (-1)^{l-1} \binom{l}{t}\binom{t}{i}\binom{t-1}{i}.$$

Elementary estimates on binomial coefficients yield

$$\left|b_{\mathrm{MP}}^{(l)}\right| < \frac{2^l}{2} \sum_{i=1}^{l-1} \frac{y^i l^{2i}}{(i!)^2} < 2^{l-1}(yl^2) \sum_{i=0}^{l-1} \frac{(yl^2)^i}{(i!)^2} \leq 2^{l-1}(yl^3) \max_{0\leq i\leq l-1} \frac{(yl^2)^i}{(i!)^2}.$$

By quotient test we find that the maximal value is attained at $i_0 = [\sqrt{y}l]$. If $i_0 = 0$ or 1, then the equality (8) can be easily verified. Now suppose $i_0 \geq 2$. Then $i_0 > \sqrt{y}r - 1 \geq \sqrt{y}r/2$. Using the Stirling's bound on $n!$, given by

$$n! \geq \sqrt{2\pi n}(n/e)^n, \tag{9}$$

we obtain

$$\left|b_{\mathrm{MP}}^{(l)}\right| < 2^{l-1}(yl^3)\frac{(yl^2)^{i_0}}{4\pi\left(\sqrt{y}l/2e\right)^{2i_0}} = \frac{l^3 2^l}{8\pi}(4e^2)^{i_0} y \leq \frac{l^3 2^l}{8\pi}(4e^2)^l y.$$

This completes the proof of Lemma 3. □

To prove Theorem 1, following the method of [3], we need a lemma from probability theory, which is discussed in details in [5, Ch. XVI-3] (or see [3, Lemma 3.1]):

Lemma 4. *Let F be a probability distribution with vanishing expectation and characteristic function ϕ. Suppose that $F - G$ vanishes at $\pm\infty$ and that G has a derivative g such that $|g| \leq m$. Finally, suppose that g has a continuously differentiable Fourier transform γ such that $\gamma(0) = 1$ and $\gamma'(0) = 0$. Then, for all z and $T > 0$ we have*

$$|F(z) - G(z)| \leq \frac{1}{\pi}\int_{-T}^{T} \left|\frac{\phi(t) - \gamma(t)}{t}\right| \mathrm{d}t + \frac{24m}{\pi T}.$$

B. Proof of Theorem 1

Using notation from Section II, for each $s \in \Omega_p$, let $\lambda_1(s), \ldots, \lambda_p(s)$ be the eigenvalues of $\mathcal{G}(s)$. The characteristic function we consider is

$$\phi_{\mathcal{C}}(t) := \frac{1}{p} \sum_{k=1}^{p} \mathbb{E}\left(\exp\left(it(\lambda_k(s) - 1)\right), \Omega_p\right).$$

For the Marchenko-Pastur random variable $\mathbf{x}$ we consider

$$\gamma(t) := \mathbb{E}\left(\exp(it(\mathbf{x} - 1))\right).$$

Define for each l

$$B_l = \frac{1}{p} \sum_{k=1}^{p} \mathbb{E}\left((\lambda_k(s) - 1)^l, \Omega_p\right).$$

Expanding the l-th power we find that

$$B_l = \sum_{t=0}^{l} (-1)^{l-t} \binom{l}{t} \mathbb{E}\left(A_t(s), \Omega_p\right), \tag{10}$$

where estimates on $\mathbb{E}\left(A_t(s), \Omega_p\right)$ is provided by Theorem 2. Using the inequality

$$\left| \exp(it) - \sum_{l=0}^{r-1} \frac{(it)^l}{l!} \right| \leq \frac{|t|^r}{r!},$$

and choosing $r \geq 4$ to be even, we find that

$$\left| \phi_{\mathcal{C}}(t) - \sum_{l=0}^{r-1} \frac{(it)^l B_l}{l!} \right| \leq \frac{t^r B_r}{r!}, \tag{11}$$

and

$$\left| \gamma(t) - \sum_{l=0}^{r-1} \frac{(it)^l b_{\mathrm{MP}}^{(l)}}{l!} \right| \leq \frac{t^r\, b_{\mathrm{MP}}^{(r)}}{r!}. \tag{12}$$

We note that $B_l = b_{\mathrm{MP}}^{(l)}$ for $l = 0, 1$. For $l \geq 2$, using the expression (10) and Theorem 2, given that $d^{\perp} \geq 5$, we find

$$\left| B_l - b_{\mathrm{MP}}^{(l)} \right| \leq \sum_{t=2}^{l} \binom{l}{t} \frac{5\, t^{t+1}}{n} < \frac{15\, l^{l+1}}{n}. \tag{13}$$

In writing

$$|\phi_{\mathcal{C}}(t)-\gamma(t)| \le \left|\phi_{\mathcal{C}}(t)-\sum_{l=0}^{r-1}\frac{(it)^l B_l}{l!}\right| + \left|\gamma(t)-\sum_{l=0}^{r-1}\frac{(it)^l b_{\mathrm{MP}}^{(l)}}{l!}\right| + \left|\sum_{l=0}^{r-1}\frac{(it)^l\left(B_l-b_{\mathrm{MP}}^{(l)}\right)}{l!}\right|,$$

applying Lemma 4 and using the above estimates from (11)(12)(13) and Lemma 3, we collect terms together and finally obtain

$$|M_{\mathcal{C}}(z+1)-M_{\mathrm{MP}}(z+1)| \le \frac{r^2(8e^2T)^r y}{2\pi^2(r!)} + \frac{60\,r(Tr)^r}{\pi n(r!)} + \frac{24}{\pi^2\sqrt{y}(1-y)T}. \tag{14}$$

Finally, taking r to be a positive even integer of size

$$r \approx \frac{\log n}{\log\log n}, \text{ and } T=\frac{r}{16e^3},$$

and using the Stirling's bound (9), when n (and consequently r) is sufficiently large, it is easy to see that the first two terms on the right side of (14) can be both bounded by $\frac{\log\log n}{\log n}$, while the third term is

$$\frac{24\cdot 16\cdot e^3}{\pi^2\sqrt{y}(1-y)r} \approx \frac{782\cdot\log\log n}{\sqrt{y}(1-y)\cdot\log n}.$$

Combining these terms completes the proof of Theorem 1. □

IV. The analysis of W_γ

Let $\gamma : [0, l_\gamma] \to [1, p]$ be a closed path with $V_\gamma = \gamma([0, l_\gamma]) = \{z_a : 1 \le a \le v_\gamma\}, v_\gamma = |V_\gamma|$ and $I_a = \gamma^{-1}(z_a)$. Denote by W_γ the number of solutions $(t_0, t_1, \ldots, t_{l_\gamma-1})$ such that $1 \le t_0, t_1, \ldots, t_{l_\gamma-1} \le n$ and

$$\sum_{u\in I_a}\left(\mathbf{h}_{t_u}-\mathbf{h}_{t_{u-1}}\right)=\mathbf{0}, \quad \forall\ 1\le a\le v_\gamma,$$

here $\mathbf{h}_i$ denotes the i-th row of the matrix H, whose rows are all distinct by assumption, and the indices shall be considered modulo l_γ, i.e., $t_{-1} = t_{l_\gamma-1}$. The purpose of this section is to study W_γ, which is crucial in the proof of Theorem 2.

Definition 5. The closed path γ is called "reduced" if $v_\gamma = l_\gamma = 1$, or if $v_\gamma \ge 2$ and the following two conditions are satisfied:

(i). each $|I_a| \ge 2$, hence $l = \sum_a |I_a| \ge 2v \ge 4$;

(ii). each I_a does not contain consecutive indices, that is, $\gamma(u) \ne \gamma(u+1), \forall u$.

We first study W_γ when γ is reduced.

A. Study of W_γ for γ reduced

Let γ be a reduced closed path with $l = l_\gamma \geq 1$ and $v = v_\gamma \geq 1$. If $v_\gamma = l_\gamma = 1$, then trivially we have

$$W_\gamma = n.$$

Now suppose that $v_\gamma \geq 2$. For each I_a, define $I'_a := I_a - \{1\} = \{u - 1 \pmod{l_\gamma - 1} : u \in I_a\}$. For any $1 \leq a \leq v_\gamma$, the equation corresponding to I_a is

$$\sum_{u \in I_a} \mathbf{h}_{t_u} - \sum_{u \in I'_a} \mathbf{h}_{t_u} = \mathbf{0}. \tag{15}$$

We shall write down the equations (15) for $1 \leq a \leq v_\gamma$ as a matrix with respect to the variables $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}, \ldots, \mathbf{h}_{t_{l-1}}$, given in the same ordered.

Since $\cup_a I_a$ is a partition of $[0, l-1]$, and each I_a does not contain consecutive elements, there are distinct indices, which we may say 1 and v, such that $0 \in I_1$ and $1 \in I_v$. Hence $k - 1 \in I'_0$, and the row vector corresponding to the equation of I_1 with respect to $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}, \ldots, \mathbf{h}_{t_{l-1}}$ is of shape

$$[1, *, \cdots, *, -1].$$

Now let u_2 be the smallest index in the set $\cup_{2 \leq a \leq v-1} (I_a \cup I'_a)$. We must have $u_2 \geq 1$, and $u_2 \in I'_a$ for some $2 \leq a \leq v - 1$, because if otherwise, then $u_2 = 0$, which contradicts the fact that $0 \in I_1$ and $1 \in I_v$. We may reorder the indices and say $u_2 \in I'_2$. Hence $u_2 + 1 \in I_2$, and the row vector corresponding to the equation of I_2 with respect to $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}, \ldots, \mathbf{h}_{t_{l-1}}$ is of shape

$$[0 \cdots 0, -1, 1, *, \cdots, *, 0],$$

where the first non-zero entry "-1" appears at the u_2-th column.

Now let u_3 be the smallest index in the set $\cup_{3 \leq a \leq v-1} (I_a \cup I'_a)$. Similarly we must have $u_3 \geq u_2 + 1$, and $u_3 \in I'_a$ for some $3 \leq a \leq v - 1$. We reorder the indices and say $u_3 \in I'_3$. Then $u_3 + 1 \in I_3$, and the row vector corresponding to the equation of I_3 with respect to $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}, \ldots, \mathbf{h}_{t_{l-1}}$ is of shape

$$[0 \cdots 0, 0 \cdots 0, -1, 1, *, \cdots, *, 0],$$

where the first non-zero entry "-1" appears at the u_3-th column.

We can continue this process up to $a = v-1$ because each row contains at least two non-zero entries. Clearly the row vectors corresponding to the equations I_a for $1 \leq a \leq v-1$ form an upper triangular matrix with rank $v-1$. So the number of free variables is $l-v+1$. This proves that $W_\gamma \leq n^{l-v+1}$. Actually we shall do much better.

Since $l \geq 2v$, and each row vector corresponding to $I_a, 1 \leq a \leq v-1$ with respect to $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}, \ldots, \mathbf{h}_{t_{l-1}}$ contains at least two 1's, we may find $l-v$ free variables, say they are $t_v, \ldots, t_{l-1}$ after reordering the indices, so that for any given values of $t_v, \ldots, t_{l-1}$ from 1 to n, solving the equations (15) becomes looking for $1 \leq t_0, \ldots, t_{v-1} \leq n$ such that

$$\begin{aligned} \mathbf{h}_{t_i} &= \mathbf{v}_i, \quad \forall\, 2 \leq i \leq v-1, \\ \mathbf{h}_{t_0} + \mathbf{h}_{t_1} &= \mathbf{v}_1, \end{aligned}$$

where the vectors $\mathbf{v}_i$ are linear combinations of the rows of H, depending only on $t_v, \ldots, t_{l-1}$. Clearly the number of solutions for $t_i, 2 \leq i \leq v-1$ is at most one. One only needs to consider t_0, t_1.

If $\mathbf{v}_1 = \mathbf{0}$, this enforces a new relation on $t_v, \ldots, t_{l-1}$ which were free before, hence the number of such $(t_v, \ldots, t_{l-1})$'s with $\mathbf{v}_1 = \mathbf{0}$ is at most n^{l-v-1}. On the other hand, for each given t_0, there is at most one value t_1 such that $\mathbf{h}_{t_0} + \mathbf{h}_{t_1} = \mathbf{0}$. Hence the total number of solutions of t_i's for this case is at most n^{l-v}. Let us define

$$A_{\mathbf{v}} = |\left\{(t_0, t_1) : 1 \leq t_0, t_1 \leq n, \text{ and } \mathbf{h}_{t_0} + \mathbf{h}_{t_1} = \mathbf{v}\right\}|.$$

We have just proved that

$$W_\gamma \leq n^{l-v} \left(1 + \sup_{\mathbf{v} \neq \mathbf{0}} A_{\mathbf{v}}\right). \tag{16}$$

Now for a fixed $\mathbf{v} \neq \mathbf{0}$, note that if $t_0 = t_1$, the equation $2\mathbf{h}_t = \mathbf{v}$ has at most one solution for $1 \leq t \leq n$. So we have

$$A_{\mathbf{v}} \leq 1 + 2B_{\mathbf{v}}, \tag{17}$$

where $B_{\mathbf{v}}$ is the cardinality of the set

$$\mathcal{B}_{\mathbf{v}} = \left\{(t_0, t_1) : 1 \leq t_0 < t_1 \leq n, \text{ and } \mathbf{h}_{t_0} + \mathbf{h}_{t_1} = \mathbf{v}\right\}.$$

If $B_{\mathbf{v}} \geq 2$, then for any distinct elements $(t_0, t_1), (t_0', t_1') \in \mathcal{B}_{\mathbf{v}}$, we conclude that t_0, t_1, t_0', t_1' are all distinct

and

$$\mathbf{h}_{t_0} + \mathbf{h}_{t_1} - \mathbf{h}_{t'_0} - \mathbf{h}_{t'_1} = \mathbf{0}.$$

This gives a weight 4 codeword in $\mathcal{C}^{\perp}$ with entries $1, 1, -1, -1$ at the t_0, t_1, t'_0 and t'_1-th places respectively. From it we may multiply elements of $\mathrm{GF}(q) - \{0\}$ to get new weight 4 codewords. Now suppose that A is the number of weight 4 codewords of $\mathcal{C}^{\perp}$. The above argument shows that

$$A \geq (q-1)\binom{B_{\mathbf{v}}}{2} = \frac{q-1}{2} B_{\mathbf{v}} \left(B_{\mathbf{v}} - 1\right).$$

Hence we have

$$B_{\mathbf{v}} \leq \sqrt{\frac{2A}{q-1} + \frac{1}{4}} + \frac{1}{2}\,.$$

In relation to (17) and (16) we conclude that if $v_\gamma \geq 2$,

$$W_\gamma \leq C_A\, n^{l_\gamma - v_\gamma}, \tag{18}$$

where

$$C_A = 3 + 2\sqrt{\frac{2A}{q-1} + \frac{1}{4}}\,. \tag{19}$$

B. An example

To illuminate the combinatorial nature of solving W_γ in general, it may be useful to consider an example first.

Let $l_\gamma = 9$, and γ define the partition

$$\{0, 1, \ldots, 8\} = \{0, 1, 2, 7\} \cup \{3, 5, 8\} \cup \{4\} \cup \{6\}.$$

So $v_\gamma = 4$. Then W_γ is the number of solutions $(t_0, t_1, \ldots, t_8)$ such that $1 \leq t_0, t_1, \ldots, t_8 \leq n$ and the following four equations hold simultaneously:

$$\mathbf{h}_{t_0} + \mathbf{h}_{t_1} + \mathbf{h}_{t_2} + \mathbf{h}_{t_7} = \mathbf{h}_{t_8} + \mathbf{h}_{t_0} + \mathbf{h}_{t_1} + \mathbf{h}_{t_6} \tag{20}$$

$$\mathbf{h}_{t_3} + \mathbf{h}_{t_5} + \mathbf{h}_{t_8} = \mathbf{h}_{t_2} + \mathbf{h}_{t_4} + \mathbf{h}_{t_7} \tag{21}$$

$$\mathbf{h}_{t_4} = \mathbf{h}_{t_3} \tag{22}$$

$$\mathbf{h}_{t_6} = \mathbf{h}_{t_5} \tag{23}$$

Clearly one equation is redundant: we can always remove one and keep the rest.

Consider (20), we find that $\mathbf{h}_{t_0}, \mathbf{h}_{t_1}$ can be canceled out on both sides. Hence t_0 and t_1 are free and can be removed, and (20) becomes

$$\mathbf{h}_{t_2} + \mathbf{h}_{t_7} \quad = \quad \mathbf{h}_{t_8} + \mathbf{h}_{t_6} \tag{24}$$

Consider (22). Since the rows of H are all distinct, this implies that $t_3 = t_4$, and under this restriction, $\mathbf{h}_{t_3}$ and $\mathbf{h}_{t_4}$ are also canceled out on both sides of (21). Then $t_3 = t_4$ is also a free variable and can be removed.

Consider (23). Clearly we have $t_5 = t_6$, but this is not a free variable: replacing t_5 by t_6, we find that $W_\gamma = n^3 W_{\gamma'}$, where $W_{\gamma'}$ is the number of solutions (t_2, t_6, t_7, t_8) such that $1 \le t_2, t_6, t_7, t_8 \le n$ and the equation (24) is satisfied.

The γ' can be reinterpreted as a closed path. It is a reduced path with $l_{\gamma'} = 4, v_{\gamma'} = 2$, hence the quantity $W_{\gamma'}$ can be estimated by (18), so we conclude that

$$W_\gamma \le n^3 \, C_A \, n^{l_{\gamma'} - v_{\gamma'}} = C_A \, n^5.$$

C. Study of W_γ in general

As illustrated by the previous example, we shall isolate variables from the equations related to W_γ, and removing these variables would result in a new but simpler closed path γ', and three different situations may arise and need to be examined carefully.

We use some notation. For a closed path $\gamma : [0, l_\gamma] \to [1, p]$, the terms V_γ, v_γ and I_a's are as before. γ yields a loop $t_0, t_1, \cdots, t_{u-1}, t_u, t_{u+1}, \cdots, t_{l-2}, t_{l-1}, t_0$, according to which we say that t_{u-1} and t_u are consecutive in γ, and $t'_u := t_{u-1}$ is the left neighbor of t_u (as usual t_{l-1} is the left neighbor of t_0). If we remove t_u from γ, then in the resulting γ', the loop is $t_0, \ldots, t_{u-1}, t_{u+1}, \ldots, t_{l_\gamma - 1}, t_0$, hence $l_{\gamma'} = l - 1$, and the left neighbor of t_{u+1} becomes t_{u-1}, but all other relations in terms of "left neighbors" stay the same.

1) Case 1. Removing consecutive elements: Suppose that there are consecutive elements in I_a for some a, say, for example $u, u+1 \in I_a$. The equation with respect to I_a is

$$\cdots + \mathbf{h}_{t_u} + \mathbf{h}_{t_{u+1}} + \cdots = \cdots + \mathbf{h}_{t_{u-1}} + \mathbf{h}_{t_u} + \cdots .$$

Clearly $\mathbf{h}_{t_u}$ can be canceled out on both sides of the equation, and it does not appear in any other equations with respect to I_b, $b \neq a$. Let γ' be the closed path by removing t_u, then t_{u-1} becomes the left neighbor of t_{u+1} in γ' and all other relations in terms of "neighbors" remain the same. Hence we have

$$\text{Case 1}: \qquad l_{\gamma'} = l - 1, \;\; v_{\gamma'} = v_\gamma, \;\; W_\gamma = nW_{\gamma'}.$$

In $W_{\gamma'}$, we may rename the variables so that $\gamma' : [0, l_{\gamma'}] \to [1, p]$ is a closed path with variables $t_0, \ldots, t_{l_{\gamma'-1}}$.

2) Case 2. Removing "leaves": For a closed path γ, the vertex $u \in I_a$ is called a "leaf" if $I_a = \{u\}$ and $\gamma(u-1) = \gamma(u+1) \neq \gamma(u)$. Hence $u-1, u+1 \in I_b$ for some $b \neq a$. The equation with respect to I_a is

$$\mathbf{h}_{t_u} = \mathbf{h}_{t_{u-1}} \Longrightarrow t_u = t_{u-1}. \tag{25}$$

The equation with respect to I_b is

$$\cdots + \mathbf{h}_{t_{u-1}} + \mathbf{h}_{t_{u+1}} + \cdots = \cdots + \mathbf{h}_{t_{u-2}} + \mathbf{h}_{t_u} + \cdots . \tag{26}$$

Assuming (25), then $\mathbf{h}_{t_u}$ and $\mathbf{h}_{t_{u-1}}$ can be canceled out trivially on both sides of (26). Hence we have solved that $t_u = t_{u-1}$, which can be removed from the variables. Let γ' be the resulting closed path. Removing both t_u, t_{u-1} from (25), it is clear that in γ', t_{u-2} becomes the left neighbor of t_{u+1} and all other relations in terms of "neighbors" remain the same. We have

$$\text{Case 2}: \qquad l_{\gamma'} = l - 2, \;\; v_{\gamma'} = v_\gamma - 1, \;\; W_\gamma = nW_{\gamma'}.$$

3) Case 3. Removing "transition" vertices: For a closed path γ, the vertex $u \in I_a$ is called a "transition" vertex if $I_a = \{u\}$ and $\gamma(u-1), \gamma(u), \gamma(u+1)$ are all distinct. Say $u-1 \in I_b$ and $u+1 \in I_c$, where a, b, c are all distinct. The equation with respect to I_a is still

$$\mathbf{h}_{t_u} = \mathbf{h}_{t_{u-1}} \Longrightarrow t_u = t_{u-1}. \tag{27}$$

The equations with respect to I_b, I_c are

$$\cdots + \mathbf{h}_{t_{u-1}} + \cdots \;\; = \;\; \cdots + \mathbf{h}_{t_{u-2}} + \cdots \tag{28}$$

$$\cdots + \mathbf{h}_{t_{u+1}} + \cdots \;\; = \;\; \cdots + \mathbf{h}_{t_u} + \cdots \tag{29}$$

Assuming (27), that is, replacing t_u by t_{u-1}, then (28) stays the same but (29) becomes

$$\cdots + \mathbf{h}_{t_{u+1}} + \cdots \;\; = \;\; \cdots + \mathbf{h}_{t_{u-1}} + \cdots$$

which means that by removing t_u, in the resulting γ', t_{u-1} becomes the left neighbor of t_{u+1} and all the other relations in terms of "neighbors" remain the same. So we have

$$\text{Case 3}: \qquad l_{\gamma'} = l-1, \;\; v_{\gamma'} = v_\gamma - 1, \;\; W_\gamma = W_{\gamma'}.$$

D. Conclusion on W_γ

In conclusion, suppose that altogether we perform u, v, and $w(\geq 0)$ times of Case 1, Case 2 and Case 3 reductions respectively on γ, maybe in different orders and combinations, to finally arrive at, after reordering the variables, a closed path $\gamma' : [0, l_{\gamma'}] \to [1, p]$ with $l_{\gamma'}, v_{\gamma'} \geq 1$, on which we could not do any of the reductions as described above. Then by definition γ' is a reduced path, and we also have

$$l_{\gamma'} = l_\gamma - u - 2v - w, \;\; v_{\gamma'} = v_\gamma - v - w, \;\; W_\gamma = n^{u+v} W_{\gamma'} \; . \tag{30}$$

There are two cases:

Case 1. If $v_{\gamma'} = l_{\gamma'} = 1$, then $W_{\gamma'} = n$. Hence in this case $W_\gamma = n^{l_\gamma - v_\gamma + 1}$.

Case 2. If $v_{\gamma'} \geq 2$, then $W_{\gamma'} \leq C_A \, n^{l_{\gamma'} - v_{\gamma'}}$ by (18). We have in this case $W_\gamma \leq C_A \, n^{l_\gamma - v_\gamma}$.

Denote by Γ the set of all the γ's that can be reduced to **Case 1**. We conclude that

$$\begin{cases} W_\gamma = n^{l_\gamma - v_\gamma + 1} & : \quad \text{if } \gamma \in \Gamma; \\ 0 \leq W_\gamma \leq C_A \, n^{l_\gamma - v_\gamma} & : \quad \text{if } \gamma \notin \Gamma, \end{cases}$$

E. Combinatorial structure of Γ

Finally we need to prove the identity

$$\sum_{\substack{\gamma \in \Gamma \subset \Pi_{l,p}/\Sigma_p \\ v_\gamma = v}} 1 = \frac{1}{v} \binom{l}{v-1} \binom{l-1}{v-1}. \tag{31}$$

The theory of random matrices has been extensively studied (see [1], [9]), and the above identity might be a well-known fact. Actually the left hand side appears naturally in the standard proof of the Marchenko-Pastur law for random matrices. Since we can not find a reference, we may sketch a proof here.

Let $X = (\mathbf{x}_{ij}) \in \mathbb{R}^{p\times n}$ be a random matrix where $\mathbf{x}_{ij}$'s are i.i.d, $\mathbb{E}(\mathbf{x}_{ij}) = 0, \mathbb{E}(\mathbf{x}_{ij}^2) = 1$ and $p < n$. Define

$$S = \frac{1}{n} XX^T.$$

Then

$$\frac{1}{p}\,\mathbb{E}\left(\mathrm{Tr}(S^l)\right) = \frac{1}{pn^l}\sum_{\gamma,\tau}\mathbb{E}\left(\mathbf{x}_{\gamma(0)t_0}\mathbf{x}_{\gamma(1)t_0}\mathbf{x}_{\gamma(1)t_1}\mathbf{x}_{\gamma(2)t_1}\ldots\mathbf{x}_{\gamma(l-1)t_{l-1}}\mathbf{x}_{\gamma(0)t_{l-1}}\right) = \frac{1}{pn^l}\sum_{\gamma,\tau}\mathbb{E}\left(\gamma,\tau\right),$$

where the sum is over all maps $\gamma \in \Pi_{l,p}$ and all $\tau := \{t_i\}_{i=0}^{l-1} \in [1,n]^l$. Now this corresponds to a directed loop on a bipartite graph from the vertex set $\{\gamma(0),\ldots,\gamma(l-1)\}$ to the vertex set $\{t_0,\ldots,t_{l-1}\}$ with $2l$ steps. As the standard proof goes, each edge must appear at least twice, otherwise $\mathbb{E}(\gamma,\tau) = 0$. Hence we have at most l edges in the graph, and at most $l+1$ vertices in the skeleton. The optimal situation, that is, graphs with exactly l edges and $l+1$ vertices, or "double trees" will give the main contribution. Terms arising from other configuration of graphs are negligible and can be ignored. The standard result on counting such "double trees" is that, for each $1 \le v \le l$, the number of double tree shapes with v vertices in γ (i.e., $v_\gamma = v$) and $l - v + 1$ vertices in τ is given by the right hand side of (31) (see [1, page 20, Exercise 2.1.18]). A little thought about properties of Γ concludes that the left hand side of (31) also counts the total number of such double trees. The finishes the proof of the identity (31). $\square$

REFERENCES

[1] G. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge studies in advanced mathematics 118, Cambridge Univ. Press, 2010.

[2] B. Babadi, S. S. Ghassemzadeh, and V. Tarokh, "Group randomness properties of pseudo-noise and Gold sequences," presented at the Canadian Workshop on Information Theory, 2011.

[3] B. Babadi and V. Tarokh, "Spectral distribution of random matrices from binary linear block codes," *IEEE Trans. Inform. Theory*, vol. 57, no. 6, pp. 3953–3962, 2011.

[4] B. Babadi and V. Tarokh, "Spectral distribution of product of pseudorandom matrices formed from binary block codes," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 970–978, 2013.

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. Hoboken, NJ: Wiley, 1991, vol. 2.

[6] R. Gold, "Maximal recursive sequences with 3-valued recursive crosscorrelation functions (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 154–156, 1968.

[7] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland Mathematical Library, 1988.

[8] V. A. Marchenko and L. A. Pastur, "The distribution of eigenvalues in certain sets of random matrices," Math. Sb., vol. 72, pp. 507–536, 1967.

[9] M. L. Mehta, *Random matrices*, Pure and Applied Mathematics, Vol. 142, Third Edition, Academic Press, 2004.

[10] S. Pafka, M. Potters, and I. Kondor, *Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization 2004* [Online]. Available: arxiv: cond-mat/0402573

[11] V. M. Sidel'nikov, "Weight spectrum of binary Bose-Chaudhuri-Hoquinghem codes," Probl. Inf. Transm., vol. 7:1, pp. 11–17, 1971.

[12] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications.* Hanover, MA: Now Publishers Inc., 2004, Foundations and Trends in Communications and Information Theory.