Bernoulli **27**(2), 2021, 1077–1102
https://doi.org/10.3150/20-BEJ1266

Limiting behavior of large correlated Wishart matrices with chaotic entries

SOLESNE BOURGUIN[1], CHARLES-PHILIPPE DIEZ[2,*] and CIPRIAN A. TUDOR[2,†]

[1]*Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, MA 02215, USA. E-mail: bourguin@math.bu.edu*

[2]*CNRS, Laboratoire Paul Painlevé, Université de Lille, UMR 8524, F-59655 Villeneuve d'Ascq, France. E-mail: *charles-philippe.diez@univ-lille.fr; †ciprian.tudor@math.univ-lille.fr*

We study the fluctuations, as $d, n \to \infty$, of the Wishart matrix $\mathcal{W}_{n,d} = \frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T$ associated to a $n \times d$ random matrix $\mathcal{X}_{n,d}$ with non-Gaussian entries. We analyze the limiting behavior in distribution of $\mathcal{W}_{n,d}$ in two situations: when the entries of $\mathcal{X}_{n,d}$ are independent elements of a Wiener chaos of arbitrary order and when the entries are partially correlated and belong to the second Wiener chaos. In the first case, we show that the (suitably normalized) Wishart matrix converges in distribution to a Gaussian matrix while in the correlated case, we obtain its convergence in law to a diagonal non-Gaussian matrix. In both cases, we derive the rate of convergence in the Wasserstein distance via Malliavin calculus and analysis on Wiener space.

Keywords: Wishart matrix; multiple stochastic integrals; Malliavin calculus; Stein's method; Rosenblatt process; fractional Brownian motion; high-dimensional regime

1. Introduction

Random matrix theory plays an important role in various areas of applications, including statistical physics, engineering sciences, signal processing or mathematical finance. The various tools that can be used to study random matrices come from different branches of mathematics, such as combinatorics, non-commutative algebra, geometry, spectral analysis and, of course, probability and statistics. We focus on a special type of random matrices, called *Wishart matrices*, which have been introduced in Wishart [21]. Given a $n \times d$ random matrix $\mathcal{X}_{n,d} = (X_{ij})_{1 \le i \le n, 1 \le j \le d}$ with real entries, its associated Wishart matrix $\mathcal{W}_{n,d} = (W_{ij})_{1 \le i,j \le n}$ is the symmetric $n \times n$ matrix $\mathcal{W}_{n,d} = \frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T$ ($\mathcal{X}^T$ being the transpose of the matrix $\mathcal{X}$). The class of Wishart matrices constitutes a special class of sample covariance matrices with applications in multivariate analysis or statistical theory, see, for example, the surveys (Bishop, Del Moral and Niclas [1], Johnstone [10], Rasmussen and Williams [17]). The limiting behavior of this type of random matrices, as d goes to infinity and n is fixed (which is referred to as the classical or finite dimensional regime) or when both n, d tend to infinity (usually called the high dimensional regime), has been studied by many authors. The starting point of this analysis is the situation where the entries of the matrix $\mathcal{X}_{n,d}$ are i.i.d. and n is fixed. In this case, the Wishart matrix associated to $\mathcal{X}_{n,d}$ converges almost surely, as $d \to \infty$, to the $n \times n$ identity matrix $\mathcal{I}_n$ by the strong law of large numbers and the renormalized Wishart matrix $\sqrt{d}(\mathcal{W}_{n,d} - \mathcal{I}_n)$ satisfies a Central Limit Theorem (CLT in the sequel). Later, due to the increasing need of handling large data sets, several authors investigated the high dimensional regime, when the matrix size n also goes to infinity. Different strategies have been considered in this case. A classical approach is based on the study of the empirical spectral distribution and of the eigenvalues of $\mathcal{W}_{n,d}$. It is well known that if $n, d \to \infty$ such that $n/d \to c \in (0, \infty)$, then the empirical spectral distribution of the Wishart matrix converges weakly to the so-called Marchenko–Pastur distribution (see Marčenko and Pastur [11]). A more recent approach consists in analyzing the distance in distribution (for example, under the total variation distance or Wasserstein

1350-7265 © 2021 ISI/BS

distance) between the renormalized Wishart matrix $\sqrt{d}(\mathcal{W}_{n,d} - \mathcal{I}_n)$ and its limiting distribution when d and n are large. This approach has been used in, for example, Bubeck et al. [4], Bubeck and Ganguly [5], Jiang and Li [9], Rácz and Richey [16], Nourdin and Zheng [13]. It has been discovered that the distance (in the Wasserstein or total variation sense) between the distribution of the renormalized Wishart matrix and its limiting distribution (when this limit is Gaussian, which happens in all the cases except when the entries have a strong enough correlation, see Nourdin and Zheng [13]), as $n, d \to \infty$, is of order less than n^3/d. In the above references, several situations have been studied: the entries of the initial matrix $\mathcal{X}_{n,d}$ are independent and Gaussian (see Bubeck et al. [4], Jiang and Li [9], Rácz and Richey [16]), the entries are independent and not necessarily Gaussian (they are supposed to have a log-concave distribution in Bubeck and Ganguly [5]) or the entries are Gaussian and partially correlated (see Nourdin and Zheng [13]). While in most references the proofs are based on entropy or moments analysis, in Nourdin and Zheng [13] the authors use the recent Stein–Malliavin calculus (see Nourdin and Peccati [12]).

Our purpose is to use the techniques of Malliavin calculus and analysis on Wiener space in order to generalize the above results in two directions. First, we start with an $n \times d$ matrix $\mathcal{X}_{n,d}$ whose entries are independent (not necessarily identically distributed) elements of Wiener chaoses of arbitrary order. That is, we assume that for every $1 \le i \le n$ and for every $1 \le j \le d$,

$$X_{ij} = I_{q_i}(f_{ij}), \tag{1.1}$$

with $f_{ij} \in \mathfrak{H}^{\odot q_i}$, where $q_i \ge 1$ and the maximum of the q_i's is bounded by an integer number N_0 for every $1 \le i \le n$. In (1.1), I_q denotes the multiple Wiener integral of order q with respect to an isonormal process W. Assume that the entries have the same second and fourth moments, that is, for every $1 \le i \le n$ and $1 \le j \le d$,

$$\mathbb{E}\big(X_{ij}^2\big) = q_i! \| f_{ij} \|^2_{\mathfrak{H}^{\otimes q_i}} = 1 \quad \text{and} \quad \mathbb{E}\big(X_{ij}^4\big) = m_4.$$

In this situation, we obtain the convergence in law of the corresponding renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d} = (\widetilde{W}_{ij})_{1\le i,j\le n}$ with entries $\widetilde{W}_{ij} = \sqrt{d} W_{ij}$ for $1 \le i, j \le n$ to the GOE (Gaussian Orthogonal Ensemble) matrix $\mathcal{Z}_n$ given by (3.7). This is a symmetric random matrix $\mathcal{Z}_n = (Z_{ij})_{1\le i,j\le n}$ whose diagonal elements follow the distribution $Z_{ii} \sim N(0, m_4 - 1)$ while the non-diagonal entries are such that $Z_{ij} \sim N(0,1)$ if $1 \le i < j \le n$ and $Z_{ij} = Z_{ji}$ if $1 \le i < j \le n$, the variables $\{Z_{ij} : i \le j\}$ being independent.

The study of Wishart matrices based on an initial matrix $\mathcal{X}_{n,d}$ with independent elements in (potentially different) Wiener chaoses is motivated by the following facts. As mentioned above, Wishart matrices can be viewed as sample covariance matrices and the elements of the matrix $\mathcal{X}_{n,d}$ can be interpreted as the data. In recent years, the statistical inference based on observations belonging to Wiener chaoses of arbitrary order has been intensively studied (see, among others, Chronopoulou, Tudor and Viens [6], Clausel et al. [7], Pipiras and Taqqu [15], Tudor [18]). Another motivation is related to the concept of universality, which has been tremendously studied for random matrices by many authors (see e.g. Edelman, Guionnet and Péché [8] and the references therein). Loosely speaking, the notion of universality implies to understand the behavior of random matrices with entries from a general (non necessarily Gaussian) distribution and to see if the behavior displayed by Gaussian matrices still holds in the general case.

We actually show that, when $n, d \to \infty$, the distance between the renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d} = (\widetilde{W}_{ij})_{1\le i,j\le n}$ and the GOE matrix is of order less that n^3/d. This generalizes the results of Bubeck et al. [4], Bubeck and Ganguly [5], Jiang and Li [9], Rácz and Richey [16]. More precisely, we prove the following result.

Theorem 1.1. *Consider the renormalized Wishart matrix* $\widetilde{\mathcal{W}}_{n,d}$ *with entries given by* (3.6). *Then for every* $n \geq 1$, $\widetilde{\mathcal{W}}_{n,d}$ *converges in distribution componentwise, as* $d \to \infty$, *to the matrix* $\mathcal{Z}_n$ *given by* (3.7). *Moreover, there exists a positive constant* C *such that for every* $n, d \geq 1$,

$$d_W(\widetilde{\mathcal{W}}_{n,d}, \mathcal{Z}_n) \leq C\sqrt{\frac{n^3}{d}}, \tag{1.2}$$

where d_W *denotes the Wasserstein distance defined in Section* 2.1.

Another direction of study is to start with a matrix $\mathcal{X}_{n,d}$ whose elements are non-Gaussian and partially correlated. As pointed out in for example, Bubeck and Ganguly [5], obtaining an approximation result without the assumption of independence represents a natural question which has been a subject of wide interest. We will assume that these entries are elements of the second Wiener chaos, correlated on the same row, with the correlation being given by the increments of the Rosenblatt process (see Section 4 for the definition and basic properties of this stochastic process). More precisely, the entries of the matrix $\mathcal{X}_{n,d} = (X_{ij})_{1\leq i\leq n, 1\leq j\leq}$ are given by $X_{ij} = Z_j^{H,i} - Z_{j-1}^{H,i}$, where $Z^{H,i}$, $1 \leq i \leq n$ are n-independent Rosenblatt processes with the same Hurst parameter $H \in (\frac{1}{2}, 1)$. The definition and basic properties of the Rosenblatt process are recalled in Section 4. This stochastic process is a non-Gaussian self-similar process with stationary increments and long-memory. Due to these properties, it found several applications in various areas (hydrology, finance, interned traffic analysis, and more). For more details on the theoretical aspects and practical applications of the Rosenblatt process, we refer to the monographs Pipiras and Taqqu [15], Tudor [18].

Note that the correlation structure of the Rosenblatt process is the same as the one of the fractional Brownian motion (fBm). In this sense, the correlation on the rows of the matrix $\mathcal{X}_{n,d}$ considered in our work is the same as in Nourdin and Zheng [13] (where the entries are increments of the fBm). Nevertheless, the non-Gaussian character of the entries brings more complexity and leads to a different behavior of the associated Wishart matrix. Actually, we show that the renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d} = (\widetilde{W}_{ij})_{1\leq i,j\leq n}$ with $\widetilde{W}_{ij} = c_{1,H}^{-1} d^{1-H} W_{ij}$ (the constant $c_{1,H}$ is defined in (4.7)) converges to a diagonal matrix whose diagonal entries are random variables distributed according to the Rosenblatt distribution and we are also able to quantify the distance associated to this limit theorem. Our result can be stated as follows.

Theorem 1.2. *Let* $\widetilde{\mathcal{W}}_{n,d}$ *be the renormalized Wishart matrix* (4.10) *and let* $\mathcal{R}_n^H$ *be the diagonal matrix with entries given by* (4.12). *Then, for every* $n \geq 1$, *the random matrix* $\widetilde{\mathcal{W}}_{n,d}$ *converges componentwise in distribution, as* $d \to \infty$, *to the matrix* $\mathcal{R}_n^H$. *Moreover, there exists a positive constant* C *such that as* $n, d \geq 1$,

$$d_W\big(\widetilde{\mathcal{W}}_{n,d}, \mathcal{R}_n^H\big) \leq C \begin{cases} nd^{\frac{1}{2}-H} & \text{if } H \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ n\sqrt{\log(d)}\,d^{-\frac{1}{4}} & \text{if } H = \frac{3}{4}, \\ nd^{H-1} & \text{if } H \in \left(\frac{3}{4}, 1\right), \end{cases}$$

where d_W *denotes the Wasserstein distance defined in Section* 2.1.

In the case of independent entries, the proof of our main result is based on the Stein–Malliavin calculus and the characterization of independent random variables in Wiener chaos while when the entries of the initial matrix $\mathcal{X}_{n,d}$ are correlated, we use the properties of random variables in the second Wiener chaos and in particular the behavior of the increments of the Rosenblatt process.

The paper is organized as follows. In Section 2, we recall several facts related to the distance between the probability distributions of random matrices and random vectors, as well as the basics of Wiener space analysis and Malliavin calculus. In Section 3, we analyze the fluctuations of the Wishart matrix constructed from a matrix with independent entries in an arbitrary Wiener chaos, while in Section 4 we treat the situation where the elements of the starting matrix $\mathcal{X}_{n,d}$ are non-Gaussian and partially correlated.

2. Preliminaries

In this preliminary part, we recall some facts related to the concept of distance between the probability distributions of random matrices and random vectors and we introduce the tools of the Malliavin calculus needed in the sequel.

2.1. Distances between random matrices

We will use the Wasserstein distance between two random matrices taking values in $\mathcal{M}_n(\mathbb{R})$, which denotes the space of $n \times n$ real matrices. Given two $\mathcal{M}_n(\mathbb{R})$-valued random matrices $\mathcal{X}$ and $\mathcal{Y}$, the Wasserstein distance between them is given by

$$d_W(\mathcal{X}, \mathcal{Y}) = \sup_{\|g\|_{\mathrm{Lip}} \le 1} \left| \mathbb{E}\big(g(\mathcal{X})\big) - \mathbb{E}\big(g(\mathcal{Y})\big) \right|,$$

where the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$ of $g\colon \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}$ is defined by

$$\|g\|_{\mathrm{Lip}} = \sup_{A \ne B \in \mathcal{M}_n(\mathbb{R})} \frac{|g(A) - g(B)|}{\|A - B\|_{\mathrm{HS}}},$$

with $\|\cdot\|_{\mathrm{HS}}$ denoting the Hilbert–Schmidt norm on $\mathcal{M}_n(\mathbb{R})$.

With this definition at hand, we recall the definition of the notion of ϕ-closeness between random matrices.

Definition 2.1. For every $n \ge 1$, let $\{\mathcal{A}_{n,d} \colon d \ge 1\}$ and $\{\mathcal{B}_{n,d} \colon d \ge 1\}$ be two families of $n \times n$ random matrices. Let $\phi\colon \mathbb{N} \times \mathbb{N} \to \mathbb{R}_+$ be given. Then, $\mathcal{A}_{n,d}$ is said to be ϕ-close to $\mathcal{B}_{n,d}$ if $d_W(\mathcal{A}_{n,d}, \mathcal{B}_{n,d})$ converges to zero as $n, d \to \infty$ and $\phi(n,d) \to 0$.

We will also make use of the Wasserstein distance between random vectors, defined analogously as in the matrix case. Namely, if X, Y are two n-dimensional random vectors, then the Wasserstein distance between them is defined to be

$$d_W(X, Y) = \sup_{\|g\|_{\mathrm{Lip}} \le 1} \left| \mathbb{E}\big(g(X)\big) - \mathbb{E}\big(g(Y)\big) \right|, \tag{2.1}$$

where the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}}$ of $g\colon \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\|g\|_{\mathrm{Lip}} = \sup_{x \ne y \in \mathbb{R}^n} \frac{|g(x) - g(y)|}{\|x - y\|_{\mathbb{R}^n}},$$

with $\|\cdot\|_{\mathbb{R}^n}$ denoting the Euclidean norm on $\mathbb{R}^n$.

If $\mathcal{X} = (X_{ij})_{1\le i,j\le n}$ is an $n \times n$ symmetric random matrix, we associate to it its "half-vector" defined to be the $n(n+1)/2$-dimensional random vector

$$\mathcal{X}^{\text{half}} = (X_{11}, X_{12}, \dots, X_{1n}, X_{22}, X_{23}, \dots, X_{2n}, \dots, X_{nn}). \tag{2.2}$$

It turns out that, in the case of two symmetric matrices, the Wasserstein distance between said matrices can be bounded from above by a constant multiple of the Wasserstein distance between their associated half-vectors. More specifically, we have the following lemma (see Nourdin and Zheng [13], Lemma 2.2).

Lemma 2.1. *Let $\mathcal{X}, \mathcal{Y}$ be two symmetric random matrices with values in $\mathcal{M}_n(\mathbb{R})$. Then*

$$d_W(\mathcal{X}, \mathcal{Y}) \le \sqrt{2} d_W\big(\mathcal{X}^{\text{half}}, \mathcal{Y}^{\text{half}}\big),$$

where $\mathcal{X}^{\text{half}}$, $\mathcal{Y}^{\text{half}}$ are the associated half-vectors defined in (2.2).

2.2. Elements of Malliavin calculus

We briefly describe the main tools from analysis on Wiener space that we will need in this paper. For a complete treatment of this topic, we refer the reader to the monographs Nualart [14] or Nourdin and Peccati [12].

Let $\mathfrak{H}$ be a real separable Hilbert space and $\{W(h) \colon h \in \mathfrak{H}\}$ an isonormal Gaussian process indexed by it, that is, a centered Gaussian family of random variables such that $\mathbb{E}(W(h)W(g)) = \langle h, g\rangle_{\mathfrak{H}}$. Denote by I_n the multiple Wiener (or Wiener–Itô) stochastic integral of order $n \ge 0$ with respect to W (see Nualart [14], Section 1.1.2). The mapping I_n is actually an isometry between the Hilbert space $\mathfrak{H}^{\odot n}$ (symmetric tensor product) equipped with the scaled norm $\frac{1}{\sqrt{n!}}\|\cdot\|_{\mathfrak{H}^{\otimes n}}$ and the Wiener chaos of order n, which is defined as the closed linear span of the random variables

$$\big\{H_n\big(W(h)\big) \colon h \in \mathfrak{H}, \|h\|_{\mathfrak{H}} = 1\big\},$$

where H_n is the n-th Hermite polynomial given by $H_0 = 1$ and for $n \ge 1$

$$H_n(x) = \frac{(-1)^n}{n!} \exp\left(\frac{x^2}{2}\right) \frac{d^n}{dx^n}\left(\exp\left(-\frac{x^2}{2}\right)\right), \quad x \in \mathbb{R}.$$

Multiple Wiener integrals enjoy the following isometry property: for any integers $m, n \ge 1$,

$$\mathbb{E}\big(I_n(f)I_m(g)\big) = \mathbb{1}_{\{n=m\}} n! \langle \tilde{f}, \tilde{g}\rangle_{\mathfrak{H}^{\otimes n}}, \tag{2.3}$$

where $\tilde{f}$ denotes the symmetrization of f and we recall that $I_n(f) = I_n(\tilde{f})$.

Recall the multiplication formula satisfied by multiple Wiener integrals: for any integers $n, m \ge 1$, and any $f \in \mathfrak{H}^{\odot n}$ and $g \in \mathfrak{H}^{\odot m}$, it holds that

$$I_n(f)I_m(g) = \sum_{r=0}^{n\wedge m} r! \binom{n}{r}\binom{m}{r} I_{m+n-2r}(f \otimes_r g), \tag{2.4}$$

where the r-th contraction of f and g is defined by, for $0 \le r \le m \wedge n$,

$$f \otimes_r g = \sum_{i_1,\dots,i_r=1}^{\infty} \langle f, e_{i_1} \otimes \cdots \otimes e_{i_r}\rangle_{\mathfrak{H}^{\otimes r}} \otimes \langle g, e_{i_1} \otimes \cdots \otimes e_{i_r}\rangle_{\mathfrak{H}^{\otimes r}}, \tag{2.5}$$

with $\{e_i : i \ge 1\}$ denoting a complete orthonormal system in $\mathfrak{H}$.

Recall that any square integrable random variable F which is measurable with respect to the σ-algebra generated by W can be expanded into an orthogonal sum of multiple Wiener integrals:

$$F = \sum_{n=0}^{\infty} I_n(f_n), \tag{2.6}$$

where $f_n \in \mathfrak{H}^{\odot n}$ are (uniquely determined) symmetric functions and $I_0(f_0) = \mathbb{E}(F)$.

Let L denote the Ornstein–Uhlenbeck operator, whose action on a random variable F with chaos decomposition (2.6) and such that $\sum_{n=1}^{\infty} n^2 n! \| f_n \|^2_{\mathfrak{H}^{\otimes n}} < \infty$ is given by

$$LF = -\sum_{n=1}^{\infty} n I_n(f_n).$$

For $p > 1$ and $\alpha \in \mathbb{R}$ we introduce the Sobolev–Watanabe space $\mathbb{D}^{\alpha,p}$ as the closure of the set of polynomial random variables with respect to the norm

$$\|F\|_{\alpha,p} = \left\| (I - L)^{\frac{\alpha}{2}} F \right\|_{L^p(\Omega)},$$

where I represents the identity operator. We denote by D the Malliavin derivative that acts on smooth random variables of the form $F = g(W(h_1), \dots, W(h_n))$, where g is a smooth function with compact support and $h_i \in \mathfrak{H}$, $1 \le i \le n$. Its action on such a random variable F is given by

$$DF = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \big(W(h_1), \dots, W(h_n) \big) h_i.$$

The operator D is closable and continuous from $\mathbb{D}^{\alpha,p}$ into $\mathbb{D}^{\alpha-1,p}(\mathfrak{H})$.

3. Random matrices with independent chaotic entries

In this section, we consider random matrices $\mathcal{X}_{n,d} = (X_{ij})_{1 \le i \le n, 1 \le j \le d}$ with independent entries belonging to arbitrary order Wiener chaoses associated with an isonormal Gaussian process $W = \{W(h) \colon h \in \mathfrak{H}\}$ as introduced in Section 2.2. Moreover, we assume that the elements on the same row of the matrix $\mathcal{X}_{n,d}$ belong to the same Wiener chaos, while the order of the chaos may change from one row to another. In other words, we assume that for every $1 \le i \le n$ and for every $1 \le j \le d$,

$$X_{ij} = I_{q_i}(f_{ij}), \tag{3.1}$$

with $f_{ij} \in \mathfrak{H}^{\odot q_i}$, where the integer numbers q_i for every $1 \le i \le n$ are all in the set $\{1, 2, \dots, N_0\}$ with $N_0 \ge 1$ being an integer. Here and in the sequel, I_q denotes the multiple Wiener integral of order q with respect to W introduced in Section 2.2.

We do not assume that the entries have the same probability distribution, only that they have the same second and fourth moments, that is, for every $1 \le i \le n$ and $1 \le j \le d$,

$$\mathbb{E}\big(X_{ij}^2\big) = q_i! \| f_{ij} \|^2_{\mathfrak{H}^{\otimes q_i}} = 1 \quad \text{and} \quad \mathbb{E}\big(X_{ij}^4\big) = m_4. \tag{3.2}$$

Consider the centered Wishart matrix (which is what will be referred to as Wishart matrix in the sequel) $\mathcal{W}_{n,d} = (W_{ij})_{1 \le i,j \le n}$ defined by

$$\mathcal{W}_{n,d} = \frac{1}{d} \mathcal{X}_{n,d} \mathcal{X}_{n,d}^T - \mathcal{I}_n, \tag{3.3}$$

where $\mathcal{I}_n$ denotes the identity matrix of $\mathcal{M}_n(\mathbb{R})$, and $\mathcal{X}^T$ stands for the transpose of the matrix $\mathcal{X}$. Note that the Wishart matrix $\mathcal{W}_{n,d}$ is a symmetric $n \times n$ matrix and its entries can be explicit as

$$W_{ii} = \frac{1}{d}\sum_{k=1}^{d}\left(X_{ik}^2 - 1\right), \quad i = 1, \dots, n \tag{3.4}$$

and

$$W_{ij} = \frac{1}{d}\sum_{k=1}^{d} X_{ik}X_{jk}, \quad 1 \le i \neq j \le n. \tag{3.5}$$

Note that the independence of the entries X_{ij} and assumption (3.2) yield, for any $1 \le i \le n$,

$$\mathbb{E}\left(W_{ii}^2\right) = \frac{1}{d^2}\sum_{k=1}^{d} E\left(\left(X_{ik}^2 - 1\right)^2\right) = \frac{m_4 - 1}{d},$$

and for $1 \le i \neq j \le n$,

$$\mathbb{E}\left(W_{ij}^2\right) = \frac{1}{d^2}\sum_{k=1}^{d} \mathbb{E}\left(X_{ik}^2\right)\mathbb{E}\left(X_{jk}^2\right) = \frac{1}{d}.$$

Based on this observation, we define the renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d} = (\widetilde{W}_{ij})_{1 \le i,j \le n}$ as

$$\widetilde{W}_{ij} = \sqrt{d}\,W_{ij} \tag{3.6}$$

for all $1 \le i, j \le n$.

Also, consider the GOE matrix $\mathcal{Z}_n = (Z_{ij})_{1 \le i,j \le n}$ with entries given by

$$\begin{cases} Z_{ii} \sim N(0, m_4 - 1) & \text{for } 1 \le i \le n, \\ Z_{ij} \sim N(0, 1) & \text{for } 1 \le i < j \le n, \\ Z_{ij} = Z_{ji} & \text{for } 1 \le j < i \le n, \end{cases} \tag{3.7}$$

where the entries $\{Z_{ij} : i \le j\}$ are independent.

Remark 3.1. Note that proving Theorem 1.1 entails proving that the matrices $\widetilde{\mathcal{W}}_{n,d}$ and $\mathcal{Z}_n$ are ϕ-close for $\phi(n,d) = \frac{n^3}{d}$ (as introduced in Definition 2.1).

As pointed out in Section 2.1, assessing the Wasserstein distance between symmetric random matrices can be shifted to the problem of estimating the Wasserstein distance between associated random vectors (see Lemma 2.1). In our context, a helpful result in this direction is Nourdin and Peccati [12], Theorem 6.1.1, which we restate here for convenience.

Theorem 3.1 (Theorem 6.1.1 in Nourdin and Peccati [12]). *Fix $m \ge 2$, and let $F = (F_1, \dots, F_m)$ be a centered m-dimensional random vector with $F_i \in \mathbb{D}^{1,4}$ for every $i = 1, \dots, m$. Let $C \in \mathcal{M}_m(\mathbb{R})$ be a symmetric and positive definite matrix, and let $Z \sim N_m(0, C)$. Then,*

$$d_W(F, Z) \le \|C^{-1}\|_{\mathrm{op}}\|C\|_{\mathrm{op}}^{1/2}\sqrt{\sum_{i,j=1}^{m}\mathbb{E}\left(\left(C_{ij} - \langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}\right)^2\right)},$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm on $\mathcal{M}_m(\mathbb{R})$.

3.1. Independent random variables in Wiener chaos

This section prepares the proof of Theorem 1.1 by providing results related to the independence of multiple Wiener integrals. By a standard argument based on the fact that separable Hilbert spaces are isometrically isomorphic, we may assume, when it serves the clarity of our exposition, that $\mathfrak{H} = L^2(T, \mathcal{B}, \mu)$ where μ is a σ-finite measure without atoms.

Recall that the entries of the matrix $\mathcal{X}$, on which our Wishart matrices are based, are independent multiple Wiener integrals of possibly different orders. The independence of random variables in Wiener chaos can be characterized in terms of their kernels via the celebrated Üstünel-Zakai criterion (see Üstünel and Zakai [20]), which we will intensively make use of in the sequel. We recall the criterion here for convenience.

Theorem 3.2 (Üstünel-Zakai Üstünel and Zakai [20]). *For any $n, m \geq 1$, let $f \in \mathfrak{H}^{\otimes n}$ and $g \in \mathfrak{H}^{\otimes m}$. The multiple Wiener integrals $I_n(f)$ and $I_m(g)$ are independent if and only if*

$$f \otimes_1 g = 0 \quad \textit{almost everywhere on } \mathfrak{H}^{\otimes m+n-2}. \tag{3.8}$$

Remark 3.2. Relation (3.8) also implies that

$$f \otimes_r g = 0 \quad \text{almost everywhere on } \mathfrak{H}^{\otimes m+n-2r}$$

for all $1 \leq r \leq n \wedge m$.

We will also need the notion of strong independence of random variables introduced in Bourguin and Tudor [2] (to which we refer for various properties of strongly independent random variables).

Definition 3.1. Two random variables X and Y with Wiener chaos decomposition

$$X = \sum_{n=0}^{\infty} I_n(f_n) \quad \text{and} \quad Y = \sum_{m=0}^{\infty} I_m(g_m),$$

where $f_n \in \mathfrak{H}^{\odot n}$, $g_m \in \mathfrak{H}^{\odot m}$ for every $n, m \geq 0$, are said to be strongly independent if every chaos component of X is independent of every chaos component of Y, that is, for every $n, m \geq 0$, the random variables $I_n(f_n)$ and $I_m(g_m)$ are independent.

The following lemma assesses the strong independence of squares of chaotic random variables.

Lemma 3.1. *Let $X = I_n(f)$, $f \in \mathfrak{H}^{\odot n}$ and $Y = I_m(g)$, $g \in \mathfrak{H}^{\odot m}$ be independent. Then, the random variables X^2 and Y^2 are strongly independent.*

Proof. By the product formula for multiple Wiener integrals (2.4),

$$X^2 = \sum_{r_1=0}^{n} r_1! \binom{n}{r_1}^2 I_{2n-2r_1}(f \otimes_{r_1} f)$$

and

$$Y^2 = \sum_{r_2=0}^{m} r_2! \binom{m}{r_2}^2 I_{2m-2r_2}(g \otimes_{r_2} g).$$

It suffices to show that for every $0 \le r_1 \le n-1$ and $0 \le r_2 \le m-1$, the random variables $I_{2n-2r_1}(f \otimes_{r_1} f)$ and $I_{2m-2r_2}(g \otimes_{r_2} g)$ are independent, which by (3.8) is equivalent to

$$(f \tilde{\otimes}_{r_1} f) \otimes_1 (g \tilde{\otimes}_{r_2} g) = 0 \tag{3.9}$$

almost everywhere on $\mathfrak{H}^{\otimes 2n+2m-2r_1-2r_2}$. By the definition of contractions (2.5), with $\mathfrak{S}_n$ denoting the group of permutations of $\{1, \dots, n\}$, we have

$$\begin{aligned}
&(f \tilde{\otimes}_{r_1} f)(t_1, \dots, t_{2n-2r_1}) \\
&\quad = \frac{1}{(2n-2r_1)!} \sum_{\sigma \in \mathfrak{S}_{2n-2r_1}} \int_{T^{r_1}} f(u_1, \dots, u_{r_1}, t_{\sigma(1)}, \dots, t_{\sigma(n-r_1)}) \\
&\qquad \times f(u_1, \dots, u_{r_1}, t_{\sigma(n-r_1+1)}, \dots, t_{\sigma(2n-2r_1)})\, du_1 \cdots du_{r_1}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&(g \tilde{\otimes}_{r_2} g)(t_1, \dots, t_{2m-2r_2}) \\
&\quad = \frac{1}{(2m-2r_2)!} \sum_{\tau \in \mathfrak{S}_{2m-2r_2}} \int_{T^{r_2}} g(u_1, \dots, u_{r_2}, t_{\tau(1)}, \dots, t_{\tau(m-r_2)}) \\
&\qquad \times g(u_1, \dots, u_{r_2}, t_{\tau(m-r_2+1)}, \dots, t_{\tau(2m-2r_2)})\, du_1 \cdots du_{r_2}.
\end{aligned}$$

Hence, we can write

$$\begin{aligned}
&\bigl((f \tilde{\otimes}_{r_1} f) \otimes_1 (g \tilde{\otimes}_{r_2} g)\bigr)(t_1, \dots, t_{2n-2r_1+2m-2r_2-2}) \\
&\quad = \int_T (f \tilde{\otimes}_{r_1} f)(t_1, \dots, t_{2n-2r_1-1}, x) \\
&\qquad \times (g \tilde{\otimes}_{r_2} g)(t_{2n-2r_1}, \dots, t_{2n-2r_1+2m-2r_2-2}, x)\, dx.
\end{aligned} \tag{3.10}$$

Note that for a symmetric function $h \in \mathfrak{H}^{\odot n}$, it holds that

$$\tilde{h}(t_1, \dots, t_{n-1}, x) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_{n-1}} \sum_{i=1}^{n} h(t_{\sigma(1)}, \dots, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, \dots, t_{\sigma(n-1)}),$$

so that by plugging the above identity into (3.10), we get

$$\begin{aligned}
&\bigl[(f \tilde{\otimes}_{r_1} f) \otimes_1 (g \tilde{\otimes}_{r_2} g)\bigr](t_1, \dots, t_{2n-2r_1+2m-2r_2-2}) \\
&\quad = \frac{1}{(2n-2r_1-1)!(2m-2r_2-1)!} \\
&\qquad \times \sum_{\sigma \in \mathfrak{S}_{2n-2r_1-1}, \tau \in \mathfrak{S}_{2m-2r_2-1}} \sum_{i=1}^{2n-2r_1} \sum_{j=1}^{2m-2r_2} \int_T (f \\
&\qquad \otimes_{r_1} f)(t_{\sigma(1)}, \dots, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, \dots, t_{\sigma(2n-2r_1-1)}) \\
&\qquad \times (g \otimes_{r_2} g)(t_{\tau(1)}, \dots, t_{\tau(j-1)}, x, t_{\tau(j+1)}, \dots, t_{\tau(2m-2r_2-1)})\, dx.
\end{aligned}$$

To obtain (3.9), it suffices to show that for all $1 \le i \le 2n - 2r_1$ and $1 \le j \le 2m - 2r_2$,

$$\int_T (f \otimes_{r_1} f)(t_{\sigma(1)}, \dots, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, \dots, t_{\sigma(2n-2r_1-1)}) \\ \times (g \otimes_{r_2} g)(t_{\tau(1)}, \dots, t_{\tau(j-1)}, x, t_{\tau(j+1)}, \dots, t_{\tau(2m-2r_2-1)})\, dx = 0 \tag{3.11}$$

almost everywhere with respect to $t_1, \dots, t_{2n+2m-2r_1-2r_2-2}$.

Assume that $1 \le i \le n - r_1$ and $1 \le j \le m - r_2$ (the other cases can be dealt with in the same way). Then, we have

$$\begin{aligned} &\int_T (f \otimes_{r_1} f)(t_{\sigma(1)}, \dots, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, \dots, t_{\sigma(2n-2r_1-1)}) \\ &\quad \times (g \otimes_{r_2} g)(t_{\tau(1)}, \dots, t_{\tau(j-1)}, x, t_{\tau(j+1)}, \dots, t_{\tau(2m-2r_2-1)})\, dx \\ &= \int_T \int_{T^{r_1}} du_1 \cdots du_{r_1} f(u_1, \dots, u_{r_1}, x, t_{\sigma(1)}, \dots, t_{\sigma(n-r_1-1)}) \\ &\quad \times f(t_{\sigma(n-r_1)}, \dots, f(t_{\sigma(2n-2r_1-1)}) \\ &\quad \times \int_{T^{r_2}} dv_1 \cdots dv_{r_2} g(v_1, \dots, v_{r_2}, x, t_{\tau(1)}, \dots, t_{\tau(m-r_2-1)}) \\ &\quad \times g(t_{\tau(m-r_2)}, \dots, t_{\tau(2m-2r_2-1)})\, dx. \end{aligned}$$

Now, for almost every $u_1, \dots, u_{r_1}, v_1, \dots, v_{r_2}, t_{\sigma(1)}, \dots, t_{\sigma(n-r_1-1)}, t_{\tau(1)}, \dots, t_{\tau(m-r_2-1)}$, (3.8) implies that

$$\int_T f(u_1, \dots, u_{r_1}, x, t_{\sigma(1)}, \dots, t_{\sigma(n-r_1-1)}) g(v_1, \dots, v_{r_2}, x, t_{\tau(1)}, \dots, t_{\tau(m-r_2-1)})\, dx = 0,$$

which implies (3.11) and in turn (3.9). □

The following lemma is the statement of Bourguin and Tudor [2], Lemma 2.

Lemma 3.2. *Let X, Y be centered, strongly independent random variables in $\mathbb{D}^{1,2}$. Then*

$$\langle DX, -DL^{-1}Y\rangle_{\mathfrak{H}} = \langle DY, -DL^{-1}X\rangle_{\mathfrak{H}} = 0.$$

We prove another consequence of strong independence needed later in the paper.

Lemma 3.3. *Let X, Y be strongly independent random variables in $\mathbb{D}^{1,2}$.*

(i) *The random variables $\langle DX, -DL^{-1}X\rangle_{\mathfrak{H}}$ and $\langle DY, -DL^{-1}Y\rangle_{\mathfrak{H}}$ are strongly independent.*
(ii) *The random variables X and $\langle DY, -DL^{-1}Y\rangle_{\mathfrak{H}}$ are strongly independent.*

Proof. Let us prove (i) (the proof of (ii) follows in a similar way by the same arguments, and an analogous result has been proved in Bourguin and Tudor [2], Lemma 1). Assume

$$X = \sum_{n=0}^{\infty} I_n(f_n) \quad \text{and} \quad Y = \sum_{m=0}^{\infty} I_m(g_m),$$

where $f_n \in \mathfrak{H}^{\odot n}$ and $g_m \in \mathfrak{H}^{\odot m}$ for every $n, m \geq 0$. Then, we have

$$D_\theta X = \sum_{n\geq 1} n I_{n-1}\big(f_n(\cdot,\theta)\big) \quad \text{and} \quad -D_\theta L^{-1} X = \sum_{n\geq 1} I_{n-1}\big(f_n(\cdot,\theta)\big),$$

where $I_{n-1}(f_n(\cdot,\theta))$ denotes the multiple Wiener integral of the function

$$(t_1,\dots,t_{n-1}) \mapsto f_n(t_1,\dots,t_{n-1},\theta).$$

Then, it holds that

$$\begin{aligned}\big\langle DX, -DL^{-1}X\big\rangle_{\mathfrak{H}} &= \sum_{n_1,n_2=1}^{\infty} n_1 \int_T I_{n_1-1}\big(f_{n_1}(\cdot,x)\big) I_{n_2-1}\big(f_{n_2}(\cdot,x)\big)\,dx \\ &= \sum_{n_1,n_2=1}^{\infty} n_1 \sum_{r=0}^{n_1\wedge n_2-1} \binom{n_1}{r}\binom{n_2}{r} I_{n_1+n_2-2r-2}(f_{n_1}\tilde{\otimes}_{r+1} f_{n_2}).\end{aligned}$$

Similarly,

$$\big\langle DY, -DL^{-1}Y\big\rangle_{\mathfrak{H}} = \sum_{m_1,m_2=1}^{\infty} m_1 \sum_{r=0}^{m_1\wedge m_2-1} \binom{m_1}{r}\binom{m_2}{r} I_{m_1+m_2-2r-2}(g_{m_1}\tilde{\otimes}_{r+1} g_{m_2}).$$

The conclusion is obtained if we prove that for every $0 \leq r_1 \leq n_1 \wedge n_2 - 1$ and for every $0 \leq r_2 \leq m_1 \wedge m_2 - 1$, the random variables $I_{n_1+n_2-2r-2}(f_{n_1}\tilde{\otimes}_{r_1+1} f_{n_2})$ and $I_{m_1+m_2-2r-2}(g_{m_1}\tilde{\otimes}_{r_2+1} g_{m_2})$ are independent, or equivalently, that

$$(f_{n_1}\tilde{\otimes}_{r_1+1} f_{n_2}) \otimes_1 (g_{m_1}\tilde{\otimes}_{r_2+1} g_{m_2}) = 0 \quad \text{a.e.} \tag{3.12}$$

Since for every $n, m \geq 0$, we have $f_n \otimes_1 g_m = 0$ almost everywhere on T^{m+n-2}, (3.12) follows from the proof of Lemma 3.1. □

Let us illustrate what the above results on strong independence imply about the entries of the matrix $\mathcal{X}_{n,d}$. We begin by introducing some notation. For every $1 \leq i \leq n$ and for every $1 \leq k, l \leq d$, we define

$$F_{ikl} = \big\langle D\big(X_{ik}^2 - 1\big), -DL^{-1}\big(X_{il}^2 - 1\big)\big\rangle_{\mathfrak{H}}. \tag{3.13}$$

We then have the following lemma.

Lemma 3.4. *Let the above notation prevail.*

(i) *If $k \neq l$, $F_{ikl} = 0$ almost surely.*
(ii) *For every $k, l = 1, \dots, d$ with $k \neq l$, the random variables F_{ikk} and F_{ill} are independent.*

Proof. By Lemma 3.1, X_{ik}^2 and X_{il}^2 are strongly independent random variables. Lemma 3.2 yields (i), and Lemma 4 implies (ii). □

3.2. Proof of Theorem 1.1

This subsection is dedicated to the proof of Theorem 1.1. We restate it here for convenience.

Theorem 1. *Consider the renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d}$ with entries given by (3.6). Then for every $n \geq 1$, $\widetilde{\mathcal{W}}_{n,d}$ converges in distribution componentwise, as $d \to \infty$, to the matrix $\mathcal{Z}_n$ given by (3.7). Moreover, there exists a positive constant C such that for every $n, d \geq 1$,*

$$d_W(\widetilde{\mathcal{W}}_{n,d}, \mathcal{Z}_n) \leq C\sqrt{\frac{n^3}{d}}.$$

Proof. Lemma 2.1 combined with Theorem 3.1 implies that we need to estimate the quantity

$$\mathbb{E}\big(\big(\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ab}\rangle_{\mathfrak{H}} - \mathbb{E}(Z_{ij}Z_{ab})\big)^2\big)$$

for every $1 \leq i, j, a, b \leq n$ with $i \leq j$ and $a \leq b$, and Z_{ij} as in (3.7). Note that $\mathbb{E}(Z_{ii}^2) = m_4 - 1$, $\mathbb{E}(Z_{ij}^2) = 1$ if $i \neq j$, and $\mathbb{E}(Z_{ij}Z_{ab}) = 0$ if $(i, j) \neq (a, b)$.

Step 1: calculation of $\mathbb{E}((\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}} - (m_4 - 1))^2)$.

By (3.4) and the strong independence proved in Lemma 3.4, for every $1 \leq i \leq n$, it holds that

$$\begin{aligned}
\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}} &= \frac{1}{d}\sum_{k,l=1}^{d}\langle D(X_{ik}^2 - 1), -DL^{-1}(X_{il}^2 - 1)\rangle_{\mathfrak{H}} \\
&= \frac{1}{d}\sum_{k=1}^{d}\langle D(X_{ik}^2 - 1), -DL^{-1}(X_{ik}^2 - 1)\rangle_{\mathfrak{H}} = \frac{1}{d}\sum_{k=1}^{d} F_{ikk},
\end{aligned}$$

where F_{ikk} is given by (3.13). Since for every $G \in \mathbb{D}^{1,2}$, $\mathbb{E}(G^2) = \mathbb{E}(\langle DG, -DL^{-1}G\rangle_{\mathfrak{H}})$, we can write, using (3.2),

$$\begin{aligned}
\mathbb{E}(\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}}) &= \frac{1}{d}\sum_{k=1}^{d}\mathbb{E}(\langle D(X_{ik}^2 - 1), -DL^{-1}(X_{ik}^2 - 1)\rangle_{\mathfrak{H}}) \\
&= \frac{1}{d}\sum_{k=1}^{d}\mathbb{E}\big((X_{ik}^2 - 1)^2\big) = m_4 - 1.
\end{aligned}$$

Hence, we can write

$$\begin{aligned}
&\mathbb{E}\big(\big(\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}} - (m_4 - 1)\big)^2\big) \\
&\quad = \mathbb{E}\big(\big(\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}} - \mathbb{E}(\langle D\widetilde{W}_{ii}, -DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}})\big)^2\big) \\
&\quad = \frac{1}{d^2}\mathbb{E}\left(\left(\sum_{k=1}^{d}\big(F_{ikk} - \mathbb{E}(F_{ikk})\big)\right)^2\right) \\
&\quad = \frac{1}{d^2}\sum_{k=1}^{d}\mathbb{E}\big(\big(F_{ikk} - \mathbb{E}(F_{ikk})\big)^2\big). \qquad (3.14)
\end{aligned}$$

We claim that for every $1 \le i \le n$ and $1 \le k \le d$,

$$\mathbb{E}\big(\big(F_{ikk} - \mathbb{E}(F_{ikk})\big)^2\big) \le C(i),$$

where $C(i) > 0$ is a constant depending on i, but not on k. In order to prove this, we will make use of the Wiener chaos decomposition of F_{ikk}, together with (3.8) and assumption (3.2). From (3.1) and the product formula (2.4), for every $1 \le i \le n$ and $1 \le k \le d$, it holds that

$$X_{ik}^2 = \sum_{r=0}^{q_i} r! \binom{q_i}{r}^2 I_{2q_i-2r}(f_{ik} \otimes_r f_{ik}),$$

$$D_\theta\big(X_{ik}^2 - 1\big) = \sum_{r=0}^{q_i-1} r! \binom{q_i}{r}^2 (2q_i - 2r) I_{2q_i-2r-1}\big((f_{ik} \otimes_r f_{ik})(\cdot,\theta)\big),$$

and

$$-D_\theta L^{-1}\big(X_{ik}^2 - 1\big) = \sum_{r=0}^{q_i-1} r! \binom{q_i}{r}^2 I_{2q_i-2r-1}\big((f_{ik} \otimes_r f_{ik})(\cdot,\theta)\big).$$

This yields, for every $1 \le i \le n$ and $1 \le k \le d$,

$$\begin{aligned}
F_{ikk} &= \big\langle D\big(X_{ik}^2 - 1\big), -DL^{-1}\big(X_{ik}^2 - 1\big)\big\rangle_{\mathfrak{H}} \\
&= \sum_{r_1,r_2=0}^{q_i-1} r_1! r_2! \binom{q_i}{r_1}^2 \binom{q_i}{r_2}^2 (2q_i - 2r_1) \\
&\quad \times \big\langle I_{2q_i-2r_1-1}(f_{ik} \otimes_{r_1} f_{ik}), I_{2q_i-2r_2-1}(f_{ik} \otimes_{r_2} f_{ik})\big\rangle_{\mathfrak{H}} \\
&= \sum_{r_1,r_2=0}^{q_i-1} r_1! r_2! \binom{q_i}{r_1}^2 \binom{q_i}{r_2}^2 (2q_i - 2r_1) \\
&\quad \times \sum_{p=0}^{(2q_i-2r_1)\wedge(2q_i-2r_2)-1} p! \binom{2q_i - 2r_1 - 1}{p} \binom{2q_i - 2r_2 - 1}{p} \\
&\quad \times I_{4q_i-2r_1-2r_2-2(p+1)}\big((f_{ik} \otimes_{r_1} f_{ik}) \otimes_{p+1} (f_{ik} \otimes_{r_2} f_{ik})\big),
\end{aligned}$$

and hence

$$\begin{aligned}
&F_{ikk} - \mathbb{E}(F_{ikk}) \\
&\quad = \sum_{r_1,r_2=0}^{q_i-1} \mathbb{1}_{\{r_1 \ne r_2\}} r_1! r_2! \binom{q_i}{r_1}^2 \binom{q_i}{r_2}^2 (2q_i - 2r_1) \\
&\qquad \times \big\langle I_{2q_i-2r_1-1}(f_{ik} \otimes_{r_1} f_{ik}), I_{2q_i-2r_2-1}(f_{ik} \otimes_{r_2} f_{ik})\big\rangle_{\mathfrak{H}} \\
&\quad = \sum_{r_1=0}^{q_i-1} \sum_{r_2=0}^{q_i-1} r_1! r_2! \binom{q_i}{r_1}^2 \binom{q_i}{r_2}^2 (2q_i - 2r_1)
\end{aligned}$$

$$
\begin{aligned}
&\times \sum_{p=0}^{(2q_i-2r_1)\wedge(2q_i-2r_2)-1} p!\binom{2q_i-2r_1-1}{p}\binom{2q_i-2r_2-1}{p} \\
&\times I_{4q_i-2r_1-2r_2-2(p+1)}\big((f_{ik}\otimes_{r_1} f_{ik})\otimes_{p+1}(f_{ik}\otimes_{r_2} f_{ik})\big) \\
&+\sum_{r=0}^{q_i-1} r!^2\binom{q_i}{r}^4(2q_i-2r)\sum_{p=0}^{2q_i-2r-2} p!\binom{2q_i-2r-1}{p}^2 \\
&\times I_{4q_i-4r-2(p+1)}\big((f_{ik}\otimes_r f_{ik})\otimes_{p+1}(f_{ik}\otimes_r f_{ik})\big).
\end{aligned}
\tag{3.15}
$$

Now, using the isometry property (2.3) of multiple Wiener integrals together with the bounds $\|\tilde{f}\|_{\mathfrak{H}^{\otimes n}} \le \|f\|_{\mathfrak{H}^{\otimes n}}$ and $\|f\otimes_r g\|_{\mathfrak{H}^{\otimes(2n-2r)}} \le \|f\|_{\mathfrak{H}^{\otimes n}}\|g\|_{\mathfrak{H}^{\otimes n}}$ for every $f,g\in\mathfrak{H}^{\otimes n}$ and $0\le r\le n$, we can write

$$
\begin{aligned}
&\mathbb{E}\Big(I_{4q_i-2r_1-2r_2-2(p+1)}\big((f_{ik}\otimes_{r_1} f_{ik})\otimes_{p+1}(f_{ik}\otimes_{r_2} f_{ik})\big)^2\Big) \\
&\quad = c(q_i,r_1,r_2,p)\big\|(f_{ik}\tilde{\otimes}_{r_1} f_{ik})\tilde{\otimes}_{p+1}(f_{ik}\tilde{\otimes}_{r_2} f_{ik})\big\|^2_{\mathfrak{H}^{\otimes(4q_i-2r_1-2r_2-2(p+1))}} \\
&\quad \le c(q_i,r_1,r_2,p)\|f_{ik}\tilde{\otimes}_{r_1} f_{ik}\|^2_{\mathfrak{H}^{\otimes(2q_i-2r_1)}}\|f_{ik}\tilde{\otimes}_{r_2} f_{ik}\|^2_{\mathfrak{H}^{\otimes(2q_i-2r_2)}} \\
&\quad \le c(q_i,r_1,r_2,p)\|f_{ik}\|^8_{\mathfrak{H}^{\otimes q_i}} \\
&\quad \le c(q_i,r_1,r_2,p),
\end{aligned}
\tag{3.16}
$$

where $c(q_i,r_1,r_2,p)$ is a strictly positive constant depending on q_i, r_1, r_2, p but not on k. Now, in (3.15), we use the isometry property (2.3) together with (3.16) to obtain, for every $1\le i\le n$ and for every $1\le k\le d$,

$$
\mathbb{E}\big(\big(F_{ikk}-\mathbb{E}(F_{ikk})\big)^2\big)\le C(i),
$$

where $C(i)>0$ is a constant (depending only on q_i). Therefore, using the above inequality and (3.14) yields

$$
\mathbb{E}\Big(\big(\langle D\widetilde{W}_{ii},-DL^{-1}\widetilde{W}_{ii}\rangle_{\mathfrak{H}}-(m_4-1)\big)^2\Big)=\frac{1}{d^2}\sum_{k=1}^{d}\mathbb{E}\big(\big(F_{ikk}-\mathbb{E}(F_{ikk})\big)^2\big)\le\frac{C(i)}{d}.
\tag{3.17}
$$

Step 2: calculation of $\mathbb{E}((\langle D\widetilde{W}_{ij},-DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}}-1)^2)$ with $i<j$.

Assume $1\le i<j\le n$. In this case, by (3.1), the product formula (2.4) as well as (3.8), we have for every $1\le k\le d$,

$$
X_{ik}X_{jk}=I_{q_i+q_j}(f_{ik}\otimes f_{jk}),
$$

so that $X_{ik}X_{jk}$ is an element of the (q_i+q_j)-th Wiener chaos. Consequently,

$$
-DL^{-1}(X_{ik}X_{jk})=\frac{1}{q_i+q_j}D(X_{ik}X_{jk})
$$

and

$$
\langle D\widetilde{W}_{ij},-DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}}=\frac{1}{d(q_i+q_j)}\sum_{k,l=1}^{d}\langle D(X_{ik}X_{jl}),-DL^{-1}(X_{ik}X_{jk})\rangle_{\mathfrak{H}}
$$

$$= \frac{1}{d(q_i + q_j)} \sum_{k=1}^{d} \|D(X_{ik}X_{jk})\|_{\mathfrak{H}}^2$$

$$= \frac{1}{d(q_i + q_j)} \sum_{k=1}^{d} \big(X_{ik}^2 \|DX_{jk}\|_{\mathfrak{H}}^2 + X_{jk}^2 \|DX_{ik}\|_{\mathfrak{H}}^2\big).$$

On the other hand, since $\mathbb{E}(\|DX_{ik}\|_{\mathfrak{H}}^2) = q_i$ and $\mathbb{E}(\|DX_{jk}\|_{\mathfrak{H}}^2) = q_j$, we have

$$\mathbb{E}\big(\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}}\big) = \frac{1}{d(q_i + q_j)} \sum_{k=1}^{d} \big(\mathbb{E}\big(X_{ik}^2\big)\mathbb{E}\big(\|DX_{jk}\|_{\mathfrak{H}}^2\big) + \mathbb{E}\big(X_{jk}^2\big)\mathbb{E}\big(\|DX_{ik}\|_{\mathfrak{H}}^2\big)\big) = 1$$

and thus, writing $1 = \frac{q_i}{q_i + q_j} + \frac{q_j}{q_i + q_j}$,

$$\big|\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}} - 1\big| \le \frac{1}{d(q_i + q_j)} \left|\sum_{k=1}^{d} \big(X_{ik}^2 \|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)\right| + \frac{1}{d(q_i + q_j)} \left|\sum_{k=1}^{d} \big(X_{jk}^2 \|DX_{ik}\|_{\mathfrak{H}}^2 - q_i\big)\right|$$

and

$$\mathbb{E}\big(\big|\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}} - 1\big|^2\big) \le \frac{2}{d^2(q_i + q_j)^2} \mathbb{E}\left(\left|\sum_{k=1}^{d} \big(X_{ik}^2 \|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)\right|^2\right) + \frac{2}{d^2(q_i + q_j)^2} \mathbb{E}\left(\left|\sum_{k=1}^{d} \big(X_{jk}^2 \|DX_{ik}\|_{\mathfrak{H}}^2 - q_i\big)\right|^2\right). \tag{3.18}$$

The two summands above can be estimated in a similar way, so we only cover the first one. By the independence of the entries and Bourguin and Tudor [2], Lemma 1, we have

$$\begin{aligned}
\mathbb{E}\left(\left|\sum_{k=1}^{d} \big(X_{ik}^2 \|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)\right|^2\right) &\le 2\mathbb{E}\left(\left|\sum_{k=1}^{d} X_{ik}^2 \big(\|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)\right|^2\right) \\
&\quad + 2\mathbb{E}\left(\left|\sum_{k=1}^{d} q_j \big(X_{ik}^2 - 1\big)\right|^2\right) \\
&= 2\sum_{k=1}^{d} \mathbb{E}\big(\big[X_{ik}^2\big(\|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)\big]^2\big) \\
&\quad + 2\sum_{k=1}^{d} q_j^2 \mathbb{E}\big(\big(X_{ik}^2 - 1\big)^2\big)
\end{aligned}$$

$$= 2\sum_{k=1}^{d} \mathbb{E}\big(\big(\|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)^2\big)m_4$$
$$+ 2d(m_4 - 1). \tag{3.19}$$

Writing

$$\|DX_{jk}\|_{\mathfrak{H}}^2 - q_j = q_j^2 \sum_{r=0}^{q_j-2} r! \binom{q_j - 1}{r}^2 I_{2q_j-2r-2}(f_{ik}\tilde{\otimes}_{r+1} f_{ik})$$

and estimating the L^2-norm as in the proof of (3.16) yields

$$\mathbb{E}\big(\big(\|DX_{jk}\|_{\mathfrak{H}}^2 - q_j\big)^2\big) \le C(j), \tag{3.20}$$

where $C(j)$ is a constant depending solely on q_j. By (3.18), (3.19) and (3.20), we get

$$\mathbb{E}\big(\big(\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ij}\rangle_{\mathfrak{H}} - 1\big)^2\big) \le \frac{C(i,j)}{d}, \tag{3.21}$$

where $C(i,j)$ is a positive constant depending only on q_i and q_j.

Step 3: calculation of $\mathbb{E}((\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ab}\rangle_{\mathfrak{H}})^2)$ *with* $(i,j) \ne (a,b)$.

Let $1 \le i, j, a, b \le n$ with $i \le j$, $a \le b$ and $(i,j) \ne (a,b)$. If i, j, a, b are all distinct, then we have

$$\begin{aligned}
\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ab}\rangle_{\mathfrak{H}} &= \frac{1}{dq_a}\sum_{k,l=1}^{d} \langle D(X_{ik}X_{jk}), D(X_{al}X_{bl})\rangle_{\mathfrak{H}} \\
&= \frac{1}{dq_a}\sum_{k,l=1}^{d} \langle X_{ik}DX_{jk} + X_{jk}DX_{ik}, X_{al}DX_{bl} + X_{bl}DX_{al}\rangle_{\mathfrak{H}} \\
&= \frac{1}{dq_a}\sum_{k,l=1}^{d} \big(X_{ik}X_{al}\langle DX_{jk}, DX_{bl}\rangle_{\mathfrak{H}} + X_{ik}X_{bl}\langle DX_{jk}, DX_{al}\rangle_{\mathfrak{H}} \\
&\quad + X_{jk}X_{al}\langle DX_{ik}, DX_{bl}\rangle_{\mathfrak{H}} \\
&\quad + X_{jk}X_{bl}\langle DX_{ik}, DX_{al}\rangle_{\mathfrak{H}}\big) = 0,
\end{aligned} \tag{3.22}$$

since all the scalar products vanish according to Lemma 3.2.

The remaining cases, namely $\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ib}\rangle_{\mathfrak{H}}$ with $j \ne b$ and $\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{aj}\rangle_{\mathfrak{H}}$ with $i \ne a$ can all be dealt with in a similar manner. For instance, if $j \ne b$, assuming $i < j$ and $i < b$, we can write

$$\begin{aligned}
\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ib}\rangle_{\mathfrak{H}} &= \frac{1}{2dq_i}\sum_{k=1}^{d} \langle D(X_{ik}X_{jk}), D(X_{ik}X_{bk})\rangle_{\mathfrak{H}} \\
&= \frac{1}{2dq_i}\sum_{k=1}^{d} X_{jk}X_{bk}\|DX_{ik}\|_{\mathfrak{H}}^2.
\end{aligned}$$

Similarly, we also have

$$\mathbb{E}\Big(\big(\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{aj}\rangle_{\mathfrak{H}}\big)^2\Big) = \frac{C(i)}{d^2}\sum_{k=1}^{d}\mathbb{E}\big(\|DX_{ik}\|_{\mathfrak{H}}^4\big) \leq \frac{C(i)}{d}, \tag{3.23}$$

where the above equality and inequality are derived similarly as for what was done for the bound appearing in (3.20).

An application of Lemma 2.1 together with Theorem 3.1 yields

$$d_W(\widetilde{\mathcal{W}}_{n,d}, \mathcal{Z}_n) \leq \sqrt{2}C\sqrt{\sum_{i,j,a,b=1}^{n}\mathbb{E}\Big(\big(\langle D\widetilde{W}_{ij}, -DL^{-1}\widetilde{W}_{ab}\rangle_{\mathfrak{H}} - \mathbb{E}(Z_{ij}Z_{ab})\big)^2\Big)},$$

where $C > 0$ is the constant appearing in Theorem 3.1. Since for a, b, i, j all distinct, the corresponding above summands vanish according to (3.22), we have only $n^4 - n(n-1)(n-2)(n-3) \leq 6n^3$ summands. By (3.17), (3.21) and (3.23), all these non-vanishing summands are bounded by $\frac{C}{d}$, where $C > 0$ denotes a generic constant resulting from the aggregation of the $C(i)$ and $C(i, j)$ constants appearing in the previous steps of the proof. This yields (1.2) and concludes the proof. □

4. Random matrices with correlated second chaos entries

In this section, we consider the case where the entries of the matrix $\mathcal{X}_{n,d}$ are allowed to be correlated. As in the previous section, let $\mathcal{X}_{n,d} = (X_{ij})_{1\leq i\leq n, 1\leq j\leq d}$ be a $n \times d$ random matrix whose entries are given by the increments of a Rosenblatt process, which lives in the second Wiener chaos. The choice of dealing with the second chaos in the case of correlated entries comes both from the accrued importance of the second chaos in applications, as well as from technical considerations of keeping the involved combinatorics at a reasonable level for our exposition. The Rosenblatt process $(Z_t^H)_{t\geq 0}$ with self-similarity parameter $H \in (\frac{1}{2}, 1)$ is defined by, for every $t \geq 0$,

$$Z_t^H = I_2(L_t), \tag{4.1}$$

where I_2 denote the multiple Wiener integral of order two with respect to a Brownian motion $(B_t)_{t\in\mathbb{R}_+}$ and the kernel L_t is given by, for every $y_1, y_2 \in \mathbb{R}$ and $t \geq 0$,

$$L_t(y_1, y_2) = d(H)\mathbb{1}_{[0,t]^2}(y_1, y_2)\int_{y_1\vee y_2}^{t}\partial_1 K^{\frac{H+1}{2}}(u, y_1)\partial_1 K^{\frac{H+1}{2}}(u, y_1)\, du, \tag{4.2}$$

where

$$d(H) = \frac{1}{H+1}\sqrt{\frac{2(2H-1)}{H}} \tag{4.3}$$

and for $t > s$,

$$K^H(t, s) = c(H)s^{\frac{1}{2}-H}\int_s^t (u-s)^{H-\frac{3}{2}}u^{H-\frac{1}{2}}\, du,$$

with $c(H) = \sqrt{(\frac{H(2H-1)}{\beta(2-2H, H-\frac{1}{2})})}$, where β denotes the beta function (see e.g. Nualart [14]).

The kernel L_t belongs to $L^2(\mathbb{R}_+^2)$ for every $t \geq 0$. The Rosenblatt process Z^H is H-self similar, has stationary increments and long memory. We refer to the monographs Pipiras and Taqqu [15] or Tudor [18] for its basic properties. In particular, it has the same covariance as the fractional Brownian motion, that is, for any $s, t \geq 0$,

$$\mathbb{E}(Z_t^H Z_s^H) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

A random variable with the same distribution as Z_1^H will be called a *Rosenblatt random variable*.

Let us now define the entries of the matrix $\mathcal{X}_{n,d}$. Let $B = (B^1, \dots, B^n)$ denote a d-dimensional Brownian motion and define

$$Z_t^{H,i} = I_2^i(L_t), \tag{4.4}$$

where I_q^i denotes the multiple Wiener integral of order q with respect to the Brownian motion B^i for any $1 \leq i \leq n$. Then, by (4.1), the processes $(Z_t^{H,i})_{t\geq 0}$, $1 \leq i \leq n$ are independent Rosenblatt processes with the same Hurst parameter (or self-similarity parameter) $H \in (\frac{1}{2}, 1)$. For any $i \geq 1$, denote $f_i = L_i - L_{i-1}$.

Consider the random matrix $\mathcal{X}_{n,d} = (X_{ij})_{1\leq i\leq n, 1\leq j\leq d}$ with entries given by

$$X_{ij} = I_2^i(f_j) = Z_j^{H,i} - Z_{j-1}^{H,i} \tag{4.5}$$

for every $1 \leq i \leq n$ and $1 \leq j \leq d$, with $Z^{H,i}$ given by (4.4). This means that all the entries have the same distribution, the ones on different columns are independent and those on the same rows are correlated according to the correlation structure of the increments of the Rosenblatt process. Since the covariance of the Rosenblatt process coincides with that of the fractional Brownian motion, the correlation structure of our matrix is the same as in Nourdin and Zheng [13] (where the entries are given by the increments of the fractional Brownian motion). Despite this fact, the non-Gaussian character will yield a different limiting behavior of the associated Wishart matrix.

More precisely, we have for every $1 \leq 1, k \leq n$ and $1 \leq j, l \leq d$,

$$\mathbb{E}(X_{ij} X_{kl}) = \mathbb{1}_{\{i=k\}} \rho_H(j-l),$$

where ρ_H denotes the correlation function of the Rosenblatt process (or that of the fractional Brownian motion) given by, for $k \in \mathbb{Z}$,

$$\rho_H(k) = \frac{1}{2}\left(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}\right). \tag{4.6}$$

In particular, for $1 \leq i \leq n$ and $1 \leq j \leq d$,

$$\mathbb{E}(X_{ij}^2) = 2!\langle f_i, f_j\rangle_{L^2(\mathbb{R}_+^2)} = 1.$$

4.1. Rosenblatt limiting distribution

Consider the Wishart matrix $\mathcal{W}_{n,d}$ obtained from $\mathcal{X}_{n,d}$ as in (3.3), where $\mathcal{X}_{n,d}$ is now given by (4.5). Recall that the entries of the Wishart matrix are given by (3.4) and (3.5). We start by analyzing the asymptotic behavior in distribution of each element of the Wishart matrix. This will be related to the limiting behavior of the quadratic variations of the Rosenblatt process. Consider the constant $c_{1,H}$ given by

$$c_{1,H} = 4d(H), \tag{4.7}$$

with $d(H)$ given by (4.3). Let us recall the following result from Tudor and Viens [19].

Theorem 4.1. *Let $(Z_t^H)_{t\geq 0}$ be a Rosenblatt process. Define, for $d \geq 1$,*

$$V_d = c_{1,H}^{-1} d^{-H} \sum_{k=0}^{d-1} \left[\frac{(Z_{\frac{k+1}{d}}^H - Z_{\frac{k}{d}}^H)^2}{d^{-2H}} - 1 \right]. \tag{4.8}$$

Then, the sequence $(V_d)_{d\geq 1}$ converges in $L^2(\Omega)$, as $d \to \infty$, to the Rosenblatt random variable Z_1^H.

Let us first study the limiting behavior, as $d \to \infty$, of the diagonal terms of the Wishart matrix $\mathcal{W}_{n,d}$.

Proposition 4.1. *For $1 \leq i \leq n$, let W_{ii} be given by (3.4), and let*

$$\widetilde{W}_{ii} = c_{1,H}^{-1} d^{1-H} W_{ii},$$

where $c_{1,H}$ is the constant defined in (4.7). Then, for every $1 \leq i \leq n$,

$$\widetilde{W}_{ii} \to Z_1^{H,i}$$

in $L^2(\Omega)$ as $d \to \infty$.

Proof. By the scaling property of the Rosenblatt process and (4.5), we have, for every $1 \leq i \leq n$,

$$\begin{aligned} W_{ii} &= \frac{1}{d} \sum_{k=1}^{d} \bigl(X_{ik}^2 - 1\bigr) = \frac{1}{d} \sum_{k=1}^{d} \bigl((Z_{k+1}^{H,i} - Z_k^{H,i})^2 - 1\bigr) \\ &\overset{\mathcal{D}}{=} \frac{1}{d} \sum_{k=0}^{d-1} \left(\frac{(Z_{\frac{k+1}{d}}^{H,i} - Z_{\frac{k}{d}}^{H,i})^2}{d^{-2H}} - 1 \right) = c_{1,H} d^{H-1} V_d^i, \end{aligned}$$

where $\overset{\mathcal{D}}{=}$ denotes equality in distribution, and for $1 \leq i \leq n$,

$$V_d^i = c_{1,H}^{-1} d^{-H} \sum_{k=0}^{d-1} \left(\frac{(Z_{\frac{k+1}{d}}^{H,i} - Z_{\frac{k}{d}}^{H,i})^2}{d^{-2H}} - 1 \right). \tag{4.9}$$

The conclusion follows from Theorem 4.1. □

As far as the convergence of the non-diagonal terms of the Wishart matrix (3.3), we have the following result. It shows that the square mean of the non-diagonal terms of the renormalized Wishart matrix is dominated by the square mean of the diagonal terms. Intuitively, this happens because the mean square of the non-diagonal terms involves the increments of two independent Rosenblatt processes.

Proposition 4.2. *For $1 \leq i, j \leq n$ with $i \neq j$, let W_{ij} be given by (3.5), and define*

$$\widetilde{W}_{ij} = c_{1,H}^{-1} d^{1-H} W_{ij}, \tag{4.10}$$

where $c_{1,H}$ denotes the constant defined in (4.7). Then, for every $1 \leq i, j \leq n$,

$$\widetilde{W}_{i,j} \to 0$$

in $L^2(\Omega)$ *as* $d \to \infty$, *and*

$$\mathbb{E}\big(\widetilde{W}_{ij}^2\big) \le C \begin{cases} d^{1-2H} & \textit{if } H \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ \log(d) d^{-\frac{1}{2}} & \textit{if } H = \frac{3}{4}, \\ d^{2H-2} & \textit{if } H \in \left(\frac{3}{4}, 1\right), \end{cases} \tag{4.11}$$

where $C > 0$ *denotes a generic constant.*

Proof. By self-similarity and (4.5),

$$W_{ij} = \frac{1}{d} \sum_{k=1}^{d} X_{ik} X_{jk} = \frac{1}{d} \sum_{k=0}^{d-1} \big(Z_{k+1}^{H,i} - Z_k^{H,i}\big)\big(Z_{k+1}^{H,j} - Z_k^{H,j}\big),$$

so that, for every $1 \le i, j \le n$,

$$\begin{aligned} \mathbb{E}\big(\widetilde{W}_{ij}^2\big) &= c_{1,H}^{-2} d^{-2H} \mathbb{E}\left(\left(\sum_{k=0}^{d-1} \big(Z_{k+1}^{H,i} - Z_k^{H,i}\big)\big(Z_{k+1}^{H,j} - Z_k^{H,j}\big)\right)^2\right) \\ &= c_{1,H}^{-2} d^{-2H} \sum_{k,l=0}^{d-1} \mathbb{E}\big(\big(Z_{k+1}^{H,i} - Z_k^{H,i}\big)\big(Z_{l+1}^{H,i} - Z_l^{H,i}\big)\big) \\ &\quad \times \mathbb{E}\big(\big(Z_{k+1}^{H,j} - Z_k^{H,j}\big)\big(Z_{l+1}^{H,j} - Z_l^{H,j}\big)\big) \\ &= c_{1,H}^{-2} d^{-2H} \sum_{k,l=0}^{d-1} \rho_H\big(|k-l|\big)^2 \\ &\le c_{1,H}^{-2} d^{1-2H} \sum_{v \in \mathbb{Z}} \rho_H\big(|v|\big)^2 \left(1 - \frac{|v|}{n}\right) 1_{(|v|<n)}, \end{aligned}$$

where ρ_H is given by (4.6). The fact that $\rho_H(|k|)$ behaves as $H(2H-1)|k|^{2H-2}$ as $|k| \to \infty$ concludes the proof. □

4.2. Proof of Theorem 1.2

In this section, we pave the way to the proof of Theorem 1.2 by stating and proving some preparatory results, making use of the results established in the previous subsection to do so. Theorem 1.2 is restated for convenience at the end of the section right before its proof.

Consider the renormalized Wishart matrix $\widetilde{\mathcal{W}}_{n,d}$ defined in (4.10). By Propositions 4.1 and 4.2, its limit in distribution is an $n \times n$ diagonal matrix, denoted by $\mathcal{R}_n^H = (R_{ij}^H)_{1 \le i,j \le n}$, with independent diagonal entries given by, for all $1 \le i \le n$,

$$R_{ii}^H = Z_1^{H,i}. \tag{4.12}$$

Given that what we need is to estimate the Wasserstein distance between $\widetilde{\mathcal{W}}_{n,d}$ and $\mathcal{R}_n^H$, we start with the observation that, due to the scaling property of the Rosenblatt process, we have

$$d_W\big(\widetilde{\mathcal{W}}_{n,d}, \mathcal{R}_n^H\big) = d_W\big(\mathcal{V}_{n,d}, \mathcal{R}_n^H\big),$$

where the matrix $\mathcal{V}_{n,d} = (V_{ij})_{1\le i,j\le n}$ is given by

$$\begin{cases} V_{ii} = V_d^i & \text{for } 1 \le i \le n, \\ V_{ij} = c_{1,H}^{-1} d^H \displaystyle\sum_{k=0}^{d-1} \big(Z^{H,i}_{\frac{k+1}{d}} - Z^{H,i}_{\frac{k}{d}}\big)\big(Z^{H,j}_{\frac{k+1}{d}} - Z^{H,j}_{\frac{k}{d}}\big) & \text{for } 1 \le i \ne j \le n, \end{cases}$$

where V_d^i was defined in (4.9). By the definition of the Wasserstein distance (2.1),

$$d_W\big(\widetilde{\mathcal{W}}_{n,d}, \mathcal{R}_n^H\big) = d_W\big(\mathcal{V}_{n,d}, \mathcal{R}_n^H\big) \le \sqrt{\sum_{i,j=1}^n \mathbb{E}\big(\big(V_{ij} - R_{ij}^H\big)^2\big)} \tag{4.13}$$

with $R_{ij}^H = 0$ if $i \ne j$ and R_{ii}^H given by (4.12).

The estimates for the terms with $i \ne j$ in the right-hand side of (4.13) will follow from Proposition 4.2. The next proposition provides estimates for the diagonal summands of the right-hand side of (4.13).

Proposition 4.3. *Let V_d be given by* (4.8). *Then, it holds that*

$$\mathbb{E}\big(\big|V_d - Z_1^H\big|^2\big) \le C \begin{cases} d^{1-2H} & \text{if } H \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ \log(d) d^{-\frac{1}{2}} & \text{if } H = \frac{3}{4}, \\ d^{2H-2} & \text{if } H \in \left(\frac{3}{4}, 1\right), \end{cases} \tag{4.14}$$

where $C > 0$ denotes a generic constant.

Proof. For $0 \le k \le d-1$, we have

$$Z^H_{\frac{k+1}{d}} - Z^H_{\frac{k}{d}} = I_2(L_{\frac{k+1}{n}} - L_{\frac{k}{n}}),$$

where L is the kernel defined in (4.2). By the product formula for multiple Wiener integrals (2.4), we can decompose V_d as the sum of two terms, one in the fourth Wiener chaos and one in the second Wiener chaos. Namely,

$$\begin{aligned} V_d &= c_{1,H}^{-1} d^H \sum_{k=0}^{d-1} \big[I_4\big((L_{\frac{k+1}{n}} - L_{\frac{k}{n}})^{\otimes 2}\big) + 4 I_2\big((L_{\frac{k+1}{n}} - L_{\frac{k}{n}}) \otimes_1 (L_{\frac{k+1}{n}} - L_{\frac{k}{n}})\big)\big] \\ &= T_{4,d} + T_{2,d}. \end{aligned} \tag{4.15}$$

The estimation of the $L^2(\Omega)$-norm of the term $T_{4,d}$ has been done in Tudor and Viens [19]. This term has no contribution to the limit of V_d and using Tudor and Viens [19], Equations (3.15)–(3.17), yields

$$\mathbb{E}\big(T_{4,d}^2\big) \leq C \begin{cases} d^{1-2H} & \text{if } H \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ \log(d) d^{-\frac{1}{2}} & \text{if } H = \frac{3}{4}, \\ d^{2H-2} & \text{if } H \in \left(\frac{3}{4}, 1\right). \end{cases}$$

The summand $T_{2,d}$ appearing in (4.15) converges in $L^2(\Omega)$ to Z_1^H. This has also been proved in Tudor and Viens [19], but we still need to evaluate the rate of this convergence. We can write $T_{2,d} = I_2(h_d)$, with

$$h_d(y_1, y_2) = 4c_{1,H}^{-1} d^H \sum_{k=0}^{d-1} \big((L_{\frac{k+1}{n}} - L_{\frac{k}{n}}) \otimes_1 (L_{\frac{k+1}{n}} - L_{\frac{k}{n}})\big). \tag{4.16}$$

Hence,

$$\mathbb{E}\big(\big|T_{2,d} - Z_1^H\big|^2\big) = \mathbb{E}\big(|T_{2,d}|^2\big) - 2\mathbb{E}\big(T_{2,d} Z_1^H\big) + \mathbb{E}\big(\big|Z_1^H\big|^2\big). \tag{4.17}$$

On one hand, Tudor and Viens [19], Equation (3.11), yields

$$\begin{aligned} \mathbb{E}\big(|T_{2,d}|^2\big) &= 2\|h_d\|_{L^2(\mathbb{R}_+^2)}^2 \\ &= 2c_{1,H}^{-2} d(H)^4 \big(H(H+1)\big)^4 d^{2H} \sum_{i,j=0}^{d-1} \int_{\frac{i}{d}}^{\frac{i+1}{d}} \int_{\frac{i}{d}}^{\frac{i+1}{d}} \int_{\frac{j}{d}}^{\frac{j+1}{d}} \int_{\frac{j}{d}}^{\frac{j+1}{d}} |u-v|^{H-1} \\ &\quad \times \big|u'-v'\big|^{H-1} \big|u-u'\big|^{H-1} \big|v-v'\big|^{H-1} \, du \, dv \, du' \, dv' \\ &= H(2H-1) e(H) d^{-2H} \sum_{i,j=0}^{d-1} \int_{[0,1]^4} |u-v|^{H-1} \big|u'-v'\big|^{H-1} \\ &\quad \times \big|u-u'+i-j\big|^{H-1} \big|v-v'+i-j\big|^{H-1} \, du \, dv \, du' \, dv', \end{aligned}$$

where $e(H)$ is a constant given by

$$e(H) = \frac{H^2(H+1)^2}{4}. \tag{4.18}$$

On the other hand, using the fact that $2\|L_1\|_{L^2(\mathbb{R}_+^2)}^2 = 1$ yields

$$\begin{aligned} \mathbb{E}\big(\big|Z_1^H\big|^2\big) &= 2\|L_1\|_{L^2(\mathbb{R}_+^2)}^2 \\ &= H(2H-1) \int_0^1 \int_0^1 |u-v|^{2H-2} \, du \, dv \\ &= H(2H-1) \sum_{i,j=0}^{d-1} \int_{\frac{i}{d}}^{\frac{i+1}{d}} \int_{\frac{j}{d}}^{\frac{j+1}{d}} |u-v|^{2H-2} \, du \, dv \end{aligned}$$

$$= H(2H-1)d^{-2H}\sum_{i,j=0}^{d-1}\int_{[0,1]^2}|u-v+i-j|^{2H-2}\,du\,dv$$

$$= 1.$$

Furthermore, note that (4.16) and (4.2) imply

$$\begin{aligned}\mathbb{E}\big(T_{2,d}Z_1^H\big) &= 2\langle h_N, L_1\rangle_{L^2(\mathbb{R}_+^2)}\\ &= H(2H-1)f(H)d^H\sum_{i=0}^{d-1}\int_{\frac{i}{d}}^{\frac{i+1}{d}}\int_{\frac{i}{d}}^{\frac{i+1}{d}}\int_0^1|u-v|^{H-1}\big|u-u'\big|^{H-1}\\ &\quad\times\big|v-u'\big|^{H-1}\,du\,dv\,du'\\ &= H(2H-1)f(H)d^H\sum_{i,j=0}^{d-1}\int_{\frac{i}{d}}^{\frac{i+1}{d}}\int_{\frac{i}{d}}^{\frac{i+1}{d}}\int_{\frac{j}{d}}^{\frac{j+1}{d}}|u-v|^{H-1}\big|u-u'\big|^{H-1}\\ &\quad\times\big|v-u'\big|^{H-1}\,du\,dv\,du'\\ &= H(2H-1)f(H)d^{-2H}\sum_{i,j=0}^{d-1}\int_{[0,1]^3}|u-v|^{H-1}\big|u-u'+i-j\big|^{H-1}\\ &\quad\times\big|v-u'+i-j\big|^{H-1}\,du\,dv\,du',\end{aligned}$$

where $f(H)$ is a constant given by

$$f(H) = \frac{H+1}{2(2H-1)}. \tag{4.19}$$

Now, (4.17) becomes

$$\begin{aligned}&\mathbb{E}\big(\big|T_{2,d}-Z_1^H\big|^2\big)\\ &\quad= H(2H-1)d^{-2H}e(H)\sum_{i,j=0}^{d-1}\bigg[\int_{[0,1]^4}|u-v|^{H-1}\big|u'-v'\big|^{H-1}\big|u-u'+i-j\big|^{H-1}\\ &\qquad\times\big|v-v'+i-j\big|^{H-1}\,du\,dv\,du'\,dv'\\ &\qquad-2f(H)\int_{[0,1]^3}|u-v|^{H-1}\big|u-u'+i-j\big|^{H-1}\big|v-u'+i-j\big|^{H-1}\,du\,dv\,du'\\ &\qquad+\int_{[0,1]^2}|u-v+i-j|^{2H-2}\,du\,dv\bigg]\\ &\quad\le Cd^{1-2H}e(H)\sum_{k\in\mathbb{Z}}\bigg[\int_{[0,1]^4}|u-v|^{H-1}\big|u'-v'\big|^{H-1}\big|u-u'+k\big|^{H-1}\\ &\qquad\times\big|v-v'+k\big|^{H-1}\,du\,dv\,du'\,dv'\end{aligned}$$

$$-2f(H)\int_{[0,1]^3}|u-v|^{H-1}\left|u-u'+k\right|^{H-1}\left|v-u'+k\right|^{H-1}\,du\,dv\,du'$$
$$+\int_{[0,1]^2}|u-v+k|^{2H-2}\,du\,dv\bigg]. \tag{4.20}$$

Now, Tudor and Viens [19], Lemma 5 (see also Clausel et al. [7], Lemma 2), together with the definition of $e(H)$ given in (4.18) yields

$$\int_{[0,1]^4}|u-v|^{H-1}\left|u'-v'\right|^{H-1}\left|u-u'+k\right|^{H-1}\left|v-v'+k\right|^{H-1}\,du\,dv\,du'\,dv'$$
$$=e(H)^{-1}k^{2H-2}+O\big(k^{2H-2}\big). \tag{4.21}$$

Similarly,

$$\int_{[0,1]^3}|u-v|^{H-1}\left|u-u'+k\right|^{H-1}\left|v-u'+k\right|^{H-1}\,du\,dv\,du'=f(H)^{-1}k^{2H-2}+o\big(k^{2H-2}\big), \tag{4.22}$$

where $f(H)$ is given by (4.19). Finally, Breton and Nourdin [3], Proof of Proposition 3.1, yields

$$\int_{[0,1]^2}|u-v+k|^{2H-2}\,du\,dv=k^{2H-2}+o\big(k^{2H-2}\big). \tag{4.23}$$

Combining (4.21), (4.22) and (4.23) implies that the sum over $k\in\mathbb{Z}$ in (4.20) converges. Hence,

$$\mathbb{E}\big(\left|T_{2,d}-Z_1^H\right|^2\big)\le Cd^{1-2H},$$

and since, by (4.15),

$$\mathbb{E}\big(\left|V_d-Z_1^H\right|^2\big)=\mathbb{E}\big(|T_{4,d}|^2\big)+\mathbb{E}\big(\left|T_{2,d}-Z_1^H\right|^2\big),$$

we obtain (4.14). □

We are now ready to provide the proof of Theorem 1.2, which we restate here for convenience.

Theorem 2. *Let $\widetilde{\mathcal{W}}_{n,d}$ be the renormalized Wishart matrix (4.10) and let $\mathcal{R}_n^H$ be the diagonal matrix with entries given by (4.12). Then, for every $n\ge 1$, the random matrix $\widetilde{\mathcal{W}}_{n,d}$ converges componentwise in distribution, as $d\to\infty$, to the matrix $\mathcal{R}_n^H$. Moreover, as $n,d\ge 1$, there exists a positive constant C such that*

$$d_W\big(\widetilde{\mathcal{W}}_{n,d},\mathcal{R}_n^H\big)\le C\begin{cases} nd^{\frac{1}{2}-H} & \text{if } H\in\left(\frac{1}{2},\frac{3}{4}\right),\\ n\sqrt{\log(d)}\,d^{-\frac{1}{4}} & \text{if } H=\frac{3}{4}m\\ nd^{H-1} & \text{if } H\in\left(\frac{3}{4},1\right).\end{cases}$$

Proof of Theorem 1.2. The conclusion follows from combining relation (4.13) with Propositions 4.2 and 4.3. Indeed, the summands with $i\neq j$ in (4.13) have been estimated in Proposition 4.2 (see (4.11)), while the diagonal terms of (4.13) are estimated by (4.14). □

Acknowledgements

S. Bourguin was supported in part by the Simons Foundation grant 635136. C. Tudor was supported in part by the Labex CEMPI (ANR-11-LABX-0007-01) and MATHAMSUD project SARC (19-MATH-06).

References

[1] Bishop, A., Del Moral, P. and Niclas, A. (2018). An introduction to Wishart matrix moments. *MAL* **2** 97–208.

[2] Bourguin, S. and Tudor, C.A. (2011). Cramér theorem for gamma random variables. *Electron*. *Commun*. *Probab*. **16** 365–378. MR2819659 https://doi.org/10.1214/ECP.v16-1639

[3] Breton, J.-C. and Nourdin, I. (2008). Error bounds on the non-normal approximation of Hermite power variations of fractional Brownian motion. *Electron*. *Commun*. *Probab*. **13** 482–493. MR2447835 https://doi.org/10.1214/ECP.v13-1415

[4] Bubeck, S., Ding, J., Eldan, R. and Rácz, M.Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms* **49** 503–532. MR3545825 https://doi.org/10.1002/rsa.20633

[5] Bubeck, S. and Ganguly, S. (2018). Entropic CLT and phase transition in high-dimensional Wishart matrices. *Int*. *Math*. *Res*. *Not*. *IMRN* **2** 588–606. MR3801440 https://doi.org/10.1093/imrn/rnw243

[6] Chronopoulou, A., Tudor, C.A. and Viens, F.G. (2011). Self-similarity parameter estimation and reproduction property for non-Gaussian Hermite processes. *Commun*. *Stoch*. *Anal*. **5** 161–185. MR2808541 https://doi.org/10.31390/cosa.5.1.10

[7] Clausel, M., Roueff, F., Taqqu, M.S. and Tudor, C. (2014). Asymptotic behavior of the quadratic variation of the sum of two Hermite processes of consecutive orders. *Stochastic Process*. *Appl*. **124** 2517–2541. MR3192505 https://doi.org/10.1016/j.spa.2014.02.013

[8] Edelman, A., Guionnet, A. and Péché, S. (2016). Beyond universality in random matrix theory. *Ann*. *Appl*. *Probab*. **26** 1659–1697. MR3513602 https://doi.org/10.1214/15-AAP1129

[9] Jiang, T. and Li, D. (2015). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J*. *Theoret*. *Probab*. **28** 804–847. MR3413957 https://doi.org/10.1007/s10959-013-0519-7

[10] Johnstone, I.M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians*. *Vol*. *I* 307–333. Zürich: Eur. Math. Soc. MR2334195 https://doi.org/10.4171/022-1/13

[11] Marčenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat*. *Sb*. (*N.S.*) **72** 507–536. MR0208649

[12] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. MR2962301 https://doi.org/10.1017/CBO9781139084659

[13] Nourdin, I. and Zheng, G. (2018). Asymptotic behavior of large Gaussian correlated Wishart matrices. Available at arXiv:1804.06220.

[14] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. *Probability and Its Applications* (*New York*). Berlin: Springer. MR2200233

[15] Pipiras, V. and Taqqu, M.S. (2017). *Long-Range Dependence and Self-Similarity*. *Cambridge Series in Statistical and Probabilistic Mathematics* **45**. Cambridge: Cambridge Univ. Press. MR3729426

[16] Rácz, M.Z. and Richey, J. (2019). A smooth transition from Wishart to GOE. *J*. *Theoret*. *Probab*. **32** 898–906. MR3959632 https://doi.org/10.1007/s10959-018-0808-2

[17] Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2514435

[18] Tudor, C.A. (2013). *Analysis of Variations for Self-Similar Processes*: *A Stochastic Calculus Approach*. *Probability and Its Applications* (*New York*). Cham: Springer. MR3112799 https://doi.org/10.1007/978-3-319-00936-0

[19] Tudor, C.A. and Viens, F.G. (2009). Variations and estimators for self-similarity parameters via Malliavin calculus. *Ann*. *Probab*. **37** 2093–2134. MR2573552 https://doi.org/10.1214/09-AOP459

[20] Üstünel, A.S. and Zakai, M. (1989). On independence and conditioning on Wiener space. *Ann. Probab.* **17** 1441–1453. MR1048936

[21] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

Received February 2020 and revised August 2020