ALEA, Lat. Am. J. Probab. Math. Stat. 18, 1625–1641 (2021)
DOI: 10.30757/ALEA.v18-59

Limit behavior for Wishart matrices with Skorohod integrals

Charles-Philippe Diez and Ciprian A. Tudor

CNRS, Université de Lille
Laboratoire Paul Painlevé
UMR 8524F-59655 Villeneuve d'Ascq, France.
E-mail address: charles-philippe.diez@univ-lille.fr , ciprian.tudor@univ-lille.fr

Abstract. We consider a $n \times d$ random matrix $\mathcal{X}_{n,d}$ whoses entries can be expressed as Skorohod integrals. By using the techniques of the Malliavin calculus, we study the fluctuations under the Wasserstein distance, as $n, d \to \infty$, of the renormalized Wishart matrix

$$\mathcal{W}_{n,d} = \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - \mathcal{I}_n\right),$$

where $\mathcal{I}_n$ is the $n \times n$ identity matrix.

1. Introduction

Consider a $n \times d$ random matrix $\mathcal{X}_{n,d} = (X_{i,j})_{1\leq i\leq n, 1\leq j\leq d}$. We can associate to it the so-called (renormalized) Wishart matrix $\mathcal{W}_{n,d} = \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - \mathcal{I}_n\right)$ where $\mathcal{I}_n$ is the $n \times n$ identity matrix and $\mathcal{X}^T$ denotes the transpose of the matrix $\mathcal{X}$. This matrix has been introduced by Wishart in the twenties in Wishart (1928) and it has numerous applications in multivariate analysis or statistics. Of particular interest is to understand its asymptotic behavior when the dimensions d and n are large. Assume that the entries of the starting matrix $\mathcal{X}_{n,d}$ are independent and identically distributed, with zero mean and unit variance. Then it is easy to see that for fixed $n \geq 1$, the entries of the matrix $\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T$ converge, as $d \to \infty$, to the entries of the so-called $n \times n$ GOE matrix $\mathbb{Z}_n$ (Gaussian Orthogonal Ensemble) which is a random matrix with Gaussian elements given by (3.17) with $m_4 = 3$ (this matrix belongs to the larger class of the so-called Wigner matrices). Moreover, the renormalized Wishart matrix $\mathcal{W}_{n,d}$ satisfies a Central Limit Theorem (CLT in the sequel) when $d \to \infty$ and n is fixed. A relatively recent research direction on random matrices is the study of the limit behavior in distribution of the Wishart matrix in the "high-dimensional regime", i.e. when both sizes n and d tend to infinity. This research is motivated by the need to handle large data sets nowadays.

Received by the editors March 12th, 2021; accepted June 23th, 2021.

2010 *Mathematics Subject Classification.* 82C44, 60K35, 60G70.

Key words and phrases. Random dynamics, random environments, scaling limit.

C-P. D. acknowledges partial support from the Labex CEMPI (ANR-11- LABX-0007-01) and C. T. acknowledges partial support from the Labex CEMPI (ANR-11- LABX-0007-01) and MATHAMSUD project SARC (19-MATH-06).

A possible way to analyze the behavior of the Wishart matrix in the high-dimensional regime is to evaluate the Wasserstein distance between its probability distribution and the law of its limiting matrix when the sizes n and d are large enough. It was first discovered (independently) in Bubeck et al. (2016) and Jiang and Li (2015) that, if the entries of the starting matrix $\mathcal{X}_{n,d}$ are independent Gaussian random variables, then the Wasserstein distance between the associated renormalized Wishart matrix $\mathcal{W}_{n,d}$ and its limit (the Wigner matrix (3.17) with $m_4 = 3$) is of order less than $\sqrt{\frac{n^3}{d}}$ which means that $\mathcal{W}_{n,d}$ is "close" to the Wigner matrix when $\frac{n^3}{d}$ tends to zero. Since then, these results have been extended in mainly two directions: by supposing that these entries are independent but with a possibly non- Gaussian distribution, or by assuming that the entries of the initial matrix are (at least partially) correlated. Concerning the second line of research, we mention the recents works Nourdin and Zheng (2019), Bourguin et al. (2021), Diez and Tudor (2021) in which the authors assume that the correlation between the elements of the initial matrix are correlated and the correlation is related to the correlation structure of the increments of the fractional Brownian motion or of the Hermite process. Our work concerns the first direction of study: we start with a matrix with independent entries, not identically distributed, and we assume that these entries follow a very general probability law. As mentioned before, several works treated this question: besides the case of Gaussian entries studied in Bubeck et al. (2016), Jiang and Li (2015) (see also Rácz and Richey, 2019 for results on the phase transition from the Wishart to the limit matrix and Bubeck and Ganguly, 2018 for a discussion of the optimality of the estimates), we mention the paper Bubeck and Ganguly (2018) for entries with a log-concave distribution, the work Mikulincer (2020) also for entries with log-concave distribution but only column independent, the work Bourguin et al. (2021) for the situation when the entries belong to a Wiener chaos of arbirary order, or Fang and Koike (2021+) for entries with a general distribution, assuming only the finiteness of their sixth moment.

Our purpose is to extend these results on the behavior of Wishart matrices to a more general situation by using the techniques of the Malliavin calculus. We will assume that the entries of the initial matrix $\mathcal{X}_{n,d}$ are independent random variables which can be expressed as Skorohod (or divergence) integrals. This covers a very general case since basically any centered square integrable random variable can be expressed as a Skorohod integral. Our results are not covered by the findings in Fang and Koike (2021+) due to the following aspects. Firstly, in Fang and Koike (2021+) the authors measure the Wasserstein distance between the law of the Wishart matrix $\mathcal{W}_{n,d}$ viewed as vector $(W_{i,j}, 1 \leq i < j \leq n)$ and its limit. Notice that the random vector considered in Fang and Koike (2021+) does not include the diagonal terms of the Wishart matrix. This is due to the fact that they use an approach based on the Stein method for exchangeable pairs, which does not allow to include the diagonal of the Wishart matrix. Actually, including the diagonal is not trivial, see Mikulincer (2020) (in this reference the author uses the log-concavity of the law of the entries). Secondly, we work with a different distance (the so-called d_2-distance) which is not necessarily defined via Lipschitz functions. We detailed the relation between our findings and those in Fang and Koike (2021+) in Remark 3.7 and we noticed that in some particular cases, our estimates can be more optimal when the diagonal of the Wishart matrix is considered.

We need in addition some regularity assumptions (in the sense of Malliavin calculus) for the integrands of the Skorohod integrals which define the entries the matrix $\mathcal{X}_{n,d}$ and we will also use the hypothesis of strong independence for the entries of the initial matrix $\mathcal{X}_{n,d}$ which means more than the usual independence. The strong independence of two square integrable random variables F and G actually means that any component of the chaos expansion of F is independent of any component of the chaos expansion of G. This is the price to pay in order to keep a very general form for the entries of the starting matrix. Under these assumptions, we are able to evaluate the distance (the so-called d_2-distance defined later) between the random vector associated to the Wishart matrix and its limit in distribution and then to deduce the Wasserstein distance (in the matrix sense) between the renormalized Wishart matrix and its limiting Gaussian matrix. We use criteria from the recent Stein-Malliavin calculus and we exploit the strong independence of the entries, which leads to several simplifications in the calculations. Actually, we will prove that this d_2-distance between the distribution of the random vector associated to the Wishart matrix (including the diagonal) and its limit is less than $Cn^2\sqrt{\frac{n^3}{d}}$ and then that the Wasserstein distance between the renormalized Wishart matrix $\mathcal{W}_{n,d}$ and its limit is less $C\frac{n^{\frac{9}{4}}}{d^{\frac{1}{4}}}$, meaning that we lose some speed with respect to the standard bound $\sqrt{\frac{n^3}{d}}$. This is due to the following fact: in a first step we majorate the Wasserstein distance between the laws of the random matrices $\mathcal{W}_{n,d}$ and $\mathbb{Z}_n$ by the Wasserstein distance between their so-called associated half-vectors (defined in Section 2.2). But there are no criteria to estimate directly the Wasserstein distance between random vectors whose components are Skorohod integrals and we need to bound it by a new distance (the d_2 distance defined in Section 2.2) in order to have some estimates.

Our paper is structured as follows. In Section 2 we describe the basic tools from Malliavin calculus which are needed in our work as well as the distances between random matrices and random vectors. Section 3 is the core of our work: here we introduce our random matrices, we present the assumptions and prepare, state and prove the main result concerning the limit in distribution, under the high-dimensional regime, of the renormalized Wishart matrix. In Section 4 we present some examples where our theoretical results can be applied.

2. Preliminaries

This section contains the basic tools from Malliavin calculus needed in our work (the monographs Nourdin and Peccati, 2012 and Nualart, 2006 contain a more complete exposition). We also define the distances between random matrices and random vectors which are used in the sequel.

2.1. *Malliavin derivative.* Let $T \subset \mathbb{R}$ be a nonempty set and denote by $L^2_S(T^p)$ the set of real-valued symmetric square integrable functions on T^p. Let $(B_t)_{t\in T}$ be a Wiener process. Denote by $B(\varphi) := \int_T \varphi_s dB_s$ the Wiener integral of $\varphi \in H := L^2(T, \mathcal{B}(T), \lambda)$ with respect to the Brownian motion B. We denoted by λ the Lebesgue measure and $\mathcal{B}(T)$ stands for the Borel subsets of T. The family $(B(\varphi), \varphi \in H)$ forms an isonormal process, i.e. a Gaussian family of centered random variables such that

$$\mathbf{E}B(\varphi_1)B(\varphi_2) = \langle \varphi_1, \varphi_2\rangle_H = \int_T \varphi_1(s)\varphi_2(s)ds$$

for any $\varphi_1, \varphi_2 \in H$.

Denote I_n the multiple stochastic integral with respect to B (see Nualart, 2006). This I_n is actually an isometry between the Hilbert space $H^{\odot n}$(symmetric tensor product) equipped with the scaled norm $\sqrt{n!}\|\cdot\|_{H^{\otimes n}}$ and the Wiener chaos of order n which is defined as the closed linear span of the random variables $H_n(B(\varphi))$ where $\varphi \in H, \|\varphi\|_H = 1$ and H_n is the Hermite polynomial of degree $n \geq 1$

$$H_n(x) = (-1)^n \exp\left(\frac{x^2}{2}\right)\frac{d^n}{dx^n}\left(\exp\left(-\frac{x^2}{2}\right)\right), \quad x \in \mathbb{R}. \tag{2.1}$$

The isometry of multiple integrals can be written as: if $\tilde{f}$ denotes the symmetrization of the function f, for m, n positive integers,

$$\begin{aligned} \mathbf{E}\left(I_n(f)I_m(g)\right) &= n!\langle \tilde{f}, \tilde{g}\rangle_{H^{\otimes n}} \quad \text{if } m = n, \\ \mathbf{E}\left(I_n(f)I_m(g)\right) &= 0 \quad \text{if } m \neq n. \end{aligned} \tag{2.2}$$

We will use the product formula for multiple stochastic integrals, if $f \in L^2_S(T^m)$ and $g \in L^2_S(T^n)$, then

$$I_m(f)I_n(g) = \sum_{r=0}^{m\wedge n} r!\binom{m}{r}\binom{n}{r} I_{m+n-2r}(f \otimes_r g) \tag{2.3}$$

where for $r = 0, ..., m \wedge n$, the contraction $f \otimes_r g$ is the function in $L^2(T^{m+n-2r})$ given by

$$(f \otimes_r g)(t_1, ..., t_{m+n-2r}) = \int_{T^r} f(u_1, ..., u_r, t_1, ..., t_{m-r})g(u_1, ..., u_r, t_{m-r+1}, ..., t_{m+n-2r})du_1...du_r. \tag{2.4}$$

Notice that $f \otimes_r g$ is not necesarily a symmetric function (even if f, g are symmetric) and we will denote by $f\tilde{\otimes}_r g$ its symmetrization.

Let $\mathcal{S}$ be the class of smooth functionals of the form

$$F = f(B_{t_1}, .., B_{t_n}), \quad t_1, .., t_n \in T, \tag{2.5}$$

with $f \in C^\infty(\mathbb{R}^n)$ with at most polynomial growth (for f and its derivatives). For the random variable (2.5) we define its Malliavin derivative with respect to B by

$$D_t F = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(B_{t_1}, .., B_{t_n})1_{[0,t_i]}(t), \quad t \in T.$$

The operator D is an unbounded closable operator and it can be extended to the closure of $\mathcal{S}$ with respect to the Malliavin -Sobolev norm

$$\|F\|^p_{k,p} = \mathbf{E}|F|^p + \sum_{j=1}^k \mathbf{E}\|D^{(j)}F\|^p_{L^2(T^j)}, \quad F \in \mathcal{S}, p \geq 2, k \geq 1. \tag{2.6}$$

where $D^{(j)}$ stands for the jth iterated Malliavin derivative. This closure will be denoted by $\mathbb{D}^{k,p}$.

The Skorohod integral integral, denoted by δ, is the adjoint operator of D. Its domain is

$$Dom(\delta) = \left\{u \in L^2(T \times \Omega), \mathbf{E}\left|\int_T u_s D_s F ds\right| \leq C\|F\|_2\right\}$$

and we have the duality relationship

$$\mathbf{E}F\delta(u) = \mathbf{E}\int_T D_s F u_s ds, \quad F \in \mathcal{S}, u \in Dom(\delta). \tag{2.7}$$

We set $\mathbb{L}^{k,p} = L^p\left(T; \mathbb{D}^{k,p}\right), k \geq 1, p \geq 2$. This set is a subset of $Dom(\delta)$ and it is endowed with the norm

$$\|u\|^p_{k,p} = \int_T \left[\mathbf{E}|u_t|^p + \sum_{j=1}^k \mathbf{E}\|D_{s_1,...,s_k} u_t\|^p_{L^2(T^j)}\right] dt.$$

We recall the Meyer's inequality, for $u \in \mathbb{L}^{k,p}$ with $k \geq 1, p \geq 2$ (see e.g. Nualart, 2006, Proposition 1.5.4)

$$\|\delta(u)\|_{k-1,p} \leq C_p\|u\|_{k,p}. \tag{2.8}$$

We also recall (see e.g. Lemma 1 in Tudor and Yoshida, 2018) that if $F \in \mathbb{D}^{k,p}$ then $D(-L)^{-1}F \in \mathbb{L}^{k+1,p}$ and

$$\|D(-L)^{-1}F\|_{k+1,p} \leq C_{p,k}\|F\|_{k,p} \tag{2.9}$$

where $(-L)^{-1}$ denotes the pseudo-inverse of the Ornstein-Uhlenbeck operator L, which satisfies $(-L)^{-1}I_n F = \frac{1}{n}I_n(f)$ if $n \geq 1$ and $f \in L^2_S(T^n)$.

The Malliavin derivative D acts on the Wiener chaos as an annihilation operator: if $F = I_n(f)$ with $f \in L^2(T^n)$ symmetric, then $D_t F = nI_{n-1}(f(\cdot, t))$ where "$\cdot$" stands for $n-1$ variables in T.

2.2. *Distances.* Let us recall the definition of some distances between random matrices and random vectors. Let $\mathcal{X},\mathcal{Y}$ be two random matrices with values in $\mathcal{M}_n(\mathbb{R})$, $n\geq 1$ (the set of $n\times n$ matrices with real entries). We will denote by d_W the Wasserstein distance between the probability distributions of $\mathcal{X}$ and $\mathcal{Y}$. That is,

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

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

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

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

If X,Y are two random vectors in $\mathbb{R}^n$, we will consider the d_2-distance between their probability distributions

$$d_2(X,Y) = \sup_{\|h''\|_\infty\leq 1} |\mathbf{E}h(X) - \mathbf{E}h(Y)| \tag{2.10}$$

where

$$\|h''\|_\infty = \sup_{x\in\mathbb{R}^n}\sup_{1\leq i,j\leq n}\left|\frac{\partial^2 h}{\partial x_i\partial x_j}(x)\right|.$$

If $\mathcal{X}=(X_{i,j})_{1\leq i,j\leq n}$ 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

$$\mathcal{X}^{\mathrm{half}} = \left(X_{1,1}, X_{1,2}\ldots, X_{1,n}, X_{2,2}, X_{2,3},\ldots, X_{2,n},\ldots, X_{n,n}\right). \tag{2.11}$$

It is possible to bound the Wasserstein distance between two random matrices by a constant times the square root of the d_2-distance between their associated half random vectors as follows.

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

$$d_W(\mathcal{X},\mathcal{Y}) \leq 4n^{\frac{1}{4}}\sqrt{d_2(\mathcal{X}^{\mathrm{half}},\mathcal{Y}^{\mathrm{half}})}, \tag{2.12}$$

where $\mathcal{X}^{\mathrm{half}},\mathcal{Y}^{\mathrm{half}}$ are the associated half-vectors defined in (2.11) and $C>0$.

Proof: The inequality (2.12) is obtained by combining the results in Lemma 2.2 and Proposition 4.4 in Nourdin and Zheng (2019). □

3. The behavior of the Wishart matrix

Here we introduce the starting random matrix $\mathcal{X}_{n,d}$ and the assumptions on its entries. Then we state and prove some auxiliary results concerning the strong independence, which will be used in the final part for the proof of the main result.

3.1. *The starting matrix and the assumptions on its entries.* We will consider a $n\times d$ random matrix $\mathcal{X}_{n,d}=(X_{i,j})_{1\leq i\leq n, 1\leq j\leq d}$ whose entries are centered, square integrable, independent and they are written in a very general form, as an infinite sum of multiple stochastic integrals with respect to an isonormal process (see Section 2.1). More precisely,

$$X_{i,j} = \sum_{p\geq 1} I_p(f_p^{(i,j)}) \tag{3.1}$$

with $f_p^{(i,j)}\in L^2_S(T^p)$ for every $1\leq i\leq n, 1\leq j\leq d$ and for every $p\geq 1$. Notice that we can express the entries $X_{i,j}$ as Skorohod integrals

$$X_{i,j} = \delta(u_{i,j}) \text{ with } u_{i,j}(t) = \sum_{p\geq 1} I_{p-1}(f_p^{(i,j)}(\cdot,t)),\quad t\in T$$

where " $\cdot$ " stands for $p-1$ variables. Actually we have

$$u_{i,j}(t) = D_t(-L)^{-1}X_{i,j}, \quad t \in T \tag{3.2}$$

where D is the Malliavin derivative, L denotes the Ornstein-Uhlenbeck operator with respect to B and $(-L)^{-1}$ its pseudo-inverse (see Section 2.1). The bound (2.9) assures that $u_{i,j}$ is Skorohod integrable for every $1 \leq i \leq n, 1 \leq j \leq d$.

We will make the following hypothesis:

- **H1:** We will assume that $(X_{i,j}, 1 \leq i \leq n, 1 \leq j \leq d)$ are strongly independent random variables. That means that every chaos component of $X_{i,j}$ is independent of every chaos component of $X_{k,l}$ if $(i,j) \neq (k,l)$, i.e. $I_p(f_p^{(i,j)})$ and $I_q(f_q^{(k,l)})$ are independent for every $p, q \geq 1$ and for every $(i,j) \neq (k,l)$.
- **H2:** For every $1 \leq i \leq n, 1 \leq j \leq d$, the random variables $X_{i,j}$ have the same second and fourth moments (without loss of generality, the second moment is assumed to be 1),

$$\mathbf{E}X_{i,j}^2 = 1 \tag{3.3}$$

  and

$$\mathbf{E}X_{i,j}^4 = m_4. \tag{3.4}$$

- **H3:** The processes $(u_{i,j}(t), t \in T)$ are sufficiently regular in the sense of Malliavin calculus, more precisely, for some $p \geq 8$, for every $1 \leq i \leq n, 1 \leq j \leq d$, $u_{i,j} \in \mathbb{L}^{2,p}$ and

$$\|u_{i,j}\|_{2,p} \leq C_p \tag{3.5}$$

  with $C_p > 0$ an universal constant depending only on p.

3.2. *Some technical results.* This paragraph is devoted to the proof of some technical results needed later. These results concern some inequalities for the Malliavin-Sobolev norms and some consequences of the strong independence assumption.

We start with the following crucial lemma, well-known in the Stein-Malliavin calculus.

Lemma 3.1. *Let $F = \sum_{p\geq 1} I_p(f_p)$ with $f_p \in L_S^2(T^p)$ be a centered random variable in $\mathbb{D}^{k,p}$ with $k \geq 1, p \geq 2$. Then $u = D(-L)^{-1}F \in \mathbb{L}^{k+1,p}$ and*

$$F = \delta D(-L)^{-1}(F). \tag{3.6}$$

In particular $u = D(-L)^{-1}F$ given by $u(t) = \sum_{p\geq 1} I_{p-1}(f_p(\cdot, t))$ for every $t \in T$.

Proof: Let $p \geq 2$. The fact that $u = D(-L)^{-1}F$ belongs to $\mathbb{L}^{1,p}$ follows from the inequality (2.9) while the identity (3.6) is well-known (see e.g. Nourdin and Peccati, 2012). □

Lemma 3.2. *Let $u, v \in \mathbb{L}^{2,2p}$ with $p \geq 1$. Then $\delta(u)\delta(v) \in \mathbb{D}^{1,p}$ and*

$$\|\delta(u)\delta(v)\|_{1,p} \leq C_p \|u\|_{2,2p} \|v\|_{2,2p}.$$

Proof: Let $p \geq 2$. We use the definition of the norm in $\mathbb{D}^{1,p}$, the derivation rule for D and the inequality $(a+b)^q \leq 2^{q-1}(a^q + b^q)$ for $q \geq 1$, we obtain

$$\begin{aligned}
\|\delta(u)\delta(v)\|_{1,p}^{p} &= \mathbf{E}\,|\delta(u)\delta(v)|^{p}+\mathbf{E}\left(\int_{T}(D_{s}(\delta(u)\delta(v)))^{2}ds\right)^{\frac{p}{2}}\\
&= \mathbf{E}\,|\delta(u)\delta(v)|^{p}+\mathbf{E}\left(\int_{T}[\delta(u)D_{s}\delta(v)+\delta(v)D_{s}\delta(u)]^{2}\,ds\right)^{\frac{p}{2}}\\
&\leq \mathbf{E}\,|\delta(u)\delta(v)|^{p}+\mathbf{E}\left(2\int_{T}([\delta(u)D_{s}\delta(v)]^{2}+[\delta(v)D_{s}\delta(u)]^{2})ds\right)^{\frac{p}{2}}\\
&\leq \mathbf{E}\,|\delta(u)\delta(v)|^{p}+2^{\frac{p}{2}}\mathbf{E}\left(\int_{T}(\delta(u)D_{s}\delta(v))^{2}ds+\int_{T}(\delta(v)D_{s}\delta(u))^{2}ds\right)^{\frac{p}{2}}\\
&\leq \mathbf{E}\,|\delta(u)\delta(v)|^{p}+2^{p-1}\mathbf{E}\left|\delta(u)^{p}\left(\int_{T}(D_{s}\delta(v))^{2}ds\right)^{\frac{p}{2}}\right|\\
&\quad +2^{p-1}\mathbf{E}\left|\delta(v)^{p}\left(\int_{T}(D_{s}\delta(u))^{2}ds\right)^{\frac{p}{2}}\right|\\
&\leq (\mathbf{E}\,|\delta(u)|^{2p})^{\frac{1}{2}}(\mathbf{E}\,|\delta(v)|^{2p})^{\frac{1}{2}}\\
&\quad +2^{p-1}\left(\mathbf{E}\,|\delta(u)|^{2p}\right)^{\frac{1}{2}}\left(\mathbf{E}\left(\int_{T}(D_{s}\delta(v))^{2}ds\right)^{p}\right)^{\frac{1}{2}}\\
&\quad + 2^{p-1}\left(\mathbf{E}\,|\delta(v)|^{2p}\right)^{\frac{1}{2}}\left(\mathbf{E}\left(\int_{T}(D_{s}\delta(u))^{2}ds\right)^{p}\right)^{\frac{1}{2}}. \qquad (3.7)
\end{aligned}$$

Notice that

$$\mathbf{E}\,|\delta(u)|^{2p}\leq\|\delta(u)\|_{1,2p}^{2p}\leq C_{p}\|u\|_{2,2p}^{2p} \qquad (3.8)$$

where the last inequality is obtained via Meyer's inequality (2.8). Clearly a similar bound will hold for v. Also, from the definition of the norm in $\mathbb{D}^{1,2p}$,

$$\mathbf{E}\left(\int_{T}(D_{s}\delta(u))^{2}ds\right)^{p}\leq C_{p}\|\delta(u)\|_{1,2p}^{2p}\leq C_{p}\|u\|_{2,2p}^{2p}. \qquad (3.9)$$

Then the conclusion follows by plugging (3.8) and (3.9) into (3.7). □

Let us now state and prove some results concerning the strongly independent random variables. Let us recall a key result from Üstünel and Zakai (1989) concerning the independence of multiple stochastic integrals. For $n,m\geq 1$, let $f\in L^{2}_{S}(T^{n})$ and $g\in L^{2}_{S}(T^{m})$. The multiple Wiener integrals $I_{n}(f)$ and $I_{m}(g)$ are independent if and only if (recall the definition (2.4) of the contraction)

$$f\otimes_{1}g=0 \text{ almost everywhere on } T^{m+n-2}. \qquad (3.10)$$

Relation (3.10) implies that for $r=1,...,n\wedge m$,

$$f\otimes_{r}g=0 \text{ almost everywhere on } T^{m+n-2r}. \qquad (3.11)$$

Lemma 3.3. *Consider the random variables $F=\sum_{p\geq 1}I_{p}(f_{p})$ and $G=\sum_{q\geq 1}I_{q}(g_{q})$ with $f_{p},g_{p}\in L^{2}_{S}(T_{p})$ for every $p\geq 1$. Assume that $F,G\in\mathbb{D}^{1,4}$ and that they are strongly independent. Then*

(1) *The random variables F^{2} and G^{2} are strongly independent.*
(2) *Let $u=D(-L)^{-1}F$ and $v=D(-L)^{-1}G$. Then*

$$\langle u,v\rangle_{L^{2}(T)}=\langle u,DG\rangle_{L^{2}(T)}=\langle v,DF\rangle_{L^{2}(T)}=0 \text{ almost surely.}$$

(3) *Let* $u = D(-L)^{-1}F$ *and* $v = D(-L)^{-1}G$. *Then the random variables* $\langle DF, u\rangle_{L^2(T)}$ *and* $\langle DG, v\rangle_{L^2(T)}$ *are strongly independent.*

(4) *Let* $H = \sum_{p\geq 1} I_p(h_p)$, $J = \sum_{q\geq 1} I_q(j_q)$ *with* $h_p, j_p \in L^2_S(T^p)$ *for* $p \geq 1$ *be two other random variables in* $\mathbb{D}^{1,4}$. *Assume that* F, G, H, J *are mutually strongly independent. Then the random variables* FG *and* HJ *are strongly independent.*

Proof: For point 1., by the product formula (2.3),

$$F^2 = \sum_{p_1,p_2\geq 1}\sum_{r=0}^{p_1\wedge p_2} r!\binom{p_1}{r}\binom{p_2}{r} I_{p_1+p_2-2r}(f_{p_1}\otimes_r f_{p_2})$$

and

$$G^2 = \sum_{p_1,p_2\geq 1}\sum_{r=0}^{p_1\wedge p_2} r!\binom{p_1}{r}\binom{p_2}{r} I_{p_1+p_2-2r}(g_{p_1}\otimes_r g_{p_2})$$

To obtain the conclusion, it suffices to show that for every $p_1, p_2, q_1, q_2 \geq 1$ and for every $r_1 = 0, ..., p_1\wedge p_2$, $r_2 = 0, ..., q_1 \wedge q_2$,

$$(f_{p_1}\tilde{\otimes}_{r_1} f_{p_2}) \otimes_1 (g_{q_1}\tilde{\otimes}_{r_2} g_{q_2}) = 0 \text{ a.e. on } T^{p_1+p_2+q_1+q_2-2r_1-2r_2-2}$$

and this follows by Lemma 3.1 in Bourguin et al. (2021).

Let us prove point 2. For every $t \in T$, we have

$$u(t) = \sum_{p\geq 1} I_{p-1}(f_p(\cdot, t)) \text{ and } v(t) = \sum_{q\geq 1} I_{q-1}(g_q(\cdot, t)).$$

It suffices to show that for every $p, q \geq 1$,

$$\int_T I_{p-1}(f_p(\cdot,t)) I_{q-1}(g_q(\cdot,t))dt = 0 \text{ almost surely.}$$

Again by the product formula (2.3),

$$\begin{aligned}
&\int_T I_{p-1}(f_p(\cdot,t)) I_{q-1}(g_q(\cdot,t))dt \\
&= \int_T dt \sum_{r=0}^{(p-1)\wedge(q-1)} r!\binom{p-1}{r}\binom{q-1}{r} I_{p+q-2r-2}(f_p(\cdot,t)\otimes g_q(\cdot,t)) \\
&= \sum_{r=0}^{(p-1)\wedge(q-1)} r!\binom{p-1}{r}\binom{q-1}{r} I_{p+q-2r-2}(f_p\otimes_{r+1} g_q)
\end{aligned}$$

and by (3.11), for every $g \geq 0$, $f_p \otimes_{r+1} g_q = 0$ almost everywhere on $T^{p+q-2r-2}$. Point 3. is a consequence of Lemma 3.3 in Bourguin et al. (2021).

Let us show the last point of the statement. By the product formula and the strongly independance, we remain with the simple expressions

$$FG = \sum_{p_1,p_2\geq 1} I_{p_1+p_2}(f_{p_1}\otimes g_{p_2})$$

and

$$HJ = \sum_{p_1,p_2\geq 1} I_{p_1+p_2}(h_{p_1}\otimes j_{p_2})$$

Thus, it suffices to show that for every $p_1, p_2, q_1, q_2 \geq 1$

$$(f_{p_1}\tilde{\otimes} g_{p_2}) \otimes_1 (h_{q_1}\tilde{\otimes} j_{q_2}) = 0 \text{ a.e. on } T^{p_1+p_2+q_1+q_2-2} \tag{3.12}$$

where

$$(f_{p_1}\tilde{\otimes}g_{p_2})(t_1,\dots,t_{p_1+p_2})$$
$$=\frac{1}{(p_1+p_2)!}\sum_{\sigma\in\mathfrak{S}_{p_1+p_2}} f(t_{\sigma(1)},...,t_{\sigma(p_1)})g(t_{\sigma(p_1+1)},...,t_{\sigma(p_1+p_2)}).$$

Similarly,

$$(h_{q_1}\tilde{\otimes}j_{q_2})(t_1,\dots,t_{q_1+q_2})$$
$$=\frac{1}{(q_1+q_2)!}\sum_{\sigma\in\mathfrak{S}_{q_1+q_2}} h(t_{\sigma(1)},...,t_{\sigma(q_1)})j(t_{\sigma(q_1+1)},...,t_{\sigma(q_1+q_2)})$$

Hence, we can write via (2.4),

$$\left(\left(f_{p_1}\tilde{\otimes}g_{p_2}\right)\otimes_1\left(h_{q_1}\tilde{\otimes}j_{q_2}\right)\right)(t_1,\dots,t_{p_1+p_2+q_1+q_2-2})$$
$$=\int_T (f_{p_1}\tilde{\otimes}g_{p_2})(t_1,\dots,t_{p_1+p_2-1},x)(h_{q_1}\tilde{\otimes}j_{q_2})(t_{p_1+p_2},\dots,t_{p_1+p_2+q_1+q_2-1},x)dx. \tag{3.13}$$

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

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

so that by plugging the above identity into (3.13), we get

$$\left[\left(f_{p_1}\tilde{\otimes}g_{p_2}\right)\otimes_1\left(h_{q_1}\tilde{\otimes}j_{q_1}\right)\right](t_1,\dots,t_{p_1+p_2+q_1+q_2-2})$$
$$=\frac{1}{(p_1+p_2)!(q_1+q_2)!}\sum_{\sigma\in\mathfrak{S}_{p_1+p_2-1},\tau\in\mathfrak{S}_{q_1+q_2-1}}\sum_{i=1}^{p_1+p_2}\sum_{j=1}^{q_1+q_2}$$
$$\int_T (f_{p_1}\otimes g_{p_2})(t_{\sigma(1)},\dots,t_{\sigma(i-1)},x,t_{\sigma(i+1)},\dots,t_{\sigma(p_1+p_2-1)})$$
$$(h_{q_1}\otimes j_{q_1})(t_{\tau(1)},\dots,t_{\tau(j-1)},x,t_{\tau(j+1)},\dots,t_{\tau(q_1+q_2-1)})dx.$$

To obtain (3.12), we prove that for all $1\le i\le p_1+p_2$ and $1\le j\le q_1+q_2$,

$$\int_T (f_{p_1}\otimes g_{p_2})(t_{\sigma(1)},\dots,t_{\sigma(i-1)},x,t_{\sigma(i+1)},\dots,t_{\sigma(p_1+p_2-1)}) \tag{3.14}$$
$$(h_{q_1}\otimes j_{q_2})(t_{\tau(1)},\dots,t_{\tau(j-1)},x,t_{\tau(j+1)},\dots,t_{\tau(q_1+q_2-1)})dx=0$$

almost everywhere with respect to $t_1,\dots,t_{p_1+p_2+q_1+q_2-2}$.

Assume that $1\le i\le p_1$ and $1\le j\le q_1$ (the other cases can be dealt with in the same way). Then, we have

$$\int_T (f_{p_1}\otimes g_{p_2})(t_{\sigma(1)},\dots,t_{\sigma(i-1)},x,,t_{\sigma(i+1)},\dots,t_{\sigma(p_1+p_2-1)})$$
$$(h_{q_1}\otimes j_{q_2})(t_{\tau(1)},\dots,t_{\tau(j-1)},x,t_{\tau(j+1)},\dots,t_{\tau(q_1+q_2-1)})dx$$
$$=\int_T f_{p_1}(t_{\sigma(1)},\dots,t_{\sigma(i-1)},x,t_{\sigma(i+1)},\dots,t_{\sigma(p_1-1)})g_{p_2}(t_{\sigma(p_1)},\dots,t_{\sigma(p_1+p_2-1)})$$
$$h_{q_2}(t_{\tau(1)},\dots t_{\tau(j-1)},x,t_{\tau(q_1-1)})j_{q_2}(t_{\tau(q_1)},\dots,t_{\tau(q_1+q_2-1)})dx.$$

Now, the strong indepedence and so the fact that the contraction of f and h vanishes, implies for almost every $t_{\sigma(1)},\dots,t_{\sigma(p_1+p_2-1)},t_{\tau(1)},\dots,t_{\tau(q_1+q_1-1)}$ (see (3.10))

$$
\begin{aligned}
& \int_T f_{p_1}(t_{\sigma(1)}, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, ..., t_{\sigma(p_1-1)}) g_{p_2}(t_{\sigma(p_1)}, \dots, t_{\sigma(p_1+p_2-1)}) \\
& \times h_{q_1}(t_{\tau(1)}, \dots, t_{\tau(j-1)}, x, t_{\sigma(j+1)}, ..., t_{\sigma(q_1-1)}) j_{q_2}(t_{\tau(q_1)}, \dots, t_{\tau(q_1+q_2-1)}) dx \\
= \;& g_{p_2}(t_{\sigma(p_1)}, \dots, t_{\sigma(p_1+p_2-1)}) j_{q_2}(t_{\tau(q_1)}, \dots, t_{\tau(q_1+q_2-1)}) \\
& \times \int_T f_{p_2}(t_{\sigma(1)}, \dots, t_{\sigma(i-1)}, x, t_{\sigma(i+1)}, ..., t_{\sigma(p_1-1)}) h_{q_1}(t_{\tau(1)}, \dots, t_{\tau(j-1)}, x, t_{\sigma(j+1)}, ..., t_{\sigma(q_1-1)}) dx \\
= \;& 0
\end{aligned}
$$

which concludes the proof. □

3.3. *The Wishart matrix and its asymptotic behavior.* We introduce the (renormalized) Wishart matrix $\mathcal{W}_{n,d} = (W_{i,j})_{1\leq i,j\leq n}$ associated to the starting matrix $\mathcal{X}_{n,d}$ whose entries are given in (3.1),

$$
\mathcal{W}_{n,d} = \sqrt{d}\left(\frac{1}{d}\mathcal{X}_{n,d}\mathcal{X}_{n,d}^T - \mathcal{I}_n\right)
$$

where "^{T}" denotes the transpose and $\mathcal{I}_n$ the identity $n\times n$ matrix. Its components are

$$
W_{i,i} = \frac{1}{\sqrt{d}}\sum_{k=1}^{d}(X_{i,k}^2 - 1), \text{ for } 1\leq i\leq n \tag{3.15}
$$

and

$$
W_{i,j} = \frac{1}{\sqrt{d}}\sum_{k=1}^{d} X_{i,k}X_{j,k} \text{ for } 1\leq i,j\leq n, i\neq j. \tag{3.16}
$$

The independence of the components of the matrix $\mathcal{X}_{n,d}$ and the assumptions (3.3), (3.4) imply

$$
\mathbf{E}W_{i,i}^2 = m_4 - 1 \text{ and } \mathbf{E}W_{i,j}^2 = 1 \text{ for } 1\leq i,j\leq n, i\neq j.
$$

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

$$
\begin{cases}
Z_{i,i} \sim N(0, m_4 - 1) & \text{for } 1\leq i\leq n \\
Z_{i,j} \sim N(0,1) & \text{for } 1\leq i<j\leq n\ , \\
Z_{i,j} = Z_{j,i} & \text{for } 1\leq j<i\leq n
\end{cases} \tag{3.17}
$$

where the entries $(Z_{i,j}\colon i\leq j)$ are (mutually) independent. Clearly, by the standard Central Limit Theorem, the Wishart matrix $\mathcal{W}_{n,d}$ converges componentwise in distribution, as $d\to\infty$ to the Wigner matrix $\mathbb{Z}_n$ when n is fixed. We are interested to evaluate the distance between $\mathcal{W}_{n,d}$ and the Wigner matrix when both dimensions n,d are large enough.

Our main tool to evaluate the distance between the Wishart and Wigner matrices is the following result (Proposition 2.3 in Huang et al., 2020, see also relation (4.6) and the footnote (6) in Nourdin and Zheng, 2019).

Proposition 3.4. *Let $F = (F_1, ..., , F_m)$ be a random vector with $F_i = \delta(u_i), u_i \in Dom(\delta)$ and $F_i \in \mathbb{D}^{1,2}$ for every $1\leq i\leq m$. Let Z be a centered Gaussian vector with covariance matrix $C = (C_{i,j})_{1\leq i,j\leq m}$. Then*

$$
d_2(F,Z) \leq \frac{m}{2}\sqrt{\sum_{i,j=1}^{m}\mathbf{E}\left(C_{i,j} - \langle DF_i, u_j\rangle_{L^2(T)}\right)^2}.
$$

Let us observe that the elements of the Wishart matrix $\mathcal{W}_{n,d}$ can be written as Skorohod integrals. Indeed, since for every $1 \leq k \leq d$ and $1 \leq i,j \leq n$ with $i \neq j$ the random variables $X_{i,k}^2 - 1$ and $X_{i,k}X_{j,k}$ are centered (by assumption (3.3)) and sufficiently regular (by **H3**). Then by Lemma 3.2, we have

$$X_{i,k}^2 - 1 = \delta D(-L)^{-1}(X_{i,k}^2 - 1), \;\; 1 \leq i \leq n, 1 \leq k \leq d. \tag{3.18}$$

Consequently, the diagonal entries of the Wishart matrix can be expressed as, for every $1 \leq i \leq n$,

$$W_{i,i} = \delta(V_{i,i}) \text{ with } V_{i,i} = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} D(-L)^{-1}(X_{i,k}^2 - 1). \tag{3.19}$$

Similarly, for every $1 \leq i,j \leq n$ with $i \neq j$, we have

$$W_{i,j} = \delta(V_{i,j}) \text{ with } \frac{1}{\sqrt{d}} \sum_{k=1}^{d} D(-L)^{-1} X_{i,k}X_{j,k}. \tag{3.20}$$

Since all the elements of $\mathcal{W}_{n,d}$ can be expressed as Skorohod integrals, we will apply Proposition 3.4 in order to evaluate the d_2-distance between the half vectors associated to the $\mathcal{W}_{n,d}$ and to the Wigner matrix $\mathbb{Z}_n$. To this end we need to calculate and to evaluate the quantity

$$\mathbf{E}\left(\langle DW_{i,j}, V_{a,b}\rangle_{L^2(T)} - \mathbf{E}(Z_{i,j}Z_{a,b})\right)^2 \tag{3.21}$$

for every $1 \leq i,j,a,b \leq n$ with $i \leq j$ and $a \leq b$. The processes $V_{i,j}$ are those defined in (3.19) and (3.20) respectively.

If $1 \leq i,j,a,b \leq n$, we denote by $M_{i,j,a,b}$ the subset of $\{1,2,..,n\}^4$ such that $i \leq j, a \leq b$ and $\{i,j\} \cap \{a,b\} = \emptyset$. The quantity (3.21) is estimated in the below result.

Proposition 3.5. *Assume* **H1**-**H3**. *Let $(W_{i,j}, 1 \leq i,j \leq n)$ be given by (3.15), (3.16). Then for $1 \leq i,j,a,b \leq n$ with $i \leq j, a \leq b$,*

$$\mathbf{E}\left(\langle DW_{i,j}, V_{a,b}\rangle - \mathbf{E}Z_{i,j}Z_{a,b}\right)^2 \leq C\frac{1}{d} \textit{ if } (i,j,a,b) \notin M_{i,j,a,b}$$

and

$$\langle DW_{i,j}, V_{a,b}\rangle - \mathbf{E}Z_{i,j}Z_{a,b} = 0 \textit{ if } (i,j,a,b) \in M_{i,j,a,b}.$$

Proof: Assume $(i,j,a,b) \notin M_{i,j,a,b}$. We separate the proof into the following cases: $(i = j = a = b), (i = a \neq j = b)$ and $(i = a$ or $j = b)$. Note that $\mathbf{E}\left(Z_{i,i}^2\right) = m_4 - 1$, $\mathbf{E}\left(Z_{i,j}^2\right) = 1$ if $i \neq j$, and $\mathbf{E}\left(Z_{i,j}Z_{a,b}\right) = 0$ if $(i,j) \neq (a,b)$.

Let us consider first the case $i = j = a = b$. We have,

$$\langle DW_{i,i}, V_{i,i}\rangle \;\; = \;\; \frac{1}{d}\sum_{k,l=1}^{d} \langle D(X_{i,k}^2 - 1), D(-L)^{-1}(X_{i,l}^2 - 1)\rangle_{L^2(T)}.$$

Notice that $X_{i,k}^2$ is strongly independent, for $k \neq l$, by $X_{i,l}^2$ by Lemma 3.3, point 1. Therefore, by Lemma 3.3, point 2. we have

$$\langle D(X_{i,k}^2 - 1), D(-L)^{-1}(X_{i,l}^2 - 1)\rangle_{L^2(T)} = 0 \text{ if } 1 \leq k,l \leq d, k \neq l.$$

Thus we can write

$$\langle DW_{i,i}, V_{i,i}\rangle = \frac{1}{d}\sum_{k=1}^{d} \langle D(X_{i,k}^2 - 1), D(-L)^{-1}(X_{i,k}^2 - 1)\rangle_{L^2(T)}.$$

On the other hand, by the duality formula (2.7), (3.18) and assumptions (3.3), (3.4),

$$\begin{aligned} &\mathbf{E}\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)} = \mathbf{E}(X_{i,k}^2-1)\delta(D(-L)^{-1}(X_{i,k}^2-1)) \\ = \;& \mathbf{E}\left(X_{i,k}^2-1\right)^2 = m_4-1. \end{aligned}$$

Thus

$$\begin{aligned} &\mathbf{E}\left(\langle DW_{i,i}, V_{i,i}\rangle - (m_4-1)\right)^2 \\ = \;& \frac{1}{d^2}\mathbf{E}\left(\sum_{k=1}^d \langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right. \\ & \left.-\mathbf{E}\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right)^2 \\ = \;& \frac{1}{d^2}\sum_{k,l=1}^d \mathbf{E}\left(\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right. \\ & \left.-\mathbf{E}\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right) \\ & \times\left(\langle D(X_{i,l}^2-1), D(-L)^{-1}(X_{i,l}^2-1)\rangle_{L^2(T)} - \mathbf{E}\langle D(X_{i,l}^2-1), D(-L)^{-1}(X_{i,l}^2-1)\rangle_{L^2(T)}\right) \\ = \;& \frac{1}{d^2}\sum_{k=1}^d \mathbf{E}\left(\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right. \\ & \left.-\mathbf{E}\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right)^2. \end{aligned}$$

We used the fact that for $k \neq l$, the random variables $X_{i,k}^2$ and $X_{i,l}^2$ are strongly independent (Lemma 3.3, point 1.) and also the fact that $\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}$ and $\langle D(X_{i,l}^2-1), D(-L)^{-1}(X_{i,l}^2-1)\rangle_{L^2(T)}$ are independent (Lemma 3.3, point 3.)

It suffices to show that for every i,k,

$$\mathbf{E}\left(\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right)^2 \leq C$$

with $C>0$ an universal constant (not depending on i,k). We have

$$\begin{aligned} &\mathbf{E}\left(\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right)^2 \\ \leq \;& \mathbf{E}\left(\|D(X_{i,k}^2-1)\|_{L^2(T)}^2 \|D(-L)^{-1}(X_{i,k}^2-1)\|_{L^2(T)}^2\right) \\ \leq \;& \left(\mathbf{E}\|D(X_{i,k}^2-1)\|_{L^2(T)}^4\right)^{\frac{1}{2}} \left(\mathbf{E}\|D(-L)^{-1}(X_{i,k}^2-1)\|_{L^2(T)}^4\right)^{\frac{1}{2}} \\ \leq \;& C\|X_{i,k}^2-1\|_{1,4}^2 \|D(-L)^{-1}(X_{i,k}^2-1)\|_{1,4}^2 \end{aligned} \tag{3.22}$$

and by Lemma 3.2 and (2.9),

$$\begin{aligned} &\mathbf{E}\left(\langle D(X_{i,k}^2-1), D(-L)^{-1}(X_{i,k}^2-1)\rangle_{L^2(T)}\right)^2 \\ \leq \;& \|\delta(u_{i,k})^2-1\|_{1,4}^2 \|\delta(u_{i,k})^2-1\|_{L^4(\Omega)}^2 \\ \leq \;& C(\|u_{i,k}\|_{2,8}^8+1) \leq C \end{aligned}$$

due to (3.5).

Now, let us assume $i=a\neq j=b$ and compute the term $\langle DW_{i,j}, V_{i,j}\rangle_{L^2(T)}$ with $i\neq j$ where $W_{i,j}, V_{i,j}$ are given by (3.16) and (3.20) respectively.

Since the random variables $X_{i,k}X_{j,k}$ are centered for every $1\leq i\leq n, 1\leq k\leq d$, we have using Lemma 3.2

$$X_{i,k}X_{j,k} = \delta(D(-L)^{-1}(X_{i,k}X_{j,k}))$$

and by Lemma 3.3 point 4. $X_{i,k}X_{j,k}$ and $X_{i,l}X_{j,l}$ are strongly independant for $k \neq l$. The same Lemma 3.3 point 2. implies

$$\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,l}X_{j,l})\rangle_{L^2(T)} = 0 \text{ if } 1 \leq k \neq l \leq d$$

and consequently

$$\langle DW_{i,j}, V_{i,j}\rangle = \frac{1}{d}\sum_{k=1}^{d}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)} \tag{3.23}$$

Moreover, using again the duality formula, the strongly independence and assumption (3.3),

$$\begin{aligned} & \mathbf{E}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)} \\ = \; & \mathbf{E}(X_{i,k}X_{j,k}\delta(D(-L)^{-1}(X_{i,k}X_{j,k})) = \mathbf{E}\left(X_{i,k}^2X_{j,k}^2\right) = 1. \end{aligned}$$

Thus

$$\begin{aligned} & \mathbf{E}\left(\langle DW_{i,j}, V_{i,j}\rangle - 1\right)^2 \\ = \; & \frac{1}{d^2}\mathbf{E}\left(\sum_{k=1}^{d}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right. \\ & \left. -\mathbf{E}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right)^2 \\ = \; & \frac{1}{d^2}\sum_{k,l=1}^{d}\mathbf{E}\left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right. \\ & \left. -\mathbf{E}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right) \\ & \times\left(\langle D(X_{i,l}X_{j,l}), D(-L)^{-1}(X_{i,l}X_{j,l})\rangle_{L^2(T)} - \mathbf{E}\langle D(X_{i,l}X_{j,l}), D(-L)^{-1}(X_{i,l}X_{j,l})\rangle_{L^2(T)}\right) \\ = \; & \frac{1}{d^2}\sum_{k=1}^{d}\mathbf{E}\left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right. \\ & \left. -\mathbf{E}\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right)^2. \end{aligned}$$

We also used the fact that for $k \neq l$, the random variables $\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}$ and $\langle D(X_{i,l}X_{j,l}), D(-L)^{-1}(X_{i,l}X_{j,l})\rangle_{L^2(T)}$ are also strongly independent due to point 3. in Lemma 3.3.

It suffices to show that for every $i \neq j, k$, with $C > 0$ not depending on these parameters,

$$\mathbf{E}\left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right)^2 \leq C.$$

We have

$$\begin{aligned} & \mathbf{E}\left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right)^2 \\ \leq \; & \mathbf{E}\left(\|D(X_{i,k}X_{j,k})\|_{L^2(T)}^2\|D(-L)^{-1}(X_{i,k}X_{j,k})\|_{L^2(T)}^2\right) \\ \leq \; & \left(\mathbf{E}\|D(X_{i,k}X_{j,k})\|_{L^2(T)}^4\right)^{\frac{1}{2}}\left(\mathbf{E}\|D(-L)^{-1}(X_{i,k}X_{j,k})\|_{L^2(T)}^4\right)^{\frac{1}{2}} \\ \leq \; & \|X_{i,k}X_{j,k}\|_{1,4}^2\|D(-L)^{-1}(X_{i,k}X_{j,k})\|_{1,4}^2 \end{aligned}$$

and by Lemma 3.2 and (2.9), by proceeding as for the bound (3.22),

$$\mathbf{E}\left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{j,k})\rangle_{L^2(T)}\right)^2 \leq C_p\|u_{i,k}\|_{2,8}^4\|u_{j,k}\|_{2,8}^4 \leq C$$

where the last inequality is due to the assumption **H3**.

Next case we have to deal is when $i = a$, $j \neq b$, (by symmetry we can deduce the case $i \neq a$ $j = b$). Using the same arguments as above and assuming $i < j$ and $i \leq b$, we get

$$\begin{aligned} \langle DW_{i,j}, V_{i,b} \rangle &= \frac{1}{d} \sum_{k,l=1}^{d} \langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,l}X_{b,l}) \rangle_{L^2(T)} \\ &= \frac{1}{d} \sum_{k=1}^{d} \langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{b,k}) \rangle_{L^2(T)} \end{aligned}$$

.

Moreover by the same bounds as above, the following inequality is verified for every (i, j, b, k) such as (i, j, b) are all distinct

$$\mathbf{E} \left(\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{i,k}X_{b,k}) \rangle_{L^2(T)} \right)^2 \leq C$$

The next case is when $(i, j, a, b) \in M_{i,j,a,b}$,

$$\begin{aligned} \langle DW_{i,j}, V_{a,b} \rangle &= \frac{1}{d} \sum_{k,l=1}^{d} \langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{a,l}X_{b,l}) \rangle_{L^2(T)} \\ &= \frac{1}{d} \sum_{k,l=1}^{d} \langle D(\delta(u_{i,k})\delta(u_{j,k})), D(-L)^{-1}(\delta(u_{a,l})\delta(u_{b,l}) \rangle_{L^2(T)}. \end{aligned}$$

We have, by using Lemma 3.3, points 2 and 4., $\forall\ 1 \leq k, l \leq d$

$$\langle D(X_{i,k}X_{j,k}), D(-L)^{-1}(X_{a,l}X_{b,l}) \rangle_{L^2(T)} = 0 \text{ if } (i, j, a, b) \in M_{i,j,a,b}$$

Hence $\forall\ (i, j, a, b) \in M_{i,j,a,b}$,

$$\langle DW_{i,j}, V_{a,b} \rangle = 0$$

□

The above result allows to evaluate the d_2-distance between the half-vector associated to the Wishart matrix $\mathcal{W}_{n,d}$ and its limit. Recall that the half-vector associated to a matrix is defined by (2.11).

Theorem 3.6. *Assume that $\mathcal{W}_{n,d}$ has the entries given by (3.15) and (3.16) and consider the Wigner matrix (3.17). Then for every $n, d \geq 1$, and for any function $h : \mathbb{R}^{\frac{n(n+1)}{2}} \to \mathbb{R}$ with bounded second partial derivative,*

$$\left| \mathbf{E}h(\mathcal{W}_{n,d}^{\text{half}}) - \mathbf{E}h(\mathbb{Z}_n^{\text{half}} \right| \leq Cn^2 \sqrt{\frac{n^3}{d}} \tag{3.24}$$

and consequently

$$d_2 \left(\mathcal{W}_{n,d}^{\text{half}}, \mathbb{Z}_n^{\text{half}} \right) \leq Cn^2 \sqrt{\frac{n^3}{d}}.$$

Proof: It suffices to use Proposition 3.4 and the estimates in Proposition 3.5, by noticing that the dimension of the vectors $\mathcal{W}_{n,d}^{\text{half}}$ and $\mathbb{Z}_n^{\text{half}}$ is $\frac{n(n+1)}{2}$ and the cardinal of the set $\overline{M_{i,j,a,b}}$ (the complement of the set $M_{i,j,a,b}$) is less than $6n^3$. □

Remark 3.7. Let us denore by $\overline{\mathcal{W}}_{n,d}^{\text{half}}$ and $\overline{\mathbb{Z}}_n^{\text{half}}$ the following random vectors:

$$\overline{\mathcal{W}}_{n,d}^{\text{half}} = (W_{1,2}, \ldots, W_{1,n}, W_{2,3}, \ldots, W_{2,n}, \ldots, W_{n-1,n})$$

and

$$\overline{\mathbb{Z}}_n^{\text{half}} = (Z_{1,2}, \ldots, Z_{1,n}, Z_{2,3}, \ldots, Z_{2,n}, \ldots, Z_{n-1,n}).$$

That is, the components of the above vectors are the components of $\mathcal{W}_{n,d}^{\text{half}}$ and $\mathbb{Z}_n^{\text{half}}$ without the diagonal terms. It follows from Theorem 1.2 in Fang and Koike (2021+) that for any suitable function $h : \mathbb{R}^{\frac{n(n-1)}{2}} \to \mathbb{R}$ (in particular h is Lipschitz but it also satisfies additional conditions),

$$\left|\mathbf{E}h(\overline{\mathcal{W}}_{n,d}^{\text{half}}) - \mathbf{E}h(\overline{\mathbb{Z}}_n^{\text{half}})\right| \leq C\sqrt{\frac{n^3}{d}}. \tag{3.25}$$

Notice that the above estimate (3.25) does not take into account the diagonal of the matrice $\mathcal{W}_{n,d}$, which is not trivial to include (actually, in Mikulincer (2020) one uses the log-concavity of the distribution of the entries in order to include the diagonal terms). Moreover, if we would use the triangle inequality (which is not probably, the most optimal choice),

$$\begin{aligned}\left|\mathbf{E}h(\mathcal{W}_{n,d}^{\text{half}}) - \mathbf{E}h(\mathbb{Z}_n^{\text{half}})\right| &\leq \left|\mathbf{E}h(\mathcal{W}_{n,d}^{\text{half}}) - \mathbf{E}h(\overline{\mathcal{W}}_{n,d}^{\text{half}})\right| \\ &\quad + \left|\mathbf{E}h(\overline{\mathcal{W}}_{n,d}^{\text{half}}) - \mathbf{E}h(\overline{\mathbb{Z}}_n^{\text{half}})\right| + \left|\mathbf{E}h(\overline{\mathbb{Z}}_n^{\text{half}}) - \mathbf{E}h(\mathbb{Z}_n^{\text{half}})\right| \end{aligned} \tag{3.26}$$

where we kept the notation $\overline{\mathcal{W}}_{n,d}^{\text{half}}$ for the random vector
$(0, W_{1,2}, \dots, W_{1,n}, W_{2,1}, 0, W_{2,3}, \dots, W_{2,n}, \dots, W_{n-1,n}, 0)$. It is easy to see that, since h is Lipschitz,

$$\begin{aligned}&\left|\mathbf{E}h(\mathcal{W}_{n,d}^{\text{half}}) - \mathbf{E}h(\overline{\mathcal{W}}_{n,d}^{\text{half}})\right| \leq C\mathbf{E}\left[\left(\sum_{i=1}^{n} W_{i.i}^2\right)^{\frac{1}{2}}\right] \\ &\leq C\left[\mathbf{E}\left(\sum_{i=1}^{n} W_{i.i}^2\right)\right]^{\frac{1}{2}} = \frac{1}{d}\left[\mathbf{E}\sum_{i=1}^{n}\left(\sum_{k=1}^{d}(X_{i,k}^2 - 1)\right)^2\right]^{\frac{1}{2}} = (m_4 - 1)\sqrt{n}\end{aligned} \tag{3.27}$$

and a similar estimate holds for $\left|\mathbf{E}h(\overline{\mathbb{Z}}_n^{\text{half}}) - \mathbf{E}h(\mathbb{Z}_n^{\text{half}})\right|$. By plugging (3.25) and (3.27) into (3.26), we would get $\left|\mathbf{E}h(\mathcal{W}_{n,d}^{\text{half}}) - \mathbf{E}h(\mathbb{Z}_n^{\text{half}})\right| \leq C\left(\sqrt{n} + \sqrt{\frac{n^3}{d}}\right)$ which is, for large enough d, a worse estimate than (3.24). We also refer to Remark 1.5 in Fang and Koike (2021+) for similar estimates when the covariance matrix of $\mathbb{Z}_n$ is not invertible.

We state our main result.

Theorem 3.8. *Consider the Wishart matrix $\mathcal{W}_{n,d}$ with entries (3.15) and (3.16) and assume* **H1** *-* **H3***. Then for every $n \geq 1$, the matrix $\mathcal{W}_{n,d}$ converges compontwise in law, as $d \to \infty$, to the Wigner matrix Z_n defined by (3.17) and for every $n, d \geq 1$, there exists $C > 0$ such that*

$$d_W(\mathcal{W}_{n,d}, \mathbb{Z}_n) \leq C\frac{n^{\frac{9}{4}}}{d^{\frac{1}{4}}}$$

Proof: By Lemma 2.1 we have, since the dimension of the half-vector associated to $\mathcal{W}_{h,d}$ is $\frac{n(n+1)}{2}$,

$$d_W(\mathcal{W}_{n,d}, \mathbb{Z}_n) \leq Cn^{\frac{3}{2}}\left(\sum_{i,j,k,l=1}^{n}\mathbf{E}\left(C_{i,j} - \langle D\tilde{W}_{i,j}, V_{k,l}\rangle_{L^2(T)}\right)^2\right)^{\frac{1}{4}}$$

and by Proposition 3.5, we obtain, since the cardinal of the set $\overline{M_{i,j,a,b}}$ is less than $6n^3$,

$$d_W(\mathcal{W}_{n,d}, \mathbb{Z}_n) \leq C\frac{n^{\frac{9}{4}}}{d^{\frac{1}{4}}}.$$

□

In the literature, one usually says that $\mathcal{W}_{n,d}$ is "Φ-close to $\mathbb{Z}_n$", where $\Phi_{n,d} = \frac{n^{\frac{9}{4}}}{d^{\frac{1}{4}}}$. As commented in the introduction, we have lost some speed with respect to the classical bound $C\sqrt{\frac{n^3}{d}}$. This is due to the need to use the d_2 distance between random vectors.

4. Examples

We present few examples of random matrices to which our main result can be applied.

4.1. *Random entries in a finite sum of Wiener chaoses.* Let us consider a starting matrix $\mathcal{X}_{n,d} = (X_{i,j})_{1\leq i\leq n, 1\leq j\leq d}$ such that every random variable $X_{i,j}$ can be expanded into a finite sum of Wiener chaos, i.e. for every $1 \leq i \leq n$ and $1 \leq j \leq d$ we have, with $N \geq 1$ integer

$$X_{i,j} = \sum_{k=1}^{N} I_{q_{i,j}^{(k)}}\left(f_{i,j}^{k}\right) \text{ with } f_{i,j}^{(k)} \in L_S^2(T^{q_{i,j}^{(k)}})$$

where $q_{i,j}^{(k)} \geq 1$ are interger numbers for every $1 \leq i \leq n, 1 \leq j \leq d$ and $1 \leq k \leq N$. Assume that the family of random variables $(F_{i,j}, 1 \leq i \leq n, 1 \leq j \leq d)$ are independent where

$$F_{i,j} = \left(I_{q_{i,j}^{(k)}}, k = 1, ..., N\right).$$

This ensures the strong indeoendence of the entries of the matrix $\mathcal{X}_{n,d}$ (assumption **H1**). We need to assume (3.3) and (3.4) in order that **H2** holds true. Moreover, it is well-known that the assumption **H3** is satisfied for variables in a finite sum of Wiener chaoses (actually we have $X_{i,j}, D(-L)^{-1}X_{i,j} \in \mathbb{D}^{k,p}$ for every $k \geq 1$ and $p \geq 2$).

This example contains as a particular case a result in Bourguin et al. (2021) (where the entries of $\mathcal{X}_{n,d}$ are assume to be in a Wiener chaos of fixed order).

4.2. *Explicit probability laws in a finite sum of Wiener chaoses.* A particular case of the previous example is when the elements $X_{i,j}$ have the same probability distribution. For example, we can consider

$$X_{i,j} = \frac{1}{\sqrt{3}}\left(W(h_{i,j}) + W(g_{i,j})^2 - 1\right)$$

where W is a Wiener process and $(h_{i,j}, g_{i,j}, 1 \leq i \leq n, 1 \leq j \leq d)$ constitutes a family of orthogonal elements in $L^2(T)$. We can also write

$$X_{i,j} = \frac{1}{\sqrt{3}}\left(I_1(h_{i,j}) + I_2(g_{i,j}^{\otimes 2})\right).$$

Then $(X_{i,j}, 1 \leq i \leq n, 1 \leq j \leq d)$ is a family of strongly independent random variable and assumptions **H2**-**H3** are also verified.

4.3. *Random variables with an infinite chaos expansion.* It is possible to provide examples of random variables with an infinite chaotic decomposition which satisfy **H1**-**H3**. Let for $1 \leq i \leq n, 1 \leq j \leq d$

$$Y_{i,j} = e^{W(A_{i,j}) - \frac{1}{2}} - 1 \text{ and } X_{i,j} = \frac{Y_{i,j}}{(\mathbf{E}Y_{i,j}^2)^{\frac{1}{2}}}$$

where $A_{i,j}$ are disjoint intervals of length one and $W(A_{i,j}) = \int_{A_{i,j}} dW_s$. Then for every $1 \le i \le n, 1 \le j \le d$, we have (see e.g. Nualart, 2006)

$$X_{i,j} = \sum_{n\ge1} \frac{1}{n!} I_n(1_{A_{i,j}}^{\otimes n})$$

and therefore $(X_{i,j}, 1 \le i \le n, 1 \le j \le d)$ is a family of strongly independent random variables with infinite chaos expansion. It is easy to see that **H2** holds true while to check **H3**, we can use for example the bound (2.9).

Acknowledgements

The authors would like to thank an anonymous referee for having carefully read the preliminary version and for her/his valuable comments.

References

Bourguin, S., Diez, C.-P., and Tudor, C. A. Limiting behavior of large correlated Wishart matrices with chaotic entries. *Bernoulli*, **27** (2), 1077–1102 (2021). MR4255227.

Bubeck, S., Ding, J., Eldan, R., and Rácz, M. Z. Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms*, **49** (3), 503–532 (2016). MR3545825.

Bubeck, S. and Ganguly, S. Entropic CLT and phase transition in high-dimensional Wishart matrices. *Int. Math. Res. Not. IMRN*, (2), 588–606 (2018). MR3801440.

Diez, C.-P. and Tudor, C. A. Noncentral limit theorem for large Wishart matrices with Hermite entries. *J. Stoch. Anal.*, **2** (1), Art. 2, 13 (2021). MR4213892.

Fang, X. and Koike, Y. New Error Bounds in Multivariate Normal Approximations with Exchangeable Pairs with Applications to Wishart matrices and Fourth Moment Theorems (2021+). To appear in *Ann. of Appl. Probab.*

Huang, J., Nualart, D., and Viitasaari, L. A central limit theorem for the stochastic heat equation. *Stochastic Process. Appl.*, **130** (12), 7170–7184 (2020). MR4167203.

Jiang, T. and Li, D. Approximation of rectangular beta-Laguerre ensembles and large deviations. *J. Theoret. Probab.*, **28** (3), 804–847 (2015). MR3413957.

Mikulincer, D. A CLT in Stein's distance for generalized Wishart matrices and higher order tensors. *Preprint* (2020).

Nourdin, I. and Peccati, G. *Normal approximations with Malliavin calculus. From Stein's method to universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge (2012). ISBN 978-1-107-01777-1. MR2962301.

Nourdin, I. and Zheng, G. Asymptotic behavior of large Gaussian correlated Wishart matrices. *Preprint* (2019).

Nualart, D. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition (2006). ISBN 978-3-540-28328-7; 3-540-28328-5. MR2200233.

Rácz, M. Z. and Richey, J. A smooth transition from Wishart to GOE. *J. Theoret. Probab.*, **32** (2), 898–906 (2019). MR3959632.

Tudor, C. A. and Yoshida, N. High order asymptotic expansion for Wiener functionals. *Preprint* (2018).

Üstünel, A. S. and Zakai, M. On independence and conditioning on Wiener space. *Ann. Probab.*, **17** (4), 1441–1453 (1989). MR1048936.

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