Universal Gaussian fluctuations of non-Hermitian matrix ensembles

Ivan Nourdin, Giovanni Peccati

Universal Gaussian fluctuations of non-Hermitian matrix ensembles

by Ivan Nourdin[1] and Giovanni Peccati[2]
Université Paris VI and Université Paris Ouest

Abstract. In the paper [19], written in collaboration with Gesine Reinert, we proved a universality principle for the Gaussian Wiener chaos. In the present work, we aim at providing an original example of application of this principle in the framework of random matrix theory. More specifically, by combining the result in [19] with some combinatorial estimates, we are able to prove multi-dimensional central limit theorems for the spectral moments (of arbitrary degrees) associated with random matrices with real-valued i.i.d. entries, satisfying some appropriate moment conditions. Our approach has the advantage of yielding, without extra effort, bounds over classes of smooth (i.e., thrice differentiable) functions. Moreover, it allows to deal directly with discrete distributions.

Key words: Central limit theorems; Eigenvalues; Fourth-moment criteria; Invariance principles; Non-Hermitian random matrices; Normal approximation; Spectral moments; Universality.

2000 Mathematics Subject Classification: 60F05; 60G15; 60H05; 60H07.

1 Introduction

1.1 Overview and main results

In the paper [19], written in collaboration with Gesine Reinert, we proved several *universality results*, involving sequences of random vectors whose components have the form of finite homogeneous sums based on sequences of independent random variables. Roughly speaking, our main finding implied that, in order to study the normal approximations of homogeneous sums (and under suitable moment conditions) it is always possible to replace the original sequence with an i.i.d. Gaussian family. The power of this approach resides in the fact that homogeneous sums associated with Gaussian sequences are indeed elements of the so-called *Wiener chaos*, so that normal approximations can be established by means of the general techniques developed in [18, 22, 23] – that are based on a powerful interaction between standard Gaussian analysis, *Malliavin calculus* (see e.g. [21]) and *Stein's method* (see e.g. [6]). Moreover, in the process one always recovers uniform bounds over suitable classes of smooth functions.

The aim of this paper is to introduce these techniques into the realm of random matrix theory. More specifically, our goal is to use the universality principles developed in [19], in

[1]Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, Boîte courrier 188, 4 Place Jussieu, 75252 Paris Cedex 5, France. Email: ivan.nourdin@upmc.fr

[2]Equipe Modal'X, Université Paris Ouest – Nanterre la Défense, 200 Avenue de la Rpublique, 92000 Nanterre, and LSTA, Université Paris VI, France. Email: giovanni.peccati@gmail.com

order to prove the forthcoming Theorem 1.1, which consists in a multidimensional central limit theorem (CLT) for traces of non-Hermitian random matrices with i.i.d. real-valued entries. More precisely, let X be a centered real random variable, having unit variance and with finite moments of all orders, that is, $E(X) = 0$, $E(X^2) = 1$ and $E|X|^n < \infty$ for every $n \geqslant 3$. We consider a doubly indexed collection $\mathbf{X} = \{X_{ij} : i, j \geqslant 1\}$ of i.i.d. copies of X. For every integer $N \geqslant 2$, we denote by X_N the $N \times N$ random matrix

$$X_N = \left\{ \frac{X_{ij}}{\sqrt{N}} : i, j = 1, ..., N \right\}, \tag{1.1}$$

and by $\mathrm{Tr}(\cdot)$ and X_N^k, respectively, the usual trace operator and the kth power of X_N.

Theorem 1.1 *Let the above notation prevail. Fix $m \geqslant 1$, as well as integers*

$$1 \leqslant k_1 < \ldots < k_m.$$

Then, the following holds.

(i) *As $N \to \infty$,*

$$\left(\mathrm{Tr}(X_N^{k_1}) - E\left[\mathrm{Tr}(X_N^{k_1})\right], \ldots, \mathrm{Tr}(X_N^{k_m}) - E\left[\mathrm{Tr}(X_N^{k_m})\right] \right) \xrightarrow{\text{Law}} \left(Z_{k_1}, ..., Z_{k_m}\right), \tag{1.2}$$

where $\mathbf{Z} = \{Z_k : k \geqslant 1\}$ denotes a collection of real independent centered Gaussian random variables such that, for every $k \geqslant 1$, $E(Z_k^2) = k$.

(ii) *Write $\beta = E|X|^3$. Suppose that the function $\varphi : \mathbb{R}^m \to \mathbb{R}$ is thrice differentiable and that its partial derivatives up to the order three are bounded by some constant $B < \infty$. Then, there exists a finite constant $C = C(\beta, B, m, k_1, ..., k_m)$, not depending on N, such that*

$$\left| E\left[\varphi\left(\frac{\mathrm{Tr}(X_N^{k_1}) - E[\mathrm{Tr}(X_N^{k_1})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_1}))}}, \ldots, \frac{\mathrm{Tr}(X_N^{k_m}) - E[\mathrm{Tr}(X_N^{k_m})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_m}))}} \right) \right] \right. \tag{1.3}$$
$$\left. - E\left[\varphi\left(\frac{Z_{k_1}}{\sqrt{k_1}}, ..., \frac{Z_{k_m}}{\sqrt{k_m}} \right) \right] \right| \leqslant C\, N^{-1/4}.$$

Remark 1.2 1. We chose to state and prove Theorem 1.1 in the case of non-Hermitian matrices with *real-valued* entries, mainly in order to facilitate the connection with the universality results proved in [19]. However, our techniques may be extended to the case where the random variable X is *complex-valued* and with finite absolute moments of every order. This line of research will be pursued elsewhere. One should also note that, differently from [25], in the present paper we do not use any technique coming from complex analysis.

2. Fix an integer $K \geqslant 2$ and assume that $E|X|^{2K} < \infty$, while higher moments are allowed to be possibly infinite. By inspection of the forthcoming proof of Theorem 1.1, one sees that the CLT (1.2) as well as the bound (1.4) continue to hold, as long as the integers $k_1, ..., k_m$ verify $k_j \leqslant K$ for $j = 1, ..., m$.

3. In a similar vein as at the previous point, by imposing adequate uniform bounds on moments one can easily adapt our techniques in order to deal with random matrices whose entries are independent but not identically distributed. One crucial fact supporting this claim is that the universality principles of Section 2 hold for collections of independent, and not necessarily identically distributed, random variables.

4. For non-Hermitian matrices, limits of moments are not sufficient to provide an exhaustive description of the limiting spectral measure or of the fluctuations around it. Rather, one would need to consider polynomials in the eigenvalues and their complex conjugates. These quantities cannot be represented using traces of powers of X_N, so that our approach cannot be extended to this case.

The upper bound appearing in Theorem 1.1-(ii) is based on the forthcoming Theorem 2.7, which hinges in turn on some intricated combinatorial estimates taken from [19]. However, when X is a standard Gaussian random variable one can directly use the results of [20, Theorem 3.5 and Lemma 3.7], and obtain an even better rate of convergence (in the stronger Wasserstein distance). This fact is described in the following statement (the proof is left to the reader).

Proposition 1.3 *Assume that X is a centered standard Gaussian random variable. Suppose that the function $\varphi : \mathbb{R}^m \to \mathbb{R}$ is differentiable and that its (first) partial derivative is bounded by some constant $B < \infty$. Then, there exists a finite constant $C = C(B, m, k_1, ..., k_m)$, not depending on N, such that*

$$\left| E\left[\varphi\left(\frac{\mathrm{Tr}(X_N^{k_1}) - E[\mathrm{Tr}(X_N^{k_1})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_1}))}}, \ldots, \frac{\mathrm{Tr}(X_N^{k_m}) - E[\mathrm{Tr}(X_N^{k_m})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_m}))}}\right)\right] - E\left[\varphi\left(\frac{Z_{k_1}}{\sqrt{k_1}}, ..., \frac{Z_{k_m}}{\sqrt{k_m}}\right)\right] \right| \leqslant C\, N^{-1/2}. \tag{1.4}$$

1.2 Discussion

In this section we compare our Theorem 1.1 with some related results proved in the existing probabilistic literature.

1. In the paper [25], Rider and Silverstein proved the following CLT.

Theorem 1.4 *Let X be a* complex *random variable such that $E(X) = E(X^2) = 0$, $E(|X|^2) = 1$, $E(|X|^k) \leqslant k^{\alpha k}$, $k \geqslant 3$ (for some $\alpha > 0$) and* $\mathrm{Re}(X)$, $\mathrm{Im}(X)$ *possess a*

joint bounded density. For $N \geqslant 2$, let X_N be defined as in (1.1). Consider the space $\mathcal{H}$ of functions $f : \mathbb{C} \to \mathbb{C}$ which are analytic in a neighborhood of the disk $|z| \leqslant 4$ and otherwise bounded. Then, as $N \to \infty$, the random field

$$\{\mathrm{Tr}(f(X_N)) - E\left[\mathrm{Tr}(f(X_N))\right] : \ f \in \mathcal{H}\}$$

converges in the sense of finite-dimensional distributions (f.d.d.) to the centered complex-valued Gaussian field $\{Z(f) : \ f \in \mathcal{H}\}$, whose covariance structure is given by

$$E[Z(f)\overline{Z(g)}] = \int_{\mathbb{U}} f'(z)\overline{g'(z)}\frac{d^2z}{\pi}. \tag{1.5}$$

Here, $\mathbb{U} = \{z \in \mathbb{C} : |z| \leqslant 1\}$ is the unit disk, and d^2z/π stands for the uniform measure on $\mathbb{U}$ (in other words, $d^2z = dxdy$ for $x, y \in \mathbb{R}$ such that $z = x + iy$).

By using the elementary relations: for every integers $n, m \geqslant 0$,

$$\frac{1}{\pi}\int_{\mathbb{U}} z^n\overline{z}^m d^2z = \begin{cases} (n+1)^{-1} & \text{if } m = n \\ 0 & \text{otherwise,} \end{cases}$$

one sees that our Theorem 1.1 can be reformulated by saying that

$$\{\mathrm{Tr}(f(X_N)) - E\left[\mathrm{Tr}(f(X_N))\right] : \ f \in \mathrm{Pol}(\mathbb{C})\} \xrightarrow{\text{f.d.d.}} \{Z(f) : \ f \in \mathrm{Pol}(\mathbb{C})\}, \tag{1.6}$$

where the covariance structure of $\{Z(f) : \ f \in \mathrm{Pol}(\mathbb{C})\}$ is given by (1.5). It follows that Theorem 1.1 *roughly* agrees with Theorem 1.4. however, we stress that the framework of [25] is different from ours, since the findings therein cannot be applied to the real case due to the assumption that real and imaginary parts of entries must possess a joint bounded density. In addition, also note that (differently from [25]) we do not introduce in the present paper any requirement on the absolute continuity of the law of the real random variable X, so that the framework of our Theorem 1.1 contemplates every discrete random variable with values in a finite set and with unit variance.

2. One should of course compare the results of this paper with the CLTs involving traces of *Hermitian* random matrices, like for instance Wigner random matrices. One general reference in this direction is the fundamental paper by Anderson and Zeitouni [3], where the authors obtain CLTs for traces associated with large classes of (symmetric) band matrix ensembles, using a version of the classical method of moments based on graph enumerations. It is plausible that some of the findings of the present paper could be also deduced from a suitable extension of the combinatorial devices introduced in [3] to the case of non-Hermitian matrices. However, proving Theorem 1.1 using this kind of techniques would require estimates for arbitrary joint moments of traces, whereas our approach merely requires the computation of variances and fourth moments. Also, the findings of [3] do not allow to directly deduce bounds such as (1.4). We refer the reader e.g. to Guionnet [14] or to Anderson *et al.* [2], and the references therein, for a detailed overview of existing

asymptotic results for large Hermitian random matrices.

3. The general statement proved by Chatterjee in [5, Theorem 3.1] concerns the normal approximation of linear statistics of random matrices that are possibly non-Hermitian. However, the techniques used by the author require that the entries can be re-written as smooth transformations of Gaussian random variables. In particular, the findings of [5] do not apply to discrete distributions. On the other hand, the results of [5] also provide uniform bounds (based on Poincaré-type inequalities and in the total variation distance) for one-dimensional CLTs. Here, we do not introduce any requirements on the absolute continuity of the law of the real random variable X, and we get bounds for *multi*-dimensional CLTs.

4. Let us denote by $\{\lambda_j(N) : j = 1, ..., N\}$ the complex-valued (random) eigenvalues of X_N, repeated according to their multiplicities. Theorem 1.1 deals with the spectral moments of X_N, that are defined by the relations:

$$N \times \int z^k d\mu_{X_N}(z) = \sum_{j=1}^{N} \lambda_j(N)^k = \mathrm{Tr}(X_N^k), \quad N \geqslant 2, \quad k \geqslant 1, \tag{1.7}$$

where μ_{X_N} denote the spectral measure of X_N. Recall that

$$\mu_{X_N}(\cdot) = \frac{1}{N} \sum_{j=1}^{N} \delta_{\lambda_j(N)}(\cdot), \tag{1.8}$$

where $\delta_z(\cdot)$ denotes the Dirac mass at z, and observe that one has also the alternate expression

$$\mathrm{Tr}(X_N^k) = N^{-\frac{k}{2}} \sum_{i_1,...,i_k=1}^{N} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}. \tag{1.9}$$

It follows that our Theorem 1.1 can be seen as a partial (see Remark 1.2 (4) above) characterization of the Gaussian fluctuations associated with the so-called *circular law*, whose most general version has been recently proved by Tao and Vu:

Theorem 1.5 (Circular law, see [29]) *Let X be a complex-valued random variable, with mean zero and unit variance. For $N \geqslant 2$, let X_N be defined as in (1.1). Then, as $N \to \infty$, the spectral measure μ_{X_N} converges almost surely to the uniform measure on the unit disk $\mathbb{U} = \{z \in \mathbb{C} : |z| \leqslant 1\}$. The convergence takes place in the sense of the vague topology.*

To see why Theorem 1.1 concerns fluctuations around the circular law, one can proceed as follows. First observe that, since $E(X^2) = 1$ and $E(X^4) < \infty$ by assumption, one can use a result by Bai and Yin [4, Theorem 2.2] stating that, with probability one,

$$\limsup_{N \to \infty} \max_{j=1,...,N} |\lambda_j(N)| \leqslant 1. \tag{1.10}$$

Now fix a polynomial $p(z)$. Elementary considerations yield that, since (1.10) and the circular law are in order, with probability one

$$\frac{1}{N}\mathrm{Tr}(p(X_N)) \to \frac{1}{\pi}\int_{\mathbb{U}} p(z)d^2z = p(0). \tag{1.11}$$

On the other hand, it is not difficult to see that, for every $k \geqslant 1$ and as $N \to \infty$,

$$E\left[\int z^k d\mu_{X_N}(z)\right] = E\left[\frac{1}{N}\mathrm{Tr}(X_N^k)\right] \to 0$$

(one can use e.g. the same arguments exploited in the second part of the proof Proposition 3.1 below). This implies in particular, for every complex polynomial p,

$$E\left[\frac{1}{N}\mathrm{Tr}(p(X_n))\right] \to p(0) = \frac{1}{\pi}\int_{\mathbb{U}} p(z)d^2z. \tag{1.12}$$

By (1.11) and (1.12), one has therefore that the quantities $\frac{1}{N}\mathrm{Tr}(p(X_N))$ and $E(\frac{1}{N}\mathrm{Tr}(p(X_N)))$ both converge to $p(0)$, and (1.6) ensures that, for N sufficiently large, the difference

$$\mathrm{Tr}(p(X_N)) - Np(0) - [E\left(\mathrm{Tr}(p(X_N))\right) - Np(0)]$$

has approximately a centered Gaussian distribution with variance $\frac{1}{\pi}\int_{\mathbb{U}} |p'(z)|^2 d^2z$. Equivalently, one can say that the random variable $\frac{1}{N}\mathrm{Tr}(p(X_N))$ tends to concentrate around its mean as N goes to infinity, and (1.6) describes the Gaussian fluctuations associated with this phenomenon.

On the other hand, one crucial feature of the proof of the circular law provided in [29] is that it is based on a universality principle. This result basically states that, under adequate conditions, the distance between the spectral measures of (possibly perturbed) non-Hermitian matrices converges systematically to zero, so that Theorem 1.5 can be established by simply focussing on the case where X is complex Gaussian (this is the so-called Ginibre matrix ensemble, first introduced in [13]). It is interesting to note that our proof of Theorem 1.1 is also based on a universality result. Indeed, we shall show that the relevant part of the vector on the LHS of (1.2) (that is, the part not vanishing at infinity) has the form of a collection of homogeneous sums with fixed orders. This implies that the CLT in (1.2) can be deduced from the results established in [19], where it is proved that the Gaussian Wiener chaos has a universal character with respect to Gaussian approximations. Roughly speaking, this means that, in order to prove a CLT for a vector of general homogeneous sums, it is sufficient to consider the case where the summands are built from an i.i.d. Gaussian sequence. This phenomenon can be seen as a further instance of the so-called Lindeberg invariance principle for probabilistic approximations, and stems from powerful approximation results by Rotar' [27] and Mossel *et al.* [17]. See the forthcoming Section 2 for precise statements.

5. We finish this section by listing and discussing very briefly some other results related to Theorem 1.1, taken from the existing probabilistic literature.

- In Rider [24] (but see also Forrester [10]), one can find a CLT for (possibly discontinuous) linear statistics of the eigenvalues associated with complex random matrices in the Ginibre ensemble. This partially builds on previous findings by Costin and Lebowitz [7].

- Reference [26], by Rider and Virag, provides further insights into limit theorems involving sequences in the complex Ginibre ensemble. In particular, one sees that relaxing the assumption of analyticity on test functions yields a striking decomposition of the variance of the limiting noise, into the sum of a "bulk" and of a "boundary" term. Another finding in [26] is an asymptotic characterization of characteristic polynomials, in terms of the so-called *Gaussian free field*.

- Finally, one should note that the Gaussian sequence $\mathbf{Z}$ in Theorem 1.1 also appears when dealing with Gaussian fluctutations of vectors of traces associated with large, Haar-distributed unitary random matrices. See e.g. [8] and [9] for two classic references on the subject.

1.3 Proof of Theorem 1.1: the strategy

In order to prove (1.2) (and (1.4) as well), we use an original combination of techniques, which are based both on the universality results of [19] and on combinatorial considerations. The aim of this section is to provide a brief outline of this strategy.

For $N \geqslant 1$, write $[N] = \{1, ..., N\}$. For $k \geqslant 2$, let us denote by $D_N^{(k)}$ the collection of all vectors $\mathbf{i} = (i_1, \ldots, i_k) \in [N]^k$ such that all pairs (i_a, i_{a+1}), $a = 1, \ldots, k$, are different (with the convention that $i_{k+1} = i_1$), that is, $\mathbf{i} \in D_N^{(k)}$ if and only if $(i_a, i_{a+1}) \neq (i_b, i_{b+1})$ for every $a \neq b$. Now consider the representation given in (1.9) and, after subtracting the expectation, rewrite the resulting expression as follows:

$$\begin{aligned} & \mathrm{Tr}(X_N^k) - E\left[\mathrm{Tr}(X_N^k)\right] \\ = \; & N^{-\frac{k}{2}} \sum_{i_1,\ldots,i_k=1}^{N} \left(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1} - E[X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}]\right) \qquad (1.13) \\ = \; & N^{-\frac{k}{2}} \sum_{\mathbf{i} \in D_N^{(k)}} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1} \\ & \qquad + N^{-\frac{k}{2}} \sum_{\mathbf{i} \notin D_N^{(k)}} \left(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1} - E[X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}]\right). \qquad (1.14) \end{aligned}$$

Our proof of (1.2) is based on the representation (1.13)–(1.14), and it is divided in two (almost independent) parts.

I. In Section 3, we shall prove that the following multi-dimensional CLT takes place for

every integers $2 \leqslant k_1 < ... < k_m$:

$$\left(N^{-1/2} \sum_{i=1}^{N} X_{ii},\ N^{-\frac{k_1}{2}} \sum_{\mathbf{i} \in D_N^{(k_1)}} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k_1} i_1},\ \ldots \right. \tag{1.15}$$

$$\left. \ldots,\ N^{-\frac{k_m}{2}} \sum_{\mathbf{i} \in D_N^{(k_m)}} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k_m} i_1} \right) \xrightarrow{\text{Law}} \big(Z_1, Z_{k_1}, ..., Z_{k_m}\big),$$

for $\mathbf{Z} = \{Z_i : i \geqslant 1\}$ as in Theorem 1.1. In order to prove (1.15), we apply the universality result obtained in [19] (and stated in a convenient form in the subsequent Section 2). This result roughly states that, in order to show (1.15) in full generality, it is sufficient to consider the special case where the collection $\mathbf{X} = \{X_{ij} : i, j \geqslant 1\}$ is replaced by an i.i.d. centered Gaussian family $\mathbf{G} = \{G_{ij} : i, j \geqslant 1\}$, whose elements have unit variance. In this way, the components of the vector on the LHS of (1.15) become elements of the so-called *Gaussian Wiener chaos* associated with $\mathbf{G}$: it follows that one can establish the required CLT by using the general criteria for normal approximations on a fixed Wiener chaos, recently proved in [18, 22, 23]. Note that the results of [18, 22, 23] can be described as a "simplified method of moments": in particular, the proof of (1.15) will require the mere computation of quantities having the same level of complexity of covariances and fourth moments.

II. In Section 4, we shall prove that the term (1.14) vanishes as $N \to \infty$, that is, for every $k \geqslant 2$,

$$N^{-\frac{k}{2}} \sum_{\mathbf{i} \notin D_N^{(k)}} \big(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1} - E[X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}] \big) \to 0 \quad \text{in } L^2(\Omega). \tag{1.16}$$

The proof of (1.16) requires some subtle combinatorial analysis, that we will illustrate by means of graphical devices, known as *diagrams*. Some of the combinatorial arguments and ideas developed in Section 4 should be compared with the two works by Geman [11, 12].

Then, the upper bound (1.4) will be deduced in Section 4.4 from the estimates obtained at the previous steps.

The rest of the paper is organized as follows. In Section 2 we present the universality results proved in [19], in a form which is convenient for our analysis. Section 3 contains a proof of (1.15), whereas Section 4 deals with (1.16).

2 Main tool: universality of Wiener chaos

In what follows, every random object is defined on an adequate common probability space $(\Omega, \mathscr{F}, P)$. The symbols E and 'Var' denote, respectively, the expectation and the variance associated with P. Also, given a finite set B, we write $|B|$ to indicate the cardinality of B. Finally, given numerical sequences a_N, b_N, $N \geqslant 1$, we write $a_N \sim b_N$ whenever $a_N/b_N \to 1$ as $N \to \infty$.

We shall now present a series of invariance principles and central limit theorems involving sequences of homogeneous sums. These are mainly taken from [19] (Theorem 2.2), [23] (Theorem 2.4) and [22] (Theorem 2.6). Note that the framework of [19] is that of random variables indexed by the set of positive integers. Since in this paper we mainly deal with random variables indexed by *pairs* of integers (i.e., matrix entries) we need to restate some of the findings of [19] in terms of random variables indexed by a general (fixed) discrete countable set A.

Definition 2.1 (Homogeneous sums) Fix an integer $k \geqslant 2$. Let $\mathbf{Y} = \{Y_a : a \in A\}$ be a collection of square integrable and centered independent random variables, and let $f : A^k \to \mathbb{R}$ be a *symmetric function vanishing on diagonals* (that is, $f(a_1, ..., a_k) = 0$ whenever there exists $k \neq j$ such that $a_k = a_j$), and assume that f has finite support. The random variable

$$Q_k(f, \mathbf{Y}) = \sum_{a_1,...,a_k \in A} f(a_1, ..., a_k) Y_{a_1} \cdots Y_{a_k} = \sum_{\{a_1,...,a_k\} \subset A^k} k! f(a_1, ..., a_k) Y_{a_1} \cdots Y_{a_k} \tag{2.17}$$

is called the *homogeneous sum*, of order k, based on f and $\mathbf{Y}$. Clearly, $E[Q_k(f, \mathbf{Y})] = 0$ and also, if $E(Y_a^2) = 1$ for every $a \in A$, then

$$E[Q_k(f, \mathbf{Y})^2] = k! \|f\|_k^2, \tag{2.18}$$

where, here and for the rest of the paper, we set

$$\|f\|_k^2 = \sum_{a_1,...,a_k \in A} f^2(a_1, ..., a_k).$$

Now let $\mathbf{G} = \{G_a : a \in A\}$ be a collection of i.i.d. centered Gaussian random variables with unit variance. We recall that, for every k and every f, the random variable $Q_k(f, \mathbf{G})$ (defined according to (2.17)) is an element of the kth *Wiener chaos* associated with $\mathbf{G}$. See e.g. Janson [15] for basic definitions and results on the Gaussian Wiener chaos. The next result, proved in [19], shows that sequences of random variables of the type $Q_k(f, \mathbf{G})$ have a *universal character* with respect to normal approximations. The proof of Theorem 2.2 is based on a powerful interaction between three techniques, namely: the *Stein's method* for probabilistic approximations (see e.g. [6]), the *Malliavin calculus of variations* (see e.g. [21]), and a general Lindeberg-type invariance principle recently proved by Mossel *et al.* in [17].

Theorem 2.2 (Universality of Wiener chaos, see [19]) *Let* $\mathbf{G} = \{G_a : a \in A\}$ *be a collection of standard centered i.i.d. Gaussian random variables, and fix integers* $m \geqslant 1$ *and* $k_1, ..., k_m \geqslant 2$. *For every* $j = 1, ..., m$, *let* $\{f_N^{(j)} : N \geqslant 1\}$ *be a sequence of functions such that* $f_N^{(j)} : A^{k_j} \to \mathbb{R}$ *is symmetric and vanishes on diagonals. We also suppose that, for every* $j = 1, ..., m$, *the support of* $f_N^{(j)}$, *denoted by* $\mathrm{supp}(f_N^{(j)})$, *is such that* $|\mathrm{supp}(f_N^{(j)})| \to \infty$, *as* $N \to \infty$. *Define* $Q_{k_j}(f_N^{(j)}, \mathbf{G})$, $N \geqslant 1$, *according to (2.17). Assume that, for every* $j = 1, ...m$, *the following sequence of variances is bounded:*

$$E[Q_{k_j}(f_N^{(j)}, \mathbf{G})^2], \quad N \geqslant 1. \tag{2.19}$$

Let V *be a* $m \times m$ *non-negative symmetric matrix, and let* $\mathscr{N}_m(0, V)$ *indicate a* m*-dimensional centered Gaussian vector with covariance matrix* V. *Then, as* $N \to \infty$, *the following two conditions are equivalent.*

(1) *The vector* $\{Q_{k_j}(f_N^{(j)}, \mathbf{G}) : j = 1, ..., m\}$ *converges in law to* $\mathscr{N}_m(0, V)$.

(2) *For every sequence* $\mathbf{X} = \{X_a : a \in A\}$ *of independent centered random variables, with unit variance and such that* $\sup_a E|X_a|^3 < \infty$, *the law of the vector* $\{Q_{k_j}(f_N^{(j)}, \mathbf{X}) : j = 1, ..., m\}$ *converges to the law of* $\mathscr{N}_m(0, V)$.

Note that Theorem 2.2 concerns only homogeneous sums of order $k \geqslant 2$: it is easily seen (see e.g. [19, Section 1.6.1]) that the statement is indeed false in the case $k = 1$. However, if one considers sums with a specific structure (basically, verifying some Lindeberg-type condition) one can embed sums of order one into the previous statement. A particular instance of this fact is made clear in the following statement, whose proof (combining the results of [19] with the main estimates of [17]) is standard and therefore omitted.

Proposition 2.3 *For* $m \geqslant 1$, *let the kernels* $\{f_N^{(j)} : N \geqslant 1\}$, $j = 1, ..., m$, *verify the assumptions of Theorem 2.2. Let* $\{a_i : i \geqslant 1\}$ *be an infinite subset of* A, *and assume that condition (1) in the statement of Theorem 2.2 is verified. Then, for every sequence* $\mathbf{X} = \{X_a : a \in A\}$ *of independent centered random variables, with unit variance and such that* $\sup_a E|X_a|^3 < \infty$, *as* $N \to \infty$ *the law of the vector* $\{W_N\,;\, Q_{k_j}(f_N^{(j)}, \mathbf{X}) : j = 1, ..., m\}$, *where* $W_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_{a_i}$, *converges to the law of* $\{N_0\,;\, N_j : j = 1, ..., m\}$, *where* $N_0 \sim \mathscr{N}(0, 1)$, *and* $(N_1, ..., N_m) \sim \mathscr{N}_m(0, V)$ *denotes a centered Gaussian vector with covariance* V, *and independent of* N_0.

Theorem 2.2 and Proposition 2.3 imply that, in order to prove a CLT involving vectors of homogeneous sums based on some independent sequence $\mathbf{X}$, it suffices to replace $\mathbf{X}$ with an i.i.d. Gaussian sequence $\mathbf{G}$. In this way, one obtains a sequence of random vectors whose components belong to a fixed Wiener chaos. We now present two results, showing that proving CLTs for this type of random variables can be a relatively easy task: indeed, one can apply some drastic simplification of the method of moments. The first statement deals with multi-dimensional CLTs and shows that, in a Gaussian Wiener chaos setting, componentwise convergence to Gaussian always implies joint convergence. See also [1] for some connections with Stokes formula.

Theorem 2.4 (Multidimensional CLTs on Wiener chaos, see [19, 23]) *Let the family* $\mathbf{G} = \{G_a : a \in A\}$ *be i.i.d. centered standard Gaussian and, for* $j = 1, ..., m$, *define the sequences* $Q_{k_j}(f_N^{(j)}, \mathbf{G})$, $N \geqslant 1$, *as in Theorem 2.2 (in particular, the functions* $f_N^{(j)}$ *verify the same assumptions as in that theorem). Suppose that, for every* $i, j = 1, ..., m$, *as* $N \to \infty$

$$E\big[Q_{k_i}(f_N^{(i)}, \mathbf{G}) \times Q_{k_j}(f_N^{(j)}, \mathbf{G})\big] \to V(i, j), \tag{2.20}$$

where V *is a* $m \times m$ *covariance matrix. Finally, assume that* W_N, $N \geqslant 1$, *is a sequence of* $\mathscr{N}(0, 1)$ *random variables with the representation*

$$W_N = \sum_{a \in A} w_N(a) \times G_a,$$

where the weights $w_N(a)$ *are zero for all but a finite number of indices* a, *and* $\sum_{a \in A} w_N(a)^2 = 1$. *Then, the following are equivalent:*

(1) *The random vector* $\{W_N\,;\, Q_{k_j}(f_N^{(j)}, \mathbf{G}) : j = 1, ..., m\}$ *converges in law to* $\{N_0\,;\, N_j : j = 1, ..., m\}$, *where* $N_0 \sim \mathscr{N}(0, 1)$, *and* $(N_1, ..., N_m) \sim \mathscr{N}_m(0, V)$ *denotes a centered Gaussian vector with covariance* V, *and independent of* N_0.

(2) *For every fixed* $j = 1, ..., m$, *the sequence* $Q_{k_j}(f_N^{(j)}, \mathbf{G})$, $N \geqslant 1$, *converges in law to* $Z \sim \mathscr{N}\big(0, V(j, j)\big)$, *that is, to a centered Gaussian random variable with variance* $V(j, j)$.

The previous statement implies that, in order to prove CLTs for vectors of homogeneous sums, one can focus on the componentwise convergence of their (Gaussian) Wiener chaos counterpart. The forthcoming Theorem 2.6 shows that this type of one-dimensional convergence can be studied by focussing exclusively on fourth moments. To put this result into full use, we need some further definitions.

Definition 2.5 Fix $k \geqslant 2$. Let $f : A^k \to \mathbb{R}$ be a (not necessarily symmetric) function vanishing on diagonals and with finite support. For every $r = 0, ..., k$, the *contraction* $f \star_r f$ is the function on A^{2d-2r} given by

$$\begin{aligned} &f \star_r f(a_1, ..., a_{2d-2r}) \\ &= \sum_{(x_1, ..., x_r) \in A^r} f(a_1, ..., a_{k-r}, x_1, ..., x_r) f(a_{k-r+1}, ..., a_{2d-2r}, x_1, ..., x_r). \end{aligned} \tag{2.21}$$

Observe that (even when f is symmetric) the contraction $f \star_r f$ is not necessarily symmetric and not necessarily vanishes on diagonals. The canonical symmetrization of $f \star_r f$ is written $f \widetilde{\star}_r f$.

Theorem 2.6 (The simplified method of moments, see [22]) *Fix* $k \geqslant 2$. *Let* $\mathbf{G} = \{G_a : a \in A\}$ *be an i.i.d. centered standard Gaussian family. Let* $\{f_N : N \geqslant 1\}$ *be a*

sequence of functions such that $f_N : A^k \to \mathbb{R}$ is symmetric and vanishes on diagonals. Suppose also that $|\mathrm{supp}(f_N)| \to \infty$, as $N \to \infty$. Assume that

$$E[Q_k(f_N, \mathbf{G})^2] \to \sigma^2 > 0, \quad \text{as } N \to \infty. \tag{2.22}$$

Then, the following three conditions are equivalent, as $N \to \infty$.

(1) *The sequence $Q_k(f_N, \mathbf{G})$, $N \geqslant 1$, converges in law to $Z \sim \mathscr{N}(0, \sigma^2)$.*

(2) *$E[Q_k(f_N, \mathbf{G})^4] \to 3\sigma^4$.*

(3) *For every $r = 1, ..., k-1$, $\|f_N \star_r f_N\|_{2k-2r} \to 0$.*

Finally, we present a version of Theorem 2.2 with bounds, that will lead to the proof of Theorem 1.1-(ii) provided in Section 4.4.

Theorem 2.7 (Universal bounds, see [19]) *Let $\mathbf{X} = \{X_a : a \in A\}$ be a collection of independent centered random variables, with unit variance and such that $\beta := \sup_a E|X_a|^3 < \infty$. Fix integers $m \geqslant 1$, $k_m > ... > k_1 \geqslant 2$. For every $j = 1, ..., m$, let $f^{(j)} : A^{k_j} \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Define $Q^j(\mathbf{X}) := Q_{k_j}(f^{(j)}, \mathbf{X})$ according to (2.17), and assume that $E[Q^j(\mathbf{X})^2] = 1$ for all $j = 1, \ldots, m$. Also, assume that $K > 0$ is given such that $\sum_{a \in A} \max_{1 \leqslant j \leqslant m} \mathrm{Inf}_a(f^{(j)}) \leqslant K$, where*

$$\mathrm{Inf}_a(f^{(j)}) = \sum_{\{a_2, ..., a_{k_j}\} \subset A^{k_j}} f^{(j)}(a, a_2, \ldots, a_{k_j})^2 = \frac{1}{(k_j - 1)!} \sum_{a_2, ..., a_{k_j} \in A} f^{(j)}(a, a_2, \ldots, a_{k_j})^2.$$

Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a thrice differentiable function such that $\|\varphi''\|_\infty + \|\varphi'''\|_\infty < \infty$, with $\|\varphi^{(k)}\|_\infty = \max_{|\alpha| = k} \frac{1}{\alpha!} \sup_{z \in \mathbb{R}^m} |\partial^\alpha \varphi(z)|$. Then, for $Z = (Z^1, \ldots, Z^m) \sim \mathscr{N}_m(0, I_m)$ (standard Gaussian vector on $\mathbb{R}^m$), we have

$$\begin{aligned} \big|E[\varphi(Q^1(\mathbf{X}), \ldots, Q^m(\mathbf{X}))] - E[\varphi(Z)]\big| &\leqslant \|\varphi''\|_\infty \left(\sum_{i=1}^m \Delta_{ii} + 2 \sum_{1 \leqslant i < j \leqslant m} \Delta_{ij} \right) \\ &\quad + K \|\varphi'''\|_\infty \left(\beta + \sqrt{\frac{8}{\pi}} \right) \left[\sum_{j=1}^m (16\sqrt{2}\beta)^{\frac{k_j - 1}{3}} k_j! \right]^3 \sqrt{\max_{1 \leqslant j \leqslant m} \max_{a \in A} \mathrm{Inf}_a(f^{(j)})}, \end{aligned}$$

where Δ_{ij}, $1 \leqslant i \leqslant j \leqslant m$, is given by

$$\begin{aligned} &\frac{k_j}{\sqrt{2}} \sum_{r=1}^{k_j - 1} (r-1)! \binom{k_i - 1}{r-1} \binom{k_j - 1}{r-1} \sqrt{(k_i + k_j - 2r)!} \big(\|f^{(i)} \star_{k_i - r} f^{(i)}\|_{2r} + \|f^{(j)} \star_{k_j - r} f^{(j)}\|_{2r} \big) \\ &\qquad + \mathbf{1}_{\{k_i < k_j\}} \sqrt{k_j! \binom{k_j}{k_i}} \|f^{(j)} \star_{k_j - k_i} f^{(j)}\|_{2k_i}. \end{aligned}$$

We finish this section by a useful result, which shows how the *influence* $\mathrm{Inf}_a f$ of $f : A^k \to \mathbb{R}$ can be bounded by the norm of the contraction of f of order $k-1$:

Proposition 2.8 *Let $f : A^k \to \mathbb{R}$ be a symmetric function vanishing on diagonals. Then*

$$(k-1)! \max_{a\in A} \mathrm{Inf}_a(f) := \max_{a\in A} \sum_{a_2,\ldots,a_k\in A} f(a,a_2,\ldots,a_k)^2 \leqslant \|f \star_{k-1} f\|_2.$$

Proof. We have

$$\begin{aligned}
\|f \star_{k-1} f\|_2^2 &= \sum_{a,b\in A} \left[\sum_{a_2,\ldots,a_k\in A} f(a,a_2,\ldots,a_k) f(b,a_2,\ldots,a_k)\right]^2 \\
&\geqslant \sum_{a\in A} \left[\sum_{a_2,\ldots,a_k\in A} f^2(a,a_2,\ldots,a_k)\right]^2 \\
&\geqslant \max_{a\in A} \left[\sum_{a_2,\ldots,a_k\in A} f^2(a,a_2,\ldots,a_k)\right]^2 \\
&= \left[(k-1)! \max_{a\in A} \mathrm{Inf}_a(f)\right]^2.
\end{aligned}$$

□

As a consequence of Theorem 2.7 and Proposition 2.8, we immediately get the following result.

Corollary 2.9 *Let $\mathbf{X} = \{X_a : a \in A\}$ be a collection of independent centered random variables, with unit variance and such that $\beta := \sup_a E|X_a|^3 < \infty$. Fix integers $m \geqslant 1$, $k_m > ... > k_1 \geqslant 1$. For every $j = 1, ..., m$, let $\{f_N^{(j)} : N \geqslant 1\}$ be a sequence of functions such that $f_N^{(j)} : A^{k_j} \to \mathbb{R}$ is symmetric and vanishes on diagonals. Define $Q_N^j(\mathbf{X}) := Q_{k_j}(f_N^{(j)}, \mathbf{X})$ according to (2.17), and assume that $E[Q_N^j(\mathbf{X})^2] = 1$ for all $j = 1, \ldots, m$ and $N \geqslant 1$. Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a thrice differentiable function such that $\|\varphi''\|_\infty + \|\varphi'''\|_\infty < \infty$. If, for some $\alpha > 0$, $\|f_N^{(j)} \star_{k_j - r} f_N^{(j)}\|_{2r} = O(N^{-\alpha})$ for all $j = 1, \ldots, m$ and $r = 1, \ldots, k_j - 1$, then, by noting $(Z^1, \ldots, Z^m)$ a centered Gaussian vector such that $E[Z^i Z^j] = 0$ if $i \neq j$ and $E[(Z^j)^2] = 1$, we have*

$$\left|E[\varphi(Q_N^1(\mathbf{X}), \ldots, Q_N^m(\mathbf{X}))] - E[\varphi(Z^1, \ldots, Z^m)]\right| = O(N^{-\alpha/2}).$$

3 Gaussian fluctuations of non-diagonal trace components

Our aim in this section is to prove the multidimensional CLT (1.15), by using the universality results presented in Section 2. To do this, we shall use an auxiliary collection $\mathbf{G} = \{G_{ij} : i, j \geqslant 1\}$ of i.i.d. copies of a $\mathcal{N}(0,1)$ random variable.

As in Section 1.3, for a given integer $k \geqslant 2$, we write $D_N^{(k)}$ to indicate the set of vectors $\mathbf{i} = (i_1, \ldots, i_k) \in [N]^k$ such that all the elements (i_a, i_{a+1}), $a = 1, \ldots, k$, are different in pairs (with the convention that $i_{k+1} = i_1$). We have the following preliminary result:

Proposition 3.1 *For any fixed integer $k \geqslant 2$,*

$$N^{-k/2} \sum_{\mathbf{i} \in D_N^{(k)}} G_{i_1 i_2} \ldots G_{i_k i_1} \xrightarrow{\text{Law}} Z_k \sim \mathscr{N}(0, k) \quad \textit{as } N \to \infty.$$

Remark 3.2 When $k = 1$, the conclusion of the above proposition continues to be true, since in this case we obviously have

$$N^{-1/2} \sum_{i=1}^{N} G_{ii} \sim \mathscr{N}(0, 1).$$

Proof of Proposition 3.1: The main idea is to use the results of Section 2, in the special case $A = \mathbb{N}^2$, that is, A is the collection of all pairs (i, j) such that $i, j \geqslant 1$. Observe that

$$N^{-k/2} \sum_{\mathbf{i} \in D_N^{(k)}} G_{i_1 i_2} \ldots G_{i_k i_1} = Q_k(f_{k,N}, \mathbf{G}),$$

with $f_{k,N} : ([N]^2)^k \to \mathbb{R}$ the symmetric function defined by

$$f_{k,N} = \frac{1}{k!} \sum_{\sigma \in \mathfrak{S}_k} f_{k,N}^{(\sigma)}, \tag{3.23}$$

where we used the notation

$$f_{k,N}^{(\sigma)}\big((a_1, b_1), \ldots, (a_k, b_k)\big) = N^{-k/2} \sum_{\mathbf{i} \in D_N^{(k)}} \mathbf{1}_{\{i_{\sigma(1)} = a_1,\, i_{\sigma(1)+1} = b_1\}} \ldots \mathbf{1}_{\{i_{\sigma(k)} = a_k,\, i_{\sigma(k)+1} = b_k\}}, \tag{3.24}$$

and $\mathfrak{S}_k$ denotes the set of all permutations of $[k]$. Hence, by virtue of Theorem 2.6, to prove Proposition 3.1 it is sufficient to accomplish the following two steps: (*Step 1*) prove that property (3) (with $f_{k,N}$ replacing f_N) in the statement of Theorem 2.6 takes place, and (*Step 2*) show that relation (2.22) (with $f_{k,N}$ replacing f_N) is verified.

Step 1. Let $r \in \{1, \ldots, k-1\}$. For $\sigma, \tau \in \mathfrak{S}_k$, we compute

$$\begin{aligned} & f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\big((x_1, y_1), \ldots, (x_{2k-2r}, y_{2k-2r})\big) \qquad (3.25) \\ = \;& N^{-k} \sum_{\mathbf{i},\mathbf{j} \in D_N^{(k)}} \mathbf{1}_{\{i_{\sigma(1)} = x_1,\, i_{\sigma(1)+1} = y_1\}} \ldots \mathbf{1}_{\{i_{\sigma(k-r)} = x_{k-r},\, i_{\sigma(k-r)+1} = y_{k-r}\}} \\ & \qquad \times \mathbf{1}_{\{j_{\tau(1)} = x_{k-r+1},\, j_{\tau(1)+1} = y_{k-r+1}\}} \ldots \mathbf{1}_{\{j_{\tau(k-r)} = x_{2k-2r},\, j_{\tau(k-r)+1} = y_{2k-2r}\}} \\ & \qquad \times \mathbf{1}_{\{i_{\sigma(k-r+1)} = j_{\tau(k-r+1)},\, i_{\sigma(k-r+1)+1} = j_{\tau(k-r+1)+1}\}} \ldots \mathbf{1}_{\{i_{\sigma(k)} = j_{\tau(k)},\, i_{\sigma(k)+1} = j_{\tau(k)+1}\}}. \end{aligned}$$

We now want to assess the quantity $\|f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\|_{2k-2r}^2$. To do this, we exploit the representation (3.25) in order to write such a squared norm as a sum over $([N]^k)^4$: as a consequence, one deduces that $\|f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\|_{2k-2r}^2 \leqslant |F_N^{(r,\sigma,\tau)}|\, N^{-2k}$ where $F_N^{(r,\sigma,\tau)}$ is the subset of $([N]^k)^4$ composed of those quadruplets $(\mathbf{i},\mathbf{j},\mathbf{a},\mathbf{b})$ such that

$$\begin{aligned}
&i_{\sigma(1)} = a_{\sigma(1)}, \quad i_{\sigma(1)+1} = a_{\sigma(1)+1}, \ldots, \quad i_{\sigma(k-r)} = a_{\sigma(k-r)}, \quad i_{\sigma(k-r)+1} = a_{\sigma(k-r)+1} \\
&j_{\tau(1)} = b_{\tau(1)}, \quad j_{\tau(1)+1} = b_{\tau(1)+1}, \ldots, \quad j_{\tau(k-r)} = b_{\tau(k-r)}, \quad j_{\tau(k-r)+1} = b_{\tau(k-r)+1} \\
&i_{\sigma(k-r+1)} = j_{\tau(k-r+1)}, \quad i_{\sigma(k-r+1)+1} = j_{\tau(k-r+1)+1}, \ldots, \quad i_{\sigma(k)} = j_{\tau(k)}, \quad i_{\sigma(k)+1} = j_{\tau(k)+1} \\
&a_{\sigma(k-r+1)} = b_{\tau(k-r+1)}, \quad a_{\sigma(k-r+1)+1} = b_{\tau(k-r+1)+1}, \ldots, \quad a_{\sigma(k)} = b_{\tau(k)}, \quad a_{\sigma(k)+1} = b_{\tau(k)+1}.
\end{aligned} \tag{3.26}$$

It is immediate that, among the equalities in (3.26), the $2k$ equalities appearing in the forthcoming display (3.27) are pairwise disjoint (that is, an index appearing in one of the equalities does not enter into the others):

$$\begin{aligned}
&i_{\sigma(1)} = a_{\sigma(1)}, \ldots, i_{\sigma(k-r)} = a_{\sigma(k-r)}, \quad j_{\tau(1)} = b_{\tau(1)}, \ldots, j_{\tau(k-r)} = b_{\tau(k-r)} \\
&i_{\sigma(k-r+1)} = j_{\tau(k-r+1)}, \ldots, \quad i_{\sigma(k)} = j_{\tau(k)}, \quad a_{\sigma(k-r+1)} = b_{\tau(k-r+1)}, \ldots, a_{\sigma(k)} = b_{\tau(k)}.
\end{aligned} \tag{3.27}$$

Hence, the cardinality of $F_N^{(r,\sigma,\tau)}$ is less than N^{2k}, from which we infer that $\|f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\|_{2k-2r}^2$ is bounded by 1. This is not sufficient for our purposes, since we need to show that $\|f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\|_{2k-2r}^2$ tends to zero as $N \to \infty$. To prove this, it is sufficient to extract from (3.26) one supplementary equality which is not already written in (3.27). We shall prove that this equality exists by contradiction. Set $L = \{\sigma(s) : 1 \leqslant s \leqslant k-r\}$ and $R = \{\sigma(s)+1 : 1 \leqslant s \leqslant k-r\}$ (with the convention that $k+1 = 1$). Now assume that $R = L$. Then $\sigma(1)+1 \in R$ also belongs to L, so that $\sigma(1)+2 \in R$. By repeating this argument, we get that $L = R = [k]$, which is a contradiction because $r \geqslant 1$. Hence, $R \neq L$. In particular, the display (3.26) implies at least one relation involving two indices that are not already coupled in (3.27). This yields that the cardinality of $F_N^{(r,\sigma,\tau)}$ is at most N^{2k-1}, and consequently that $\|f_{k,N}^{(\sigma)} \star_r f_{k,N}^{(\tau)}\|_{2k-2r}^2 \leqslant N^{-1}$. This fact implies immediately that the norms $\|f_N \star_r f_N\|_{2k-2r}$, $r = 1, \ldots, k-1$, verify

$$\|f_N \star_r f_N\|_{2k-2r} = O(N^{-1/2}), \tag{3.28}$$

and tend to zero as $N \to \infty$. In other words, we have proved that condition (3) in the statement of Theorem 2.6 is met.

Step 2. We have

$$\mathrm{Var}\left(N^{-k/2} \sum_{\mathbf{i} \in D_N^{(k)}} G_{i_1 i_2} \ldots G_{i_k i_1} \right) = N^{-k} \sum_{\mathbf{i},\mathbf{j} \in D_N^{(k)}} E[G_{i_1 i_2} \ldots G_{i_k i_1} G_{j_1 j_2} \ldots G_{j_k j_1}].$$

For fixed $\mathbf{i},\mathbf{j} \in D_N^{(k)}$, observe that the expectation $E[G_{i_1 i_2} \ldots G_{i_k i_1} G_{j_1 j_2} \ldots G_{j_k j_1}]$ can only be zero or one. Moreover, it is one if and only if, for all $s \in [k]$, there is exactly one $t \in [k]$

such that $(i_s, i_{s+1}) = (j_t, j_{t+1})$. In this case, we define $\sigma \in \mathfrak{S}_k$ as the bijection of $[k]$ into itself which maps each s to the corresponding t and we have, for all $s \in [k]$,

$$i_s = j_{\sigma(s)} = j_{\sigma(s-1)+1}. \tag{3.29}$$

To summarize, one has that $\mathrm{Var}\left(N^{-k/2}\sum_{\mathbf{i}\in D_N^{(k)}} G_{i_1 i_2}\ldots G_{i_k i_1}\right)$ equals

$$N^{-k}\sum_{\sigma\in\mathfrak{S}_k}\left|\left\{(\mathbf{i},\mathbf{j})\in (D_N^{(k)})^2 :\ (i_s,i_{s+1}) = (j_{\sigma(s)}, j_{\sigma(s)+1}) \ \text{ for all } s\in[k]\right\}\right|. \tag{3.30}$$

If $\sigma \in \mathfrak{S}_k$ is such that $\sigma(s) = \sigma(s-1)+1$ for all s (it is easily seen that there are exactly k permutations verifying this property in $\mathfrak{S}_k$), we get k different conditions by letting s run over $[k]$ in (3.29), so that

$$\left\{(\mathbf{i},\mathbf{j})\in (D_N^{(k)})^2 :\ (i_s,i_{s+1}) = (j_{\sigma(s)}, j_{\sigma(s)+1}) \ \text{ for all } s\in[k]\right\} \sim N^k, \quad \text{as } N\to\infty.$$

In contrast, if $\sigma \in \mathfrak{S}_k$ is *not* such that $\sigma(s) = \sigma(s-1)+1$ for all s, then by letting s run over $[k]$, one deduces from (3.29) at least $k+1$ different conditions, so that, in this case,

$$\left\{(\mathbf{i},\mathbf{j})\in (D_N^{(k)})^2 :\ (i_s,i_{s+1}) = (j_{\sigma(s)}, j_{\sigma(s)+1}) \ \text{ for all } s\in[k]\right\} = o(N^k), \quad \text{as } N\to\infty.$$

Taking into account these two properties together with the representation (3.30), we deduce that the variance of

$$N^{-k/2}\sum_{\mathbf{i}\in D_N^{(k)}} G_{i_1 i_2}\ldots G_{i_k i_1}$$

tends to k as $N\to\infty$. It follows that the required property (2.22) in Theorem 2.6 (with $\sigma^2 = k$) is met.

The proof of Proposition 3.1 is concluded. □

The *multidimensional* version of Proposition 3.1 reads as follows:

Proposition 3.3 *Fix $m \geqslant 1$, as well as integers $k_m > \ldots > k_1 \geqslant 2$. Then, as $N\to\infty$,*

$$\left(N^{-1/2}\sum_{i=1}^N G_{ii},\ N^{-\frac{k_1}{2}}\sum_{\mathbf{i}\in D_N^{(k_1)}} G_{i_1 i_2}\cdots G_{i_{k_1} i_1},\ \ldots \right. \tag{3.31}$$

$$\left. \ldots,\ N^{-\frac{k_m}{2}}\sum_{\mathbf{i}\in D_N^{(k_m)}} G_{i_1 i_2}\cdots G_{i_{k_m} i_1}\right) \xrightarrow{\text{Law}} \left(Z_1, Z_{k_1}, ..., Z_{k_m}\right),$$

where $\mathbf{Z} = \{Z_k : k\geqslant 1\}$ denotes a collection of independent centered Gaussian random variables such that, for every $k\geqslant 1$, $E(Z_k^2) = k$.

Proof: It is an application of Theorem 2.4, in the following special case:

- $w_N(i,j) = \frac{1}{\sqrt{N}}$, if $i = j \leqslant N$ and $w_N(i,j) = 0$ otherwise;
- V is equal to the diagonal matrix such that $V(a,b) = 0$ if $a \neq b$ and $V(a,a) = k_a$, for $a = 1, ..., m$;
- for $j = 1, ..., m$, $f_N^{(j)} = f_{k_j,N}$, where we used the notation (3.23).

Indeed, in view of Proposition 3.1, one has that condition (2) in the statement of Theorem 2.4 is satisfied. Moreover, for fixed $a \neq b$ and since $\mathbf{G}$ consists of a collection of independent and centered (Gaussian) random variables, it is clear that, for all N,

$$E\left[\sum_{\mathbf{i}\in D_N^{(k_a)}} G_{i_1 i_2}\ldots G_{i_{k_a} i_1} \times \sum_{\mathbf{j}\in D_N^{(k_b)}} G_{j_1 j_2}\ldots G_{j_{k_b} j_1}\right] = 0,$$

so that condition (2.20) is met. The proof is concluded. □

By combining Proposition 3.3 and Proposition 2.3, we can finally deduce the following general result for non-diagonal trace components.

Corollary 3.4 *For $N \geqslant 2$, let X_N be the $N \times N$ random matrix given by (1.1), where the reference random variable X has mean zero, unit variance and finite absolute third moment. Fix $m \geqslant 1$, as well as integers $2 \leqslant k_1 < \ldots < k_m$. Then, the CLT (1.15) takes place, with $\mathbf{Z} = \{Z_k : k \geqslant 1\}$ denoting a sequence of independent centered Gaussian random variables such that, for every $k \geqslant 1$, $E(Z_k^2) = k$.*

Remark 3.5 In order to prove Corollary 3.4, one only needs the existence of third moments. Note that, as will become clear in the following Section 4, moments of higher orders are necessary for our proof of (1.16).

4 The remainder: combinatorial bounds on partitioned chains and proof of Theorem 1.1

Fix an integer $k \geqslant 2$. From section 1.3, recall that $D_N^{(k)}$ denotes the subset of vectors $\mathbf{i} = (i_1, \ldots, i_k) \in [N]^k$ such that all the elements (i_a, i_{a+1}), $a = 1, \ldots, k$, are different in pairs (with the convention that $i_{k+1} = i_1$). From the Introduction, recall that X is a centered random variable, having unit variance and with finite moments of all orders. Let also $\mathbf{X} = \{X_{ij} : i, j \geqslant 1\}$ be a collection of i.i.d. copies of X. In the present section, our aim is to prove (1.16), that is

Proposition 4.1 *For every $k \geqslant 2$, as $N \to \infty$,*

$$\mathrm{Var}\left(N^{-k/2} \sum_{\mathbf{i} \notin D_N^{(k)}} \left[X_{i_1 i_2} \dots X_{i_k i_1} - E(X_{i_1 i_2} \dots X_{i_k i_1}) \right] \right) = O(N^{-1}). \tag{4.32}$$

The proof of Proposition 4.1 is detailed in Section 4.4, and builds on several combinatorial estimates derived in Sections 4.2–4.3. To ease the reading of the forthcoming material, we now provide an intuitive outline of this proof.

Remark on notation. Given an integer $k \geqslant 2$, we denote by $\mathcal{P}(k)$ the collection of all partitions of $[k] = \{1, ..., k\}$. Recall that a partition $\pi \in \mathcal{P}(k)$ is an object of the type $\pi = \{B_1, ..., B_r\}$, where the B_j's are disjoint and non-empty subsets of $[k]$, called *blocks*, such that $\cup_{j=1,...,r} B_j = [k]$. Given $a, x \in [k]$ and $\pi \in \mathcal{P}(k)$, we write $a \overset{\pi}{\sim} x$ whenever a and x are in the same block of π. We also use the symbol $\hat{1}$ to indicate the one-block partition $\hat{1} = \{[k]\}$ (this is standard notation from combinatorics – see e.g. [28]). In this section, for the sake of simplicity and because k is fixed, we write D_N instead of $D_N^{(k)}$.

4.1 Sketch of the proof of Proposition 4.1

Our starting point is the following elementary decomposition:

$$[N]^k \setminus D_N = \bigcup_{\pi \in \mathcal{Q}(k)} A_N(\pi),$$

where $\mathcal{Q}(k)$ stands for the collection of all partitions of $[k]$ containing at least one block of cardinality $\geqslant 2$, and $A_N(\pi)$ is the collection of all vectors $\mathbf{i} \in [N]^k$ such that the equality $(i_a, i_{a+1}) = (i_x, i_{x+1})$ holds if and only if $a \overset{\pi}{\sim} x$. Using this decomposition, one sees immediately that, in order to show (4.32), it is sufficient to prove that, for each *fixed* $\pi \in \mathcal{Q}(k)$, the quantity

$$\begin{aligned} &\mathrm{Var}\left(N^{-k/2} \sum_{\mathbf{i} \in A_N(\pi)} \left[X_{i_1 i_2} \dots X_{i_k i_1} - E(X_{i_1 i_2} \dots X_{i_k i_1}) \right] \right) \qquad (4.33) \\ &= N^{-k} \sum_{(\mathbf{i},\mathbf{j}) \in A_N(\pi) \times A_N(\pi)} \left[E(X_{i_1 i_2} \dots X_{i_k i_1} X_{j_1 j_2} \dots X_{j_k j_1}) - E(X_{i_1 i_2} \dots X_{i_k i_1}) E(X_{j_1 j_2} \dots X_{j_k j_1}) \right] \end{aligned}$$

is $O(N^{-1})$, as $N \to \infty$. Let $G_N(\pi)$ denote the subset of pairs $(\mathbf{i}, \mathbf{j}) \in A_N(\pi) \times A_N(\pi)$ such that the following non-vanishing condition is in order:

$$E(X_{i_1 i_2} \dots X_{i_k i_1} X_{j_1 j_2} \dots X_{j_k j_1}) - E(X_{i_1 i_2} \dots X_{i_k i_1}) E(X_{j_1 j_2} \dots X_{j_k j_1}) \neq 0. \tag{4.34}$$

Hence

$$\begin{aligned} &\mathrm{Var}\left(N^{-k/2} \sum_{\mathbf{i} \in A_N(\pi)} \left[X_{i_1 i_2} \dots X_{i_k i_1} - E(X_{i_1 i_2} \dots X_{i_k i_1}) \right] \right) \qquad (4.35) \\ &= N^{-k} \sum_{(\mathbf{i},\mathbf{j}) \in G_N(\pi)} \left[E(X_{i_1 i_2} \dots X_{i_k i_1} X_{j_1 j_2} \dots X_{j_k j_1}) - E(X_{i_1 i_2} \dots X_{i_k i_1}) E(X_{j_1 j_2} \dots X_{j_k j_1}) \right]. \end{aligned}$$

Due to the finite moment assumptions for X, and by appling the generalized Hölder inequality, it is clear that, for a generic pair $(\mathbf{i},\mathbf{j})$,

$$\left|E(X_{i_1i_2}\dots X_{i_ki_1}X_{j_1j_2}\dots X_{j_kj_1})-E(X_{i_1i_2}\dots X_{i_ki_1})E(X_{j_1j_2}\dots X_{j_kj_1})\right| \leqslant 2\,E(|X|^{2k}) < \infty.$$

It follows that, in order to prove that the sum in (4.35) is $O(N^{-1})$, it is enough to show that

$$\left|G_N(\pi)\right| \leqslant \Theta(k,\pi)N^{k-1}, \tag{4.36}$$

for some constant $\Theta(k,\pi)$ not depending on N. Our way of proving (4.36) is to show that, if $(\mathbf{i},\mathbf{j})$ denotes a *generic* element of $G_N(\pi)$, then, necessarily, there exists at least $k+1$ equalities between the $2k$ indices $i_1,\dots,i_k,\ j_1,\dots,j_k$ of $(\mathbf{i},\mathbf{j})$. Note that by 'equality' we just mean the existence of two *different* integers $a,b\in[k]$ such that $i_a=i_b$ or $j_a=j_b$, or the existence of two integers $a,b\in[k]$ such that $i_a=j_b$. Proving this fact implies that the $2k$ indices of a generic elements $(\mathbf{i},\mathbf{j})$ of $G_N(\pi)$ have at most $k-1$ *degrees of freedom* (see Point **7** of Section 4.2 for a precise definition), so that (4.36) holds immediately — the constant $\Theta(k,\pi)$ merely counting the number of ways in which the $k+1$ equalities can be consistently distributed among the indices composing $(\mathbf{i},\mathbf{j})$. In order to extract these $k+1$ equalities between the $2k$ indices of a generic element $(\mathbf{i},\mathbf{j})$ of $G_N(\pi)$, we will consider two cases, according as the partition $\pi\in\mathcal{Q}(k)$ contains at least one singleton or not.

Case A: No singletons in π. By definition of $A_N(\pi)$, and due to the absence of singleton in π, we already see that there are at least $k/2$ or $(k+1)/2$ (according to the evenness of k) equalities between the k indices of $\mathbf{i}$ (resp. $\mathbf{j}$). Moreover, the non-vanishing condition (4.34) implies that there is at least one further equality between one index of $\mathbf{i}$ and one index of $\mathbf{j}$. So, we proved the existence of $k+1$ equalities between the $2k$ indices of $(\mathbf{i},\mathbf{j})$, and the proof of (4.36) in the Case A is done.

Case B: At least one singleton in π. Let S denote the collection of the singleton(s) of π. In order for (4.34) to be true, observe that, for all $s\in S$, we must have $(j_s,j_{s+1})=(i_a,i_{a+1})$ for some $a\in[k]$. In particular, this means that there exist $|S|$ equalities of the type $j_s=i_a$ for the indices composing $(\mathbf{i},\mathbf{j})$. Also, by definition of the objects we are dealing with, for all $t\in[k]\setminus S$, we must have $(i_t,i_{t+1})=(i_a,i_{a+1})$ for some a, different from t, in the same π-block as t. Of course, the same must hold with i replaced by j. Hence, in order for (4.36) to be true, it remains to produce one equality between indices that has not been already considered. We mentioned above that for all $t\in[k]\setminus S$, there exists a, different from t and in the same block as t, such that $j_t=j_a$. Hence, to conclude it remains to show that we have $j_t=j_a$ for at least one integer t belonging to $[k]\setminus S$ and one integer a *not* belonging to the same block as t. Since, by assumption, π contains at least one singleton and one block of cardinality $\geqslant 2$ (indeed, $\pi\in\mathcal{Q}(k)$), without loss of generality (up to relabeling the indices according to a cyclic permutation of $[k]$), we can assume that S contains the singleton $\{k\}$. Consider now the singleton $\{s^*\}$ of S, where s^* is defined as the greatest of the integers m such that $\{m\}$ is adjacent from the right to a block, say B_{u*}, of cardinality

$\geqslant 2$. For a particular example of this situation, see the diagram in Fig. 1, where each row represents the same partition of [7] having $s^* = 6$ (see Point **3**. in the subsequent Section 4.2 for a formal construction of diagrams). To finish the proof, once again we split it into two cases:

Case B1: The block B_{u*} contains two consecutive integers. This assumption implies that $j_x = j_t = j_{t+1}$ for all $x, t \in B_{u*}$. Since $\{a\}$ is adjacent from the right to B_{u*}, we have $j_a = j_t$ for all $t \in B_{u*}$, which is exactly what we wanted to show.

Case B2: The block B_{u*} does not contain two consecutive integers. Fig. 7 is an illustrative example of such situation, where each row represents the same partition of [8], with $s^* = 7$. As we see on this picture, we have necessarily $j_7 = j_5$, yielding the desired additional equality, which could not be extracted from the previous discussion. In Section 4.3, it is shown that this line of reasoning can be extended to general situations.

Remark 4.2 The sketch given above contains all the main ideas entering in the proof of Proposition 4.1. The reader not interested in technical combinatorial details, can then go directly to Section 4.4, where the proof of Theorem 1.1 is concluded. The subsequent Sections 4.2–4.3 fill the gaps of the above sketch, by providing exact definitions as well as complete formal arguments leading to the estimate (4.32).

4.2 Definitions

In the following list, we introduce some further definitions that are needed for the analysis developed in the rest of this section.

1. Fix integers $N, k \geqslant 2$. A *chain* c of length $2k$, built from $[N]$, is an object given by the juxtaposition of $2k$ pairs of integers of the type

$$c = (i_1, i_2)(i_2, i_3)...(i_k, i_1)(j_1, j_2)(j_2, j_3)...(j_k, j_1), \tag{4.37}$$

where $i_a, j_x \in [N]$, for $a, x = 1, ..., k$. The class of all chains of length $2k$ built from $[N]$ is denoted by $C(2k, N)$. As a notational convention, we will use the letter i to write the first k pairs in the chain, and the letter j to write the remaining ones. For instance, an element of $C(6, 5)$ (that is, a chain of length 6 built from the set $\{1, 2, 3, 4, 5\}$) is

$$c = (1, 5)(5, 1)(1, 1)(3, 3)(3, 3)(3, 3),$$

where $i_1 = 1$, $i_2 = 5$, $i_3 = 1$, $j_1 = j_2 = j_3 = 3$. According to the graphical conventions given below (at Point 3 of the present list) we will sometimes say that $(i_1, i_2)(i_2, i_3)...(i_k, i_1)$ and $(j_1, j_2)(j_2, j_3)$ $...(j_k, j_1)$ are, respectively, the *upper sub-chain* and the *lower sub-chain* associated with the chain c in (4.37). For instance, in the previous example the upper sub-chain is $(1, 5)(5, 1)(1, 1)$, whereas the lower one is $(3, 3)(3, 3)(3, 3)$. We shall say that (i_l, i_{l+1}) is the lth pair in the upper sub-chain of c (and similarly for the elements of the

lower sub-chain). We shall sometimes call i_a the *left index* of the pair (i_a, i_{a+1}). Also, we use the convention $i_{k+1} = i_1$ and $j_{k+1} = j_1$. Of course, a chain is completely determined by the left indices of its pairs.

2. Let $\pi \in \mathcal{P}(k)$ be a partition of $[k]$. Recall that, for $a, b \in [k]$, we write $a \overset{\pi}{\sim} b$ to indicate that a and b belong to the same block of π. We say that a chain c as in (4.37) *has partition* π if, for every $a, b \in [k]$, the following double implications take place: (i) $(i_a, i_{a+1}) = (i_b, i_{b+1})$ if and only if $a \overset{\pi}{\sim} b$, and (ii) $(j_a, j_{a+1}) = (j_b, j_{b+1})$ if and only if $a \overset{\pi}{\sim} b$. In other words, a chain has partition π if and only if the partitions of $[k]$ induced by the identical pairs in its upper and lower sub-chain are both equal to π, that is (with the notation of Section 4.1), if and only if $(i_1, ..., i_k), (j_1, ..., j_k) \in A_N(\pi)$. For instance, take $k = 4$ and $\pi = \{\{1, 3\}, \{2, 4\}\}$. Then, the following chain built from [3] has partition π:

$$c = (1, 2)(2, 1)(1, 2)(2, 1)(3, 1)(1, 3)(3, 1)(1, 3).$$

Note the 'only if' part in the definition given above, implying that, if a chain has partition π and if x and y are not in the same block of π, then necessarily $(i_x, i_{x+1}) \neq (i_y, i_{y+1})$ and $(j_x, j_{x+1}) \neq (j_y, j_{y+1})$. This yields in particular that a chain cannot have two different partitions.

3. Given $k \geqslant 2$, we shall sometimes represent a generic chain with partition $\pi \in \mathcal{P}(k)$ by means of *diagrams*. These diagrams are mnemonic devices composed of an upper row and a lower row, of k dots each. These rows represent, respectively, the upper and lower sub-chain of a given chain, in such a way that the lth dot (from left to right) in the upper (resp. lower) row corresponds the lth pair in the upper (resp. lower) sub-chain. Each block B of the partition π is represented by two closed curves: the first one is drawn around the dots of the upper row corresponding to the pairs (i_a, i_{a+1}) verifying $a \in B$; the second one is drawn around the dots of the lower row corresponding to those (j_x, j_{x+1}) verifying $x \in B$. The resulting diagram is the superposition of two identical combinations of dots and curves. Note that the shape of the diagram does not depend on N. For instance, the diagram in Fig. 1 corresponds to the case $k = 7$, and $\pi = \{\{1, 4, 5\}, \{2\}, \{3\}, \{6\}, \{7\}\}$,[3] whereas the diagram in Fig. 2 corresponds to $k = 6$ and the one-block partition $\hat{1} = \{[6]\}$.

4. In general, given a chain c as in (4.37) with partition $\pi = \{B_1, ..., B_r\}$ as at Point 2 of the present list, we shall say the the block B_u of the upper sub-chain *corresponds* to the block B_v of the lower sub-chain, whenever $(i_a, i_{a+1}) = (j_x, j_{x+1})$ for every $a \in B_u$ and every $x \in B_v$. Note that one given block B_u in the upper sub-chain cannot correspond to more than one block in the lower sub-chain. For $\pi = \{B_1, ..., B_r\} \in \mathcal{P}(k)$, we shall now define a class of chains $C_\pi(2k, N) \subset C(2k, N)$, whose elements have partition π and are

[3] A chain with partition π as in Fig. 1 is

$$c = (1, 1)(1, 2)(2, 1)(1, 1)(1, 1)(1, 3)(3, 1)(1, 1)(1, 4)(4, 1)(1, 1)(1, 1)(1, 5)(5, 1).$$

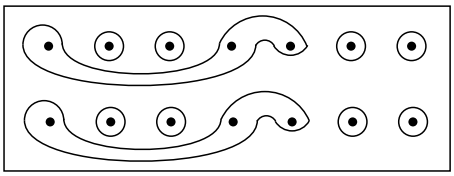

Figure 1: a chain with a five-block partition

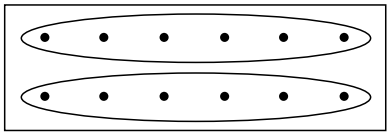

Figure 2: a chain with a one-block partition

characterized by two facts: the associated upper and lower sub-chains have at least one pair in common, and "no singletons are left on their own". Formally, the class $C_\pi(2k,N)$ is defined as follows (recall that we use the letter i for the elements of the upper sub-chain, and the letter j for the elements of the lower sub-chain). (i) If $|B_t| \geqslant 2$ for every $t=1,...,r$, then $C_\pi(2k,N)$ is the collection of all chains of partition π verifying that there exists $a,x \in [k]$ such that the block B_a in the upper sub-chain corresponds to the block B_x in the lower sub-chain. (ii) If π contains at least one singleton, then $C_\pi(2k,N)$ is the collection of all chains of partition π such that every singleton in the upper (resp. lower) sub-chain corresponds to a block of the lower (resp. upper) subchain, that is: for every $\{a\} \in \pi$, there exists $u=1,...,r$ such that $(i_a, i_{a+1}) = (j_l, j_{l+1})$ for every $l \in B_u$, and, for every $\{x\} \in \pi$, there exists $v=1,...,r$ such that $(j_x, j_{x+1}) = (j_s, j_{s+1})$ for every $s \in B_v$. For instance, if $k=3$ and $\pi = \{[3]\}$, then one element of $C_\pi(6,5)$ is

$$c = (5,5)(5,5)(5,5)(5,5)(5,5)(5,5).$$

If $k=6$ and $\pi = \{\{1,2,3\},\{4\},\{5\},\{6\}\}$, then one element of $C_\pi(12,5)$ is

$$c = (1,1)(1,1)(1,1)(1,2)(2,5)(5,1)(2,2)(2,2)(2,2)(2,5)(5,1)(1,2).$$

5. Fix $k,N \geqslant 2$, as well as a partition $\pi = \{B_1,...,B_r\} \in \mathcal{P}(k)$. Given two subsets $U,V \subset [r]$ such that $|U| = |V|$, let $R : U \to V : u \mapsto R(u)$ be a bijection from U onto V. We shall denote by $C_\pi^R(2k,N)$ the subset of $C_\pi(2k,N)$ composed of those chains $c \in C_\pi(2k,N)$ such that the block B_u in the upper sub-chain corresponds to the block $B_{R(u)}$ in the lower sub-chain. When $U=\{u\}$ and $V=\{v\}$ are singletons, we shall simply write $C_\pi^{u,v}(2k,N)$ to indicate the set of those $c \in C_\pi(2k,N)$ such that the block B_u in the upper sub-chain corresponds to the block B_v in the lower sub-chain. For instance, the chain

$$c_1 = (1,1)(1,1)(1,2)(2,5)(5,1)(2,2)(2,2)(2,5)(5,1)(1,2)$$

is an element of $C_\pi^R(10,4)$, where $\pi = \{B_1, B_2, B_3, B_4\} = \{\{1,2\},\{3\},\{4\},\{5\}\}$, $U = V = \{2,3,4\}$, and $R(2) = 4$, $R(3) = 2$ and $R(4) = 3$. The chain

$$c_2 = (3,3)(3,3)(3,3)(3,3)$$

belongs to $C_{\hat{1}}^{1,1}(4,3)$, where $\hat{1} = \{B_1\} = \{[2]\}$. Note that the definition of $C_\pi^R(2k,N)$ does not give any information concerning the blocks of the upper and lower sub-chains that do not belong, respectively, to the domain and the image of R. In other words, for a chain $c \in C_\pi^R(2k,N)$, one can have that the block B_u in the upper sub-chain corresponds to the block B_v in the lower sub-chain even if $u \notin U$ and $v \notin V$. For instance, the chain

$$c = (1,1)(1,1)(1,2)(2,5)(5,1)(1,1)(1,1)(1,2)(2,5)(5,1)$$

is counted as an element of $C_\pi^R(10,4)$, where

$$\pi = \{B_1, B_2, B_3, B_4\} = \{\{1,2\},\{3\},\{4\},\{5\}\},$$

$U = V = \{2,3,4\}$, and $R(u) = u$, for $u = 2,3,4$.

6. Fix $k, N \geqslant 2$, as well as a partition $\pi = \{B_1, ..., B_r\} \in \mathcal{P}(k)$. Given a bijection $R : U \to V$ as at Point 5 above, we shall represent a generic element of the class $C_\pi^R(2k,N)$ by means of a diagram built as follows: first (i) draw the diagram associated with the class $C_\pi(2k,N)$, as explained at Point 3 of the present list, then (ii) for every pair of blocks B_u and B_v such that $u \in U$, $v \in V$ and $v = R(u)$ (note that B_u is in the upper sub-chain, and B_v in the lower sub-chain), draw a segment linking a representative element of B_u with a representative element of B_v. For instance, the class $C_\pi^R(10,N)$, associated with the chain c_1 appearing at Point 5 above, is represented by the diagram appearing in Fig. 3, whereas the chain c_2 is associated with the class $C_{\hat{1}}^{1,1}(4,3)$, whose diagram is drawn in Fig. 4.

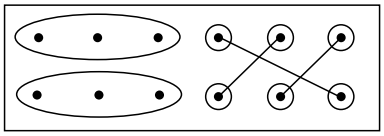

Figure 3: a chain with three pairs of corresponding singletons

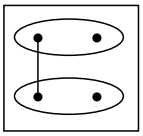

Figure 4: a chain with two corresponding blocks

7. Fix $k, N \geqslant 2$ and let $C \subset C(2k, N)$ be a generic subset of $C(2k, N)$. Let $q = 1, ..., 2k$ be an integer. We say that C has *at most q degrees of freedom* (or, equivalently, that C *has at most q free indices*) if there exists two subsets $D, E \subset [k]$ such that $|D| \geqslant 1$ and the following two properties are verified: (i) $|D| + |E| \leqslant q$, and (ii) for every[4] $x_D = \{x_a : a \in D\} \in [N]^{|D|}$ and every $y_E = \{y_b : b \in E\} \in [N]^{|E|}$, there exists *at most* one chain c as in (4.37) such that $i_a = x_a$ for every $a \in D$ and $j_b = y_b$ for every $b \in E$. Note that our definition contemplates the possibility that $E = \emptyset$, and in this case the role of $y_E = \emptyset$ is immaterial. In other words, the class C has at most q degrees of freedom if every $c \in C$ is completely determined by those i_a in the upper sub-chain such that $a \in D$ and those j_b in the lower sub-chain such that $b \in E$. For instance, it is easily seen the class $C(2k, N)$ has (exactly) $2k$ degrees of freedom. Another example is the diagram in Fig. 5, which corresponds to the case $k = 6$, $\pi = \{\{1, 2\}, \{3, 5\}, \{4, 6\}\}$ and $u = v = 1$. One sees that, for every N, specifying i_1, i_4 and j_4 completely identifies a chain inside the class $C_\pi^{1,1}(12, N)$, which has therefore three degrees of freedom.[5]

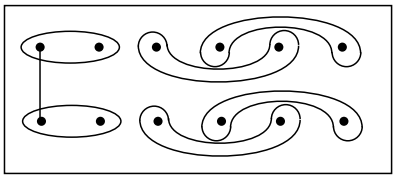

Figure 5: a class with three degrees of freedom

The proof of the two (useful) results contained in the next statement is elementary and omitted.

Lemma 4.3 *Fix $k, N \geqslant 2$.*

(1) *Let $q = 1, ..., 2k$. Assume that a generic class $C \subset C(2k, N)$ has at most q degrees of freedom. Then, $|C| \leqslant N^q$.*

(2) *Let $\hat{1} = \{[k]\}$ be the one-block partition of $[k]$. Then, the class $C_{\hat{1}}(2k, N)$ contains only "constant" chains of the type (4.37) such that $(i_1, i_2) = (i_a, i_{a+1}) = (j_x, j_{x+1})$, for every $a = 2, ..., k$ and every $x = 1, ..., k$. It follows that $|C_{\hat{1}}(2k, N)| = N$.*

Lemma 4.3 will be used in the subsequent section.

4.3 Combinatorial upper bounds

We keep the notation introduced in the previous section. The following statement, which is the key element for proving Proposition 4.1, contains the main combinatorial estimate of the paper.

[4]As indicated by our notation, we regard x_D and y_E as vectors, respectively in $[N]^{|D|}$ and $[N]^{|E|}$, by endowing D and E with the natural ordering induced by the ordering on $[k]$.

[5]Indeed, one has necessarily that $i_1 = i_2 = i_3 = i_5 = j_1 = j_2 = j_3 = j_5$, $i_4 = i_6$ and $j_4 = j_6$.

Proposition 4.4 *Fix $k, N \geqslant 2$, and let $\pi = \{B_1, ..., B_r\} \in \mathcal{P}(k)$ be a partition containing at least one block of cardinality $\geqslant 2$. Let the class $C_\pi(2k, N)$ be defined as at Point* **4.** *of the previous section. Then, there exists a finite constant $\Theta(k, \pi) \geqslant 0$, depending only on k and π (and not on N), such that*

$$|C_\pi(2k, N)| \leqslant \Theta(k, \pi) \times N^{k-1}. \tag{4.38}$$

Proof: We shall consider separately the two cases

A. For every $v = 1, ..., r$, $|B_v| \geqslant 2$.

B. The partition π contains at least one singleton.

Case A. When $k = 2, 3$, the only partition meeting the needed requirements is $\hat{1}$. According to Lemma 4.3-(2), $|C_{\hat{1}}(2k, N)| = N$, so that the claim is proved, and we shall henceforth assume that $k \geqslant 4$. Start by observing that $r \leqslant k/2$. Moreover, the class $C_\pi(2k, N)$ contains only chains such that at least one block in the upper sub-chain corresponds to a block in the lower sub-chain, which yields in turn that

$$C_\pi(2k, N) = \bigcup_{u,v=1}^{r} C_\pi^{u,v}(2k, N),$$

where we adopted the notation introduced at Point **5.** of Section 4.2. This implies the crude estimate

$$|C_\pi(2k, N)| \leqslant \sum_{u,v=1}^{r} |C_\pi^{u,v}(2k, N)|. \tag{4.39}$$

According to Lemma 4.3-(1), it is now sufficient to prove that each class $C_\pi^{u,v}(2k, N)$ has at most $2r-1$ degrees of freedom: indeed, (4.39) together with the fact that $2r-1 \leqslant k-1$ would imply relation (4.38), with $\Theta(k, \pi) = r^2 \leqslant k^2/4$. Fix $u, v \in \{1, ..., r\}$. To prove that $C_\pi^{u,v}(2k, N)$ has at most $2r-1$ degrees of freedom, we shall build two sets $D, E \subset [k]$ as follows. For every $s = 1, ..., r$, choose an element of the block B_s, and denote this element by a_s. Then, define

$$D = \{a_s : s = 1, ..., r\}, \;\; E = D \backslash \{a_v\},$$

where '\' denotes the difference between sets. We now claim that, for every $x_D = \{x_a : a \in D\} \in [N]^{|D|}$ and every $y_E = \{y_b : b \in E\} \in [N]^{|E|}$, there exists at most one chain $c \in C_\pi^{u,v}(2k, N)$ as in (4.37) such that $i_a = x_a$ for every $a \in D$ and $j_b = y_b$ for every $b \in E$. To prove this fact, suppose that such a chain c exists, and assume that there exists another chain

$$c' = (i'_1, i'_2)(i'_2, i'_3)...(i'_k, i'_1)(j'_1, j'_2)(j'_2, j'_3)...(j'_k, j'_1)$$

verifying this property and such that $c' \in C_\pi^{u,v}(2k, N)$. The following hold: (a) for every $s = 1, ..., r$ and every $a \in B_s$, one has that $i'_a = x_{a_s} = i_{a_s} = i_a$, (b) for every $s \neq v$ and every $a \in B_s$, $j'_a = y_{a_s} = j_{a_s} = j_a$ and (c) for $s = v$ and every $a \in B_v$,

$$j'_a = j'_{a_v} = i'_{a_u} = x_{a_u} = i_{a_u} = j_{a_v} = j_a.$$

As a consequence, $c' = c$. Since $|D| + |E| = 2r - 1$, this concludes the proof of Proposition 4.4 in the Case A.

Case B. We shall denote by S the collection of the singleton(s) of π, that is the subset of $[k]$ composed of those indices a such that $\{a\} \in \pi$. Note that $|S| > 0$ by assumption. We also write P for the collection of the indices $u \in [r]$ such that $|B_u| \geqslant 2$. Note that P is a subset of $[r]$, whereas $S \subset [k]$. Note also that the set $[r] \backslash P$ is the collection of all those $v \in [r]$ such that B_v is a singleton. Clearly,

$$|P| = r - |S| \leqslant \frac{k - |S|}{2}.$$

By exploiting the cyclic nature of sub-chains, we can always assume, without loss of generality, that S contains the singleton $\{k\}$. Since P is not empty, this entails that there exists at least one singleton of π that is adjacent from the right to a block of cardinality at least two. Formally, this means that there exists $s^* \in S$ and $u^* \in P$ such that $s^* - 1 \in B_{u^*}$. We shall distinguish two cases

B1. The block B_{u^*} contains two consecutive integers.

B2. The block B_{u^*} does not contain two consecutive integers.

(*Proof under B1.*) The situation of B1 is illustrated in Fig. 6, where $k = 9$,

$$\pi = \{B_1, ..., B_7\} = \{\{1\}, \{2\}, \{3, 6, 7\}, \{4\}, \{5\}, \{8\}, \{9\}\},$$

and one can take $s^* = 8$, $u^* = 3$, and the two consecutive integers in B_{u^*} are 6 and 7.

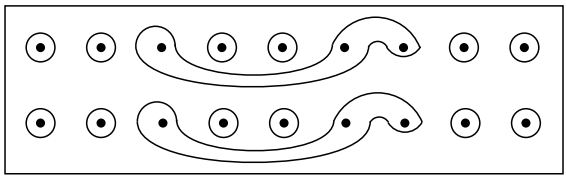

Figure 6: a singleton is adjacent to a 3-block with two consecutive elements

Since each element of $C_\pi(2k, N)$ is such that every singleton in a given sub-chain corresponds to a block in the opposite sub-chain, we have that

$$C_\pi(2k, N) = \bigcup_{R \in \mathcal{R}} C_\pi^R(2k, N), \tag{4.40}$$

where we adopted the same notation as at Point **5.** of Section 4.2, and the union runs over the class $\mathcal{R}$ of all bijections $R : U \to V$ such that both U and V contain the set $[r]\backslash P$, and every pair $(u, R(u))$ is such that at least one of the two blocks B_u and $B_{R(u)}$ is a singleton. This entails the estimate

$$|C_\pi(2k, N)| \leqslant \sum_{R\in\mathcal{R}} |C_\pi^R(2k, N)|. \tag{4.41}$$

To conclude the proof, we shall show that every class $C_\pi^R(2k, N)$ appearing in (4.41) has at most $k-1$ degrees of freedom: indeed, this fact together with Lemma 4.3-(1) yields the desired conclusion (4.38), with the constant $\Theta(k, \pi) = |\mathcal{R}|$ (note that the definition of $\mathcal{R}$ does not depend on N) . To prove that $C_\pi^R(2k, N)$ has at most $k-1$ degrees of freedom, we define two sets $D, E \subset [k]$ as follows. For every $s = 1, ..., r$, choose an element of the block B_s, and denote this element by a_s. Then, define

$$D = \{a_s : s = 1, ..., r\}, \;\; E = D\backslash \left\{\{a_{u^*}\} \cup \{a_s : s \in [r]\backslash P\}\right\}.$$

In other words, E is obtained by subtracting from D the singleton(s) and the representative element of the block B_{u^*}, that is, of the block adjacent to $\{s^*\}$. We now want to prove that, for every $x_D = \{x_a : a \in D\} \in [N]^{|D|}$ and every $y_E = \{y_b : b \in E\} \in [N]^{|E|}$, there is at most one chain $c \in C_\pi^R(2k, N)$ as in (4.37) such that $i_a = x_a$ for every $a \in D$ and $j_b = y_b$ for every $b \in E$. To show this, assume that such a chain c exists, and suppose that there exists another chain

$$c' = (i'_1, i'_2)(i'_2, i'_3)...(i'_k, i'_1)(j'_1, j'_2)(j'_2, j'_3)...(j'_k, j'_1)$$

verifying this property and such that $c' \in C_\pi^R(2k, N)$ and $c' \neq c$. By construction of the sets D and E, all the indices composing the upper chain are completely determined by the choice of x_D, whereas the choice of y_E determines the indices j_x such that either x is a singleton or $x \in B_v$ for some block B_v of cardinality $\geqslant 2$ and such that $v \neq u^*$. This entails in turn that, necessarily since $c' \neq c$, one has that $j'_x \neq j_x$ for every $x \in B_{u^*}$. This is absurd. Indeed, since B_{u^*} contains two consecutive integers, one has that $j'_x = j'_{x+1}$ and $j_x = j_{x+1}$ for every $x \in B_{u^*}$; it follows that, since $\{s^*\}$ is adjacent from the right to B_{u^*} and therefore $s^* - 1 \in B_{u^*}$,

$$j'_x = j'_{s^*-1} = j'_{s^*} = y_{s*} = j_{s^*} = j_{s^*-1} = j_x,$$

which is indeed a contradiction. Since

$$|D| + |E| = r + |P| - 1 \leqslant \frac{k - |S|}{2} + |S| + \frac{k - |S|}{2} - 1 = k - 1,$$

the proof is concluded.

(*Proof under B2.*) Since B_{u^*} does not contain two consecutive integers and $|B_{u^*}| \geqslant 2$, we deduce the existence of a block $B_{\overline{u}} \in \pi$, which is different from B_{u^*} and $\{s^*\}$, enjoying the

following "interlacement property": there exists an integer $a \in [k]$ such that $a+1 < s^*-1$, $a \in B_{u^*}$ and $a+1 \in B_{\overline{u}}$. The block $B_{\overline{u}}$ can be either a singleton or a block with two or more elements. This situation is illustrated in Fig. 7, corresponding to the case $k = 8$ and $\pi = \{B_1, ..., B_5\} = \{\{1,2\}, \{3,5\}, \{4,6\}, \{7\}, \{8\}\}$. Here, $s^* = 7$, $B_{u^*} = B_3 = \{4,6\}$, $B_{\overline{u}} = B_2 = \{3,5\}$ and $a = 4$.

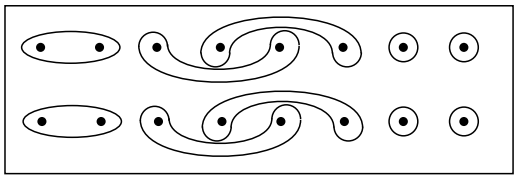

Figure 7: a singleton is adjacent to a 2-block with no consecutive elements

The crucial remark is now that, for a chain c as in (4.37) with partition π, one has that $i_{s^*} = i_{a+1}$. Indeed, a and s^*-1 both belong to B_{u^*}, and therefore $(i_{s^*-1}, i_{s^*}) = (i_a, i_{a+1})$. Since $a+1 \in B_{\overline{u}}$, this fact yields in particular that, $i_x = i_{s^*}$ for every $x \in B_{\overline{u}}$, that is, the left indices associated with $B_{\overline{u}}$ are completely determined by the choice of i_{s^*}. By the same argument, one shows that $j_{s^*} = j_{a+1}$. The rest of the proof is similar to the case B1. First, we observe that the representation (4.40), with $\mathcal{R}$ defined exactly as for B1, continues to be true, from which we deduce the estimate (4.41). It is now sufficient to show that each class $C_\pi^R(2k, N)$ has at most $k-1$ degrees of freedom. To do this, one chooses a representative element from each block $B_s \in \pi$, noted a_s, and then defines the sets

$$D = \{a_s : s = 1, ..., r,\ s \neq \overline{u}\},\ \ E = D \backslash \{a_s : s \in [r] \backslash P\},$$

that is, D is built by selecting one element from each block of π, except for $B_{\overline{u}}$, and E is obtained by subtracting from D all the remaining indices a such that $\{a\}$ is a singleton of π. One has that

$$|D| + |E| \leqslant k-1. \tag{4.42}$$

Indeed, $|D| = r-1 = |P|+|S|-1 \leqslant \frac{k-|S|}{2}+|S|-1$, and then one has to consider two cases: either (a) $B_{\overline{u}}$ is a singleton, from which it follows that $|E| = |D| - (|S|-1) \leqslant \frac{k-|S|}{2}$, or (b) $B_{\overline{u}}$ is not a singleton, yielding $|E| = |D| - |S| \leqslant \frac{k-|S|}{2} - 1$. In these two cases, (4.42) is then in order. To conclude, it remains to show that, for every $x_D = \{x_a : a \in D\} \in [N]^{|D|}$ and every $y_E = \{y_b : b \in E\} \in [N]^{|E|}$, there is at most one chain $c \in C_\pi^R(2k, N)$ as in (4.37) such that $i_a = x_a$ for every $a \in D$ and $j_b = y_b$ for every $b \in E$. To see this, assume that such a chain c exists, and observe that, due to the above considerations, the choice of x_D completely determines the upper sub-chain of c, as well as those indices j_x in the lower sub-chain such that $\{x\}$ is a singleton of π or (whenever $B_{\overline{u}}$ is not a singleton) such that $x \in B_{\overline{u}}$. Since the remaining left indices in the lower sub-chain of c are determined by the choice of y_E, the claim is proved. In view of (4.42), this shows that $C_\pi^R(2k, N)$ has at most $k-1$ free indices. This concludes the proof of Proposition 4.4.

As an illustration of the above arguments, one can consider the diagram in Fig. 8, that is constructed from the situation in Fig. 7 by selecting $U = V = \{2, 3, 4, 5\}$ and $R(2) = 4$, $R(3) = 5$, $R(4) = 2$ and $R(5) = 3$. In particular, it is easily seen that fixing i_4, i_7 and i_8 completely identifies a chain c inside the class $C_\pi^R(16, N)$, that has therefore three degrees of freedom.

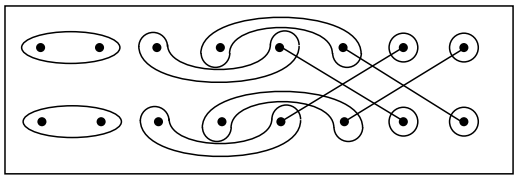

Figure 8: a class with three free indices

□

4.4 Proofs of Proposition 4.1 and Theorem 1.1

Proof of Proposition 4.1: We take up the notation introduced in Section 4.1. In view of Proposition 4.4, in order to prove relation (4.36) (and therefore Proposition 4.1), it is sufficient to show that, for every $\pi \in \mathcal{Q}(k)$, each pair $(\mathbf{i}, \mathbf{j}) \in G_N(\pi)$ is such that the corresponding chain $(i_1, i_2)...(i_k, i_1)(j_1, j_2)...(j_k, j_1)$ is an element of $C_\pi(2k, N)$, from which one deduces $|G_N(\pi)| \leqslant |C_\pi(2k, N)| \leqslant \Theta(k, \pi)N^{k-1}$. To show the desired property, it is enough to prove that, for every pair $(\mathbf{i}, \mathbf{j}) \in A_N(\pi) \times A_N(\pi)$ such that the chain $(i_1, i_2)...(i_k, i_1)(j_1, j_2)...(j_k, j_1)$ *is not* in $C_\pi(2k, N)$, one has that $(\mathbf{i}, \mathbf{j}) \notin G_N(\pi)$. By definition of $C_\pi(2k, N)$, we have to examine two cases. Start by considering a partition $\pi \in \mathcal{Q}(k)$ not containing any singleton: if $(\mathbf{i}, \mathbf{j}) \in A_N(\pi) \times A_N(\pi)$ is such that $(i_1, i_2)...(i_k, i_1)(j_1, j_2)...(j_k, j_1) \notin C_\pi(2k, N)$, then the random variables $X_{i_a i_{a+1}}$ indexed by the upper sub-chain are independent of those indexed by the lower sub-chain, and consequently

$$E(X_{i_1 i_2} \dots X_{i_k i_1} X_{j_1 j_2} \dots X_{j_k j_1}) = E(X_{i_1 i_2} \dots X_{i_k i_1}) E(X_{j_1 j_2} \dots X_{j_k j_1}),$$

yielding $(\mathbf{i}, \mathbf{j}) \notin G_N(\pi)$. On the other hand, if $\pi \in \mathcal{Q}(k)$ contains a singleton and if $(\mathbf{i}, \mathbf{j})$ is such that $(i_1, i_2)...(i_k, i_1)(j_1, j_2)...(j_k, j_1) \notin C_\pi(2k, N)$, then there exists $a = 1, ..., k$ such that $X_{i_a i_{a+1}}$ or $X_{j_a j_{a+1}}$ is independent of all the other variables indexed by the elements of the chain. This gives

$$E(X_{i_1 i_2} \dots X_{i_k i_1} X_{j_1 j_2} \dots X_{j_k j_1}) = E(X_{i_1 i_2} \dots X_{i_k i_1}) E(X_{j_1 j_2} \dots X_{j_k j_1}) = 0,$$

thus proving the required property $(\mathbf{i}, \mathbf{j}) \notin G_N(\pi)$. The proof is finished.

□

Proof of Theorem 1.1-(i): By virtue of the representation (1.13)–(1.14) and of Proposition 4.1, one sees that, for every $2 \leqslant k_1 < ... < k_m$, the limit in distribution of the vector

$$\left(\mathrm{Tr}(X_N), \mathrm{Tr}(X_N^{k_1}) - E\left[\mathrm{Tr}(X_N^{k_1})\right], ..., \mathrm{Tr}(X_N^{k_m}) - E\left[\mathrm{Tr}(X_N^{k_m})\right]\right)$$

coincides with the limit in distribution of

$$\left(N^{-1/2}\sum_{i=1}^{N} X_{ii},\ \ N^{-\frac{k_1}{2}}\sum_{\mathbf{i}\in D_N^{(k_1)}} X_{i_1i_2}X_{i_2i_3}\cdots X_{i_{k_1}i_1},\ \ldots,\ N^{-\frac{k_m}{2}}\sum_{\mathbf{i}\in D_N^{(k_m)}} X_{i_1i_2}X_{i_2i_3}\cdots X_{i_{k_m}i_1}\right),$$

so that the desired conclusion follows from Corollary 3.4. □

Proof of Theorem 1.1-(ii): For the simplicity of exposition, we assume that $k_1 \geqslant 2$, the proof when $k_1 = 1$ being completely similar and easier. We have, using the notation $D_N^{(k)}$ introduced in the beginning of Section 1.3 and using (1.14),

$$\left| E\left[\varphi\left(\frac{\mathrm{Tr}(X_N^{k_1}) - E[\mathrm{Tr}(X_N^{k_1})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_1}))}}, \ldots, \frac{\mathrm{Tr}(X_N^{k_m}) - E[\mathrm{Tr}(X_N^{k_m})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_m}))}}\right)\right] - E\left[\varphi\left(\frac{Z_{k_1}}{\sqrt{k_1}}, ..., \frac{Z_{k_m}}{\sqrt{k_m}}\right)\right]\right| \leqslant A_N + B_N,$$

where, by writing $\mathrm{Var}(\mathrm{Tr}(X_N^{k_j})) = C_j(N)$,

$$A_N = \left| E\left[\varphi\left(\frac{1}{C_1(N)^{1/2}N^{\frac{k_1}{2}}}\sum_{\mathbf{i}\in D_N^{(k_1)}} X_{i_1i_2}\ldots X_{i_{k_1}i_1}, \ldots, \frac{1}{C_m(N)^{1/2}N^{\frac{k_m}{2}}}\sum_{\mathbf{i}\in D_N^{(k_m)}} X_{i_1i_2}\ldots X_{i_{k_m}i_1}\right)\right] - E\left[\varphi\left(\frac{Z_{k_1}}{\sqrt{k_1}}, ..., \frac{Z_{k_m}}{\sqrt{k_m}}\right)\right]\right|$$

and

$$B_N = \left| E\left[\varphi\left(\frac{1}{C_1(N)^{1/2}N^{\frac{k_1}{2}}}\sum_{\mathbf{i}\in D_N^{(k_1)}} X_{i_1i_2}\ldots X_{i_{k_1}i_1}, \ldots, \frac{1}{C_m(N)^{1/2}N^{\frac{k_m}{2}}}\sum_{\mathbf{i}\in D_N^{(k_m)}} X_{i_1i_2}\ldots X_{i_{k_m}i_1}\right)\right] - E\left[\varphi\left(\frac{\mathrm{Tr}(X_N^{k_1}) - E[\mathrm{Tr}(X_N^{k_1})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_1}))}}, \ldots, \frac{\mathrm{Tr}(X_N^{k_m}) - E[\mathrm{Tr}(X_N^{k_m})]}{\sqrt{\mathrm{Var}(\mathrm{Tr}(X_N^{k_m}))}}\right)\right]\right|.$$

By combining Corollary 2.9 with the computations made in the proof of Proposition 3.1, we immediately get that $A_N = O(N^{-1/4})$. For B_N, we can write

$$\begin{aligned} |B_N| &\leqslant K\|\varphi'\|_\infty \sum_{j=1}^{m} E\left[N^{-\frac{k_j}{2}} \left| \sum_{\mathbf{i}\notin D_N^{(k_j)}} \left(X_{i_1 i_2} \dots X_{i_{k_j} i_1} - E[X_{i_1 i_2} \dots X_{i_{k_j} i_1}] \right) \right| \right] \\ &\leqslant K\|\varphi'\|_\infty \sum_{j=1}^{m} \sqrt{ \mathrm{Var}\left(N^{-\frac{k_j}{2}} \left| \sum_{\mathbf{i}\notin D_N^{(k_j)}} \left(X_{i_1 i_2} \dots X_{i_{k_j} i_1} - E[X_{i_1 i_2} \dots X_{i_{k_j} i_1}] \right), \right| \right) }, \end{aligned}$$

for some constant K not depending on N, so that $B_N = O(N^{-1/2}) = O(N^{-1/4})$ by Proposition 4.1.

□

Acknowledgments. We thank Paul Bourgade, Mikhail Lifshits, Gesine Reinert and Brian Rider for helpful discussions.

References

[1] H. Airault, P. Malliavin and F. Viens (2009). Stokes formula on the Wiener space and n-dimensional Nourdin-Peccati analysis. To appear in *J. Funct. Anal.*

[2] G. Anderson, A. Guionnet and O. Zeitouni (2009). An introduction to random matrices. Cambridge University Press.

[3] G. Anderson and O. Zeitouni (2006). A CLT for a band matrix model. *Probab. Theory Related Fields* **134**(2), 283338.

[4] Z.D. Bai and Y.Q. Yin (1986). Limiting behaviour of the norm products of random matrices and two problems of Geman-Hwang. *Probab. Theory Rel. Fields* **73**, 555–569.

[5] S. Chaterjee (2009). Fluctuation of eigenvalues and second order Poincar inequalities. *Probab. Theory Rel. Fields* **143**, 1–40.

[6] L.H.Y. Chen and Q.-M. Shao (2005). Stein's method for normal approximation. In: *An Introduction to Stein's Method* (A.D. Barbour and L.H.Y. Chen, eds), Lecture Notes Series No.**4**, Institute for Mathematical Sciences, National University of Singapore, Singapore University Press and World Scientific 2005, 1–59.

[7] A. Costin and J.L. Lebowitz (1995). Gaussian fluctuations in random matrices. *Physical Review Letters* **75**, 69–72.

[8] P. Diaconis and S.E. Evans (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353**, no. 7, 2615-2633.

[9] P. Diaconis and M. Shahshahani (1994). On the eigenvalues of random matrices. *J. Appl. Probab.* **31**, 49-62.

[10] P.J. Forrester (1999). Fluctuation formula for complex random matrices. *J. Phys. A* **32**, 159–163.

[11] S. Geman (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261.

[12] S. Geman (1986). The spectral radius of large random matrices. *Ann. Probab.* **14** 1318–1328.

[13] J. Ginibre (1965). Statistical ensembles of complex, quaternion and real matrices. *J. Math. Phys.* **6**, 440–449

[14] A. Guionnet (2008). *Large random matrices: lectures on macroscopic asymptotics.* LNM **1957**. Springer.

[15] S. Janson (1997). *Gaussian Hilbert Spaces.* Cambridge University Press, Cambridge.

[16] D. Marinucci and G. Peccati (2008). High-frequency asymptotics for subordinated stationary fields on an Abelian compact group. *Stochastic Processes and their Applications* **118**(4), 585-613.

[17] E. Mossel, R. O'Donnell and K. Oleszkiewicz (2008). Noise stability of functions with low influences: invariance and optimality. To appear in: *Ann. Math.*.

[18] I. Nourdin and G. Peccati (2009). Stein's method on Wiener chaos. *Probab. Theory Rel. Fields* **145**, no. 1, 75–118.

[19] I. Nourdin, G. Peccati and G. Reinert (2009). Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. Preprint.

[20] I. Nourdin, G. Peccati and A. Réveillac (2009). Multivariate normal approximation using Stein's method and Malliavin calculus. To appear in: *Ann. Inst. H. Poincaré Probab. Statist.*.

[21] D. Nualart (2006). *The Malliavin calculus and related topics.* Springer Verlag, Berlin, Second edition, 2006.

[22] D. Nualart and G. Peccati (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** (1), 177–193.

[23] G. Peccati and C.A. Tudor (2005). Gaussian limits for vector-valued multiple stochastic integrals. *Séminaire de Probabilités XXXVIII*, LNM **1857**. Springer-Verlag, Berlin Heidelberg New York, pp. 247–262.

[24] B. Rider (2004). Deviations from the circular law. *Probab. Theory Related Fields* **130**, 337–367

[25] B. Rider and J. Silverstein (1986). Gaussian fluctuations for non-Hermitian random matrix ensembles. *Ann. Probab.* **34**(6), 2118–2143

[26] B. Rider and B. Virág (2007). The noise in the circular law and the Gaussian free field. *Int. Math. Res. Not.* 2, Art. ID rnm006.

[27] V. I. Rotar' (1979). Limit theorems for polylinear forms. *J. Multivariate Anal.* **9**, 511–530.

[28] R. Stanley (1997). *Enumerative combinatorics, Vol. 1.* Cambridge University Press.

[29] T. Tao and V. Vu (2008). Random matrices: Universality of ESD and the Circular Law (with an appendix by M. Krishnapur). Preprint.

[30] T. Tao and V. Vu (2009). On the permanent of random Bernoulli matrices. *Adv. Math.* **220**, 657–669.