On quantum dynamics and statistics of vectors

1999 J. Phys. A: Math. Gen. 32 6547

(http://iopscience.iop.org/0305-4470/32/37/306)

J. Phys. A: Math. Gen. **32** (1999) 6547–6571. Printed in the UK PII: S0305-4470(99)05490-6

On quantum dynamics and statistics of vectors

M De Cock, M Fannes and P Spincemaille

Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Heverlee, Belgium

E-mail: mieke.decock@fys.kuleuven.ac.be, mark.fannes@fys.kuleuven.ac.be and pascal.spincemaille@fys.kuleuven.ac.be

Received 25 June 1999

Abstract. Given a sequence of vectors in a Hilbert space, we propose to use the spectrum of the associated Gram matrix as a tool for extracting statistical information on the sequence. We examine two simple models in some detail: the fractional shift where the sequence is generated by a deterministic unitary dynamics and random normalized vectors in a high-dimensional space chosen at a given density. In both cases, the limiting eigenvalue distribution of the Gram matrix is explicitly found. We relate our results to the notion of growth entropy and recover in the stochastic case the eigenvalue distribution of the Wishart matrices.

1. Introduction

Our concern in this paper is to extract some statistical information from sequences of normalized vectors in a Hilbert space. Such sequences arise naturally in quantum dynamics, either by a stroboscopic observation of the usual Schrödinger-type evolution of wavefunctions or by kicked evolutions where one records the state of the system immediately after each kick. Much wilder sequences, partly randomly generated, are also interesting in situations where it is hopeless to study the deterministic evolution.

It is common practice to analyse the time evolution of an initial wavefunction in terms of Wigner or Husimi functions, at least in situations allowing for such a description; e.g. quantum mechanics in terms of the usual position and momentum operators or spins with total angular momentum j. Coherent states often play a distinguished role. One also tries to relate geometrical properties of eigenfunctions of the Hamiltonian, such as the location of nodes, to orbits of the limiting classical system when $\overline{h}$ tends to zero (or j to infinity). Many references can be found in [3].

We shall not consider in this paper specific structures of a quantum dynamical system related to its classical limit, but rather consider statistical properties of time sequences of vectors, analogous to the relative frequencies of letters in long words. Unlike classical letters, which belong to a discrete set and have distance 1 from one another, quantum states are somewhat fuzzy and can almost coincide. The natural separation between two states φ and ψ is given by $1 - |\langle\varphi, \psi\rangle|^2$. It turns out that a notion of relative frequency can be obtained by considering the spectrum of the Gram matrix built on a sequence of vectors. One can, in particular, distinguish on the basis of the spectrum of the Gram matrix between initial states with short periods and states that wander wildly through the Hilbert space. This is argued in section 2.

0305-4470/99/376547+25$30.00

In the remainder of the paper, we consider essentially two models: a deterministic evolution which can be considered as a fractional shift and a random evolution obtained by choosing independently normalized vectors in an N-dimensional space according to the uniform distribution on the unit sphere. We compute in both cases the eigenvalue distribution of the Gram matrix when the length of the sequence tends to infinity.

In the deterministic case, it provides us with a notion of effective dimension occupied by the initial state during its evolution. This dimension, which is usually strictly smaller than the algebraic dimension, can be called entropic. In fact, we relate it to the growth entropy introduced in [14]. This is dealt with in section 4.

In the random case, the eigenvalue distribution of the Gram matrix of sequences of length τN tends as $N \to \infty$ to a deterministic distribution which was found by Marchenko and Pastur while studying the spectral distribution of the Wishart ensemble [10]. The parameter $\tau > 0$ can be interpreted as a rescaled time, which is short in comparison to the scale at which quantum effects are dominant but large with respect to the level separation in the spectrum of the Gram matrix. The non-classical limiting distribution is therefore a semi-classical trace of the quantum character of the initial system [4].

The limiting distribution is obtained in terms of its Stieltjes transform. In section 6, the expectation of the limiting eigenvalue distribution is computed in a combinatorial way. Section 7, written with L Pastur, contains the main result on the random case, namely the weak convergence with probability one of the eigenvalue distribution to its deterministic limit. The proof relies on an operator-analytic method: the resolvents of the Gram matrices are shown to converge in probability to the Stieltjes transform of the limiting measure.

2. Motivation and definitions

Truly quantum dynamical systems with compact phase space are finite-dimensional in virtue of the uncertainty principle. As each state occupies a same volume $\overline{h}$, the dimension of their Hilbert space of states is $N = 1/\overline{h}$. Planck's constant has here a rather symbolic meaning: for d-dimensional systems it is the dth power of the actual Planck constant, while $\overline{h} = 1/(2j+1)$ for a spin with angular momentum j. A quantum evolution in discrete time, also called kicked evolution, is determined by a unitary Floquet operator u. In the Schrödinger picture $\varphi \mapsto u\varphi$ is the evolution between two consecutive kicks. We face the problem of studying time sequences $\Phi = (f, uf, \ldots)$ generated by a Floquet operator u as it acts repeatedly on an initial condition f.

Intuitively, one expects the overlap $\langle f, u^n f\rangle$ between the initial state f and its time-evolved states to be small if the dynamics is sufficiently irregular. Quite the other extreme happens in case of a very regular dynamics: then $u^T f$ will be similar to f for some $T \ll N$.

In the following, we denote by $\mathcal{H}$ an N-dimensional Hilbert space with scalar product $\langle .\,, .\rangle$. We consider a sequence Φ of K vectors in $\mathcal{H}$:

$$\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_K).$$

Using the inner product on $\mathcal{H}$, we define a $K \times K$ Hermitian matrix Γ^Φ as follows:

$$\Gamma^\Phi := [\langle \varphi_i, \varphi_j\rangle]_{i,j=1,\ldots,K}. \tag{1}$$

This matrix we call the *Gram matrix*, built with the sequence of vectors Φ. If there is no confusion possible, we leave out the superscript Φ. Γ is positive semi-definite and its rank is the maximum number of independent vectors in Φ, i.e. the dimension of the space spanned by the vectors in Φ. In particular, Γ is non-singular if and only if the vectors $\varphi_1, \ldots, \varphi_K$ are independent. The spectrum of Γ, $\Sigma(\Gamma)$, is independent of the order of the φ_j in Φ and of

multiplying the φ_j with a complex number of modulus one. Moreover, it has a clear geometric interpretation: let $\{\gamma_1, \gamma_2, \ldots, \gamma_K\}$ be the eigenvalues of Γ, repeated according to multiplicity; then, its ℓth elementary symmetric invariant

$$\varepsilon_\ell := \sum_{\substack{\Lambda\subset\{1,\ldots,K\}\\ \#(\Lambda)=\ell}} \prod_{j\in\Lambda} \gamma_j$$

is equal to the sum of the squares of the ℓ-volumes of the parallelotopes spanned by $\{\varphi_j | j \in \Lambda\}$, where Λ runs through the size-ℓ subsets of $\{1, 2, \ldots, K\}$. For proofs, see [7]. Here we shall use Gram matrices for statistical purposes, namely for analysing properties of sequences of vectors, e.g. as generated by a quantum dynamics.

Let us for a moment consider the classical case, i.e. a sequence of natural numbers $\boldsymbol{i} = (i(1), i(2), \ldots, i(K))$, $i(j) \in \{1, \ldots, N\}$, which we identify with the sequence $\boldsymbol{e} := (e_{i(1)}, e_{i(2)}, \ldots, e_{i(K)})$ for an orthonormal basis $\{e_1, e_2, \ldots, e_N\}$ of the Hilbert space. Grouping $e_{i(\ell)}$ with equal index j, the Gram matrix is block-diagonal with blocks $E(j)$ of the type

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}.$$

The dimension of $E(j)$ is precisely the multiplicity $m(j)$ of j in $\boldsymbol{i}$. As the spectrum of $E(j)$ consists of the non-degenerate eigenvalue $m(j)$ and the $(m(j)-1)$-fold degenerated eigenvalue zero, we find that $\Sigma(\Gamma)$ determines precisely the amount of different numbers appearing in $\boldsymbol{i}$ with their multiplicity. The spectrum of the Gram matrix of a very regular sequence $\boldsymbol{i}$ will consist of a few large naturals and a highly degenerated zero, while for sequences with many different indices the spectrum will be concentrated on small natural numbers appearing with high multiplicities.

The same interpretation remains valid for the general non-commutative case: a Gram matrix with spectrum concentrated around small numbers points to a vector wandering wildly through the Hilbert space of the system and is a sign of chaotic behaviour. More regular motion, such as precession or slow diffusion, signals its presence by large eigenvalues and a high degeneracy of eigenvalues close to zero. In contrast to the classical case however, the non-zero eigenvalues of a Gram matrix are generically non-degenerate and eigenvalues are no longer limited to natural numbers so that a same vector can now be visited a fractional number of times.

To illustrate briefly the meaning of the spectrum of the Gram matrix, we consider a simple dynamics, which we shall study in more detail later. Consider the space $\mathcal{H} = \mathcal{L}^2(\mathbb{T}, \mathrm{d}\theta/2\pi)$ of square integrable functions on the one-dimensional torus on which we define

$$u : \mathcal{H} \to \mathcal{H} : f \mapsto uf(\theta) := \mathrm{e}^{\mathrm{i}\theta} f(\theta) \qquad 0 \leqslant \theta < 2\pi. \tag{2}$$

Starting from an initial condition $\varphi_1 = f$, we compute $\varphi_2 = uf, \varphi_3 = u^2 f, \ldots, \varphi_K = u^{K-1} f$ and construct the Gram matrix Γ^u of this sequence of vectors:

$$\Gamma^u_{\ell,m} = \langle u^{\ell-1} f, u^{m-1} f\rangle = \langle f, u^{m-\ell} f\rangle$$

where $m, \ell = 1, \ldots, K$.

If we choose for initial condition the constant function $f = 1$, then the dynamics generates an orthonormal set in $\mathcal{H}$ and the spectrum of the Gram matrix consists of the K-fold degenerated value one. There is strictly no overlap between vectors at different times. If we choose, e.g., for initial condition $f(\theta) = \sqrt{2}\chi_{[0,\pi]}(\theta)$, the characteristic function of $[0, \pi]$, then we are dealing

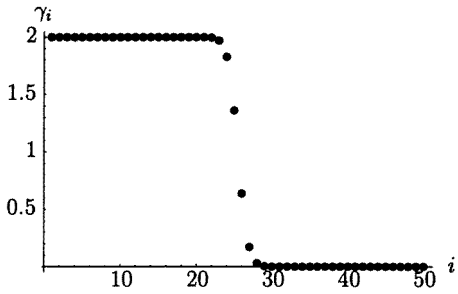

Figure 1. Eigenvalues of the Gram matrix for the dynamics (2) ($K = 50$).

with the 'square root' of the previous shift. For this initial condition, the matrix elements of Γ^u can easily be computed:

$$\Gamma^u_{\ell,m} = \begin{cases} 1 & \text{if} \quad m = \ell \\ 0 & \text{if} \quad m - \ell \text{ even and not equal to 0} \\ 2\mathrm{i}/\pi(m-l) & \text{if} \quad m - \ell \text{ odd.} \end{cases}$$

Computing numerically the eigenvalues of Γ is not difficult. For instance, for $K = 50$ they are plotted in figure 1. We see that approximately half the number of eigenvalues lie around two, the other half lie close to zero and a few are in between these two values. Roughly speaking, this means that every vector is met approximately twice.

3. The Stieltjes transform

The first objective in studying Gram matrices is to obtain the asymptotic eigenvalue distribution when the dimension of the matrix tends to infinity. For every natural K we have a self-adjoint K-dimensional matrix Γ^Φ with eigenvalues $\lambda_1^K, \ldots, \lambda_K^K$, to which we associate the empirical eigenvalue distribution

$$\rho_K(\mathrm{d}x) := \frac{1}{K}\sum_{i=1}^{K} \delta(x - \lambda_i^K)\,\mathrm{d}x \tag{3}$$

on $\mathbb{R}$. ρ_K is also called integrated density of states or eigenvalue counting measure. We are interested in finding the limiting eigenvalue distribution

$$\rho(\Delta) := \lim_{K\to\infty} \rho_K(\Delta) = \lim_{K\to\infty} \frac{1}{K}\#\{\lambda_i^K \in \Delta\}$$

where Δ is any Borel subset of $\mathbb{R}$, i.e. in finding the weak limit of ρ_K as $K \to \infty$.

Instead of directly finding the limiting distribution, we shall focus on its Stieltjes transform. For any probability measure ν on $\mathbb{R}$ the complex function $I(z, \nu)$ defined by

$$I(z, \nu) := \int_{\mathbb{R}} \nu(\mathrm{d}t)\frac{1}{t - z} \qquad \mathrm{Im}\,(z) \neq 0$$

is called the *Stieltjes transform* of ν, where $z \mapsto I(z, \nu)$

- is analytic on $\mathbb{C} \setminus \mathbb{R}$
- satisfies $\mathrm{Im}\,(I(z, \nu)) \cdot \mathrm{Im}\,(z) > 0$ for $\mathrm{Im}\,(z) \neq 0$ and (4)
- behaves asymptotically as $I(z, \nu) = -1/z + \mathrm{o}((\mathrm{Im}\,z)^{-1})$.

These three conditions are also sufficient to guarantee that I is the Stieltjes transform of a probability measure. The following result [13] (Perron–Frobenius) states the inverse operation in terms of the distribution function $n(x) := \nu(]-\infty, x])$ of the probability measure ν. Recall that n

- is monotonically increasing
- tends to zero at $-\infty$ and to one at ∞ and
- is right-continuous.

For $t_0 < t_1$

$$\tfrac{1}{2}(n(t_1) + n(t_1^-)) - \tfrac{1}{2}(n(t_0) + n(t_0^-)) = \lim_{\epsilon \downarrow 0} \frac{1}{2\pi \mathrm{i}} \int_{t_0}^{t_1} \mathrm{d}t \, (I(t + \mathrm{i}\epsilon, \nu) - I(t - \mathrm{i}\epsilon, \nu)) \tag{5}$$

where $n(t^-)$ means $\lim_{s \uparrow t} n(s)$, which exists by the monotonicity of n. In the case of a ν that is absolutely continuous with respect to the Lebesgue measure and with a Hölder-continuous density ρ, one finds

$$\lim_{\epsilon \downarrow 0} I(x \pm \mathrm{i}\epsilon, \nu) = \lim_{\epsilon \downarrow 0} \int_{\mathbb{R}} \mathrm{d}t \, \frac{\rho(t)}{t - x \mp \mathrm{i}\epsilon} = P \int_{\mathbb{R}} \mathrm{d}t \, \frac{\rho(t)}{t - x} \pm \mathrm{i}\pi\rho(x)$$

where $P \int$ indicates the Cauchy principal value. It follows immediately that

$$\rho(x) = \lim_{\epsilon \downarrow 0} \frac{1}{2\pi \mathrm{i}} (I(x + \mathrm{i}\epsilon, \nu) - I(x - \mathrm{i}\epsilon, \nu)).$$

Using the Stieltjes transform of a probability measure rather than the characteristic function (Fourier transform) can be quite convenient, especially in the case of probability measures with compact support. There is a direct connection between the Stieltjes transform of the empirical eigenvalue distribution (3) ρ_K and the resolvent $G_K(z)$ of Γ^Φ through the spectral theorem

$$I(z, \rho_K) := \int_{\mathbb{R}} \rho_K(\mathrm{d}\lambda) \frac{1}{\lambda - z} = \frac{1}{K} \operatorname{Tr} G_K(z) \tag{6}$$

where

$$G_K(z) := (\Gamma^\Phi - z)^{-1}.$$

Weak convergence of ρ_K as $K \to \infty$ can then be studied in terms of pointwise convergence of the $I(z, \rho_K)$ for $\operatorname{Im}(z) \neq 0$.

The Stieltjes transform of the limiting eigenvalue distribution may be obtained as a solution of an algebraic equation, as we do in section 7. A second way to obtain the Stieltjes transform of the limiting eigenvalue distribution is through the calculation of moments. This is based on the following calculation:

$$\begin{aligned} I(z, \rho_K) &= \int_{\mathbb{R}} \rho_K(\mathrm{d}t) \frac{1}{t - z} \\ &= -\sum_{n=0}^{\infty} \frac{1}{z^{n+1}} \int_{\mathbb{R}} \rho_K(\mathrm{d}t) t^n = -\sum_{n=0}^{\infty} \frac{m_n(\rho_K)}{z^{n+1}} \end{aligned}$$

where $m_n(\rho_K)$ is the nth moment of the measure ρ_K

$$m_n(\rho_K) := \int_{\mathbb{R}} \rho_K(\mathrm{d}t) t^n = \frac{1}{K} \operatorname{Tr}(\Gamma^\Phi)^n.$$

We assume now that all the following limits exist and are finite:

$$m_n := \lim_{K \to \infty} m_n(\rho_K). \tag{7}$$

Is there a probability measure having these numbers as its moments? This problem is known as the *moment problem* [1, 13]. In general, an equivalent condition for a given sequence $(m_n)_n$

to form a sequence of moments of a probability measure is the positivity of the determinants ($p \in \mathbb{N}$) [12, 13]

$$\begin{vmatrix} m_0 & m_1 & \dots & m_p \\ m_1 & m_2 & \dots & m_{p+1} \\ \dots & & \ddots & \\ m_p & m_{p+1} & \dots & m_{2p} \end{vmatrix}.$$

This is a condition for the existence of a probability measure and it is obviously met by the limiting moments of (7). The moment problem is called *determined* if this measure is unique, in all other cases it is called indeterminate. There exist several sufficient conditions for determinacy on the sequence $(m_n)_n$: e.g., if the moment problem for a given sequence has a solution of which the spectrum is a bounded set, then it is determined (i.e. this solution is unique). A general discussion of the moment problem can be found in [1, 13], in which the construction of the corresponding measure is also discussed.

Theorem 1 ([13]). *If ρ is a solution of the moment problem, then $I(z, \rho)$ is analytic in the upper complex plane* $\operatorname{Im}(z) > 0$. $\operatorname{Im} I(z, \rho) > 0$ *on the same half-plane and is asymptotically represented by the series $-\sum_0^\infty m_n z^{-n-1}$ in any sector $\epsilon \leqslant \arg(z) \leqslant \pi - \epsilon$ with $0 < \epsilon < \pi/2$: i.e., in these sectors one can write for $z \to \infty$*

$$I(z, \rho) + \left(\frac{m_0}{z} + \dots + \frac{m_{n-1}}{z^n} \right) = o(z^{-n}) \qquad n = 1, 2, \dots. \tag{8}$$

Conversely, if a complex function $f(z)$ is analytic in the upper complex plane $\operatorname{Im}(z) > 0$, $\operatorname{Im} f(z) > 0$ *and $f(z)$ is asymptotically represented by the series $-\sum_0^\infty m_n z^{-n-1}$ in any sector $0 < \epsilon \leqslant \arg(z) \leqslant \pi - \epsilon$ with $0 < \epsilon < \pi/2$, then there exists a unique solution ρ of the moment problem such that $f(z) = I(z, \rho)$.*

It is clear that the solution of the moment problem is now reduced to finding a function $f(z)$ which has the asymptotic representation of above. Via the Perron–Frobenius inversion theorem (5) one can then recover the probability measure ρ. Back to our situation, where we assumed the moments to be finite, we can then use the following theorem by Fréchet and Shohat.

Theorem 2.

(i) Let the probability measures ρ_K possess finite moments $m_n^K := \int_{\mathbb{R}} \rho_K(\mathrm{d}t) t^n$ for $n, K \in \mathbb{N}$. Assume that the limits $m_n := \lim_{K\to\infty} m_n^K$ exist for each n, then the m_n are the moments of a probability measure ρ and, if this ρ is unique, then ρ_K converges weakly to ρ.

(ii) A sequence $(\rho_K)_K$ of probability measures converges weakly to a probability measure ρ if and only if the sequence $(I(z, \rho_K))_K$ of their corresponding Stieltjes transforms converges uniformly on compact subsets of the upper complex half-plane to a function $f(z)$ that satisfies (4). Moreover, f is the Stieltjes transform of ρ.

4. The Gram matrix and approximation entropies

In this section, we examine in more detail the example (2) of the unitary evolution that was already mentioned at the end of section 2. We compute the distribution of the spectrum of Gram's matrix Γ^u when the dimension, i.e. the number of time steps, goes to infinity. We recall Voiculescu's growth entropy and connect it to the spectrum of Γ^u and finally compute the entropy for this specific example.

We consider $\mathcal{H} = \mathcal{L}^2(\mathbb{T}, \mathrm{d}\theta/2\pi)$ on which $u : \mathcal{H} \to \mathcal{H}$ acts like $(uf)(\theta) := \mathrm{e}^{\mathrm{i}\theta} f(\theta)$. The matrix elements of the Gram matrix Γ are then given by

$$\Gamma_{\ell,m} = g(\ell - m) := \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \, \mathrm{e}^{\mathrm{i}(m-\ell)\theta} \, |f(\theta)|^2 \tag{9}$$

for a given initial condition f. Note that Γ is constant along parallels to the diagonal. We call $\Gamma(K)$ the $K \times K$ matrix obtained by cutting out the first K rows and columns of Γ, i.e.

$$\Gamma(K)_{\ell,m} := g(\ell - m) \qquad \ell, m = 1, 2, \ldots, K.$$

The limiting eigenvalue distribution of the $\Gamma(K)$ can be obtained through the following version of Szegö's theorem [6].

Theorem 3. *Denote by $(h\hat{}(n))_{n\in\mathbb{Z}}$ the Fourier coefficients of a real-valued function $h \in \mathcal{L}^\infty(\mathbb{T})$ and for $K = 1, 2, \ldots$ let $T(K)$ be the $K \times K$ matrix $[h\hat{}(j-i)]_{i,j=1,2,\ldots,K}$, then the empirical eigenvalue distribution ρ_K of $T(K)$ converges weakly to ρ when $K \to \infty$, where the measure ρ is determined by*

$$\rho(\Delta) = \frac{1}{2\pi} \int_{h^{-1}(\Delta)} \mathrm{d}\theta$$

on any Borel subset Δ of $\mathbb{T}$. That is, for any continuous complex function F defined on the essential support of h, one has

$$\lim_{K\to\infty} \frac{1}{K} \sum_{j=1}^{K} F(\gamma_j^K) = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \, F(h(\theta))$$

where the γ_j^K are the eigenvalues of $T(K)$.

We shall use theorem 3, with $h = |f|^2$, to relate Voiculescu's approximation entropy with the spectral distribution of the Gram matrix. This approximation entropy is introduced as a growth entropy [14], reflecting the mathematical idea of finding the minimal growth in dimension of a finite-dimensional subspace that is needed to follow within a small error bar the evolution of an initial condition under the dynamics of the system. This is to be contrasted with observational entropies [2, 5] that are introduced on the basis of measurements on the system, such as computing transition probabilities between states at different times, as is done in the construction of Gram matrices.

The Hilbert space version of the approximation entropy is introduced as follows: consider a complex Hilbert space $\mathcal{H}$ of infinite dimension. For Φ a finite subset of $\mathcal{H}$ and $A \subset \mathcal{H}$, we write $\Phi \subset_\delta A$ if $\|\Phi - A\| < \delta$. For a finite subset Φ of $\mathcal{H}$ and $\varepsilon > 0$, the approximate dimension of Φ is

$$\dim(\Phi, \varepsilon) := \inf\{\dim(V) | V \text{ finite-dimensional subspace of } \mathcal{H} \text{ and } \Phi \subset_\varepsilon V\}$$

and the Voiculescu approximation entropy, given the limit exists, is

$$S^V(\varepsilon) := \lim_{K\to\infty} \frac{1}{K} \dim(\Phi, \varepsilon).$$

We want to make a connection between this entropy and the spectrum of the Gram matrix Γ^Φ generated with the finite set of vectors in the Hilbert space. So, starting from an initial condition $f \in \mathcal{H}$, we generate $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_K)$ by $\varphi_i = u^{i-1} f$ where u is a unitary acting on $\mathcal{H}$. The eigenvalues of Γ^Φ, arranged in increasing order and repeated according to their multiplicity, are γ_j with corresponding eigenvectors g_j

$$\Gamma^\Phi g_j = \gamma_j g_j \qquad j = 1, \ldots, K.$$

We put

$$C_\Phi^\varepsilon := \#\{\lambda | \lambda \in \Sigma(\Gamma^\Phi), \lambda \geqslant \varepsilon\} \tag{10}$$

and

$$S(\varepsilon) := \lim_{K\to\infty} \frac{1}{K} C_\Phi^\varepsilon.$$

$S(\varepsilon)$ is a Boltzmann-like entropy in the sense that it measures the extra volume per time unit occupied by the Gram matrix and we shall prove that, in the limit $\varepsilon \downarrow 0$, both entropies coincide.

Lemma 1. *Let $\Phi = (\varphi_1, \varphi_2, \ldots, \varphi_K)$ be a K-tuple of normalized vectors in a Hilbert space $\mathcal{H}$ and Γ^Φ the corresponding Gram matrix as in (1) with eigenvalues γ_j in ascending order and repeated according to multiplicity; then*

$$\inf_{V,\dim V=\ell} \max_{j=1,\ldots,K} \|\varphi_j - V\| \geqslant \frac{1}{\sqrt{K}} \sqrt{\sum_{j=1}^{K-l} \gamma_j}.$$

Proof. It is obvious that we can limit ourselves to ℓ-dimensional subspaces V of the subspace of $\mathcal{H}$ spanned by Φ. Suppose the infimum is reached for $V = V_0$ and call it ε. So,

$$\inf_{V,\dim V=\ell} \max_{j=1,\ldots,K} \|\varphi_j - V\| = \varepsilon.$$

If we denote by $[V]$ the projector on the space V, we can write

$$\|\varphi_j - V_0\|^2 = 1 - \|[V_0]\varphi_j\|^2$$

and so

$$\varepsilon^2 \geqslant 1 - \frac{1}{K} \sum_{j=1}^{K} \|[V_0]\varphi_j\|^2.$$

In order to minimize the right-hand side, we have to maximize the sum. Thus, we should check that

$$\sum_{j=1}^{K} \|[V_0]\varphi_j\|^2 \leqslant \sup_{V,\dim V=\ell} \sum_{j=1}^{K} \|[V]\varphi_j\|^2.$$

Suppose $\{\psi_\alpha\}_{\alpha=1,\ldots,\ell}$ is an orthonormal basis for V, which we may assume to be contained in the space spanned by Φ, and $\psi_\alpha = \sum_{j=1}^{K} \lambda_{\alpha_j} \varphi_j$. Let $\{e_1, \ldots, e_K\}$ be the canonical basis in $\mathbb{C}^K$. The assumption that $\{\psi_\alpha\}_{\alpha=1,\ldots,K}$ is orthonormal is equivalent to $\langle \lambda_\alpha, \Gamma^\Phi \lambda_\beta \rangle = \delta_{\alpha,\beta}$ where $\lambda_\alpha := \sum_{j=1}^{K} \lambda_{\alpha_j} e_j$. It is easy to see that

$$\sum_{j=1}^{K} \|[V]\varphi_j\|^2 = \sum_{\alpha=1}^{\ell} \|\Gamma^\Phi \lambda_\alpha\|^2.$$

If we call $\mu_\alpha = \Gamma^{\frac{1}{2}} \lambda_\alpha$, we see that

$$\sup_{V,\dim V=\ell} \sum_{j=1}^{K} \|[V]\varphi_j\|^2 = \sup_{\langle \mu_\alpha, \mu_\beta \rangle = \delta_{\alpha,\beta}} \sum_{\alpha} \|\Gamma^{\frac{1}{2}} \mu_\alpha\|^2.$$

By the mini–max principle, this is always greater than or equal to the sum of the ℓ largest eigenvalues of Γ^Φ. □

Theorem 4. *Suppose $S^V(\varepsilon)$ and $S(\varepsilon)$ are continuous functions of ε and suppose that the limits* $\lim_{\varepsilon\downarrow 0} S^V(\varepsilon)$ *and* $\lim_{\varepsilon\downarrow 0} S(\varepsilon)$ *exist. Then*

$$\lim_{\varepsilon\downarrow 0} S^V(\varepsilon) = \lim_{\varepsilon'\downarrow 0} S(\varepsilon').$$

Proof. In a first step, we prove that $S^V(\sqrt{\varepsilon}) \leqslant S(\varepsilon), \forall\varepsilon$. Using the same notations as in the proof of lemma 1, we therefore define the vectors

$$\eta_j := \sum_k \langle e_k, g_j\rangle \varphi_k$$

and the projector

$$P_\varepsilon := \sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1} |\eta_j\rangle\langle\eta_j|.$$

Then we prove that $\|\varphi_\ell - P_\varepsilon\varphi_\ell\| \leqslant \sqrt{\varepsilon}, \forall \ell = 1, \ldots, K$. This implies that every vector $\varphi_\ell \in \Phi$ can be approximated by an element of the C_Φ^ε-dimensional subspace (10) spanned by the η_j.

$$\begin{aligned}
\|\varphi_\ell - P_\varepsilon\varphi_\ell\|^2 &= \|\varphi_\ell\|^2 - \sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_j\rangle - \sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}\langle\varphi_\ell,\eta_j\rangle\langle\eta_j,\varphi_\ell\rangle \\
&\quad + \sum_{j,j',\gamma_j\geqslant\varepsilon,\gamma_{j'}\geqslant\varepsilon} \gamma_j^{-1}\gamma_{j'}^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_{j'}\rangle\langle\eta_{j'},\eta_j\rangle \\
&= \|\varphi_\ell\|^2 - 2\sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_j\rangle \\
&\quad + \sum_{j,j',\gamma_j\geqslant\varepsilon,\gamma_{j'}\geqslant\varepsilon} \gamma_j^{-1}\gamma_{j'}^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_{j'}\rangle \sum_{k,k'}\langle e_k, g_j\rangle\langle g_{j'}, e_{k'}\rangle\langle\varphi_{k'},\varphi_k\rangle \\
&= \|\varphi_l\|^2 - 2\sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_j\rangle \\
&\quad + \sum_{j,j',\gamma_j\geqslant\varepsilon,\gamma_{j'}\geqslant\varepsilon} \gamma_j^{-1}\gamma_{j'}^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_{j'}\rangle\langle g_{j'},\Gamma g_j\rangle \\
&= \|\varphi_\ell\|^2 - \sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}\langle\eta_j,\varphi_\ell\rangle\langle\varphi_\ell,\eta_j\rangle \\
&= \left\langle\varphi_\ell, \left(\mathbb{1} - \sum_{j,\gamma_j\geqslant\varepsilon} \gamma_j^{-1}|\eta_j\rangle\langle\eta_j|\right)\varphi_\ell\right\rangle.
\end{aligned}$$

Because $\{g_1, \ldots, g_n\}$ is an orthonormal basis, we have that

$$\|\varphi_\ell\|^2 = \langle\varphi_\ell,\varphi_\ell\rangle = \Gamma_{\ell,\ell} = \langle e_\ell, \Gamma e_\ell\rangle = \sum_j \langle e_\ell, g_j\rangle\langle g_j, \Gamma e_\ell\rangle$$

but

$$\left|\sum_{j,\gamma_j<\varepsilon}\langle e_\ell, g_j\rangle\langle g_j,\Gamma e_\ell\rangle\right| \leqslant \sum_{j,\gamma_j<\varepsilon} |\langle e_\ell, g_j\rangle||\langle g_j,\Gamma e_\ell\rangle| \leqslant \varepsilon\sum_j |\langle e_\ell, g_j\rangle|^2 \leqslant \varepsilon.$$

Secondly, we prove $\lim_{\varepsilon'\downarrow 0} S^V(\varepsilon') \geqslant \lim_{\varepsilon\downarrow 0} S(\varepsilon)$. Suppose that this is not true. This means that locally, in the neighbourhood of zero, figure 2 reflects the relation between both entropies, where $S^V(\varepsilon) = S(g(\varepsilon))$.

Assuming $\lim_{\varepsilon\downarrow 0} S(\varepsilon) > \lim_{\varepsilon'\downarrow 0} S^V(\varepsilon')$ means assuming $g(0) > 0$. Because of the definition of g, it follows that $S^V(0) = S(g(0))$ or

$$\begin{aligned}
&\inf\{\dim V | V \text{ finite-dimensional subspace of } \mathcal{H}, \Phi \subset_0 V\} \\
&\quad = \#\{\gamma|\gamma\in\Sigma(\Gamma) \text{ and } \gamma \geqslant g(0)\} + o(K) \\
&\quad = C_\Phi^{g(0)} + o(K) =: s(K).
\end{aligned}$$

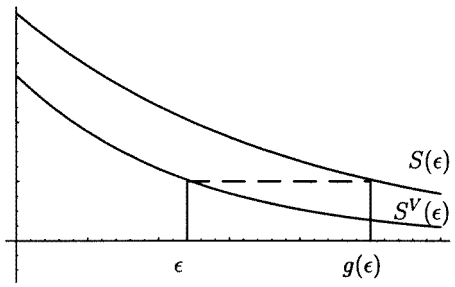

Figure 2. Definition of $g(\epsilon)$.

So, if we take $s(K)$ dimensions to approximate, we make no error. On the other hand we know, because of lemma 1, that the error is at least

$$\sqrt{\frac{1}{K}\sum_{i=1}^{K-s(K)}\gamma_i} = \sqrt{\frac{1}{K}\left(\sum_{i=1}^{K-C_\Phi^{g(0)}}\gamma_i - \sum_{K-C_\Phi^{g(0)}+o(K)+1}^{K-C_\Phi^{g(0)}}\gamma_i\right)}.$$

As we assumed $g(0)$ to be strictly positive, we know that there is a $\widetilde{g(0)}$, $0 < \widetilde{g(0)} < g(0)$ such that $S(\widetilde{g(0)}) = (S(0) + S^V(0))/2$. This means that the first term in the square root is always greater than or equal to $\widetilde{g(0)}(C_\Phi^{\widetilde{g(0)}} - C_\Phi^{g(0)})/K$ which in the limit $K \to \infty$ goes to $\widetilde{g(0)}(S(0) - S^V(0))/2$. The second term in the square root tends to zero, because it contains a non-extensive number of eigenvalues. This means that we end up with a contradiction. □

Finally, we want to compute $S(\varepsilon)$ for the unitary u defined in (2). To compute $S(\varepsilon)$, we should count how many eigenvalues of Γ built with $(f, uf, \ldots)$ are larger than ε. This can be simply done by using theorem 3. It suffices to consider a sequence of continuous functions with support, say $[1/2n, \|f\|_\infty^2]$, and which are equal to one on $[1/n, \|f\|_\infty^2]$; then

$$\lim_{\varepsilon\downarrow 0} S(\varepsilon) = \frac{1}{2\pi}|\text{Ess. Supp}(f)|$$

where 'Ess. Supp(f)' is the essential support of f.

5. A random vector model

In this section, we focus on a stochastic dynamics: instead of considering a unitary repeatedly acting on an initial condition, we study the spectrum of a Gram matrix constructed with randomly generated vectors $(\varphi_1, \ldots, \varphi_K)$ in an N-dimensional Hilbert space $\mathcal{H}$. We can see the case $\tau := K/N \ll 1$ as mimicking, in a way, the situation of a discrete time chaotic dynamics. Indeed, for $K \ll N$, we can expect that there is almost no overlap between the different vectors as we think will be the case in the chaotic situation. By increasing τ however, the chance to pick overlapping vectors increases. This should give rise to more and more small and large values in the spectrum of Γ^Φ. As for Gram matrices constructed with an initial vector f and its time evolutions $u^n f$, one can expect that after N time steps, so N vectors in an N-dimensional Hilbert space, in a way, all degrees of freedom are used and the space is completely filled.

The spectrum of Γ^Φ is, of course, a random object but it turns out that, in the limit $K, N \to \infty$ and τ constant, the spectral distribution tends with probability one to a deterministic limit given by the Marchenko–Pastur probability distribution ρ_τ.

We start the section by describing what is meant by randomly generated vectors and consider briefly the classical analogue. We then turn to the determination of the limiting eigenvalue distribution of the random Gram matrices in closed form and present two different approaches. Both of them rely on the computation of the Stieltjes transform of the limiting eigenvalue distribution. A first head-on computation provides us with a recursion relation between limiting moments from which we determine the generating function; the second approach consists of obtaining a closed equation for the Stieltjes transform considering the resolvent of the Gram matrix.

The Gram matrix is constructed with random vectors independently picked in a finite (N)-dimensional Hilbert space according to a same N-dependent distribution μ_N. There are several possibilities for describing such a probability distribution and we shall use the two following: either specify the distribution of the components of the vector with respect to a given basis or specify for $k = 1, 2, \ldots$ the density matrices

$$\int \mu_N(\mathrm{d}\varphi)|\varphi^{\otimes k}\rangle\langle\varphi^{\otimes k}|$$

on $\mathbb{C}^N$. Here, we shall only consider the unique probability distribution on the surface of the unit sphere in $\mathbb{C}^N$ that is invariant under all unitary transformations. This probability measure, called uniform, will from here on be denoted by μ_N. The two descriptions above now become either

$$\int \mu_N(\mathrm{d}\varphi)|\varphi^{\otimes k}\rangle\langle\varphi^{\otimes k}| = \frac{1}{\binom{N+k-1}{k}} P_k^{\mathrm{sym}} \tag{11}$$

where P_k^{sym} denotes the projector onto the fully symmetric subspace of the k-fold tensor power of $\mathbb{C}^N$, or

$$\varphi = \frac{x}{\|x\|} \tag{12}$$

where x is an N-dimensional complex Gaussian random vector with identically and independently distributed components of mean zero and variance one. By this we mean that each of the components x_α of x, $\alpha = 1, \ldots, N$, is of the form $x_\alpha = y_\alpha + \mathrm{i}z_\alpha$ where y_α and z_α are independent random variables, normally distributed, with mean zero and common arbitrary variance $\frac{1}{2}$. Due to the special form of the Gaussian distribution, we see that in terms of generalized spherical coordinates, the distribution of x factorizes into a product of a radial part and an angular part. From (12) it is immediately clear that the distribution of φ is uniform on the unit sphere in $\mathbb{C}^N$.

Before considering the Gram matrix of randomly generated vectors, we consider the classical example presented at the beginning of section 2 which was used to give some intuitive feeling about spectra of Gram matrices.

Consider a discrete set with N elements $\{1, \ldots, N\}$. Out of this set, we choose randomly K numbers, independently and identically distributed according to the probability distribution $\Lambda = (\lambda_1, \ldots, \lambda_N)$. How many times j is chosen will be denoted by k_j. The draw $(k_1, k_2, \ldots, k_N)$, $k_1 + k_2 + \cdots + k_N = K$, where we choose k_j times j, can be realized in

$$\binom{K}{k_1, k_2, \ldots, k_N}$$

different ways. The chance of this configuration to appear is $\lambda_1^{k_1} \ldots \lambda_N^{k_N}$. Associating with j the unit vector e_j in the orthonormal basis $\{e_1, e_2, \ldots, e_N\}$, a particular draw $\boldsymbol{i} = (i_1, \ldots, i_k)$ defines a Gram matrix with matrix elements

$$\Gamma_{k,\ell} = \langle e_{i_k}, e_{i_\ell}\rangle = \delta_{i_k, i_\ell}.$$

The spectrum $\Sigma(\Gamma)$ of Γ determines precisely how many different numbers appear in i together with their multiplicities. An interesting question is: How many different points are visited on average? In order to compute the expectation of the number of different elements appearing in a particular choice, we consider

$$g : \{0, \ldots, K\} \to \{0, 1\} : g(k) = 1 \qquad \text{for} \quad k \geqslant 1$$
$$g(0) = 0.$$

The expectation of the number of different points which are chosen, which we denote by #, equals

$$\mathbb{E}(\#) = \sum_{(k_1,\ldots,k_N)} \{g(k_1) + \cdots + g(k_N)\} \lambda_1^{k_1} \ldots \lambda_N^{k_N} \binom{K}{k_1, k_2, \ldots, k_N}$$
$$= N - \sum_{j=1}^{N} (1 - \lambda_j)^K.$$

Taking the distribution uniform, i.e. $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 1/N$, and $K/N = \tau$, we find

$$\lim_{K\to\infty} \frac{1}{K} \mathbb{E}(\#) = \frac{1}{\tau}(1 - \mathrm{e}^{-\tau}).$$

It is easy to check that, in this classical case, in the limit $N \to \infty$, the eigenvalues of Γ are Poisson distributed.

In the following, we shall be interested in the limiting eigenvalue distribution of the *random Gram matrix ensemble* of density $\tau \geqslant 0$ which is defined as follows. Consider an N-dimensional Hilbert space $\mathcal{H}$, in which we choose randomly $K = \tau N$ vectors $\Phi = \{\varphi_1, \ldots, \varphi_K\}$ according to the uniform distribution μ_N (11), (12). With these vectors we construct the Gram matrix (1)

$$\Gamma^{\Phi}_{i,j} = \langle \varphi_i, \varphi_j \rangle$$

and we are interested in the limiting eigenvalue distribution of Γ^{Φ} when $N \to \infty$ and τ remains constant.

6. Moments of random Gram matrices

As was explained in section 3, we can calculate the limiting eigenvalue distribution of the ensemble of random Gram matrices by computing the generating function (8) for the moments. So, we shall compute

$$m_n := \lim_{N\to\infty} \frac{1}{K} \mathbb{E}_N(\mathrm{Tr}\, \Gamma^n)$$

where $\mathbb{E}_N$ means expectation with respect to the probability measure μ_N. It turns out that these moments can be expressed in terms of the number of non-crossing partitions on a circle. This will provide us with a closed equation for the generating function of the moments and with the limiting eigenvalue distribution through the Perron–Frobenius inversion formula. Once we have an explicit expression for the generating function (5), we can also explicitly compute the moments of the limiting distribution. We start by introducing and defining some of the notions we need later on.

A *partition of a set* $S = \{1, \ldots, s\}$, $s \in \mathbb{N}$, *on a circle* is defined as follows: consider a regular s-gon on a circle. Choose a starting point and label it with 1, label the following points by $2, 3, \ldots, s$, in the clockwise direction. A partition is now a drawing of polygons with these points as vertices. It is called *non-crossing* if none of the edges of the different

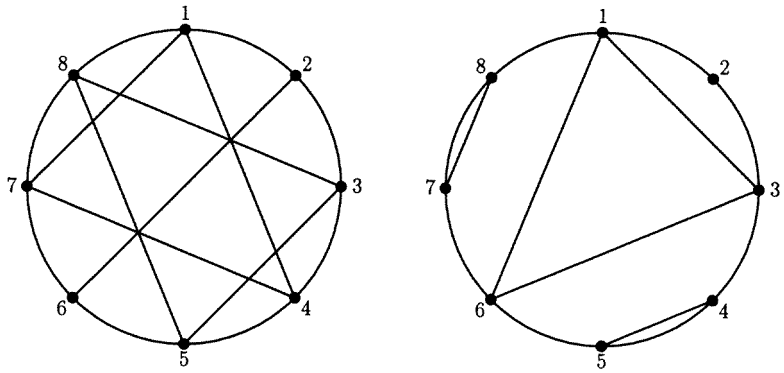

Figure 3. An example of a crossing and a non-crossing partition.

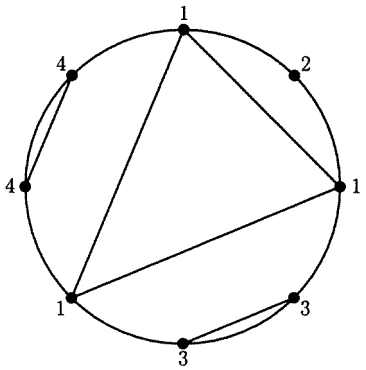

Figure 4. The non-crossing partition of figure 3 relabelled.

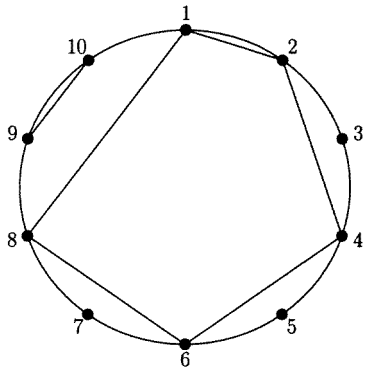

Figure 5. In this partition, the first polygon has corners $\{1, 2\}$, $\{4\}$, $\{6\}$ and $\{8\}$. So $k_1 = 2$ and $k_2 = k_3 = k_4 = 1$, $n_1 = n_2 = n_3 = 1$, $n_4 = 2$ and $j_1 = j_2 = j_3 = j_4 = 1$.

polygons cross each other (see, e.g., figure 3). The set of non-crossing partitions of S having exactly j connected components is denoted by $\Lambda(s, j)$ and their number by $c(s, j)$.

For simplicity later on, we relabel the s points on the circle in the following way: all points of the polygon to which 1 belongs get the label 1. Then we turn in the clockwise direction around the circle. The first point we meet which did not get labelled 1 we label by 2, as we do also for all points belonging to the polygon of 2. We continue this until we have relabelled all points. As there were j different polygons, we end up with a drawing carrying j different indices, one for each polygon (see, e.g., figure 4).

Consider a partition in $\Lambda(n, k)$. We denote by k_1 the interval of points on the circle to which the 'top corner' of the first polygon, i.e. the polygon containing the original point 1, belongs; we denote by k_2 the interval to which the second corner of the first polygon belongs, and so on. By top corner, we mean the set of points belonging to the first polygon that all lie next to each other, containing the original point 1 and not being separated by a point of another connected component of the partition. By second corner we mean the first set of points of the first polygon carrying a label 1 and lying next to each other, that we meet after turning in the clockwise direction around the circle after the top corner. In general, the jth corner is the set of points carrying a label 1 and lying next to each other that we meet when continuing in clockwise direction after the $(j-1)$th corner. n_i is the number of points in between the ith

and the $(i+1)$th corner of the first polygon; j_i is the number of different indices appearing in these n_i points. An explicit example is given figure 5. In this formulation, two drawings which are essentially the same but rotated through an angle of $2\pi s/N$ are considered as two different figures.

Theorem 5. *Using the notations just introduced and putting*

$$c(n, j) = 0 \qquad \text{if} \quad n < 1$$
$$c(n, j) = 0 \qquad \text{if} \quad j < 1 \quad \text{and} \quad j > n$$

the number $c(n, k)$ of elements in $\Lambda(n, k)$ is determined by the relations

$$c(n, k) = \sum{}' k_1 c(n_1, j_1) \dots c(n_r, j_r) \tag{13}$$

and

$$c(n, 1) = 1 \qquad \text{for} \quad n = 1, 2, \dots .$$

The summation in (13) extends over all r-tuples $(k_1, k_2, \dots, k_r)$ and $(n_1, n_2, \dots, n_r)$ such that $k_1 + k_2 + \cdots + k_r + n_1 + \cdots + n_r = n$, all $k_j > 0$ and $j_1 + \cdots + j_r + 1 = k$.

Proof. In order to prove the theorem, we have to count in how many different ways we can put k symbols on n points on a circle, remembering that there is a starting point (the original point 1) and a direction in which we follow the circle. If we call the polygon with relabelled indices 1 the first polygon of the partition, we know because of the non-crossing condition that if a label (different from 1) appears between two corners of the first polygon it cannot appear between two other corners. This means that we can consider all points between the ith and the $(i+1)$th corner of the first polygon as points of a circle where we have to put j_i points on n_i places. This can be done in $c(n_i, j_i)$ different ways, while the top corner can be chosen in k_1 different ways. Summing over all possibilities gives us the above expression. □

Theorem 6. *The limiting nth moment of the eigenvalue distribution of the random Gram matrix ensemble at density τ is given by*

$$m_n = \sum_{j=1}^{n} c(n, j)\tau^{j-1}$$

where $c(n, j)$ is the number of non-crossing partitions of the set $\{1, \dots, n\}$ on a circle where we use exactly j different symbols.

Proof. We start by computing the trace of the nth power of Γ

$$\operatorname{Tr}\Gamma^n = \sum_{i_1,\dots,i_n} \Gamma_{i_1 i_2}\Gamma_{i_2 i_3}\dots\Gamma_{i_n i_1} = \sum_{i_1,\dots,i_n} \langle\varphi_{i_1}, \varphi_{i_2}\rangle \dots \langle\varphi_{i_n}, \varphi_{i_1}\rangle.$$

Taking the average of the trace means that we integrate with respect to the probability measure μ_N:

$$\mathbb{E}_N(\operatorname{Tr}\Gamma^n) = \sum_{i_1,\dots,i_n=1}^{K} \int \mu_N(\mathrm{d}\varphi_{i_1}) \dots \mu_N(\mathrm{d}\varphi_{i_n}) \langle\varphi_{i_1}, \varphi_{i_2}\rangle \dots \langle\varphi_{i_n}, \varphi_{i_1}\rangle.$$

Each term in the sum can be seen as a partition of n points on a circle, where to each point we assign a number in $\{1, \dots, K\}$ (see, e.g., figure 6).

The claim of the theorem is that only those terms in the sum giving rise to a non-crossing partition contribute to the moments of Γ. We first prove that the contribution of the non-crossing partitions is exactly the expression of the theorem. Later we show that the crossing partitions do not contribute.

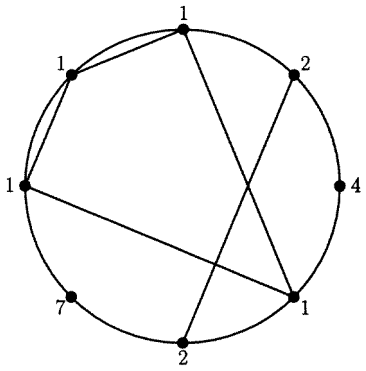

Figure 6. $n = 8$: $i_1 = 1, i_2 = 2, i_3 = 4, i_4 = 1, i_5 = 2, i_6 = 7, i_7 = 1, i_8 = 1$.

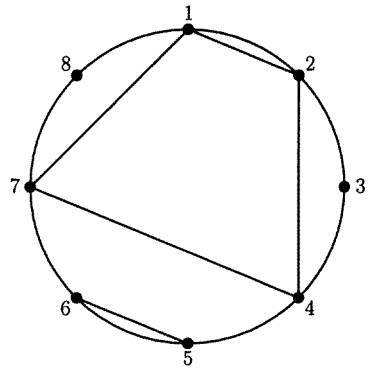

Figure 7. $n = 8$.

Suppose, first, that we have a term in which exactly j different symbols and no crossings appear. By drawing them on a circle we see only single points and non-crossing chords (see, e.g., figure 7).

In computing the integral for such a configuration, we start by integrating over an index appearing only once on the circle, say i_s:

$$\int \mu_N(\mathrm{d}\varphi_{i_1}) \dots \mu_N(\mathrm{d}\varphi_{i_s}) \dots \mu_N(\mathrm{d}\varphi_{i_n}) \langle \varphi_{i_1}, \varphi_{i_2} \rangle \dots \langle \varphi_{i_{s-1}}, \varphi_{i_s} \rangle \langle \varphi_{i_s}, \varphi_{i_{s+1}} \rangle \dots \langle \varphi_{i_n}, \varphi_{i_1} \rangle. \tag{14}$$

Because

$$\int \mu_N(\mathrm{d}\varphi) |\varphi\rangle\langle\varphi| = \frac{1}{N} \mathbb{1}$$

we have that (14) equals

$$\frac{1}{N} \int \mu_N(\mathrm{d}\varphi_{i_1}) \dots \widehat{\mu_N(\mathrm{d}\varphi_{i_s})} \dots \mu_N(\mathrm{d}\varphi_{i_n}) \langle \varphi_{i_1}, \varphi_{i_2} \rangle \dots \langle \varphi_{i_{s-1}}, \varphi_{i_{s+1}} \rangle \dots$$

where $\widehat{\mu_N}$ indicates that we have integrated over this variable. We can do this for all indices appearing only once, collecting for each of them a factor $1/N$.

After having eliminated all single points, only indices appearing at least twice remain. But, because of the non-crossing condition, equal indices now have to lie next to each other and these can be contracted to one single point because φ is normalized

$$\int \mu_N(\mathrm{d}\varphi_{i_1}) \dots \langle \varphi_{i_1}, \varphi_{i_2} \rangle \dots \langle \varphi_{i_{t-1}}, \varphi_{i_t} \rangle \langle \varphi_{i_t}, \varphi_{i_t} \rangle \langle \varphi_{i_t}, \varphi_{i_{t+1}} \rangle \dots$$
$$= \int \mu_N(\mathrm{d}\varphi_{i_1}) \dots \langle \varphi_{i_1}, \varphi_{i_2} \rangle \dots \langle \varphi_{i_{t-1}}, \varphi_{i_t} \rangle \langle \varphi_{i_t}, \varphi_{i_{t+1}} \rangle \dots .$$

So, we end up with a circle, again with single points on it, which we can integrate, each of them yielding a factor $1/N$. After this integration, we end up with indices appearing at least thrice but lying next to each other. So, they can again be contracted. This procedure can be continued until we have a circle with only single points. We can again integrate over the remaining indices, each of them giving a factor $1/N$, except for the last one where we have

$$\int \mu_N(\mathrm{d}\varphi_{i_r}) \langle \varphi_{i_r}, \varphi_{i_r} \rangle = 1.$$

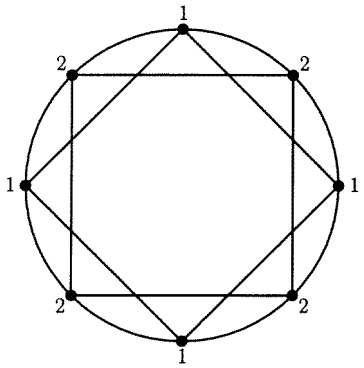

Figure 8. $n = 8$.

So, eventually, we see that a non-crossing partition where exactly j different symbols appear, gives a contribution $1/N^{j-1}$. Because there are exactly $c(n, j)K(K-1)\dots(K-j+1)$ of these partitions, their contribution is

$$c(n, j)\frac{K^j}{N^{j-1}}.$$

Taking all non-crossing partitions into account, we find

$$\lim_{N\to\infty}\frac{1}{K}\sum_{j=1}^{n}c(n, j)\frac{K^j}{N^{j-1}} = \sum_{j=1}^{n}c(n, j)\tau^{j-1}.$$

It remains to be shown that crossing partitions do not contribute to the moments. So, suppose that j different indices appear in the scalar product and that at least one crossing occurs. We prove the statement by induction. For $j = 2$ and one crossing appearing, it is true. Indeed, the most complicated case is the one of two $n/2$-gones (e.g., figure 8)

$$\int \mu_N(\mathrm{d}\varphi_{i_1})\mu_N(\mathrm{d}\varphi_{i_2})\mu_N(\mathrm{d}\varphi_{i_1})\dots\underbrace{\langle\varphi_{i_1}, \varphi_{i_2}\rangle\langle\varphi_{i_2}, \varphi_{i_1}\rangle\langle\varphi_{i_1}, \varphi_{i_2}\rangle\dots\langle\varphi_{i_2}, \varphi_{i_1}\rangle}_{n}$$

$$= \int \mu_N(\mathrm{d}\varphi_{i_1})\dots\langle\varphi_{i_1}\otimes\dots\otimes\varphi_{i_1}, \varphi_{i_2}\otimes\dots\varphi_{i_2}\rangle\langle\underbrace{\varphi_{i_2}\otimes\dots\varphi_{i_2}}_{n/2}, \underbrace{\varphi_{i_1}\otimes\dots\otimes\varphi_{i_1}}_{n/2}\rangle$$

$$= \frac{1}{N(N+1)\dots(N+n/2-1)}.$$

So

$$\lim_{N\to\infty}\frac{1}{K}\frac{K(K-1)}{N^{n/2}} = \lim_{N\to\infty}\frac{K}{N^{n/2}} = 0 \qquad (n \geqslant 4).$$

Suppose that the statement is true for $j-1$ indices appearing, then it also holds for j indices. Indeed, start by integrating over one of the indices appearing in a crossing and suppose this index appears t times. Because of the crossing condition, $t \geqslant 2$. This gives rise to a factor $1/N^t$. By the symmetrization procedure, we end up with

$$\frac{1}{N^t}\left(\sum \text{ cycles in which } j-1 \text{ indices appear on } n-t \text{ points}\right).$$

Herein, the maximal contribution $1/(N^t N^{j-2})$ comes from the non-crossing cycles

$$\frac{1}{K}\frac{K^j}{N^{t+j-2}} = \frac{K^{j-1}}{N^{t+j-2}} \leqslant \frac{K^{j-1}}{N^j} = \frac{\tau^{j-1}}{N}$$

which for $N \to \infty$ goes to zero. □

Next, we derive a closed expression for the generating function $F(t, \tau)$ of the limiting moments

$$F(t, \tau) := 1 + \sum_{n=1}^{\infty} m_n t^n = 1 + \sum_{n=1}^{\infty} t^n \sum_{k=1}^{n} c(n, k) \tau^{k-1}. \tag{15}$$

Theorem 7. *The generating function of the limiting moments of the eigenvalue distribution of the random Gram matrix ensemble at density τ is given by*

$$F(t, \tau) = \frac{1 - t + \tau t - \sqrt{(1 - t + \tau t)^2 - 4\tau t}}{2\tau t}$$

for $\tau > 0$ and $|t| < 1/(1 + \sqrt{\tau})^2$.

Proof. Let $|t| < 1$. The recurrence relation of theorem 5 gives

$$\begin{aligned} F(t, \tau) &= 1 + \sum_{n=1}^{\infty} t^n \sum_{k=1}^{n} c(n, k) \tau^{k-1} \\ &= 1 + \sum_{n=1}^{\infty} t^n + \sum_{n=1}^{\infty} t^n \sum_{k=2}^{n} c(n, k) \tau^{k-1} \\ &= 1 + \frac{t}{1 - t} + \sum_{n=1}^{\infty} t^n \sum_{k=2}^{n} \tau^{k-1} \sum{}' k_1 c(n_1, j_1) \dots c(n_r, j_r). \end{aligned}$$

The primed sum in the last line is over all r-tuples $(k_1, k_2, \dots, k_r)$ and $(n_1, n_2, \dots, n_r)$ such that $k_1 + k_2 + \dots + k_r + n_1 + \dots + n_r = n$, all $k_j > 0$ and $j_1 + \dots + j_r + 1 = k$. The last term on the right-hand side is equal to

$$\begin{aligned} &\sum_{r=1}^{\infty} \tau^r \sum_{k_1=1}^{\infty} k_1 t^{k_1} \left(\sum_{k_2=1}^{\infty} t^{k_2} \right)^{r-1} \sum_{n_1, \dots n_r = 1}^{\infty} \sum_{j_1, \dots, j_r = 1}^{n_1, \dots, n_r} t^{n_1} \tau^{j_1} c(n_1, j_1) \dots t^{n_r} \tau^{j_r} c(n_r, j_r) \\ &\quad = \sum_{r=1}^{\infty} \tau^r \frac{t}{(1 - t)^2} \left(\frac{t}{1 - t} \right)^{r-1} (F(t, \tau) - 1)^r \\ &\quad = \frac{1}{1 - t} \sum_{r=1}^{\infty} \left(\frac{\tau t (F(t, \tau) - 1)}{1 - t} \right)^r \\ &\quad = \frac{\tau t (F(t, \tau) - 1)}{(1 - t)(1 - t - \tau t (F(t, \tau) - 1))}. \end{aligned}$$

Of course, this last expression is only true if

$$\left| \frac{\tau t (F(t, \tau) - 1)}{1 - t} \right| < 1. \tag{16}$$

This will be checked at the end of this calculation. Substituting this back in the expression for $F(t, \tau)$, we get

$$F(t, \tau) = \frac{1}{1 - t - \tau t (F(t, \tau) - 1)}.$$

This results in a quadratic equation in $F(t, \tau)$ with solutions

$$F_{\pm}(t, \tau) = \frac{1 - t + \tau t \pm \sqrt{(1 - t + \tau t)^2 - 4\tau t}}{2\tau t}.$$

Because

$$\lim_{t\to 0} t F_+(t,\tau) = \frac{1}{\tau}$$

we see that $F_+(t,\tau)$ cannot be of the form of (15). From the explicit expression of $F(t,\tau)$, it can now be checked that the function

$$G(t,\tau) := \frac{\tau t(F(t,\tau)-1)}{1-t}$$

satisfies $|G(t,\tau)| < G(1/(1+\sqrt{\tau})^2,\tau) = (2+\sqrt{\tau})^{-1} < 1$ so that indeed (16) is satisfied. □

Now we apply the ideas of section 3. In order to solve the moment problem, we construct a function f by setting $f(z) := -1/z\ F(1/z,\tau)$ for $|z| > (1+\sqrt{\tau})^2$. For these z we can write

$$f(z) = \frac{-z-\tau+1+\sqrt{(z+\tau-1)^2-4\tau z}}{2\tau z}. \tag{17}$$

For all other $z \in \mathbb{C}$ we define $f(z)$ as the analytic continuation, where possible. As a result, we get a function which is analytic in $\mathbb{C}\setminus[(1-\sqrt{\tau})^2,(1+\sqrt{\tau})^2]$. This function has the properties mentioned in theorem 1, so we can conclude that f is the Stieltjes transform of the limiting eigenvalue measure. Via the Perron–Frobenius inversion theorem (5), one can now prove the following theorem.

Theorem 8. *The expectation of the empirical eigenvalue distribution of the ensemble of random Gram matrices at density τ converges weakly to the Marchenko–Pastur distribution [10]*

$$\rho_\tau = \begin{cases} \delta(t-1) & \text{if} \quad \tau = 0 \\ \sigma(t,\tau) & \text{if} \quad 0 < \tau \leqslant 1 \\ \dfrac{\tau-1}{\tau}\delta(t) + \sigma(t,\tau) & \text{if} \quad \tau > 1 \end{cases} \tag{18}$$

with

$$\sigma(t,\tau) = \begin{cases} \dfrac{\sqrt{4\tau t-(t+\tau-1)^2}}{2\pi\tau t} & (1-\sqrt{\tau})^2 \leqslant t \leqslant (1+\sqrt{\tau})^2 \\ 0 & \text{otherwise.} \end{cases}$$

A three-dimensional plot of this distribution is shown in figure 9. The solid curve represents the weight of the δ function appearing in this distribution for $\tau > 1$.

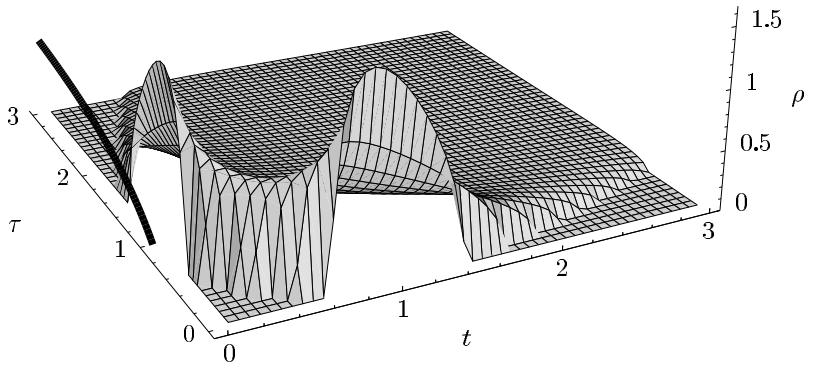

Figure 9. The Marchenko–Pastur distribution.

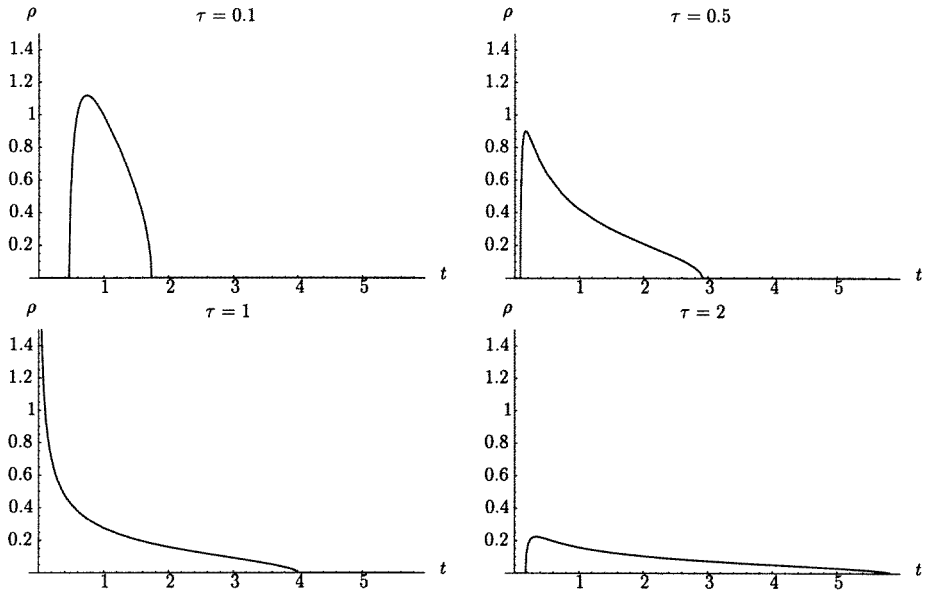

Figure 10. Cross sections of the Marchenko–Pastur distribution.

Several cross sections of this figure are shown in figure 10. For very small τ, we must choose relatively few vectors in a large space. This will often lead to almost orthogonal choices and therefore ρ_τ will be concentrated around one. When τ increases to one, there is a fair chance of many vectors overlapping and the support of ρ will simultaneously extend towards zero, which is a lower bound of its support, and to larger positive values. When $\tau > 1$ there will almost surely be a sizeable degree of linear dependence responsible for a high multiplicity of the eigenvalue zero. This results in the atom of ρ_τ with weight $(\tau - 1)/\tau$ (not shown in figure 10).

Either theorem 7 or 8 allows us to compute explicitly the moments of the limiting eigenvalue distribution. They are given by

$$m_n = \sum_{k=1}^{n} \frac{1}{n-k+1}\binom{n}{k}\binom{n-1}{k-1}\tau^{k-1}.$$

For $\tau = 1$, the moments of the limiting eigenvalue distribution are precisely the Catalan numbers

$$C_{2j} = \sum_{k=1}^{n} c(n,k) = \frac{1}{j+1}\binom{2j}{j}.$$

For general τ, the first few limiting moments are

$$m_1 = 1$$
$$m_2 = 1 + \tau$$
$$m_3 = 1 + 3\tau + \tau^2$$
$$m_4 = 1 + 6\tau + 6\tau^2 + \tau^3$$
$$m_5 = 1 + 10\tau + 20\tau^2 + 10\tau^3 + \tau^4$$
$$\ldots.$$

Theorem 8 can be considerably strengthened. In fact, the empirical distributions themselves converge with probability one to the Marchenko–Pastur distribution. In order to prove this, we must essentially control the fluctuations of the moments of the empirical eigenvalue distributions well enough when $K \to \infty$. The next section proves this convergence by using a very different approach, namely by obtaining an algebraic equation for the limiting distribution.

7. The resolvent of random Gram matrices†

In this section, we apply the method proposed in [8], where it is used to compute the limiting eigenvalue distribution of certain ensembles of random matrices. The following is an adaptation of that proof. We start from the description of the random Gram matrix ensemble in terms of vectors obtained by normalizing N-dimensional vectors with independent complex components, distributed according to a fixed normal distribution (11).

Let $x_1, \ldots, x_K$ be complex N-dimensional Gaussian vectors with i.i.d. components of mean zero and variance one

$$x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{Nj} \end{pmatrix} = (x_{\alpha j})_{\alpha=1,\ldots,N} \tag{19}$$

and put $\varphi_j := x_j/\|x_j\|$, $j = 1, \ldots, K$. Define $\boldsymbol{X} = (x_1, \ldots, x_K)$. So $\boldsymbol{X}$ is an $N \times K$ matrix containing the K Gaussian vectors as columns. Set $M = \boldsymbol{X}^*\boldsymbol{X}$, then we have that $M_{ij} = \langle x_i, x_j \rangle$. Finally, let $R := \operatorname{diag}(\|x_j\|)$, the diagonal matrix of dimension K with the norms of the x_j as its diagonal elements. It is clear that the Gram matrix corresponding to $\Phi = (\varphi_1, \ldots, \varphi_K)$ is

$$\Gamma = \Gamma^{\Phi} = R^{-1} M R^{-1}$$

and its resolvent is

$$G = G(z) = (\Gamma - z)^{-1} = R \Lambda R$$

with $\Lambda := (M - zR^2)^{-1}$.

We want to calculate the limit of $g_N(z) := \operatorname{Tr} G(z)/K$ when $N, K \to \infty$ while the ratio $\tau = K/N$ remains constant. This limit is the Stieltjes transform of the limiting eigenvalue distribution (6). The idea is to establish an equation for $\mathbb{E}(g_N(z))$ which has a non-trivial $N \to \infty$ limit.

We need the following facts:

- *The resolvent identity:* let A, B be square matrices and $z \in \mathbb{C}$ and suppose that $(A - z)^{-1}$ and $(B - z)^{-1}$ exist, then

$$(B - z)^{-1} = (A - z)^{-1} - (B - z)^{-1}(B - A)(A - z)^{-1}. \tag{20}$$

- Let ξ be a complex Gaussian random variable with $\mathbb{E}(\xi) = \mathbb{E}(\xi^2) = 0$ and $\mathbb{E}(|\xi|^2) = 1$ and let $(\xi, \overline{\xi}) \mapsto f(\xi, \overline{\xi})$ be a continuously differentiable function growing polynomially at infinity, a property which it shares with its derivatives, then

$$\mathbb{E}(\xi f(\xi, \overline{\xi})) = \mathbb{E}(|\xi|^2)\mathbb{E}\left(\frac{\partial f}{\partial \overline{\xi}}\right) = \mathbb{E}\left(\frac{\partial f}{\partial \overline{\xi}}\right) \tag{21}$$

  which can be checked by using partial integration.

† This section was written with the assistance of L Pastur, of Université Paris 7/ Denis Diderot, UMR 9994, Paris 75251, France.

- Let $x \mapsto A(x)$ be a differentiable matrix-valued function and suppose that A is invertible, then
$$\frac{\mathrm{d}}{\mathrm{d}x}A^{-1} = -A^{-1}\frac{\mathrm{d}A}{\mathrm{d}x}A^{-1}.$$
In particular,
$$\frac{\partial}{\partial A_{ij}}(A^{-1})_{pq} = -(A^{-1})_{pi}(A^{-1})_{jq}. \tag{22}$$
- The resolvent $G(z) := (\Gamma - z)^{-1}$ of an Hermitian matrix Γ satisfies
$$|G(z)_{ij}| \leqslant \|G(z)\| \leqslant |\mathrm{Im}\,(z)|^{-1} \qquad i, j = 1, \ldots, K. \tag{23}$$
We now establish a closed equation for $\lim_{N\to\infty}\mathbb{E}(g_N(z))$.

Theorem 9. *For $R \in \mathbb{R}^+$, define*
$$U_R := \left\{ z \in \mathbb{C} \,\middle|\, |\mathrm{Im}\,(z)| > 2 + 6\tau \text{ and } \frac{|z|}{|\mathrm{Im}\,(z)|} < R \right\}. \tag{24}$$
Then $\lim_{N\to\infty}\mathbb{E}(g_N(z))$ exists for $z \in U_R$ and is equal to the unique solution f of the equation
$$\tau f(z)^2 + \left(1 - \frac{1}{z} + \frac{\tau}{z}\right) f(z) + \frac{1}{z} = 0$$
which satisfies $\mathrm{Im}\,(z)\mathrm{Im}\,(f(z)) \geqslant 0$.

Proof. First we use the resolvent identity (20) to write
$$\begin{aligned} g_N(z) = \frac{1}{K}\mathrm{Tr}\,G &= -\frac{1}{z} + \frac{1}{zK}\mathrm{Tr}(\Gamma G) \\ &= -\frac{1}{z} + \frac{1}{zK}\mathrm{Tr}(M\Lambda) \\ &= -\frac{1}{z} + \frac{1}{zK}\sum_{i,j=1}^{K}\sum_{\alpha=1}^{N}\overline{x_{\alpha i}}x_{\alpha j}\Lambda_{ji}. \end{aligned}$$
We take the expectation value of this with respect to the Gaussian variables and use (21) to obtain
$$\begin{aligned} \mathbb{E}(g_N(z)) &= -\frac{1}{z} + \frac{1}{zK}\sum_{i,j=1}^{K}\sum_{\alpha=1}^{N}\mathbb{E}(\overline{x_{\alpha i}}x_{\alpha j}\Lambda_{ji}) \\ &= -\frac{1}{z} + \frac{1}{zK}\sum_{i,j=1}^{K}\sum_{\alpha=1}^{N}\mathbb{E}\left(\frac{\partial}{\partial\overline{x_{\alpha j}}}\overline{x_{\alpha i}}\Lambda_{ji}\right). \end{aligned}$$
Using (22) this becomes
$$\begin{aligned} \mathbb{E}(g_N(z)) &= -\frac{1}{z} + \frac{1}{zK}\mathbb{E}\left(N\,\mathrm{Tr}\,\Lambda - \mathrm{Tr}\,\Lambda\,\mathrm{Tr}\,\Lambda M + z\sum_{j=1}^{K}\Lambda_{jj}(\Lambda M)_{jj}\right) \\ &= -\frac{1}{z} + \frac{1}{zK}\mathbb{E}\left(N\,\mathrm{Tr}\,R^{-2}G - \mathrm{Tr}\,R^{-2}G\,\mathrm{Tr}\,G\Gamma + z\sum_{j=1}^{K}(R^{-2}G)_{jj}(G\Gamma)_{jj}\right). \end{aligned}$$

Introducing generalized spherical coordinates, every complex N-dimensional vector $(z_1, \ldots, z_N)$ corresponds to a real $2N$-dimensional vector $(r, \theta_1, \ldots, \theta_{2N-1})$ with $r = \sum_j |z_j|^2$. The above expectation value is a sum of expectation values of the form

$\mathbb{E}(R(r)\Theta(\theta_1,\ldots,\theta_{2N-1}))$ and because of the special form of the Gaussian distribution function this is equal to $\mathbb{E}(R(r))\mathbb{E}(\Theta(\theta_1,\ldots\theta_{2N-1}))$. So we can write

$$\mathbb{E}(g_N(z)) = -\frac{1}{z} + \frac{1}{zK}\mathbb{E}\left(\frac{1}{r^2}\right)\mathbb{E}\left(N\,\mathrm{Tr}\,G - \mathrm{Tr}\,G\,\mathrm{Tr}\,G\Gamma + z\sum_{j=1}^{K} G_{jj}(G\Gamma)_{jj}\right).$$

As

$$\mathbb{E}\left(\frac{1}{r^2}\right) = \frac{1}{N-1}$$

we end up with

$$\mathbb{E}(g_N(z)) = -\frac{1}{z} + \frac{1}{z}\frac{N}{N-1}\mathbb{E}\left(g_N(z) - \tau g_N(z) - z\tau g_N(z)^2 + \frac{z}{N}g_N(z) + \frac{z^2}{N}d_N(z)\right) \qquad (25)$$

with $\tau = K/N$ as before and

$$d_N(z) := \frac{1}{K}\sum_{j=1}^{K}(G_{jj})^2.$$

Because $|g_N(z)| \leqslant 1/|\mathrm{Im}\,(z)| < 2 + 6\tau$ for every N we see that the sequence $g_N(z)$, for a fixed $z \in U_R$ and a fixed τ, is confined within a compact set. Thus there exists a convergent subsequence. Moreover, every convergent subsequence has the same limit if we show that the limit is the unique solution f of the equation

$$\tau f(z)^2 + \left(1 - \frac{1}{z} + \frac{\tau}{z}\right) f(z) + \frac{1}{z} = 0$$

which satisfies $\mathrm{Im}\,(z)\mathrm{Im}\,(f(z)) > 0$. That this is indeed the case follows from (25) and the estimates

$$\left|\frac{1}{N-1}\mathbb{E}(g_N(z))\right| \leqslant \frac{1}{N-1}\frac{1}{|\mathrm{Im}\,(z)|}$$
$$\left|\frac{z}{N-1}\mathbb{E}(d_N(z))\right| \leqslant \frac{|z|}{N-1}\frac{1}{|\mathrm{Im}\,(z)|^2}$$

and

$$\mathbb{E}(g_N(z)^2) = \mathbb{E}(g_N(z))^2 + \mathrm{O}(N^{-2}).$$

The first two inequalities are based on (23) and the third one is proven in lemma 2. □

In view of the continuity of the one-to-one correspondence between probability measures and their Stieltjes transforms as expressed in theorem 2(ii), theorem 9 implies the weak convergence of the expectations $\mathbb{E}(\rho_K)$ to the Marchenko–Pastur distribution (18). That is, theorem 9 implies theorem 8.

We shall now prove that, not only do the expectations of the empirical eigenvalue distributions converge to the limit given in theorem 8, but that the distributions themselves converge to the same limit with probability one. This allows one, in particular, to use this distribution for statistical purposes as argued in sections 1, 2 and 4. To this end, we first introduce a probability space on which all the Gram matrices are defined simultaneously. This space is just the infinite product space generated by the double infinite sequence of the i.i.d. Gaussian random variables $\{x_{\alpha j}\}_{\alpha,j=1}^{\infty}$ which were introduced in (19) for α, $j = 1, 2, \ldots, K$. Using this space as the space of realizations of the Gram matrices we can now formulate the following theorem.

Theorem 10. *The empirical eigenvalue distribution of the ensemble of random Gram matrices at density τ converges weakly with probability one to the Marchenko–Pastur distribution ρ_τ.*

Proof. Fix $z \in U_R$ and $\varepsilon > 0$. By Tchebyshev's inequality and lemma 2, there exists a C such that

$$\begin{aligned}\text{Prob}\{|g_N(z) - \mathbb{E}(g_N(z))| > \varepsilon\} &\leqslant \frac{1}{\varepsilon^2}(\mathbb{E}(|g_N(z)|^2) - |\mathbb{E}(g_N(z))|^2) \\ &\leqslant \frac{C}{\varepsilon^2 N^2}.\end{aligned}$$

The Borel–Cantelli lemma and theorem 9 then imply that the sequence $(g_N(z))_N$ of random complex numbers converges with probability one to $f(z)$ as defined in (17). We now use the following result to conclude that the sequence of random analytic functions g_N converges uniformly with probability one to the limit f on any compact subset of U_R: suppose that we are given a sequence of analytic functions, defined on a common compact domain D. If the sequence converges on a countable set of points having an accumulation point in D to an analytic function defined on the same domain, then it converges uniformly to the same limit on D. In view of the continuity of the one-to-one correspondence between probability measures and their Stieltjes transforms (theorem 2(ii)), we can conclude that the empirical eigenvalue distributions of the Gram matrices converge with probability one to the limiting measure ρ_τ of theorem 8. □

By mimicking the proof of the Glivenko–Cantelli theorem on the uniform convergence with probability one of the empirical distribution functions of i.i.d. random variables to their probability law [9], one can show that the sequence of random distributions functions $\rho_K(\lambda) := \rho_K(]-\infty, \lambda])$ converges uniformly with probability one to the distribution function $\rho_\tau(\lambda) := \rho_\tau(]-\infty, \lambda])$. It now remains to prove the following lemma.

Lemma 2. *For $z \in U_R$ as defined in (24)*

$$\mathbb{E}(|g_N(z)|^2) = |\mathbb{E}(g_N(z))|^2 + \mathrm{O}(N^{-2}).$$

Proof. Define

$$\gamma_N(z) := g_N(z) - \mathbb{E}(g_N(z)).$$

It follows that $\mathbb{E}(|\gamma_N(z)|^2) = \mathbb{E}(g_N(z)\gamma_N(\overline{z}))$. Note that $\overline{g_N(z)} = g_N(\overline{z})$. Now use the resolvent identity (20) again to write

$$\mathbb{E}(g_N(z)g_N(\overline{z})) = -\frac{1}{z}\mathbb{E}(g_N(\overline{z})) + \frac{1}{z}\mathbb{E}\left(\frac{1}{K}\operatorname{Tr}\Gamma G(z)g_N(\overline{z})\right).$$

Once more, we apply (21) to the last expectation on the right-hand side. Using the same notations and a similar calculation as in the proof of theorem 9 results in

$$\begin{aligned}\mathbb{E}(g_N(z)g_N(\overline{z})) = -\frac{1}{z}\mathbb{E}(g_N(\overline{z})) + \frac{1}{z(N-1)}\mathbb{E}(&Ng_N(z)g_N(\overline{z}) - Kg_N(z)g_N(\overline{z}) \\ &-zg_N(z)^2 g_N(\overline{z}) + zg_N(z)g_N(\overline{z}) + z^2 d_N(z)g_N(\overline{z}) + h(z,\overline{z}))\end{aligned} \tag{26}$$

with

$$\begin{aligned}h(z,\overline{z}) = \frac{1}{K}g_N(\overline{z}) + \frac{z}{K^2}\sum_{j=1}^{K} G(z)_{jj}G(\overline{z})_{jj} - \frac{1}{K^2}\operatorname{Tr}G(\overline{z})^2 \\ -\frac{z}{K^2}\operatorname{Tr}G(\overline{z})G(z)G(\overline{z}) + \frac{\overline{z}}{K^2}\operatorname{Tr}G(\overline{z})^2 + \frac{z\overline{z}}{K^2}\sum_{j=1}^{K} G(z)_{jj}(G(\overline{z})^2)_{jj}.\end{aligned}$$

In the first term of the right-hand side of (26) we insert (25) via the substitution of $1/z$. This results in

$$\begin{aligned}\mathbb{E}(|\gamma_N(z)|^2) = {} & \frac{1}{z}\frac{N}{N-1}\mathbb{E}(|\gamma_N(z)|^2) - \frac{1}{z}\frac{K}{N-1}\mathbb{E}(|\gamma_N(z)|^2) \\ & -\frac{K}{N-1}\mathbb{E}(g_N(z)^2\gamma_N(z)) + \frac{1}{N-1}\mathbb{E}(|\gamma_N(z)|^2) \\ & +\frac{z}{N-1}\mathbb{E}(d_N(z)\gamma_N(\overline{z})) + \frac{1}{z(N-1)}\mathbb{E}(h(z,\overline{z})). \end{aligned} \tag{27}$$

Now, we estimate the different terms in this expression. It is easy to see that

$$\mathbb{E}(g_N(z)^2\gamma_N(\overline{z})) = \mathbb{E}(g_N(z)|\gamma_N(z)|^2) + \mathbb{E}(g_N(z))\mathbb{E}(|\gamma_N(z)|^2)$$

and thus

$$|\mathbb{E}(g_N(z)^2\gamma_N(\overline{z}))| \leqslant \frac{2}{|\mathrm{Im}\,(z)|}\mathbb{E}(|\gamma_N(z)|^2).$$

Using the Cauchy–Schwarz inequality, we get

$$\begin{aligned}|\mathbb{E}(d_N(z)\gamma_N(z))| &\leqslant \mathbb{E}(|d_N(z)|^2)^{1/2}\mathbb{E}(|\gamma_N(z)|^2)^{1/2} \\ &\leqslant \frac{1}{|\mathrm{Im}\,(z)|^2}\mathbb{E}(|\gamma_N(z)|^2)^{1/2}.\end{aligned}$$

Finally, remembering that $|\mathrm{Im}\,(z)| \leqslant |z|$, we get

$$\left|\mathbb{E}\left(\frac{1}{z(N-1)}h(z,\overline{z})\right)\right| \leqslant \frac{1}{(N-1)K}\left(\left(3+\frac{|z|}{|\mathrm{Im}\,(z)|}\right)\frac{1}{|\mathrm{Im}\,(z)|^2} + \frac{2}{|\mathrm{Im}\,(z)|^3}\right).$$

Combining these inequalities with (27), we end up with

$$a\mathbb{E}(|\gamma_N(z)|^2) + b\mathbb{E}(|\gamma_N(z)|^2)^{1/2} + c \leqslant 0 \tag{28}$$

with

$$\begin{aligned}a &= \frac{(N-2)|\mathrm{Im}\,(z)| - N - 3K}{|\mathrm{Im}\,(z)|(N-1)} \\ b &= -\frac{|z|}{|\mathrm{Im}\,(z)|^2(N-1)} \\ c &= -\frac{1}{\tau N(N-1)}\left(\left(3+\frac{|z|}{|\mathrm{Im}\,(z)|}\right)\frac{1}{|\mathrm{Im}\,(z)|^2} + \frac{2}{|\mathrm{Im}\,(z)|^2}\right).\end{aligned}$$

If we define the set U_R as in (24), then $z \in U_R$ implies $a > 0$ and $b^2 - 4ac > 0$. So (28) implies

$$\mathbb{E}(|\gamma_N(z)|^2)^{1/2} \leqslant \frac{C(\tau, R)}{N}$$

with $C(\tau, R)$ a finite constant independent of z. We conclude that for $z \in U_R$

$$\mathbb{E}(|\gamma_N(z)|^2) \leqslant \frac{C(\tau, R)^2}{N^2}.$$

□

Note added in proof. Shortly after this work was finished, we became aware of an article of Oravecz and Petz [11] in which the moments of the Marchenko–Pastur distribution were derived. They prove the convergence in expectation of the moments of the empirical eigenvalue distribution of the Wishart matrices to those of the Marchenko–Pastur distribution. This ensemble of matrices is slightly different from ours. They use another combinatorial argument due to their description of the ensemble of random matrices.

Acknowledgments

It is a pleasure to thank L Pastur for his interest and very helpful comments, especially in connection with section 7. M De Cock and P Spincemaille acknowledge financial support from FWO-project G.0239.96.

References

[1] Akhiezer N I 1965 *The Classical Moment Problem* (Oliver and Boyd)
[2] Alicki R and Fannes M 1994 Defining quantum dynamical entropy *Lett. Math. Phys.* **32** 75–82
[3] Casati G and Chirikov B V 1995 *Quantum Chaos* (Cambridge: Cambridge University Press)
[4] Casati G and Chirikov B V 1995 Quantum chaos: unexpected complexity *Physica* D **86** 220
[5] Connes A, Narnhofer H and Thirring W 1987 Dynamical entropy of C*-algebras and von Neumann algebras *Commun. Math. Phys.* **112** 691–719
[6] Grenander U and Szegö G 1958 *Toeplitz Forms and their Applications* (Berkeley, CA: University of California Press)
[7] Horn R A and Johnson Ch R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
[8] Khorunzhy A, Khoruzhenko B and Pastur L 1996 Asymptotic properties of large random matrices with independent entries *J. Math. Phys.* **37** 5033
[9] Loève M 1979 *Probability Theory* vol I (Berlin: Springer)
[10] Marchenko V and Pastur L 1967 The eigenvalue distribution in some ensembles of random matrices *Math. USSR Sbornik* **1** 457–83
[11] Oravecz F and Petz D 1997 On the eigenvalue distribution of some symmetric random matrices *Acta Sci. Math. (Szeged)* **63** 383–95
[12] Reed M and Simon B 1975 *Methods Of Modern Mathematical Physics 2: Fourier Analysis, Self-Adjointness* (New York: Academic)
[13] Shohat J A and Tamarkin J D 1963 *The Problem of Moments* (Providence, RI: American Mathematical Society)
[14] Voiculescu D 1992 Dynamical approximation entropies and topological entropy in operator algebras *Commun. Math. Phys.* **144** 443–90