ESAIM: PROCEEDINGS AND SURVEYS, November 2023, Vol. 74, p. 137-157
Editors: Marguerite Zani, Clément Marteau - Journées MAS 2020, Fédération Siméon Poisson (Orléans-Tours),

RANDOM MATRICES AND RANDOM GRAPHS *

Mireille Capitaine[1], Simon Coste[2], Franck Gabriel[3], Pascal Maillard[4] and Cecile Mailler[5]

Abstract. We collect recent results on random matrices and random graphs. The topics covered are: fluctuations of the empirical measure of random matrices, finite-size effects of algorithms involving random matrices, characteristic polynomial of sparse matrices and Voronoi tesselations of split trees.

Résumé. Nous rassemblons des résultats récents sur les matrices et graphes aléatoires. Les sujets abordés sont : fluctuations de la mesure empirique de matrices aléatoires, effets de taille finie d'algorithmes impliquant des matrices aléatoires, polynôme caractéristique de matrices diluées et diagrammes de Voronoï d'arbres de fragmentation (*split trees*).

Introduction

We present here some recent results on random matrices and random graphs. These results were presented in the session "Random matrices and random graphs" of the *Journées MAS 2021*, which consisted of four talks by Mireille Capitaine, Simon Coste, Franck Gabriel and Cécile Mailler. The talks covered a wide range of topics, reflecting the diversity in these fields. The present article contains extended abstracts of these talks. We now give an overview of their content.

In Section 1, Mireille Capitaine reports on a central limit theorem [7] for the Stieltjes transform of the empirical spectral measure of random matrices which are of the form $X_N := P(W_N, D_N)$, where W_N is a complex Wigner matrix, D_N is a real diagonal matrix and P is a self-adjoint polynomial in two non-commuting variables. The section starts by recalling basic concepts in random matrices and free probability and includes many references, making it accessible to non-experts. It then presents previous works on central limit theorems for random matrices of the form $W_N + D_N$. In order to pass to random matrices of the form $P(W_N, D_N)$, the authors of [7] make use of a so-called *linearization procedure*, which converts a general noncommutative polynomial with complex coefficients into a linear polynomial with matrix coefficients, and which is briefly presented. In order to work with this, one requires an extension of the notion of freeness, called *freeness*

* *FG acknowledges partial support from grant ERC SG CONSTAMIS. CM is grateful to EPSRC for support through the fellowship EP/R022186/1; she and her co-authors also acknowledge support from DFG through the scientific network Stochastic Processes on Evolving Networks.*

[1] CNRS, Institut de Mathématiques de Toulouse (CNRS UMR 5219) and Université de Toulouse 3 Paul Sabatier

[2] Centre Sciences des Données, Ecole Normale Supérieure

[3] Ecole Polytechnique Fédérale de Lausanne

[4] Institut de Mathématiques de Toulouse (CNRS UMR 5219) and Université de Toulouse 3 Paul Sabatier and Institut Universitaire de France (IUF),

[5] Department of Mathematical Sciences, The University of Bath

with amalgamation, which is briefly outlined. The statement of the central limit theorem obtained using these methods concludes the section.

In Section 2, Franck Gabriel reports on two works concerning algorithms using random matrix approximations in order to calculate certain infinite-dimensional objects. The first one [38] concerns the so-called Kernel Regression Method — a least-square regression method involving a kernel K and a regularization parameter λ. A Monte-Carlo algorithm for this method has been proposed in [50], based on a finite-dimensional random approximation of the kernel K. The authors of [38] show that this algorithm creates in fact a bias, leading to an *effective* regularization parameter $\lambda_{\text{eff}} \neq \lambda$. They also propose an algorithm to overcome this bias. In the second part, based on [3], Gabriel presents the notion of *freeness with amalgamation over the diagonal* and explains how it can be used to express the limiting spectrum of an additive perturbation of a large permutation-invariant random matrix. An effective Monte-Carlo algorithm can be deduced from this relation. Numerical simulations illustrate the quality of this algorithm compared to naïve density estimates based on the empirical spectrum.

In Section 3, Simon Coste reports on a convergence result [21] for the characteristic polynomial of the adjacency matrix of a large sparse Erdős-Rényi digraph, i.e. the random $n \times n$ matrix A_n with iid Bernoulli(d/n) entries, with d fixed or going to infinity slower than $n^{o(1)}$. He proves that the *reverse characteristic polynomial* $P_n = \det(I_n - zA_n)$ almost surely converges to a certain random holomorphic function in a disk around the origin. This limit is called the *Poisson holomorphic chaos* and admits an infinite product representation involving Poisson-distributed random variables. As a byproduct, results on the extremal eigenvalues are derived. In the limit $d \to \infty$, one recovers the so-called Gaussian holomorphic chaos, previously defined and studied in the dense case.

In Section 4, Cécile Mailler reports on results on Voronoi cells in random split trees [29]. Given a graph G on n vertices and k vertices $v_1, \dots, v_k$ in G, the Voronoi tesselation of G centered on $v_1, \dots, v_k$ is a partition $(V_1, \dots, V_k)$ of the vertices such that for every $v \in V_i$, the distance of v to v_i is not larger than to any other vertex v_j, $j \neq i$, breaking ties according to an arbitrary rule. An influential conjecture of Chapuy [19] states that for a large family of random planar maps, i.e. planar graphs embedded in a compact surface, the vector $(|V_1|/n, \dots, |V_k|/n)$ of the rescaled sizes of the Voronoi cells converges in law to a Dirichlet$(1, \dots, 1)$ random vector. Weak forms of this conjecture have subsequently been proven, but the full conjecture remains open to this date. As so often, trees provide an interesting playground for testing the scope of this conjecture, and indeed, it has been proven [1] that the conjecture is true for random trees converging to the so-called Brownian continuous random tree. The authors of [29] explore whether the result still holds in the case of *split trees*. These have a quite different shape, for example, their height is typically logarithmic in n. The results of [29] are described precisely in this article, including a brief sketch of the proof techniques.

1. Fluctuations of the Stieltjes transform of the empirical spectral measure of a selfadjoint polynomial in a Wigner matrix and a deterministic diagonal matrix

Mireille Capitaine

This is based on a recent joint work with Serban Belinschi, Sandrine Dallaporta and Maxime Février [7]. The question of fluctuations of linear statistics $\sum_{i=1}^{N} \varphi(\lambda_i)$ of eigenvalues $\lambda_1, ..., \lambda_N$ of $N \times N$ Hermitian random matrices, associated to test functions $\varphi : \mathbb{R} \to \mathbb{C}$, has attracted a lot of attention in the past decades. Authors established CLT under various assumptions on the matrix and on the test function φ. A line of attack to study linear spectral statistics, based on Stieltjes transforms, has been developed in [6] (for Wigner matrices) and [5] (for sample covariance matrices); this approach has its roots in the works [35] and [44]. So this approach corresponds to test functions $\varphi_z, z \in \mathbb{C} \setminus \mathbb{R}$, $\varphi_z : x \mapsto \frac{1}{z-x}$. CLT could possibly be then extended to a wider class of test functions by Shcherbina's extension density argument [54], Cauchy's theorem [6]...

In this note, we investigate this line of attack of Stieltjes transform for general polynomial matrix models in a Wigner matrix and a real deterministic diagonal matrix. We first present a gradual build-up of the works in the

lineage of our results and then explain the methodology based on a linearization procedure and operator-valued free subordination properties that allowed us to establish our CLT.

1.1. Presentation of the Model

The complex algebra $\mathbb{C}\langle t_1, \ldots, t_n\rangle$ of polynomials with complex coefficients in n noncommuting indeterminates $t_1, \ldots, t_n$ becomes a $*$-algebra by anti-linear extension of $(t_{i_1} t_{i_2} \cdots t_{i_l})^* = t_{i_l} \cdots t_{i_2} t_{i_1}$, $i_1, \ldots, i_l = 1, \ldots, n$, $l \in \mathbb{N}$. We consider, on a probability space, a sequence of random matrices

$$X_N := P(W_N, D_N), \quad N \in \mathbb{N},$$

where:

(1) $P \in \mathbb{C}\langle t_1, t_2\rangle$ is a selfadjoint polynomial in two noncommuting indeterminates that really involves both indeterminates;
(2) entries $\{W_{ij}\}_{1 \leq i \leq j \leq N}$ of the $N \times N$ Hermitian matrix W_N are independent random variables;
(3) off-diagonal entries $\{W_{ij}\}_{1 \leq i < j \leq N}$ of W_N are identically distributed complex random variables such that, for some $\varepsilon > 0$, $\mathbb{E}[|\sqrt{N} W_{ij}|^{6(1+\varepsilon)}] \leq C_6$. We assume that $\mathbb{E}[W_{ij}] = 0$ and that

$$\sigma_N^2 := \mathbb{E}[|W_{ij}|^2] \geq 0, \quad \theta_N := \mathbb{E}[W_{ij}^2] \in \mathbb{C}, \quad \kappa_N := \mathbb{E}[|W_{ij}|^4] - 2\sigma_N^4 - |\theta_N|^2 \in \mathbb{R},$$

satisfy

$$\lim_{N \to +\infty} N\sigma_N^2 = \sigma^2 > 0, \quad \lim_{N \to +\infty} N\theta_N = \theta \in \mathbb{R}, \quad \lim_{N \to +\infty} N^2 \kappa_N = \kappa \in \mathbb{R}.$$

The assumption $\theta \in \mathbb{R}$ means that correlations between the real and imaginary parts of off-diagonal entries of W_N are negligible.
(4) diagonal entries $\{W_{ii}\}_{1 \leq i \leq N}$ of W_N are identically distributed real random variables such that, for some $\varepsilon > 0$, $\mathbb{E}[|\sqrt{N} W_{ii}|^{4(1+\varepsilon)}] \leq C_4$. We assume that $\mathbb{E}[W_{ii}] = 0$ and that $\tilde{\sigma}_N^2 := \mathbb{E}[W_{ii}^2] \geq 0$ satisfies $\lim_{N \to +\infty} N\tilde{\sigma}_N^2 = \tilde{\sigma}^2 > 0$;
(5) D_N is a $N \times N$ deterministic real diagonal matrix. We assume that $\sup_{N \in \mathbb{N}} \|D_N\| < \infty$ and that, for some Borel probability measure ν on $\mathbb{R}$, $\nu_N := \frac{1}{N} \sum_{\lambda \in \mathrm{sp}(D_N)} \delta_\lambda$ weakly converges towards ν. Here, we use the notation $\mathrm{sp}(A)$ for the (multi)set of eigenvalues (counted with their algebraic multiplicity) of a square matrix A.

We will also assume that all entries of W_N are almost surely bounded by δ_N, where $(\delta_N)_{N \in \mathbb{N}}$ is a sequence of positive numbers slowly converging to 0 (at rate less than $N^{-\epsilon}$ for any $\epsilon > 0$); this may be assumed without loss of generality to establish our CLT by truncation-centering-homogeneization arguments.

Throughout this note, for $p \in \mathbb{N}$, $M_p(\mathbb{C})$ is the set of $p \times p$ matrices with complex entries, I_p is the identity matrix, Tr_p denotes the trace on $M_p(\mathbb{C})$ and $id_p : M_p(\mathbb{C}) \to M_p(\mathbb{C})$ is the identity map. Moreover, for a probability measure μ on $\mathbb{R}$, $\mathrm{supp}(\mu)$ denotes the support of μ and $g_\mu : z \in \mathbb{C} \setminus \mathrm{supp}(\mu) \mapsto \int \frac{1}{z-x} d\mu(x)$ is the Stieltjes transform of μ.

1.2. Previous results

1.2.1. *CLT for the trace of the resolvent of a Wigner matrix*

The following theorem comes as a result of the work of several authors [35], [44], [6], [4], although these authors made additional assumptions on the entries of the Wigner matrix.

Theorem 1.1. *Let W_N be as defined in Section 1.1. For any $z \in \mathbb{C} \setminus \mathbb{R}$, when N goes to infinity, the random variable $\mathrm{Tr}_N(zI_N - W_N)^{-1} - \mathbb{E}\left(\mathrm{Tr}_N(zI_N - W_N)^{-1}\right)$ converges in distribution towards a centered Gaussian variable $\mathcal{N}$, such that*

$$\mathbb{E}\left(|\mathcal{N}|^2\right) = C(z, \bar{z}),\ \mathbb{E}\left(\mathcal{N}^2\right) = C(z, z),\ C(z_1, z_2) := \frac{\partial^2}{\partial z_1 \partial z_2} \gamma(z_1, z_2), \quad z_1, z_2 \in \mathbb{C} \setminus \mathbb{R},$$

$$\gamma(z_1,z_2) = -\log\left[1-\sigma^2 T_{\{z_1,z_2\}}\right] - \log\left[1-\theta T_{\{z_1,z_2\}}\right] + \frac{\kappa}{2}T^2_{\{z_1,z_2\}} + (\tilde{\sigma}^2-\sigma^2-\theta)T_{\{z_1,z_2\}},$$

with $T_{\{z_1,z_2\}} := g_{\mu_\sigma}(z_1)g_{\mu_\sigma}(z_2)$, *where* g_{μ_σ} *is the Stieltjes transform of the so-called semi-circular distribution:*

$$\frac{d\mu_\sigma}{dx}(x) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2-x^2}\,1_{[-2\sigma,2\sigma]}(x).$$

The strategy of the proof of [4, 6] consists in writing the centered trace of the resolvent of a Wigner matrix as the sum of martingale differences with respect to the filtration generated by the upper left corners of the Wigner matrix and then in using a classical CLT for martingale differences (e.g. Theorem 35.12 in [14]).

1.2.2. *CLT for the trace of the resolvent of deformed Wigner matrices*

It turns out that the so-called free probability theory allows for the extension of Theorem 1.1 to deformed Wigner matrices. We refer to [46] and [62] for an introduction to free probability theory and to [11, 43, 56] for the definitions and main properties of the additive free convolution of two probability measures μ and ν on $\mathbb{R}$, denoted by $\mu \boxplus \nu$. The free convolution of probability measures has an important property, called subordination, which can be stated as follows [13, 58]: there exists an analytic map $\omega_{\mu,\nu} : \mathbb{C}^+ \to \mathbb{C}^+$ such that

$$\forall z \in \mathbb{C}^+, g_{\mu\boxplus\nu}(z) = g_\nu(\omega_{\mu,\nu}(z)).$$

Free probability theory and random matrix theory are closely related. Indeed the purely algebraic concept of free relation of noncommutative random variables can be also modeled by random matrix ensembles if the matrix size goes to infinity. In the lineage of Voiculescu's [57] pioneering work, Dykema [31] and Ryan [53] established the asymptotic freeness of $(W_N)_N$ and $(D_N)_N$ defined in Section 1.1, that is, for any polynomial Q in two noncommutative variables,

$$\mathbb{E}\left(\frac{1}{N}\mathrm{Tr}_N(Q(W_N,D_N))\right) \longrightarrow_{N\to+\infty} \phi(Q(x,d)),$$

where x, d are free selfadjoint noncommutative variables in a $\mathcal{C}^*$-algebra $(\mathcal{A},\phi)$, the distribution of x (resp. d) is the semicircular distribution μ_σ (resp. ν). In particular the limiting mean empirical spectral distribution of $W_N + D_N$ is $\mu_\sigma \boxplus \nu$. Moreover, we have this free subordination property: $\forall z \in \mathbb{C}^+, g_{\mu_\sigma\boxplus\nu}(z) = g_\nu(\omega(z))$, the subordination function ω being given explicitly by $\omega(z) = z - \sigma^2 g_{\mu_\sigma\boxplus\nu}(z)$ [12]. The following result follows from [26], [42].

Theorem 1.2. *Let* W_N *and* D_N *be as defined in Section 1.1. For any* $z \in \mathbb{C}\setminus\mathbb{R}$, *when* N *goes to infinity, the random variable* $\mathrm{Tr}_N(zI_N - W_N - D_N)^{-1} - \mathbb{E}\left(\mathrm{Tr}_N(zI_N - W_N - D_N)^{-1}\right)$ *converges in distribution towards a centered Gaussian variable* $\mathcal{N}$ *such that*

$$\mathbb{E}\left(|\mathcal{N}|^2\right) = C(z,\bar{z}),\ \mathbb{E}\left(\mathcal{N}^2\right) = C(z,z) \text{ with } C(z_1,z_2) := \frac{\partial^2}{\partial z_1 \partial z_2}\gamma(z_1,z_2), \quad z_1,z_2 \in \mathbb{C}\setminus\mathbb{R},$$

$$\gamma(z_1,z_2) = -\log\left[1-\sigma^2 T_{\{z_1,z_2\}}\right] - \log\left[1-\theta T_{\{z_1,z_2\}}\right] + \frac{\kappa}{2}T^2_{\{z_1,z_2\}} + (\tilde{\sigma}^2-\sigma^2-\theta)T_{\{z_1,z_2\}},$$

$T_{\{z_1,z_2\}} = \int \frac{\nu(dx)}{(\omega(z_1)-x)(\omega(z_2)-x)}$ *and* ω *is the subordination function* $\omega(z) = z - \sigma^2 g_{\mu_\sigma\boxplus\nu}(z)$.

1.3. CLT for traces of resolvents of polynomials in W_N and D_N

Denote by $\mathcal{H}(\mathbb{C}\setminus\mathbb{R})$ the space of complex analytic functions on $\mathbb{C}\setminus\mathbb{R}$, endowed with the topology of uniform convergence on compact sets. We equip it with the topological Borel σ-field. The aim of our work is to extend Theorem 1.2 by studying the convergence of the following centered $\mathcal{H}(\mathbb{C}\setminus\mathbb{R})$-valued random variable

$$\xi_N(z) = \mathrm{Tr}_N\left((z-X_N)^{-1}\right) - \mathbb{E}\left(\mathrm{Tr}_N\left((z-X_N)^{-1}\right)\right)$$

where $X_N = P(W_N, D_N)$ is the polynomial matrix model defined in Section 1.1. This investigation is achieved by a methodology based on a linearization procedure and operator-valued free subordination properties.

A powerful tool to deal with non-commutative polynomials in random matrices or in operators is the so-called linearization trick that goes back to Haagerup and Thørbjornsen [37] in the context of operator algebras and random matrices. Roughly speaking, the idea is to relate the spectral properties of a general polynomial in operators to a linear polynomial but with matrix coefficients. Here, we use the procedure introduced by Anderson in [2].

Definition 1.3. *Let P be in the complex algebra $\mathbb{C}\langle t_1, \ldots, t_k\rangle$ of polynomials with complex coefficients in k noncommuting indeterminates $t_1, \ldots, t_k$.*

$$L_P := \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix} \in M_m(\mathbb{C}) \otimes \mathbb{C}\langle t_1, \ldots, t_k\rangle,$$

where $m \in \mathbb{N}$, $Q \in M_{m-1}(\mathbb{C}) \otimes \mathbb{C}\langle t_1, \ldots, t_k\rangle$ is invertible, u is a row vector and v is a column vector, both of size $m-1$, is called a **linearization of P**, *if there are matrices $\gamma_0, \gamma_1, \ldots, \gamma_k \in M_m(\mathbb{C})$, such that $L_P = \gamma_0 \otimes 1 + \gamma_1 \otimes t_1 + \cdots + \gamma_k \otimes t_k$ and $P = -uQ^{-1}v$.*

Proposition 1.4 (Anderson [2])**.** *Any polynomial $P \in \mathbb{C}\langle t_1, \ldots, t_k\rangle$ admits a linearization L_P. If P is selfadjoint then P admits a selfadjoint linearization L_P. There is no uniqueness but there exists an explicit algorithm for finding one that we will call canonical.*

Note that for any selfadjoint polynomial $P \in \mathbb{C}\langle t_1, t_2\rangle$ and any selfadjoint linearization $L_P = \gamma_0 \otimes 1 + \gamma_1 \otimes t_1 + \gamma_2 \otimes t_2 \in M_m(\mathbb{C}\langle t_1, t_2\rangle)$,

$$\mathrm{Tr}_N\left((z - P(W_N, D_N))^{-1}\right) = \mathrm{Tr}_m \otimes \mathrm{Tr}_N\left[\left((ze_{11} - \gamma_0) \otimes I_N - \gamma_1 \otimes W_N - \gamma_2 \otimes D_N\right)^{-1}(e_{11} \otimes I_N)\right] \quad (1)$$

where e_{11} denotes the m by m matrix whose only nonzero entry equals 1 and occurs in the first row and first column. Thus, this linearization trick converts our initial general noncommutative polynomial with complex coefficients into a linear polynomial with matrix coefficients and thus allows adapting the strategy based on CLT for martingale differences and Schur's inversion formula previously used for additively deformed Wigner matrices in [26].

This investigation requires a so-called matrix-valued free probability theory. There exists an extension, operator-valued free probability theory, in which, roughly speaking, scalars are replaced by elements of a subalgebra $\mathcal{B}$, the linear form is replaced by a $\mathcal{B}$-valued conditional expectation and freeness is replaced by the so-called freeness with amalgamation over $\mathcal{B}$. We refer to [59], [61], [10]. Let x, d be free selfadjoint noncommutative variables in a $\mathcal{C}^*$-algebra $(\mathcal{A}, \phi)$ such that the distribution of x (resp. d) is μ_σ (resp. ν). According to [48], $\gamma_1 \otimes x$ and $\gamma_2 \otimes d$ are free with amalgamation over $M_m(\mathbb{C})$. Moreover, it turns out that the $M_m(\mathbb{C})$-valued subordination function defined for $\beta \in M_m(\mathbb{C})$, $\Im m(\beta) > 0$, by

$$id_m \otimes \phi\left((\beta \otimes 1_{\mathcal{A}} - \gamma_1 \otimes x - \gamma_2 \otimes d)^{-1}\right) = id_m \otimes \phi\left((\omega_m(\beta) \otimes 1_{\mathcal{A}} - \gamma_2 \otimes d)^{-1}\right),$$

is explicitly given by

$$\omega_m(\beta) = \beta - \gamma_1 id_m \otimes \phi\left[(\beta \otimes 1_{\mathcal{A}} - \gamma_1 \otimes x - \gamma_2 \otimes d)^{-1}\right]\gamma_1 \in M_m(\mathbb{C}) \quad (2)$$

and extends as an analytic map $z \mapsto \omega_m(ze_{11} - \gamma_0)$ to $\mathbb{C} \setminus \mathrm{supp}(\mu_{P(x,d)})$.

We are now in position to state our result.

Theorem 1.5. *Let $P \in \mathbb{C}\langle t_1, t_2\rangle$ be a selfadjoint polynomial in two noncommuting indeterminates, that really involves both indeterminates. Let $L_P = \gamma_0 \otimes 1 + \gamma_1 \otimes t_1 + \gamma_2 \otimes t_2 \in M_m(\mathbb{C}\langle t_1, t_2\rangle)$ be the selfadjoint canonical*

linearization of P and ω_m defined by (2). Let W_N and D_N be defined as in Section 1.1 and define for any $z \in \mathbb{C}\backslash\mathbb{R}$, $\xi_N(z) = \mathrm{Tr}_N\left((z - P(W_N, D_N))^{-1}\right) - \mathbb{E}\left(\mathrm{Tr}_N\left((z - P(W_N, D_N))^{-1}\right)\right)$. The sequence $(\xi_N)_{N\in\mathbb{N}}$ of $\mathcal{H}(\mathbb{C}\backslash\mathbb{R})$-valued random variables converges in distribution towards a complex centred Gaussian process $\{\mathcal{G}(z), z \in \mathbb{C}\setminus\mathbb{R}\}$ defined by $\overline{\mathcal{G}(z)} = \mathcal{G}(\bar{z})$ and $\mathbb{E}\left(\mathcal{G}(z_1)\mathcal{G}(z_2)\right) = \frac{\partial^2}{\partial z_1 \partial z_2}\gamma(z_1, z_2), z_1, z_2 \in \mathbb{C}\setminus\mathbb{R}$,

$$\begin{aligned}\gamma(z_1, z_2) = &-\mathrm{Tr}_m \otimes \mathrm{Tr}_m\left\{\log\left[\mathrm{id}_m \otimes \mathrm{id}_m - \sigma^2 T_{\{z_1,z_2\}}\right](I_m \otimes I_m)\right\}\\ &- \mathrm{Tr}_m \otimes \mathrm{Tr}_m\left\{\log\left[\mathrm{id}_m \otimes \mathrm{id}_m - \theta T_{\{z_1,z_2\}}\right](I_m \otimes I_m)\right\}\\ &+ \frac{\kappa}{2}\mathrm{Tr}_m \otimes \mathrm{Tr}_m\{T^2_{\{z_1,z_2\}}(I_m \otimes I_m)\} + (\tilde{\sigma}^2 - \sigma^2 - \theta)\mathrm{Tr}_m \otimes \mathrm{Tr}_m\{T_{\{z_1,z_2\}}(I_m \otimes I_m)\},\end{aligned}$$

where $T_{\{z_1,z_2\}} : M_m(\mathbb{C}) \otimes M_m(\mathbb{C}) \to M_m(\mathbb{C}) \otimes M_m(\mathbb{C})$ is defined for $u \in M_m(\mathbb{C}) \otimes M_m(\mathbb{C})$ by

$$T_{\{z_1,z_2\}}(u) = \int_{\mathbb{R}} ((\omega_m(z_1 e_{11} - \gamma_0) - t\gamma_2)^{-1}\gamma_1 \otimes I_m)u(I_m \otimes \gamma_1(\omega_m(z_2 e_{11} - \gamma_0) - t\gamma_2)^{-1})d\nu(t).$$

Note that a highly non-trivial first task is to prove that the logarithms involved in the definition of $\gamma(z_1, z_2)$ are well defined, making use of the contractivity of analytic self-maps on hyperbolic domains to establish the following

Proposition 1.6. *For any $z_1, z_2 \in \mathbb{C}\setminus\mathbb{R}$, the spectrum of the operator*

$$T_{\{z_1,z_2\}} : M_m(\mathbb{C}) \otimes M_m(\mathbb{C}) \to M_m(\mathbb{C}) \otimes M_m(\mathbb{C})$$

is included in the open disk of radius σ^{-2}.

2. Tweaking Algorithms in Finite-Size Random Matrix Approximations

Franck Gabriel

Discrete approximations (e.g. Riemann sum approximation, Euler method or Monte Carlo methods) are essential when sampling or calculating various quantities on a computer. In this note, we consider two situations where a random matrix approximation can be used to estimate quantities of interest.

Given a family of "large" objects $\mathcal{A} = (a_k)_{k\in K}$ about which we wish to compute some observables $\mathcal{O}(\mathcal{A})$, we approximate $\mathcal{A}$ using some families of random matrices $\mathcal{A}_N = (a_k^N)_{k\in K}$ of smaller size or rank. A naive approach consists in approximating the observables $\mathcal{O}(\mathcal{A})$ by $\mathcal{O}(\mathcal{A}_N)$, i.e. using the same algorithm, but implemented on the random matrix approximation. This may lead to some bias which we study in two distinct settings: in kernel methods (Section 2.1) and in the computation of limits of empirical eigenvalue distributions (Section 2.2). Understanding the bias provides us with alternative algorithms whose underlying motivation is to suppress the bias created by the finite-size random matrix approximation.

The results presented in Section 2.1 on Gaussian features are taken from [38] with C. Hongler, A. Jacot, B. Şimşek and F. Spadaro. Section 2.2 on permutation invariant random matrices is based on the collaborative work [3] with B. Au, A. Cébron, A. Dahlqvist and C. Male.

2.1. Kernel Method: Gaussian Random Features and Effective Ridge

We explain how one can implement approximations of kernel methods (Section 2.1.1) using random features (Section 2.1.2), by essentially approximating the Gram matrix of the kernel by a smaller rank random matrix. This creates an implicit bias (Section 2.1.2) which can be partially corrected (Section 2.1.3).

2.1.1. *Kernel Regression Method: from Kernel and Gram Matrix to Predictor*

Given N distinct data points $(x_i)_{i=1}^N \in \mathbb{R}^d$ with respective labels $(y_i)_{i=1}^N$, a kernel K and a ridge $\lambda \geq 0$, the kernel regression method with kernel K and ridge λ provides an explicit function f_λ^K such that $f_\lambda^K(x_i) \simeq y_i$ for $i = 1, \ldots, N$. Specifically, let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a positive semi-definite kernel: for any finite subset $Z \subset \mathbb{R}^d$, the Gram matrix $K_{Z,Z} = (K(z_i, z_j))_{z_i, z_j \in Z}$ of K is a symmetric positive semi-definite matrix. We assume that $\lambda > 0$. The predictor $f_\lambda^K \in \mathrm{Span}\,(K_{x_i}, i = 1, \ldots, N)$, where $K_{x_i} : x \mapsto K(x_i, x)$, minimizes over $\mathrm{a} = (a_i)_{i=1}^N$ the cost:

$$\mathcal{C}(f_\mathrm{a}) = \frac{1}{N}\sum_{i=1}^N (f_\mathrm{a}(x_i) - y_i)^2 + \frac{\lambda}{N}\mathrm{a}^T K_{X,X}\mathrm{a},$$

with $X = \{x_1, \ldots, x_N\}$ and $f_\mathrm{a} = \sum_{i=1}^N a_i K_{x_i}$. From the first-order condition, we get the following.

Lemma 2.1. *The Kernel regression predictor with kernel K and ridge λ is $f_\lambda^K(x) = K_{x,X}\left[K_{X,X} + \lambda\mathrm{Id}\right]^{-1}\mathrm{y}$, where $K_{x,X} = (K(x, x_i))_{i=1,\ldots,N}$, $K_{X,X}$ is the Gram matrix and $\mathrm{y} = (y_i)_{i=1,\ldots,N}$ is the column vector of labels.*

Although kernel predictor has a simple expression, its use in practice is usually challenging, especially if the cardinality of the data set is large, since one has to invert the $N \times N$ matrix $K_{X,X} + \lambda\mathrm{Id}$. Actually, if the kernel has finite-dimensional feature space, the computation of the kernel predictor is much simpler.

Remark 2.2. *The so-called* kernels with finite-dimensional feature space *can be obtained by considering a feature map $\phi : \mathbb{R}^d \to \mathbb{R}^P$ with $\phi(x) = (\phi_1(x), \ldots, \phi_P(x))$, and defining $K(x, x') = \phi(x)^T\phi(x')$. One can express the quantities involved in the definition of $f_\lambda^K(x)$ using the $N \times P$ data matrix $(\boldsymbol{\phi}_X)_{i,j} = \phi_j(x_i)$: $K_{x,X} = \phi(x)\boldsymbol{\phi}_X^T$, and $K_{X,X} = \boldsymbol{\phi}_X\boldsymbol{\phi}_X^T$. In this setting, the predictor $f_\lambda^K(x) = \phi(x)\boldsymbol{\phi}_X^T[\boldsymbol{\phi}_X\boldsymbol{\phi}_X^T + \lambda\mathrm{Id}]^{-1}\mathrm{y}$ is also equal to:*

$$f_\lambda^K = \phi(x)\left[\boldsymbol{\phi}_X^T\boldsymbol{\phi}_X + \lambda\mathrm{Id}\right]^{-1}\boldsymbol{\phi}_X^T\mathrm{y}. \tag{3}$$

One has then to invert a $P \times P$ matrix: the computation complexity is independent of the number of data points and yields an efficient way to compute f_λ^K when $P \ll N$. Note that, in this setting, f_λ^K can also be obtained by minimizing $\mathcal{C}(f_\theta) = \frac{1}{N}\sum_{i=1}^N (f_\theta(x_i) - y_i)^2 + \frac{\lambda}{N}\theta^T\theta$, over the space of functions of the form $f_\theta = \sum_{i=1}^P \theta_i\phi_i$.

2.1.2. *Random Features Method: implicit bias and effective ridge.*

Given a general kernel K, the random features models of Rahimi and Recht in [50] consist of considering random approximations of K by kernels with finite-dimensional feature space, i.e. considering a P dimensional random process $\phi = (\phi_1, \ldots, \phi_P)$ on $\mathbb{R}^n$ such that $\lim_{P\to\infty}\phi(x)^T\phi(y) = K(x, y)$. The random features predictor $f_{\lambda,P}^{RF}$ is the kernel estimator obtained from the kernel $K_P(x, y) := \phi(x)^T\phi(y)$, and with ridge λ. The random features models are thus randomized approximations of the kernel methods aimed at easing the computational challenges. Using the law of large numbers, one can exhibit a simple example of such random approximation:

Theorem 2.3. *Let $\varphi_1, \ldots, \varphi_P$ be i.i.d. centered random processes, each with covariance K. Let $f_{\lambda,P}^{RF}$ be the predictor obtained with the random features $\varphi_1/\sqrt{P}, \ldots, \varphi_P/\sqrt{P}$. As $P \to \infty$, for any $x \in \mathbb{R}^d$, $f_{\lambda,P}^{RF}(x) \to f_\lambda^K(x)$.*

By approximating the Gram matrix $K_{X,X}$ by a random matrix of the form $\boldsymbol{\phi}_X\boldsymbol{\phi}_X^T$ (and similarly for $K_{x,X}$), we obtain an efficient way to approximate the kernel predictor. Yet, as advertised in the introduction, using the same algorithm for $\boldsymbol{\phi}_X\boldsymbol{\phi}_X^T$ as for $K_{X,X}$ might not be the best option. Indeed, since the algorithm is not a linear transformation of $\boldsymbol{\phi}_x\boldsymbol{\phi}_X^T$ and $\boldsymbol{\phi}_X\boldsymbol{\phi}_X^T$, the predictor might be biased: $\mathbb{E}[f_{\lambda,P}^{RF}(x)] \neq f_\lambda^K(x)$.

In [38], we consider as random features P i.i.d. Gaussian random processes $\varphi_1, \ldots, \varphi_P$, which are centered and each with covariance K. We then show that there is an implicit regularisation due to the finite sampling:

Theorem 2.4 (Theorem 4.1 of [38])**.** *For $N, P > 0$, $x \in \mathbb{R}^d$, and $\lambda > 0$, $\mathbb{E}[f_{\lambda,P}^{RF}(x)] \simeq f_{\lambda_\mathrm{eff}}^K(x)$ where the* effective ridge *$\lambda_\mathrm{eff} > \lambda$ is the unique positive solution of*

$$\lambda_\mathrm{eff} = \lambda + \lambda_\mathrm{eff}\frac{1}{P}\mathrm{Tr}\left[K_{X,X}\left(K_{X,X} + \lambda_\mathrm{eff}\mathrm{Id}\right)^{-1}\right].$$

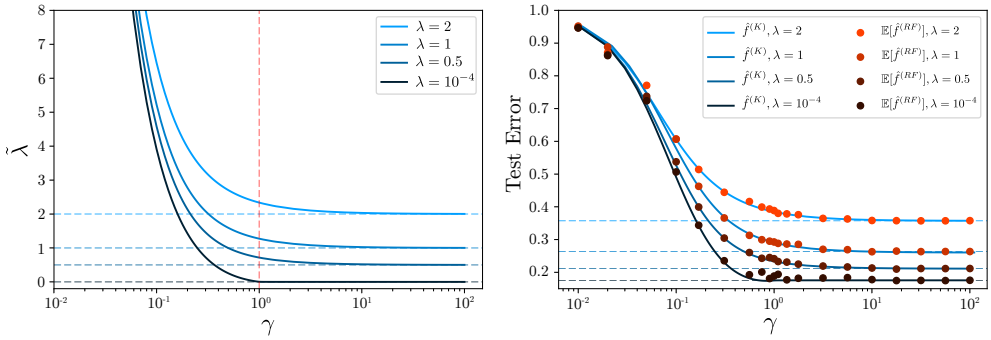

FIGURE 1. **Left:** Effective ridge as a function of $\gamma := P/N$ for different values of λ. **Right:** Comparison of the test errors of $\mathbb{E}[f^{RF}_{\lambda,P}]$ and $f^K_{\lambda_{\text{eff}}}$.

Hence, as illustrated in Figure 1, while the random feature method was intended to approximate f^K_λ, we are actually approximating $f^K_{\lambda_{\text{eff}}}$. The connection between $f^{RF}_{\lambda,P}$ and $f^K_{\lambda_{\text{eff}}}$ can be further refined. First we obtain an explicit bound on the difference between $\mathbb{E}[f^{RF}_{\lambda,P}(x)]$ and $f^K_{\lambda_{\text{eff}}}(x)$. Besides, assuming that the labels $y_1, \dots, y_N$ were obtained from an existing but unknown function f^*, i.e. $f^*(x_i) = y_i$ for $i = 1, \dots, N$, and that $x_1, \dots, x_n$ and future datapoints are i.i.d. samples of a probability measure $\mathcal{D}$ on $\mathbb{R}^d$, the difference between the test errors of $\mathbb{E}[f^{RF}_{\lambda,P}]$ and $f^K_{\lambda_{\text{eff}}}$ can also be bounded (see an illustration of this in Figure 1). At last, by showing an upper bound on the variance of $f^{RF}_{\lambda,P}$, we can control the difference between expected test errors of $f^{RF}_{\lambda,P}$ and $f^K_{\lambda_{\text{eff}}}$.

2.1.3. *Improved Algorithm*

From the previous discussion, we propose an algorithm to partially correct the regularization bias of the finite sampling in the Gaussian random features method.

Algorithm for an approximation of $f^K_\lambda(x)$

(1) Sample $\varphi_1, \dots, \varphi_P$ i.i.d. centered Gaussian processes on $\{x_1, \dots, x_N, x\}$ with covariance kernel K.
(2) Consider $\mu = \lambda - \frac{\lambda}{P}\text{Tr}[K_{X,X}(K_{X,X} + \lambda\text{Id})^{-1}]$.
(3) Return $f^{RF}_\mu(x) = \phi(x)\left[\boldsymbol{\phi}_X^T\boldsymbol{\phi}_X + \mu\text{Id}\right]^{-1}\boldsymbol{\phi}_X^T\text{y}$ where $\phi(x) = (\varphi_1(x)/\sqrt{P}, \dots, \varphi_P(x)/\sqrt{P})$, the data matrix is $(\boldsymbol{\phi}_X)_{i,j} = \varphi_j(x_i)/\sqrt{P}$, and $\text{y} = (y_1, \dots, y_N)$ is the vector of labels.

2.2. **Large permutation invariant random matrices**

We now illustrate how the same idea developed in the previous section can be implemented in the setting of large random matrices. Permutation-invariant symmetry leads to a "rigidity" property in the large size limit of mixed observables: independent random matrices become asymptotically free with amalgamation over the diagonal (Section 2.2.2). When sampling finite size random matrices, this freeness property is only approximately satisfied and a fixed-point algorithm can be implemented in order to correct this bias (Section 2.2.3).

2.2.1. *Eigenvalue distribution, symmetries and notions of freeness*

When we speak of a family of Hermitian random matrices $\boldsymbol{A} = (A_N)_{N\in\mathbb{N}}$, we assume that for any $N \in \mathbb{N}$, A_N is a random $N \times N$ Hermitian matrix in $\mathcal{M}_N(L^{\infty-})$, i.e. a matrix for which all moments of the form $\mathbb{E}[\prod_{k=1}^n A_{i_1,j_1}]$ exist and are finite. The empirical eigenvalues distribution of A_N is $\mu_{A_N} = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$, where $(\lambda_i)_{i=1,\dots,N}$ are the eigenvalues of A_N. Given a family $\boldsymbol{A}$ which, for example, plays the role of a signal, it is natural to investigate the effect of an additive perturbation. If $\boldsymbol{B}$ is a family of Hermitian random matrices, independent of $\boldsymbol{A}$, assuming that both $\lim_{N\to\infty}\mu_{A_N}$ and $\lim_{N\to\infty}\mu_{B_N}$ exist as $N \to \infty$, how can we obtain numerically $\lim_{N\to\infty}\mu_{A_N+B_N}$ (if it exists)? A naive, yet natural algorithm consists in: (1) selecting a large N,

(2) sampling one or many realizations of A_N and B_N, (3) computing the spectral distribution of $A_N + B_N$ for each sample (or average the spectral distributions over the samples if needed).

In Section 2.2.3, we provide an algorithm which allows to correct a bias in the previous algorithm which appears due to the finite size of the samples. To explain this bias, we review some theoretical results about the limit of $\mu_{A_N+B_N}$ when one of the families satisfies some symmetry: unitary and permutation invariance.

Unitary invariant matrices. When one of the two families is invariant under conjugation by the unitary group, for example, if for any unitary matrix U_N, the matrix $U_N A_N U_N^*$ has the same law as A_N, the fact that the limit of $\mu_{A_N+B_N}$ exists (in expectation, probability or almost surely depending on some technical assumptions), is a consequence of a line of works, e.g. [18, 20, 25, 46, 49, 57, 60], related with free probability theory. The measure μ_{A_N} can be encoded using the so-called moments of A_N since for any $k \geq 0$,

$$\int z^k d\mu_{A_N}(z) = \operatorname{tr}\left[A_N^k\right],$$

where $\operatorname{tr} = {}^{\operatorname{Tr}}/_N$ is the normalized trace. Hence, the study of $\mu_{A_N+B_N}$ can be reduced to the study of the moments $\operatorname{tr}\left[(A_N + B_N)^k\right]$, or more generally of the mixed moments $\operatorname{tr}\left[\prod_i P_i(A_N)Q_i(B_N)\right]$ where P_i and Q_i are polynomials in $\mathbb{C}\langle X\rangle$. If $\boldsymbol{A}$ is invariant under conjugation by the unitary group and if $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent, then the mixed moments of A_N and B_N converge and the matrices A_N and B_N are asymptotically free:

Theorem 2.5. *Assume that for any k, $\operatorname{tr}\left[A_N^k\right]$ and $\operatorname{tr}\left[B_N^k\right]$ converge almost surely and that $\boldsymbol{A}$ is invariant by conjugation by the unitary group, then, the matrices A_N and B_N are almost surely asymptotically free in the following sense: for any monomials $P_1, \ldots, P_n \in \mathbb{C}\langle X\rangle$, for any $C_{N,1}, \ldots, C_{N,n} \in \{A_N, B_N\}$,*

$$\operatorname{tr}\left[\prod_{i=1}^{n} \left[P_i(C_{N,i}) - \operatorname{tr}\left[P_i(C_{N,i})\right] \mathrm{Id}\right]\right]$$

converges to zero almost surely whenever $C_{N,i} \neq C_{N,i+1}$ for $i = 1, \ldots, n-1$.

Remark 2.6. *Note that the asymptotic freeness of A_N and B_N allows one to recover the asymptotic moments of $\mu_{A_N+B_N}$ from the asymptotic moments of A_N and B_N. For example, applying the asymptotic freeness property with $P_1(X) = P_2(X) = X$, yields $\int z^2 \mu_{A_N+B_N}(dz) = \operatorname{tr}[(A_N+B_N)^2] \sim \operatorname{tr}[A_N^2] + \operatorname{tr}[B_N^2] + 2\operatorname{tr}[A_N]\operatorname{tr}[B_N]$. More generally, as a corollary of Theorem 2.5, there exists a universal function F such that*

$$\left(\lim_{N\to\infty} \operatorname{tr}\left[(A_N + B_N)^k\right]\right)_{k\geq 0} = F\left[\left(\lim_{N\to\infty} \operatorname{tr}\left[A_N^k\right]\right)_{k\geq 0}, \left(\lim_{N\to\infty} \operatorname{tr}\left[B_N^k\right]\right)_{k\geq 0}\right].$$

The function F can be obtained as followed. First, $\operatorname{tr}((A_N + B_N)^k) = \sum_{\substack{n_1,\ldots,n_\ell \\ m_1,\ldots,m_\ell}} \operatorname{tr}\left(\prod_{i=1}^{\ell} A_N^{n_i} B_N^{m_i}\right)$ where the sum is over the non-negative integers $n_1, \ldots, n_\ell$ and $m_1, \ldots, m_\ell$ such that $\sum_{i=1}^{\ell}(n_i + m_i) = k$. The freeness property allows then to compute by induction the limit of $\operatorname{tr}\left(\prod_{i=1}^{\ell} A_N^{n_i} B_N^{m_i}\right)$ in terms of the limit of the moments of A_N and B_N. Indeed, $\operatorname{tr}\left(\prod_{i=1}^{\ell} \left(A_N^{n_i} - \operatorname{tr}\left(A_N^{n_i}\right)\right)\left(B_N^{m_i} - \operatorname{tr}\left(B_N^{m_i}\right)\right)\right) \underset{N\to\infty}{\longrightarrow} 0$, thus, by expanding the product, the limit of $\operatorname{tr}\left(\prod_{i=1}^{\ell} A_N^{n_i} B_N^{m_i}\right)$ can be computed as a linear combination of terms of the form

$$\lim_{N\to\infty} \prod_{i=1}^{u} \operatorname{tr}\left(A_N^{\ell_i}\right) \prod_{i=1}^{v} \operatorname{tr}\left(B_N^{k_i}\right) \operatorname{tr}\left(\prod_{i=1}^{\ell'} A_N^{p_i} B_N^{q_i}\right)$$

where $\sum_{i=1}^{\ell'}(p_i + q_i) < k$. By induction, the limit of $\operatorname{tr}\left(\prod_{i=1}^{\ell'} A_N^{p_i} B_N^{q_i}\right)$ can itself be written as a function of the limiting moments of A_N and B_N. This shows the existence of the function F and allows one to compute it by induction.

For finite size N, the matrices are not free: the moments of $A_N + B_N$ are not obtained by applying F to the moments of A_N and B_N. This hints at the fact that one could actually use the asymptotic freeness theorem to correct the "finite-dimensional" bias in the interaction between A_N and B_N. Doing so, one obtains a new algorithm for the approximation of $\lim_{N\to\infty} \mu_{A_N+B_N}$: one should find a measure μ such that for any k,

$$\int z^k d\mu(z) = F\left[\left(\operatorname{tr}\left[A_N^k\right]\right)_k, \left(\operatorname{tr}\left[B_N^k\right]\right)_k\right]$$

where A_N and B_N are realizations of the random matrices. One is faced with two problems in order to implement this algorithm: (1) the condition should hold for an infinite number of k, (2) one has to find the measure μ. In fact, there exists a fixed-point algorithm (explained in a more general setting in Section 2.2.3), based on the notion of Stieltjes transform, which allows one to handle the two difficulties raised.

Permutation invariant matrices. The unitary invariance is a powerful symmetry, but which is not satisfied by simple models such as random permutations or adjacency matrices of random graphs. A weaker symmetry was proposed and studied in [45]: permutation invariance. The family $\boldsymbol{A}$ is *permutation invariant* if for any permutation matrix S_N, the matrix $S_N A_N S_N^{-1}$ has the same law as A_N. It was then shown in [45] using graph observables and in [33,34] using a dual point of view, i.e. partitions observables and cumulants, that:

- two families of random matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ which are independent and permutation invariant are not necessarily asymptotically free in expectation,
- there exist families of observables $\mathcal{O}(A_N)$, $\mathcal{O}(B_N)$, and $\mathcal{O}(A_N + B_N)$ which generalize the notion of expected moments $\mathbb{E}[\operatorname{tr}(A_N^k)]$, $\mathbb{E}[\operatorname{tr}(B_N^k)]$ and $\mathbb{E}[\operatorname{tr}((A_N + B_N)^k)]$, and a universal function T such that $\lim_{N\to\infty} \mathcal{O}(A_N + B_N) = T\left[\lim_{N\to\infty} \mathcal{O}(A_N), \lim_{N\to\infty} \mathcal{O}(B_N)\right]$,
- one can understand the limit of the expected eigenvalues distribution using observables in A_N and observables in B_N, but, in general, one cannot do so using only the moments $\mathbb{E}\left[\operatorname{tr}(A_N^k)\right]$ and $\mathbb{E}\left[\operatorname{tr}(B_N^k)\right]$.

The fact that asymptotic freeness is lost when one weakens the symmetry seems to hinder us from using a fixed-point algorithm, as in the unitary invariance case, in order to obtain an approximation of $\lim_{N\to\infty} \mu_{A_N+B_N}$ using samples of A_N and B_N. Besides, the observables cannot be naturally indexed by integers, hence it seems complicated to define a version of the Stieltjes transform suited to these observables: this impedes us from using the analytical tools necessary to obtain such fixed-point algorithm. In fact, in [3], we show that asymptotic freeness can be recovered, at the cost of having to consider non-commutative conditional expectation and the notion of freeness associated with it. As a result, using the operator-valued Cauchy transform, we propose a fixed-point algorithm in order to approximate the limiting eigenvalues distribution of $A_N + B_N$.

2.2.2. *Large permutation invariant random matrices and freeness over the diagonal*

Asymptotic freeness appears when one considers random matrices in the space of matrices $\mathcal{M}_N$ endowed with the non-commutative expectation $\operatorname{tr} : \mathcal{M}_N \to \mathbb{C}$ with $M \mapsto \frac{1}{N}\operatorname{Tr}(M)$. Actually, $\mathcal{M}_N$ can be endowed with another interesting structure: the structure of operator-valued probability space. If $\mathcal{D}_N \subset \mathcal{M}_N$ is the space of diagonal matrices, the map $\Delta : \mathcal{M}_N \to \mathcal{D}_N$ such that $M \mapsto D = (\delta_{i,j} M_{i,j})_{i,j}$ satisfies that for any $D_1, D_2 \in \mathcal{D}_N$, and any $M \in \mathcal{M}_N$, $\Delta(D_1 M D_2) = D_1 \Delta(M) D_2$.

Remark 2.7. *The map Δ can be regarded as a non-commutative conditional expectation. Note that* $\operatorname{tr}(\Delta(M)) = \operatorname{tr}(M)$*, hence the knowledge of the family of diagonal matrices $\Delta(M^k)$ is sufficient in order to recover the moments of the eigenvalues distribution of M as* $\operatorname{tr}(\Delta(M^k)) = \operatorname{tr}(M^k) = \int z^k d\mu_M(z)$.

Hence, the limit of $\mu_{A_N+B_N}$ can be deduced from the study of $\Delta((A_N + B_N)^k)$ whose asymptotic behaviour can fortunately be derived from those of $\Delta(A_N^\ell), \Delta(B_N^\ell)$ for $\ell \leq k$. Indeed, Theorem 2.5 essentially holds for permutation invariant matrices if one replaces the non-commutative expectation tr by the conditional non-commutative expectation Δ.

Theorem 2.8 (Theorem 1.3 of [3])**.** *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be independent families of permutation invariant random matrices. Under some technical assumptions on $\boldsymbol{A}$ and $\boldsymbol{B}$, the families $\boldsymbol{A}$ and $\boldsymbol{B}$ are asymptotically free with*

amalgamation over $\mathcal{D}_N$: for any monomials $P_1, \dots, P_n \in \mathcal{D}_N \langle X \rangle$, any $C_{N,1}, \dots, C_{N,n} \in \{A_N, B_N\}$, the matrix

$$\epsilon_N = \Delta \left[\prod_{i=1}^{n} [P_i(C_{N,i}) - \Delta [P_i(C_{N,i})]] \right],$$

converges to zero in normalized Schatten p-norm for any $p \in [1, \infty[$ (i.e $\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}[(\epsilon_N \epsilon_N^*)^{\frac{p}{2}}]] = 0$) whenever $C_{N,i} \neq C_{N,i+1}$ for $i = 1, \dots, n-1$.

Remark 2.9. *Note that the asymptotic freeness with amalgamation over $\mathcal{D}_N$ allows one to recover the asymptotic moments of $\mu_{A_N+B_N}$ from $\Delta(A_N)$ and $\Delta(B_N)$. For example, $\Delta[(A_N - \Delta(A_N))(B_N - \Delta(B_N))] \simeq 0$ implies that $\Delta(A_N B_N) \simeq \Delta(A_N)\Delta(B_N)$, hence $\mathrm{tr}[A_N B_N] = \mathrm{tr}[\Delta(A_N B_N)] \simeq \mathrm{tr}[\Delta(A_N)\Delta(B_N)]$. This yields:*

$$\int z^2 \mu_{A_N+B_N}(dz) \sim \mathrm{tr}\left[A_N^2\right] + \mathrm{tr}\left[B_N^2\right] + 2\mathrm{tr}\left[\Delta(A_N)\Delta(B_N)\right].$$

More generally, there exists a function F_Δ such that $\left(\Delta\left[(A_N + B_N)^k\right]\right)_k \simeq F_\Delta\left[\left(\Delta\left[A_N^k\right]\right)_k, \left(\Delta\left[B_N^k\right]\right)_k\right]$.

For finite size N, the matrices are not asymptotically free with amalgamation. This hints at a new algorithm in order to obtain an approximation of $\lim_{N\to\infty} \mu_{A_N+B_N}$: given a realization of A_N, B_N, find μ such that

$$\int z^k d\mu(z) = \mathrm{tr}\left[F_\Delta\left[\left(\Delta\left[A_N^k\right]\right)_k, \left(\Delta\left[B_N^k\right]\right)_k\right]\right],$$

holds for any k. Again, one is faced with the same two problems as in the unitary invariance setting. In order to handle the two difficulties raised, one can use the operator-valued Cauchy transform:

Definition 2.10. *The operator-valued Cauchy transform of a $N \times N$ Hermitian matrix M is the function*

$$G_M : \mathcal{D}_N^+ \to \mathcal{D}_N^-, \qquad Z \mapsto \Delta\left[(Z - M)^{-1}\right],$$

where $\mathcal{D}_N^\pm = \{Z \in \mathcal{D}_N : \pm\Im(Z) = \pm\frac{Z - Z^*}{2i} > 0\}$. The H-transform of M is $H_M : Z \mapsto G_M(Z)^{-1} - Z$.

Remark 2.11. *Note that $\mathrm{tr}[G_M(z\mathrm{Id}_N)]$ is the Stieltjes transform at z of μ_M. Hence, the measure μ_M can be obtained by applying the Stieltjes-Perron inversion formula (Theorem X.6.1 of [30]) to $z \mapsto \mathrm{tr}[G_M(z\mathrm{Id}_N)]$.*

2.2.3. *Fixed-point algorithm*

If A_N and B_N were actually free with amalgamation over $\mathcal{D}_N$, then one could compute $G_{A_N+B_N}$ as follows (see [10]). For any $Z \in \mathcal{D}^+$, we consider the unique solution $\Omega(Z)$ to

$$\Omega(Z) = H_{B_N}(H_{A_N}(\Omega(Z)) + Z) + Z$$

which can be computed by iterating the map $M \mapsto H_{B_N}(H_{A_N}(M) + Z) + Z$. Then $G_{A_N+B_N} = G_{A_N} \circ \Omega$.

Since the measure $\mu_{A_N+B_N}$ can be obtained from $G_{A_N+B_N}$ (Remark 2.11), we get the following algorithm that enables the correction of the "finite-dimensional" bias in the interaction between A_N and B_N and provides an approximation of $\lim_{N\to\infty} \mu_{A_N+B_N}$. For the sake of simplicity, we provide the algorithm when the limiting measure has a density at x; the general case can be obtained using the Stieltjes-Perron inversion formula.

Algorithm for an approximation of the density of $\lim_{N\to\infty} \mu_{A_N+B_N}$ at x.

(1) Consider a large N, small $\epsilon > 0$, a small threshold $\mathrm{t} > 0$, and sample A_N and B_N.
(2) Set $\Omega_0 = (x + i\epsilon)\mathrm{Id}_N = Z$.
(3) Compute $\Omega_{n+1} = H_{B_N}(H_{A_N}(\Omega_n) + Z) + Z$ until $\|\Omega_{n+1} - \Omega_n\| < \mathrm{t}$ for some $n = n_0$.
(4) Return $-\frac{1}{\pi}\Im\left(\frac{1}{N}\mathrm{Tr}[(\Omega_{n_0} - A_N)^{-1}]\right)$.

In Figure 2, the matrix A_N is a block variance GUE and B_N is a corrupted permutation (details in Section 3.2 of [3]). The histogram is consistent with the blue line, thereby illustrating the fact that they are asymptotically free with amalgamation over $\mathcal{D}_N$, while the red line confirms that they are not asymptotically free.

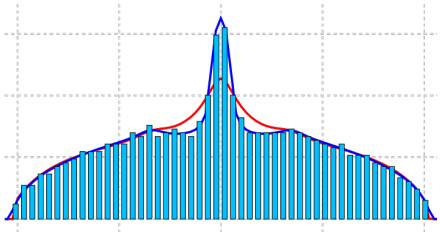

FIGURE 2. Example of limiting eigenvalues distribution of $A_N + B_N$. **Histogram:** eigenvalues histogram of $A_N + B_N$. **Blue:** Fixed-point algorithm. **Red:** limiting eigenvalues distribution of $A_N + C_N$ when $C_N \sim B_N$ is asymptotically free from A_N.

3. SPARSE MATRICES: CONVERGENCE OF THE REVERSE CHARACTERISTIC POLYNOMIAL

Simon Coste

Spectral properties of non-Hermitian random matrices can have different behaviors depending on their degree of sparsity. These properties are now well understood for dense matrices with iid entries; a well-known example is the Circular Law [16], for which the optimal sparsity threshold is known [8, 52]. However, when the matrices in question are very sparse, with a fixed number of non-zero entries on each row, possibly with dependencies, the problem becomes different and more challenging. In this contribution based on Coste [21], we show how the limiting spectral objects in this sparse regime are no longer Gaussian, but Poisson.

3.1. Random, sparse matrices and their characteristic polynomial

The random matrix model we consider in this note is as follows: A_n is a square $n \times n$ matrix whose n^2 entries are independent Bernoulli(d/n) random variables, where d is a fixed positive number. This non-Hermitian matrix arises, for example, as the adjacency matrix of a directed Erdős-Rényi graph G with mean in-degree and mean out-degree d. Its *empirical spectral distribution* is the atomic measure defined by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A_n)} \tag{4}$$

where $|\lambda_i(A_n)| \geqslant \cdots \geqslant |\lambda_n(A_n)|$ are the complex eigenvalues of A_n ordered by decreasing modulus. Not much is known on the asymptotic behaviour of μ_n. However, it turns out that the reverse characteristic polynomial of A_n has a tractable asymptotic behaviour. More precisely, we study $q_n(z) = \det(I - zA_n)$, a sequence of random complex polynomials with real coefficients. The k-th coefficient of q_n, say Δ_k, is given by

$$\Delta_k = \frac{(-1)^k}{k!} P_k(\mathrm{tr}(A_n^1), \ldots, \mathrm{tr}(A_n^k)) \tag{5}$$

for some polynomial P_k — its expression involves the Bell polynomials. A natural method to study the asymptotic behaviour of q_n, inspired by [9, 17] who studied the dense case, consists in identifying the limiting distribution of the traces of powers of A_n, which will directly prove the convergence of the coefficients of q_n towards the corresponding limit. This first step can be performed thanks to combinatorial methods which are of great use in trace methods: in short, the trace of A_n^k counts the number of directed closed paths in the graph

G, of length k. It turns out that the dominant contribution to this counting comes from the paths that are simple cycles of length h, possibly crossed multiple times (consequently, h must divide k); and the number of these small cycles is asymptotically Poisson thanks to a rare-events-theorem type analysis. The limit of $\mathrm{tr}(A_n^k)$ is given in the following definition and theorem.

Definition 3.1. *Let $d > 0$, and let $(Y_\ell : \ell \in \mathbb{N}^*)$ be a family of independent random variables, with $Y_\ell \sim \mathrm{Poi}(d^\ell/\ell)$. We define a family of (non-independent) random variables by*

$$X_k := \sum_{\ell | k} \ell Y_\ell \qquad\qquad (k \in \mathbb{N}^*) \tag{6}$$

where $a|b$ means that b is a nonzero multiple of a.

Theorem 3.2 (trace asymptotics)**.** *For every integer k, the following joint weak convergence holds:*

$$(\mathrm{tr}(A_n^1), \dots, \mathrm{tr}(A_n^k)) \xrightarrow[n\to\infty]{\text{law}} (X_1, \dots, X_k). \tag{7}$$

This settles the limiting behaviour of the coefficients Δ_k of q_n: thanks to (5), we see that Δ_k converges towards $c_k := (-1)^k P_k(X_1, \dots, X_k)/k!$ and it is tempting to infer that q_n converges towards the random function $F(z) := \sum_n c_n z^n$, in some topology like the natural topology of uniform convergence of compact sets. This is true, but with a caveat: even if the q_n are polynomials, hence entire functions, the function F is not entire.

Theorem 3.3. *The random series*

$$F(z) = \sum_{n=0}^{\infty} c_n z^n \tag{8}$$

is almost surely uniformly convergent in the disk $D(0, 1/d)$ and can be analytically extended to the disk $D(0, 1/\sqrt{d})$, in which it almost surely has exactly one zero at the location $z_d = 1/d$. This zero is simple.

We endow the set $\mathbb{H}_{1/\sqrt{d}}$ of analytic functions on $D(0, 1/\sqrt{d})$ with the topology of uniform convergence over compact sets. Then, the sequence of random polynomials (q_n) weakly converges in this topology towards F.

This mode of convergence is strictly stronger than the convergence of coefficients (aka, finite-dimensional convergence) displayed before, and its proof needs an extra ingredient. A classical argument coming back to Shirai [55] reduces this mode of convergence to (i) the convergence of coefficients and (ii) the tightness of (q_n). This tightness can be hard to prove. In general, it is sufficient to prove uniform bounds on the L^2 norm of the coefficients of q_n: typically, if one wants to prove that (q_n) is tight in $\mathbb{H}_{d^{-1/2}}$, one can simply prove that $\mathbb{E}[|\Delta_k|^2] \leqslant c/d^k$ for some c not depending of k and n. This was almost trivial to check in the dense, centered case studied in [17]; but for non-centered, sparse models, there is no general recipe. For the moment, the simplest way to do this is to use an alternative expression to (5), which is

$$\Delta_k = \sum_I \det(I)$$

where the sum is over all principal sub-matrices of A_n of size $k \times k$. In the model under investigation for A_n, with independent Bernoulli(d/n) entries, it turns out that certain determinant identities naturally appear in this expression, that allowed for a simple proof of this tightness. In another model of sparse random matrices (sums-of-permutations), a technical analysis of this expression could also lead to a proof of tightness [23]; but there is, for the moment, no general method.

There are various paths for exploring Theorem 7. The first one is to see to which extent does this theorem provide information on the asymptotic behaviour of the eigenvalues $\lambda_i(A_n)$; the second one is on a better understanding of F.

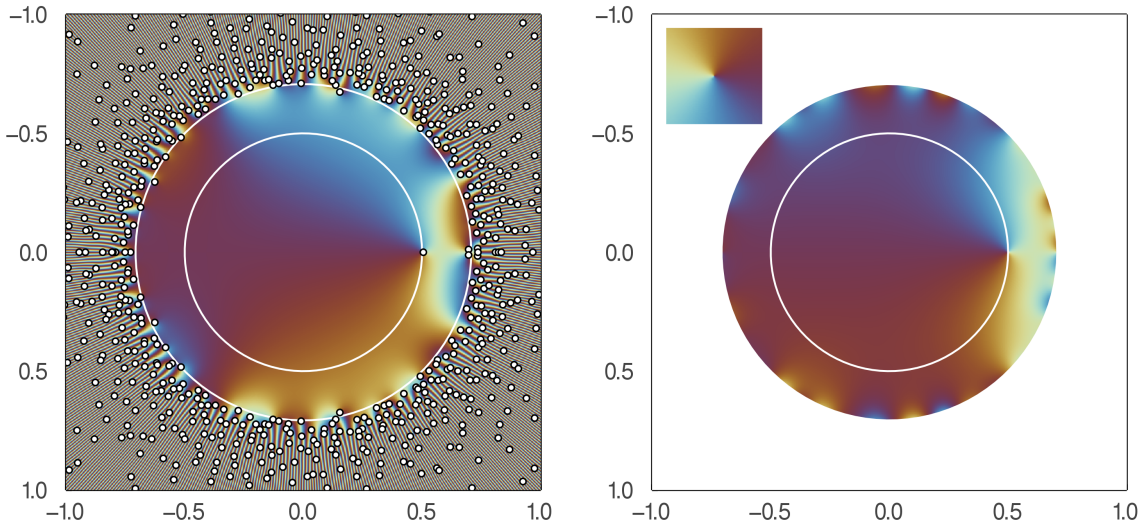

FIGURE 3. An illustration of Theorem 3.3 when $d > 1$. The color scheme used for these domain colourings is depicted in the small inset of the right picture. Note that in both cases, the depicted function has real coefficients, hence the apparent skew-symmetry with respect to the real axis: we have $f(\bar{z}) = \overline{f(z)}$. **Left** is the domain colouring of $z \mapsto \det(I - zA)$, where A is an $n \times n$ random matrix with independent entries equal to 1 with probability d/n and 0 otherwise ($n = 500$ and $d = 2$). The inverse eigenvalues of A are in white and the two circles have radius $1/d$ and $1/\sqrt{d}$. **Right** is the colouring of the random analytic function F in (8). What we see inside $D(0, 1/\sqrt{d})$ in the left picture converges in distribution towards what we see in the right picture.

3.2. **Asymptotic behaviour of eigenvalues**

The convergence of q_n towards F does not imply the convergence of μ_n towards some measure. Indeed, this convergence $q_n \to F$ is only proved inside the disk $D(0, 1/\sqrt{d})$, and the roots of q_n in this set are the inverses of the eigenvalues of A_n outside $D(0, \sqrt{d})$, the 'extremal' eigenvalues. But obtaining the asymptotic behaviour of these eigenvalues is in itself an important topic in random matrix theory. These extremal eigenvalues often capture deep properties of the underlying matrix (or, in this case, directed graph): we refer for example to the celebrated Alon-Friedman theorem for context [15, 32].

In our case, the convergence mode of Theorem 7 implies the convergence of the zeroes, another simple result from [55]. As mentioned, since F has almost surely one zero at $1/d$, it implies the following result: for every ε, with probability going to 1 as $n \to \infty$,

(1) The largest eigenvalue $\lambda_1(A_n)$ is within distance ε to $d = 1/(1/d)$;
(2) All the other eigenvalues of A_n are contained in the disk $D(0, \sqrt{d} + \varepsilon)$.

This theorem implies that if μ_n converges towards some measure, this measure will have support inside the disk $D(0, \sqrt{d})$. It is supposed that this result is sharp, in that there is at least one eigenvalue of A_n with modulus greater than $\sqrt{d} - \varepsilon$; but for the moment, this is not proved. This result was previously known from the papers [17, 24], where it was used in various statistical contexts; but the proof therein was extremely technical. The new proof method shown in this note is considerably simpler and elegant.

3.3. **The Poisson Holomorphic Chaos**

The random analytic function F, pictured in the right panel of Figure 3, is defined as a random series; however, a few alternative representations are shown in [22], notably the infinite product representation

$$F(z) = \prod_{\ell=1}^{\infty} (1 - z^\ell)^{Y_\ell}$$

where Y_ℓ was defined in Definition 3.1, and the exponential representation

$$F(z) = \exp\left\{ -\sum_{n=1}^{\infty} X_n \frac{z^n}{n} \right\}. \tag{9}$$

For comparison, the limiting functions when A_n is a matrix with iid entries (not sparse), studied for example in [17, 47, 51], are typically like (9) but with the X_k being iid Gaussian random variables, real or complex depending on the model. What happens at the border of the radius of convergence is crucial; in all these cases, this border is entirely composed of singularities, so that the trace of F on it cannot be defined as a function, but only as a generalized function. When the X_n are complex Gaussians as in [47], this distribution is called the Gaussian Holomorphic Chaos; in our case, this trace, noted PHC_d and formally defined thereafter, is called Poisson Holomorphic Chaos. That this object is really a Poisson analog of the GHC bears no doubt, and it was recently checked in [23] that when $d \to \infty$, PHC_d converges towards the GHC.

Definition 3.4. *The Poisson Holomorphic Chaos of index $d > 1$, noted PHC_d, is the random distribution on $\mathbb{T}_{d^{-1/2}}$ almost surely defined by*

$$(\mathrm{PHC}_d, \varphi) = \lim_{r \to d^{-1/2}} \frac{1}{2\pi} \int_0^{2\pi} F(r\mathrm{e}^{it}) \varphi(r\mathrm{e}^{it}) \mathrm{d}t \tag{10}$$

where φ is any trigonometric polynomial on $\mathbb{T}_{d^{-1/2}}$.

Proposition 3.5. *Let $d > 1$. Almost surely, the random distribution PHC_d is s-Sobolev for every $s < -1/2$.*

This gives a hint on the regularity of PHC_d; it can be supposed that, just like the Gaussian Holomorphic Chaos, PHC_d is not s-Sobolev for $s > -1/2$; however, the simplest way to prove this fact seems to rely on the study of the distribution of the total mass of PHC_d, defined as $\lim_{r\to\infty} \int |F(r\mathrm{e}^{it})|^2 \mathrm{d}t$ — for the Gaussian Holomorphic Chaos, this is given by the Fyodorov-Bouchaud formula, a difficult result. It is not known if there is a similar formula in our Poisson case.

In general, studying the properties of F requires a good understanding of the properties of the coefficients c_n (called *secular coefficients*, see [28]). We saw that these coefficients are polynomials in the X_k (hence of the Y_ℓ), but this expression is difficult to manipulate; however, we have access to their moments by means of a combinatorial analysis. For every integer $k > 0$, we note Odd_k the set of nonempty subsets of $[k] = \{1, \dots, k\}$ with an odd number of elements, and Even_k the set of nonempty subsets of $[k] = \{1, \dots, k\}$ with an even number of elements.

Theorem 3.6. *For any $z_1, \dots, z_k$, one has*

$$\mathbf{E}[F(z_1) \cdots F(z_k)] = \frac{\prod_{S \in \mathrm{Odd}_k} (1 - d \prod_{s \in S} z_s)}{\prod_{S \in \mathrm{Even}_k} (1 - d \prod_{s \in S} z_s)}. \tag{11}$$

To give a few examples,

$$\mathbf{E}[F(z)] = 1 - zd$$
$$\mathbf{E}[F(y)F(z)] = \frac{(1-dy)(1-dz)}{1-dyz}$$
$$\mathbf{E}[F(x)F(y)F(z)] = \frac{(1-dx)(1-dy)(1-dz)(1-dxyz)}{(1-dxy)(1-dxz)(1-dyz)}.$$

The formula given above is our analog of the generating-function formula for the Gaussian Holomorphic Chaos in [47]. Therein, the combinatorial interpretation of the secular coefficients was easily linked with the enumeration of magic squares; we do not have a combinatorial interpretation for (11) at the moment.

4. Voronoi cells in random split trees

Cécile Mailler

This section is based on recent work with Alexander Drewitz (University of Cologne) and Markus Heydenreich (Ludwig Maximilian University of Munich), see [29].

4.1. **Introduction**

Consider a large graph $\mathcal{G}$, from which we choose k vertices uniformly at random, $U_1, \dots, U_k$. The *Voronoi cell* $\mathrm{Vor}(U_j)$ of U_j consists of those vertices that are closer in graph distance to U_j than to any of the other chosen vertices $\{U_i \colon i = 1, \dots, k; i \neq j\}$, with an arbitrary rule to break ties. We are studying the vector of proportional sizes

$$\left(\frac{|\mathrm{Vor}(U_1)|}{n}, \dots, \frac{|\mathrm{Vor}(U_k)|}{n} \right),$$

in the limit as $n \to \infty$, where $n = |\mathrm{Vor}(U_1)| + \cdots + |\mathrm{Vor}(U_k)|$ denotes the total number of vertices.

In [1], Addario-Berry, Angel, Chapuy, Fusy, and Goldschmidt prove that, if $\mathcal{G}$ is a tree taken uniformly at random among all n-node trees, then, in distribution as $n \to +\infty$,

$$\left(\frac{|\mathrm{Vor}(U_1)|}{n}, \dots, \frac{|\mathrm{Vor}(U_k)|}{n} \right) \to \mathrm{Dirichlet}(1, \dots, 1). \tag{12}$$

In other words, the vector of renormalised Voronoi cell sizes converges to the uniform distribution on the simplex. Addario-Berry et al. prove in fact a stronger result since it applies to any sequence of random trees whose scaling limit is Aldous' continuous random tree. Guitter [36] proved that the same limiting theorem holds in the case when $\mathcal{G}$ is a random planar map of genus 0 and $k = 2$. Both Addario-Berry et al's and Guitter's works were motivated by the conjecture of Chapuy [19] that (12) holds for *all random embedded graphs of fixed genus.*

In [29], we look at the distribution of the Voronoi cells of k uniform nodes in a random *split tree.* Split trees are a family of rooted trees introduced by Devroye [27] and later extended by Janson [41] who allowed trees of unbounded degrees: this family includes classical random trees such as the binary search tree, the random recursive tree, the preferential attachment tree (also called PORT for "plane oriented recursive tree"). In our main result, we prove that the largest of the Voronoi cells of k uniform nodes in an n-node split tree contains a proportion 1 of all nodes. We are also able to prove that the second, third, ..., k-th largest Voronoi cells each contains an order $n \exp(-\mathrm{const}\sqrt{\log n})$ of all vertices. In [29], we also show that this result holds when edges of the tree are given random i.i.d. lengths (of finite variance, or heavy-tailed but with finite mean), and defining the Voronoi cells with respect to the distance induced by these edge-lengths instead of the graph distance; we do not cover this case here.

The results of [29] are in contrast with the findings of [1] for the uniform random tree equipped with the graph distance: the distribution of the sizes of Voronoi cells is balanced in the case of the uniform random tree

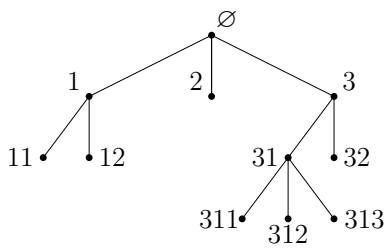

FIGURE 4. The 3-ary tree $\{\varnothing, 1, 2, 3, 11, 12, 31, 32, 311, 312, 313\}$. Node 312 is the "second child of the first child of the third child of the root", its parent is node 31, its siblings are 311 and 313. The last common ancestor of 32 and 313 is 3.

(and other trees whose scaling limit is the CRT), while we show a "winner takes it all" behaviour in the case of split trees. This difference in behaviour should not be surprising since it is well-known that split trees have a very different shape from the uniform random trees (and other random trees whose scaling limit is the CRT). Similarly to [1] conjecturing that their result generalises to maps that scale to the random Brownian map, one might expect that the behaviour we prove for random split trees might also be exhibited by other graphs such as preferential attachment graphs and other scale-free models such as the configuration model. However, our proofs cannot be straightforwardly generalised.

4.2. **Trees and random split trees: definitions**

In this section, we use the Ulam-Harris definition of m-ary trees: let $m \in \mathbb{N}$ and

$$\mathcal{D}_m = \{1, 2, \dots, m\}^* = \{\varnothing, 1, 2, \dots, m, 11, 12, \dots 1m, \dots\},$$

be the set of all finite words on the alphabet $\{1, 2, \dots, m\}$. We further consider the case of infinitary trees, where $m = \infty$ and $\mathcal{D}_\infty = \mathbb{N}^*$. We henceforth formulate our results for finite and infinite m in a unified fashion (unless stated explicitly); finite tuples, such as in (13) below, should be interpreted as infinite sequences whenever $m = \infty$.

Definition 4.1. *An m-ary tree is a subset t of $\mathcal{D}_m$ such that for all $w = w_1 \cdots w_\ell \in t$, all the prefixes of w are in t, i.e. for all $i \in \{0, \dots, \ell\}$ one has $w_1 \cdots w_i \in t$. (See Figure 4 for an example of a 3-ary tree.)*

We now define a probability distribution on the set of m-ary trees: it is the distribution of "split trees" first introduced by Devroye [27], but generalised to possibly infinite arity as in [41]. Let ν be a probability distribution on the set

$$\Sigma_m = \{(v_1, \dots, v_m) \in [0,1]^m \colon \sum_{i=1}^{m} v_i = 1\}, \tag{13}$$

and $(\boldsymbol{Y}(w))_{w \in \mathcal{D}_m}$ be a family of i.i.d. ν-distributed random vectors. For each node $w = w_1 \cdots w_\ell \in \mathcal{D}_m$, we let $Z_w = Y_{w_\ell}(\overleftarrow{w})$, where $\overleftarrow{w}$ is the parent of w, i.e. $\overleftarrow{w} = w_1 \cdots w_{\ell-1}$ and with $Y_{w_\ell}(\overleftarrow{w})$ denoting the w_ℓ-th coordinate of the vector $\boldsymbol{Y}(\overleftarrow{w})$ (see Figure 5 for an example: $\boldsymbol{Y}(3) = (.1, .4, .5)$ and thus $Z_{32} = .4$).

We also let $(X_n)_{n \geq 0}$ be a sequence of i.i.d. random variables uniformly distributed on $[0, 1]$, and independent from the sequence $(\boldsymbol{Y}(w))_{w \in \mathcal{D}_m}$.

Finally, given a tree t, we denote by ∂t the nodes of $\mathcal{D}_m$ that are not in t but whose parent is in t, and we call the elements of this set the "leaves" of t. It is not hard to see that if t has n nodes, then ∂t has cardinality $(m-1)n + 1$ (see Figure 5).

We can now define the sequence $(\tau_n)_{n \geq 1}$ of random trees recursively as follows.

- the tree τ_1 is defined to consist of the root only, i.e. $\tau_1 = \{\varnothing\}$.
- for $n \geq 1$, given τ_n, we define τ_{n+1} as the tree obtained by adding one node to τ_n as follows:

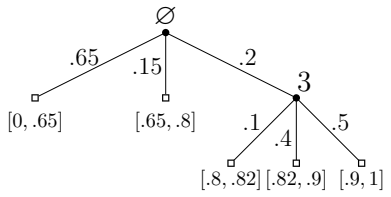

FIGURE 5. A realisation of the 3-ary split tree τ_2, here we have $\tau_2 = \{\varnothing, 3\}$. The labels on the edges represent the values of $(\boldsymbol{Y}(w))_{w\in\tau_2}$: for example, $\boldsymbol{Y}(\varnothing) = (.65, .15, .2)$. The value of Z_w is thus the label on the edge from w to its parent: for example, $Z_{31} = .1$. The nodes that are marked by a square are the elements of $\partial\tau_2$, underneath each leaf is written the corresponding part in the partition used to build τ_3. For example, the part corresponding to 32 is of length $Z_3 Z_{32} = .2 \times .4 = .08$.

- We subdivide the interval $[0, 1]$ in subintervals indexed by $\partial\tau_n$ of respective lengths $\prod_{\varnothing\neq v\preccurlyeq w} Z_v$, for all $w \in \partial\tau_n$. (Note that, by definition, $\sum_{w\in\partial\tau_n} \prod_{\varnothing\neq v\preccurlyeq w} Z_v = 1$; see Figure 5 for an example, and observe that some points form part of several intervals.)
- We set $\xi(n+1) = w$ if $X_{n+1} \in [0, 1]$ belongs to the part indexed by w of this partition of $[0, 1]$, and finally set $\tau_{n+1} = \tau_n \cup \{\xi(n+1)\}$; note that this is well defined almost surely.

The sequence of random trees $(\tau_n)_{n\geq 1}$ is called the *random split tree of split distribution* ν (which we recall is the distribution of the $\boldsymbol{Y}(w)$'s).

This definition incorporates a variety of different random trees that are classical in the literature:

- If $m = 2$ and ν is the distribution of $(Y, 1-Y)$, where Y is uniform on $[0, 1]$, then $(\tau_n)_{n\geq 1}$ is the *random binary search tree* (see [27, Table 1]).
- If ν is the uniform distribution on the simplex Σ_m for m finite, then $(\tau_n)_{n\geq 1}$ is the *random m-ary increasing tree* (see [27, Table 1]).
- If $m = \infty$ and ν is GEM$(0, 1)$[1] on Σ_∞, then $(\tau_n)_{n\geq 1}$ is the *random recursive tree* (see [41, Cor. 1.2]).
- If $m = \infty$ and ν is GEM$(1/2, 1/2)$, then $(\tau_n)_{n\geq 1}$ is the *random preferential attachment tree* (see [41, Cor. 1.3]).

4.3. **Voronoi cells: definition**

In [29], our aim is to investigate the sizes of the Voronoi cells corresponding to k nodes taken uniformly at random in the n-node random split tree τ_n defined in Subsection 4.2. We consider the graph distance in the graph whose nodes are all elements of $\mathcal{D}_m$, and where there is an edge between two nodes if and only if one is the parent of the other.

Definition 4.2. *Let $u_1, \ldots, u_k$ be k nodes in an m-ary tree t. We define the Voronoi cells of $u_1, \ldots, u_k$ as follows: for all $1 \leq i \leq k$,*

$$\mathrm{Vor}_t^i(u_1, \ldots, u_k) = \{w \in t\colon d(w, u_i) \leq d(w, u_j) \text{ for } j = 1, \ldots, i-1 \text{ and } d(w, u_i) < d(w, u_j) \text{ for } j = i+1, \ldots, n\}.$$

We say that $\mathrm{Vor}_t^i(u_1, \ldots, u_k)$ is the Voronoi cell of u_i (with respect to $u_1, \ldots, u_k$).

[1] The GEM(α, θ) is the distribution of the random infinite vector $(P_i)_{i\geq 1}$, where $P_i = Z_i \prod_{j=1}^{i-1}(1 - Z_j)$ and Z_j are independent with distribution $Z_j \sim \mathrm{Beta}(1-\alpha, \theta + j\alpha)$, see e.g. [41, Section 2].

Remark 4.3. *The idea of Definition 4.2 is that* $\mathrm{Vor}_t^i(u_1,\dots,u_k)$ *contains all the nodes that are closer to* u_i *than to any of the other* u_j*'s for the graph distance on* t*. The difference between '*$<$*' and '*$\leq$*' induces a simple rule to break ties (in case of equal distances, the vertex with smaller index is preferred). However, since the number of boundary vertices is of constant order, the choice we make about how to break ties has no impact on our results.*

4.4. **Main result**

Our main result holds under the following hypothesis on the split-vector distribution ν:

(A) If $(Y_1,\dots,Y_m)\sim\nu$, U is a uniform random variable on $[0,1]$, and[2]

$$\bar{Y}=\sum_{i=1}^{m} Y_i \mathbf{1}\{\sum_{j=1}^{i-1} Y_j \leq U < \sum_{j=1}^{i} Y_j\} \tag{14}$$

is the size-biased version of the marginals of ν, then $\mu:=\mathbb{E}[\log 1/\bar{Y}]>0$ and $\sigma^2:=\mathrm{Var}(\log \bar{Y})<+\infty$.

The assumption that $\mu>0$ excludes the trivial case when the n-node split tree is almost surely equal to a line of n nodes hanging under each other under the root. The assumption that $\sigma^2<+\infty$ gives some control over the moments of the split vectors: this assumption is used in the proof when applying laws of large numbers and of the iterated logarithm, as well as central limit theorems to sum of independent copies of $\log \bar{Y}$.

Theorem 4.4. *Let* ν *be a probability distribution on* Σ_m*, and* $(\tau_n)_{n\geq 1}$ *be the random split tree of split distribution* ν*.*

For each $n\geq 1$*, let* $U_1(n),\dots,U_k(n)$ *be* k *nodes taken uniformly at random among the* n *nodes of* τ_n*; we let* $V_{(1)}(n)\geq\ldots\geq V_{(k)}(n)$ *be the sizes of their Voronoi cells in* τ_n *with respect to the graph distance, ordered in decreasing order.*

Under Assumption (A)*, in distribution when* $n\to+\infty$*,*

$$\frac{1}{\sqrt{\log n}}(\log(V_{(2)}(n)/n),\dots,\log(V_{(k)}(n)/n)) \Rightarrow \frac{1}{2\sqrt{\mu}}(\Psi_{(1)}-\Psi_{(2)},\dots,\Psi_{(1)}-\Psi_{(k)}), \tag{15}$$

where $\Psi_{(1)}\leq\cdots\leq\Psi_{(k)}$ *is the order statistics of* k *i.i.d. random variables whose distribution is* $\mathcal{N}(0,\sigma^2)$*.*

In words, the above amounts to the fact that the second, third, ..., kth largest component each occupies a proportion of roughly $\exp\{-\Psi\sqrt{\log n}\}$ of the vertices, where Ψ is some explicit positive random variable. This implies that asymptotically and in distribution, the entire mass is allocated to the largest component (which, by construction, belongs to the vertex closest to the root). The allocation for split trees is therefore qualitatively very different from the allocation in the universality class of the continuum random tree, where the limit of the proportions of the masses is known to be uniform [1].

4.5. **Ideas of the proofs**

To prove our main result, we prove two results that may be of independent interest because they give information of the typical shape of a random split tree:

(1) The profile of a random tree is the distribution of the height (distance to the root) of a node taken uniformly at random in the tree. If the tree is random then its profile is a random measure.

Proposition 4.5. *Let* $(\tau_n)_{n\geq 1}$ *be the random split tree of split distribution* ν*, and let, for all integer* n*,* $\pi_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{|\nu_i|}$ *be the random profile of* τ_n*, where we recall that* $|\nu_i|$ *is the height of the node inserted at time* i *in* $(\tau_n)_{n\geq 1}$*. If* ν *satisfies Assumption* (A1)*, then*

$$\pi_n(\ \cdot\ \sqrt{(\log n)/\mu^3}+(\log n)/\mu)\to\pi_\infty=\mathcal{N}(0,1), \tag{16}$$

[2]By convention, we set $\sum_{i=1}^{0}a_i=0$ for each sequence $(a_i)_{i\geq 0}$ of real numbers.

in probability as $n \to +\infty$, on the space of probability measures on $\mathbb{R}$ equipped with the topology of weak convergence.

(2) Fringe trees are subtrees that are rooted at an ancestor of a node taken uniformly at random in the tree (or at the uniform node itself). Oftentimes, this ancestor is chosen to be at constant distance of the uniform node (see, e.g. [40] and the references therein). In [29, Proposition 2.6], we extend this definition to allow the ancestor to be at distance to the uniform node that tends to infinity with n, the number of nodes in the whole split tree.

The authors are grateful to two anonymous referees for detailed comments improving the exposition of the paper.

References

[1] L. Addario-Berry, O. Angel, G. Chapuy, É. Fusy, and C. Goldschmidt. Voronoi tessellations in the CRT and continuum random maps of finite excess. *SODA 2018*, 933–946, 2018.
[2] G. W. Anderson. Convergence of the largest singular value of a polynomial in independent Wigner matrices. Ann. Probab., 41(3B):2103–2181, 2013.
[3] B. Au, G. Cébron, A. Dahlqvist, F. Gabriel, and C. Male. Freeness over the diagonal for large random matrices. Annals of Probability, 49 (1). pp. 157-179. 2021.
[4] Z. Bao and J. Xie. CLT for linear spectral statistics of Hermitian Wigner matrices with general moment conditions. Theory Probab. Appl., 60(2):187–206, 2016.
[5] Z.D. Bai and J.W. Silverstein. Ann. Probab. 32 (2004), no. 1A, 553–605.
[6] Z. D. Bai and J. Yao. On the convergence of the spectral empirical process of Wigner matrices. Bernoulli, 11(6):1059–1092, 2005.
[7] S. T. Belinschi, M. Capitaine, S. Dallaporta, M. Février. Fluctuations of the Stieltjes transform of the empirical spectral distribution of selfadjoint polynomials in Wigner and deterministic diagonal matrices. https://arxiv.org/abs/2107.10031.
[8] A. Basak, M. Rudelson, et al. The circular law for sparse non-Hermitian matrices. *Annals of Probability*, 47(4):2359–2416, 2019.
[9] A. Basak and O. Zeitouni. Outliers of random perturbations of Toeplitz matrices with finite symbols. *Probability Theory and Related Fields*, 178(3):771–826, 2020.
[10] S. T. Belinschi, T. Mai, and R. Speicher. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. J. Reine Angew. Math. 732 21–53, 2017.
[11] H. Bercovici and D. Voiculescu. Free convolution of measures with unbounded support, Indiana Univ. Math. J., 42, 733–773, 1993.
[12] P. Biane. On the free convolution with a semi-circular distribution. Indiana Univ. Math. J., 46(3), 705–718, 1997.
[13] P. Biane. Processes with free increments, Math. Z. 227, 143–174, 1998.
[14] P. Billingsley. (1995). Probability and Measure, 3rd ed. Wiley, New York. MR1324786.
[15] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.
[16] C. Bordenave and D. Chafaï. Around the circular law. *Probability surveys*, 9:1–89, 2012.
[17] C. Bordenave, S. Coste, and R.R. Nadakuditi. Detection thresholds in very sparse matrix completion. *arXiv preprint arXiv:2005.06062*, 2020.
[18] M. Capitaine and M. Casalis. Geometric interpretation of the cumulants for random matrices previously defined as convolutions on the symmetric group. Lecture Notes in Mathematics, Séminaire de Probabilités XLI, 1934:93–119, 2008.
[19] G. Chapuy. On tessellations of random maps and the t_g-recurrence. *Probability Theory and Related Fields*, 174(1-2):477–500, 2019.
[20] B. Collins. Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability, Int. Math. Res. Not., no. 17, 953–982, 2003.
[21] S. Coste. Sparse matrices: convergence of the characteristic polynomial seen from infinity, `https://arxiv.org/abs/2106.00593`
[22] S. Coste. The spectral gap of sparse random digraphs. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 57, pages 644–684. Institut Henri Poincaré, 2021.
[23] S. Coste, G. Lambert, and Y. Zhu. The characteristic polynomial of sums of random permutations and regular digraphs, 2022.
[24] S. Coste and L. Stephan. A simpler spectral approach for clustering in directed networks, 2021.
[25] B. Collins and P. Sniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. Com. in Math. Phys., 264(3):773–795, 2004.
[26] S. Dallaporta and M. Février. Fluctuations of linear spectral statistics of deformed Wigner matrices. arXiv:1903.11324, 2019.
[27] L. Devroye. Universal limit laws for depths in random trees. *SIAM Journal on Computing*, 28(2):409–432, 1998.

[28] P. Diaconis and A. Gamburd. Random matrices, magic squares and matching polynomials. *Electron. J. Combin.*, 11(2):Research Paper 2, 26, 2004/06.
[29] A. Drewitz, M. Heydenreich, and C. Mailler. Voronoi cells in random split trees. `https://arxiv.org/abs/2103.09784`.
[30] N. Dunford and J.T. Schwartz. Linear Operators II: Spectral Theory, Interscience, 1963.
[31] K. Dykema. On certain free product factors via an extended matrix model. J. Funct. Anal., 112(1):31–60, 1993.
[32] Joel Friedman. *A proof of Alon's second eigenvalue conjecture and related problems.* American Mathematical Soc., 2008.
[33] F. Gabriel. Combinatorial theory of permutation-invariant random matrices I: partitions, geometry and renormalization. arXiv:1503.02792, 2015.
[34] F. Gabriel. Combinatorial theory of permutation-invariant random matrices II: cumulants, freeness and Lévy processes. arXiv:1510.01046, 2015.
[35] V. L. Girko, Spectral theory of random matrices (Nauka, Moscow, 1988 (in Russian)).
[36] E. Guitter. On a conjecture by Chapuy about Voronoi cells in large maps. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(10):103401, 2017.
[37] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. Ann. of Math. (2), 162(2):711–775, 2005.
[38] A. Jacot, B. Simsek, F. Spadaro, C. Hongler, and F. Gabriel. Implicit Regularization of Random Feature Models. ICML, 2020.
[39] A. Jacot, B. Simsek, F. Spadaro, C. Hongler, and F. Gabriel. Kernel Alignment Risk Estimator: Risk Prediction from Training Data. NeurIPS, 2020.
[40] C. Holmgren and S. Janson. Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees. *Probability Surveys*, 14:53–154, 2017.
[41] Svante Janson. Random recursive trees and preferential attachment trees are random split trees. *Combinatorics, Probability and Computing*, 28(1):81–99, 2019.
[42] H. C. Ji and Lee, O. Ji. Gaussian fluctuations for linear spectral statistics of deformed Wigner matrices. Random Matrices Theory Appl. 9 (2020), no. 3, 2050011, 79 pp.
[43] H. Maassen. Addition of freely independent random variables. J. Funct. Anal., 106(2), 409–438, 1992.
[44] A. M. Khorunzhy, B. A. Khoruzhenko, and L. A. Pastur. Asymptotic properties of large random matrices with independent entries. J. Math. Phys., 37(10):5033–5060, 1996.
[45] C. Male. Traffic distributions and independence: permutation invariant random matrices and the three notions of independence. Memoirs of the American Mathematical Society, 2021.
[46] J. A. Mingo and R. Speicher. Free Probability and Random Matrices, Fields Institute Communications 35, Springer Nature, 2017.
[47] J. Najnudel, E. Paquette, and N. Simm. Secular coefficients and the holomorphic multiplicative chaos, 2020.
[48] A. Nica, D. Shlyakhtenko, and R. Speicher. Operator-Valued Distributions. I. Characterizations of Freeness. Internat. Math. Res. Notices, 29, 1509–1538, 2002.
[49] A. Nica and R. Speicher. Lectures on the Combinatorics of Free Probability, Cambridge Univ. Press, 2006.
[50] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In Advances in neural information processing systems, pp. 1177–1184, 2008.
[51] B. Rider and B. Virág. The noise in the circular law and the Gaussian free field. *International Mathematics Research Notices*, 2007, 2007.
[52] M. Rudelson and K. Tikhomirov. The sparse circular law under minimal assumptions. *Geometric and Functional Analysis*, 29(2):561–637, 2019.
[53] Ø. Ryan. On the limit distributions of random matrices with independent or free entries. Commun. Math. Phys., 193(3):595–626, 1998.
[54] M. Shcherbina. Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. Zh. Mat. Fiz. Anal. Geom., 7(2):176–192, 197, 199, 2011.
[55] T. Shirai. Limit theorems for random analytic functions and their zeros: Dedicated to the late professor Yasunori Okabe (functions in number theory and their probabilistic aspects). *RIMS Kokyuroku Bessatsu*, 34:335–359, 2012.
[56] D. Voiculescu. Addition of certain noncommuting random variables, J. Funct. Anal. 66, 323–346, 1986.
[57] D. Voiculescu. Limit laws for random matrices and free products, Invent. Math. 104, 201–220, 1991.
[58] D. Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory. I, Comm. Math. Phys. 155, 71–92, 1993.
[59] D. Voiculescu. Operations on certain non-commutative operator-valued random variables. Recent advances in operator algebras (Orléans, 1992). Astérisque (1995), no. 232, 243–275, 1995.
[60] D. Voiculescu. A strengthened asymptotic freeness result for random matrices with applications to free entropy, Internat. Math. Res. Notices, no. 1, 41–63, 1998.
[61] D. Voiculescu. The coalgebra of the free difference quotient and free probability, Internat. Math. Res. Notices, no. 2, 79–106, 2000.
[62] D. Voiculescu, K. J. Dykema, and A. Nica. Free random variables, American Mathematical Society, Providence, RI, 1992.