Extreme matrices or how an exponential map links classical and free extreme laws

Jacek Grela* and Maciej A. Nowak†
Institute of Theoretical Physics and Mark Kac Complex Systems Research Centre, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland

(Received 27 March 2020; accepted 3 July 2020; published 10 August 2020)

Using our proposed approach to describe extreme matrices, we find an explicit *exponentiation* formula linking the classical extreme laws of Fréchet, Gumbel, and Weibull given by the Fisher-Tippet-Gnedenko classification and free extreme laws of free Fréchet, free Gumbel, and free Weibull of Ben Arous and Voiculescu. We also develop an extreme random matrix formalism, in which refined questions about extreme matrices can be answered. In particular, we demonstrate explicit calculations for several more or less known random matrix ensembles, providing examples of all three free extreme laws. Finally, we present an exact mapping, showing the equivalence of free extreme laws to the Peak-over-Threshold method in classical probability.

DOI: 10.1103/PhysRevE.102.022109

I. INTRODUCTION

Extreme value theory in classical probability is the prominent application of probability calculus for several problems seeking extreme values for a large number of random events. Its power comes from universality according to the Fisher-Tippett-Gnedenko classification [1] which permits only three statistical laws of extremes: the Gumbel distribution, the Fréchet distribution, and the Weibull distribution. Beyond applications of extreme value theory in physics in the theory of disordered systems [2], seminal applications include insurance, finances, hydrology, neuroscience, biology, computer science, and several others [3–7].

Since the seminal work of Ref. [8], random matrix theory has become one of the most universal probabilistic tools in physics and in several multidisciplinary applications [9]. In the limit when the size of the matrix tends to infinity, random matrix theory bridges to free probability theory, which can be viewed as an operator valued (i.e., noncommutative) analog of the classical theory of probability [10,11]. Both calculi exhibit striking similarities. Wigner's semicircle law can be viewed as an analog of normal distribution, Marçenko-Pastur spectral distribution for Wishart matrices is an analog of Poisson distribution in classical probability calculus, and Bercovici-Pata bijection [12] is an analog of Lévy stable processes classification for heavy-tailed distributions. It is therefore tempting to ask the question: How far can we extend the analogies between these two formalisms?

In particular, do we have an analog of extreme values limiting distributions for the spectra of very large random matrices, i.e., does the Fisher-Tippett-Gnedenko classification exist in free probability? The positive answer to this crucial question was provided more then a decade ago by Ben Arous and Voiculescu [13]. Using operator techniques, they proved that free probability theory also has three limiting extreme distributions: the free Gumbel, the free Fréchet, and the free Weibull distribution. The functional form of these limiting distributions differs from the classical probability case, but, surprisingly, the domains of attraction are the same as in their classical counterparts. The authors of Ref. [13] state several properties of newly found free extreme laws like their representation in terms of certain generalized Pareto distribution or relation with a Balkema–de Haan–Pickands [14,15] classification in the classical probability theory of exceedances.

Despite various connections between classical and free calculi, an explicit link between extreme laws was lacking. In this paper we establish an *exponentiation* formula between laws by comparing and contrasting existing our proposed approaches to free extreme laws.

A. Main results

First, we propose two approaches to study extremes: the *thinning* method, having its root in classical extreme value theory, and an extreme random matrices scheme based on random matrix theory. With two previously studied frameworks due to Ben-Arous and Voiculescu (new based on free probability and old Peak-over-Threshold statistics), we carefully establish interrelations between them summarized in Fig. 1. Whereas for the first three approaches equivalency is straightforward, explaining the relation with the Peak-over-Threshold method is both unique and nontrivial. In the end, despite some specialization of each approach, all considered frameworks are equivalent, i.e., respective cumulative distribution functions (CDFs) agree.

Second, we describe in brief the merits of the two approaches. The *thinning* approach, in contrast the to free-probabilistic method, both is intuitive and encompasses classical and free extreme events. On the other hand, extreme random matrices present a framework to which the matrix aspect of objects is accessible and presents its applicability in Fig. 3 below when the number $r \gg N$ of extremized matrices is much larger than their sizes N.

*jacekgrela@gmail.com
†maciej.a.nowak@uj.edu.pl

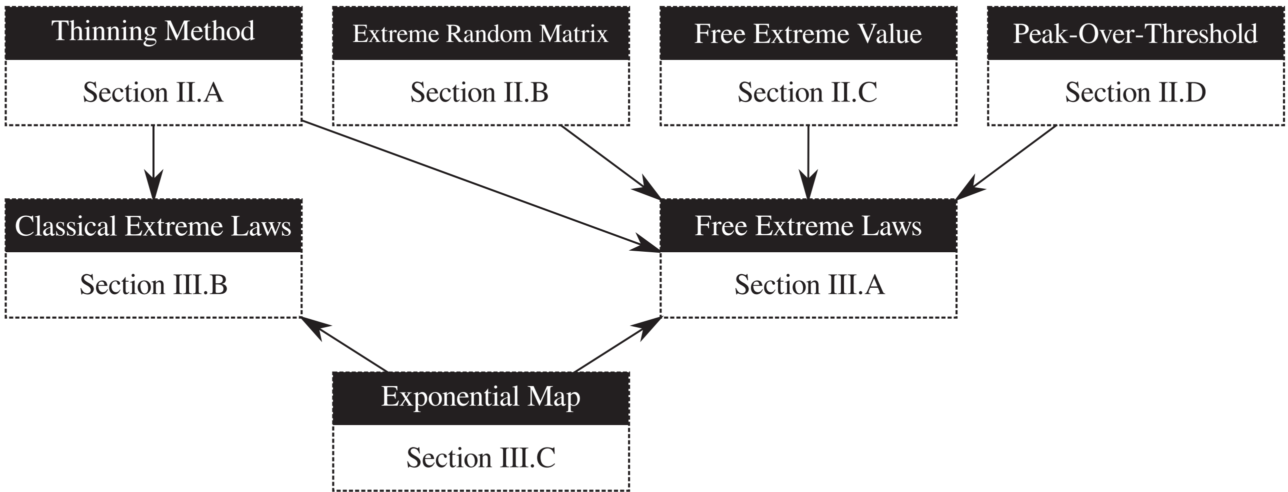

FIG. 1. Flow diagram of the discussed frameworks in relation to families of extreme laws.

The main result (5) in the thinning approach is the formula for a CDF of a large maximum matrix (understood as distribution of its eigenvalues) obtained by extremizing r large matrices:

$$\mathbf{F}_r(x) = r[f(x) - \alpha_r]\theta[f(x) - \alpha_r],$$

where f is the CDF of a single large matrix, parameter $\alpha_r = \frac{r-1}{r}$, and θ is a Heaviside step function. The result is simply a *truncated* single matrix CDF f; see Fig. 2 for an example of the semicircle law.

In the extreme random matrix framework, the CDF of picking the *largest* out of r matrices is given by Eq. (10) where now, however, the matrices are of finite size N:

$$F_{N,r}(x) = \frac{1}{N}\sum_{n=0}^{N-1}(N-n)\sum_{\substack{j_1\cdots j_r=0\\ j_1+\cdots+j_r=n}}^{n}\prod_{l=1}^{r}E_N(j_l;x).$$

The gap functions $E_N(j;x)$ are probability functions that the jth largest eigenvalues are greater than x while $N-j$ are smaller than x. When x is far from critical points like edges of the spectrum, two CDFs agree: $\lim_{N\to\infty} F_{N,r}(x) = \mathbf{F}_r(x)$.

Third, based on the thinning method encompassing both classical and free worlds, we find an *exponential* map (33) relating CDFs of classical extreme laws $F^{\rm class}(x)$ with free extreme laws $F^{\rm free}(x)$:

$$F^{\rm class}(x) = t(x)\exp\left[\frac{F^{\rm free}(x)}{T(x)} - 1\right],$$

where $t(x) = \theta(x), 1, 1$ and $T(x) = \theta(x-1), \theta(x), \theta(x+1)$ are step functions for Fréchet, Gumbel, and Weibull classes, respectively. The step function in the denominator is understood formally so as to cancel out the corresponding term in the free CDF $F^{\rm free}$. This proves the one-to-one correspondence between extreme laws in both probability calculi [i.e., in classical and in matricial (free) ones].

Last, due to the operational and calculational simplicity of the thinning approach, we give several explicit examples of free extreme laws for several random matrix ensembles. As most spectral densities occurring in random matrix theory

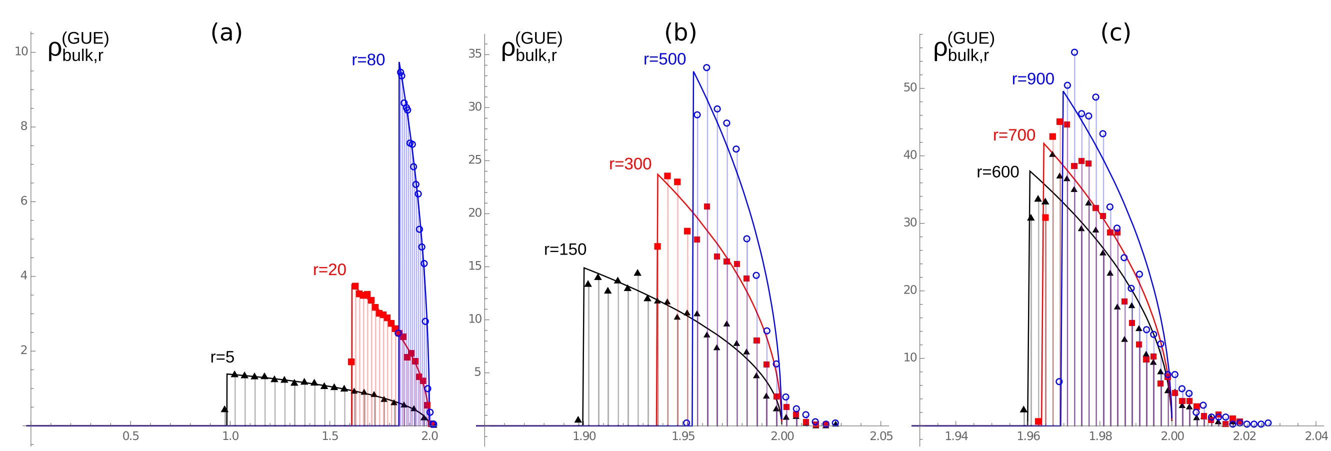

FIG. 2. Analytical (lines) and numerical (points) density of eigenvalues $\rho^{\rm (GUE)}_{{\rm bulk},r}(x) = \frac{d}{dx}F^{\rm (GUE)}_{{\rm bulk},r}(x)$ given by Eqs. (23) and (24) where we have plotted r matrices of size 500×500 each drawn from an GUE defined by jPDF (21). The change of r parameter from small [(a), $r \ll N$] through medium [(b), $r \sim N$] to plot (c) for large $r > N$ shows a transition where tail-like spectral features leaks out beyond the bulk boundary at $x = 2$.

have finite supports, the free Weibull category is most numerous. The free Fréchet class is represented by spectral densities found through Bercovici-Pata construction being analogues of Lévy heavy-tailed distributions. Finally, we report in detail on one, quite exotic example (the so-called free Gaussian distribution) found to lie within the free Gumbel class, and we comment on the link of such a distribution to plasma physics.

II. APPROACHES TO EXTREMES

In this section we describe the status of all four approaches to extreme events and compare existing one and those proposed by us. The authors of Ref. [13] have defined free extreme values using free probability and also noticed an unexpected connection to Peak-over-Threshold statistics. In this work we introduce two distinct but related frameworks based on random matrix theory and classical extreme value theory: the thinning method and extreme random matrices. All four approaches with their distinctive features and the notation we use are summarized below:

Thinning method (using statistics and extreme value theory)

(1) Works for asymptotically large matrices
(2) Connects classical and free extreme laws
(3) Applicable not only to eigenvalues

The central object is the CDF $\mathbf{F}_{m,n}$ and its asymptotic form $\mathbf{F}_r$ given by Eqs. (4) and (5), respectively.

Extreme random matrices (using random matrix theory)

(1) Works for finite matrices
(2) Can address questions beyond bulk

The central object is the CDF $F_{N,r}$ and its asymptotic form F_r given by Eqs. (10) and (14), respectively.

Free extreme values (using free probability)

(1) Formulated in operator language
(2) Not limited to matrices

The central object is the CDF $F_{H^{\vee r}}$ given by Eq. (19).

Peak-over-Threshold method (using statistics and extreme value theory)

The central object is the probability $\mathcal{P}_{\rm POT}$ given by Eq. (20).

We will describe mostly the first two approaches and later compare them with the third and fourth. For ease of presentation we try to clearly delineate each approach as our goal is to highlight connections (and to large extent equivalency) between the arising extreme laws. To this end, in all three descriptions we focus on a common quantity of the CDF.

First, we describe an approach based on statistics and extreme value theory where we focus on order statistics. We study the CDF of a fraction of the largest i.i.d. random variables, which we name *the thinning procedure*. This approach is applicable to any random variables, not only related to random matrices. In particular, it does degenerate to the usual order statistics.

Second, we focus on a second approach based on the random matrix perspective where we define en extreme CDF $F_{N,r}$ as an average over the joint probability distribution function of the underlying ensemble of r random matrices. In particular, we obtain general asymptotic results (as the matrix size N goes to infinity) valid both in the bulk of the spectrum and near the (soft) spectral edge. We also investigate the limit where the number of matrices r goes to infinity and a certain type of double-extreme distribution emerges.

The third point of view is based on Refs. [13,16] and has its source in free probability. We present a result of Ref. [17] for the eigenvalue CDF of maximizing r Hermitian random matrices. This framework can be generalized to a general operator language.

Last, we describe a Peak-over-Threshold, method which focuses on statistical study of random events only upon exceeding a certain threshold. It is somewhat unrelated to the previous ones as it is not concerned with extreme matrices per se. However, the excess distribution function $\mathcal{P}_{POT}$ is related to the CDF studied in the first three approaches, and we explain this connection.

A. Thinning method

Extremes of random numbers are described by *order statistics*. Given a set of m random variables $\{x_1, \dots, x_m\}$, we rearrange them in an descending order $\{x_{(1)}, \dots, x_{(m)}\}$. As an example, for a set of such variables the following inequalities hold true:

$$\begin{array}{ccccccccc} x_{(1)} & \geqslant & x_{(2)} & \geqslant & \cdots & \geqslant & x_{(m-1)} & \geqslant & x_{(m)}, \\ x_3 & \geqslant & x_{m-1} & \geqslant & \cdots & \geqslant & x_6 & \geqslant & x_2. \end{array}$$

Typically one is interested in the extreme events and studies a particular element in the ordered set $\{x_{(1)}, \dots, x_{(m)}\}$—either the largest $x_{(1)}$ or the smallest one $x_{(m)}$. One can study also the distributions of a subset of the ordered set: the n largest or smallest values.

In all these cases, the cumulative distribution function (or CDF) for the kth-order statistic $x_{(k)}$ of a sample of m variables is given by [18]

$$\mathcal{P}^{(m)}(x_{(k)} < x) = \sum_{i=0}^{k-1} \sum_{\{\sigma\}} \underbrace{\int_{-\infty}^{x} dx_{\sigma(1)} \cdots \int_{-\infty}^{x} dx_{\sigma(m-i)}}_{m-i} \times \sum_{\{\delta\}} \underbrace{\int_{x}^{\infty} dx_{\delta(1)} \cdots \int_{x}^{\infty} dx_{\delta(i)}}_{i} \mathbf{P}(x_1, \dots, x_m), \tag{1}$$

where $\sum_{\{\delta\}}$ is the summation over i combinations of rN indices and $\sum_{\{\sigma\}}$ is $rN - i$ combinations of the remaining $rN - i$ elements. This formula is easy to understand by positioning all particles on a line and considering only configurations where the particle with kth largest position is on the left of the barrier centered at x (the meaning of the condition $x_{(k)} < x$). There are k possible scenarios satisfying this condition—when the number of particles on the right side of the barrier varies from $k-1$ to 0 which results in the summation $\sum_{i=0}^{k-1}$. Each term in the sum describes one such eventuality; the i integrals $\int_x^\infty$ place particles on the right side of the barrier, whereas the rest $rN - i$ integrals $\int_{-\infty}^x$ position the remaining ones on the left side. The only additional thing we take into account is labeling the particles, which results in summation over all possible combinations σ, δ.

The joint probability density function (PDF) $\mathbf{P}$ describes any set of correlated or uncorrelated random variables. In particular, for i.i.d. uncorrelated variables $\mathbf{P}(x_1,\dots,x_m) = \prod_{i=1}^m p(x_i)$ we find $\int_{-\infty}^y p(t)\,dt = f(x)$ and $\int_x^\infty p(t)\,dt = 1 - f(x)$ which produces a well-known CDF for the kth largest order statistic:

$$\mathcal{P}^{(m)}(x_{(k)} < x) = \sum_{i=0}^{k-1} \binom{m}{i} [1 - f(x)]^i [f(x)]^{m-i}. \tag{2}$$

In the special case $k = 1$, the distribution function of the largest value for $k = 1$ is just $\mathcal{P}^{(m)}(x_{(1)} < x) = [f(x)]^m$. Now we turn to describing a thinning procedure which takes a finite fraction of largest variables.

A thinning procedure applied to order statistics is to consider the following problem: draw m i.i.d. variables $\{x_1,\dots,x_m\}$ from the parent PDF $p(x)$ and CDF $f(x)$, pick out the n largest ones $\{x_{(1)}\cdots x_{(n)}\}$, and look at their distribution. What will be the resulting PDF and CDF? We find the thinned CDF $\mathbf{F}_{m,n}$ of the n largest values selected out of m values as a normalized sum of the first n-order statistics given by Eq. (2):

$$\mathbf{F}_{m,n}(x) = \frac{1}{n} \sum_{k=1}^{n} \mathcal{P}^{(m)}(x_{(k)} < x). \tag{3}$$

In the i.i.d. case, we plug Eq. (2) into formula (3):

$$\mathbf{F}_{m,n}(x) = \frac{1}{n} \sum_{k=0}^{n-1} (n-k) \binom{m}{k} [1 - f(x)]^k [f(x)]^{m-k}. \tag{4}$$

Define the ratio $r = \frac{m}{n}$ so we take the $m, n \to \infty$ limit such that r remains fixed. An asymptotic form of the thinned CDF $\lim_{m,n\to\infty} \mathbf{F}_{m,n}(x) = \mathbf{F}_r(x)$ is found in Sec. A of the the Supplemental Material [19]:

$$\mathbf{F}_r(x) = r[f(x) - \alpha_r]\theta[f(x) - \alpha_r], \tag{5}$$

where $\alpha_r = \frac{r-1}{r}$ and θ is a step function. It gives a CDF of a thinned population where from m random elements we pick the $n < m$ largest ones with ratio $r = m/n$. The above definition should *not* be confused with the thinning in classical extreme statistics where a subsample is picked out according to a probabilistic criterion [20].

Interpretation of the asymptotic thinned CDF $\mathbf{F}_r$ is clear: picking n largest values out of m does not modify the shape of the parent distribution $f(x)$ but truncates it up to a point x_* such that $f(x_*) = \alpha_r$. The point x_* is known in statistics as the last of the r-quantile and gives the point where the fraction of values smaller than x_* is $\alpha_r = \frac{r-1}{r}$. Importantly, since the large n, m limit was taken, the fraction α_r takes all real numbers between (0,1).

We stress that above discussion is purposely not restricted to matrix eigenvalues because it is applicable to general random variables. In particular, above we have addressed the simplest case of the completely uncorrelated case where a joint PDF (jPDF) factorizes, which still has applications to matrices, as we will see later. Besides that, in the next section we deal with another important class of jPDFs with matrix eigenvalues as coupled or correlated random variables arising within random matrix theory.

B. Extreme random matrices

Unlike real numbers, finding extremes in the space of matrices cannot be done easily due to lack of natural ordering. To circumvent that we instead define ordering in the space of *random* matrices. This is due to an existing natural identification between the random matrix and its eigenvalues—in almost all matrix PDFs the eigenvectors completely decouple. Thus, we can disregard them completely and define extreme matrices based on eigenvalues alone. The procedure is straightforward:

(1) Take r random matrices each of size $N \times N$, collect all Nr eigenvalues, and

(2) Pick out N largest ones; these form the largest or extreme random matrix representation.

The main drawback of this definition is that the resulting extreme random matrix will contain a mixture of eigenvalues from several initial matrices.

We now introduce some useful notation and then continue describing the above approach mathematically. First, denote $\{\lambda_1^{(i)}\cdots\lambda_N^{(i)}\}$ to be the set of eigenvalues of the ith matrix drawn from the most general jPDF for the ith individual matrix:

$$P_N^{(i)}\big(\lambda_1^{(i)},\dots,\lambda_N^{(i)}\big).$$

In general, these distribution functions could differ between matrices; however, in what follows we consider an i.i.d. case. All Nr eigenvalues are ordered in the following way:

$$\begin{array}{ccccccccc} \lambda_1^{(1)} & \cdots & \lambda_N^{(1)} & \lambda_1^{(2)} & \cdots & \lambda_1^{(r)} & \cdots & \lambda_N^{(r)} \\ \downarrow & \cdots & \downarrow & \downarrow & \cdots & \downarrow & \cdots & \downarrow \\ x_1 & \cdots & x_N & x_{N+1} & \cdots & x_{(r-1)N+1} & \cdots & x_{rN} \end{array} \tag{6}$$

and so the total jPDF is a product of single-matrix distributions:

$$\mathbf{P}(x_1,\dots,x_{rN}) = \prod_{i=1}^{r} P_N^{(i)}(x_{(i-1)N+1},\dots,x_{iN}). \tag{7}$$

Last, we rearrange all variables $\{x_{(1)},\dots,x_{(rN)}\}$ in descending order $x_{(1)} > x_{(2)} > \cdots > x_{(rN)}$ so that we pick out only the N largest ones, i.e., $x_{(1)},\dots,x_{(N)}$.

It is important to note how such rearrangements cast our current matrix problem into a thinning framework introduced before in Sec. II A with a special form of a correlated jPDF given by Eq. (7) and substituting $n \to N$, $m \to rN$.

1. Extreme CDF $F_{N,r}$

The CDF for the kth-order statistic $\mathcal{P}_r(x_{(k)} < x) = \langle\theta(x - x_{(k)})\rangle$ is given by Eq. (1) with $m \to Nr$:

$$\mathcal{P}_{N,r}(x_{(k)} < x) = \sum_{i=0}^{k-1} \sum_{\{\sigma\}} \underbrace{\int_{-\infty}^{x} dx_{\sigma(1)} \cdots \int_{-\infty}^{x} dx_{\sigma(rN-i)}}_{rN-i} \times \sum_{\{\delta\}} \underbrace{\int_{x}^{\infty} dx_{\delta(1)} \cdots \int_{x}^{\infty} dx_{\delta(i)}}_{i} \mathbf{P}(x_1,\dots,x_{rN}), \tag{8}$$

where $\sum_{\{\delta\}}$ is the summation over i combinations of rN indices and $\sum_{\{\sigma\}}$ is $rN-i$ combinations of the remaining $rN-i$ elements. Its validity was explained in Sec. II A.

Since we study the density of the N largest eigenvalues, we define an extreme CDF as a normalized sum of terms (8) in analogy with Eq. (3):

$$F_{N,r}(x) = \frac{1}{N}\sum_{k=1}^{N}\mathcal{P}_{N,r}(x_{(k)} < x). \quad (9)$$

Next we make use of symmetries in the eigenvalue jPDFs and find the following form (see details in the Supplemental Material [19]):

$$F_{N,r}(x) = \frac{1}{N}\sum_{n=0}^{N-1}(N-n)\sum_{\substack{j_1\cdots j_r=0\\ j_1+\cdots+j_r=n}}^{n}\prod_{l=1}^{r}E_N(j_l;x), \quad (10)$$

where the kth gap function $E_N(k;x)$ of finding exactly k (of N) particles in an interval $(x,+\infty)$ is defined as

$$E_N(k;x) = \binom{N}{k}\int_x^{\infty}d\lambda_1\cdots d\lambda_k\int_{-\infty}^{x}d\lambda_{k+1}\cdots d\lambda_N P_N, \quad (11)$$

with an eigenvalue jPDF $P_N(\lambda_1\cdots\lambda_N)$. The formula (10) is already expressed entirely in terms of well-known objects in random matrix theory. It is exact for any value of both N and r, though explicit forms of gap functions are not known, and we consider several important cases.

For $r=1$, Eq. (10) reduces to $F_{N,1}(x) = \frac{1}{N}\sum_{n=0}^{N-1}(N-n)E_N(n;x)$, and in Sec. B of the Supplemental Material [19] we show how it is in turn given in terms of spectral density $\rho_N^{(1)}$ (or the one-point correlation function):

$$F_{N,1}(x) = \frac{1}{N}\int_{-\infty}^{x}\rho_N^{(1)}(y)\,dy, \quad (12)$$

which simply means that $F_{N,1}$ is the spectral CDF. It is hardly surprising since inspecting N out of the N largest eigenvalues should reduce exactly to quantities related with spectral density itself.

2. Bulk and edge limiting forms of $F_{N,r}$

We turn to describe various limiting forms of the extreme CDF $F_{N,r}(x)$ given by Eq. (10). We address mostly cases when the argument x is far from the edge of the matrix spectrum (the bulk regime), when x becomes close to the spectral edge (soft edge regime), and last we comment on the double scaling limit when $r\sim N$.

a. In the bulk. We first evaluate $F_{N,r}$ in the bulk. To this end, in Sec. A of the Supplemental Material [19], we calculate the asymptotic form of the gap function $E_{\rm bulk}(k;x) = e^{-N[1-f(x)]}\frac{\{N[1-f(x)]\}^k}{k!}$ and plug it into Eq. (10) so that $F_{N,r}(x)\sim F_{{\rm bulk},r}(x)$ reads

$$F_{{\rm bulk},r} = \frac{e^{-Nr(1-f)}}{N}\sum_{n=0}^{N-1}(N-n)\sum_{j_1\cdots j_r=0}^{n}\frac{[N(1-f)]^n}{j_1!\cdots j_r!},$$

where for brevity we skipped the argument x, and the multiple sum is over indices such that $j_1+\cdots+j_r=n$ and $f(x)$ is the asymptotic spectral CDF related to the asymptotic spectral density $\rho_{\rm bulk}^{(1)}$. We take out the exponent and powers outside of the sums, and compute the constrained multiple sum as $\sum_{j_1\cdots j_r=0}^{n}\frac{1}{j_1!}\cdots\frac{1}{j_r!} = \frac{r^n}{n!}$, and so the extreme CDF in the bulk reads

$$F_{{\rm bulk},r}(x) = e^{-N(1-f)}\frac{1}{N}\sum_{n=0}^{N-1}(N-n)\frac{1}{n!}[Nr(1-f)]^n. \quad (13)$$

In Sec. A of the Supplemental Material [19] we find an asymptotic form of this sum:

$$F_{{\rm bulk},r}(x) = r[f(x)-\alpha_r]\theta[f(x)-\alpha_r], \quad (14)$$

where $\alpha_r = \frac{r-1}{r}$. We stress that the formula is valid in the bulk and f is the matrix CDF $f(x) = \frac{1}{N}\int_{-\infty}^{x}\rho_{\rm bulk}^{(1)}(y)\,dy$.

We emphasize that the current formula found within the matrix setup is the same as Eq. (5) found in the thinning approach. At first glance this is a very surprising result, since in the current matrix case, we inspect CDFs of highly correlated eigenvalues, while in Eq. (5) we restricted the thinned approach to independent random variables! The key is in understanding the bulk region properly as an effective macroscopic picture of eigenvalues where all correlations are absorbed into the spectral density alone. Hence the macroscopic and bulk point of view are indeed equivalent.

This relation also marks the limitations of the thinning method, which holds only when the underlying eigenvalues are typical, i.e., drawn from spectral PDFs. In the following we comment on a case for which this assumption does not hold.

b. Near the edge. Near the edge we lack an explicit formula due to strong correlations rendering the formula for gap functions intractable. Instead, we define $\lim_{N\to\infty}F_{N,r}(x_{\rm edge}+\sigma N^{-\alpha}) = F_{{\rm edge},r}(\sigma)$ (for example, $\alpha=2/3$ in the soft edge regime) and present its implicit integral representation:

$$F_{{\rm edge},r} = \lim_{N\to\infty}\oint_{\Gamma(0)}\frac{dz}{2\pi iNz}\sum_{k=1}^{\infty}\sum_{n=0}^{k-1}\left[\sum_{j=0}^{n}\frac{E_{\rm edge}(j;\sigma)}{z^{n-j}}\right]^r, \quad (15)$$

where $\Gamma(0)$ is a contour encircling $z=0$ counterclockwise and $E_{\rm edge}$ is found in Sec. A of the Supplemental Material [19]. Equation (15) is found simply from Eq. (10) since the contour integral is an alternative representation of the constraint $j_1+\cdots+j_r=n$ present in the multiple sum.

This equation is an implicit form of what we tentatively call a free Airy CDF. The name stems from the $r=1$ case where $\frac{d}{dx}F_{N,1}(x)|_{x=x_{\rm edge}+\sigma/N^{-\alpha}}$ in general describes the spectral edge oscillations of the Airy type. Hence, for general r we expect a similar oscillatory pattern to emerge.

c. Large sample limit $r\to\infty$ in the bulk and the edge regimes. Last, we consider the limit of large sample, i.e., when $r\to\infty$ or when the number of matrices we maximize grows. In the bulk, behavior is simple since CDF is localized inside an interval $x\in(f^{-1}(\alpha),x_*)$ where x_* is the rightmost edge point in the spectrum. As we increase r, since $f^{-1}(\alpha)\to f^{-1}(1)=x_*$, the result is a step function placed at the rightmost edge of the spectrum x_*:

$$\lim_{r\to\infty}F_{{\rm bulk},r}(x) = \theta(x-x_*). \quad (16)$$

This agrees with the intuition that as we draw from an increasing pool of matrices, the result becomes a degenerate matrix with N eigenvalues equal to x_* as these are the maximal attainable eigenvalues of a matrix when studied in the bulk regime. Typically x_* is also the edge point x_{edge}. In the following we will consider such case.

Near the edge $x_* = x_{\text{edge}}$, we observe fluctuations of the maximal eigenvalue position, and so it is no longer fixed. Moreover, we expect a natural transition point to happen around $r = N$ as then the pool of $rN \sim N^2$ eigenvalues becomes large enough so that almost surely the N largest are almost all extremes from the single-matrix point of view. Detailed treatment of this transition could be done through setting $r = \rho N$ in formula (15), it is intractable already for fixed and finite r. Instead, we offer a less rigorous but both intuitively and numerically backed approach.

We continue with the discussion about the nature of eigenvalues as they are sifted through the maximalization procedure. For concreteness, we set $r = \rho N$ with integer $\rho \geqslant 1$. Then each chosen eigenvalue has on average ρ single-matrix extremes to maximize over—we look for extremes among the (single-matrix) extremes. In other words, all nonextreme eigenvalues are almost always disregarded when looking at a pool numerous enough (large r). This situation describes an emergent picture—for large enough r, instead of picking N out of $N^2\rho$ eigenvalues, we pick N out of $N\rho$ *extreme* eigenvalues. Crucially, it is also tractable as the joint emergent law for extreme eigenvalues $\mathbf{P}$ completely decouples, and we go through all the steps in the derivation of CDF (10) with $F_{N,N\rho}(x_{\text{edge}} + \sigma N^{-\alpha}) \sim \bar{F}_{\text{edge},\rho}(\sigma)$:

$$\bar{F}_{\text{edge},\rho}(\sigma) = \sum_{n=0}^{N-1} \frac{N-n}{N} \sum_{j_1 \ldots j_\rho = 0}^{n} \prod_{i=1}^{\rho} \binom{n}{j_i} \times [1 - E_{\text{edge}}(0;\sigma)]^{j_i} E_{\text{edge}}(0;\sigma)^{n-j_i},$$

where the summand now consists only of $E_{\text{edge}}(0;\sigma)$ describing the CDF of the maximal eigenvalue Tracy-Widom formula [21], while the multiple sum is subject to a constraint $j_1 + \cdots + j_\rho = n$. We simplify the above formula similarly to Eq. (13) by using an identity $\sum^{n}_{\substack{j_1 \cdots j_\rho = 0 \\ j_1 + \cdots + j_\rho = n}} = \binom{N\rho}{n}$, and so multiple sums read $\bar{F}_{\text{edge},\rho}(\sigma) = \sum_{n=0}^{N-1} \frac{N-n}{N} \binom{N\rho}{n} E_{\text{edge}}(0;\sigma)^{n(\rho-1)} [1 - E_{\text{edge}}(0;\sigma)]^n$. In Sec. A of the Supplemental Material [19] we compute its large N asymptotics:

$$\bar{F}_{\text{edge},\rho}(\sigma) = \rho[E_{\text{edge}}(0;\sigma) - \alpha_\rho]\theta[E_{\text{edge}}(0;\sigma) - \alpha_\rho], \quad (17)$$

where $\alpha_\rho = \frac{\rho - 1}{\rho}$. The formula admits a form identical to the bulk extreme CDF $F_{\text{bulk},r}$ given by Eq. (14) and thinned CDF (5) upon substituting $f(x) \to E_{\text{edge}}(0,\sigma)$ and $r \to \rho$. Although it was derived for integer ρ, the above equation can be computed for any value ρ. In Fig. 3 we present numerical experiments in the case of a Gaussian unitary ensemble where we clearly observe regions of validity for formulas (14) and (17).

It is important to note that Eq. (17) does not give a free Tracy-Widom distribution despite consisting of a Tracy-Widom formula. A correct way of finding such law is possible however, one would need to go beyond order statistics (8) and define global gap functions, which we find as both interesting yet unsolved problem.

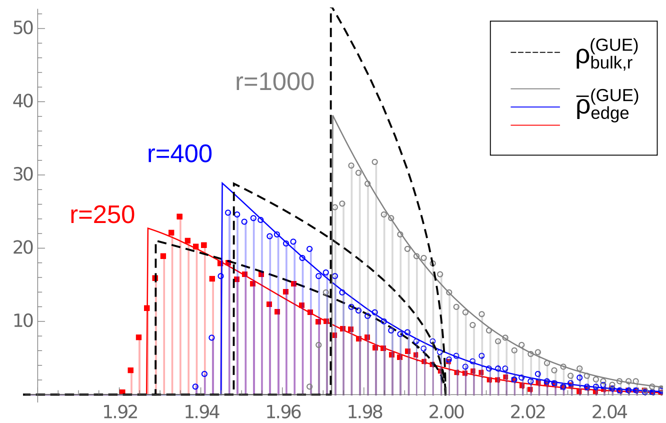

FIG. 3. Numerical and analytic plots of eigenvalues comprising solid lines for the edge density, dashed black lines for bulk density, and points for the histograms. Density of the eigenvalues in bulk is given by $\rho^{(\text{GUE})}_{\text{bulk},r}(x) = \frac{d}{dx} F^{(\text{GUE})}_{\text{bulk},r}(x)$ and Eqs. (23) and (24), whereas the edge formula holds with CDF given by Eq. (25). Histograms are plotted for matrices of size 100×100 each drawn from the GUE defined by jPDF (21). We present only large r parameters $r > N$ where the spectrum deviates considerably from the bulk description and is in turn described well by the edge formula.

C. Free extreme values

Free extreme values were introduced in general operator language in Refs. [13,16], and the special case of extreme matrices was discussed in detail in Ref. [17]. In the following we restrict ourselves to user-friendly operational definitions applicable to random matrices.

We first define an operational definition of max operation for random Hermitian matrices $H_a \vee H_b$: given $2N$ eigenvalues of H_a, H_b, we pick out the N largest eigenvalues and form the spectrum of $H_a \vee H_b$. Since a random matrix is unitarily invariant, eigenvalues alone fully specify the matrix. We state the maximal law for asymptotically large matrices given in Refs. [13,17]. The main result we need from these papers is that of the asymptotic eigenvalue CDF of the maximum $H_a \vee H_b$ of two random matrices H_a, H_b:

$$F_{H_a \vee H_b}(x) = \max\big(0, f_{H_a}(x) + f_{H_b}(x) - 1\big), \quad (18)$$

where $f_H(x) = \int_{-\infty}^{x} dt\, \rho_{\text{bulk},H}(t)$ is the spectral CDF of the corresponding bulk PDF $\rho_{\text{bulk},H}(t) = \lim_{N\to\infty} \frac{1}{N} \langle \sum_{i=1}^{N} \delta(\lambda_i - t) \rangle$. Single-matrix definitions of CDF and PDF are the same as to those found in Sec. II A and II B with an additional subscript denoting the underlying matrix H.

A special case of Eq. (18), for a maximum of r i.i.d. matrices each with eigenvalue CDF $f_H(x)$ we have

$$F_{H^{\vee r}}(x) = \max(0, r f_H(x) - (r-1)), \quad (19)$$

where $H^{\vee r} = \underbrace{H \vee \cdots \vee H}_{r\text{ terms}}$.

D. Peak-over-Threshold method

The method is closely related to the notion of *exceedances*, which arise conditioned on the event that the random variable X is larger than some threshold u. For $t \geqslant u$, the exceedance distribution function $\mathrm{F}_{[u]}(t)$ is then

$$\mathrm{F}_{[u]}(t) = \mathcal{P}(X < t|X > u) = \frac{\mathcal{P}(X < t, X > u)}{\mathcal{P}(X > u)} = \frac{f(t) - f(u)}{1 - f(u)},$$

where we used the usual definition of conditional probability $\mathcal{P}(A|B) = \mathcal{P}(A, B)/\mathcal{P}(B)$ and CDF $f(x) := \mathcal{P}(X < x)$. The Peak-over-Threshold method (POT) developed in Refs. [15,22] in turn looks at excess distribution functions of events X above some threshold u:

$$\mathcal{P}_{\mathrm{POT}}(X < u + t|X > u) = \frac{f(u+t) - f(u)}{1 - f(u)}. \quad (20)$$

An excess of t is therefore a variant of an exceedance shifted by the threshold u, i.e., $\mathrm{F}_{[u]}(t + u)$.

E. Extreme GUE example

In order to highlight both similarities and differences of first three frameworks we work out explicitly the case of a Gaussian unitary ensemble (GUE). The model is properly rescaled with jPDF given by

$$P(H)dH \sim \exp\left(-\frac{N}{2}\mathrm{Tr}H^2\right)dH, \quad (21)$$

so that the bulk spectral density is contained within an interval $(-2, 2)$ and is given by the Wigner semicircle law $\rho^{(\mathrm{GUE})}_{\mathrm{bulk}}(y) = \frac{1}{2\pi}\sqrt{4 - y^2}$. Its CDF reads $f^{(\mathrm{GUE})}_{\mathrm{bulk}}(x) = \int_{-2}^{x} \rho^{(\mathrm{GUE})}_{\mathrm{bulk}}(y)\, dy$:

$$f^{(\mathrm{GUE})}_{\mathrm{bulk}}(x) = \frac{1}{2} + \frac{1}{4\pi}x\sqrt{4 - x^2} + \frac{1}{\pi}\arcsin\frac{x}{2}. \quad (22)$$

a. GUE through a thinning method. First, we look at the GUE example from the point of view of a thinning procedure; i.e., we think of taking m i.i.d. random variables each drawn from a CDF (22). Then we look at n largest variables and study the resulting quantity in the $m, n \to \infty$ limit such that $m/n = r$ remains fixed. This results in Eq. (5) adapted to our example:

$$\mathbf{F}^{(\mathrm{GUE})}_{r} = r\big[f^{(\mathrm{GUE})}_{\mathrm{bulk}}(y) - \alpha_r\big]\theta\big[f^{(\mathrm{GUE})}_{\mathrm{bulk}}(y) - \alpha_r\big], \quad (23)$$

where $\alpha_r = \frac{r-1}{r}$.

To check the above formula we collect several numerical experiments in Fig. 2 (left plot) and find perfect agreement for $N = 500$, $r = 5, 20, 80$. For larger values of pool size r, the center and right plots of Fig. 2 become more concentrated around $x = 2$ (notice the scales on the x axis), i.e., in an interval $x \in [(f^{(\mathrm{GUE})}_{\mathrm{bulk}})^{-1}(\alpha_r), 2]$. For α_r sufficiently close to 1, bulk asymptotics becomes less relevant as $(f^{(\mathrm{GUE})}_{\mathrm{bulk}})^{-1}(\alpha_r) \to 2$, and the thinning approach is no longer applicable.

b. GUE as an extreme random matrix. We look at r random matrices each drawn from jPDF (21) and inspect their spectra. Then we pick N largest eigenvalues and form an extreme CDF given by Eq. (14):

$$F^{(\mathrm{GUE})}_{\mathrm{bulk},r} = r\big[f^{(\mathrm{GUE})}_{\mathrm{bulk}}(x) - \alpha\big]\theta\big[f^{(\mathrm{GUE})}_{\mathrm{bulk}}(x) - \alpha\big], \quad (24)$$

where $x \in (-2, 2)$ and with the function $f^{(\mathrm{GUE})}_{\mathrm{bulk}}$ given by Eq. (22). As we discussed previously, thinning method and extreme matrices in the bulk are equivalent and so are CDFs (23) and (24). For large pool sizes r both lose their applicability, and we enter the edge scaling regime.

For larger values of r, the center and right plots of Fig. 2 become more concentrated around $x = 2$, and the soft-edge asymptotics become dominant when α_r is such that $2 - (f^{(\mathrm{GUE})}_{\mathrm{bulk}})^{-1}(\alpha_r) \sim N^{-2/3}$ or when $r \sim N$. In Fig. 2 we notice this transition through the appearance of tail-like features missed altogether by the bulk formula and captured by the implicit formula (15). We pass to values $r = 500, 600, 700, 900$ with $N = 500$ and enter into region of increasing tails reaching beyond the bulk boundary at $x = 2$ where we have at our disposal an explicit formula for the extreme CDF given by Eq. (17). In our GUE example it reads

$$\bar{F}^{(\mathrm{GUE})}_{\mathrm{edge},\rho}(\sigma) = \rho[E(0;\sigma) - \alpha_\rho]\theta[E(0;\sigma) - \alpha_\rho], \quad (25)$$

where $E(0;\sigma)$ is the CDF of the $\beta = 2$ Tracy-Widom law and $\alpha_\rho = \frac{\rho-1}{\rho}$. We tentatively name it a *double* extreme law as it sieves out extremes among extremes. In Fig. 3 we show how well the large r formula (25) fits the simulations and juxtapose it with the bulk formula (24).

c. GUE as a free extreme value. We move on to describe the GUE example from the point of view of free extreme values. The CDF for the maximum is given by (19) for a GUE:

$$F_{H^{\vee r}}(x) = \max\big[0, r f^{(\mathrm{GUE})}_{\mathrm{bulk}}(x) - (r-1)\big]. \quad (26)$$

We use a formula $\max(0, x) = x\theta(x)$, which results in a formula equal to Eq. (23) or (24) obtained in the previous sections within the thinning and extreme random matrix framework.

Similarly as in the thinning method, the bulk level is the only level of detail we can access through free probability. The large matrix size limit is taken implicitly so that no results related to edge-like phenomena have natural counterparts within this framework.

III. CONNECTIONS BETWEEN EXTREME LAWS

In this section, we start from recapitulating already observed links as well the the new ones, between two known and two alternative approaches to free (or matrix) extremes given in Sec. II. Then we establish notation for the classical extreme laws of Fréchet, Weibull, and Gumbel and recall its free analogs introduced in Ref. [13] to prepare the ground for one of the main results of present work: the *exponentiation* formula relating both worlds. The last part is devoted to explicit calculations of free extreme laws in models related with random matrices.

A. Relating CDFs between frameworks

Although the first three frameworks of the thinning method, extreme random matrices, and free extreme values discussed in Sec. II start off from slightly different initial considerations, they arrive at the same CDFs given by Eqs. (5), (14), and (19). Extreme random matrices are equivalent only in the bulk, while CDFs for the free extreme values require only minor reformulation using the identities $\theta(af) = \theta(f)$ for $a > 0$ and $\max(0, f) = f\theta(f)$.

The Peak-over-Threshold approach requires some work. We present how the POT excess distribution function given by Eq. (20),

$$\mathcal{P}_{\mathrm{POT}}(X < u+t|X > u) = \frac{f(u+t) - f(u)}{1 - f(u)},$$

is related to the thinned CDF (5) $\mathbf{F}_r(x) = r[f(x) - \alpha_r]\theta[f(x) - \alpha_r]$.

It is evident that both methods study extremes: the POT method looks at values above some threshold, whereas the thinning approach focuses on a fraction r of largest values drawn from the sample of m observations. Thus, we relate the POT threshold u to the thinning parameter r:

$$f(u) = 1 - \frac{1}{r(u)}, \tag{27}$$

via the known CDF $f(x)$. This one-to-one relation dictates where one should position the threshold u in order to capture a fraction r of values in the sample. This relation is strict in the limit of large samples as only then do the intersample fluctuations vanish. By the same reason, this relation makes sense for any real value of r. In the statistics literature, the threshold value u is set to be an r quantile of the CDF f.

Now we show how with a quantile relation (27) between threshold u and thinning size r, the thinned CDF given by Eq. (5) has a form of the POT excess distribution function $\mathcal{P}_{\mathrm{POT}}$ given by Eq. (20). Therefore, in Eq. (5) we plug $r \to r(u)$ and evaluate the function at an exceedance level $x = u + t$:

$$\mathbf{F}_{r(u)}(u+t) = r(u)\left[f(u+t) - 1 + \frac{1}{r(u)}\right] \times \theta\left[f(u+t) - 1 + \frac{1}{r(u)}\right].$$

We next plug in (27) and find $f(u+t) - 1 + \frac{1}{r(u)} = f(u+t) - f(u)$ along with $\theta[f(u+t) - f(u)] = \theta(t)$ as f is a monotonic function. Finally, we obtain

$$\mathbf{F}_{r(u)}(u+t) = \frac{f(t+u) - f(u)}{1 - f(u)}\theta(t),$$

which re-creates the POT excess distribution function given by Eq. (20) with an implicit assumption that $t > 0$. This expression is also exactly that of Definition 7.2 given in Ref. [13]. So the POT is equal to the remaining three approaches,

$$\mathcal{P}_{\mathrm{POT}}(X < u+t|X > u) = \mathbf{F}_{r(u)}(u+t), \tag{28}$$

through a change of variables; instead of sample size r and spectral parameter x used in thinning, free extreme values, or extreme random matrices, we inspect the threshold u with an excess parameter t. Parameters r and u are nonlinearly related through the CDF and Eq. (27).

B. Classic and free extreme laws

a. Classic extreme laws. We first revise the classical extreme laws arising when inspecting the distribution of the largest value $y_{(1)}$ in the sample of m i.i.d. variables. The thinning approach encompasses this case upon setting $n = 1$ in the CDF (4) and studying the $m \to \infty$ limit:

$$\lim_{m\to\infty} \mathbf{F}_{m,1}(a_m + b_m x) = F^{\mathrm{class}}(x), \tag{29}$$

with m dependent constants a_m and b_m representing *centering* and *scaling*, respectively. By the Fisher-Typpett-Gnedenko theorem, there exist three limiting forms of $F^{\max}(x)$ depending on the properties of the CDF $f(x)$ as summarized in Table I (see Ref. [23] for a pedagogical review).

b. Free extreme laws. Highly similar free extreme laws exist for the CDF of noncommutative (or free) random variables defined as the limit of the formula given by Eq. (19):

$$\lim_{r\to\infty} F_{H^{\vee r}}(a_r + b_r x) = F^{\mathrm{free}}(x), \tag{30}$$

with some scaling and centering constants a_r, b_r. The classical and free extreme laws are highly similar—they admit the same domains of attraction, constants a_r, b_r, and properties of the parent distributions. The functional forms for extreme CDF's are, however, different and summarized in Table II.

Since the free CDF (19) coincides exactly with both the thinning CDF (5) and bulk extreme CDF (14), the same limiting extreme laws follow. The POT method, however, is slightly more contrived due to a change of variables (27) used in deriving relation (28). In Sec. B of the Supplemental Material [19] we work out how this variable change results in new scaling constants and the same extreme laws.

C. *Exponentiation* explains relation between classical and free extreme laws

Although classical and free (as well as POT) extreme CDFs have different functional forms, they seem to be related by a

TABLE I. Summary of the three classical extreme laws of Gumbel, Fréchet, and Weibull. Functional inverse of the CDF f is denoted by f^{-1} and $\alpha_n = \frac{n-1}{n}$.

Name	Gumbel	Fréchet	Weibull
Properties of PDF $p(x) = f'(x)$	Tails falls off faster than any power of x	$p(x)$ falls off as $\sim x^{-(\gamma+1)}$ and is infinite	$p(x)$ is finite, $p(x) = 0$ for $x > x_+$; $p(x) \sim (x - x_+)^{-\gamma-1}$
Maximal CDF	$F_I^{\mathrm{class}}(x) = \exp(-e^{-x}), x \in \mathbb{R}$	$F_{II}^{\mathrm{class}}(x) = \begin{cases} 0, & x < 0 \\ \exp(-x^{-\gamma}), & x > 0 \end{cases}$	$F_{III}^{\mathrm{class}}(x) = \begin{cases} \exp(-(-x)^{\gamma}), & x < 0 \\ 1, & x > 0 \end{cases}$
a_n	$f^{-1}(\alpha_n)$	0	x_+
b_n	$f^{-1}(\alpha_{ne}) - a_n$	$f^{-1}(\alpha_n)$	$x_+ - f^{-1}(\alpha_n)$

TABLE II. Summary of free extreme laws along with concrete examples computed in Sec. III C 1. Functions $t(x)$, $T(x)$ are step functions used in the exponentiation map (33).

Name	Free Gumbel	Free Fréchet	Free Weibull
CDF	$F_I^{\text{free}}(x) = \begin{cases} 0, & x<0 \\ 1-e^{-x}, & x>0 \end{cases}$	$F_{II}^{\text{free}}(x) = \begin{cases} 0, & x<1 \\ 1-x^{-\gamma}, & x>1 \end{cases}$	$F_{III}^{\text{free}}(x) = \begin{cases} 0, & x<-1 \\ 1-(-x)^{\gamma}, & x\in(-1,0) \\ 1, & x>0 \end{cases}$
$t(x)$	1	$\theta(x)$	1
$T(x)$	$\theta(x)$	$\theta(x-1)$	$\theta(x+1)$
Examples	Free Gauss (Ex. 3)	Free Cauchy $\gamma=1$ (Ex. 2) Free Lévy-Smirnov $\gamma=1/2$ (Ex. 6)	Wigner's semicircle $\gamma=3/2$ (Ex. 1) Marçenko-Pastur $\gamma=3/2$ (Ex. 4) Free arcsine $\gamma=1/2$ (Ex. 5)

striking expression:

$$F^{\text{free}}(x) \approx 1 + \ln F^{\text{class}(x)} \quad \text{or}$$
$$F^{\text{class}}(x) \approx \exp[F^{\text{free}}(x) - 1], \tag{31}$$

Such a relation between POT and classical extreme laws has been observed in classical probability; see, e.g., Ref. [3]. Unfortunately, it is valid only for the functional forms (see Tables I and II) and not for whole functions because their domains do not simply match up. Using the thinning method we now derive a slightly modified formula (31) which corrects these domain inconsistencies. We stress that the thinning approach exemplified in formula (3) for the CDF $\mathbf{F}_{m,n}$ is indispensable and encompasses both classical and free worlds. The former is attained when $n = 1, m \to \infty$, while the latter as $n, m \to \infty$ with $m/n = r$ fixed and then taking $r \to \infty$.

The classical extreme laws F^{class} are found as limits of CDF (3) for $n = 1$ and in the $m \to \infty$ limit:

$$\mathbf{F}_{m,1}(x) = [f(x)]^m \theta[f(x)], \tag{32}$$

with the CDF $f(x)$ and where we added a step function as f is a positive function.

The free extreme laws F^{free} on the other side arise from the asymptotic thinned CDF given by Eq. (5) where we rename $r \to m$ and write explicitly $\alpha_m = \frac{m-1}{m}$:

$$\mathbf{F}_m(x) = m\left[f(x) - 1 + \frac{1}{m}\right]\theta\left[f(x) - 1 + \frac{1}{m}\right].$$

To find the correct formula relating extreme laws, we first combine both formulas and afterwards compute the $m \to \infty$ limit. The exponential map (31) relating free and classical extreme laws reads

$$F^{\text{class}}(x) = t(x) \exp\left[\frac{F^{\text{free}}(x)}{T(x)} - 1\right], \tag{33}$$

where step functions t and T are given in Table II. Details of this derivation are given in Sec. C of the Supplemental Material [19].

The presence of step functions in the denominator is a formal notation which becomes evident by rewriting free CDFs of Table II with the use of step functions:

The Gumbel domain gives $F_I^{\text{free}}(x) = \theta(x)(1 - e^{-x})$

The Fréchet domain gives $F_{II}^{\text{free}}(x) = \theta(x-1)(1 - x^{-\gamma})$

The Weibull domain gives $F_{III}^{\text{free}}(x) = \theta(x+1)[1 - \theta(-x)(-x)^{\gamma}]$

and realizing how their pure functional forms can be expressed as a ratio.

a. POT extreme laws. Finally, the POT formalism is in the same region as free extreme laws, and so formally all formulas shown in this section hold also for the POT approach. Details about different scaling constants are given in Sec. B of the Supplemental Material [19].

1. Examples

Finally, due to the operational simplicity of the thinning theorem, we are able to give several explicit examples of free extreme laws following from spectral densities of large random matrix models. The majority of well-known models are defined on a finite spectral support, like Wigner's semicircle or the Marçenko-Pastur distribution. Via the exponentiation argument, they belong therefore to the free Weibull class, which we show with an explicit calculation. The free Fréchet class is more subtle, since the spectral density has to vanish as a power law and the support is not limited. The so-called Bercovici-Pata construction [12], being the analog of Lévy heavy-tailed distributions in classical probability, provides explicit examples. We consider two exotic random matrix models, corresponding to the free Cauchy and free Lévy-Smirnov distribution, and by explicit calculation we show that they realize extreme statistics of the free Fréchet class. The last class, the free Gumbel distribution, turned out to be the most demanding to find, despite being relatively common in classical probability, as realized, e.g., by Gaussian or Poisson distributions. Here as an example we used recent work [24,25]. This last example shows that the standard practice of calling the Wigner semicircle a "free Gaussian" has to be used with care.

In what follows we use the extreme CDF found in all the discussed frameworks, and we denote it jointly as

$$\mathbf{F}_r(x) = r[f(x) - \alpha_r]\theta[f(x) - \alpha_r],$$

while the corresponding PDF is defined as

$$\mathbf{p}_r(x) = \frac{d}{dx}\mathbf{F}_r(x) = r\rho(x)\theta[f(x) - \alpha_r],$$

with density $\rho(x) = f'(x)$.

a. Example 1: Wigner's semicircle law (free Weibull domain). An example of the GUE discussed in Sec. II E belongs to the free Weibull domain. To show this, we choose the scaling parameters $a_r = 2$, $b_r = a_r - f^{-1}(\alpha)$ with the CDF given by Eq. (22), set $x = a_r + b_r\tilde{x}$, and find $\theta[f(x) - \alpha_r] = \theta[x - f^{-1}(\alpha_r)] = \theta\{[2 - f^{-1}(\alpha_r)](\tilde{x}+1)\} = \theta(\tilde{x}+1)$. With the approximation $f^{-1}(\alpha_r) \sim 2 - (\frac{3\pi}{2r})^{\frac{2}{3}}$ we find the extreme PDF:

$$\lim_{r\to\infty} \mathbf{p}_r(a_r + b_r\tilde{x})b_r\, d\tilde{x} = \tfrac{3}{2}(-\tilde{x})^{\frac{1}{2}}\theta(\tilde{x}+1)\theta(-\tilde{x})\, d\tilde{x},$$

where the second Heaviside θ function arises by truncating the semicircle law as $\theta(2-x) = \theta(-\tilde{x})$. The extreme CDF therefore reads

$$F_{III}^{\rm GUE}(x) = \begin{cases} 0, & x < -1 \\ 1 - (-x)^{3/2}, & x \in (-1, 0) \\ 1, & x > 0 \end{cases}$$

and is an example of the free Weibull distribution of Table II with parameter $\gamma = 3/2$.

b. Example 2: Free Cauchy (free Fréchet domain). In free probability, there exists the whole class of spectral distributions, which are stable under the free convolution, *modulo* the affine transformation. They form exactly the analog of Lévy heavy (fat) tail distributions in classical probability theory. This one-to-one analogy is called the Bercovici-Pata bijection [12]. As the simplest example in the free probability context, the following PDF and CDF are considered:

$$\rho_C(x) = \frac{1}{\pi}\frac{1}{1+x^2},$$
$$f_C(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x), \quad f_C^{-1}(x) = -\cot(x\pi).$$

This is the symmetric, spectral Cauchy distribution. The realization of such free heavy-tailed ensembles is nontrivial; e.g., the potential, which by the entropic argument yields the Cauchy spectrum, reads explicitly [26]

$$V(\lambda) = \tfrac{1}{2}\ln(\lambda^2 + 1),$$

so it is nonpolynomial [note that for the Gaussian ensembles $V(\lambda) \sim \lambda^2$]. However, to get the extreme law we do not need at any time the form of the potential. According to Table I, we choose $a_r = 0, b_r = f_C^{-1}(\alpha_r)$, compute $\theta[a_r + b_r\tilde{x} - f_C^{-1}(\alpha_r)] = \theta(\tilde{x} - 1)$, and find the PDF:

$$\lim_{r\to\infty} \mathbf{p}_r(a_r + b_r\tilde{x})b_r\, d\tilde{x} = \frac{1}{\tilde{x}^2}\theta(\tilde{x} - 1)\, d\tilde{x}.$$

Upon integration the extreme CDF in turn reads

$$F_{II}^{\rm C}(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-1}, & x > 1 \end{cases}$$

and belongs to the free Fréchet class of Table II with $\gamma = 1$.

c. Example 3: Free Gaussian (free Gumbel domain). To apply our procedure for this case, we have to choose the spectral distribution whose tails fall faster than any power of x. We can use the powerful result [24,27], noticing that the normal distribution is freely infinitely divisible. This implies that there exists a random $N \times N$ matrix ensemble, whose spectrum in the large N limit approaches the normal distribution. An entropic argument can even help to find the shape of the confining potential yielding such a distribution [25],

$$V(\lambda) = c + \frac{\lambda^2}{2}\,{}_2F_2\left(1, 1; \frac{3}{2}, 2; -\frac{\lambda^2}{2}\right), \tag{34}$$

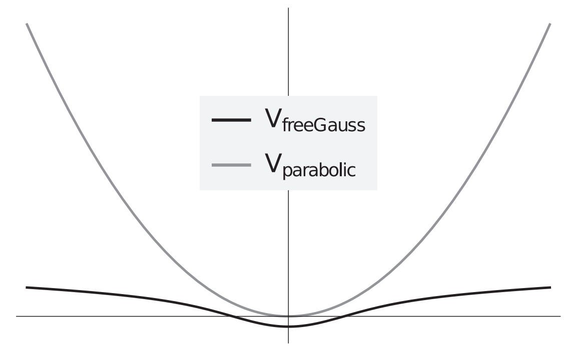

FIG. 4. Plot showing the weak confinement property of free Gaussian potential (34) in comparison with the GUE parabolic shape $V_{\rm parabolic} = \lambda^2/2$. The former results in a bell-shaped spectral density with infinite support, while the latter is a prime example of semicircular spectral density with finite support.

where $c = -\frac{\gamma + \log 2}{2}$ and the potential is a solution to $V(\lambda) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\ln|x - \lambda|\, dx$. In Fig. 4 we plot potential $V(\lambda)$ against a parabolic function showing its weakly confining property. It is interesting to note that the Green's function corresponding to the free Gaussian is equal, *modulo* the sign, to the famous and well-studied plasma dispersion function Z [28], which allows us, e.g., to study the subtle asymptotics of the resolvent. Luckily, in our thinning model we do not need the shape of the potential to find the free extreme laws. The resulting PDF, CDF, and inverse CDF (quantile) for the spectral normal distribution read, respectively,

$$\rho_{\rm G}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},$$
$$f_{\rm G}(x) = \frac{1}{2}[1 + {\rm erf}(x/\sqrt{2})], \quad f_{\rm G}^{-1}(x) = \sqrt{2}{\rm erf}^{-1}(2x - 1).$$

According to Table I, we set $a_r = f_{\rm G}^{-1}(1 - 1/r), b_r = f_{\rm G}^{-1}[1 - 1/(er)] - f_{\rm G}^{-1}(1 - 1/r)$, and so with $x = a_r + b_r\tilde{x}$ we have to perform the limit

$$\lim_{r\to\infty} \mathbf{p}_r(a_r + b_r\tilde{x})b_r\, d\tilde{x} = \lim_{r\to\infty} \frac{r}{\sqrt{2\pi}}b_r e^{-[a_r + b_r\tilde{x}]^2/2}\, d\tilde{x}.$$

The limit is subtle, since the inverse error function develops the singularity when its argument approaches unity:

$${\rm erf}^{-1}(z)|_{z\to 1} \sim \frac{1}{\sqrt{2}}\sqrt{\ln[g(z)] - \ln\{\ln[g(z)]\}}$$

with $g(z) = \frac{2}{\pi(z-1)^2}$. We set $r = \sqrt{2\pi}e^{u/2}$ and find an asymptotic series for both scaling parameters $a_r \sim \sqrt{u - \ln u}$ and $b_r \sim \sqrt{2 + u - \ln(2+u)} - \sqrt{u - \ln u}$. These asymptotic expansions result in $a_r^2 \sim u - \ln u$, $a_r b_r \sim 1$, $b_r^2 \sim \frac{1}{u}$ and $\ln b_r \sim -\frac{1}{2}\ln u$, which makes all the divergent terms cancel out and only the $a_r b_r \sim 1$ survives, yielding

$$\lim_{r\to\infty} \frac{r}{\sqrt{2\pi}}b_r e^{-[a_r + b_r\tilde{x}]^2/2}\, d\tilde{x} = e^{-\tilde{x}}\theta(\tilde{x})\, d\tilde{x},$$

which in turn gives the CDF:

$$F_I^{\rm G}(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x > 0 \end{cases},$$

an instance of the free Gumbel domain of Table II.

Other examples include, e.g., several free infinite divisible γ distributions [29], with the simplest $\rho(x) = e^{-x}$ for nonnegative x. Since $f(x) = 1 - e^{-x}$, the application of scaling and centering formulas from Table I yields, $a_r = \ln r$ and $b_r = 1$, which trivially reproduces the free Gumbel CDF.

We stress that the same functional form of the PDF may lead to either a classical or free extreme law, depending if the PDF represents the one-dimensional, classical probability or represents the spectral PDF of the ensemble of asymptotically large matrices. Examples 1, 2, and 3 show it explicitly, for each domain: Weibull, Fréchet and Gumbel, respectively. Additional examples 4, 5, and 6 listed in Table II are calculated explicitly in Sec. D of the Supplemental Material [19].

IV. CONCLUSIONS AND OUTLOOK

In this study of extreme matrices we have devised a *thinning* method, which, in contrast with previous approaches, is able to bridge the gap between classical extreme values and free (or matrix) extreme values. Through this link we establish an explicit exponentiation map between classical extreme laws of Weibull, Fréchet, and Gumbel and free Weibull, free Fréchet, and free Gumbel. Moreover, we show that also Peak-over-Threshold method is related to a thinning approach through a simple change of variables. Finally, we provide an approach of extreme random matrices, which in turn enables refined questions about the spectra of extreme matrices. The thinning method provides an operational language which we elucidate by showing several explicit examples of extreme laws of random-matrix-inspired models.

Studies of extreme matrices started relatively recently, and so many questions remain unanswered. Among the most promising is a free analog of the Tracy-Widom law or free Airy-type behavior near the spectral edge of the extreme matrix.

ACKNOWLEDGMENTS

The authors appreciate discussions with M. Bożejko, S. Majumdar, and D. V. Voiculescu, and comments from P. Warchoła nd W. Tarnowski. They also thank Ilan Roth for pointing out the connection of the Green's function for a free Gaussian to the famous plasma dispersion function Z. The research was supported by the MAESTRO DEC-2011/02/A/ST1/00119 grant of the Polish National Science Centre (NCN) (2017-2018) and by the TEAMNET POIR.04.04.00-00-14DE/18-00 grant of the Foundation for Polish Science (2019-2020).

[1] R. A. Fisher and L. H. C. Tippett, Proc. Cambridge Phil. Soc. **24**, 180 (1928); B. V. Gnedenko, Ann. Math. **44**, 423 (1943).

[2] J.-Y. Fortin and M. Clusel, J. Phys. A: Math. Theor. **48**, 183001 (2015), and references therein; G. Biroli, J.-P. Bouchaud, and M. Potters, J. Stat. Mech. (2007) P07019.

[3] R. D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*, 2nd ed. (Birkhäuser Verlag, Basel, 2001).

[4] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **73**, 016130 (2006).

[5] I. Calvo, J. C. Cuchi, J. G. Esteve, and F. Falceto, Phys. Rev. E **86**, 041109 (2012).

[6] S. Jalan and S. K. Dwivedi, Phys. Rev. E **89**, 062718 (2014).

[7] P. Greulich and B. D. Simons, Phys. Rev. E **98**, 050401(R) (2018).

[8] J. Wishart, Biometrika **20A**, 32 (1928).

[9] G. Akemann, J. Baik, and P. Di Francesco (eds.), *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford, 2011).

[10] D. V. Voiculescu, Invent. Math **104**, 201 (1991); D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables* (American Mathematical Society, Providence, RI, 1992).

[11] A. Nica and R. Speicher, Amer. J. Math. **118**, 799 (1996).

[12] H. Bercovici and D. V. Voiculescu, Ind. Univ. Math. J. **42**, 733 (1993); H. Bercovici and V. Pata, Ann. Math. **149**, 1023 (1999), Appendix by P. Biane.

[13] G. Ben Arous and D. V. Voiculescu, Ann. Probab. **34**, 2037 (2006).

[14] A. A. Balkema and L. de Haan, Ann. Probab. **2**, 792 (1974).

[15] J. I. Pickands, Ann. Statist. **3**, 119 (1975).

[16] G. Ben Arous and V. Kargin, Probab. Theory Relat. Fields **147**, 161 (2010).

[17] F. Benaych-Georges and T. Cabanal-Duvillard, J. Theor. Probab. **23**, 447 (2010).

[18] S. N. Majumdar, A. Pal, and G. Schehr, Phys. Rep. **840**, 1 (2020).

[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.102.022109 for details of the proof of the exponentiation formula, several technical details connected to derived equations, a primer on free extreme calculus, and some additional examples of random matrix ensembles belonging to free extreme statistics.

[20] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes* (CRC Press, Boca Raton, FL, 2003).

[21] C. A. Tracy and H. Widom, Commun. Math. Phys. **159**, 151 (1994).

[22] R. L. Smith, Ann. Statist. **15**, 1174 (1987).

[23] P. Vivo, Eur. J. Phys. **36**, 055037 (2015).

[24] S. Belinschi, M. Bożejko, F. Lehner, and R. Speicher, Adv. Math. **226**, 3677 (2011); M. Bożejko and T. Hasebe, Prob. Math. Stat. **33**, 363 (2013).

[25] M. Tierz, arXiv:cond-mat/0106485.

[26] Z. Burda, R. A. Janik, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed, Phys. Rev. E **65**, 021106 (2002).

[27] M. Anshelevich, S. T. Belinschi, M. Bożejko, and F. Lehner, Math. Res. Lett. **17**, 905 (2010).

[28] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function—The Hilbert Transform of the Gaussian* (Academic Press, New York, 1961).

[29] T. Hasebe, Electron. J. Probab. **19**, 33 (2014).