J Stat Phys (2016) 163:303–323
DOI 10.1007/s10955-016-1485-0

The Real Ginibre Ensemble with $k = O(n)$ Real Eigenvalues

Luis Carlos García del Molino[1,2] · Khashayar Pakdaman[1] · Jonathan Touboul[2,3] · Gilles Wainrib[4]

Received: 20 February 2015 / Accepted: 18 February 2016 / Published online: 1 March 2016

Abstract We consider the ensemble of real Ginibre matrices conditioned to have positive fraction $\alpha > 0$ of real eigenvalues. We demonstrate a large deviations principle for the joint eigenvalue density of such matrices and introduce a two phase log-gas whose stationary distribution coincides with the spectral measure of the ensemble. Using these tools we provide an asymptotic expansion for the probability $p^n_{\alpha n}$ that an $n \times n$ Ginibre matrix has $k = \alpha n$ real eigenvalues and we characterize the spectral measures of these matrices.

Keywords Real Ginibre matrices · Large deviations · Log-gas

1 Introduction

Random matrices constitute a central topic in modern probability theory [1–3] and an important tool for an increasing number of applications, from physics [4,5] to biology [6–8] or engineering [9–11].

The present study deals with specific properties of a canonical family of random matrices, the real Ginibre ensemble. Under a reference probability $\mathbb{P}$, the entries of such matrices are i.i.d. normal random variables with variance $\frac{1}{n}$ where n is the matrix size. In particular we intend to look deeper into the properties of real Ginibre matrices with anomalously large number of real eigenvalues, which are still largely unknown. These constraints are drastic for the random matrices, and affect the shape of the distribution of the eigenvalues as illustrated

✉ Luis Carlos García del Molino
garciadelmolino@gmail.com

1 Institut Jacques Monod, CNRS UMR 7592, Université Paris Diderot, Paris Cité Sorbonne, 750205 Paris, France

2 Mathematical neuroscience Team, CIRB-Collège de France (CNRS UMR 7241, INSERM U1050, UPMC ED 158, MEMOLIFE PSL), 11 place Marcelin Berthelot, 75005 Paris, France

3 MYCENAE Team, INRIA Paris-Rocquencourt, 11 place Marcelin Berthelot, 75005 Paris, France

4 Département d'Informatique (DATA), Ecole Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France

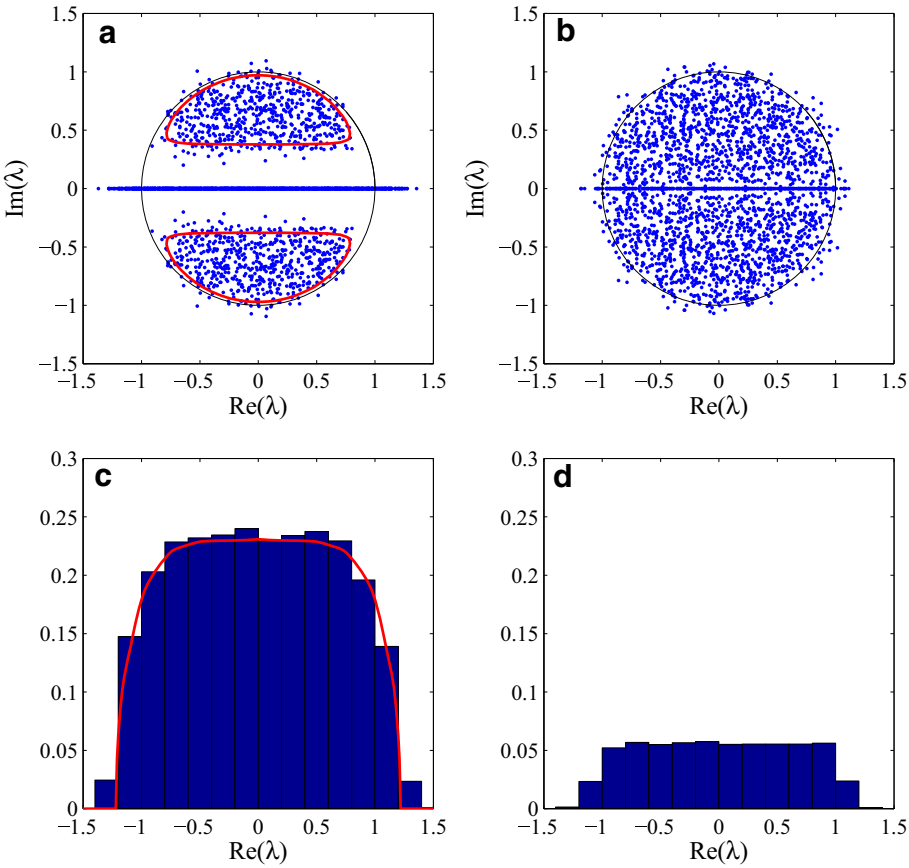

Fig. 1 *Top* Superposition of 50 spectra of 50×50 Ginibre matrices conditioned to have 26 real eigenvalues (**a**) and unconditioned (**b**). *Black lines* correspond to the unit circle and red lines correspond to our estimation of the support of the complex part of the spectrum in the large n limit. *Bottom* Histogram of the real eigenvalues of 200 50×50 Ginibre matrices conditioned to have 26 real eigenvalues (**c**) and unconditioned (**d**). Histograms are normalized with respect to n. *Red line* corresponds to our estimation of the distribution of real eigenvalues in the large n limit (Color figure online)

in Fig. 1. We aim at characterizing the eigenvalue distribution of random matrices when the number of real eigenvalues k is proportional to the matrix size n.

Our work reveals that in this regime, the empirical spectral measure (ESD) markedly departs from the one of the unconditioned ensemble. We establish that when $n \to \infty$, the ESD converges to a limit that is supported by both the real line and the complex plane. We characterize the macroscopic properties of this limit and analyze its microscopic organization. In the process, we also obtain an estimate for the probability that an $n \times n$ matrix has k real eigenvalues p_k^n with $k = O(n)$ as $n \to \infty$.

Literature Review Before presenting our main results, we concisely review some relevant past results in random matrix theory. Characterizing the spectral properties of the real Ginibre ensemble has been an active field of research. A milestone in this direction was the computation, first in [12] and later in [13], of the joint probability density of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $n \times n$ real Ginibre matrices:

$$p[\lambda_1, \cdots, \lambda_n] = C_n \prod_{i>j} |\lambda_i - \lambda_j| \left(\prod_{i=1}^{n} \exp(-n\lambda_i^2) \mathrm{erfc}\left(|\lambda_i - \lambda_i^*| \sqrt{n/2} \right) \right)^{1/2} \tag{1}$$

where the asterisk denotes the complex conjugate and C_n is the normalization constant. We denote in the sequel by Q^n the associated probability measure on $\mathbb{C}^n$ (i.e., the probability with density p with respect to Lebesgue's measure on $\mathbb{C}^n$).

A later breakthrough was the characterization of the correlations between eigenvalues in terms of Pfaffian processes in a series of studies [14–20]. More recently, this picture has been further augmented by the description of the distribution of the spectral radius and of the largest real eigenvalue of this ensemble of matrices [21].

Much work has also been devoted to the characterization of the spectrum of real Ginibre matrices in the limit of $n \to \infty$. We denote by $\hat{\mu}^n$ the empirical spectral distribution of such matrices of size $n \times n$ defined as:

$$\hat{\mu}^n = \frac{1}{n} \sum_{i=1}^{n} \delta(\lambda_i) \tag{2}$$

where $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of the matrix. For M^n a real Ginibre matrix, it is now well known that $\hat{\mu}^n$ converges to the uniform distribution on the unit disk as $n \to \infty$, a result known as circular law:

$$\hat{\mu}^n \rightharpoonup \frac{1}{\pi} \mathbb{1}_{|x| \le 1} \, . \tag{3}$$

For real Ginibre matrices, this result was first demonstrated in [13]. It has been now established that the circular law is universal in the sense that convergence of $\hat{\mu}^n$ to the uniform distribution on the unit disk holds for matrices composed of i.i.d. random variables with zero mean and $1/n$ variance [22]. Furthermore, local properties, such as correlations between eigenvalues are also universal in the $n \to \infty$ limit for matrices with independent elements with exponentially decaying distribution and moments matching the normal distribution up to fourth order [23]. See also [24,25] for universality results in dimension one, as well as rigidity results in dimension one and two [26–30].

Large deviations principles (LDP) for Gaussian random matrices were derived in [31,32], where it was shown that the sequence of empirical measures $\{\hat{\mu}^n\}_{n\to\infty}$ of M^n satisfy a LDP with speed n^2 and rate function

$$\mathcal{I}[\mu] = \frac{1}{2} \left(\int x^2 d\mu(x) - \int \int \log|x - y| d\mu(x) d\mu(y) \right) - K \tag{4}$$

for symmetric Gaussian ensembles (i.e. Gaussian Orthogonal Ensemble) and the real Ginibre ensemble respectively. In the former, the map $\mathcal{I}$ acts on probability measures on the real axis $\mathcal{M}_1^+(\mathbb{R})$ and $K = \frac{3}{8} + \frac{1}{4}\log 2$. In the later, the map $\mathcal{I}$ acts on probability measures on $\mathbb{C}$ symmetrical with respect to complex conjugation: $\mathcal{M}_1^S(\mathbb{C})$, and $K = \frac{3}{8}$.

The distinctive feature of the spectrum of real Ginibre matrices is that it has a non-zero probability of having real eigenvalues. As shown in [13,33,34], the empirical spectral distribution of finite real Ginibre matrices has a singularity on the real line because there is a positive probability of having real eigenvalues. As the matrix size n goes to infinity, this singularity disappears because the expected number of real eigenvalues is of order $\sqrt{n}$. The first numerical reports on this scaling appeared in [35] and a rigorous proof for the average number of real eigenvalues and higher order expansions in [33]. Its universality was established in [23]. Recent studies have provided a more detailed analysis of the distribution

of real eigenvalues of real Ginibre matrices, notably the inter eigenvalue gap distribution [36–38], which takes on an approximately semi-Poisson form in the bulk.

In [13] the author introduced the probabilities p_k^n for a real Ginibre matrix of size n to have k real eigenvalues and provided numerical estimates for some special cases of p_k^n and exact expressions for the specific case $k = n$. Integral expressions for these probabilities for small n were derived[39], enlarging the range of values that could be numerically computed. An exact expression for $k_n = n - 2$ and its large n asymptotic behavior was derived in [40] using the integrable structure of the real Ginibre ensemble and Pfaffian properties. Recently, fine asymptotic estimates of p_k^n for k small ($k = o(\log(n)/\sqrt{n})$) were analytically derived in [41]. To our knowledge, there is no asymptotic expression of p_k^n for large n and general k.

Another very efficient method for the study of the spectra of the Gaussian β ensembles was proposed in [42]. This pioneering work made a deep analogy between 1d log-gases, i.e. freely moving charged particles with quadratic confinement and logarithmic repulsion, and the eigenvalues of matrix-valued symmetric real Ornstein–Uhlenbeck processes. In detail, the equilibrium distribution of a one-dimensional log-gas at an inverse temperature β is precisely the distribution of the eigenvalues of the Gaussian β ensemble (symmetric, hermitian or quaternionic random matrices), and the equilibrium density of a two-dimensional log-gas is identical to the distribution of eigenvalues of the complex Ginibre ensemble. This link between interacting particle systems and spectra of random matrices has proved an essential tool to demonstrate properties of the spectrum of random matrices even in cases with extremely low probability. As an example, the use of the log-gas for symmetric matrices was instrumental in the characterization the spectrum of random matrices with anomalous densities [43,44], with applications to data analysis.

From the mathematical viewpoint, the existence and uniqueness of solutions to 1d log-gas systems as well as the convergence as the system size goes to infinity were proved in [45] and for a more general class of gases in [46]. For more on log-gases we refer to [1].

Methods and Summary of the Main Results Our methods rely on the derivation of a LDP for the Ginibre ensemble conditioned on the proportion of real eigenvalues $\alpha = k/n$. Our contribution here is to extend the LDP in [31,32] to the situation that interpolates between the two cases presented to allow for measures that are supported both on the real axis and the complex plane. We find that such matrices satisfy a LDP with rate n^2 and rate function $\mathcal{I}$, where $K = \frac{3}{8}$.

From our LDP, we are able to show that when $k/n \to \alpha$ and $n \to \infty$, $\frac{1}{n^2} \log p_k^n$ scales as $\mathcal{I}[\mu_\alpha] - K$ where μ_α is the minimizer of the rate function $\mathcal{I}$ on the set $\mathcal{M}_\alpha = \{\nu \in \mathcal{M}_1^S(\mathbb{C}) \; ; \; \nu(\mathbb{R}) = \alpha\}$. In particular, for the case $\alpha = 1$ one can see that

$$\frac{1}{n^2} \log p_n^n \to -\mathcal{I}[\mu_1] = -\frac{1}{4} \log 2 \, ,$$

which coincides asymptotically with the exact formula derived in [13] and with the formula of p_{n-2}^n derived in [40]; actually, our result shows that this logarithmic equivalent is valid for p_{n-2r}^n for any $r \in \mathbb{N}$ and not only for $r = 0$ or 1. However, obtaining a closed form expression for the minimizer μ_α is not straightforward. Nonetheless, we are able to derive a precise qualitative picture of the support and shape of the minimizer through the use of a constrained optimization problem [47].

To gain a deeper understanding on the minimizer, we next introduce and investigate the log-gas whose stationary distribution corresponds to the eigenvalue distribution of the class of matrices we are studying. In contrast with the existing literature, this log-gas is neither one nor two dimensional: it is a mixture, in the complex plane, of the two types of gas, one

fraction of the particles being confined on a singular region of the plane. We use this gas to obtain numerically for various values of α, approximations of the distribution μ_α.

The above steps characterize the macroscopic properties of the limit distribution μ_α. We complement these by a description of microscopic features. To this end, we use renormalization techniques inspired from hydrodynamics and the Ginzburg–Landau theory [48–50] that were extended to investigate the microscopic organization of particles in log-gases with applications to the distribution of eigenvalues of the Ginibre ensemble. This approach has unveiled in particular a crystallization phenomenon in one dimension, and led to conjecture that particles in two-dimensional log-gases organize according to a regular triangular lattice in the zero temperature limit [48,49]. We readily apply these methods to our mixed-type problem.

Organization of the Paper Since the complex eigenvalues of real matrices come in pairs of complex conjugates, matrices where k and n have different parity have probability 0. For this reason through the text we assume that the number of real eigenvalues k has the same parity as n. Also we introduce the notation $\overline{\alpha n}$ for the closest integer to αn with the same parity as n.

The article is organized as follows. In Sect. 2, we establish a specific LDP for real Ginibre matrices conditioned to have $k = \overline{\alpha n}$ real eigenvalues and the asymptotic estimation of p_k^n in Sect. 3. We characterize more precisely the form of the distribution of real and complex eigenvalues minimizer in Sect. 4. We introduce and analyze in Sect. 5 the 1d 2d log-gas whose stationary distribution is identical to the eigenvalues of a Ginibre matrix constrained on having a specific number of real eigenvalues. Finally, in Sect. 6, we derive the renormalized energy for the mixed gas and discuss its implications in terms of the distributions of the particles in the zero temperature limit.

2 Large Deviations Principle for $k = \overline{\alpha n}$ and Analysis of the Rate Function

Consider $M_{k_n}^n \in \mathbb{R}^{n\times n}$ with $n \in \mathbb{N}$ a sequence of real Ginibre random matrices with k_n real eigenvalues, such that k_n has the same parity as n and $k_n/n \to \alpha \in [0, 1]$. The large deviations principles shown in [31,32] correspond to $\alpha \in \{0, 1\}$. We now show that they can be extended to $\alpha \in (0, 1)$. To this purpose, we define $\mathcal{M}_\alpha$ as the subset of $\mathcal{M}_1^S(\mathbb{C})$ (i.e., probability measures symmetrical with respect to the complex conjugation on $\mathbb{C}$) exactly charging a mass α to the real line:

$$\mathcal{M}_\alpha := \{\mu \in \mathcal{M}_1^S(\mathbb{C}) : \mu(\mathbb{R}) = \alpha\}$$

For fixed $\alpha \in (0, 1)$ and finite n, the space $\mathcal{M}_\alpha$ contains empirical measures of matrices of size n only if αn is an integer value. Therefore, this space is slightly too small in order to understand the convergence properties of the spectrum of sequences of random matrices asymptotically charging non-trivial mass on the real axis: the spectral density of matrices of size n can only charge a mass proportional to $1/n$ to the real axis. In order to take into account these fluctuations of the mass on the real axis as n is increased, we introduce the decreasing sequence of spaces:

$$\mathcal{M}_\alpha^n = \left\{\mu \in \mathcal{M}_1^S(\mathbb{C}) : |\mu(\mathbb{R}) - \alpha| \le \frac{1}{n}\right\}.$$

The limit of this sequence is exactly $\mathcal{M}_\alpha$, and for any $n \in \mathbb{N}$, the space $\mathcal{M}_\alpha^n$ contains all empirical spectral densities of matrices with a number of real eigenvalues $k = \overline{\alpha n}$. Throughout the paper we will use the classical Lévy distance, which provides a metric for the topology of

weak convergence.[1] We will be interested in sequences of random matrices $(M^n)_{n\geq 0}$ whose spectral density belongs to $\mathcal{M}_\alpha^n$. These matrices satisfy the following estimates:

Theorem 1 *For any $\nu \in \mathcal{M}_\alpha$, we have:*

$$\begin{cases} \lim\limits_{\delta \searrow 0} \limsup\limits_{n\to\infty} \dfrac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n \cap B(\nu, \delta)\right] \leq -\mathcal{I}[\nu] \\ \lim\limits_{\delta \searrow 0} \liminf\limits_{n\to\infty} \dfrac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n \cap B(\nu, \delta)\right] \geq -\mathcal{I}[\nu] \, . \end{cases}$$

where $B(\nu, \delta)$ is the Lévy ball of radius δ centered at ν and $K = \frac{3}{8}$.

This property is more general than a Large Deviation Principle (LDP) in that in allows to take into account the constraint on the asymptotic proportion of real eigenvalues, which vary with the sequence index. We denote the half complex plane as $\mathbb{H} = \{z \in \mathbb{C};\, \Im(z) > 0\}$.

Proof The proof is based on a combination of evaluations and methods proposed in [31,32] in order to prove large deviations principles for the Wishart or Ginibre ensembles. The first inequality (upper bound) can be readily deduced from the corresponding inequalities in [31,32] by noting that $\mathbb{P}[\hat{\mu}^n \in \mathcal{M}_\alpha^n \cap B(\nu, \delta)] \leq \mathbb{P}[\hat{\mu}^n \in B(\nu, \delta)]$.

The lower bound is slightly more complex. In [31], the authors propose an original construction of a particular set of points on the real line, from which they construct a measure whose probability compares to the rate function and lower bounds the probability we aim at controlling. This construction was generalized in [32] where the points now belong to $\mathbb{H}$. For our purposes, a mixed construction both on the real line and on $\mathbb{H}$ proves necessary to ensure the property stated in the theorem.

Let us start by introducing the notations that are useful in the rest of the proof. We consider $\nu \in \mathcal{M}_\alpha$ a probability measure symmetric with respect to complex conjugation and charging a mass equal to α on the real line. We denote $\nu_R = \frac{1}{\alpha}\mathrm{Tr}_{\mathbb{R}}(\nu)$ the trace of the measure ν on the real axis renormalized to obtain a probability distribution, and ν_C the probability distribution obtained from renormalizing the trace of ν on $\mathbb{H}$.

We now make explicit the mixed construction of a set of points on $\mathbb{C}$ whose empirical measure will belong to $\mathcal{M}_\alpha^n$ for some integer n, and with which we will be able to deduce the lower bound expected. Let us thus fix $n \in \mathbb{N}$, define $k = \overline{\alpha n}$ and $l = (n-k)/2$ corresponding respectively to the number of points in $\mathbb{R}$ and $\mathbb{H}$ of empirical measures belonging to $\mathcal{M}_\alpha^n$. The construction, schematically described in Fig. 2, proceeds as follows. By the continuity properties of the rate function proved in [32], we can assume that ν_C has no atom with continuous and everywhere positive density with respect to Lebesgue's measure on $\mathbb{H}$, and we can define a square $\mathcal{H}$ in $\mathbb{H}$ with mass at least $(1 - 1/n)$. Similarly for ν_R on $\mathbb{R}$, one can define a bounded interval $\mathcal{R}$ with mass at least equal to $(1 - 1/n)$. The square $\mathcal{H}$ (resp. the interval $\mathcal{R}$) can be decomposed into D disjoint squares $\{B_l\}_{l\in\{1,\cdots,D\}}$ (resp. δ disjoint intervals) of length proportional to $1/\sqrt{n}$ (resp. $1/n$), in each of which are set a fixed number of complex (resp. real) points based on the density ν in each of these intervals. These points are denoted $(\tilde{\lambda}_1 < \cdots < \tilde{\lambda}_k) \in \mathbb{R}^k$ and $(\tilde{Z}_1, \cdots, \tilde{Z}_l) \in \mathbb{H}^l$ (see [31,32] for the details of the construction). The important information on these points is that:

[1] For the sake of completeness, we recall that the distance between two probability measures (μ, ν) on $\mathbb{C}$ is defined as:

$$d_L(\mu, \nu) = \inf \left\{ \delta > 0 \,;\; \mu(A) \leq \nu(A^\delta) \text{ and } \nu(A) \leq \mu(A^\delta) \quad \forall A \in \mathcal{B}(\mathbb{C}) \right\}$$

where $\mathcal{B}(\mathbb{C}$ is the Borel algebra of $\mathbb{C}$ and for $A \in \mathcal{B}(\mathbb{C})$, $A^\delta = \{x;\, d(x, A) \leq \delta\}$.

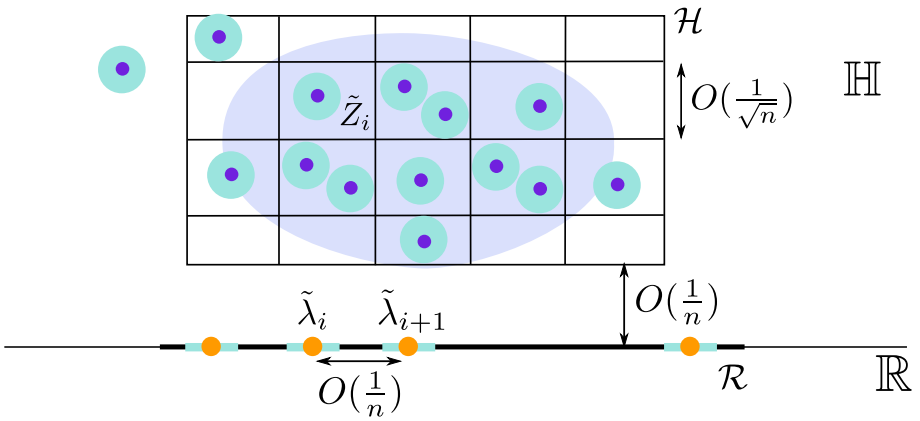

Fig. 2 The artificial empirical measure constructed for the lower bound in the ball centered at ν. The density of ν charges both $\mathbb{H}$ (*light blue* region) and a part on $\mathbb{R}$ (*bold black line*). The square $\mathcal{H} \subset \mathbb{H}$ contains in its interior a mass greater $(1-\alpha)(1-1/n)$ and is partitioned into smaller squares of typical size of order $1/\sqrt{n}$. Atoms $\tilde{Z}_i$ are *blue circles*, surrounded by *green balls*of size ε/n, and together form the space D_ε. Along the real line, *orange circles* are the atoms $\tilde{\lambda}_i$ and *green* intervals are of amplitude δ (Color figure online)

(i) $\tilde{\lambda}_i$ are the boundaries of the intervals (remark that the λ_i could alternatively be fixed as quantiles of the distribution ν_R),
(ii) the number of points $\tilde{Z}_j$ is related to the mass contained in the square they are contained in,
(iii) the distance to the boundary of the square, as well as distances between two points, are lower bounded by $C/\sqrt{n}$ for some constant C.

Note that such a construction yields an empirical measure whose mass is smaller that one. This is completed by adding some points outside $\mathcal{H}$, satisfying the same conditions of distance to boundaries and between points. Thus the constructed empirical distribution is close from ν, in the sense that the distance (in total variation) between these two measures is arbitrarily small as soon as taking into account a sufficient number of points $(\tilde{\lambda}_i)$ and a sufficiently fine partition of $\mathbb{H}$. Around these points, we can define small intervals $[\tilde{\lambda}_i - \delta/2, \tilde{\lambda}_i + \delta/2]$ and for sufficiently small $\varepsilon > 0$, non-overlapping balls centered at $\tilde{Z}^i$ with radius ε/n. The union of these balls is denoted D^ε. From this construction, following exactly the same algebra as in the pure real or complex case, we obtain:

$$
\begin{aligned}
&Q^n(B(\nu,\delta)\cap\mathcal{M}_\alpha^n)\\
&\geq B_n D_n \exp\Big((-l+\frac{1}{\sqrt{l}})\sum_{j=1}^{l/2}|\tilde{Z}_j|^2+\sum_{i\neq j}\log|\tilde{Z}_i-\tilde{Z}_j||\tilde{Z}_i^*-\tilde{Z}_j^*|-l(l+k)\log(1-\frac{2\varepsilon}{C})\\
&\quad -k\sum_{i=1}^{k}(|\tilde{\lambda}_i|+\delta)^2+\sum_{i\neq j}\log|\tilde{\lambda}_i-\tilde{\lambda}_j|+2\sum_{i=1}^{k}\sum_{j=1}^{l/2}\log|\tilde{\lambda}_i-\tilde{Z}_j|\Big)\\
&\geq B_n D_n \exp\Big((k+1)^2\int\log|x-y|d\nu_R(x)d\nu_R(y)-(k+1)^2\int x^2 d\nu_R(x)\\
&\quad -l\sum_{i=1}^{l/2}|\tilde{Z}_j|^2+\frac{l^2}{2}\int\int\log|x-y|d\nu_C(x)d\nu_C(y)+2kl\int\int\log|x-y|d\nu_C(x)d\nu_R(y)+R(\delta,\varepsilon,n)\Big)
\end{aligned}
$$

where C denotes a constant independent of n (possibly depending on the parameters δ and ε), B_n a constant depending on all parameters such that $\log(B_n) = o(n^2)$, D_n a constant satisfying the same property as in the proof of the upper bound ($1/n^2 \log(D_n) \to K$), and $R(\delta, \varepsilon, n)$ a term tending to zero when δ or ε tend to zero, and is negligible compared to n^2 as $n \to \infty$ (this term is obtained explicitly). From this expression, it is now easy to see that :

$$\begin{aligned}
&\frac{1}{n^2}\log(Q^n(B(v,\delta)\cap\mathcal{M}_\alpha^n))\\
&\quad\geq \frac{\log(C_n)}{n^2}+\frac{(k+1)^2}{n^2}\left(\int|x|^2dv_R(x)+\int\int\log|x-y|dv_R(x)dv_R(y)\right)\\
&\quad\quad+\frac{l^2}{n^2}\left(\int|x|^2dv_C(x)+\int\int\log|x-y|dv_C(x)dv_C(y)\right)\\
&\quad\quad+\frac{2kl}{n^2}\int\int\log|x-y|dv_C(x)dv_C(y)+o(1)
\end{aligned}$$

and therefore in the limit $n \to \infty$ we conclude that:

$$\begin{aligned}
&\lim_{\delta\searrow 0}\liminf_{n\to\infty}\frac{1}{n^2}\log(Q^n(B(v,\delta)\cap\mathcal{M}_\alpha^n))\\
&\quad\geq K-\alpha\int|x|^2dv_R(x)+\alpha^2\int\int\log|x-y|dv_R(x)dv_R(y)\\
&\quad\quad-(1-\alpha)\int|x|^2dv_C(x)+(1-\alpha)^2\int\int\log|x-y|dv_C(x)dv_C(y)\Big)\\
&\quad\quad+2\alpha(1-\alpha)\int\int\log|x-y|dv_C(x)dv_C(y)\\
&\quad=\mathcal{I}[\alpha v_R+(1-\alpha)v_C]
\end{aligned}$$

which ends the proof. □

Similarly to a LDP, the above theorem readily ensures the following

Theorem 2 *The rate function $\mathcal{I}$ has a unique minimizer μ_α on the space $\mathcal{M}_\alpha$. The sequence $\hat{\mu}^n \in \mathcal{M}_\alpha^n$ of empirical measures of the spectrum of Ginibre matrices conditioned on having $k = \overline{\alpha n}$ real eigenvalues converges in the Lévy topology towards μ_α.*

Proof We start by noting that the space $\mathcal{M}_\alpha$ is convex since any convex combination of elements of $\mathcal{M}_\alpha$ remains in $\mathcal{M}_\alpha$. Moreover, it is a closed subset of $\mathcal{M}_1^S(\mathbb{C})$. The convexity of $\mathcal{I}$ [31,32] ensures that the map $\mathcal{I}$ restricted to $\mathcal{M}_\alpha$ remains lower-semicontinuous, and has compact level sets in $\mathcal{M}_\alpha$. In particular, it implies that the map $\mathcal{I}$ is strictly convex on $\mathcal{M}_\alpha$, guaranteeing that there exists a unique minimum μ_α in $\mathcal{M}_\alpha$. In order to ensure that sequences of empirical processes in $\mathcal{M}_\alpha^n$ converge towards μ_α exponentially fast with speed n^2, we further need to ensure that these sequences are tight. That property is, again, a consequence of the analogous properties proved in the case of one-dimensional log-gases [31] and on the the unconstrained Ginibre ensemble [32] (seen as a consequence of the first result). Here, the elegant proof proposed in [32] readily extends to our constrained case and allows ensuring the exponential tightness of sequences of random matrices with spectral distribution constrained to belong to $\mathcal{M}_\alpha^n$, as we now outline. Defining for $K \in [0,1]$ and $r > 0$ the compact set $A_{K,r}^n$, subset of $\mathcal{M}_\alpha^n$, as:

$$A_{K,r}^n = \left\{\mu \in \mathcal{M}_\alpha^n \;;\; \mu(B_r^c) > K\right\},$$

where B_r is the complex centered ball with radius r, and $A_{K,r} = \{\mu \in \mathcal{M}_1^S(\mathbb{C}) \ ; \ \mu(B_r^c) > K\}$ it is easy to see, using the inequality $\log(|x-y|) \leq 2 + (|x|^2 + |y|^2)/4$, that:

$$Q^n(\hat{\mu}^n \in A_{K,r}^n) \leq g_n \int \prod i = 1^k d\lambda_i \prod_{j=1}^{l} dx_j dy_j \mathbb{1}_{A_{K,r}} \exp\left(-\frac{n^2}{4}\int (|z|^2 + |z'|^2) d\tilde{\mu}(z,z')\right)$$

$$\leq g_n \exp\left(-\frac{Kn^2r^2}{4}\right)$$

where $\tilde{\mu}$ is the empirical measure with atoms at λ_i and $x_j \pm \mathbf{i}y_j$ and $\log(g_n) = O(n^2)$. This inequality readily yields the desired of exponential tightness including the constraint that empirical distributions belong to $\mathcal{M}_\alpha^n$.

We have therefore built up all necessary ingredients for proving convergence of our constrained ensembles. Indeed, classical Large Deviations theory show exponentially fast convergence of a sequence of empirical measures under the condition that the sequence is exponentially tight and satisfies a LDP with good rate function (the sequence converges towards the unique minimum of the rate function). Here, analogous properties were proved under the constraint that each element of the sequence belongs to $\mathcal{M}_\alpha^n$, and the classical proof ensures convergence of the constrained measures towards the minimum of the rate function on $\mathcal{M}_\alpha$. □

This method of proof of the convergence of the sequence of empirical measures $\hat{\mu}^n$ to μ_α the unique minimizer of $\mathcal{I}$ on $\mathcal{M}_\alpha$ goes beyond the case of constraining the Ginibre ensemble to the number of real eigenvalues, and provides an account for so-called *log-gas method* which proved very efficient for the understanding of rare events in the Gaussian and Wishart β-ensembles [43,44] or the Ginibre ensemble [51].

In our purpose to characterize Ginibre matrices with prescribed proportion of real eigenvalues, we therefore need to find the minimizer μ_α of the rate function in $\mathcal{M}_\alpha$. If this is possible, we can access (i) to the typical distribution of eigenvalues under our constraint and (ii) to the probability of these events at leading order, logarithmically equivalent to $\exp(-n^2\mathcal{I}[\mu_\alpha])$. Our efforts will therefore now be devoted to the characterization of these distributions.

3 Asymptotic Behavior of $p_{\alpha n}^n$

We now use the LDP in order to characterize the asymptotic behavior of the probability $p_{\alpha n}^n$ of having asymptotically $k \sim \alpha n$ real eigenvalues. For M^n a random matrix of size $n \times n$ from the real Ginibre ensemble, we recall that

$$p_k^n = \mathbb{P}\left[M^n \text{ has } k \text{ real eigenvalues}\right] = \mathbb{P}\left[\hat{\mu}^n(\mathbb{R}) = k/n\right] \tag{5}$$

where $\hat{\mu}^n$ is the empirical spectral distribution of M^n. To account for the discrete nature of the mass on the real line and for the parity of the number of complex eigenvalues, we consider the following quantity:

$$p_{\alpha n}^n = \mathbb{P}\left[|\hat{\mu}^n(\mathbb{R}) - \alpha| \leq 1/n\right] = \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n\right] . \tag{6}$$

In order to characterize this quantity, it is tempting to apply directly the large deviation principle of [32, Theorem 1.1] on the set $A = \mathcal{M}_\alpha$ which is closed with empty interior. One therefore obtains the upper bound:

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n(\mathbb{R}) \in A\right] \leq -\inf_{\nu \in A} \mathcal{I}[\nu] = -\mathcal{I}[\mu_\alpha] . \tag{7}$$

Here one faces two difficulties. First, because of finite size effects the quantity $\mathbb{P}[\hat{\mu}^n(\mathbb{R}) \in A]$ is not exactly equal to $\mathbb{P}[\hat{\mu}^n(\mathbb{R}) \in \mathcal{M}_\alpha^n]$. In fact, if α is irrational $\log \mathbb{P}[\hat{\mu}^n(\mathbb{R}) \in A] = -\infty$ for any n. Second, even if this problem is solved, only an upper bound holds.

A much better estimate can be achieved using the results of Theorem 1:

Corollary 1 *For any $\alpha \in (0, 1)$, we have:*

$$\lim_{n\to\infty} \frac{1}{n^2} \log p^n_{\alpha n} = \lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n\right] = -\mathcal{I}[\mu_\alpha] \tag{8}$$

Proof In the proof of Theorem 1 we have shown that for any $\nu \in \mathcal{M}_\alpha$ and $\delta > 0$:

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n \cap B(\nu, \delta)\right] \leq - \inf_{\mu \in \mathcal{M}_\alpha \cap B(\nu,\delta)} \mathcal{I}[\mu] \tag{9}$$

and

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n \cap B(\nu, \delta)\right] \geq -\mathcal{I}[\nu] \tag{10}$$

And therefore, letting $\delta \to \infty$, we obtain:

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n\right] \leq - \inf_{\mu \in \mathcal{M}_\alpha} \mathcal{I}[\mu] \tag{11}$$

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\hat{\mu}^n \in \mathcal{M}_\alpha^n\right] \geq -\mathcal{I}[\nu] \tag{12}$$

$$\geq - \inf_{\mu \in \mathcal{M}_\alpha} \mathcal{I}[\mu] \tag{13}$$

readily proving result announced in Eq. (8). □

This result shows that the probabilities $p^n_{\alpha n}$ decrease as e^{-n^2}. For $\alpha = 1$, we know that μ_1 is the semi-circular law, and $\mathcal{I}[\mu_1]$ can be readily computed and is equal to $\log(2)/4$ (it is exactly the difference between the constant term in the rate function for Hermitian matrices and that for non-Hermitian matrices, see e.g. [31,32]). This shows that $\log p^n_{\varphi(n)} \sim -\frac{n^2}{4} \log 2$ for any map $\varphi : \mathbb{N} \mapsto \mathbb{N}$ with $\varphi(n)/n \to 1$. In particular, $\log p^n_{n-2r} \sim -\frac{n^2}{4} \log 2$ for any $r \in \mathbb{N}$. Furthermore the continuity and convexity properties of $\alpha \mapsto \mathcal{I}[\mu_\alpha]$ proven below imply the continuity and convexity of the map $\alpha \mapsto \lim_{n\to\infty} p^n_{\alpha n}$.

4 Properties of the Minimizer μ_α

We investigate the qualitative properties of the distribution μ_α, the large n limit spectral distribution of the Ginibre ensemble with asymptotic proportion α of real eigenvalues.

4.1 Continuity and Monotonicity of $\alpha \mapsto \mu_\alpha$

In this subsection, we establish continuity of the minimizer and of the minimum of the rate function on the subspaces $\mathcal{M}_\alpha$ with respect to the parameter α, as well as the convexity and monotonicity of the minimum.

Proposition 1 *1. The mappings $\alpha \in [0, 1] \mapsto \mu_\alpha$ and $\alpha \in [0, 1] \mapsto \mathcal{I}[\mu_\alpha]$ are continuous.*
2. The mapping $\alpha \in [0, 1] \mapsto \mathcal{I}[\mu_\alpha]$ is convex and increasing.

Proof 1. We define:

$$\mathcal{J}_\alpha[\mu] := \begin{cases} \mathcal{I}[\mu] \text{ if } \mu \in \mathcal{M}_\alpha \\ +\infty \text{ otherwise.} \end{cases} \tag{14}$$

$\mathcal{I}$ has a unique minimizer in $\mathcal{M}_\alpha$. In order to show the continuity properties in α we will use the theory of Γ-convergence [52, Theorem 1.21, p. 29]. In detail we consider a sequence $(\alpha_n)_{n\geq 0} \in [0,1]^{\mathbb{N}}$ converging towards $\alpha \in [0,1]$. For each α_n, there exists a minimum μ_{α_n} of the map $\mathcal{J}_{\alpha_n}$. To show that $\mu_{\alpha_n} \to \mu_\alpha$, we will show that the sequence of functionals $\mathcal{J}_{\alpha_n}$ Γ-converges to $\mathcal{J}_\alpha$. This amounts to proving a few inequalities on the limits of the sequence of functions $\mathcal{J}_{\alpha_n}$ together with an equi-coerceness property of the sequence $(\mathcal{J}_{\alpha_n})_n$. In detail, it suffices to prove that the sequence of processes satisfies the following properties:

(a) for any $\mu \in \mathcal{M}$ and any converging sequence $\alpha_n \to \alpha$ with $\nu_{\alpha_n} \to \mu$, the following inequality holds:

$$\liminf_{n\to\infty} \mathcal{J}_{\alpha_n}\left[\nu_{\alpha_n}\right] \geq \mathcal{J}_\alpha[\mu] \tag{15}$$

(b) for any $\mu \in \mathcal{M}$, there exists a sequence $\alpha_n \to \alpha$ and $\nu_{\alpha_n} \in \mathcal{M}_{\alpha_n}$ converging to μ such that:

$$\limsup_{n\to\infty} \mathcal{J}_{\alpha_n}\left[\nu_{\alpha_n}\right] \leq \mathcal{J}_\alpha[\mu] \tag{16}$$

(c) (Equicoerciveness): for all $n \in \mathbb{N}$ and for all $t > 0$ there exists a compact set $K_t \subset \mathcal{M}^1(\mathbb{C})$ such that $\{\nu : \mathcal{J}_{\alpha_n}[\nu] \leq t\} \subset K_t$.

We now prove that the three assumptions indeed hold in our case.

(a) This property is a consequence of the regularity properties of good rate functions. Indeed, let $\mu \in \mathcal{M}$ be a fixed measure. If $\mu \notin \mathcal{M}_\alpha$, then for n large enough, ν_{α_n} cannot be in $\mathcal{M}_{\alpha_n}$, the above lim inf is infinite and inequality (15) is trivial. Otherwise, if $\mu \in \mathcal{M}_\alpha$, the above inequality is actually a direct consequence of the lower-semicontinuity of $\mathcal{I}$: we can assume that $\nu_{\alpha_n} \in \mathcal{M}_{\alpha_n}$ (or at least a subsequence), so that $\mathcal{J}_{\alpha_n}[\nu_{\alpha_n}] = \mathcal{I}[\nu_{\alpha_n}]$ and then one uses the lower-semicontinuity of $\mathcal{I}$ to obtain (15).

(b) Given a measure $\mu \in \mathcal{M}$, the second step amounts to finding a sequence $\alpha_n \to \alpha$ and a sequence of measures $\nu_{\alpha_n} \in \mathcal{M}_{\alpha_n}$ with $\nu_{\alpha_n} \to \mu$ such that the inequality (16) holds. For $\mu \notin \mathcal{M}_\alpha$, the inequality holds for any sequence ν_{α_n} since $\mathcal{J}_\alpha[\mu] = +\infty$. Otherwise, μ can be written as $\mu = \alpha\mu_R + (1-\alpha)\mu_C$, and for any sequence $\alpha_n \to \alpha$, one defines $\nu_{\alpha_n} = \alpha_n\mu_R + (1-\alpha_n)\mu_C$. Then, $\nu_{\alpha_n} \in \mathcal{M}_{\alpha_n}$ and $\mathcal{J}_{\alpha_n}[\nu_{\alpha_n}] = \mathcal{I}[\nu_{\alpha_n}]$. Therefore, one obtains (16) by continuity of $a \mapsto \mathcal{I}[a\mu_R + (1-a)\mu_C]$.

(c) The equicoerciveness property is a consequence of the fact that $\mathcal{I}$ is a good rate function, hence has compact level level-sets. Therefore, for any $t \in \mathbb{R}$, there exists a compact K_t such that $\{\mathcal{I} \leq t\} \subset K_t$. We thus have $\{\mathcal{J}_{\alpha_n} \leq t\} = \{\mathcal{I} \leq t\} \cap \mathcal{M}_{\alpha_n} \subset K_t$.

These three results proved the Γ-convergence of $\mathcal{J}_{\alpha_n}$ towards $\mathcal{J}_\alpha$, and therefore of μ_{α_n} to μ_α for any sequence $\alpha_n \to \alpha$, which concludes the proof of the first point of the proposition.

2. We now prove the convexity and monotony of $\alpha \mapsto \mathcal{I}[\mu_\alpha]$. Let $\alpha, \alpha' \in [0,1]$. Since $\mu \mapsto \mathcal{I}[\mu]$ is convex, we have:

$$\mathcal{I}\left[t\mu + (1-t)\nu\right] \leq t\mathcal{I}[\mu] + (1-t)\mathcal{I}[\nu] \text{ for } t \in [0,1],\ \mu \in \mathcal{M}_\alpha \text{ and } \nu \in \mathcal{M}_{\alpha'} \tag{17}$$

Therefore, taking the infimum on the above equation shows that $\alpha \mapsto \mathcal{I}[\mu_\alpha]$ is convex. Since, $\mathcal{I}[\mu_0] < \mathcal{I}[\mu_1]$ and $\mathcal{I}[\mu_\alpha] \geq \mathcal{I}[\mu_0]$, we conclude that the continuous mapping $\alpha \mapsto \mathcal{I}[\mu_\alpha]$ is necessarily increasing.

□

Remark 1 Since $\alpha \mapsto \mathcal{I}[\mu_\alpha]$ is increasing, the minimum of the rate function conditioned on $\mathcal{M}_\alpha$ is equal to the minimum on $\mathcal{M}_\alpha^+ = \{\mu \in \mathcal{M}_1^S(\mathbb{C}) : \mu(\mathbb{R}) \geq \alpha\}$:

$$\inf_{\mathcal{M}_\alpha} \mathcal{I}[\mu] = \inf_{\mathcal{M}_\alpha^+} \mathcal{I}[\mu] \tag{18}$$

With this continuity properties in hand, we return to the characterization of the qualitative features of μ_α.

4.2 Qualitative Description of μ_α

We are now interested in characterizing the distribution μ_α. The large deviation principle defines this distribution as the minimum of $\mathcal{I}$, which can therefore be characterized through the computation of the differential of $\mathcal{I}$ in the space $\mathcal{M}_\alpha$. This differential is a linear operator, acting on the space of signed measures h symmetrical with respect to the real axis and such that $h(\mathbb{C}) = 0$, with traces h^R on the real axis and h^C on $\mathbb{H}$ satisfying $h^R(\mathbb{R}) = h^C(\mathbb{H}) = 0$. Denoting $\alpha\mu_\alpha^R$ the trace of μ_α on the real axis and $\frac{1-\alpha}{2}\mu_\alpha^C$ on $\mathbb{H}$ (with these definitions, $\mu_\alpha^R(\mathbb{R}) = \mu_\alpha^C(\mathbb{H}) = 1$), we find that:

$$d_\mu \mathcal{I}[h] = \frac{1}{2}\int_{\mathbb{C}} |z|^2 dh(z) - \int \log|z - z'| d\mu(z) dh(z')$$

which can be rewritten as:

$$\partial_{\mu^R}\mathcal{I}[\mu] = \int \left\{ \alpha|x|^2 + \alpha^2 \int \log|x - x'|^2 d\mu^R(x') + \alpha(1-\alpha)\int \log|x - z|^2 d\mu^C(z) \right\} h^R(dx)$$

$$\begin{aligned}\partial_{\mu^C}\mathcal{I}[\mu] = \int \Bigg\{ &(1-\alpha)|z|^2 + \alpha(1-\alpha)\int \log|z - x|^2 d\mu^R(x) \\ &+ (1-\alpha)^2 \int \left(\log|z - z'|^2 + \log|z - z'^*|^2\right) d\mu^C(z') \Bigg\} h^C(dz)\,.\end{aligned}$$

Necessarily, provided that these differentials are bounded at μ_α, the linear behavior described by the above equation fails (indeed, the freedom to replace h by $-h$ would contradict the minimality of $\mathcal{I}$ on $\mathcal{M}_\alpha$ at μ_α). Therefore the differential operator is equal to zero for any acceptable measure h, implying the following proposition:

Proposition 2 *The measure μ_α satisfies the system of integral equations defined for any $x \in \mathbb{R}$ and $z \in \mathbb{H}$ as*

$$\begin{aligned}&\alpha x^2 + \alpha^2 \int \log|x - x'|^2 d\mu_\alpha^R(x') + \alpha(1-\alpha)\int \log|x - z'|^2 d\mu_\alpha^C(z') = 0 \\ &(1-\alpha)|z|^2 + (1-\alpha)^2 \int \left(\log|z - z'|^2 + \log|z - z'^*|^2\right) d\mu_\alpha^C(z') + \alpha(1-\alpha) \\ &\quad \int \log|z - x'|^2 d\mu_\alpha^R(x') = 0.\end{aligned}$$

For the two extreme cases $\alpha \in \{0, 1\}$ we recover the circular law $\mu^R(x) = 0$, $\mu^C(z) = \frac{1}{\pi}\mathbb{1}_{|z|<1}$ for $\alpha = 0$ and the semi-circular law $\mu^R(x) = \mu_{sc}(x) = \frac{1}{\pi}\sqrt{2 - x^2}\mathbb{1}_{|x|\leq\sqrt{2}}$, $\mu^C(z) = 0$ for $\alpha = 1$.

These equations are generally too complex to be solved explicitly. It is however possible to characterize further the minimizer μ_α for intermediate values of α.

Theorem 3 *Denoting $\alpha\mu_\alpha^R$ and $\frac{1-\alpha}{2}\mu_\alpha^C$ the trace on $\mathbb{R}$ and $\mathbb{H}$ of the minimizer μ_α of $\mathcal{I}$ on $\mathcal{M}_\alpha$ for $\alpha \in [0, 1]$. The measure μ_α has the following profile:*

i. μ_α^R has a density g_α with respect to Lebesgue's measure on $\mathbb{R}$ such that $g_\alpha \in L^2(\mathbb{R})$. This density vanishes outside an interval $[-R_\alpha, R_\alpha]$ for some constant $R_\alpha > 0$

ii. μ_α^C has a constant density $\frac{2}{(1-\alpha)\pi}$ with respect to Lebesgue's measure on its support $V_\alpha \subset \mathbb{H}$. This support is a simply connected bounded open set and is such that $\bar{V}_\alpha \cap \mathbb{R} = \emptyset$.

In [53] the authors have analyzed the limiting distribution of the eigenvalues of the complex Ginibre ensemble conditioned on the event that a large proportion of the eigenvalues lies in an open subset of the complex plane. The situation we consider is similar in the sense that we have extra weight on the real line. While this is not treated explicitly in [53], the argument of proof goes through with adequate modifications.

This theorem provides us with a qualitative description of the minimizer μ_α. A more precise characterization of the minimizer requires the use of other techniques, namely numerical ones that are described below.

5 Log-Gas for Ginibre Conditioned on k

The present section introduces and analyses a log-gas whose equilibrium distribution is related to μ_α.

The log-gas approach to spectral analysis of the real Ginibre ensemble presents two new challenges with respect to previous analysis in the literature. First, the spectra of real matrices are symmetric with respect to the real axis so one has to impose this symmetry in the gas. Second, due to the positive probability of having real eigenvalues, the log-gas is composed of two interacting phases: a 2d phase supported on the complex plane and a 1d phase constrained to the real line for all timers. This last property has a very important consequence, namely that the dynamics of the whole gas depend on the quantity of real particles k with respect to the total number of particles n. Our method is based on the introduction of a mixed 1d-2d log-gas whose stationary distribution is that of the spectrum of real Ginibre matrices with k real eigenvalues.

This system is composed of n interacting particles in a two-phase gas: k of these particles are confined to the real axis and the other $2l = (n - k)$ particles form l pairs of complex conjugated particles outside the real axis. We denote the set of eigenvalues $\{\lambda_i\}_{1\leq i\leq n} \in \mathbb{C}$ with $\lambda_i = (x_i, 0)$ for $1 \leq i \leq k$, $\lambda_i = (x_i, y_i)$ for $k < i \leq k + l$ with $y_i > 0$ and $\lambda_i = \lambda_{i-l}^*$ for $k + l < i \leq n$.

5.1 Log-Gas Equations

To derive the equations for the gas dynamics for fixed k, we identify the joint probability densities (1) as a Boltzmann factor and compute the corresponding potential, following the method of [42]. One obtains the following dynamics

$$d\lambda_i(t) = \frac{\sigma}{\sqrt{n}} dB_i - \sum_i \nabla_i \Phi(\lambda_1, \ldots, \lambda_n) dt \tag{19}$$

where $\sigma = \sqrt{2}$, B_i is a real Brownian motion with $B_i = (B_i^R, 0)$ for $i \leq k$, and $B_i = (B_i^R, B_i^I)$ a complex Brownian motion for $k < i \leq k+l$. Symmetry is ensured by imposing $B_i = B_{i-l}^*$ for $k+l < i \leq n$. We denote $\nabla_i = (\partial_{x_i}, \partial_{y_i})$. The potential is given by

$$\Phi(\lambda_1, \ldots, \lambda_n) = \sum_i \sum_{j \neq i} \frac{1}{2n} V(\lambda_i, \lambda_j) + \sum_i U(\lambda_i) \tag{20}$$

with

$$\begin{aligned} V(\lambda_i, \lambda_j) &= -\log|\lambda_i - \lambda_j| \\ U(\lambda_i) &= \frac{x_i^2}{2} - \mathbb{1}_{y_i \neq 0} \left(\frac{y_i^2}{2} + \frac{1}{2n} \log\left(\operatorname{erfc}(|y_i|\sqrt{2n}) \right) \right) \end{aligned} \tag{21}$$

We will be specially interested on the $t \to \infty$ limit of the empirical measure

$$\rho_k^n(t) = \frac{1}{n} \sum_{i=1}^n \delta(\lambda_i(t)) \,. \tag{22}$$

The system (19) is well posed: we prove below that solutions neither collide nor blow up. The result is actually more general and extends to particle systems within more general confining potential U we state:

Theorem 4 *Suppose $\lambda_i(0)$ for $i = 1, \ldots, k+l$ are distinct and $\lambda_i(0) = \lambda_{i-l}^*(0)$ for $k+l < i \leq n$, that U is C^2 and the maps $(x, y) \mapsto (-x\partial_x U(x + \mathbf{i}y), -y\partial_y U(x + \mathbf{i}y))$ and $(x, y) \mapsto \Delta U(x, y)$ are upper bounded in $\mathbb{R}^2$ by some constant γ. The processes λ_i are defined by (19) up to the stopping time*

$$T = \inf\{t : \lambda_i(t) = \lambda_j(t) \text{ for some } j \neq i \text{ or } y_i = 0 \text{ for some } i > k \text{ or } \lambda_i = \infty \text{ for some } i\} \,.$$

If $\sigma^2 \leq 2$, then $\mathbb{P}[T = \infty] = 1$: there is no collision or explosion and the particle system is defined for all times.

Proof The proof follows the one in [45]. To prove that there are no collisions we show that the drift term in (19) is bounded from above. We apply Itô's formula on Φ to obtain

$$d\Phi = \frac{\sigma}{\sqrt{n}} \sum_j \nabla_j \Phi dB_j + \sum_j \left[\frac{\sigma^2}{2n} \Delta_j \Phi - (\nabla_j \Phi)^2 \right] dt \tag{23}$$

where $\Delta_j = \partial_{x_j}^2 + \partial_{y_j}^2$. Developing the drift term we have

$$\begin{aligned} &\sum_j \left[\frac{\sigma^2}{2n} \Delta_j \Phi - \left(\nabla_j \Phi\right)^2 \right] \\ &= \sum_j \left[\frac{\sigma^2}{2n} \sum_{r \neq j} \frac{1}{n} \Delta_j V(\lambda_j, \lambda_r) + \frac{\sigma^2}{2n} \Delta U(\lambda_j) \right. \\ &\quad \left. - \sum_{r \neq j} \frac{1}{n^2} \left(\nabla_j V(\lambda_j, \lambda_r)\right)^2 - \left(\nabla U(\lambda_j)\right)^2 - \sum_{r \neq j} \frac{1}{n} \nabla_j V(\lambda_j, \lambda_r)(\nabla U(\lambda_j) - \nabla U(\lambda_r)) \right] \end{aligned}$$

The term $-(\nabla U(\lambda_j))^2$ is clearly upperbounded. Moreover, our assumptions ensure that both terms $\frac{\sigma^2}{2n}\Delta U(\lambda_j)$ and the last term are upperbounded. Therefore, the only term that we have to control is

$$\sum_j \sum_{r\neq j} \frac{\sigma^2}{2n^2}\Delta_j V(\lambda_j,\lambda_r) - \frac{1}{n^2}\left(\nabla_j V(\lambda_j,\lambda_r)\right)^2.$$

Because of the noise in the system, the collisions between real and complex particles as well as between complex non conjugated particles are impossible. On the other hand, because of the confinement of some of the particles to the real line and the symmetry with respect to the real axis, collisions between two real particles and between a pair of complex conjugate particles are possible. Developing the Laplacian and the gradient squared for those interactions (real-real and complex-complex conjugate) one can see that

$$\frac{\sigma^2}{2n^2}\Delta_j V(\lambda_j,\lambda_r) - \frac{1}{n^2}\left(\nabla_j V(\lambda_j,\lambda_r)\right)^2 \leq \frac{\sigma^2-2}{2n^2}\frac{1}{|\lambda_j-\lambda_r|^2} \qquad \text{if } \lambda_j,\lambda_r\in\mathbb{R}$$

$$\frac{\sigma^2}{2n^2}\Delta_j V(\lambda_j,\lambda_j^*) - \frac{1}{n^2}\left(\nabla_j V(\lambda_j,\lambda_j^*)\right)^2 \leq \frac{\sigma^2-2}{2n^2}\frac{1}{|\lambda_j-\lambda_j^*|^2} \qquad \text{if } \lambda_j\in\mathbb{C}$$

and hence, using the same arguments as in [45], one can show that for $\sigma^2 \leq 2$ particles do not collide before an explosion.

Now it remains to show that there is no explosion in finite time. Let $R_t = \frac{1}{2n}\left(\sum_j x_j^2 + y_j^2\right) = \langle \mu_t^n, f\rangle$ where $f(\lambda) = |\lambda|^2/2$. We have to prove that $R_t < \infty$ for all t. As shown in [45] $R_t \leq R_t'$ a.s. for all t where R_t' solves

$$dR_t' = \frac{\sigma}{n}\sqrt{2R_t'}dW_t + \left(\frac{\sigma^2}{2n} + \frac{(n-1)}{n} - \gamma\right)dt$$

for W_t a Brownian motion. R_t' is a squared Bessel process which is known not to explode so R_t does not explode either. □

Remark 2 The assumptions on the confining potential U of the above theorem are clearly satisfied for U given by (21).

The result of this theorem ensures existence and uniqueness of the particle system for all times and finite n. If this system converges in time towards a non-singular stationary distribution, the limit is necessarily given by the joint pdf (1) of the eigenvalues of the Ginibre ensemble under our constraints. In this case the empirical measure of the particle system $\rho_k^n(t)$ converges in distribution as $t\to\infty$ to the empirical spectral measure $\hat{\mu}^n$ of the Ginibre ensemble with k real eigenvalues. Therefore, simulating the long term dynamics of the particle system will provide an approximation of $\hat{\mu}^n$.

5.2 Numerical Simulations

In principle, it is possible to generate numerically the spectra of Ginibre matrices conditioned on k by drawing Ginibre matrices at random and classifying them according to the number of real eigenvalues. However, for the events we consider, k deviates from its expected value and p_k^n decrease as e^{-n^2}, so that matrices with large k are not readily accessible with such a strategy. As we dispose of an explicit form of the joint probability distribution of the eigenvalues under our constraint, standard sampling methods, such as the Monte Carlo algorithm

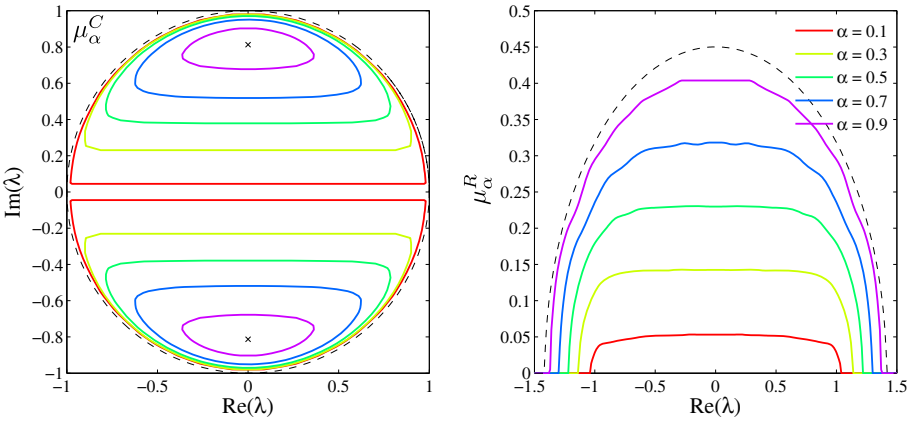

Fig. 3 *Left* Numerical estimation of the support of μ_α^C for $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The crosses correspond to the points $(0, \pm y^*)$, the *dashed line* is the unit circle. *Right* Numerical estimation of μ_α^R for $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The *dashed line* is the semi-circular law. The estimations are long time simulations of log-gases with $n = 1000$

described in Appendix, allow to access directly these rare events. Unfortunately, they are not efficient for large matrices.

The same limitations at large n arise for simulations of the log-gas given that the system is stochastic and one needs to average over several runs to smooth out fluctuations. To overcome these, we took advantage of the fact that the noise term perturbing each particle vanishes. Therefore, for estimating $n \to \infty$ spectral distributions, instead of stochastic simulations, we ran deterministic simulations of the gas. Even though the latter do not correspond to the eigenvalues of actual matrices, their $n \to \infty$ limit converges to the limit distribution of the eigenvalues, which is exactly what we seek to approximate. We checked convergence by increasing progressively the number of particles of the gas until the shapes for the complex support and the distribution of the real phase were visually indistinguishable.

Here we summarize our key findings, which can be considered educated guesses based upon known facts and numerical evidence (see Fig. 3). The first claim is that the border of the support of each connected component of the complex part of μ_α is a smooth closed curve for all $0 < \alpha < 1$. This closed curve varies continuously with α. In fact, as α goes from zero to one, the support v decreases monotonically (in the sense of strict inclusion) from the unit semi-circle in $\mathbb{H}$ to the point $z^* = (0, y^*)$ where z^* is found by minimizing the potential that a single pair of complex particles would experience interacting with $\mu^R = \mu_{sc}$ which is given by the map $z \mapsto |z^2| - \int \log|z - x| d\mu_{sc}(x)$. The coordinate y^* is hence the solution of:

$$\int \frac{1}{x^2 + (y^*)^2} d\mu_{sc}(x) = 2 \tag{24}$$

Coincidentally, the support of μ_α^R grows monotonically from $[-1, 1]$ to $[-\sqrt{2}, \sqrt{2}]$ as μ^R converges to the semi-circular law in $[-\sqrt{2}, \sqrt{2}]$.

6 Renormalized Energy and Microscopic Organization at Zero Temperature

The gas provides a precise description of the minima of the potential (20) even at temperatures that differ from that corresponding to the eigenvalues of the Ginibre ensemble. In particular, expansion of the energy of the system supplies further information on the microscopic organization of the particles at vanishing temperature. To this end, following the work for log-gases in dimension one [48], two [49] or higher [50], one shall compute the next-to-leading order terms of the energy. These terms correspond essentially to the microscopic arrangements of the particles. Since we have shown that in the regime we consider with $k = \overline{\alpha n}$, there exists a macroscopic gap between the real axis and complex eigenvalues, the energy related to the microscopic interactions of real onto the complex eigenvalues and reciprocal forces vanish in the thermodynamic limit. Once this has been noted, a direct application of the results in one and two dimensions [48,49] leads to state the following:

Proposition 3 *For large n, the equivalent energy of a Ginibre matrix conditioned on having $k \sim \alpha n$ real eigenvalues enjoys the following expansion around $\mu_\alpha = \alpha\mu_\alpha^R + (1-\alpha)\mu_\alpha^C$ the minimizer of the macroscopic energy:*

$$\mathcal{E}_{n,\alpha} = n^2\mathcal{I}(\mu_\alpha) - \frac{(1+\alpha)}{2} n\log(n) + n\Big[(1-\alpha)\frac{\kappa_2}{2\pi} + \alpha\frac{\kappa_1}{\pi}\Big] \\ -n\Big[(1-\alpha)\int_{\mathbb{C}} d\mu_\alpha^C \log(\mu_\alpha^C) + \alpha\int_{\mathbb{R}} d\mu_\alpha^R \log(\mu_\alpha^R)\Big] + o(n) \qquad (25)$$

where the κ_2 and κ_1 are universal constants related to the dimension of the spaces where μ_α^C and μ_α^R are supported.

From this expansion, the minimization at zero temperature leads to state that the real eigenvalues crystallize [48], and to the conjecture that the complex eigenvalues organize in a regular triangular lattice (called Abrikosov lattice in the superconductivity domain [54]), which is proved under the assumption that the organization is a regular lattice [49].

7 Conclusion

The main conclusions of our work can be outlined as:

1. Despite the fact that the joint probability distribution of real Ginibre matrices is always given the same compact formula above (Eq.(1)), the limit distributions of the empirical measure $\hat{\mu}^n$ strongly depends on the number k of real eigenvalues. When $k/n \to \alpha$ and $n \to \infty$, we prove that the empirical measure has a unique limit μ_α that significantly departs from the circular law.
2. The key method in establishing the above result is an LDP theorem devised to take into account both real and complex eigenvalues. Previous LDPs discarded real eigenvalues, as in the unconstrained matrices, the fraction of real eigenvalues tends to zero. While we have provided the proof in the specific case of the real Ginibre ensemble, it can readily be adapted to other situations, such as gases in higher dimensions with more general confining and repulsive potentials such as [55] or heterogeneous gases [56] with constraints (fraction of particles restrained to a subspace). It also extends to Gaussian β and Wishart ensembles conditioned on rare events (such as anomalous proportions of eigenvalues in a given interval). In this sense, our approach provides a theoretical basis

for the log-gas method used in [43,44,51,57]. These generalizations allow, for instance, to analyze the impact of anomalously large numbers of real eigenvalues on the spectra of random asymmetric matrices with some level correlations in the entries such as those in [12,35].

3. One of the consequences of the LDP was to provide an asymptotic for $\log p^n_{\alpha n}$ which scales as $-n^2\mathcal{I}[\mu_\alpha]$. To our knowledge, this is the first derivation of large n estimate for p^n_k.
4. The theoretical and numerical analysis of μ_α established that, unlike the circular law which is supported by a disk, the measure μ_α has two distinct components. The first is supported by a compact set of the complex plane that is well separated from the real line, and upon which μ_α has uniform density of mass $1-\alpha$. The second is supported by a segment in the real plane and has a density w.r.t. Lebesgue's measure with mass α. As α increases to one, the support of the former shrinks to a single point and its complex conjugate whereas the latter tends to the semi-circular law on $[-\sqrt{2}, \sqrt{2}]$.
5. The microscopic characterization of the particle distributions at zero temperature through the renormalized energy reveals that, (i) in the complex plane, particles organize in an Abrikosov lattice, similar to the unmixed 2d gas, yet (ii) on the real line, they are crystallized similarly to the Gaussian Orthogonal Ensemble in the zero temperature limit, but unlike real eigenvalues of unconstrained real Ginibre matrices.

The last two points above establish that the real Ginibre ensemble constrained by $k/n \to \alpha$ interpolates between the circular and semi-circular law as α shifts from zero to one at the macroscopic level. At the microscopic level, it is a mixture of the two extreme cases. This interpolation is distinct from the ones in which the circular law is progressively flattened on the real line with intermediate elliptic like support for the spectra [12,35]. It reveals some of the rich characteristics of real random matrices due to their spectrum containing both real and complex eigenvalues.

Acknowledgments We thank an anonymous referee for his suggestions on the proof of Theorem 1.

Appendix

Monte Carlo Algorithm for the Eigenvalues

An efficient method to approximate numerically the minimizer μ_α and the probability distribution of the proportion of real eigenvalues is to use the Metropolis–Hastings Monte Carlo algorithm. This method consists in constructing an ergodic Markov chain whose stationary distribution is given by (1). Here, we evolve a n-particles system z_t, but in contrast to the log-gas, the dynamics is now discrete, and the transition probability is based on the pdf (1): a new configuration z^* is drawn by modifying one of the eigenvalues at random and the Markov chain has a transition towards z^* if $Q^n(z^*) > Q^n(z_t)$, and otherwise according to a Bernoulli variable of parameter $\frac{Q^n(z^*)}{Q^n(z_t)}$.

When conditioning on very rare events, (here for instance, a fixed number of real eigenvalues), cases satisfying the constraints have an extremely low probability of being explored, and more refined methods need to be developed in order to access these probabilities. In the present case, the problem is considerably simplified since we dispose of an explicit form of the distribution of the eigenvalues under our constraint. Indeed, the joint probability distribution of Ginibre matrices of size n constrained on having k real eigenvalues $(\lambda_i \ ; \ i = 1 \dots k)$ (and therefore $l = (n-k)/2$ pairs of complex eigenvalues $(z_i, i = 1 \dots n-k)$) is given by:

$$\mathbb{P}\big[\lambda_1 \dots \lambda_k, z_1, \dots, z_{n-k}\big] = \tilde{C}_n \prod_{i>j} |\lambda_i - \lambda_j| \prod_{i>j} |z_i - z_j| \prod_{i,j} |\lambda_i - z_j| \times \left(\prod_{i=1}^{k} \exp(-\lambda_i^2) \prod_{i=1}^{n-k} \exp(-z_i^2) \operatorname{erfc}(|z_i - z_i^*|/\sqrt{2}) \right)^{1/2},$$

where the coefficient C_n can be found in [14].

Classical Metropolis–Hastings algorithm with Gaussian transitions preserving the nature of the system therefore allow to access directly the distribution of eigenvalues and the probability $p(n, k)$ of the event considered.

References

1. Forrester, P.J.: London Mathematical Society Monographs. Princeton University Press, Princeton (2010)
2. Tao, T.: Topics in Ramdom Matrix Theory. American Mathematical Society, Providence (2012)
3. Bordenave, C., Chafai, D.: Around the circular law. Probab. Surv. **93**, 1–89 (2012)
4. Wigner, E.P.: Lower limit for the energy derivative of the scattering phase shift. Phys. Rev. **98**(1), 145–147 (1955)
5. Auffinger, A., Ben Arous, G., Černỳ, J.: Random matrices and complexity of spin glasses. Commun. Pure Appl. Math. **66**(2), 165–201 (2013)
6. May, R.M.: Will a large complex system be stable? Nature **238**, 413–414 (1972)
7. Wainrib, G., Touboul, J.: Topological and dynamical complexity of random neural networks. Phys. Rev. Lett. **110**(11), 118101 (2013)
8. del Molino, L.C.G., Pakdaman, K., Touboul, J., Wainrib, G.: Synchronization in random balanced networks. Phys. Rev. E **88**(4), 042824 (2013)
9. Couillet, R., Debbah, M., et al.: Random Matrix Methods for Wireless Communications. Cambridge University Press, Cambridge (2011)
10. Jaeger, H., Haas, H.: Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science **304**(5667), 78–80 (2004)
11. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
12. Lehmann, N., Sommers, H.J.: Eigenvalue statistics of random real matrices. Phys. Rev. Lett. **67**, 941–944 (1991)
13. Edelman, A.: The probability that a random real Gaussian matrix has k real Eigenvalues, related distributions, and the circular law. J. Multivar. Anal. **60**, 203–232 (1997)
14. Forrester, P.J., Nagao, T.: Eigenvalue statistics of the real Ginibre enesmble. Phys. Rev. Lett. **99**, 050603 (2007)
15. Sommers, H.J.: Symplectic structure of the real Ginibre ensemble. J. Phys. A **40**(29), F671 (2007)
16. Forrester, P.J., Nagao, T.: Skew orthogonal polynomials and the partly symmetric real Ginibre ensemble. J. Phys. A **41**(37), 375003 (2008)
17. Forrester, P.J., Mays, A.: A method to calculate correlation functions for $\beta = 1$ random matrices of odd size. J. Phys. A **134**(3), 443–462 (2009)
18. Sinclair, C.: Correlation functions for $\beta = 1$ ensembles of matrices of odd size. J. Stat. Phys. **136**(1), 17–33 (2009)
19. Sommers, H.J., Wieczorek, W.: General eigenvalue correlations for the real Ginibre ensemble. J. Phys. A **41**(40), 405003 (2008)
20. Borodin, A., Sinclair, C.: The Ginibre ensemble of real random matrices and its scaling limits. Commun. Math. Phys. **291**(1), 177–224 (2009)
21. Rider, B., Sinclair, C.D., et al.: Extremal laws for the real ginibre ensemble. Ann. Appl. Probab. **24**(4), 1621–1651 (2014)
22. Tao, T., Vu, V.: Random matrices: universality of ESD and the circular law. Ann. Probab. **38**(5), 2023–2065 (2010)
23. Tao, T., Vu, V.: Random matrices: universality of local spectral statistics of non-hermitian matrices (2012). arXiv:1206.1893
24. Bourgade, P., Erdos, L., Yau, H.: Universality of general *beta*-ensembles. Duke Mathematical Journal **163**(6), 1127–1190 (2014)

25. Bourgade P., Erdos, L., Yau, H., Yin, J.: Fixed energy universality for generalized Wigner matrices (2014). arXiv:1407.5606
26. Erdös, L., Yau, H.-T., Yin, J.: Rigidity of eigenvalues of generalized wigner matrices. Adv. Math. **229**(3), 1435–1515 (2012)
27. Bourgade, P., Erdös, L., Yau, H.-T.: Bulk universality of general β-ensembles with non-convex potential. J. Math. Phys. **53**(9), 095221 (2012)
28. Bourgade, P., Erdös, L., Yau, H.-T.: Edge universality of beta ensembles. Commun. Math. Phys. **332**(1), 261–353 (2014)
29. Bourgade, P., Yau, H.-T., Yin, J.: Local circular law for random matrices. Probab. Theory Relat. Fields **159**(3–4), 545–595 (2014)
30. Bourgade, P., Yau, H.-T., Yin, J.: The local circular law ii: the edge case. Probab. Theory Relat. Fields **159**(3–4), 619–660 (2014)
31. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. Probab. Theory Relat. Fields **108**, 517–542 (1997)
32. Ben Arous, G., Zeitouni, O.: Large Deviations from the circular law. ESAIM: Probab. Stat. **2**, 123–134 (1998)
33. Edelman, A., Kostlan, E., Shub, M.: How many Eigenvalues of a random matrix are real? J. Am. Math. Soc. **7**, 247–267 (1994)
34. Ginibre, J.: Statistical ensembles of complex, quaternion, and real matrices. J. Math. Phys. **6**, 440–449 (1965)
35. Sommers, H.J., Crisanti, A., Sompolinsky, H.: Spectrum of large random asymmetric matrices. Phys. Rev. Lett. **60**, 1895 (1988)
36. Tribe, R., Zaboronski, O.: Pfaffian formulae for one dimensional coalescing and annihilating systems. Electron. J. Probab. **163**(76), 2080–2103 (2011)
37. Forrester, P.J.: Diffusion processes and the asymptotic bulk gap probability for the real Ginibre ensemble. arXiv:1306.4106
38. Beenakker, C.: Random-matrix theory of majorana fermions and topological superconductors (2014). arXiv:1407.2131
39. Kanzieper, E., Akemann, G.: Statistics of real eigenvalues in Ginibre's ensemble of random real matrices. Phys. Rev. Lett. **95**(23), 230201 (2005)
40. Akemann, G., Kanzieper, E.: Integrable structure of Ginibre's ensemble of real random matrices and a Pfaffian integration theorem. J. Stat. Phys. **129**(5–6), 1159–1231 (2007)
41. Kanzieper, E., Poplavskyi, M., Timm, C., Tribe, R., Zaboronski, O.: What is the probability that a large random matrix has no real eigenvalues? (2015). arXiv:1503.07926
42. Dyson, F.J.: A Brownian-motion model for the Eigenvalues of a random matrix. J. Math. Phys. **3**, 1191–1198 (1962)
43. Majumdar, S.N., Nadal, C., Scardicchio, A., Vivo, P.: Index distribution of gaussian random matrices. Phys. Rev. Lett. **103**(22), 220603 (2009)
44. Majumdar, S.N., Vivo, P.: Number of relevant directions in principal component analysis and wishart random matrices. Phys. Rev. Lett. **108**(20), 200601 (2012)
45. Rogers, L., Shi, Z.: Interacting Brownian particles and the Wigner law. Probab. Theory Relat. Fields **95**, 555–570 (1993)
46. Cepa, E., Lepingle, D.: Diffusing particles with electrostatic repulsion. Probab. Theory Relat. Fields **107**, 429–449 (1997)
47. Anderson, W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Cambridge University Press, Cambridge (2009)
48. Sandier, E., Serfaty, S.: 1d log gases and the renormalized energy: crystallization at vanishing temperature. Probab. Theory Relat Fields **162**, 1–52 (2014)
49. Sandier, E., Serfaty, S.: 2d coulomb gases and the renormalized energy (2012). arXiv:1201.3503
50. Rougerie, N., Serfaty, S.: Higher dimensional coulomb gases and renormalized energy functionals (2013). arXiv:1307.2805
51. Allez, R., Touboul, J., Wainrib, G.: Index distribution of the Ginibre ensemble. J. Phys. A **47**(4), 042001 (2014)
52. Braides, A.: Gamma-convergence for Beginners. Oxford University Press, Oxford (2002)
53. Armstrong, S.N., Serfaty, S., Zeitouni, O.: Remarks on a constrained optimization problem for the ginibre ensemble. Potential Anal. **41**(3), 945–958 (2014)
54. Abrikosov, A.A.: On the magnetic properties of superconductors of the second type. Sov. Phys. JETP **5**, 1174–1182 (1957)
55. Chafaï, D., Gozlan, N., Zitt, P.-A.: First order global asymptotics for confined particles with singular pair repulsion (2013). arXiv:1304.7569

56. del Molino, L.C.G., Pakdaman, K., Touboul, J.: The heterogeneous gas with singular interaction: generalized circular law and heterogeneous renormalized energy. J. Phys. A **48**(4), 045208 (2015)
57. Vivo, P., Majumdar, S.N., Bohigas, O.: Large deviations and random matrices. Acta Phys. Pol. B **38**(13), 4139 (2007)