Eigenvalue density of the Wishart matrix and large deviations[1]

Fumio Hiai[*,⋆]
Department of Mathematical Sciences
Ibaraki University
Mito, Ibaraki 310, Japan

Dénes Petz[*,◇]
Department for Mathematical Analysis
Technical University of Budapest
H-1521 Budapest XI. Sztoczek u. 2, Hungary

Abstract. A large deviation theorem is obtained for a certain sequence of random measures which includes the empirical eigenvalue distribution of Wishart matrices, as the matrix size tends to infinity. The rate function is convex and one of its ingredients is the logarithmic energy. In case of the singular Wishart matrix, the limit distribution has an atom and the rate function is infinite on absolute continuous measures.

Mathematics Subject Classification: 15A52, 60F05, 60H25

1. Introduction. It has been observed that the empirical eigenvalue density of certain random matrices converges to a non-random limit as the matrix size tends to infinity. This was observed first for Gaussian symmetric matrices, where the non-random limit, called also the integrated density of states, is the semicircle law. If T is a $p \times n$ random matrix with independent and identically distributed entries of mean zero, then $n^{-1}TT^t$ can be viewed as a sample covariance matrix of n samples of p-dimensional random vectors and it is of fundamental importance in multivariate statistical analysis. When the entries are normally distributed, the sample covariance matrix is called the Wishart matrix ([1]). It is known ([7], [17]) that the empirical distribution of the eigenvalues of the Wishart matrix converges almost surely to a non-random probability distribution as $n \to \infty$ and $p/n \to a > 0$. The limit distribution is continuous and supported on $[(1-\sqrt{a})^2, (1+\sqrt{a})^2]$ when $0 < a \le 1$, and it places a mass $1 - 1/a$ at 0 when $a > 1$. The limit is the so-called Marchenko-Pastur distribution [9]. It is remarkable that this distribution appeared also as the

[1] Published in *Infinite Dimensional Anal. Quantum Prob.* **1**(1998), 633–646
[*] Supported by OTKA F023447.
[⋆] Supported by Grant-in-Aid for Scientific Research (C)09640152.
[◇] Supported by the Canon Foundation.

limit in the free probabilistic analogue of the Poisson limit theorem ([16]). We note that [8] contains other examples of integrated density of states.

After some indications that the convergence to the integrated density of states is very fast, the first large deviation theorem for the empirical eigenvalue density was proven by Ben Arous and Guionnet [2] in the case of Gaussian random symmetric matrices. The main component of the rate function is the so-called logarithmic energy which appeared also in the work of Voiculescu [15] about random matrices and the entropy of free random variables. Afterwards, in [6] and [13], we proved further large deviation results for random unitary and Gaussian non-symmetric matrices. In this paper we apply the same method to the case of Wishart matrices. We benefit from the exact form of the joint eigenvalue density of the Wishart matrix. In fact, we treat more general random matrices than the Wishart matrix, their distribution with respect to the density of the Wishart matrix contains a continous parameter function $Q(x)$.

The paper is organized as follows. In Sec. 2 after a few preliminaries on Wishart matrices, we state a large deviation theorem for random probability measures more general than those arising from Wishart matrices. The rate function contains a logarithmic energy part but also a linear functional part. Full Sec. 3 is devoted to the proof of the theorem. We follow the strategy of Ben Arous and Guionnet, the exponential tightness and the weak large deviation principle are shown. In Sec. 4 we specialize our result to obtain the large deviation for the empirical eigenvalue distribution of the non-singular Wishart matrix. The minimizer of the rate function is the Marchenko-Pastur distribution, that is the limiting eigenvalue distribution. Finally, the singular case is treated in Sec. 5. An interesting feature of this case is that the limiting measure has an atom and the rate function is infinite on the absolute continuous measures.

2. A large deviation theorem. For each $n \in \mathbb{N}$ let $p(n) \le n$ be a positive integer. Let $T(n)$ be a $p(n) \times n$ real random matrix whose entries are independent and have the same distribution $N(0,1)$. Then the $p(n) \times p(n)$ random matrix $T(n)T(n)^t$ is known as the *non-singular Wishart matrix.* We are interested in the eigenvalues of $n^{-1}T(n)T(n)^t$, and it is known (see [1], p. 534) that the (random) eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_{p(n)})$ have the joint probability density

$$\frac{1}{Z_n} \exp\left(-\frac{n}{2}\sum_{i=1}^{p(n)} t_i\right) \prod_{i=1}^{p(n)} t_i^{(n-p(n)-1)/2} \prod_{1\le i<j\le p(n)} |t_i - t_j|\,, \tag{1}$$

where Z_n is a constant for normalization. The random discrete measure

$$P_n := \frac{1}{p(n)}\big(\delta(\lambda_1) + \delta(\lambda_2) + \cdots + \delta(\lambda_{p(n)})\big)$$

is called the *empirical eigenvalue distribution* of $n^{-1}T(n)T(n)^t$. Above $\delta(\lambda)$ denotes the Dirac measure at λ. Let $\mathcal{M}(\mathbb{R}^+)$ denote the space of all probability measures on $\mathbb{R}^+$ endowed with the weak topology (see [4]). As usual in probability theory, the random measure P_n is sometimes identified with its distribution on $\mathcal{M}(\mathbb{R}^+)$ given as $R_n(\Gamma) := \mathbf{Prob}(P_n \in \Gamma)$ for Borel subsets Γ of $\mathcal{M}(\mathbb{R}^+)$.

The subject of the paper is to obtain a large deviation theorem for the sequence (P_n). It is known ([7], [17]) that if $\lim_{n\to\infty} p(n)/n = a > 0$ then P_n converges almost everywhere to

$$\mu_a := \begin{cases} \dfrac{\sqrt{4a-(x-1-a)^2}}{2\pi a x}\chi(x)\,dx & \text{if } 0 < a \le 1, \\ (1-a^{-1})\delta(0) + \dfrac{\sqrt{4a-(x-1-a)^2}}{2\pi a x}\chi(x)\,dx & \text{if } a > 1, \end{cases} \tag{2}$$

where χ denotes the characteristic function of the interval $[(1-\sqrt{a})^2, (1+\sqrt{a})^2]$. The distribution (2) is called after *Marchenko-Pastur* due to the paper [9] (see also [12]). We refer to the monographs [3] and [4] concerning the general theory of large deviations.

Our proof will not be much more complicated if we assume a joint density more general than (1). Let $Q(x)$ be a real continuous function on $\mathbb{R}^+$ such that for any $\varepsilon > 0$

$$\lim_{x\to+\infty} x\exp(-\varepsilon Q(x)) = 0\,. \tag{3}$$

For each $n \in \mathbb{N}$ assume that the joint probability density of $(\lambda_1, \lambda_2, \dots, \lambda_{p(n)})$ is of the form

$$\nu_n := \frac{1}{Z_n}\exp\left(-n\sum_{i=1}^{p(n)} Q(t_i)\right)\prod_{i=1}^{p(n)} t_i^{\gamma(n)} \prod_{1\le i<j\le p(n)} |t_i - t_j|^{2\beta}\,, \tag{4}$$

where $\beta > 0$ is fixed and $\gamma(n) \ge 0$ depends on n. When κ_t denotes the discrete measure $p(n)^{-1}\big(\delta(t_1)+\delta(t_2)+\cdots+\delta(t_{p(n)})\big)$ for $t = (t_1, t_2, \dots, t_{p(n)})$, the empirical distribution P_n of $(\lambda_1, \lambda_2, \dots, \lambda_{p(n)})$ may be given as

$$P_n(\Gamma) = \nu_n\big(\{t \in (\mathbb{R}^+)^{p(n)} : \kappa_t \in \Gamma\}\big)$$

for every Borel subset Γ of $\mathcal{M}(\mathbb{R}^+)$. Our large deviation theorem will be shown for this sequence (P_n).

Theorem 1. Assume that $p(n)/n \to a \in (0,1]$ and $\gamma(n)/n \to \gamma > 0$ as $n \to \infty$. Then the finite limit $B := \lim_{n\to\infty} n^{-2}\log Z_n$ exists and (P_n) satisfies the large deviation principle at the scale n^{-2} with good rate function

$$I(\mu) := -a^2\beta\iint \log|x-y|\,d\mu(x)\,d\mu(y) + a\int (Q(x) - \gamma\log x)\,d\mu(x) + B \tag{5}$$

for $\mu \in \mathcal{M}(\mathbb{R}^+)$. Moreover, there exists a unique $\mu_0 \in \mathcal{M}(\mathbb{R}^+)$ such that $I(\mu_0) = 0$.

In the rate function (5) the double integral

$$\Sigma(\mu) := \iint \log|x-y|\, d\mu(x)\, d\mu(y)$$

is called the *free entropy* of μ, which arised from Voiculescu's work on free probability theory. Note that $-\Sigma(\mu)$ is the *logarithmic energy* of μ and it is familiar in potential theory.

3. Proof of theorem. To prove the theorem, let us introduce the kernel functions on $(\mathbb{R}^+)^2$ as follows:

$$F(x,y) := -a^2\beta \log|x-y| + \frac{a}{2}(Q(x)+Q(y)) - \frac{a\gamma}{2}(\log x + \log y)\,,$$

$$\tilde{F}_n(x,y) := -\frac{p(n)^2}{n^2}\beta \log|x-y| + \frac{p(n)}{2n}(Q(x)+Q(y)) - \frac{p(n)\gamma(n)}{2n^2}(\log x + \log y)\,.$$

Furthermore, for $R > 0$ their truncated versions are defined by

$$F_R(x,y) := \min\{F(x,y), R\}\,, \quad \tilde{F}_{n,R}(x,y) := \min\{F_n(x,y), R\}\,.$$

Since

$$F(x,y) \ge -\frac{a(2a\beta+\gamma)}{2}\left[\log\Big(x\exp\Big(-\frac{Q(x)}{2a\beta+\gamma}\Big)\Big) + \log\Big(y\exp\Big(-\frac{Q(y)}{2a\beta+\gamma}\Big)\Big)\right] \tag{6}$$

(and similarly for $\tilde{F}_n(x,y)$) whenever $x,y \ge 2$, it follows from (3) that $F_R(x,y)$ is bounded and continuous, so that

$$-a^2\beta\Sigma(\mu) + a\int (Q(x) - \gamma\log x)\, d\mu(x) = \iint F(x,y)\, d\mu(x)\, d\mu(y)$$

is a well-defined and lower semicontinuous functional on $\mathcal{M}(\mathbb{R}^+)$.

We know from the theory of weighted potential ([11]) that there exist unique $\mu_0, \mu_n \in \mathcal{M}(\mathbb{R}^+)$ such that

$$\iint F(x,y)\, d\mu_0(x)\, d\mu_0(y) = \inf_{\mu\in\mathcal{M}(\mathbb{R}^+)} \iint F(x,y)\, d\mu(x)\, d\mu(y)\,,$$

$$\iint \tilde{F}_n(x,y)\, d\mu_n(x)\, d\mu_n(y) = \inf_{\mu\in\mathcal{M}(\mathbb{R}^+)} \iint \tilde{F}_n(x,y)\, d\mu(x)\, d\mu(y)\,.$$

Moreover, the minimizers μ_0 and μ_n of the energy integrals $\iint F(x,y)\, d\mu(x)\, d\mu(y)$ and $\iint \tilde{F}_n(x,y)\, d\mu(x)\, d\mu(y)$ have compact supports (see [11] or [14], p. 27).

Lemma 2. For any $R > 0$, $\tilde{F}_{n,R}(x,y) \to F_R(x,y)$ uniformly as $n \to \infty$.

Proof. Using (6) for F as well as for $\tilde{F}_n$ and the assumptions on $p(n)$ and $\gamma(n)$, we can see that for any $R > 0$ there exists $\delta > 0$ such that if $(x,y) \notin [\delta, \delta^{-1}] \times [\delta, \delta^{-1}]$

then $F(x,y) \ge R$ and $\tilde F_n(x,y) \ge R$ for all n. Furthermore, since $\log x$ and $Q(x)$ are bounded on $[\delta, \delta^{-1}]$, $\delta_1 > 0$ can be chosen so that if (x,y) does not belong to

$$\Delta := \{(x,y) : \delta \le x \le \delta^{-1},\, \delta \le y \le \delta^{-1},\, |x-y| \ge \delta_1\},$$

then $F(x,y) \ge R$ and $\tilde F_n(x,y) \ge R$ for all n. It is obvious that $\tilde F_n(x,y) \to F(x,y)$ as $n \to \infty$ uniformly on Δ. Hence the conclusion follows. □

Lemma 3. (μ_n) is tight and

$$\iint F(x,y)\, d\mu_0(x)\, d\mu_0(y) \le \liminf_{n\to\infty} \iint \tilde F_n(x,y)\, d\mu_n(x)\, d\mu_n(y). \tag{7}$$

Proof. It is clear that $\iint \tilde F_n(x,y)\, d\mu_n(x)\, d\mu_n(y) \le b$ $(n \in \mathbb{N})$ for some $b < +\infty$. Also, by assumption (3) and estimate (6) for $\tilde F_n$, there exists $c < +\infty$ such that $\tilde F_n(x,y) \ge -c$ for all $x, y \in \mathbb{R}^+$ and $n \in \mathbb{N}$. For $\alpha > 0$ let

$$M_\alpha := \inf\{\tilde F_n(x,y) : n \in \mathbb{N},\, x, y \ge \alpha\}.$$

Then, by (6) for $\tilde F_n$ again, M_α can be arbitrarily large when $\alpha \to +\infty$. Since

$$b \ge M_\alpha \mu_n([\alpha,\infty))^2 - c \qquad (n \in \mathbb{N}),$$

we have $\sup_n \mu_n([\alpha,\infty)) \to 0$ as $\alpha \to +\infty$, which means the tightness of (μ_n).

Thanks to the tightness (or relative weak compactness) of (μ_n) we can choose a subsequence $(\mu_{n(m)})$ such that $\mu_{n(m)} \to \bar\mu$ weakly for some $\bar\mu \in \mathcal{M}(\mathbb{R}^+)$ and

$$\lim_{m\to\infty} \iint \tilde F_{n(m)}(x,y)\, d\mu_{n(m)}\, d\mu_{n(m)}(y) = \liminf_{n\to\infty} \iint \tilde F_n(x,y)\, d\mu_n(x)\, d\mu_n(y).$$

Then

$$\begin{aligned}\iint F(x,y)\, d\mu_0(x)\, d\mu_0(y) &\le \iint F(x,y)\, d\bar\mu(x)\, d\bar\mu(y)\\ &= \sup_{R>0} \iint F_R(x,y)\, d\bar\mu(x)\, d\bar\mu(y)\\ &= \sup_{R>0} \lim_{m\to\infty} \iint \tilde F_{n(m),R}(x,y)\, d\mu_{n(m)}(x)\, d\mu_{n(m)}(y)\\ &\qquad\qquad \text{(by Lemma 2)}\\ &\le \lim_{m\to\infty} \iint \tilde F_{n(m)}(x,y)\, d\mu_{n(m)}(x)\, d\mu_{n(m)}(y),\end{aligned}$$

showing (7). □

Lemma 4.

$$\limsup_{n\to\infty} \frac{1}{n^2} \log Z_n \le -\iint F(x,y)\, d\mu_0(x)\, d\mu_0(y). \tag{8}$$

Proof. We estimate

$$
\begin{aligned}
Z_n &= \int \cdots \int \exp\biggl(\frac{n}{p(n)} \sum_{i=1}^{p(n)} \Bigl(-Q(t_i) + \frac{\gamma(n)}{n} \log t_i\Bigr)\biggr) \\
&\qquad\qquad \times \exp\biggl(-\frac{2n^2}{p(n)^2} \sum_{1 \le i < j \le p(n)} \tilde{F}_n(t_i, t_j)\biggr) \, dt_1 \cdots dt_{p(n)} \\
&\le \biggl[\int \exp\biggl(\frac{n}{p(n)} \Bigl(-Q(x) + \frac{\gamma(n)}{n} \log x\Bigr)\biggr) \, dx\biggr]^{p(n)} \\
&\qquad\qquad \times \exp\biggl(-n^2 \iint \tilde{F}_n(x, y) \, d\mu_n(x) \, d\mu_n(y)\biggr) .
\end{aligned}
$$

Since by assumption (3)

$$
\sup_{n \ge 1} \int \exp\biggl(\frac{n}{p(n)} \Bigl(-Q(x) + \frac{\gamma(n)}{n} \log x\Bigr)\biggr) \, dx < +\infty ,
$$

the above estimate implies that

$$
\begin{aligned}
\limsup_{n \to \infty} \frac{1}{n^2} \log Z_n &\le - \liminf_{n \to \infty} \iint \tilde{F}_n(x, y) \, d\mu_n(x) \, d\mu_n(y) \\
&\le - \iint F(x, y) \, d\mu_0(x) \, d\mu_0(y)
\end{aligned}
$$

thanks to (7). □

Lemma 5. For every $\mu \in \mathcal{M}(\mathbb{R}^+)$,

$$
\begin{aligned}
&\inf_G \biggl[\limsup_{n \to \infty} \frac{1}{n^2} \log P_n(G)\biggr] \\
&\qquad \le - \iint F(x, y) \, d\mu(x) \, d\mu(y) - \liminf_{n \to \infty} \frac{1}{n^2} \log Z_n ,
\end{aligned}
\tag{9}
$$

where G runs over neighborhoods of μ in the weak topology.

Proof. For any neighborhood G of $\mu \in \mathcal{M}(\mathbb{R}^+)$ set $\tilde{G} := \bigl\{t \in (\mathbb{R}^+)^{p(n)} : \kappa_t \in G\bigr\}$. Then from the density (4) we get

$$
\begin{aligned}
P_n(G) &= \nu_n(\tilde{G}) \\
&= \frac{1}{Z_n} \int \cdots \int_{\tilde{G}} \exp\biggl(\frac{n}{p(n)} \sum_{i=1}^{p(n)} \Bigl(-Q(t_i) + \frac{\gamma(n)}{n} \log t_i\Bigr)\biggr) \\
&\qquad\qquad \times \exp\biggl(-\frac{2n^2}{p(n)^2} \sum_{1 \le i < j \le p(n)} \tilde{F}_n(t_i, t_j)\biggr) \, dt_1 \cdots dt_{p(n)} \\
&\le \frac{1}{Z_n} \biggl[\int \exp\biggl(\frac{n}{p(n)} \Bigl(-Q(x) + \frac{\gamma(n)}{n} \log x\Bigr)\biggr) \, dx\biggr]^{p(n)} \\
&\qquad\qquad \times \exp\biggl(-n^2 \inf_{\mu' \in G} \iint \tilde{F}_{n,R}(x, y) \, d\mu'(x) \, d\mu'(y) + nR\biggr)
\end{aligned}
$$

for any $R > 0$. Furthermore, by Lemma 2 we can obtain

$$\lim_{n\to\infty}\left(\inf_{\mu'\in G}\iint \tilde{F}_{n,R}(x,y)\,d\mu'(x)\,d\mu'(y)\right) = \inf_{\mu'\in G}\iint F_R(x,y)\,d\mu'(x)\,d\mu'(y)\,,$$

so that

$$\limsup_{n\to\infty}\frac{1}{n^2}\log P_n(G) \le -\inf_{\mu'\in G}\iint F_R(x,y)\,d\mu'(x)\,d\mu'(y) - \liminf_{n\to\infty}\frac{1}{n^2}\log Z_n\,.$$

Thanks to the continuity of $\mu' \mapsto \iint F_R(x,y)\,d\mu'(x)\,d\mu'(y)$ it follows that

$$\inf_G\left[\limsup_{n\to\infty}\frac{1}{n^2}\log P_n(G)\right] \le -\iint F_R(x,y)\,d\mu(x)\,d\mu(y) - \liminf_{n\to\infty}\frac{1}{n^2}\log Z_n\,.$$

which yields (9) as $R \to +\infty$. □

Lemma 6.

$$\liminf_{n\to\infty}\frac{1}{n^2}\log Z_n \ge -\iint F(x,y)\,d\mu_0(x)\,d\mu_0(y)\,, \tag{10}$$

and for every $\mu \in \mathcal{M}(\mathbb{R}^+)$

$$\begin{aligned}&\inf_G\left[\liminf_{n\to\infty}\frac{1}{n^2}\log P_n(G)\right]\\ &\qquad \ge -\iint F(x,y)\,d\mu(x)\,d\mu(y) - \limsup_{n\to\infty}\frac{1}{n^2}\log Z_n\,,\end{aligned} \tag{11}$$

where G runs over neighborhoods of μ in the weak topology.

Proof. By a standard regularization argument, we may assume that the support of μ is a finite interval $[0,R]$ and μ has a continuous density $f > 0$ on $[0,R]$. Hence $\delta \le f(x) \le \delta^{-1}$ $(0 \le x \le R)$ for some $\delta > 0$. For each $n \in \mathbb{N}$ let $0 = r_1^{(n)} < s_1^{(n)} < r_2^{(n)} < s_2^{(n)} < \ldots < r_{p(n)}^{(n)} < s_{p(n)}^{(n)} = R$ be such that

$$\int_0^{r_i^{(n)}} f(x)\,dx = \frac{i-\frac{1}{2}}{p(n)}\,,\qquad \int_0^{s_i^{(n)}} f(x)\,dx = \frac{i}{p(n)}\qquad (1 \le i \le p(n)).$$

We get

$$\frac{\delta}{2p(n)} \le s_i^{(n)} - r_i^{(n)} \le \frac{1}{2p(n)\delta}\qquad (1 \le i \le p(n)).$$

Define

$$\Delta_n := \{(t_1,\ldots,t_{p(n)}) \in (\mathbb{R}^+)^{p(n)} : r_i^{(n)} \le t_i \le s_i^{(n)},\ 1 \le i \le p(n)\}\,.$$

For any neighborhood G of μ, if n is large enough, then we have $\Delta_n \subset \{t \in (\mathbb{R}^+)^{p(n)} : \kappa_t \in G\}$ and so

$$
\begin{aligned}
P_n(G) &\ge \nu_n(\Delta_n) \\
&= \frac{1}{Z_n}\int\cdots\int_{\Delta_n} \exp\left(-n\sum_{i=1}^{p(n)} Q(t_i)\right)\prod_{i=1}^{p(n)} t_i^{\gamma(n)} \\
&\qquad\qquad \times \prod_{1\le i<j\le p(n)} (r_j^{(n)} - s_i^{(n)})^{2\beta} \int\cdots\int_{\Delta_n} dt_1\cdots dt_{p(n)} \\
&\ge \frac{1}{Z_n}\left(\frac{\delta}{2p(n)}\right)^{p(n)} \exp\left(-n\sum_{i=1}^{p(n)} \xi_i^{(n)}\right)\prod_{i=1}^{p(n)} (r_i^{(n)})^{\gamma(n)} \\
&\qquad\qquad \times \prod_{1\le i<j\le p(n)} (r_j^{(n)} - s_i^{(n)})^{2\beta}\,,
\end{aligned}
$$

where $\xi_i^{(n)} := \max\{Q(x) : r_i^{(n)} \le x \le s_i^{(n)}\}$. The following are easy to check:

$$
\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{p(n)} \xi_i^{(n)} = a\int Q(x)\,d\mu(x)\,,
$$

$$
\lim_{n\to\infty} \frac{\gamma(n)}{n^2}\sum_{i=1}^{p(n)} \log r_i^{(n)} = a\gamma\int \log x\,d\mu(x)\,,
$$

$$
\lim_{n\to\infty} \frac{2\beta}{n^2}\sum_{1\le i<j\le p(n)} \log(r_j^{(n)} - s_i^{(n)}) = a^2\beta\Sigma(\mu)\,.
$$

Therefore

$$
0 \ge \limsup_{n\to\infty} \frac{1}{n^2}\log P_n(G) \ge -\iint F(x,y)\,d\mu(x)\,d\mu(y) - \liminf_{n\to\infty}\frac{1}{n^2}\log Z_n\,,
$$

which implies (10) by taking the infimum for μ. Also, we obtain

$$
\liminf_{n\to\infty} \frac{1}{n^2}\log P_n(G) \ge -\iint F(x,y)\,d\mu(x)\,d\mu(y) - \limsup_{n\to\infty}\frac{1}{n^2}\log Z_n\,.
$$

□

It is immediate from (8) and (10) that the finite limit $B := \lim_{n\to\infty} n^{-2}\log Z_n$ exists.

Lemma 7. (P_n) is exponentially tight.

Proof. Since $Q(x) \to +\infty$ as $x \to +\infty$ by assumption (3), it is easy to see that for any $\alpha > 0$ the set

$$
K_\alpha := \left\{\mu \in \mathcal{M}(\mathbb{R}^+) : \int Q(x)\,d\mu(x) \le \alpha\right\}
$$

is compact in the weak topology. We have

$$
\begin{aligned}
P_n(K_\alpha^c) &= \nu_n\big(\{t \in (\mathbb{R}^+)^{p(n)} : \kappa_t \notin K_\alpha\}\big) \\
&= \frac{1}{Z_n} \int \cdots \int_{\left\{\frac{1}{p(n)}\sum_{i=1}^{p(n)} Q(t_i) > \alpha\right\}} \exp\left(-n \sum_{i=1}^{p(n)} Q(t_i)\right) \\
&\qquad\qquad \times \prod_{i=1}^{p(n)} t_i^{\gamma(n)} \prod_{1\le i<j\le p(n)} |t_i - t_j|^{2\beta}\, dt_1 \cdots dt_n \\
&\le \frac{1}{Z_n} \exp\left(-\frac{np(n)\alpha}{2}\right) \int \cdots \int \exp\left(-\frac{n}{2} \sum_{i=1}^{p(n)} Q(t_i)\right) \\
&\qquad\qquad \times \prod_{i=1}^{p(n)} t_i^{\gamma(n)} \prod_{1\le i<j\le p(n)} |t_i - t_j|^{2\beta}\, dt_1 \cdots dt_n\,.
\end{aligned}
$$

When $Q(x)$ is replaced by $Q(x)/2$, the fact already shown says that the finite limit

$$
\begin{aligned}
B_2 := \lim_{n\to\infty} \frac{1}{n^2} \log \int \cdots \int \exp\left(-\frac{n}{2} \sum_{i=1}^{p(n)} Q(t_i)\right) \\
\times \prod_{i=1}^{p(n)} t_i^{\gamma(n)} \prod_{1\le i<j\le p(n)} |t_i - t_j|^{2\beta}\, dt_1 \cdots dt_n
\end{aligned}
$$

exists. Hence the above estimate gives

$$
\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(K_\alpha^c) \le -B + B_2 - \frac{\alpha}{2}\,.
$$

Since $\alpha > 0$ is arbitrary, we have the conclusion. □

Proof of Theorem 1. The proof is contained already in the previous lemmas. If we set $I(\mu) := \iint F(x,y)\, d\mu(x)\, d\mu(y) + B$ for $\mu \in \mathcal{M}(\mathbb{R}^+)$, then (9) and (11) imply that

$$
\begin{aligned}
I(\mu) &= \sup_{G\in\mathcal{A},\, \mu\in G} \left[-\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G)\right] \\
&= \sup_{G\in\mathcal{A},\, \mu\in G} \left[-\liminf_{n\to\infty} \frac{1}{n^2} \log P_n(G)\right]
\end{aligned}
$$

for a base $\mathcal{A}$ of the weak topology. This together with the exponential tightness, Lemma 7, is sufficient to conclude that the large deviation principle holds for (P_n) with good rate function I. The assertion about the minimizer was stated just before Lemma 2. □

4. The case of non-singular Wishart matrices. The joint eigenvalue distribution of the (renormalized) real Wishart matrix $n^{-1}T(n)T(n)^t$ with $p(n) \le n$ is distributed according to the density (1). This is a special case of (4) when $Q(x) = x/2$, $\beta = 1/2$, and $\gamma(n) = (n - p(n) - 1)/2$. So, when $p(n)/n \to a \in (0,1]$ as $n \to \infty$, the empirical eigenvalue distribution of $n^{-1}T(n)T(n)^t$ satisfies the large deviation principle with good rate function

$$I(\mu) := -\frac{a^2}{2}\Sigma(\mu) + \frac{a}{2}\int (x - (1-a)\log x)\, d\mu(x) + B\,. \tag{12}$$

As stated in Sec. 2, it is known that the limiting distribution is the (non-atomic) Marchenko-Pastur distribution μ_a from (2). Note that if (P_n) is a sequence of random probability measures on a Polish space and it satisfies the large deviation principle with good rate function having a unique minimizer μ_0, then (P_n) converges to μ_0 in the weak topology with probability 1. This can be easily shown by using the Lévy metric for the weak topology and the Borel-Cantelli lemma. Therefore, we see that μ_a is the minimizer of the rate function (12).

Thanks to the Selberg integral formula ([10], p. 354) the normalization constant Z_n in (1) is given as

$$\begin{aligned} Z_n &= n^{-np(n)} \int_0^\infty \cdots \int_0^\infty \exp\biggl(-\sum_{i=1}^{p(n)} x_i\biggr) \prod_{i=1}^{p(n)} x_i^{n-p(n)} \\ &\qquad\qquad \times \prod_{1\le i<j\le p(n)} (x_i - x_j)^2\, dx_1 \cdots dx_{p(n)} \\ &= n^{-np(n)} \prod_{j=1}^{p(n)} j!\,(n-j)!\,. \end{aligned}$$

Hence, by using the Stirling formula, B in (12) is computed as follows:

$$\begin{aligned} 2B &= \lim_{n\to\infty} \frac{1}{n^2} \log Z_n \\ &= \lim_{n\to\infty} \Biggl[\frac{p(n)^2}{n^2} \Biggl(\frac{1}{p(n)^2} \log \prod_{j=1}^{p(n)} j! - \frac{1}{2}\log p(n) \Biggr) + \Biggl(\frac{1}{n^2} \log \prod_{j=0}^{n-1} j! - \frac{1}{2}\log n \Biggr) \\ &\qquad - \frac{(n-p(n))^2}{n^2} \Biggl(\frac{1}{(n-p(n))^2} \log \prod_{j=0}^{n-p(n)-1} j! - \frac{1}{2}\log(n-p(n)) \Biggr) \\ &\qquad + \frac{p(n)^2}{2n^2} \log \frac{p(n)}{n} - \frac{(n-p(n))^2}{2n^2} \log \frac{n-p(n)}{n} \Biggr] \\ &= -\frac{3}{4}(a^2 + 1 - (1-a)^2) + \frac{a^2}{2}\log a - \frac{(1-a)^2}{2}\log(1-a) \\ &= -\frac{3}{2}a + \frac{a^2}{2}\log a - \frac{(1-a)^2}{2}\log(1-a)\,. \end{aligned}$$

This gives the exact value of B. Since $I(\mu_a) = 0$, we can obtain

$$\Sigma(\mu_a) = -1 + \frac{1}{2}(a^{-1} + \log a + (a^{-1} - 1)^2 \log(1-a))\,. \tag{13}$$

The above arguments are summarized in the following:

Theorem 8. When $p(n)/n \to a \in (0,1]$ as $n \to \infty$, the empirical eigenvalue distribution of the Wishart matrix $n^{-1}T(n)T(n)^t$ satisfies the large deviation principle at the scale n^{-2} with good rate function

$$
\begin{aligned}
I(\mu) := -\frac{a^2}{2}\Sigma(\mu) + \frac{a}{2}\int (x-(1-a)\log x)\,d\mu(x) \\
-\frac{1}{4}(3a - a^2\log a + (1-a)^2\log(1-a))
\end{aligned}
\tag{14}
$$

for $\mu \in \mathcal{M}(\mathbb{R}^+)$. Moreover, the unique minimizer of I is the Marchenko-Pastur distribution μ_a from (2) and its free entropy is (13).

It is worthwhile to note that there is another way of determining the minimizer of the rate function. In fact, it is not difficult to compute

$$
\int \log|x-y|\,d\mu_a(x)
\begin{cases}
= \frac{1}{2a}(x-(1-a)\log x) + C & \text{if } x \in \operatorname{supp}\mu_a, \\
< \frac{1}{2a}(x-(1-a)\log x) + C & \text{if } x \in \mathbb{R}^+ \setminus \operatorname{supp}\mu_a,
\end{cases}
$$

where C is a constant. According to the generalized Frostman theorem about weighted potentials [11] (also [14], p. 27), the above condition characterizes μ_a being the minimizer of the weighted energy functional (14). (Similar maximization problems for free entropy functionals were solved in [5].) Also, we know that the rate function is convex because $\Sigma(\mu)$ is a concave functional, see [5].

Let $X(n) = (\xi_{ij}(n))$ be a $p(n) \times n$ complex random matrix such that $\mathrm{Re}\,\xi_{ij}(n)$ and $\mathrm{Im}\,\xi_{ij}(n)$ are independent and have the same distribution $N(0,1)$. The choice $Q(x) = x$, $\beta = 1$, and $\gamma(n) = n - p(n)$ in (4) yields the joint eigenvalue distribution of the complex Wishart matrix $(2n)^{-1}X(n)X(n)^*$. When $p(n)/n \to a \in (0,1]$, this satisfies the large deviation principle as well and the rate function is the above (14) multiplied by a factor 2.

5. The case of singular Wishart matrices. Next we turn to the singular case $p(n) > n$. Note that the eigenvalues of $n^{-1}T(n)T(n)^t$ are those of $n^{-1}T(n)^tT(n)$ plus $p(n) - n$ zeros. Furthermore, it is obvious that the eigenvalue distribution of $p(n)^{-1}T(n)^tT(n)$ is given by (1) again, however $p(n)$ and n are interchanged. Hence the eigenvalue distribution of $n^{-1}T(n)^tT(n)$ has the joint density

$$
\frac{1}{Z_n}\exp\left(-\frac{n}{2}\sum_{i=1}^n t_i\right)\prod_{i=1}^n t_i^{(p(n)-n-1)/2}\prod_{1\le i<j\le n}|t_i - t_j|\,. \tag{15}
$$

In this way we know that the empirical eigenvalue distribution P_n of the singular Wishart matrix $n^{-1}T(n)T(n)^t$ is written as

$$
P_n = \left(1 - \frac{n}{p(n)}\right)\delta(0) + \frac{n}{p(n)}\tilde{P}_n\,,
$$

where $\tilde{P}_n$ is a random discrete measure $n^{-1}(\delta(\lambda_1) + \delta(\lambda_2) + \cdots + \delta(\lambda_n))$ such that $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is distributed according to the density (15).

From the regular case we can see that if $\lim_{n\to\infty} p(n)/n = a > 1$ then $\tilde{P}_n$ satisfies the large deviation principle with good rate function

$$\begin{aligned}\tilde{I}(\mu) := -\frac{1}{2}\Sigma(\mu) + \frac{1}{2}\int (x - (a-1)\log x)\, d\mu(x) \\ - \frac{1}{4}(3a - a^2 \log a + (a-1)^2 \log(a-1))\end{aligned} \tag{16}$$

for $\mu \in \mathcal{M}(\mathbb{R}^+)$. The minimizer of $\tilde{I}(\mu)$ is

$$\tilde{\mu}_a := \frac{\sqrt{4a - (x-1-a)^2}}{2\pi x}\chi(x)\, dx\,,$$

where χ is the same as in (2). In fact, note that if $D_a\mu := \mu(a\,\cdot\,)$ is the dilation of $\mu \in \mathcal{M}(\mathbb{R}^+)$, then $a^{-2}\tilde{I}(\mu)$ is equal to $I(\mu)$ of (14) with a^{-1} in place of a and on the other hand $D_a\tilde{\mu}_a$ is equal to $\mu_{a^{-1}}$.

Now the large deviation theorem related to singular Wishart matrices is stated in the following:

Theorem 9. Assume that $p(n) > n$ and $p(n)/n \to a > 1$ as $n \to \infty$. Then the empirical eigenvalue distribution of $n^{-1}T(n)T(n)^t$ satisfies the large deviation principle at the scale n^{-2} with good rate function

$$I(\mu) := \begin{cases} \tilde{I}(\tilde{\mu}) & \text{if } \mu = (1-a^{-1})\delta(0) + a^{-1}\tilde{\mu},\ \tilde{\mu} \in \mathcal{M}(\mathbb{R}^+), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\tilde{I}$ is given by (16). The unique minimizer of I is μ_a (with an atom) from (2).

In the light of the discussion before the theorem was stated, it is enough to prove the following lemma.

Lemma 10. For $n \in \mathbb{N}$ let $\tilde{P}_n$ be a random probability measure on a Polish space X. Let μ_0 be a fixed probability measure on X, and $0 < \alpha_n < 1$ be such that $\alpha_n \to \alpha \in (0,1)$. If $(\tilde{P}_n)$ is exponentially tight and satisfies the large deviation principle at the scale L_n with rate function $\tilde{I}$ on $\mathcal{M}(X)$, then the sequence of random measures $(1-\alpha_n)\mu_0 + \alpha_n\tilde{P}_n$ satisfies the same with good rate function

$$I(\mu) := \begin{cases} \tilde{I}(\tilde{\mu}) & \text{if } \mu = (1-\alpha)\mu_0 + \alpha\tilde{\mu},\ \tilde{\mu} \in \mathcal{M}(X), \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Write the distribution of $\tilde{P}_n$ on $\mathcal{M}(X)$ by the same $\tilde{P}_n$. Then the distribution P_n of $(1-\alpha_n)\mu_0 + \alpha_n\tilde{P}_n$ is given by

$$P_n(\Gamma) = \tilde{P}_n\big(\{\tilde{\mu} \in \mathcal{M}(X) : (1-\alpha_n)\mu_0 + \alpha_n\tilde{\mu} \in \Gamma\}\big)$$

for Borel sets $\Gamma \subset \mathcal{M}(X)$. First we show that (P_n) is exponentially tight. For any $\varepsilon > 0$ there exists a compact $\tilde{K}_\varepsilon \subset \mathcal{M}(X)$ such that

$$\limsup_{n\to\infty} L_n \log \tilde{P}_n(\tilde{K}_\varepsilon^c) \le -\frac{1}{\varepsilon}\,.$$

By noting that the weak topology on $\mathcal{M}(X)$ is metrizable (by the Lévy metric), it is easy to see that the closure K_ε of $\bigcup_{n=1}^\infty \bigl((1-\alpha_n)\mu_0 + \alpha_n \tilde{K}_\varepsilon\bigr)$ is compact. Since $P_n(K_\varepsilon^c) \le \tilde{P}_n(\tilde{K}_\varepsilon^c)$, we get the conclusion.

Now it suffices to show that for every $\mu \in \mathcal{M}(X)$

$$\inf_G \left[\liminf_{n\to\infty} L_n \log P_n(G)\right] \ge -I(\mu)\,, \tag{17}$$

$$\inf_G \left[\limsup_{n\to\infty} L_n \log P_n(G)\right] \le -I(\mu)\,, \tag{18}$$

where G runs over neighborhoods of μ. Let $\mathcal{D}$ denote the set $\{(1-\alpha)\mu_0 + \alpha\tilde{\mu} : \tilde{\mu} \in \mathcal{M}(X)\}$. If $\mu \notin \mathcal{D}$ then $\mu(C) < (1-\alpha)\mu_0(C)$ for some closed $C \subset X$. Choose $\mu(C) < \alpha_0 < (1-\alpha)\mu_0(C)$ and define a neighborhood $G := \{\mu' \in \mathcal{M}(X) : \mu'(C) < \alpha_0\}$ of μ. Then, since $P_n(G) = 0$ if $(1-\alpha_n)\mu_0(C) > \alpha_0$, (17) and (18) hold in this case. Next assume that $\mu \in \mathcal{D}$ and $\mu = (1-\alpha)\mu_0 + \alpha\tilde{\mu}$. For any neighborhood G of μ there exists a neighborhood $\tilde{G}$ of $\tilde{\mu}$ such that $(1-\alpha_n)\mu_0 + \alpha_n\tilde{G} \subset G$ for large n and hence

$$\liminf_{n\to\infty} L_n \log P_n(G) \ge \liminf_{n\to\infty} L_n \log \tilde{P}_n(\tilde{G}) \ge -\tilde{I}(\tilde{\mu})\,.$$

This implies (17). On the other hand, for any neighborhood $\tilde{G}$ of $\tilde{\mu}$ there exists a neighborhood G of μ such that $\alpha_n^{-1}G - (\alpha_n^{-1} - 1)\mu_0 \subset \tilde{G}$ or

$$\{\tilde{\mu} \in \mathcal{M}(X) : (1-\alpha_n)\mu_0 + \alpha_n\tilde{\mu} \in G\} \subset \tilde{G}$$

for large n. Hence

$$\inf_G \left[\limsup_{n\to\infty} L_n \log P_n(G)\right] \le \inf_{\tilde{G}} \left[\limsup_{n\to\infty} L_n \log \tilde{P}_n(\tilde{G})\right] \le -\tilde{I}(\tilde{\mu})\,,$$

implying (18). □

References

[1] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Second Edition, John Wiley, New York, 1984.

[2] G. Ben Arous and A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy, *Probab. Theory Related Fields* **108** (1997), 517–542.

[3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston-London, 1993.

[4] J.D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[5] F. Hiai and D. Petz, Maximizing free entropy, *Acta Math. Hungar.*, **80**(1998), 325–346

[6] F. Hiai and D. Petz, A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices, Preprint No. 17/1997, Mathematical Institute HAS, Budapest.

[7] D. Johnson, Some limit theorems for the eigenvalues of a sample covariance matrix, *J. Multivariate Anal.* **12** (1982), 1–38.

[8] A.M. Khorunzhy and L.A. Pastur, On the eigenvalue distribution of the deformed Wigner ensemble of random matrices, in *Advances in Soviet Math.* Vol. 19, V.A. Marchenko (ed.), Amer. Math. Soc., 1994, pp. 97–127.

[9] V.A. Marchenko and L.A. Pastur, The distribution of eigenvalues in certain sets of random matrices, *Mat. Sb.* **72** (1967), 507–536.

[10] M.L. Mehta, *Random Matrices*, Academic Press, Boston, 1991.

[11] H..N. Mhaskar and E.B. Saff, Where does the sup norm of a weighted polynomial live? (A generalization of incomplete polynomials), *Constr. Approx.* **1** (1985), 71–91.

[12] F. Oravecz and D. Petz, On the eigenvalue distribution of some symmetric random matrices, *Acta Sci. Math. (Szeged)* **63** (1997), 383–395.

[13] D. Petz and F. Hiai, Logarithmic energy as entropy functional, in *Advances in Differential Equations and Mathematical Physics*, eds: E. Carlen, E.M. Harrell, M. Loss, Contemporary Mathematics, Vol **217**, Amer. Math. Soc., 1998, pp. 205–221

[14] E.B. Saff and V. Totik, *Logarithmic Potentials with External Fields*, Springer, Berlin-Heidelberg-New York, 1997.

[15] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, II, *Invent. Math.* **118** (1994), 411–440.

[16] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, CRM Monograph Ser., Vol. 1, Amer. Math. Soc., 1992.

[17] K.W. Wachter, The strong limits of random matrix spectra for sample matrices of independent elements, *Ann. Probab.* **6** (1978), 1–18.