

Large Deviations for Random Matrices

茨城大・理 日合文雄 (Fumio Hiai)

Introduction

In [Vo2], Voiculescu introduced the free entropy of a probability distribution as the minus sign of logarithmic energy from potential theory, and its extension to non-commutative multi-random variables was developed in [Vo3] in purely noncommutative setting. The free entropy theory is one of the highlights in recent breakthrough of Voiculescu's free probability theory. In [BG], Ben Arous and Guionnet obtained large deviation theorem for the empirical distribution of selfadjoint Gaussian random matrices and justified the notion of free entropy from the viewpoint of large deviation principle (LDP). But it should be mentioned that the idea from LDP already appeared in [Vo2, 3] in rather essential manner. So-called Voiculescu's heuristics in [Vo2] is a kind of LDP though of course it is not rigorous. The definition of multiple free entropy in [Vo3] was given in the framework of LDP, and the equality (up to an additive constant) of two definitions in [Vo2, 3] for single variable case is considered as a result of LDP.

In this lecture, we want to make clear the intrinsic relationship among the following three themes (see the picture at the end of Sec. 2).

- (i) Maximization problems for free entropy functionals.
- (ii) Wigner type limit theorem for the mean spectral density of random matrices.
- (iii) LDP for the empirical eigenvalue distribution of random matrices.

In Sec. 1, we present many examples of free entropy maximization problems. Important distributions in free probability theory as well as in classical theory appear as maximizers. In Sec. 2, we describe the basic framework of our LDP for random matrices. In Sec. 3, we show the LDP for the empirical eigenvalue distribution of selfadjoint Gaussian random matrices (which model a semicircular element in free probability theory). Although it was proved in [BG], our proof is more transparent in a bit more general setting. In Sec. 4, we show the similar LDP for non-selfadjoint Gaussian random matrices (which model a circular element). Finally in Sec. 5, we give a rather general LDP for unitary random matrices.

Acknowledgements. The author is grateful to D. Petz for a stimulating joint work. This lecture is mostly based on the joint work with him. He thanks M. Izumi for useful comments from several aspects.

1. Maximizing free entropy

For a probability measure μ on $\mathbb{R}$ (or $\mathbb{C}$), the *free entropy* $\Sigma(\mu)$ of μ is defined as the double integral

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

whenever the integral is meaningful. When μ is compactly supported, this integral always exists although it can be $-\infty$ (for example if μ has an atom). On the other hand, the so-called *logarithmic energy* of a signed measure ν is

$$I(\nu) = \iint \log \frac{1}{|x - y|} d\nu(x) d\nu(y),$$

which plays an important role in potential theory [La]. Note [La] that if ν is compactly supported and $\nu(1) = 0$, then $I(\nu) \geq 0$ and $I(\nu) = 0$ if and only if $\nu = 0$. The free entropy functional $\Sigma(\mu) = -I(\mu)$ is weakly upper semicontinuous and strictly concave on the set of probability measures supported on any given compact subset of $\mathbb{R}$ (or $\mathbb{C}$).

It seems remarkable that many probability distributions familiar in free probability theory [VDN] (also [HP2]) are realized as a maximizer of a free entropy functional. To deal with maximization problems for free entropy (under constraints), the following theorem from the theory of weighted potentials is quite useful. It was proved in [MS1, 2] by the adaptation of the classical Frostman method.

Let S be a closed subset in $\mathbb{R}$ (or $\mathbb{C}$). Let $\mathcal{M}(S)$ denote the set of all probability measures whose support $\text{supp}(\mu)$ is included in S . Moreover, let $w : S \rightarrow [0, \infty)$ be a weight function, which is assumed for simplicity to satisfy the following conditions:

- (i) w is continuous on S .
- (ii) $S_0 = \{x \in S : w(x) > 0\}$ has positive (inner logarithmic) capacity, i.e. $I(\mu) < +\infty$ for some probability measure μ such that $\text{supp}(\mu) \subset S_0$.
- (iii) $|x|w(x) \rightarrow 0$ as $x \in S, |x| \rightarrow \infty$, when S is unbounded.

Let $Q(x) = -\log w(x)$ and define the weighted energy functional

$$I_Q(\mu) = I(\mu) + 2 \int Q(x) d\mu(x) \quad \text{on } \mathcal{M}(S).$$

Note that $I_Q(\mu) > -\infty$ is well defined thanks to the above assumptions.

Theorem 1.1. *With the above assumptions, there exists a unique $\mu_0 \in \mathcal{M}(S)$ such that*

$$I_Q(\mu_0) = \inf\{I_Q(\mu) : \mu \in \mathcal{M}(S)\}.$$

Then $I_Q(\mu_0)$ is finite, μ_0 has finite logarithmic energy, and $\text{supp}(\mu_0)$ is compact. Furthermore, the minimizer μ_0 is characterized as $\mu_0 \in \mathcal{M}(S)$ with compact support such that for some real number B the following hold:

$$\int \log |x - y| d\mu_0(y) = Q(x) - B \quad \text{if } x \in \text{supp}(\mu_0),$$

$$\int \log |x - y| d\mu_0(y) \leq Q(x) - B \quad \text{if } x \in S \setminus \text{supp}(\mu_0).$$

In this case, $B = I_Q(\mu_0) - \int Q d\mu_0$.

In the following we list several examples of maximization problems, which can be solved by applying the above theorem. The details on (1)–(3) were given in [HP1].

(1) When S is a compact set (having positive capacity), a unique minimizer μ_S for $I(\mu)$ (or maximizer of $\Sigma(\mu)$) on $\mathcal{M}(S)$ is sometimes called the *equilibrium measure* on S . For instance, the arcsine law

$$h(x) = \frac{1}{\pi\sqrt{1-x^2}} \chi_{(-1,1)}(x)$$

is the equilibrium measure on $[-1, 1]$, because

$$\int_{-1}^1 h(y) \log |x - y| dy = -\log 2 \quad (-1 \leq x \leq 1).$$

Also this yields $\Sigma(h) = -\log 2$.

(2) For $p, r > 0$ the *Ullman distribution* has the probability density

$$v_r^{(p)}(x) = \begin{cases} \frac{p}{\pi r^p} \int_{|x|}^r \frac{t^{p-1}}{\sqrt{t^2 - x^2}} dt & \text{if } -r \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\int_{-r}^r |x|^p v_r^{(p)}(x) dx = \alpha_p r^p \quad \text{where} \quad \alpha_p = \frac{\Gamma(\frac{p+1}{2})}{2\sqrt{\pi}\Gamma(\frac{p}{2} + 1)}.$$

According to [To, pp. 12–13],

$$\int_{-r}^r v_r^{(p)}(y) \log |x - y| dy = \frac{|x|^p}{2p\alpha_p r^p} + \log \frac{r}{2} - \frac{1}{p} \quad (|x| \leq r),$$

$$\int_{-r}^r v_r^{(p)}(y) \log |x - y| dy < \frac{|x|^p}{2p\alpha_p r^p} + \log \frac{r}{2} - \frac{1}{p} \quad (|x| > r).$$

Hence $v_r^{(p)}$ is a unique maximizer of the functional

$$\Sigma(\mu) - \frac{1}{p\alpha_p r^p} \int |x|^p d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}).$$

This is equivalently formulated in the way that $v_r^{(p)}$ is a maximizer of $\Sigma(\mu)$ on $\{\mu \in \mathcal{M}(\mathbb{R}) : \int |x|^p d\mu(x) \leq \alpha_p r^p\}$. Also $\Sigma(v_r^{(p)}) = \log \frac{r}{2} - \frac{1}{2p}$. In particular, $v_r^{(2)}$ is the semicircle law $w_r = \frac{2}{\pi r^2} \sqrt{r^2 - x^2} \chi_{[-r, r]} dx$ with mean 0 and variance $r^2/4$, so that w_r maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\mathbb{R}) : \int x^2 d\mu(x) \leq r^2/4\}$ and $\Sigma(w_r) = \log \frac{r}{2} - \frac{1}{4}$.

In [BIPZ], in connection to the planar approximation to field theory, Brézin et al. considered the distribution

$$u(x) = \frac{1}{\pi} \left(\frac{1}{2} + 4ga^2 + 2gx^2 \right) \sqrt{4a^2 - x^2} \chi_{[-2a, 2a]}(x),$$

where $a^2 + 12ga^4 = 1$. This is a convex combination $a^2 w_{2a} + 12ga^4 v_{2a}^{(4)}$ and is a maximizer of the functional

$$\Sigma(\mu) - \int \left(\frac{1}{2} x^2 + gx^4 \right) d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}).$$

(3) For $p, r > 0$ define the probability density $u_r^{(p)}$ by $u_r^{(p)}(x) = v_{\sqrt{2r}}^{(2p)}(\sqrt{x})/\sqrt{x}$ on $\mathbb{R}^+$, that is,

$$u_r^{(p)}(x) = \begin{cases} \frac{p}{\pi r^p} \int_{x/2}^r \frac{t^{p-1}}{\sqrt{2tx - x^2}} dt & \text{if } 0 \leq x \leq 2r, \\ 0 & \text{otherwise.} \end{cases}$$

Then $u_r^{(p)}$ is a maximizer of the functional

$$\Sigma(\mu) - \frac{1}{p\tilde{\alpha}_p r^p} \int x^p d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}^+),$$

or equivalently, $u_r^{(p)}$ maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\mathbb{R}^+) : \int x^p d\mu(x) \leq \tilde{\alpha}_p r^p\}$, where

$$\tilde{\alpha}_p = \frac{2^{p-1} \Gamma(p + \frac{1}{2})}{\sqrt{\pi} \Gamma(p + 1)}.$$

Although this can be shown by using Theorem 1.1, a more efficient way is to transform the above (2) via the bijective correspondence $T : \mathcal{M}_s(\mathbb{R}) \rightarrow \mathcal{M}(\mathbb{R}^+)$ by $T\mu = \mu \circ \sigma^{-1}$ where $\mathcal{M}_s(\mathbb{R})$ denotes the set of symmetric probability measures on $\mathbb{R}$ and $\sigma(x) = x^2$, because $\int |x|^{2p} d\mu(x) = \int x^p d(T\mu)(x)$ and $2\Sigma(\mu) = \Sigma(T\mu)$. We have $\Sigma(u_r^{(p)}) = \log \frac{r}{2} - \frac{1}{2p}$. In particular, the distribution

$$u_2^{(1)}(x) = \frac{\sqrt{4 - x^2}}{2\pi x} \chi_{[0, 4]}(x)$$

appears as the limit distribution of the free Poisson limit theorem [VDN, p. 34].

(4) For $a > 0$ the *Marchenko-Pastur distribution* is given as

$$\mu_a = \begin{cases} \frac{\sqrt{4a - (x - 1 - a)^2}}{2\pi x} \chi(x) dx & \text{if } a \geq 1, \\ (1 - a)\delta_0 + \frac{\sqrt{4a - (x - 1 - a)^2}}{2\pi x} \chi(x) dx & \text{if } a < 1. \end{cases}$$

where χ is the characteristic function of the interval $[(1 - \sqrt{a})^2, (1 + \sqrt{a})^2]$. By using the Cauchy tranform and the Hilbert transform of μ_a it is not difficult to see that

$$\int \log |x - y| d\mu_a(y) = \frac{1}{2}(x - (a - 1) \log x) + C \quad \text{if } x \in \text{supp}(\mu_a),$$

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

where constant C is given by

$$C = \int_{(1-\sqrt{a})^2}^{(1+\sqrt{a})^2} \frac{\sqrt{4a - (x - 1 - a)^2}}{2\pi x} \log |x - 1 - a| dx \\ - \frac{1}{2}(1 + a - (a - 1) \log(1 + a)).$$

We compute

$$\begin{aligned} & \int_{(1-\sqrt{a})^2}^{(1+\sqrt{a})^2} \frac{\sqrt{4a - (x - 1 - a)^2}}{2\pi x} \log |x - 1 - a| dx \\ &= \log 2\sqrt{a} + \frac{\sqrt{a}}{\pi} \int_{-1}^1 \frac{\sqrt{1 - x^2}}{x + \frac{1+a}{2\sqrt{a}}} \log |x| dx \\ &= \log 2\sqrt{a} + \frac{1+a}{\pi} \int_0^1 \frac{\sqrt{1 - x^2}}{\frac{(1+a)^2}{4a} - x^2} \log x dx \\ &= \log 2\sqrt{a} + \frac{1+a}{\pi} \left(\frac{\pi}{2} \frac{b-1}{b} \log \frac{2b-1}{2b} - \frac{\pi}{2b} \log 2 \right) \end{aligned}$$

thanks to an integral formula obtained in [Lar], where

$$b = \frac{(1+a)^2}{4a} + \sqrt{\frac{(1+a)^4}{(4a)^2} - \frac{(1+a)^2}{4a}} = \frac{(1+a)^2 + |1 - a^2|}{4a}.$$

Therefore, C can be exactly calculated as

$$C = \begin{cases} \frac{a}{2} \log a - \frac{1+a}{2} & \text{if } a \geq 1, \\ \frac{1}{2} \log a - (1-a) \log \frac{1+a}{2} - \frac{1+a}{2} & \text{if } a < 1. \end{cases}$$

When $a \geq 1$, the weight function $w(x) = x^{(a-1)/2}e^{-x/2}$ on $\mathbb{R}^+$ satisfies conditions (i)–(iii) stated above Theorem 1.1. Hence we can apply Theorem 1.1, so that μ_a is a maximizer of the functional

$$\Sigma(\mu) + (a-1) \int \log x d\mu(x) - \int x d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}^+).$$

Also,

$$\begin{aligned} \Sigma(\mu_a) &= \frac{a}{2} - \frac{a-1}{2} \int \log x d\mu_a(x) + C \\ &= \frac{a+1}{4} \log a - \frac{a-1}{2} \log 2 - \frac{1}{2} \\ &\quad - \frac{\sqrt{a}(a-1)}{2\pi} \int_{-1}^1 \frac{\sqrt{1-x^2}}{x + \frac{1+a}{2\sqrt{a}}} \log \left(x + \frac{1+a}{2\sqrt{a}} \right) dx, \end{aligned}$$

though we do not know the exact value of $\Sigma(\mu_a)$. In particular when $a = 1$, μ_1 is $u_2^{(1)}$ in the above (3) (the case of free Poisson). When $0 < a < 1$, μ_a has an atom at 0 and of course $\Sigma(\mu_a) = -\infty$. Since $w(x)$ has a singularity at $x = 0$, one cannot directly apply Theorem 1.1. Indeed, the above functional has indefinite value $(-\infty) - (-\infty)$. Nevertheless, a certain justification can be made for μ_a to be a maximizer, while we omit the details.

(5) The *quarter-circular* distribution $\tilde{w}_r = \frac{4}{\pi r^2} \sqrt{r^2 - x^2} \chi_{[0,r]} dx$ is also interesting. From the Hilbert transform of $\tilde{w}_r$ it may be possible to find a function $Q(x)$ on $\mathbb{R}^+$ such that $\tilde{w}_r$ is a maximizer of the functional $\Sigma(\mu) - \int Q(x) d\mu(x)$ on $\mathcal{M}(\mathbb{R}^+)$. In fact, we have the principal value integral for $0 < x < 1$

$$\int_0^1 \frac{\sqrt{1-y^2}}{x-y} dy = \frac{\pi}{2}x + 1 + \sqrt{1-x^2} \log \left(\frac{1-\sqrt{1-x^2}}{x} \right).$$

But it is not so simple for further calculation. It seems more convenient (and probably more natural) to consider another type of entropy functional. For $\mu \in \mathcal{M}_s(\mathbb{R})$ let $\tilde{\mu}$ be the restriction of 2μ on $\mathbb{R}^+$. Let Q be a symmetric continuous function on $\mathbb{R}$ such that $x \exp(-Q(x)) \rightarrow 0$ as $x \rightarrow +\infty$. Then for every $\mu \in \mathcal{M}_s(\mathbb{R})$,

$$\frac{1}{2} \iint \log |x^2 - y^2| d\tilde{\mu}(x) d\tilde{\mu}(y) - \int Q(x) d\tilde{\mu}(x) = \Sigma(\mu) - \int Q(x) d\mu(x).$$

Hence μ_0 is a maximizer of $\Sigma(\mu) - \int Q(x) d\mu(x)$ on $\mathcal{M}(\mathbb{R})$ if and only if $\tilde{\mu}_0$ is a maximizer of

$$\frac{1}{2} \iint \log |x^2 - y^2| d\mu(x) d\mu(y) - \int Q(x) d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}^+).$$

So the above (2) implies that $\tilde{v}_r^{(p)} = 2v_r^{(p)}(x)\chi_{[0,r]} dx$ is a maximizer of

$$\frac{1}{2} \iint \log |x^2 - y^2| d\mu(x) d\mu(y) - \frac{1}{p\alpha_p r^p} \int x^p d\mu(x) \quad \text{on } \mathcal{M}(\mathbb{R}^+).$$

In particular, we have the quarter-circular distribution $\tilde{w}_r$ as a maximizer of $\frac{1}{2} \iint \log |x^2 - y^2| d\mu(x) d\mu(y) - \frac{2}{r^2} \int x^2 d\mu(x)$ on $\mathcal{M}(\mathbb{R}^+)$. By the way, it seems that the exact value of $\Sigma(\tilde{w}_r)$ is not known.

(6) For $R > 0$ and $p > 0$ define the distribution $\lambda_R^{(p)}$ supported on the disk $\{\zeta \in \mathbb{C} : |\zeta| \leq R\}$ by

$$\lambda_R^{(p)} = \frac{p}{2\pi R^p} d\theta \cdot r^{p-1} \chi_{[0,R]}(r) dr \quad (\zeta = re^{i\theta}).$$

In particular when $p = 2$, $\lambda_R^{(2)}$ is the uniform distribution on $\{\zeta \in \mathbb{C} : |\zeta| \leq R\}$. It is not difficult to compute

$$\int \log |\zeta - \eta| d\lambda_R^{(p)}(\eta) = \frac{|\zeta|^p}{pR^p} + \log R - \frac{1}{p} \quad (|\zeta| \leq R),$$

$$\int \log |\zeta - \eta| d\lambda_R^{(p)}(\eta) = \log |\zeta| < \frac{|\zeta|^p}{pR^p} + \log R - \frac{1}{p} \quad (|\zeta| > R).$$

Hence $\lambda_R^{(p)}$ is a maximizer of the functional

$$\Sigma(\mu) - \frac{2}{pR^p} \int |\zeta|^p d\mu(\zeta) \quad \text{on } \mathcal{M}(\mathbb{C}),$$

and it maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\mathbb{C}) : \int |\zeta|^p d\mu(\zeta) \leq R^p/2\}$. Also $\Sigma(\lambda_R^{(p)}) = \log R - \frac{1}{2p}$.

(7) Consider probability measures on the unit disk $\overline{\mathbb{D}} = \{\zeta \in \mathbb{C} : |\zeta| \leq 1\}$. For $\alpha \in \mathbb{C}$, $|\alpha| < 1$, we have the Poisson kernel measure

$$p_\alpha = \frac{1 - |\alpha|^2}{|\zeta - \alpha|^2} d\zeta \quad (\zeta = e^{i\theta}, d\zeta = d\theta/2\pi)$$

supported on the unit circle $\mathbb{T}$. Since

$$\int \log |\zeta - \eta| dp_\alpha(\eta) = \log |\zeta \bar{\alpha} - 1| \quad (\zeta \in \overline{\mathbb{D}}),$$

we see that p_α is a maximizer of the functional

$$\Sigma(\mu) - \int \log |\zeta \bar{\alpha} - 1|^2 d\mu(\zeta) \quad \text{on } \mathcal{M}(\overline{\mathbb{D}}).$$

Moreover, p_α maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\overline{\mathbb{D}}) : \int \log |\zeta \bar{\alpha} - 1| d\mu(\zeta) = \log(1 - |\alpha|^2)\}$ and $\Sigma(p_\alpha) = \log(1 - |\alpha|^2)$. In particular, as is well known, the Lebesgue (or Haar) probability measure on $\mathbb{T}$ is the equilibrium measure on $\overline{\mathbb{D}}$ (or $\mathbb{T}$).

(8) In [GW], Gross and Witten showed that the following distribution on $\mathbb{T}$ arises as the limit eigenvalue distribution in some lattice gauge theory:

$$\rho_\lambda = \begin{cases} \frac{1}{2\pi} \left(1 + \frac{2}{\lambda} \cos \theta\right) d\theta & (|\theta| \leq \pi) & \text{if } 2 \leq \lambda \leq \infty, \\ \frac{2}{\pi\lambda} \cos \frac{\theta}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{\theta}{2}} d\theta & (|\theta| \leq 2 \sin^{-1} \sqrt{\lambda/2}) & \text{if } 0 < \lambda < 2. \end{cases}$$

This means that the phase transition exists (for the limit) at $\lambda = 2$. The result in [GW] can be reformulated as follows: ρ_λ is a unique maximizer of the functional

$$\Sigma(\mu) + \frac{2}{\lambda} \int \operatorname{Re} \zeta d\mu(\zeta) \quad \text{on } \mathcal{M}(\mathbb{T}).$$

Furthermore, when $2 \leq \lambda \leq \infty$, ρ_λ maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\mathbb{T}) : \int \operatorname{Re} \zeta d\mu(\zeta) \text{ (or } \int \zeta d\mu(\zeta)) = 1/\lambda\}$ and $\Sigma(\rho_\lambda) = -1/\lambda^2$. When $0 < \lambda < 2$, ρ_λ maximizes $\Sigma(\mu)$ among $\{\mu \in \mathcal{M}(\mathbb{T}) : \int \operatorname{Re} \zeta d\mu(\zeta) \text{ (or } \int \zeta d\mu(\zeta)) = 1 - \lambda/4\}$ and $\Sigma(\rho_\lambda) = \frac{1}{2} \log \frac{\lambda}{2} - \frac{1}{4}$. Since the argument in [GW] does not seem easy to check, we include a more accessible proof below.

When $\lambda \geq 2$ the computation is straightforward. For $\zeta = e^{it}$,

$$\begin{aligned} \int \log |\zeta - \eta| d\rho_\lambda(\eta) &= \frac{1}{2\pi} \int_0^{2\pi} \left(1 + \frac{2}{\lambda} \cos(\theta + t)\right) \log |1 - e^{i\theta}| d\theta \\ &= \frac{1}{2\pi\lambda} \int_0^{2\pi} \cos(\theta + t) \log 2(1 - \cos \theta) d\theta \\ &= \frac{\cos t}{2\pi\lambda} \int_0^{2\pi} \cos \theta \log(1 - \cos \theta) d\theta \\ &= \frac{\cos t}{\pi\lambda} \int_{-1}^1 \frac{t}{\sqrt{1-t^2}} \log(1-t) dt \\ &= -\frac{\cos t}{\pi\lambda} \int_{-1}^1 \sqrt{\frac{1+t}{1-t}} dt = -\frac{1}{\lambda} \cos t. \end{aligned}$$

This implies the first assertion by Theorem 1.1. Moreover,

$$\Sigma(\rho_\lambda) = -\frac{1}{2\pi\lambda} \int_0^{2\pi} \cos t \left(1 + \frac{2}{\lambda} \cos t\right) dt = -\frac{1}{\lambda^2}.$$

If $\mu \in \mathcal{M}(\mathbb{T})$ satisfies $\int \operatorname{Re} \zeta d\mu(\zeta) = 1/\lambda$, then

$$\begin{aligned} 0 \leq I(\mu - \rho_\lambda) &= -\Sigma(\mu) - \Sigma(\rho_\lambda) + 2 \iint \log |\zeta - \eta| d\rho_\lambda(\eta) d\mu(\zeta) \\ &= -\Sigma(\mu) + \frac{1}{\lambda^2} - \frac{2}{\lambda} \int \operatorname{Re} \zeta d\mu(\zeta) = -\Sigma(\mu) - \frac{1}{\lambda^2}, \end{aligned}$$

so that $\Sigma(\mu) \leq -1/\lambda^2$ and equality occurs if and only if $\mu = \rho_\lambda$.

When $\lambda < 2$ the computation is much involved. We use the technique of the ‘‘Hilbert transform’’ on the circle. Put $\alpha = \sqrt{\lambda/2}$ and $\beta = 2 \sin^{-1} \alpha$. Define for $\zeta = e^{it}$

$$\begin{aligned}
F(t) &= \int \log |\zeta - \eta| d\rho_\lambda(\eta) \\
&= \frac{2}{\pi\lambda} \int_{-\beta}^{\beta} \cos \frac{\theta}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{\theta}{2}} \log |1 - e^{i(\theta-t)}| d\theta \\
&= \frac{2}{\pi\lambda} \int_{-\beta}^{\beta} \cos \frac{\theta}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{\theta}{2}} \log 2 \left| \sin \frac{\theta-t}{2} \right| d\theta.
\end{aligned}$$

When $|t| < \alpha$, the differential of $F(t)$ in the sense of distributions in $(-\alpha, \alpha)$ is given as

$$F'(t) = \frac{1}{\pi\lambda} \int_{-\beta}^{\beta} \cos \frac{\theta}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{\theta}{2}} \cot \frac{\theta-t}{2} d\theta.$$

(This and the integrals below mean the principal value integral.) We proceed to compute

$$\begin{aligned}
F'(t) &= \frac{2}{\pi\lambda} \int_{-\alpha}^{\alpha} \sqrt{\alpha^2 - x^2} \frac{\sqrt{1-x^2} \cos \frac{t}{2} + x \sin \frac{t}{2}}{x \cos \frac{t}{2} - \sqrt{1-x^2} \sin \frac{t}{2}} dx \\
&= \frac{1}{\pi\lambda} \int_{-\alpha}^{\alpha} \sqrt{\alpha^2 - x^2} \frac{\sin t + 4x\sqrt{1-x^2}}{x^2 - \sin^2 \frac{t}{2}} dx \\
&= \frac{\sin t}{\pi\lambda} \int_{-\alpha}^{\alpha} \frac{\sqrt{\alpha^2 - x^2}}{x^2 - \sin^2 \frac{t}{2}} dx.
\end{aligned}$$

Since the above principal value integral is equal to $-\pi$ (see [Me, p. 74]), we have $F'(t) = -\frac{1}{\lambda} \sin t$ and hence

$$F(t) = -\frac{1}{\lambda} \cos t + \text{const} \quad (|t| < \alpha).$$

Since

$$\begin{aligned}
F(0) &= \frac{2}{\pi\lambda} \int_{-\beta}^{\beta} \cos \frac{\theta}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{\theta}{2}} \log 2 \left| \sin \frac{\theta}{2} \right| d\theta \\
&= \frac{8}{\pi\lambda} \int_0^{\alpha} \sqrt{\alpha^2 - x^2} \log 2x dx \\
&= \frac{4}{\pi} \int_0^1 \sqrt{1-x^2} \log 2\alpha x dx = \frac{1}{2} \log \frac{\lambda}{2} - \frac{1}{2},
\end{aligned}$$

we have

$$F(t) = -\frac{1}{\lambda} \cos t + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} \quad (|t| < \alpha).$$

On the other hand, when $|t| > \alpha$, $F(t)$ is differentiable in usual sense and

$$\begin{aligned}
F'(t) &= \frac{\sin t}{2\pi\alpha^2} \int_{-\alpha}^{\alpha} \frac{\sqrt{\alpha^2 - x^2}}{x^2 - \sin^2 \frac{t}{2}} dx \\
&< \frac{\sin t}{2\pi\alpha^2} \int_{-\alpha}^{\alpha} \frac{\sqrt{\alpha^2 - x^2}}{x^2 - \alpha^2} dx = -\frac{1}{\lambda} \sin t.
\end{aligned}$$

Hence

$$F(t) \leq -\frac{1}{\lambda} \cos t + \text{const} \quad (|t| > \alpha).$$

Since $F(t)$ is continuous at $t = \pm\alpha$, we obtain

$$\begin{aligned} \int \log |\zeta - \eta| d\rho_\lambda(\eta) &= -\frac{1}{\lambda} \text{Re } \zeta + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} & \text{if } \zeta \in \text{supp}(\rho_\lambda), \\ \int \log |\zeta - \eta| d\rho_\lambda(\eta) &< -\frac{1}{\lambda} \text{Re } \zeta + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} & \text{if } \zeta \in \mathbf{T} \setminus \text{supp}(\rho_\lambda). \end{aligned}$$

This implies the first assertion. Moreover,

$$\begin{aligned} \Sigma(\rho_\lambda) &= -\frac{2}{\pi\lambda^2} \int_{-\beta}^{\beta} \cos t \cos \frac{t}{2} \sqrt{\frac{\lambda}{2} - \sin^2 \frac{t}{2}} dt + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} \\ &= -\frac{8}{\pi\lambda^2} \int_0^{\alpha} (1 - 2x^2) \sqrt{\alpha^2 - x^2} dx + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} \\ &= \frac{1}{2} \log \frac{\lambda}{2} - \frac{1}{4}. \end{aligned}$$

If $\mu \in \mathcal{M}(\mathbf{T})$ satisfies $\int \text{Re } \zeta d\mu(\zeta) = 1 - \lambda/4$, then

$$\begin{aligned} 0 \leq I(\mu - \rho_\lambda) &= -\Sigma(\mu) - \Sigma(\rho_\lambda) + 2 \iint \log |\zeta - \eta| d\rho_\lambda(\eta) d\mu(\zeta) \\ &\leq -\Sigma(\mu) - \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{4} + 2 \left(-\frac{1}{\lambda} \left(1 - \frac{\lambda}{4} \right) + \frac{1}{2} \log \frac{\lambda}{2} + \frac{1}{\lambda} - \frac{1}{2} \right) \\ &= -\Sigma(\mu) + \frac{1}{2} \log \frac{\lambda}{2} - \frac{1}{4}, \end{aligned}$$

so that $\Sigma(\mu) \leq \frac{1}{2} \log \frac{\lambda}{2} - \frac{1}{4}$ and equality occurs if and only if $\mu = \rho_\lambda$.

2. Introductory survey of large deviation principle

This section is a very brief introduction to the large deviation principle (LDP) for the convenience of the reader. The interested reader may consult [DZ] or [DS]. Let X be a topological space and $\mathcal{B}$ the Borel σ -field on X . Let $\{P_\epsilon\}_{\epsilon>0}$ be a family of probability measures on $(X, \mathcal{B})$. The LDP of $\{P_\epsilon\}$ characterizes the limiting behavior of P_ϵ as $\epsilon \rightarrow 0$ in terms of a rate function. Roughly speaking, the LDP may be considered as a refinement of limit theorems such as law of large numbers and central limit theorem.

The precise definition of the LDP is given as follows.

Definition 2.1. (1) A lower semicontinuous function $I : X \rightarrow [0, \infty]$ is called a *rate function*. A rate function I is said to be *good* if $\{x : I(x) \leq \alpha\}$ is compact for any $\alpha \in [0, \infty)$.

(2) $\{P_\varepsilon\}$ is said to satisfy the *LDP* with a rate function I if

$$\begin{aligned} -\inf\{I(x) : x \in \Gamma^\circ\} &\leq \liminf_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(\Gamma) \\ &\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(\Gamma) \leq -\inf\{I(x) : x \in \overline{\Gamma}\} \end{aligned}$$

for all $\Gamma \in \mathcal{B}$, where Γ° and $\overline{\Gamma}$ denote the interior and closure of Γ . This is equivalent to the following two conditions:

(i) For every open set $G \subset X$,

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(G) \geq -\inf\{I(x) : x \in G\}.$$

(ii) For every closed set $F \subset X$,

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(F) \leq -\inf\{I(x) : x \in F\}.$$

(3) $\{P_\varepsilon\}$ is said to satisfy the *weak LDP* with a rate function I if the above (i) holds and the upper bound in (ii) holds for every compact set $F \subset X$.

(4) $\{P_\varepsilon\}$ is said to be *exponentially tight* if for every $\delta > 0$ there exists a compact set $K_\delta \subset X$ such that

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(K_\delta^c) < -1/\delta.$$

This condition is trivial if X itself is compact.

Note that if I is a good rate function and the above (ii) holds, then there exists at least one point $x \in X$ such that $I(x) = 0$.

The following properties are useful to show the LDP.

Proposition 2.2. *Assume that $\{P_\varepsilon\}$ is exponentially tight. Then:*

(1) *If the above (i) holds, then the rate function I is good.*

(2) *If $\{P_\varepsilon\}$ satisfies the weak LDP with a rate function I , then it satisfies the LDP.*

(3) *$\{P_\varepsilon\}$ satisfies the weak LDP with a rate function I if*

$$\inf_G \{\limsup_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(G)\} \leq -I(x) \leq \inf_G \{\liminf_{\varepsilon \rightarrow 0} \varepsilon \log P_\varepsilon(G)\}$$

for every $x \in X$, where G runs over a neighborhood base of x .

In many cases, we treat a sequence $\{P_n\}$ of probability measures on $(X, \mathcal{B})$. For $0 < a_n \rightarrow 0$, we say that $\{P_n\}$ satisfies the LDP in the scale a_n with a rate function I if

$$\liminf_{n \rightarrow \infty} a_n \log P_n(G) \geq -\inf\{I(x) : x \in G\}$$

for every open $G \subset X$ and

$$\limsup_{n \rightarrow \infty} a_n \log P_n(F) \leq -\inf\{I(x) : x \in F\}$$

for every closed $F \subset X$. Usually $a_n = n^{-1}$ or $a_n = n^{-2}$ is used.

In the rest of this section let us explain the general scheme of the LDP for random matrices which we will discuss in subsequent sections. A *random matrix* is a matrix whose elements are random variables on a probability space. Given a probability space $(\Omega, \mathbf{P})$, set $\mathcal{L} = \bigcap_{1 \leq p < \infty} L^p(\Omega, \mathbf{P})$, the algebra of complex-valued random variables on Ω with finite moments. The expectation $E(f) = \int f d\mathbf{P}$ is a state on $\mathcal{L}$. The set of all $n \times n$ random matrices $X = [x_{ij}]_{i,j=1}^n$ such that $x_{ij} \in \mathcal{L}$ for all i, j is denoted by $M_n(\mathcal{L}) = \mathcal{L} \otimes M_n(\mathbb{C})$, which is a $*$ -algebra with a tracial state

$$\tau_n(X) = \frac{1}{n} \sum_{i=1}^n E(x_{ii}).$$

Then $(M_n(\mathcal{L}), \tau_n)$ is a noncommutative probability space suitable in random matrix theory.

For each $n \in \mathbb{N}$ let a random matrix $X(n)$ be given. If $X(n)$ is Hermitian or real symmetric, then we have the induced probability measure ν_n on $M_n(\mathbb{C})^{sa}$, the $n \times n$ Hermitian matrices, or $M_n(\mathbb{R})^{sa}$, the $n \times n$ real symmetric matrices. Assume that ν_n is invariant under transformations by $\mathcal{U}(n)$ or $\mathcal{O}(n)$. Then ν_n induces the measure $\bar{\nu}_n$ on $\mathbb{R}^n$ (the space of eigenvalues). Define a mapping Λ_n from $M_n(\mathbb{C})^{sa}$ or $M_n(\mathbb{R})^{sa}$ into $\mathcal{M}(\mathbb{R})$ by

$$\Lambda_n(A) = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)},$$

where $\lambda_1(A), \dots, \lambda_n(A)$ are the eigenvalues of A and δ_λ is the Dirac measure at λ . Now we have the probability measures P_n on $\mathcal{M}(\mathbb{R})$ such that

$$\begin{aligned} P_n(\Gamma) &= \nu_n(\Lambda_n^{-1}\Gamma) = \nu_n\left(\left\{A : \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)} \in \Gamma\right\}\right) \\ &= \bar{\nu}_n\left(\left\{t \in \mathbb{R}^n : \frac{1}{n} \sum_{i=1}^n \delta_{t_i} \in \Gamma\right\}\right) \end{aligned}$$

for every Borel set Γ of $\mathcal{M}(\mathbb{R})$. Here $\mathcal{M}(\mathbb{R})$ is endowed with the weak topology $\sigma(\mathcal{M}(\mathbb{R}), C_b(\mathbb{R}))$, $C_b(\mathbb{R})$ being the space of bounded continuous functions on $\mathbb{R}$. Note that $\mathcal{M}(\mathbb{R})$ becomes a Polish space by the Lévy metric for example. We call P_n the *empirical eigenvalue distribution* of $X(n)$.

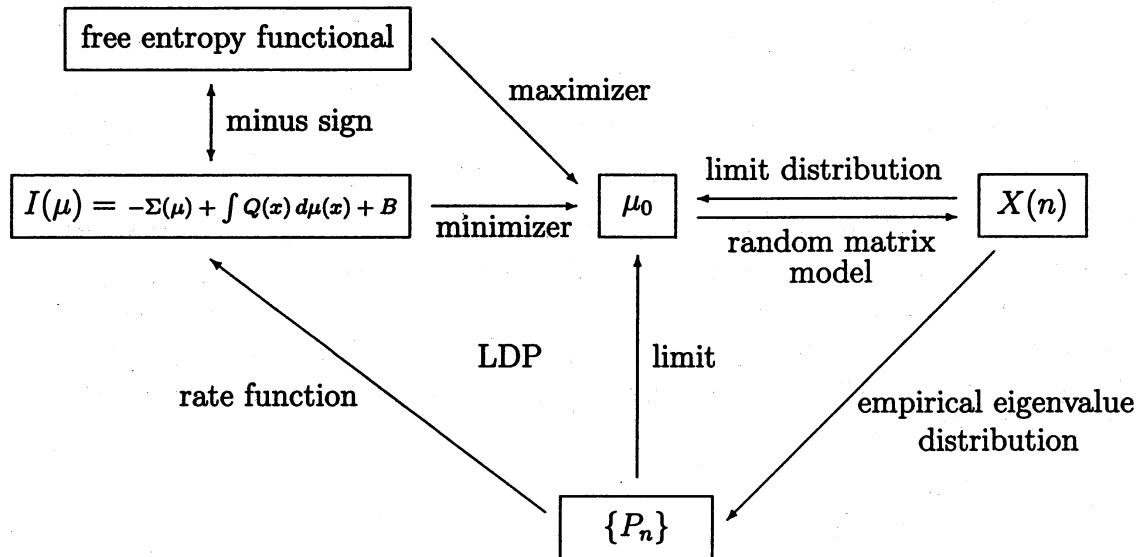
In Sec. 1 we may consider many distributions as a maximizer of a free entropy functional. It is known via Wigner type limit theorem that some of them arise as the

limit distribution μ_0 of a certain random matrix model $X(n)$. Then we may discuss the LDP for $\{P_n\}$ defined as above from $X(n)$. This is the so-called level-2 large deviation theorem [El]. Here it should be natural to expect that a rate function I is the minus sign of the free entropy functional (up to an additive constant) and so μ_0 is a unique minimizer of I with $I(\mu_0) = 0$. In this case, the LDP for $\{P_n\}$ implies that the eigenvalue distribution of $X(n)$ converges exponentially fast to the limit distribution μ_0 . Namely, for any neighborhood G of μ_0 ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G^c) = \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \nu_n \left(\left\{ A : \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)} \in G^c \right\} \right) < 0.$$

In particular, $\{P_n\}$ converges weakly to the point mass at μ_0 .

Our general strategy in the LDP problem associated with random matrices is described as follows.



3. LDP for selfadjoint Gaussian random matrices

In their paper [BG], Ben Arous and Guionnet proved a large deviation theorem for the empirical eigenvalue distribution of selfadjoint Gaussian random matrices. But their proof in [BG] is rather complicated, so that we will give a more transparent proof (in a bit more general setting) though essentially similar.

For $n \in \mathbb{N}$ let $X(n) = [x_{ij}]_{i,j=1}^n$ be an $n \times n$ random matrix such that

- (i) $X(n)$ is selfadjoint, i.e. $x_{ij} = \overline{x_{ji}}$ for $1 \leq i \leq j \leq n$,
- (ii) $\{\operatorname{Re} x_{ij} : 1 \leq i \leq j \leq n\} \cup \{\operatorname{Im} x_{ij} : 1 \leq i < j \leq n\}$ is an independent family of Gaussian random variables,

- (iii) $E(x_{ij}) = 0$ for $1 \leq i \leq j \leq n$, $E(x_{ii}^2) = 1/n$ for $1 \leq i \leq n$, and $E((\operatorname{Re} x_{ij})^2) = E((\operatorname{Im} x_{ij})^2) = 1/2n$ for $1 \leq i < j \leq n$.

Then $X(n)$ is called a *standard selfadjoint Gaussian* $n \times n$ random matrix. The probability measure ν_n on $M_n(\mathbb{C})^{sa}$ induced by $X(n)$ is invariant under $\mathcal{U}(n)$ -transformations so that it induces the measure $\bar{\nu}_n$ on $\mathbb{R}^n$. Moreover, the so-called Wigner theorem says that the mean spectral density of $X(n)$ tends to the semicircle law w_2 as $n \rightarrow \infty$ in distribution. More precisely,

$$\tau_n(X(n)^k) \rightarrow \int x^k dw_2(x) = \begin{cases} \frac{1}{k/2+1} \binom{k}{k/2} & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}$$

as $n \rightarrow \infty$ for all $k \in \mathbb{N}$. (In fact, the Wigner theorem holds true in a much milder assumption, [VDN], [HP2].) In this way, $X(n)$ becomes a standard random matrix model for w_2 . A more refined limit theorem was given by Voiculescu in [Vo1], which includes the assertion about asymptotic freeness.

Since the Hilbert-Schmidt norm of $A = [a_{ij}] \in M_n(\mathbb{C})^{sa}$ is

$$\|A\|_{HS} = \left\{ \sum_{i=1}^n a_{ii}^2 + 2 \sum_{i < j} (\operatorname{Re} a_{ij})^2 + 2 \sum_{i < j} (\operatorname{Im} a_{ij})^2 \right\}^{1/2},$$

the map $\Phi(A) = ((a_{ii})_{1 \leq i \leq n}, (\sqrt{2} \operatorname{Re} a_{ij})_{i < j}, (\sqrt{2} \operatorname{Im} a_{ij})_{i < j})$ is an isometry from $M_n(\mathbb{C})^{sa}$ with $\|\cdot\|_{HS}$ onto $\mathbb{R}^{n^2}$ with the Euclidean norm. So we have the “Lebesgue” measure λ_n on $M_n(\mathbb{C})^{sa}$ transformed from the Lebesgue measure on $\mathbb{R}^{n^2}$ via Φ . This λ_n is of course invariant under unitary transformations. It is known that the induced measure $\bar{\lambda}_n$ on $\mathbb{R}^n$ has the following joint density with respect to the Lebesgue measure $dt_1 \cdots dt_n$:

$$\frac{1}{Z'_n} \prod_{i < j} (t_i - t_j)^2 \quad \text{with} \quad Z'_n = (2\pi)^{-n(n-1)/2} \prod_{j=1}^n j!.$$

The above probability ν_n on $M_n(\mathbb{C})^{sa}$ from the standard Gaussian random matrix $X(n)$ is written as

$$\nu_n = \frac{1}{Z''_n} \exp\left(-\frac{n}{2} \operatorname{Tr} A^2\right) d\lambda_n(A),$$

so that the joint probability density of $\bar{\nu}_n$ with respect to $dt_1 \cdots dt_n$ is

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

where $Z_n = Z'_n Z''_n$ is a normalization constant.

A real symmetric standard Gaussian $n \times n$ random matrix is defined in a similar way. In this case, when ν_n is the induced measure on $M_n(\mathbb{R})^{sa}$, $\bar{\nu}_n$ has the joint density

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

See [Me] for more about joint probability density of Gaussian random matrix eigenvalues.

From now on we will treat a more general probability on $M_n(\mathbb{C})^{sa}$. Let $Q(x)$ be a real continuous function on $\mathbb{R}$ such that for any $\varepsilon > 0$

$$\lim_{|x| \rightarrow \infty} |x| \exp(-\varepsilon Q(x)) = 0. \quad (3.1)$$

This is satisfied if for instance $Q(x) = \frac{1}{2}|x|^p$ where $p > 0$. For each $n \in \mathbb{N}$ let ν_n be a $\mathcal{U}(n)$ -invariant probability measure on $M_n(\mathbb{C})^{sa}$ and assume that the induced measure $\bar{\nu}_n$ on $\mathbb{R}^n$ has the joint probability density

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

where $\beta > 0$ is fixed (independent of n) and C_n is a normalization constant, i.e.

$$Z_n = \int \cdots \int \exp\left(-n \sum_{i=1}^n Q(t_i)\right) \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n.$$

The finiteness of this integral is guaranteed by assumption (3.1). Let P_n be the empirical eigenvalue distribution on $\mathcal{M}(\mathbb{R})$ corresponding to ν_n defined in Sec. 2, so that

$$P_n(\Gamma) = \bar{\nu}_n\left(\left\{t \in \mathbb{R}^n : \frac{1}{n} \sum_{i=1}^n \delta_{t_i} \in \Gamma\right\}\right)$$

for every Borel set Γ of $\mathcal{M}(\mathbb{R})$.

Then the LDP proved by Ben Arous and Guionnet can be extended as follows:

Theorem 3.1. *The finite limit $B = \lim_{n \rightarrow \infty} n^{-2} \log Z_n$ exists and $\{P_n\}$ satisfies the LDP in the scale n^{-2} with good rate function*

$$I(\mu) = -\beta \Sigma(\mu) + \int Q(x) d\mu(x) + B. \quad (3.2)$$

Furthermore, there exists a unique $\mu_0 \in \mathcal{M}(\mathbb{R})$ such that $I(\mu_0) = 0$.

For instance, when $p > 0$ and $Q(x) = \frac{1}{2}|x|^p$, it follows from (2) in Sec. 1 that the limit distribution μ_0 is the Ullman distribution $v_r^{(p)}$ with $r = (2\beta/p\alpha_p)^{1/p}$. Since

$$\begin{aligned} B &= \beta \Sigma(v_r^{(p)}) - \frac{1}{2} \int |x|^p v_r^{(p)}(x) dx \\ &= \beta \left(\log \frac{r}{2} - \frac{1}{2p} \right) - \frac{\alpha_p r^p}{2} \\ &= \frac{\beta}{p} \log \frac{2\beta}{p\alpha_p} - \beta \log 2 - \frac{3\beta}{2p}, \end{aligned}$$

the following asymptotic limit is a by-product of the theorem:

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n = \frac{\beta}{p} \log \frac{2\beta}{p\alpha_p} - \beta \log 2 - \frac{3\beta}{2p}.$$

In particular when $p = 2$, rate function (3.2) is

$$I(\mu) = -\beta \Sigma(\mu) + \frac{1}{2} \int x^2 d\mu(x) + \frac{\beta}{2} \log \beta - \frac{3\beta}{4}$$

and $\mu_0 = w_{2\sqrt{\beta}}$. This is the case shown in [BG], and in this case, Selberg's integral formula (see [Me]) gives

$$Z_n = (2\pi)^{n/2} n^{-\frac{1}{2}(\beta n(n-1)+n)} \prod_{j=1}^n \frac{\Gamma(1+j\beta)}{\Gamma(1+\beta)}$$

and the asymptotic limit

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n = \frac{\beta}{2} \log \beta - \frac{3}{4}\beta$$

is directly obtained by the Stirling formula.

To prove the theorem, set

$$\phi(x, y) = -\beta \log |x - y| + \frac{1}{2}(Q(x) + Q(y)),$$

$$\phi_\alpha(x, y) = \min\{\phi(x, y), \alpha\} \quad \text{for } \alpha > 0.$$

Since

$$\phi(x, y) \geq -\beta \{\log(|x| \exp(-Q(x)/2\beta)) + \log(|y| \exp(-Q(y)/2\beta))\}$$

whenever $|x|, |y| \geq 2$, it follows that $\phi_\alpha(x, y)$ is bounded and continuous, so that

$$\mu \in \mathcal{M}(\mathbb{R}) \mapsto \iint \phi_\alpha(x, y) d\mu(x) d\mu(y)$$

is continuous in weak topology. Hence

$$\begin{aligned}
& -\beta \Sigma(\mu) + \int Q(x) d\mu(x) \\
& = \iint \phi(x, y) d\mu(x) d\mu(y) = \sup_{\alpha > 0} \iint \phi_\alpha(x, y) d\mu(x) d\mu(y)
\end{aligned}$$

is lower semicontinuous in weak topology on $\mathcal{M}(\mathbb{R})$. For simplicity write

$$\mu_t = \frac{1}{n} \sum_{i=1}^n \delta_{t_i} \quad \text{for } t = (t_1, \dots, t_n) \in \mathbb{R}^n.$$

The method in proving Lemmas 3.2 and 3.3 is essentially the same as Ben Arous and Guionnet's.

Lemma 3.2.

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n \leq - \inf_{\mu \in \mathcal{M}(\mathbb{R})} \iint \phi(x, y) d\mu(x) d\mu(y). \quad (3.3)$$

Proof. We get

$$\begin{aligned}
Z_n &= \int \cdots \int \exp \left(- \sum_{i=1}^n Q(t_i) \right) \\
&\quad \times \exp \left\{ - \sum_{i < j} (Q(t_i) + Q(t_j)) \right\} \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n \\
&= \int \cdots \int \exp \left(- \sum_{i=1}^n Q(t_i) \right) \exp \left\{ - 2 \sum_{i < j} \phi(t_i, t_j) \right\} dt_1 \cdots dt_n \\
&\leq \int \cdots \int \exp \left(- \sum_{i=1}^n Q(t_i) \right) \\
&\quad \times \exp \left\{ - n^2 \iint_{\{x \neq y\}} \phi(x, y) d\mu_t(x) d\mu_t(y) \right\} dt_1 \cdots dt_n \\
&\leq \exp \left\{ - n^2 \inf_{\mu} \iint_{\{x \neq y\}} \phi(x, y) d\mu(x) d\mu(y) \right\} \\
&\quad \times \int \cdots \int \exp \left(- \sum_{i=1}^n Q(t_i) \right) dt_1 \cdots dt_n \\
&= \left(\int e^{-Q(x)} dx \right)^n \exp \left\{ - n^2 \inf_{\mu} \iint \phi(x, y) d\mu(x) d\mu(y) \right\},
\end{aligned}$$

implying (3.3). □

Lemma 3.3. For every $\mu \in \mathcal{M}(\mathbb{R})$,

$$\begin{aligned}
& \inf_G \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \right\} \\
& \leq - \iint \phi(x, y) d\mu(x) d\mu(y) - \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n,
\end{aligned} \tag{3.4}$$

where G runs over a neighborhood base of μ .

Proof. For any neighborhood G of $\mu \in \mathcal{M}(\mathbb{R})$ put

$$\bar{G} = \{t \in \mathbb{R}^n : \mu_t \in G\}.$$

As in the proof of Lemma 3.2 we get

$$\begin{aligned}
P_n(G) &= \bar{\mu}_n(\bar{G}) \\
&= \frac{1}{Z_n} \int \cdots \int_{\bar{G}} \exp \left(- \sum_{i=1}^n Q(t_i) \right) \exp \left\{ -2 \sum_{i < j} \phi(t_i, t_j) \right\} dt_1 \cdots dt_n \\
&\leq \frac{1}{Z_n} \int \cdots \int_{\bar{G}} \exp \left(- \sum_{i=1}^n Q(t_i) \right) \\
&\quad \times \exp \left\{ -n^2 \iint \phi_\alpha(x, y) d\mu_t(x) d\mu_t(y) + n\alpha \right\} dt_1 \cdots dt_n \\
&= \frac{1}{Z_n} \left(\int e^{-Q(x)} dx \right)^n \exp \left\{ -n^2 \inf_{\mu' \in G} \iint \phi_\alpha(x, y) d\mu'(x) d\mu'(y) + n\alpha \right\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \\
& \leq - \inf_{\mu' \in G} \iint \phi_\alpha(x, y) d\mu'(x) d\mu'(y) - \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n.
\end{aligned}$$

Thanks to weak continuity of $\mu' \mapsto \iint \phi_\alpha(x, y) d\mu'(x) d\mu'(y)$ we get

$$\inf_G \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \right\} \leq - \iint \phi_\alpha(x, y) d\mu(x) d\mu(y) - \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n.$$

Letting $\alpha \rightarrow +\infty$ yields inequality (3.4). $\square$

Lemma 3.4. For every $\mu \in \mathcal{M}(\mathbb{R})$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n \geq - \iint \phi(x, y) d\mu(x) d\mu(y) \tag{3.5}$$

and

$$\begin{aligned}
& \inf_G \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \right\} \\
& \geq - \iint \phi(x, y) d\mu(x) d\mu(y) - \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n,
\end{aligned} \tag{3.6}$$

where G runs over a neighborhood base of μ .

Proof. It is clear that

$$\mu \in \mathcal{M}(\mathbb{R}) \mapsto \inf \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) : G \text{ a neighborhood of } \mu \right\}$$

is upper semicontinuous. Since $\phi(x, y)$ is bounded below, if $\iint \phi(x, y) d\mu(x) d\mu(y) < +\infty$ and $\mu_k = \mu([-k, k])^{-1} \chi_{[-k, k]} \mu$, then

$$\iint \phi(x, y) d\mu(x) d\mu(y) = \lim_{k \rightarrow \infty} \iint \phi(x, y) d\mu_k(x) d\mu_k(y).$$

So we may assume that μ has a compact support. For $\varepsilon > 0$ let φ_ε be a nonnegative C^∞ -function supported in $[-\varepsilon, \varepsilon]$ such that $\int \varphi_\varepsilon(x) dx = 1$, and $\varphi_\varepsilon * \mu$ be the convolution of μ with φ_ε . Thanks to concavity and upper semicontinuity of $\Sigma(\mu)$ restricted on probability measures with uniformly bounded supports, it is easy to see that

$$\Sigma(\varphi_\varepsilon * \mu) \geq \Sigma(\mu).$$

Also

$$\lim_{\varepsilon \rightarrow +0} \int Q(x) d(\varphi_\varepsilon * \mu)(x) = \int Q(x) d\mu(x).$$

Hence we may assume that μ has a continuous density with compact support. Moreover, let λ be the uniform distribution on an interval $[a, b]$ including $\text{supp}(\mu)$. Then it suffices to show the required inequalities for each $(1 - \delta)\mu + \delta\lambda$ ($0 < \delta < 1$). After all, we may assume that μ has a continuous density $f > 0$ on $\text{supp}(\mu) = [a, b]$ so that $\delta \leq f(x) \leq \delta^{-1}$ ($a \leq x \leq b$) for some $\delta > 0$.

The following proof is a modification of that of [Vo3, Proposition 4.5]. For each $n \in \mathbb{N}$ let $a = b_0^{(n)} < a_1^{(n)} < b_1^{(n)} < a_2^{(n)} < \dots < a_n^{(n)} < b_n^{(n)} = b$ be such that

$$\int_a^{a_j^{(n)}} f(x) dx = \frac{j - \frac{1}{2}}{n}, \quad \int_a^{b_j^{(n)}} f(x) dx = \frac{j}{n} \quad (1 \leq j \leq n).$$

Then it immediately follows that

$$\frac{\delta}{2n} \leq b_j^{(n)} - a_j^{(n)} \leq \frac{1}{2n\delta} \quad (1 \leq j \leq n).$$

Define

$$\Delta_n = \{(t_1, \dots, t_n) \in \mathbb{R}^n : a_j^{(n)} \leq t_j \leq b_j^{(n)}, 1 \leq j \leq n\}.$$

For any neighborhood G of μ , it is clear that

$$\Delta_n \subset \bar{G} = \{t \in \mathbb{R}^n : \mu_t \in G\}$$

for all n large enough. Therefore for large n we have

$$\begin{aligned}
P_n(G) &= \bar{\mu}_n(\bar{G}) \geq \bar{\mu}_n(\Delta_n) \\
&= \frac{1}{Z_n} \int \cdots \int_{\Delta_n} \exp\left(-n \sum_{i=1}^n Q(t_i)\right) \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n \\
&\geq \frac{1}{Z_n} \exp\left(-n \sum_{i=1}^n \xi_i^{(n)}\right) \prod_{i < j} (a_j^{(n)} - b_i^{(n)})^{2\beta} \int \cdots \int_{\Delta_n} dt_1 \cdots dt_n \\
&\geq \frac{1}{Z_n} \left(\frac{\delta}{2n}\right)^n \exp\left(-n \sum_{i=1}^n \xi_i^{(n)}\right) \prod_{i < j} (a_j^{(n)} - b_i^{(n)})^{2\beta},
\end{aligned}$$

where $\xi_i^{(n)} = \max\{Q(x) : a_i^{(n)} \leq x \leq b_i^{(n)}\}$. Now let $h : [0, 1] \rightarrow [a, b]$ be the inverse function of $t \mapsto \int_a^t f(x) dx$. Since $a_j^{(n)} = h((j - \frac{1}{2})/n)$ and $b_j^{(n)} = h(j/n)$, we get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i^{(n)} = \int_0^1 Q(h(t)) dt = \int_a^b Q(x) f(x) dx = \int Q(x) d\mu(x)$$

and

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{2}{n^2} \sum_{i < j} \log(b_j^{(n)} - a_i^{(n)}) &= 2 \iint_{0 \leq s < t \leq 1} \log(h(t) - h(s)) ds dt \\
&= \int_0^1 \int_0^1 \log |h(s) - h(t)| ds dt \\
&= \iint f(x) f(y) \log |x - y| dx dy = \Sigma(\mu).
\end{aligned}$$

Therefore

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

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \geq - \iint \phi(x, y) d\mu(x) d\mu(y) - \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n,$$

as desired. □

Lemma 3.5. *The finite limit $B = \lim_{n \rightarrow \infty} \log Z_n$ exists.*

Proof. By (3.3) and (3.5) we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n \leq - \inf_{\mu} \iint \phi(x, y) d\mu(x) d\mu(y) \leq \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n.$$

This gives the result because Theorem 1.1 says that $\mu \mapsto \iint \phi(x, y) d\mu(x) d\mu(y)$ attains the minimum. $\square$

Lemma 3.6. $\{P_n\}$ is exponentially tight.

Proof. For any $\alpha > 0$ set

$$K_\alpha = \left\{ \mu \in \mathcal{M}(\mathbb{R}) : \int Q(x) d\mu(x) \leq \alpha \right\}.$$

Since $Q(x) \rightarrow +\infty$ as $|x| \rightarrow +\infty$ by assumption (3.1), it is easy to see that

$$\sup_{\mu \in K_\alpha} (\{x : |x| \geq r\}) \rightarrow 0 \quad \text{as } r \rightarrow +\infty$$

and hence K_α is compact in weak topology (for example, consider the Lévy metric on $\mathcal{M}(\mathbb{R})$). We get

$$\begin{aligned} P_n(K^c) &= \bar{\nu} \left(\left\{ t \in \mathbb{R}^n : \frac{1}{n} \sum_{i=1}^n Q(t_i) > \alpha \right\} \right) \\ &= \frac{1}{Z_n} \int \cdots \int_{\left\{ \frac{1}{n} \sum_{i=1}^n Q(t_i) > \alpha \right\}} \exp \left(-n \sum_{i=1}^n Q(t_i) \right) \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n \\ &\leq \frac{1}{Z_n} \exp \left(-\frac{n^2 \alpha}{2} \right) \int \cdots \int \exp \left(-\frac{n}{2} \sum_{i=1}^n Q(t_i) \right) \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n. \end{aligned}$$

When $Q(x)$ is replaced by $Q(x)/2$, the finite limit

$$B_2 = \lim_{n \rightarrow \infty} \frac{1}{n^2} \log \int \cdots \int \exp \left(-\frac{n}{2} \sum_{i=1}^n Q(t_i) \right) \prod_{i < j} |t_i - t_j|^{2\beta} dt_1 \cdots dt_n$$

exists as well as (3.6). Hence the above estimate gives

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(K^c) \leq -B + B_2 - \frac{\alpha}{2}.$$

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

End of proof of Theorem 3.1. By (3.6) and Lemma 3.5,

$$I(\mu) = \iint \phi(x, y) d\mu(x) d\mu(y) + B \geq 0$$

for all $\mu \in \mathcal{M}(\mathbb{R})$. Hence I satisfies all conditions as a rate function. Now (3.4), (3.6), and Lemma 3.6 show thanks to Proposition 2.2 that $\{P_n\}$ satisfies the LDP with good rate function I . So there exists $\mu_0 \in \mathcal{M}(\mathbb{R})$ such that $I(\mu_0) = 0$. But Theorem 1.1 says that μ_0 is a unique minimizer of I on $\mathcal{M}(\mathbb{R})$. $\square$

4. LDP for non-selfadjoint Gaussian random matrices

For $n \in \mathbb{N}$ let $X(n) = [x_{ij}]_{i,j=1}^n$ be an $n \times n$ random matrix such that

- (i) $\{\operatorname{Re} x_{ij} : 1 \leq i, j \leq n\} \cup \{\operatorname{Im} x_{ij} : 1 \leq i, j \leq n\}$ is an independent family of Gaussian random variables,
- (ii) $E(x_{ij}) = 0$ and $E((\operatorname{Re} x_{ij})^2) = E((\operatorname{Im} x_{ij})^2) = 1/2n$ for $1 \leq i, j \leq n$.

Then $X(n)$ is called a *standard non-selfadjoint Gaussian* $n \times n$ random matrix. It is known [Vo1, Theorem 3.3] that $X(n)$ converges in distribution as $n \rightarrow \infty$ to a circular element. More precisely, let $(\mathcal{M}, \tau)$ be a noncommutative probability space consisting of a von Neumann algebra $\mathcal{M}$ and a faithful normal tracial state τ . An element $X \in \mathcal{M}$ is called a *circular* element if $\{(X + X^*)/\sqrt{2}, (X - X^*)/\sqrt{2}i\}$ is a free pair of selfadjoint elements with the semicircle distribution w_2 (see [VDN] for freeness). Then the above statement means that

$$\lim_{n \rightarrow \infty} \tau_n(P(X(n), X(n)^*)) = \tau(P(X, X^*))$$

for every polynomial P of two non-commuting indeterminates. In this way, $X(n)$ becomes a standard model of a circular element in free probability theory. Moreover, the limit distribution of $X(n)^*X(n)$ is the Marchenko-Pastur distribution μ_1 in (4) of Sec. 1, and that of $|X(n)| = (X(n)^*X(n))^{1/2}$ is the quarter-circular distribution $\frac{1}{\pi} \sqrt{4 - x^2} \chi_{[0,4]}(x) dx$.

The Hilbert-Schmidt norm of $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{C})$ is

$$\|A\|_{HS} = \left\{ \sum_{i,j=1}^n (\operatorname{Re} a_{ij})^2 + \sum_{i,j=1}^n (\operatorname{Im} a_{ij})^2 \right\}^{1/2},$$

and $\mathcal{M}_n(\mathbb{C})$ is isometric to $\mathbb{R}^{2n^2}$. So the “Lebesgue” measure λ_n on $\mathcal{M}_n(\mathbb{C})$ is given by

$$d\lambda_n(A) = \prod_{i,j=1}^n d(\operatorname{Re} a_{ij}) d(\operatorname{Im} a_{ij}),$$

and the probability measure ν_n on $\mathcal{M}_n(\mathbb{C})$ induced by $X(n)$ is written as

$$\nu_n = \frac{1}{Z'_n} \exp(-n \operatorname{Tr} A^* A) d\lambda_n(A).$$

Then ν_n induces the probability measure $\bar{\nu}_n$ on $\mathbb{C}^n$ (the space of eigenvalues). According to [Me, 15.1, A32, A35], the joint probability density of $\bar{\nu}_n$ is

$$\frac{1}{Z_n} \exp\left(-n \sum_{i=1}^n |\zeta_i|^2\right) \prod_{i < j} |\zeta_i - \zeta_j|^2$$

with respect to $d\zeta_1 \cdots d\zeta_n$ ($d\zeta_i$ is the Lebesgue measure on the plane), while this derivation is a bit more difficult than the selfadjoint case in Sec. 3. Here

$$Z_n = \pi^n n^{-n(n-1)/2+n} \prod_{j=1}^n j!$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n = -\frac{3}{4}.$$

Although the Gaussian random matrix $X(n)$ given above is far from normal and cannot be diagonalized, one can consider its empirical eigenvalue distribution. A bit more generally, for each $n \in \mathbb{N}$ let ν_n be a probability measure on $M_n(\mathbb{C})$ invariant under unitary transformations, and assume that the induced measure $\bar{\nu}_n$ on $\mathbb{C}^n$ has the joint probability density

$$\frac{1}{Z_n} \exp\left(-n \sum_{i=1}^n |\zeta_i|^2\right) \prod_{i < j} |\zeta_i - \zeta_j|^{2\beta},$$

where $\beta > 0$ is fixed. Define $\Lambda_n : M_n(\mathbb{C}) \rightarrow \mathcal{M}(\mathbb{C})$ by

$$\Lambda_n(A) = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)},$$

where $\lambda_i(A)$ are the eigenvalues of A . The empirical eigenvalue distribution P_n on $\mathcal{M}(\mathbb{C})$ corresponding to ν_n is given by

$$P_n(\Gamma) = \nu_n(\Lambda_n^{-1}\Gamma) = \bar{\nu}_n\left(\left\{(\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n : \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i} \in \Gamma\right\}\right)$$

for every Borel set Γ of $\mathcal{M}(\mathbb{C})$.

Then we have the large deviation theorem for $\{P_n\}$ as follows.

Theorem 4.1. *With the above assumptions,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n = \frac{\beta}{2} \log \beta - \frac{3\beta}{4},$$

and $\{P_n\}$ satisfies the LDP in the scale n^{-2} with good rate function

$$I(\mu) = -\beta \Sigma(\mu) + \int |\zeta|^2 d\mu(\zeta) + \frac{\beta}{2} \log \beta - \frac{3\beta}{4} \quad \text{on } \mathcal{M}(\mathbb{C}).$$

Furthermore, the uniform distribution on the disk $\{\zeta \in \mathbb{C} : |\zeta| \leq \sqrt{\beta}\}$ is a unique minimizer of I .

$$\phi(\zeta, \eta) = -\beta \log |\zeta - \eta| + \frac{1}{2}(|\zeta|^2 + |\eta|^2) \quad \text{for } \zeta, \eta \in \mathbb{C}.$$

Then $\phi(\zeta, \eta)$ is bounded below and

$$-\beta \Sigma(\mu) + \int |\zeta|^2 d\mu(\zeta) = \iint \phi(\zeta, \eta) d\mu(\zeta) d\mu(\eta) \quad \text{for } \mu \in \mathcal{M}(\mathbb{C}).$$

The following lemma can be proved in the same way as Lemmas 3.2 and 3.3.

Lemma 4.2.

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n \leq - \inf_{\mu \in \mathcal{M}(\mathbb{C})} \iint \phi(\zeta, \eta) d\mu(\zeta) d\mu(\eta)$$

and for every $\mu \in \mathcal{M}(\mathbb{C})$,

$$\begin{aligned} & \inf_G \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \right\} \\ & \leq - \iint \phi(\zeta, \eta) d\mu(\zeta) d\mu(\eta) - \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n, \end{aligned}$$

where G runs over a neighborhood base of μ .

Lemma 4.3. For every $\mu \in \mathcal{M}(\mathbb{C})$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n \geq - \iint \phi(\zeta, \eta) d\mu(\zeta) d\mu(\eta)$$

and

$$\begin{aligned} & \inf_G \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log P_n(G) \right\} \\ & \geq - \iint \phi(\zeta, \eta) d\mu(\zeta) d\mu(\eta) - \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log Z_n, \end{aligned}$$

where G runs over a neighborhood base of μ .

Proof. First, a suitable smoothing process can be performed as in the proof of Lemma 3.4. So we may assume that $\text{supp}(\mu) = [a, b] \times [c, d]$ and μ has a continuous density f on $[a, b] \times [c, d]$ satisfying $\delta \leq f \leq \delta^{-1}$ for some $\delta > 0$. For each $n \in \mathbb{N}$ let $m = \lfloor \sqrt{n} \rfloor$. Let $a = x_0 < x_1 < \dots < x_m = b$ be such that

$$\mu([x_{i-1}, x_i] \times [c, d]) = \frac{1}{m} \quad (1 \leq i \leq m).$$

Noting $m^2 \leq n \leq m(m+2)$ we can choose $c = y_{i,0} < y_{i,1} < \dots < y_{i,l_i} = d$ for $1 \leq i \leq m$ such that $m \leq l_i \leq m+2$, $\sum_{i=1}^m l_i = n$, and

$$\mu([x_{i-1}, x_i] \times [y_{i,j-1}, y_{i,j}]) = \frac{1}{ml_i} \quad (1 \leq i \leq m, 1 \leq j \leq l_i).$$

Arrange n pieces of rectangles $[x_{i-1}, x_i] \times [y_{i,j-1}, y_{i,j}]$ as

$$R_i^{(n)} = [a_i^{(n)}, b_i^{(n)}] \times [c_i^{(n)}, d_i^{(n)}] \quad (1 \leq i \leq n).$$

Also let

$$S_i^{(n)} = [(2a_i^{(n)} + b_i^{(n)})/3, (a_i^{(n)} + 2b_i^{(n)})/3] \times [(2c_i^{(n)} + d_i^{(n)})/3, (c_i^{(n)} + 2d_i^{(n)})/3].$$

Then we get

$$\lim_{n \rightarrow \infty} \left\{ \max_{1 \leq i \leq n} \text{diam}(R_i^{(n)}) \right\} \rightarrow 0, \quad (4.1)$$

$$\int_{S_i^{(n)}} d\zeta = \frac{1}{9} \int_{R_i^{(n)}} d\zeta \geq \frac{\delta}{9} \int_{R_i^{(n)}} f(\zeta) d\zeta \geq \frac{\delta}{9m(m+2)} \geq \frac{\delta}{27n} \quad (1 \leq i \leq n).$$

Define

$$\Delta_n = \{(\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n : \zeta_i \in S_i^{(n)}, 1 \leq i \leq n\}.$$

For any neighborhood G of μ , it is easy to check that

$$\Delta_n \subset \bar{G} = \left\{ (\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n : \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i} \in G \right\}$$

for all n large enough. For such n we have

$$\begin{aligned} P_n(G) &\geq \bar{\nu}_n(\Delta_n) \\ &\geq \frac{1}{Z_n} \int \cdots \int_{\Delta_n} \exp\left(-n \sum_{i=1}^n |\zeta_i|^2\right) \prod_{i < j} |\zeta_i - \zeta_j|^{2\beta} d\zeta_1 \cdots d\zeta_n \\ &\geq \frac{1}{Z_n} \exp\left(-n \sum_{i=1}^n \left(\max_{\zeta \in S_i^{(n)}} |\zeta_i|\right)^2\right) \\ &\quad \times \prod_{i < j} \left(\min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta|\right)^{2\beta} \int \cdots \int_{\Delta_n} d\zeta_1 \cdots d\zeta_n \\ &\geq \frac{1}{Z_n} \left(\frac{\delta}{27n}\right)^n \exp\left(-n \sum_{i=1}^n \left(\max_{\zeta \in S_i^{(n)}} |\zeta_i|\right)^2\right) \prod_{i < j} \left(\min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta|\right)^{2\beta}. \end{aligned}$$

So, to obtain the required inequalities, it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left(\max_{\zeta \in S_i^{(n)}} |\zeta_i|\right)^2 = \int |\zeta|^2 f(\zeta) d\zeta \quad (4.2)$$

and

$$\liminf_{n \rightarrow \infty} \frac{2}{n^2} \sum_{i < j} \log \left(\min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta| \right) \geq \iint f(\zeta) f(\eta) \log |\zeta - \eta| d\zeta d\eta. \quad (4.3)$$

But (4.2) is clear from (4.1). We get

$$\begin{aligned} & \iint f(\zeta) f(\eta) \log |\zeta - \eta| d\zeta d\eta \\ & \leq 2 \sum_{i < j} \int_{R_i^{(n)}} \int_{R_j^{(n)}} f(\zeta) f(\eta) \log |\zeta - \eta| d\zeta d\eta \\ & \leq 2 \sum_{i < j} \log \left(\max_{\zeta \in R_i^{(n)}, \eta \in R_j^{(n)}} |\zeta - \eta| \right) \int_{R_i^{(n)}} f(\zeta) d\zeta \int_{R_j^{(n)}} f(\eta) d\eta. \end{aligned}$$

Since

$$\max_{\zeta \in R_i^{(n)}, \eta \in R_j^{(n)}} |\zeta - \eta| \leq \text{const} \cdot \min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta|$$

and for any $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{2}{n^2} \# \left\{ (i, j) : i < j, \max_{\zeta \in R_i^{(n)}, \eta \in R_j^{(n)}} |\zeta - \eta| \leq (1 + \varepsilon) \min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta| \right\} = 1,$$

we have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left\{ \iint f(\zeta) f(\eta) \log |\zeta - \eta| d\zeta d\eta - \frac{2}{n^2} \sum_{i < j} \log \left(\min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta| \right) \right\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{2}{n^2} \sum_{i < j} \log \left(\frac{\max_{\zeta \in R_i^{(n)}, \eta \in R_j^{(n)}} |\zeta - \eta|}{\min_{\zeta \in S_i^{(n)}, \eta \in S_j^{(n)}} |\zeta - \eta|} \right) = 0, \end{aligned}$$

implying (4.3). □

End of proof of Theorem 4.1. Lemmas 4.2 and 4.3 imply as in Sec. 3 that the finite limit $B = \lim_{n \rightarrow \infty} n^{-2} \log Z_n$ exists and $\{P_n\}$ is exponentially tight, so that $\{P_n\}$ satisfies the LDP with good rate function

$$I(\mu) = -\beta \Sigma(\mu) + \int |\zeta|^2 d\mu(\zeta) + B.$$

So there exists $\mu_0 \in \mathcal{M}(\mathbb{C})$ such that $I(\mu_0) = 0$. But (6) in Sec. 1 says that μ_0 must be equal to the uniform distribution on $\{\zeta \in \mathbb{C} : |\zeta| \leq \sqrt{\beta}\}$ (i.e. $\lambda_{\sqrt{\beta}}^{(2)}$ in notation of (6) in Sec. 1) and

$$\begin{aligned} B &= \beta \Sigma(\mu_0) - \int |\zeta|^2 d\mu_0(\zeta) \\ &= \beta \left(\log \sqrt{\beta} - \frac{1}{4} \right) - \frac{\beta}{2} = \frac{\beta}{2} \log \beta - \frac{3\beta}{4}, \end{aligned}$$

completing the proof of the theorem. $\square$

When $\beta = 1$, the projection of the above μ_0 to the real or imaginary axis is the semicircle law w_1 . On the other hand, the limit distribution of standard non-selfadjoint Gaussian random matrices $X(n)$ is the distribution of a circular element X with radius 2, and the distributions of $(X + X^*)/2$ and $(X - X^*)/2i$ are $w_{\sqrt{2}}$. So the limit distribution through the eigenvalue distribution is the $1/\sqrt{2}$ -compression of the “real” limit. This is not strange because $X(n)$ is non-normal and the spectral radius is smaller than the operator norm.

Let $\mathcal{M}$ be a von Neumann algebra with a faithful normal tracial state τ . For any (non-normal) element $X \in \mathcal{M}$ one can associate a kind of “spectral measure” μ_X called the *Brown measure* for X . A noteworthy result in [Lar] is that the Brown measure of a circular element is the uniform distribution on the unit disk. This seems natural in view of Theorem 4.1.

5. LDP for unitary random matrices

An $n \times n$ unitary random matrix is a $\mathcal{U}(n)$ -valued random variable on a probability measure space. From the probability-theoretic viewpoint, this is equivalent to giving a probability measure ν_n on $\mathcal{U}(n)$ itself. When ν_n is invariant under unitary transformations, it induces the probability measure $\bar{\nu}_n$ on $\mathbb{T}^n$ (the space of eigenvalues). Then we have the empirical eigenvalue distribution P_n such that

$$\begin{aligned} P_n(\Gamma) &= \nu_n \left(\left\{ U \in \mathcal{U}(n) : \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(U)} \in \Gamma \right\} \right) \\ &= \bar{\nu}_n \left(\left\{ \zeta \in \mathbb{T}^n : \frac{1}{n} \sum_{i=1}^n \delta_{\zeta_i} \in \Gamma \right\} \right) \end{aligned}$$

for every Borel set Γ of $\mathcal{U}(n)$.

Now let $Q(\zeta)$ be a real continuous function on $\mathbb{T}$ and for each $n \in \mathbb{N}$ set a probability measure ν_n on $\mathcal{U}(n)$ as

$$\nu_n = \frac{1}{Z'_n} \exp(-n \operatorname{Tr} Q(U)) d\lambda_n(U),$$

where λ_n is the Haar probability measure on the compact group $\mathcal{U}(n)$. It is known [Me, 9.4, 10.1] that the induced measure $\bar{\lambda}_n$ on $\mathbb{T}^n$ has the following joint density with respect to $d\zeta_1 \cdots d\zeta_n$ where $d\zeta_j = d\theta_j/2\pi$ ($\zeta_j = e^{i\theta_j}$):

$$\frac{1}{n!} \prod_{i < j} |\zeta_i - \zeta_j|^2 = \frac{1}{n!} \prod_{i < j} |e^{i\theta_i} - e^{i\theta_j}|^2.$$

Hence the probability density of $\bar{\nu}_n$ with respect to $d\zeta_1 \cdots d\zeta_n$ is given by

$$\frac{1}{Z_n} \exp \left(-n \sum_{i=1}^n Q(\zeta_i) \right) \prod_{i < j} |\zeta_i - \zeta_j|^2,$$

where $Z_n = n! Z'_n$ is a normalization constant.

Then we have the LDP for $\{P_n\}$ as follows.

Theorem 5.1. *The finite limit $B = \lim_{n \rightarrow \infty} n^{-2} \log Z_n = \lim_{n \rightarrow \infty} n^{-2} \log Z'_n$ exists and $\{P_n\}$ satisfies the LDP in the scale n^{-2} with good rate function*

$$I(\mu) = -\Sigma(\mu) + \int_{\mathbf{T}} Q(\zeta) d\mu(\zeta) + B \quad \text{on } \mathcal{M}(\mathbf{T}).$$

Furthermore, there exists a unique $\mu_0 \in \mathcal{M}(\mathbf{T})$ such that $I(\mu_0) = 0$.

This can be proved more or less similarly to Theorems 3.1 and 4.1, while we will present the details elsewhere. Below we just remark a few points.

1° Since $\mathbf{T}$ is compact, the weak topology on $\mathcal{M}(\mathbf{T})$ is the weak* topology and hence the tightness of $\{P_n\}$ is automatic.

2° For every $k \in \mathbb{Z}$ and $\varepsilon > 0$ choose a weak* neighborhood G of μ_0 by

$$G = \{\mu \in \mathcal{M}(\mathbf{T}) : |\mu(\zeta^k) - \mu_0(\zeta^k)| < \varepsilon\}.$$

Then the LDP of $\{P_n\}$ says in particular that

$$P_n(G^c) = \nu_n \left(\left\{ U \in \mathcal{U}(n) : \left| \frac{1}{n} \sum_{i=1}^n \lambda_i(U)^k - \mu_0(\zeta^k) \right| \geq \varepsilon \right\} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

so that

$$\tau_n(U)^k = \int \frac{1}{n} \sum_{i=1}^n \lambda_i(U)^k d\nu_n(U) \rightarrow \mu_0(\zeta^k) \quad \text{as } n \rightarrow \infty.$$

This means that the mean spectral density of the $n \times n$ unitary random matrix distributed according to ν_n converges to μ_0 as $n \rightarrow \infty$. In this way we obtain a rather general Wigner type theorem for unitary random matrices.

3° A by-product of Theorem 5.1 together with Theorem 1.1 is that if Q is a continuous real function on $\mathbf{T}$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \int \cdots \int_{\mathbf{T}^n} \exp \left(-n \sum_{i=1}^n Q(\zeta_i) \right) \prod_{i < j} |\zeta_i - \zeta_j|^2 d\zeta_1 \cdots d\zeta_n$$

exists and it is equal to

$$\max_{\mu \in \mathcal{M}(\mathbb{T})} \left\{ \Sigma(\mu) - \int Q(\zeta) d\mu(\zeta) \right\}.$$

Finally we give two examples of the LDP for unitary random matrices corresponding to (7) and (8) in Sec. 1.

Example 5.2. For each $\alpha \in \mathbb{C}$, $|\alpha| < 1$, let $Q(\zeta) = \log |\zeta - \alpha|^2$ ($\zeta \in \mathbb{T}$). Then the probability measure ν_n on $\mathcal{U}(n)$ is given as

$$\nu_n = \frac{1}{Z'_n} \frac{d\lambda_n(U)}{\det |U - \alpha I|^{2n}}.$$

Hence

$$\bar{\nu}_n = \frac{1}{Z_n} \frac{\prod_{i < j} |\zeta_i - \zeta_j|^2}{\prod_{i=1}^n |\zeta_i - \alpha|^{2n}} d\zeta_1 \cdots d\zeta_n.$$

If P_n is the empirical eigenvalue distribution of the associated unitary random matrix, then Theorem 5.1 and (7) in Sec. 1 say that $\{P_n\}$ satisfies the LDP with rate function

$$I(\mu) = -\Sigma(\mu) + \int \log |\zeta - \alpha|^2 d\mu(\zeta) - \log(1 - |\alpha|^2) \quad \text{on } \mathcal{M}(\mathbb{T}),$$

and the Poisson kernel measure p_α is a unique minimizer of I . Also we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \int_{\mathcal{U}(n)} \frac{d\lambda_n(U)}{\det |U - \alpha I|^{2n}} = -\log(1 - |\alpha|^2).$$

It does not seem easy to directly compute the above asymptotic limit of integrals. In particular when $\alpha = 0$, the eigenvalue distribution of a unitary random matrix distributed according to the Haar measure on $\mathcal{U}(n)$ (called a *standard* unitary random matrix) converges to the Haar measure on $\mathbb{T}$.

Example 5.3. For $\lambda > 0$ let $Q(\zeta) = -\frac{2}{\lambda} \operatorname{Re} \zeta$ ($\zeta \in \mathbb{T}$). Then ν_n on $\mathcal{U}(n)$ is

$$\nu_n = \frac{1}{Z'_n} \exp \left(\frac{n}{\lambda} \operatorname{Tr} (U + U^*) \right) d\lambda_n(U),$$

and $\bar{\nu}_n$ on $\mathbb{T}^n$ is

$$\bar{\nu}_n = \frac{1}{Z_n} \exp \left(\frac{2n}{\lambda} \sum_{i=1}^n \cos \theta_i \right) \prod_{i < j} |e^{i\theta_i} - e^{i\theta_j}|^2 d\theta_1 \cdots d\theta_n.$$

By Theorem 5.1 and (8) in Sec. 1, the associated sequence of empirical eigenvalue distributions satisfies the LDP with rate function

$$I(\mu) = -\Sigma(\mu) - \frac{2}{\lambda} \int \operatorname{Re} \zeta d\mu(\zeta) + B \quad \text{on } \mathcal{M}(\mathbb{T}),$$

where

$$B = \begin{cases} \frac{1}{\lambda^2} & \text{if } \lambda \geq 2, \\ \frac{1}{2} \log \frac{\lambda}{2} + \frac{2}{\lambda} - \frac{3}{4} & \text{if } \lambda < 2, \end{cases}$$

and ρ_λ is a unique minimizer of I . Incidentally, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \int_{\mathcal{U}(n)} \exp \left(\frac{n}{\lambda} \text{Tr} (U + U^*) \right) d\lambda_n(U) = B,$$

which was calculated in [GW].

References

- [BG] G. Ben Arous and A. Guionnet, Large deviation for Wigner's law and Voiculescu's non commutative entropy, manuscript.
- [BIPZ] E. Brézin, C. Itzykson, G. Parisi and J.B. Zuber, Planar diagrams, *Comm. Math. Phys.* **59** (1978), 35–51.
- [DZ] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston-London, 1993.
- [DS] J.D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, 1989.
- [El] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag, New York, 1985.
- [GW] D.J. Gross and E. Witten, Possible third-order phase transition in the large- N lattice gauge theory, *Phys. Rev.* **D21** (1980), 446–453.
- [HP1] F. Hiai and D. Petz, Maximizing free entropy, preprint.
- [HP2] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Information*, in preparation.
- [La] N.S. Landkof, *Foundations of Modern Potential Theory* (translated from the Russian), Springer, Berlin-Heidelberg-New York, 1972.
- [Lar] F. Larsen, Status report for thesis, Univ. of Odense.
- [Me] M.L. Mehta, *Random Matrices*, Academic Press, Boston, 1991.
- [MS1] 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.

- [MS2] H.N. Mhaskar and E.B. Saff, Weighted analogues of capacity, transfinite diameter, and Chebyshev constant, *Constr. Approx.* **8** (1992), 105–124.
- [To] V. Totik, *Weighted Approximation with Varying Weight*, Lecture Notes in Math., Vol. 1569, Springer, Berlin, 1994.
- [Vo1] D. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.* **104** (1991), 201–220
- [Vo2] D. Voiculescu, The analogues of entropy and of Fisher’s information measure in free probability theory, I, *Comm. Math. Phys.* **155** (1993), 71–92.
- [Vo3] D. Voiculescu, The analogues of entropy and of Fisher’s information measure in free probability theory, II, *Invent. Math.* **118** (1994), 411–440.
- [VDN] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, CRM Monograph Ser., Vol. 1, Amer. Math. Soc., 1992.