The Analogues of Entropy and of Fisher's Information Measure in Free Probability Theory, I

Dan Voiculescu[*]

I.H.E.S., F-91440 Bures-sur-Yvette, France[**]

Dedicated to Huzuhiro Araki

Abstract. Analogues of the entropy and Fisher information measure for random variables in the context of free probability theory are introduced. Monotonicity properties and an analogue of the Cramer-Rao inequality are proved.

Introduction

In [25] we began studying operator algebra free products from the probabilistic point of view. The idea is to look at free products as an analogue of tensor products and to develop a corresponding highly noncommutative probabilistic framework where freeness is given a treatment similar to independence. We showed [25] that there is a free central limit theorem with the semicircle law playing the role of the Gaussian distribution and that there is a functor from Hilbert spaces to operator algebras, which is the free analogue of the Gaussian functor of second quantization (i.e. the Gaussian process indexed by a Hilbert space). For the addition and multiplication of bounded free random variables we introduced corresponding free convolution operations on the distributions and constructed linearizing transforms, i.e. analogues of the logarithm of the Fourier and respectively Mellin transforms [26, 27]. In this context one-parameter free convolution semigroups correspond to one-dimensional quasilinear complex conservation laws satisfied by the Cauchy-transforms of the distributions, the complex Burger equation, in particular, being the analogue of the heat equation. This free harmonic analysis has been extended to distributions of unbounded random variables and the infinitely divisible laws have been studied [26, 4, 14, 5].

The explanation for the occurrence of the semicircle law, both in the free central limit theorem and in Wigner's work on asymptotics of large random matrices, was found in [29]. We showed that asymptotically, entrywise independence of large

---

Gaussian matrices gives rise to freeness. Thus, large random matrices provide an asymptotic model for free probability theory. This has had important applications to the study of the II_1 factors of free groups [30, 19, 8, 20].

In usual probability theory the entropy $S(X)$ and Fisher's information measure $J(X)$ of a real random variable X are given by the formulae

$$S(X) = - \int p(t) \log p(t) dt \,,$$

$$J(X) = \int \frac{(p'(t))^2}{p(t)} \, dt \,,$$

where $p(t)$ is the density of the distribution $d\mu$ of X [assumed smooth in the definition of $J(X)$].

The theme of the present paper is to introduce the free entropy $\Sigma(X)$ and the free information measure $\Phi(X)$ given by

$$\Sigma(X) = \int \int p(s)p(t) \log |s - t| ds \, dt \,,$$

$$\Phi(X) = \int (p(t))^3 dt$$

and to argue that in free probability theory $\Sigma(X)$ and $\Phi(X)$ are the analogues of $S(X)$ and $J(X)$. The formulae for $\Sigma(X)$ and $\Phi(X)$ coincide with the logarithmic energy [12] of the distribution of X, while $\Phi(X)$ is just the cube of the L^3-norm of p.

The motivation for these definitions is from the asymptotic random-matrix model. Ideally, $\Sigma(X)$ should be a certain normalized limit of the entropies of the maximum entropy distributions for random matrices with spectral density given by the distribution of X. To handle this properly, however, would require advances in the theory of random matrices. Instead we proceed as follows. We first present a heuristic (nonrigorous) justification of the formula for $\Sigma(X)$ based on random matrices. Then we take the formula for $\Sigma(X)$ as a definition, derive the definition of $\Phi(X)$ so that it should relate to $\Sigma(X)$ via a free Gaussian perturbation, as $J(X)$ relates to $S(X)$ via usual Gaussian perturbations. With these definitions, we then proceed to give direct proofs (entirely rigorous) of several properties of $\Sigma(X)$, $\Phi(X)$ analogous to those of $S(X)$, $J(X)$. This includes results on monotonicity, semicontinuity and the analogue of the Cramer-Rao inequality. The monotonicity results turn out to be more general, since they hold also for Riesz energies and respectively L^p norms ($p > 1$). One-variable complex analysis topics such as univalence and subordination as well as some potential theory play an important role in our proofs. The structure of the paper is as follows. In Sect. 1 we assemble for the reader's convenience preliminary material from free probability theory.

The next section, Sect. 2 contains the heuristic justification for the definition of $\Sigma(X)$ via random matrices. It is the only non-rigorous part of this paper.

In Sect. 3 we define $\Phi(X)$ and establish the connection with $\Sigma(X)$ via free Gaussian processes, i.e. via the complex Burger equation.

In Sect. 4 we prove results on monotonicity for $\Phi(X)$ via analytic subordination results for free convolution.

The free analogue of the Cramer-rao inequality is obtained in Sect. 5 and we show that equality holds exactly for semicircle distributions.

In Sect. 6 we obtain monotonicity results for $\Sigma(X)$ and Riesz-energies. We also prove the analogue for Σ and Φ of the integral formula which relates S and J [3].

The semicontinuity properties of Σ and Φ in the W^*-algebra context are obtained in Sect. 7. We also prove inequalities for some quantities generalizing Σ.

The last section (Sect. 8), after pointing out that the monotonicity properties of Σ, Φ also hold for independent random variables we provide some infinitesimal evidence in favor of these properties being true for orthogonal algebras.

The author greatly benefitted from the stimulating scientific atmosphere at the Institut des Hautes Etudes Scientifiques, where this work was done.

1. Preliminaries

This section is devoted to recalling some definitions and results from free probability theory [25, 26, 29, 31].

A *non-commutative probability space* (A, φ) consists of a unital algebra over $\mathbb{C}$ equipped with a *state* $\varphi : A \to \mathbb{C}$, i.e. a unit-preserving linear functional. This purely algebraic frame-work can be endowed with additional structure corresponding to positivity or to weak convergence in the corresponding forms of C^*-probability spaces or W^*-probability spaces.

Random variables are elements of A, where (A, φ) is as above. To discuss more general "unbounded" random variables in the W^*-context, one may also consider more generally unbounded operators affiliated to the W^*-algebra A.

1.1. Definition. A family of subalgebras $1 \in A_\iota \subset A$ ($\mathbb{C} \in I$) in a non-commutative probability space is called *free* if $\varphi(a_1 \ldots a_n) = 0$ whenever $a_j \in A_{\iota(j)}$ with $\iota(j) \neq \iota(j+1)$ ($1 \leq j \leq n-1$) and $\varphi(a_j) = 0$ ($1 \leq j \leq n$). A family of random variables f_ι ($\iota \in I$) is free if the family of subalgebras generated by $(1, f_i)$ is free.

1.2. Note that freeness, in general, implies a high degree of non-commutativity among subalgebras. The typical example of free subalgebras being $A = \mathbb{C}[G]$ the group-ring of a group $G = G_1 * G_2$ which is the free product of two subgroups, $A_j = \mathbb{C}[G_j]$ ($j = 1, 2$) and φ the trace on $\mathbb{C}[G]$ given by $\varphi(\Sigma c_g g) = c_e$.

1.2. Definition. If $(f_\iota)_{\iota \in I}$ are random variables in (A, φ), their joint distribution μ is the functional $\mu : \mathbb{C}\langle X_\iota | \iota \in I \rangle \to \mathbb{C}$ given by $\mu = \varphi \circ h$, where $\mathbb{C}\langle X_\iota | \iota \in I \rangle$ is the free algebra with unit and generators X_ι ($\iota \in I$) and $h : \mathbb{C}\langle X_\iota | \iota \in I \rangle \to A$ the unique homomorphism with $h(X_\iota) = f_i$.

In particular for one random variable f the distribution is a linear functional $\mu_f : \mathbb{C}[X] \to \mathbb{C}$ and in the C^*-context if $f = f^*$, this functional naturally extends from the polynomials to a compactly supported probability measure on $\mathbb{R}$ (also denoted μ_f).

1.3. If $X, Y \in A$ are a free pair of random variables then the distribution of $X + Y$ depends only on the distributions of X, Y and thus there is an operation, called additive free convolution such that $\mu_{X+Y} = \mu_X \boxplus \mu_Y$. This is entirely analogous to usual convolution arising from addition of independent random variables.

This operation, via the C^*-algebra context, induces an operation of free convolution among compactly supported measures on $\mathbb{R}$ [26]. Passing to unbounded random variables this can be further extended to arbitrary probability measures on $\mathbb{R}$ [14, 5].

1.4. Theorem [26]. *If μ is the distribution of a random variable let*

$$G_\mu(z) = z^{-1} + \sum_{n \geq 1} \mu(X^n) z^{-n-1}$$

and let $K_\mu(z)$ and $R_\mu(z)$ be formal power series satisfying

$$G_\mu(K_\mu(z)) = z,$$

$$K_\mu(z) - z^{-1} = R_\mu(z).$$

If $\mu_3 = \mu_1 \boxplus \mu_2$ then

$$R_{\mu_3} = R_{\mu_1} + R_{\mu_2}.$$

1.5. If μ is a compactly supported probability measure then $G_\mu(z)$ is the Cauchy transform

$$G_\mu(z) = \int \frac{d_\mu(\zeta)}{z - \zeta}$$

and K_μ is defined by inverting G_μ locally at infinity. The extension of 1.4 to arbitrary probability measures [14, 5] is roughly that instead of inverting the Cauchy transforms in a neighborhood of infinity we will invert G_μ in a suitable domain with an angle at infinity. Note that R_μ plays the role of the logarithm of the Fourier transform for free convolution. Using Theorem 1.4 the free convolution of two probability measures is computed by finding first the Cauchy transform and then the measure via "boundary values" of the imaginary part of the Cauchy transform.

1.6. If $(\mu_t)_{t \geq 0}$ is a semigroup with respect to free convolution, i.e. $\mu_{t+s} = \mu_t \boxplus \mu_s$ then $G(t, z) = G_{\mu_t \boxplus \nu}(z)$ satisfies the complex quasilinear equation

$$\frac{\partial G}{\partial t} + \frac{\partial G}{\partial z} \varphi(G) = 0,$$

where $\varphi(z) = R_{\mu_1}(z)$ with initial data $G(0, z) = G_\nu(z)$. Note that in this context μ is the analogue of the Gaussian law if $R_\mu(z) = \alpha z$ in which case the probability measure μ has a "semi-circle" distribution. The analogue of the heat equation being then the complex Burger equation

$$\frac{\partial G}{\partial t} + \frac{\partial G}{\partial z} \alpha(G) = 0,$$

1.7. It was shown in [26, 4, 14, 5] that infinite divisibility of a probability measure μ on $\mathbb{R}$ with respect to free convolution, amounts to the requirement that $R_\mu(z)$ have an extension to an analytic function with positive imaginary part in the upper half-plane.

1.8. The last topic being asymptotic freeness for random matrices we begin with a definition. A sequence of families of random variables $(f_\iota^{(n)})_{\iota \in K}$ ($n \in \mathbb{N}$) is called asymptotically free if the distributions μ_n converge pointwise on $\mathbb{C}\langle X_\iota | \iota \in I \rangle$ to some functional μ_∞ and the $(X_\iota)_{\iota \in I}$ are a free family of random variables in $(\mathbb{C}\langle X_\iota | \iota \in I \rangle, \mu_\infty)$.

1.9. The natural framework for random matrices is to consider

$$M_n = \bigcap_{1 \leq p \leq \infty} L^p(\Sigma, \mathfrak{M}_n),$$

where $L^p(\Sigma, \mathfrak{M}_n)$ is the L^p-space over a standard measure space $(\Sigma, d\sigma)$ with values in the $n \times n$ matrices $\mathfrak{M}_n$. On M_n there is a natural trace φ_n given by

$$\varphi_n(X) = \int \tau_n(X(\omega)) d\sigma(\omega)$$

with $\tau_n = n^{-1} \operatorname{Tr}$ the normalized trace on $\mathfrak{M}_n$.

1.10. Theorem [29]. *Let*

$$Y(s,n) = (a(i,j;n,s))_{1 \leq i,j \leq n} \in M_n$$

be random matrices, $s \in I$. Assume $a(i,j;n,s) = \overline{a(j,i;n,s)}$ and that

$$\{\operatorname{Re}(a(i,j;n,s)) | 1 \leq i \leq j \leq n, s \in \mathbb{N}\} \cup \{\operatorname{Im} a(i,j;n,s) | 1 \leq i < j \leq n, s \in \mathbb{N}\}$$

are independent Gaussian random variables such that $\operatorname{Re} a(i,j;n,s)$ and $\operatorname{Im} a(i,j; n,s)$ are $(0,(2n)^{-1})$ for $1 \leq i < j \leq n$ and $a(j,j;n,s)$ is $(0,n^{-1})$. Then $(Y(s,n))_{s \in I}$ is asymptotically free as $n \to \infty$ (the limit distributions for the $Y(s,n)$'s being semicircle laws according to Wigner's classical results.

2. Random Matrix Heuristics

Let X be a random variable (in some C^*-probability space) with distribution a compactly supported measure ν on $\mathbb{R}$. The idea for defining the free entropy $\Sigma(X)$ is to approximate X by a self-adjoint random matrix X_n with distribution (in the classical sense) a probability measure σ_n on $\mathfrak{M}_n^h$ (the hermitian $n \times n$ matrices) and to define $\Sigma(X)$ as a normalized limit of the entropies of the σ_n. The motivation for this is provided by Theorem 1.10 which shows that for suitable choices of the random-matrix approximates, independence of $\mathfrak{M}_n^h$-valued random-variables corresponds asymptotically to freeness. Hence $\Sigma(X)$ defined in this way should have the desired property of behaving with respect to freeness (in particular free convolution of the distributions) in the same way as usual entropy behaves with respect to independence (and in particular convolution of the distributions).

If this random matrix "approximation" procedure would have a sharp form, one might conceive that the properties of $\Sigma(X)$ should be direct consequences of the corresponding properties of $S(X_n)$. At present this approach encounters several difficulties.

One difficulty is that choosing X_n to have the maximum entropy $U(n)$-invariant distribution for the given level density ν (i.e. the distribution introduced by Balian in [2]) the corresponding random matrices have not been studied in much detail. There are also problems with the use instead of functions of the Gaussian random matrices (like in Wigner's classical work).

The other difficulty is that we would need a substantially stronger form of our asymptotic freeness result (Theorem 1.10) so that it should guarantee that the addition of independent random-matrix approximants yields normalized entropies which relate in the limit to the normalized entropies for the random matrices with level densities given by the free convolution.

In spite of these difficulties, this approach is of such conceptual simplicity that it would be worthwhile to give it a rigorous basis.

Having made these remarks, let us pass to the formula for $\Sigma(X)$, which is based on familiar computations in random-matrix theory (see [2, 15]).

Assume $\nu = p_* \mu$, where $p : \mathbb{R} \to \mathbb{R}$ is a diffeomorphism and μ is the semicircle law. Let $X_n = p(Y_n)$, where Y_n is a Gaussian random matrix normalized as $Y(s,n)$ in Theorem 1.10. Then σ_n the distribution of X_n on $\mathfrak{M}_n^h$ is the push-forward of the Gaussian measure on $\mathfrak{M}_n^h$ via $\mathfrak{M}_n^h \ni A \to p(A) \in \mathfrak{M}_n^h$. Since these measures are

invariant with respect to conjugation by unitary matrices, it suffices to look at the densities on the sets of eigenvalues $(y_1, \ldots, y_n)$.

Up to constants the densities for the Gaussian measure, the measure σ_n and Lebesgue measure are then:

$$\exp\left(-\sum_i ny_i^2\right)\prod_{j<k}(y_j - y_k)^2\,,$$

$$\exp\left(-\sum_i nq(y_i)^2\right)\prod_{j<k}(q(y_j) - q(y_k))^2\prod_i q'(y_i)$$

and

$$\prod_{j<k}(y_j - y_k)^2\,,$$

where q is the inverse of p. Hence up to constants the entropy of σ_n is the integral of the following quantity:

$$\left(n\sum_i q(y_i)^2 - \sum_i \log q'(y_i) + \sum_{j<k}\log\frac{(y_j - y_k)^2}{(q(y_j) - q(y_k))^2}\right)$$

$$\times \exp\left(-n\sum_i q(y_i)^2\right)\prod_{j<k}(q(y_j) - q(y_k))^2\prod_i q'(y_i)\,.$$

One divides this integral by n^2 and based on the assumption that the measures on $\mathbb{R}$ giving mass $\dfrac{1}{n}$ to each y_i, will converge to ν for "most" $(y_1, \ldots, y_n)$ with respect to the measures σ_n, and hence the limit is identified with

$$\int q(y)^2 d\nu(y) + \frac{1}{2}\int\int d\nu(x)d\nu(y)\log\frac{(x-y)^2}{(q(x) - q(y)^2)}$$

$$= \int x^2 d\mu(x) + \int\int d\nu(x)d\nu(y)\log|x - y|$$

$$- \int\int d\mu(x)d\mu(y)\log|x - y|\,.$$

Leaving aside the constants we end up with

$$\int\int d\nu(x)d\nu(y)\log|x - y| \tag{2.1}$$

which is the formula for $\Sigma(X)$.

2.2. Definition. If X is a random variable whose distribution extends to a compactly supported measure ν on $\mathbb{R}$, we define the free entropy $\Sigma(X)$ to be equal to the quantity in (2.1).

We shall frequently abuse notations and write instead of $\Sigma(X)$ also $\Sigma(\nu)$ or if ν is Lebesgue-absolutely continuous and v is its density, we shall write $\Sigma(v)$. We will also use these notations whenever (2.1) makes sense, though ν may not be compactly supported.

3. $\Phi(X)$

In this section we derive the formula of the free information measure $\Phi(X)$ from the formula of the free entropy $\Sigma(X)$, a straightforward computation up to a few technicalities.

3.1 Definition. If X is a random-variable, the distribution of which extends to a compactly supported measure μ, absolutely continuous with respect to Lebesgue measure, we define the free information measure $\Phi(X)$ by

$$\Phi(X) = \int v^3(t)dt ,$$

where v is the density of μ. If μ is not Lebesgue absolutely continuous we set $\Phi(X) = \infty$

We will also write $\Phi(v)$ or $\Phi(\mu)$ and we will use this notation also in the case of unbounded support. Note that $\Phi(\mu) = \lim_{\varepsilon \downarrow 0} \Phi(\mu * P_\varepsilon)$, where P_ε is the Poisson kernel.

The lemma we prove here connecting $\Sigma(X)$ and $\Phi(X)$ will get its full power only after we prove the monotonicity results in the next sections.

3.2. Lemma. *Let X, Y be a free pair of random variables such that the distribution of X extends to a compactly supported measure on $\mathbb{R}$, while Y is a (0,1)-semicircular random variable. Then the function $[0, \infty) \ni s \to \Phi(X + \sqrt{s}Y) \in [0, \infty]$ is measurable and we have*

$$-\Sigma(X + \sqrt{T}Y) + \frac{2}{3} \int_0^T \Phi(X + \sqrt{s}Y)ds = -\Sigma(X) .$$

Proof. Let $\mu(t)$ denote $\mu_{X+\sqrt{t}Y}$ and remark that $\mu(t) = \mu(0) \boxplus \mu_{\sqrt{t}Y}$. Like for Gaussian variables, we have that $\sqrt{t}Y$ has the same distribution as a semicircular free convolution semigroup. Hence [26] the Cauchy transform $G(z, t)$ of $\mu(t)$ satisfies

$$\frac{\partial G}{\partial t} + G \frac{\partial G}{\partial z} = 0 \quad \text{for} \ \operatorname{Im} z > 0 \text{ and } t \geq 0 \text{ and } G(z, 0) = G_{\mu(0)}(z) ,$$

where $G_{\mu(0)}$ is the Cauchy transform of $\mu(0)$.

Using a free product of commutative von Neumann algebras, we may replace X, Y by an equivalent pair of self-adjoint random-variables in a finite von Neumann algebra with a trace-state. In particular since $\|X + \sqrt{t}Y\|$ for $0 \leq t \leq T$ is uniformly bounded we infer that $\operatorname{supp} \mu(t) \subset K$, where K is a compact set.

Let $-\pi^{-1}G(z, t) = u(z, t) + iv(z, t)$ so that the distributional boundary values of $v(\bullet + i\varepsilon, t)$ as $\varepsilon \downarrow 0$ are the $\mu(t)$. It will suffice to show that

$$s \to \Phi(v(\bullet + i\varepsilon, s)) \text{ is continuous}$$

and

$$\int_0^T \Phi(v(\bullet + i\varepsilon, s))ds - \sum (v(\bullet + i\varepsilon, T)) = -\sum (v(\bullet + i\varepsilon, 0)) . \qquad (*)$$

Indeed letting $\varepsilon \downarrow 0$ we infer that $s \to \Phi(v(\bullet, s))$ is measurable and

$$\int\limits_0^T \Phi(v(\bullet + i\varepsilon))ds \uparrow \int\limits_0^T \Phi(v(\bullet, s))ds \quad \text{as} \quad \varepsilon \downarrow 0 \,.$$

On the other hand, since $\operatorname{supp} \mu(s) \subseteq K$, expressing v as a Poisson integral, we have that

$$v(x + i\varepsilon, s) = O((1 + x^2)^{-1})$$

for $0 \le s \le T$ and fixed ε. Using this, together with

$$v(\bullet + i\varepsilon, s) = P_\varepsilon * \mu(s)$$

(where P_ε is the Poisson kernel) one easily gets

$$\sum (v(\bullet + i\varepsilon, s)) = \int \int v(x + i\varepsilon, s)v(y + i\varepsilon, s) \log |x - y| dx \, dy$$

$$= \int \int \log |(x - y) + 2i\varepsilon| d\mu(s)(x) d\mu(s)(y)$$

and hence also

$$\sum (v(\bullet + i\varepsilon, s)) \downarrow \sum (\mu(s))$$

as $\varepsilon \downarrow 0$.

The continuity of $\Phi(v(\bullet + i\varepsilon, s))$ can be seen as follows

$$|v(x + i\varepsilon, s) - v(x + i\varepsilon, t)|$$
$$\le |G(x + i\varepsilon, s) - G(x + i\varepsilon, t)|$$
$$= |\tau(((X + \sqrt{s}Y) - (x + iy)I)^{-1} - ((X + \sqrt{t}Y) - (x + iy)I)^{-1})|$$
$$\le C|\sqrt{s} - \sqrt{t}| (1 + x^2))^{-1} \,.$$

Finally, to prove $(*)$ we will use the differential equation for $G(z, t)$. Note that for fixed ε, we may view it as an equation on $\mathbb{R} + i\varepsilon$ since $\partial G / \partial z = \partial G / \partial x$. Thus for fixed $\varepsilon > 0$, and $0 \le s \le T$ we have

$$\left| \frac{\partial G}{\partial t} (x + i\varepsilon, s) \right| = \left| \left(G \frac{\partial G}{\partial z} \right)(x + i\varepsilon, s) \right|$$

$$\le d(x + i\varepsilon, K)^{-1} d(x + i\varepsilon, K)^{-2}$$

$$\le C(1 + |x|^3)^{-1} \,.$$

Hence, replacing ε by $\varepsilon + \delta$, we have that

$$\Sigma(v(\bullet + i(\varepsilon + \delta), s)) = \int \int v(x + i\varepsilon, s)v(y + i\varepsilon, s) \log |(x - y) + 2i\delta| dx \, dy$$

is differentiable as a function of s and

$$\frac{\partial}{\partial s}\,\Sigma(v(\bullet + i(\varepsilon + \delta)), s))$$

$$= -2\int\int (uv)_x(x + i\varepsilon, s)v(y + i\varepsilon, s)\,\mathrm{Re}\log(x - y + 2i\delta)dx\,dy$$

$$= 2\int\int (uv)\,(x + i\varepsilon, s)v(y + i\varepsilon, s)\,\mathrm{Re}(x - y + 2i\delta)^{-1}dx\,dy$$

$$= 2\int (uv)\,(x + i\varepsilon, s)v(x + i(\varepsilon + 2\delta))dx\,dy\,.$$

Thus

$$\Sigma(v(\bullet + i(\varepsilon + \delta)), T) - \Sigma(v(\bullet + i(\varepsilon + \delta)), 0)$$

$$= 2\int\limits_{0}^{T}\int (uv)\,(x + i\varepsilon, s)u(x + i(\varepsilon + 2\delta))dx\,ds$$

and there are no problems with passing to the limit $\delta \downarrow 0$, so that

$$\Sigma(v(\bullet + i\varepsilon, T)) - \Sigma(v(\bullet + i\varepsilon, 0)) = 2\int\limits_{0}^{T}\int (u^2 v)\,(x + i\varepsilon, s)dx\,ds\,.$$

To conclude the proof it suffices to use a known fact about the Hilbert transform, which we record, for further use, as the next lemma. Q.E.D.

3.3. Lemma. *Let* $v \in L^3(\mathbb{R})$ *and let* $u = -Hv$, *where* H *denotes the Hilbert transform. Then we have*

$$3\int u^2 v\,dx = \int v^3 dx\,.$$

Proof. In view of the L^3-continuity of the Hilbert transform it suffices to prove the lemma when v has compact support. Then $G(z) = \int v(x)\,(z - x)^{-1}dx$ has boundary values $-\pi(u + iv)$ and since $|G(z)| \leq (1 + |z|)^{-1}$ for $\mathrm{Re}\,z \geq \varepsilon$, we easily infer $\int G^3(x + i\varepsilon)dx = 0$. Since G is in the Hardy space H^3, we may take boundary values as $\varepsilon \downarrow 0$ and get the desired relation. Q.E.D.

3.4. Remark. The scaling behavior of Σ and Φ is given by the formulae:

$$\Sigma(aX) = \Sigma(X) + \log a\,,$$
$$\Phi(aX) = a^{-2}\Phi(X)\,,$$

where $a > 0$.

4. Monotonicity and Subordination

In this section we prove that generically free convolution leads to analytic subordination among the Cauchy transforms of probability measures on $\mathbb{R}$. This implies monotonicity results for the L^P-norms of the densities of the measures. In particular for $p = 3$ we get monotonicity for the free information and hence for the free entropy.

We will use analytic subordination in the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$. This is quite similar to the well-known result in the disk [7]. Though we have no doubt that the result for the half-plane has been known for a long time, since we could not find a suitable reference we record it as the next lemma.

4.1. Lemma. *Let $f \in L^p(\mathbb{R})$ ($1 \leq p \leq \infty$) and let $F(x + iy) = (P_y * f)(x)$ be the harmonic extension to $\mathbb{H}$ (P_y the Poisson kernel). If $\omega : \mathbb{H} \to \mathbb{H}$ is analytic and $\lim\limits_{z \to \infty} |\omega(z)| = \infty$ and $\operatorname{Im} \omega(z) \geq \operatorname{Im} z$ then we have*

$$\|F(\bullet + i\varepsilon)\|_p \geq \|F(\omega(\bullet + i\varepsilon))\|_p$$

where $\varepsilon > 0$.

Proof. Let

$$E(\bullet + iy) = P_{y-\varepsilon} * |F(\bullet + i\varepsilon)|^p$$

be the harmonic extension of $|F|^p$ from $\mathbb{R} + i\varepsilon$ to $\mathbb{H} + i\varepsilon$. It is a bounded harmonic function in $\mathbb{H} + i\varepsilon$. By Jensen's inequality $E(x + iy) \geq |F(x + iy)|^p$ if $y \geq \varepsilon$. Hence

$$\|F(\omega(\bullet + i\varepsilon))\|_p^p \leq \int E(\omega(x + i\varepsilon)) dx.$$

Since $yP_y \uparrow 1$ as $y \uparrow \infty$ we have

$$\int\limits_{-\infty}^{\infty} E(\omega(x + i\varepsilon)) dx = \lim_{y \to +\infty} (y - \varepsilon) E(\omega(iy))$$

$$\leq \lim_{y \to +\infty} \frac{y - \varepsilon}{\operatorname{Im} \omega(iy)} \cdot \operatorname{Im} \omega(iy) \cdot E(\omega(iy))$$

$$\leq \limsup_{y \to +\infty} \operatorname{Im} \omega(iy) \cdot E(\omega(iy)) \leq \int\limits_{-\infty}^{\infty} E(x + i\varepsilon) dx$$

$$= \|F(\bullet + i\varepsilon)\|_p. \quad \text{Q.E.D.}$$

The next lemma is essentially a fact about the reciprocals of Cauchy-transforms from [14] (with a simpler proof).

4.2. Lemma. *Let $F : \mathbb{H} \to \mathbb{H}$ be analytic and assume $\lim\limits_{|z| \to \infty} \operatorname{Im}(F(z) - z) = 0$. Then we have $\operatorname{Im} F(z) \geq \operatorname{Im} z$ for $z \in \mathbb{H}$. In particular if G is the Cauchy transform of a compactly supported probability measure on $\mathbb{R}$, then*

$$\operatorname{Im} \frac{1}{G(z)} \geq \operatorname{Im} z.$$

Proof. The minimum principle applied to $\mathrm{Im}((F(z) - z)$ in $\Omega = (\mathbb{H} + i\varepsilon) \cap \{z \in \mathbb{H} \,|\, |z| \leq R\}$ yields

$$\mathrm{Im}(F(z) - z) \geq \min\left(-\varepsilon, \inf_{|z|=R} \mathrm{Im}(F(z) - z)\right) \quad \text{for} \quad z \in \Omega$$

and we let $R \to \infty$. Using the expansion of $1/G$ at ∞ it is clear it satisfies the assumptions on F.   Q.E.D.

4.3. Proposition. *Let μ_j ($j = 1, 2$) be compactly supported probability measures on $\mathbb{R}$ and let $\mu_3 = \mu_1 \boxplus \mu_2$. Then we have*

$$G_3(\mathbb{H}) \subset G_1(\mathbb{H}) \cap G_2(\mathbb{H}),$$

where G_j is the Cauchy-transform of μ_j.

Proof. Let K_j, R_j be the functions that go with G_j (see Theorem 1.4) so that $R_1 + R_2 = R_3$ or equivalently $R_1 + K_2 = K_1 + R_2 = K_3$ in a neighborhood of 0. Since $R_j(G_j(z)) = K_j(G_j(z)) - (1/G_j)(z) = z - (1/G_j)(z)$ we infer that $R_j(G_j(z))$, initially defined in a neighborhood of ∞ has an analytic continuation to $\mathbb{H}$ and $\mathrm{Im}\, R_j(G_j(z)) \leq 0$ for $z \in \mathbb{H}$ by Lemma 4.2.

By $D_j = \{z \in \mathbb{H} | G'_j(z) = 0\}$ we denote the critical points of G_j and by $\Delta_j = G_j(D_j)$ the critical values and let $\Delta = \Delta_1 \cup \Delta_2 \cup \Delta_3$.

Assume $\zeta \in G_3(z)$ for some $z \in \mathbb{H}$. There is a polygonal path $\gamma : [0, 1] \to \mathbb{C}$ such that $\gamma(0) = 0$, $\gamma(1) = \zeta$, $\gamma((0, 1)) \subset (-\mathbb{H}) \backslash \Delta$ and such that K_3 has an analytic continuation along $\gamma([0, 1))$ with $K_3(\gamma([0, 1))) \cap \mathbb{R} = \emptyset$ and $\lim_{t \to 1} K_3(\gamma(t)) = z$.

Let $0 < \varepsilon < 1$ be maximal such that K_j ($j = 1, 2$) have an analytic continuation along $\gamma([0, \varepsilon))$. We shall prove $\varepsilon = 1$.

Indeed, assume $\varepsilon < 1$, since $\gamma((0, 1)) \cap \Delta = \emptyset$ analytic continuation would break down only if $\liminv_{t \to \varepsilon} \mathrm{Im}\, K_j(\gamma(t)) = 0$ or if $\limsup_{t \to \varepsilon} |K_j(\gamma(t))| = \infty$, for $j = 1$ or $j = 2$. The second possibility is immediately discarded since it would imply

$$\liminv_{t \to \varepsilon} |\gamma(t)| = \liminv_{t \to \varepsilon} |G_j(K_j(\gamma(t)))| = 0$$

which is obviously excluded.

On the other hand $R_j(z) = K_j(z) - z^{-1}$ ($j = 1, 2$) will have analytic continuation along $\gamma((0, \varepsilon))$ and since

$$\gamma((0, \varepsilon)) = G_j(K_j(\gamma((0, \varepsilon)))) \subset G_j(\mathbb{H})$$

we infer $R_j(\gamma((0, \varepsilon))) \subset -\mathbb{H}$. Hence for $0 < t < \varepsilon$,

$$\mathrm{Im}\, K_1(\gamma(t)) = \mathrm{Im}\, K_3(\gamma(t)) - \mathrm{Im}\, R_2(\gamma(t))$$
$$\geq \mathrm{Im}\, K_3(\gamma(t)) \text{ and similarly } \mathrm{Im}\, K_2(\gamma(t)) \geq \mathrm{Im}\, K_3(\gamma(t)).$$

Thus for $j = 1, 2$ we have

$$\liminv_{t \to \varepsilon} \mathrm{Im}\, K_j(\gamma(t)) \geq \liminv_{t \to \varepsilon} \mathrm{Im}\, K_3(\gamma(t)) > 0.$$

Hence K_j ($j = 1, 2$) can be analytically continued along $\gamma((0, 1))$ and

$$\liminv_{t \to 1} \mathrm{Im}\, K_j(\gamma(t)) \geq \liminv_{t \to 1} K_3(\gamma(t)) > 0.$$

Again $\limsup_{t \to 1} |K_j(\gamma(t))| = \infty$ is immediately excluded. If $\gamma(1) \notin \Delta$, K_j has analytic continuation to $\gamma(1)$. If $\gamma(1) \in \Delta_j$, since K_j may have a branch-point at $\gamma(1)$ and γ is polygonal we again infer $\lim_{t \to 1} \mathring{K}_j(\gamma(t))$ exists and $\in \mathbb{H}$. Hence

$$\zeta = \gamma(1) = \lim_{t \to 1} G_j(K_j(\gamma(t))) = G_j\left(\lim_{t \to 1} K_j(\gamma(t))\right) \in G_j(\mathbb{H}). \quad \text{Q.E.D.}$$

4.4. Proposition. *Let μ_j, G_j satisfy the assumptions of 4.3 and let Δ_j be the set of critical values of G_j. Assume $\Delta_1 \cap \Delta_2 = \emptyset$. Then there are analytic functions*

$$\omega_j : \mathbb{H} \to \mathbb{H} \text{ such that } G_3 = G_j \circ \omega_j \quad (j = 1, 2).$$

Moreover $\lim_{z \to \infty} \text{Im}(\omega_j(z) - z) = 0$.

Proof. Clearly it will suffice to show there is ω_1. Also, the additional condition involving the imaginary part of ω, at ∞ is immediate from the series expansions at ∞.

Thus in a neighborhood of ∞ the function ω_1 is defined by

$$\omega_1 = G_1^{-1} \circ G_3 = K_1 \circ K_3^{-1}$$

and we must prove that this can be analytically continued to $\mathbb{H}$. By the preceding proposition $G_3(\mathbb{H}) \subset G_1(\mathbb{H})$ and hence, it is easily seen that it will suffice to prove the following:

if γ_1, γ_3 are paths on the Riemann sphere, $\gamma_1((0, 1]) \subset \mathbb{H}\backslash(\Delta_1 \cup \Delta_2)$, $\gamma_j(0) = \infty$, $\gamma_j(1) = z_j$ $(j = 1, 3)$ such that $G_1(\gamma_1(t)) = G_3(\gamma_3(t))$ and if $G_1(z) - G_1(z_1)$ has a zero of order $p > 1$ at z_1 then $G_3(z) - G_3(z_3)$ also has a zero of order p at z_3.

To see that this is indeed so, remark that in view of the definition of G_3 via K_3, when we perform analytic continuation to define G_3 at z_3, the formula $K_3(\zeta) = K_1(\zeta) + K_2(\zeta) - \zeta^{-1}$, where ζ is in a neighborhood of $\zeta_0 = G_3(z_3)$, still holds, but in the sense of multivalued functions. The assumption $\Delta_1 \cap \Delta_2 = \emptyset$ implies that if K_1 has a branching point of order $p > 1$ at ζ_0 then ζ_0 is not a branching point for K_2 and hence ζ_0 is a branching point of order exactly p at ζ_0 for K_3, which means $G_3(z) - G_3(z_3)$ has a zero of order p at z_3. Q.E.D.

4.5. Remark. The proofs of 4.3 and 4.4 are easily extended to show that for μ_j as in 4.3 we have

$$\Delta_1 \cup \Delta_2 \supset \Delta_3 \supset (\Delta_1 \cup \Delta_2)\backslash(\Delta_1 \cap \Delta_2).$$

4.6. Remark. In case μ_1 is infinitely divisible with respect to free convolution, the function R_1 is defined in a neighborhood of $(\mathbb{C}\backslash\mathbb{R})\cup\{0\}$ (Theorem 4.3 in [26]) which implies that G_1 is univalent. Hence in this case 4.4 follows immediately from 4.3.

The analytic subordination result in 4.4 has immediately consequences for the densities of the measures.

4.7. Proposition. *Let μ_1, μ_2 be compactly supported probability measures on $\mathbb{R}$ and let $\mu_3 = \mu_1 \boxplus \mu_2$. Then, for G_j the Cauchy transforms of μ_j we have*

$$\|G_3(\bullet + i\varepsilon)\|_p \leq \|G_j(\bullet + i\varepsilon)\|_p,$$

$$\|\operatorname{Im} G_3(\bullet + i\varepsilon)\|_p \leq \|\operatorname{Im} G_j(\bullet + i\varepsilon)\|_p,$$

$$\|\operatorname{Re} G_3(\bullet + i\varepsilon)\|_p \leq \|\operatorname{Re} G_j(\bullet + i\varepsilon)\|_p,$$

where $j = 1, 2$ and $p \in (1, \infty]$. In particular, if μ_1 is Lebesgue absolutely continuous with density $v_1 \in L^p(\mathbb{R})$ ($p \in (1, \infty]$) then μ_3 is also Lebesgue absolutely continuous and its density v_3 satisfies

$$\|v_3\|_p \leq \|v_1\|_p .$$

Proof. Clearly the results in the case when μ_1 is Lebesgue absolutely continuous are obtained immediately from the inequalities on $\mathbb{R} + i\varepsilon$, so it will suffice to prove those. If μ_1, μ_2 satisfy the assumptions of Proposition 4.4 then $G_3 = G_j \circ \omega_j$ ($j = 1, 2$) and using Lemma 4.2 we find $\operatorname{Im} \omega_j(z) \geq \operatorname{Im} z$. Also clear from the expansions at ∞ is that $\lim_{z \to \infty} |\omega_j(z)| = \infty$. We apply 4.1 with ε replaced by $\varepsilon/2$ to $f = P_{\varepsilon/2} * G_j$ (respectively $P_{\varepsilon/2} * \operatorname{Im} G_j$, $P_{\varepsilon/2} * \operatorname{Re} G_j$) so that $F(z) = G_j(z + i\varepsilon/2)$ [respectively $\operatorname{Im} G_j(z + i\varepsilon/2)$, $\operatorname{Re} G_j(z + i\varepsilon/2)$] and we take $\omega(z) = \omega_j(z + i\varepsilon/2) - i\varepsilon/2$. Then $F(\omega(\bullet + i\varepsilon/2)) = G_3(\bullet + i\varepsilon)$ (and respectively the imaginary part or real part) and 4.1 gives the desired inequalities.

To prove the inequalities in general we will approximate μ_2 by a sequence $\mu_{2,n}$ ($n \in \mathbb{N}$) so that $\mu_1, \mu_{2,n}$ satisfy the assumptions of 4.4. If we define $\mu_{2,n}$ via a homothethy then $G_{2,n}(z) = a_n G_2(a_n z)$ for some real numbers $a_n \to 1$. Remark that if λ is a critical value for G_2, then $a_n \lambda$ is a critical value for $G_{2,n}$ and hence since G_1 has countably many critical values, we may choose a_n so that the critical values of G_1 and $G_{2,n}$ form disjoint sets. Thus $\mu_1, \mu_{2,n}, \mu_{3,n} = \mu_1 \boxplus \mu_{2,n}$ satisfy the assumptions of 4.4. Note also that realizing μ_1, μ_2 as distributions of a free pair of selfadjoint random variables X, Y in a C^*-probability space, $\mu_{3,n}$ is realized as the distribution of $X + a_n^{-1} Y$ and hence if $\operatorname{Im} z \geq \varepsilon$, we have

$$|G_{3,n}(z) - G_3(z)| \leq C|a_n - 1| \cdot |z|^{-1}$$

so that

$$\lim_{n \to \infty} \|G_{3,n}(\bullet + i\varepsilon) - G_3(\bullet + i\varepsilon)\|_p = 0 . \quad \text{Q.E.D.}$$

5. The Free Analogue of the Cramer-Rao Inequality

Replacing in the Cramer-Rao inequality the Fisher information measure by the free information measure, we have a free analogue for that inequality, which we prove here. We also show that equality occurs only for semi-circle laws.

5.1. Theorem. *Let $v \geq 0$, $v \in L^1(\mathbb{R}) \cap L^3(\mathbb{R})$ and assume also $\int v \, dx = 1$, $\int x^2 v \, dx < \infty$. Then with $x_0 = \int xv \, dx$ we have*

$$\left(\int v^3 \, dx \right) \left(\int (x - x_0)^2 v \, dx \right) \geq \frac{3}{4\pi^2} .$$

Proof. Replacing v by a translate, it will suffice to prove

$$\left(\int v^3 \, dx \right) \left(\int x^2 v \, dx \right) \geq \frac{3}{4\pi^2} .$$

Moreover we may assume v has compact support. To see this, let

$$v_n = (v \chi_n) \left(\int v \chi_n dx \right)^{-1} ,$$

where χ_n is the characteristic function of $[-n, n]$ and remark that the inequality for v follows from those for the v_n by letting $n \to \infty$.

Remark further that among all rearrangements of v the minimum of $\int x^2 v\, dx$ is attained when v is symmetric $(v(x) = v(-x))$ and non-increasing on the positive half-axis $(0 \leq x_1 \leq x_2 \Rightarrow v(x_1) \geq v(x_2))$. Since rearrangements leave $\int v^3 dx$ unchanged, we may assume v satisfies these two conditions. Hence the Fourier transform $\hat{v}(\xi) = \hat{v}(-\xi)$ is assumed real and symmetric.

With H denoting the Hilbert transform, using Lemma 3.3 and the Schwarz-inequality we have

$$\int v^3 \, dx \int x^2 v \, dx = 3 \int (Hv)^2 v \, dx \int x^2 v \, dx \geq 3 \left(\int v(Hv) x \, dx \right)^2.$$

On the other hand

$$\int v(Hv) x \, dx = -\frac{1}{2\pi} \int (\hat{v})'(\xi) \hat{v}(\xi) \operatorname{sign} \xi \, d\xi$$

$$= -\frac{1}{2\pi} \int_0^\infty ((\hat{v}(\xi))^2)' d\xi = \frac{1}{2\pi} (\hat{v}(0))^2$$

$$= \frac{1}{2\pi}. \quad \text{Q.E.D.}$$

5.2. Proposition. *Equality holds in the inequality of Theorem 5.1 iff v is a semicircle law, i.e.*

$$v(x) = \begin{cases} 2\pi^{-1} a^{-2} (a^2 - (x - x_0)^2)^{1/2} & \text{if } |x - x_0| \leq a \\ 0 & \text{if } |x - x_0| > a \end{cases}$$

for some $a > 0$.

Proof. From the proof of the theorem we easily infer that in order to have equality the following conditions must be satisfied:
a) v is symmetric about x_0, i.e. $v(x) = v(2x_0 - x)$,
b) $x_0 \leq x_1 \leq x_2 \Rightarrow v(x_1) \geq v(x_2)$,
c) $\lambda(x - x_0) v^{1/2} = (Hv) v^{1/2}$ for some λ.

In view of a) and b) the set on which $v > 0$ is an interval (possibly infinite) centered at x_0. Thus, for some $a \in (0, \infty]$ we will have

$$|x - x_0| < a \Rightarrow v(x) > 0,$$
$$|x - x_0| > a \Rightarrow v(x) = 0,$$
$$x - x_0 = Hv(x) \text{ almost everywhere on } (x_0 - a, x_0 + a).$$

It is easy to check that the semicircle laws satisfy these conditions. The last equality is a singular integral equation for which solution formulae are known (see [16, Chap. 11 and the references therein]). However, for our purposes here we only need to check uniqueness of a solution we know in advance and this can be done directly. It is clear by examining Fourier transforms that there is no solution if $a = \infty$ so we will assume $0 < a < \infty$. Further to simplify notations a translation on $\mathbb{R}$ will immediately reduce the discussion to the case $x_0 = 0$. Denoting by P the projection of $L^2(\mathbb{R})$ onto $L^2((-a, a))$ we will have

$$(PHv)(x) = \lambda x$$

a.e. on $(-a, a)$. Let v_0 be the semicircle law which satisfies this last equality with λ_0 and assume (v_1, λ_1) also satisfies the equality. Then with $v_2 = \lambda_1 v_0 - \lambda_0 v_1$ we have

$$PHv_2 = 0$$

or equivalently

$$(Hv_2)v_2 = 0\,.$$

Since $v_2 \in L^3$ we infer $Hv_2 \in L^3$ and hence $v_2 + iHv_2 \in H^3$ (the Hardy space) and therefore $F = (v_2 + iHv_2)^2 \in H^{3/2}$ so that $0 = 2v_2Hv_2 = \operatorname{Im} F$ and $v_2^2 - (Hv_2)^2 = \operatorname{Re} F$. Being in $H^{3/2}$ we have $\operatorname{Im} F = 0 \Rightarrow \operatorname{Re} F = 0$ so that $v_2^2 - (Hv_2)^2 = 0$ which implies $v_2 = 0$ since $v_2(Hv_2) = 0$.   Q.E.D.

The following is an immediate corollary of the result of this section.

5.3. Corollary. *Let (A, φ) be a C^*-probability space and let $a_j, b_j \in A$ ($j = 1, 2$) be random variables such that $a_j = a_j^*$, $b_j = b_j^*$, $\varphi(a_j) = \varphi(b_j) = 0$, $\varphi(a_j^2) = \varphi(b_j^2)$ and $\varphi(a_1 a_2) = \varphi(b_1 b_2) = 0$. Assume moreover the a_j and b_j have semicircle distributions and $\{b_1, b_2\}$ is free. Then: $\Phi(a_1 + a_2) \geq \Phi(b_1 + b_2)$.*

6. Monotonicity of Σ and Riesz Energies

Using the subordination results in Sect. 4 we derive here monotonicity results for the free entropy (i.e. the logarithmic energy) and the energies associated with Riesz potentials.

For a probability measure ν on $\mathbb{R}$ we consider the energies associated with the Riesz potentials [12] for $0 < \alpha < 2$,

$$I_\alpha[\nu] = \int \int |x - y|^{\alpha - 2} d\nu(x) d\nu(y)\,,$$

as well as the logarithmic energy

$$I_2[\nu] = -\int \int \log|x - y| d\nu(x) d\nu(y)\,,$$

so that $I_2[\nu] = -\Sigma(\nu)$. If ν has density f we also write $I_\alpha[f]$. The next lemma (like Lemma 4.1 from which it is derived), is unlikely to be new.

6.1. Lemma. *Let μ be a probability measure on $\mathbb{R}$ and let $F(x + iy) = (P_y * \mu)(x)$ be the harmonic function on $\mathbb{H}$ defined by the Poisson integral. If $\omega : \mathbb{H} \to \mathbb{H}$ is analytic and $\lim\limits_{z \to \infty} |\omega(z)| = \infty$, $\operatorname{Im} \omega(z) \geq \operatorname{Im} z$, then we have*

$$I_\alpha([F(\bullet + i\varepsilon)]) \geq I_\alpha([F(\omega(\bullet + i\varepsilon))])$$

for $0 < \alpha < 2$, $\varepsilon > 0$.

Proof. Let

$$I_0([\nu], \delta) = \int \int d\nu(x) d\nu(y) \left((x - y)^2 + \delta^2\right)^{-1}\,,.$$

so that

$$I_0([\nu]; \delta) = c\delta \int (P_\delta * \nu)(x)\, d\nu(x)$$

for some constant c.

Assume $0 < \alpha < 1$ and consider the function $\operatorname{Im} z^{\alpha-1}$ expressed as a Poisson integral in $\mathbb{H} + i\beta$ for some $\beta > 0$ [we choose the branch of $z^{\alpha-1}$ which is real on $(0, \infty)$]. Letting $\beta \to 0$, we obtain

$$c\left(2\cos\frac{\pi}{2}(\alpha-1)\right)^{-1} x^{-1} \int\limits_0^\infty \frac{x\varepsilon^{\alpha-2}\varepsilon}{x^2+\varepsilon^2}\,d\varepsilon = \frac{1}{x}x^{\alpha-1} = x^{\alpha-2}\,.$$

Note that this equality may be continued analytically in α so as to establish its validity for $0 < \alpha < 2$.

Using this identity to express I_α by means of the $I_0([\ ]; \delta)$ the proof of the lemma reduces to showing that

$$I_0([F(\bullet + i\varepsilon)]; \delta) \geq I_0([F(\omega(\bullet + i\varepsilon))]; \delta)\,.$$

We have

$$I_0([F(\bullet + i\varepsilon)]; \delta) = \|P_{\delta/2} * F(\bullet + i\varepsilon)\|_2^2 = \|F(\bullet + i(\varepsilon + \delta/2))\|_2^2\,,$$

$$I_0([F(\omega(\bullet + i\varepsilon))]; \delta) = \|P_{\delta/2} * F(\omega(\bullet + i\varepsilon))\|_2^2 = \|F(\omega(\bullet + i(\varepsilon + \delta/2)))\|_2^2\,,$$

and the inequality is a consequence of Lemma 4.1.   Q.E.D.

6.2. Lemma. *Let* μ, F, ω *satisfy the assumptions of* 6.1 *and assume moreover*

$$\lim_{z\to\infty} \omega(z)/z = 1\,.$$

Then we have

$$I_2([F(\bullet + i\varepsilon)]) \geq I_2([F(\omega(\bullet + i\varepsilon))])\,.$$

Proof. This follows from 6.1, the equality

$$\lim_{\alpha\to 2}(2-\alpha)^{-1}(I_\alpha([\nu]) - 1) = I_2([\nu])$$

for probability measures on $\mathbb{R}$ with L^∞-density, together with the remark that the assumptions on ω imply that

$$\int F(\omega(x + i\varepsilon))dx = 1\,.   \text{Q.E.D.}$$

6.3. Proposition. *Let* μ_1, μ_2 *be compactly supported probability measures on* $\mathbb{R}$ *and let* $\mu_3 = \mu_1 \boxplus \mu_2$. *Let further* $\mu_j(t) = P_t * \mu_j$ *so that* $\mu_j(0) = \mu_j$. *Then we have*

$$I_\alpha(\mu_3(t)) \geq I_\alpha(\mu_j(t))$$

for $j = 1, 2$, $0 < t < \infty$, *and* $0 < \alpha \leq 2$ *and also for* $\alpha = 2$ *and* $t = 0$. *In particular*

$$\Sigma(\mu_3(t)) \geq \Sigma(\mu_j(t)) \quad \text{for} \quad 0 \leq t < \infty \quad \text{and} \quad j = 1, 2\,.$$

Proof. The assertion about Σ is just the case $\alpha = 2$ of the first assertion. Note further that it suffices to prove the first assertion for $t > 0$. Indeed if $t \downarrow 0$ we have

$$-\log|x + 2it| \uparrow -\log|x|$$

and hence

$$I_2(\mu_k(t)) \uparrow I_2(\mu_k)$$

for $k = 1, 2, 3$.

The proof then proceeds along the same lines as the proof of Proposition 4.7: first one deals with the case when the critical values of the Cauchy transforms of μ_1, μ_2 are disjoint and then approximates. We leave the details to the reader.   Q.E.D.

We are now able to strengthen the results of Sect. 3.

6.4. Proposition. *Let X, Y be a free pair of random variables such that the distribution of X extends to a compactly supported measure on $\mathbb{R}$ while Y is a $(0, 1)$-semicircular random variable. Then*
a) $[0, \infty) \ni s \to \Sigma(X + \sqrt{s}Y) \in [-\infty, \infty)$ *is a concave, continuous, increasing function.*
b) *For $0 \le s \le 1$ we have*

$$\Sigma(\sqrt{s}X + \sqrt{1 - s}Y) - \Sigma(Y) = \int\limits_0^s t^{-1}\left(\tfrac{1}{2} - \Phi(\sqrt{t}X + \sqrt{1 - t}Y)\right)dt .$$

Proof. a) In the next section we will prove $\Sigma(X_1 + tX_2)$ is upper semicontinuous as a function of t if X_1, X_2 are free (and have probability measures on $\mathbb{R}$ as distributions). This together with the fact that

$$\Sigma(X + \sqrt{s + \varepsilon}Y) = \Sigma(X + \sqrt{s}Y + \sqrt{\varepsilon}Y_1) \ge \Sigma(X + \sqrt{s}Y) \text{ by 6.3},$$

where Y_1 is $(0, 1)$-semicircular and free w.r.t. (X, Y) implies the function $s \to \Sigma(X + \sqrt{s}Y)$ is increasing and continuous to the right (and hence continuous at 0). On the other hand by 4.7,

$$\Phi(X + \sqrt{s + \varepsilon}Y) = \Phi(X + \sqrt{s}Y + \sqrt{\varepsilon}Y_1) \le \Phi(X + \sqrt{s}Y)$$

so that the concavity follows from 3.2. To get continuity for $s > 0$, we may use again 3.2 and the fact that by 4.7,

$$\Phi(X + \sqrt{s}Y) \le \Phi(\sqrt{s}Y) = s^{-1}\Phi(Y)$$

and $\phi(Y) < \infty$.
b) We have

$$\Sigma(\sqrt{s}X + \sqrt{1 - s}Y) = \Sigma(X + \sqrt{s^{-1} - 1}Y) + \tfrac{1}{2}\log s .$$

Combining 3.2 and 4.7 we have

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1}(\Sigma(X + \sqrt{s + \varepsilon}Y) - \Sigma(X + \sqrt{s}Y))$$

$$= \lim_{\varepsilon \downarrow 0} \Phi(X + \sqrt{s + \varepsilon}Y) = \Phi(X + \sqrt{s}Y) \le s^{-1}\Phi(X),$$

where we used the fact that $\Phi(X + \sqrt{s}Y)$ is right continuous, being decreasing and lower semicontinuous by 7.1. Thus $\Phi(X + \sqrt{s}Y)$ is the derivative to the right of $\Sigma(X + \sqrt{s}Y)$, so that $\Sigma(X + \sqrt{s^{-1} - 1}Y)$ is left differentiable for $0 < s < 1$ and we easily get the integral formula in b), after remarking that since $\Sigma(\sqrt{s}X + \sqrt{1 - s}Y) = \tfrac{1}{2}\log(1 - s) + \Sigma(\sqrt{s(1 - s)^{-1}}X + Y)$ we have $\lim_{s \to 0} \Sigma(\sqrt{s}X + \sqrt{1 - s}Y) = \Sigma(Y)$ because of 7.1 and 6.3.   Q.E.D.

6.5 Remark. In the context of the preceding proposition we have $\Sigma(X + \sqrt{s}Y) > -\infty$ if $s > 0$, since $\Phi(X + \sqrt{s}Y) \le s^{-1}\Phi(Y)$. Even more, this gives a uniform bound $|\Sigma(X + \sqrt{s}Y)| \le C$ for all X with distribution supported in an interval of diameter $\le M$ and $|s| + |s^{-1}| \le M$. Indeed if $C(L)$ is the norm of the operator $L^3([0, L]) \ni v \to (\log|\bullet| * v)|_{[0,L]} \in L^{3/2}([0, L])$, then $|\Sigma(\mu)| \le (\Phi(\mu))^{2/3}C(\operatorname{diam} \operatorname{supp} \mu)$.

7. Free Entropy in W^*-Probability Spaces

We now move away from the analytic function machinery used in the previous sections to the operator algebra context. This is the natural context for semicontinuity properties and for looking at quantities like $\Sigma(X_1, \ldots, X_n)$ which generalize $\Sigma(X)$.

Throughout we consider $(\mathcal{M}, \tau)$ a W^*-probability space, where $\mathcal{M}$ is a von Neumann algebra and τ is a faithful normal state. We will call $(\mathcal{M}, \tau)$ "tracial" if τ is a trace-state (and hence $\mathcal{M}$ finite).

Working with unbounded operators affiliated with $\mathcal{M}$, if $X = X^* \in \mathcal{M}$ the definition of $\Sigma(X)$ can be written:

$$\Sigma(X) = (\tau \otimes \tau)(\log |X \otimes I - I \otimes X|) = \tfrac{1}{2}(\tau \otimes \tau)(\log(X \otimes I - I \otimes X)^2),$$

and we may define more generally

$$\Sigma(X_1, \ldots, X_n) = \tfrac{1}{2}(\tau \otimes \tau)(\log((X_1 \otimes I - I \otimes X_1)^2 + \ldots + (X_n \otimes I - I \otimes X_n)^2))$$

where the right hand side is the trace of an element in $\mathcal{M} \otimes \mathcal{M}^{\mathrm{op}}$, $\mathcal{M}^{\mathrm{op}}$ denoting the opposite von Neuman algebra (if $n = 1$ it doesn't matter whether the formula for $\Sigma(X)$ is w.r.t. $\mathcal{M} \otimes \mathcal{M}$ or $\mathcal{M} \otimes \mathcal{M}^{\mathrm{op}}$).

Since $\Sigma(X)$ can be viewed as a measure of "how much resistance there is to almost commutation with X," the quantity $\Sigma(X_1, \ldots, X_n)$ similarly measures "how much resistance there is to almost commutation with $X_1, \ldots, X_n$."

It will be useful to use the notations $\tau_n = \tau \otimes \ldots \otimes \tau$ (n-times), $\Delta(X) = X \otimes I - I \otimes X$ and $\varepsilon_k(X) = I \otimes \ldots \otimes X \otimes I \otimes \ldots I$ (where X is in k^{th} position and the tensor product being n-fold is specified in the context).

7.1. Proposition. *In the W^*-probability space $(\mathcal{M}, \tau)$ we have:*
a) *If $\delta \downarrow 0$ then*

$$\tfrac{1}{2}\tau_2(\log(\Delta(X)^2 + \varepsilon^2)) \downarrow \Sigma(X).$$

b) *$\Sigma(X)$ is strongly upper semi-continuous on bounded sets, i.e. if $X_n = X_n^*$ and $X_n \xrightarrow{s} X$, then*

$$\limsup_{n \to \infty} \Sigma(X_n) \leq \Sigma(X).$$

c) *Let*

$$\varphi_\delta(x, y, z) = (\delta/\pi)^3 \int (\delta^2 + (t - x)^2)^{-1}(\delta^2 + (t - y)^2)^{-1}(\delta^2 + (t - z)^2)^{-1} dt.$$

If $\delta \downarrow 0$ then

$$\tau_3(\varphi_\delta(\varepsilon_1(X), \varepsilon_2(X), \varepsilon_3(X))) \uparrow \Phi(X).$$

d) *$\Phi(X)$ is strongly lower semicontinuous on bounded sets, i.e. if $X_n = X_n^*$ and $X_n \xrightarrow{s} X$ then*

$$\liminv_{n \to \infty} \Phi(X_n) \geq \Phi(X).$$

Proof. b) and d) are immediate consequences of a) and respectively c) and the fact that continuous functional calculus is strongly continuous on bounded sets. Also a) is immediate from functional calculus.

To prove d), let ν be the measure obtained by applying τ to the spectral measure of X. Then the quantity which has to be shown to increase to $\Phi(X)$ is just

$$\int ((P_\varepsilon * \nu)(t))^3 dt$$

which is a fact we have already used.   Q.E.D.

The inequalities which appear in the study of the usual entropy are also involved in deriving inequalities for $\Sigma(X_1, \ldots, X_n)$.

7.2. Proposition. *Let $(\mathscr{M}, \tau)$ be a tracial W^*-probability space and X_j, Y_k self-adjoint random-variables in $\mathscr{M}$. Then*

(a) $\Sigma(X_1, \ldots, X_n) \geq \Sigma((X_1^2 + X_2^2)^{1/2}, X_3, \ldots, X_n)$.

(b) $e^{\Sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_m)} \geq e^{\Sigma(X_1, \ldots, X_n)} + e^{\Sigma(Y_1, \ldots, Y_m)}$.

(c) $\Sigma(X_1, \ldots, X_n) \geq \log n + n^{-1}(\Sigma(X_1) + \ldots + \Sigma(X_n))$.

Proof. a) The Wigner-Yanase-Dyson-Lieb concavity [13, 22] implies the concavity of $A \to A^t \otimes A^{1-t}$ ($A \geq 0$, $t \in [0,1]$) as an $\mathscr{M} \otimes \mathscr{M}^{\mathrm{op}}$ valued map. The case $t = 1/2$ (already proved by Wigner and Yanase) gives in particular:

$$A^{1/2} \otimes A^{1/2} + B^{1/2} \otimes B^{1/2} \leq (A+B)^{1/2} \otimes (A+B)^{1/2}.$$

We have:
$$\Delta(X_1)^2 + \Delta(X_2)^2 - (X_1^2 + X_2^2) \otimes I - I \otimes (X_1^2 + X_2^2)$$
$$= -2(X_1 \otimes X_1 + X_2 \otimes X_2)$$
$$\geq -2(|X_1| \otimes |X_1| + |X_2| \otimes |X_2|)$$
$$\geq -2(X_1^2 + X_2^2)^{1/2} \otimes (X_1^2 + X_2^2)^{1/2},$$

so that
$$\Delta(X_1)^2 + \Delta(X_2)^2 \geq \Delta((X_1^2 + X_2^2)^{1/2})^2.$$

The inequality follows from

$$0 \leq A \leq B \Rightarrow \tau(\log A) \leq \tau(\log B),$$

which is well-known.

b) Clearly this inequality will follow if we show that for $A \geq 0$, $B \geq 0$ $e^{\tau(\log(A+B))} \geq e^{\tau(\log A)} + e^{\tau(\log B)}$. To show this inequality it suffices to prove it in the case $A \geq \varepsilon I$, $B \geq \varepsilon I$, $\varepsilon > 0$. This in turn will follow from the following inequality for $\lambda \geq 0$:

$$0 \leq \frac{d}{d\lambda}(e^{\tau(\log(A+\lambda B))} - e^{\tau(\log A)} - \lambda e^{\tau(\log B)})$$

which is
$$0 \leq \tau((A+\lambda B)^{-1}B)e^{\tau(\log(A+\lambda B))} - e^{\tau(\log B)}.$$

Putting $X = -\log(A + \lambda B) + \tau(\log(A + \lambda B))$ and $Y = \log B - \tau(\log B)$ the last inequality is equivalent to

$$\tau(e^X e^Y) \geq e^{r(X) + \tau(Y)} = 1,$$

which is an immediate consequence of the Golden-Thompson inequality and Jensen's inequality.

c) This is an immediate consequence of the concavity of $\tau(\log A)$.   Q.E.D.

7.3. Remark. In finite-dimensional algebras, $\Sigma(X) = -\infty$ for all X. However there is a related quantity, which may play the role of free entropy. If $X = X^*$ is a $n \times n$ matrix and $\lambda_1, \ldots, \lambda_n$ are its eigenvalues, then it is natural to define a free entropy by

$$\sum_{1 \leq p < q \leq n} \log(\lambda_p - \lambda_q)^2 \,.$$

8. Monotonicity and Orthogonal Subalgebras

The weakest form of independence of two subalgebras $\mathscr{A}_1, \mathscr{A}_2$ in a non-commutative probability space $(\mathscr{A}, \varphi)$, is the requirement $\varphi(a_1 a_2) = \varphi(a_1)\varphi(a_2)$ for $a_j \in A_j$. In the tracial W^*-context, these are the orthogonal subalgebras of [18].

The results of the preceding sections show that $\Sigma(X + Y) \geq \Sigma(X)$, $I(X + Y) \leq I(X)$ if $X = X^* \in \mathscr{A}_1$, $Y = Y^* \in \mathscr{A}_2$ and $\mathscr{A}_1, \mathscr{A}_2$ are free. In this section we look at what happens more generally if $\mathscr{A}_1, \mathscr{A}_2$ are only orthogonal. We remark that monotonicity holds when freeness is replaced by independence. We also provide some further "infinitesimal" evidence in favour of a general monotonicity result for Σ.

8.1. Example. If $(\mathscr{M}, \tau)$ is a $\mathscr{W}^*$-probability space, $X = X^* \in \mathscr{M}$, $Y = Y^* \in \mathscr{M}$, $[X, Y] = 0$ and the algebras generated by X and Y are orthognal, then $\Sigma(X + Y) \geq \Sigma(X)$ and $\Phi(X + Y) \leq \Phi(X)$.

The assertions are equivalent to $\Sigma(\mu_1 * \mu_2) \geq \Sigma(\mu_1)$ and $\Phi(\mu_1 * \mu_2) \leq \Phi(\mu_1)$ for compactly supported probability measures on $\mathbb{R}$. The inequality for Φ is immediate from

$$\|f * \mu\|_3 \leq \|f\|_3 \,.$$

In turn the inequality for Σ follows easily from the similar fact for the Riesz energies. The Riesz-kernels are positive definite for $1 < \alpha < 2$ (the Fourier transform of $|X|^\lambda(-1 < \lambda < 0)$ being $-2\sin\dfrac{\lambda\pi}{2} \, \Gamma(\lambda + 1) |\xi|^{-\lambda-1}$ [10]). Hence $I_\alpha[\mu_1 * \mu_2] \leq I_\alpha[\mu_2]$ is just that convolution by μ_1 is contractive in the I_α-norm, which is immediate expressing μ_1 as a weak limit of convex combinations of Dirac-measures and using the lower semicontinuity of I_α [12]. By differentiation at $\alpha = 2$ we get the inequality for Σ in case μ_2 has a smooth density. The general case is then obtained via replacing μ_2 by $\mu_2 * P_\varepsilon$ and approximating $\mu_2 * P_\varepsilon$ by compactly supported measures with smooth density.

8.2. Remark. Let $(\mathscr{M}, \tau)$ be a tracial W^*-probability space and let $1 \in \mathscr{A}$, $1 \in \mathscr{B}$ be orthogonal W^*-subalgebras. Let further $X = X^* \in \mathscr{A}$, $Y = Y^* \in B$ and

$$F_\delta(\varepsilon) = \tau_2(\log((\Delta(X + \varepsilon Y))^2 + \delta^2)) \,.$$

Then we have

$$F_\delta'(0) = 0 \,,$$

$$F_\delta''(0) \geq 2\tau_2(\Delta(Y)\Delta(X)^2(\Delta(X)^2 + \delta^2)^{-1}\Delta(Y)(\Delta(X)^2 + \delta^2)) \geq 0 \,.$$

Indeed we have:

$$F_\delta'(\varepsilon) = \tau_2((\Delta(X + \varepsilon Y)^2 + \delta^2)^{-1}(\Delta(X)\Delta(Y) + \Delta(Y)\Delta(X) + 2\varepsilon\Delta(Y)^2)) \,.$$

The orthogonality of $\mathscr{A}, B$ implies the orthogonality of $\mathscr{A} \otimes \mathscr{A}, B \otimes B$. Hence we have

$$F_\delta'(0) = 2\tau_2((\Delta(X)^2 + \delta^2)^{-1}\Delta(X)\Delta(Y)) = 0 \,.$$

Further:

$$
\begin{aligned}
F_\delta''(0) &= \tau_2((\Delta(X)^2 + \delta^2)^{-1}(\Delta(X)\Delta(Y) + \Delta(Y)\Delta(X)) \\
&\quad \times (\Delta(X)^2 + \delta^2)^{-1}(\Delta(Y)\Delta(X) + \Delta(X)\Delta(Y))) \\
&\quad + 2\tau_2((\Delta(X)^2 + \delta^2)^{-1}\Delta(Y)^2) \\
&= 2\tau_2((\Delta(X)^2 + \delta^2)^{-1}\Delta(X)\Delta(Y)\,((\Delta(X)^2 + \delta^2)^{-1}\Delta(X))\Delta(Y)) \\
&\quad + 2\tau_2(\Delta(Y)\Delta(X)^2(\Delta(X)^2 + \delta^2)^{-1}\Delta(Y)\,(\Delta(X)^2 + \delta^2)^{-1}) \\
&\quad + 2\tau_2(\Delta(X)^2 + \delta^2)^{-1})\tau_2(\Delta(Y)^2) \\
&\geq -2\|(\Delta(X)^2 + \delta^2)^{-1}\Delta(X)\Delta(Y)\|_2^2 \\
&\quad + R_\delta + 2\tau_2((\Delta(X)^2 + \delta^2)^{-1})\tau_2(\Delta(Y)^2)\,,
\end{aligned}
$$

where $R_\delta = 2\tau_2(\Delta(Y)\Delta(X)^2(\Delta(X)^2 + \delta^2)^{-1}\Delta(Y)\,(\Delta(X)^2 + \delta^2)^{-1})$. Since

$$
\begin{aligned}
\|(\Delta(X)^2 + \delta^2)^{-1}\Delta(X)\Delta(Y)\|_2^2 &= \tau_2(((\Delta(X)^2 + \delta^2)^{-1}\Delta(X))^2(\Delta(Y))^2) \\
&= \tau_2((\Delta(X)\,(\Delta(X)^2 + \delta^2)^{-1})^2)\tau_2((\Delta(Y))^2) \\
&\leq \tau_2((\Delta(X)^2 + \delta^2)^{-1})\tau_2(\Delta(Y)^2)\,,
\end{aligned}
$$

we have

$$F_\delta''(0) \geq R_\delta \,.$$

Also $R_\delta \geq 0$ being the trace of a product of positive operators.

One further remark: if the spectral measure of X has no atoms, then $\lim\limits_{\delta \to 0} R_\delta = \tau_2(\Delta(X)^{-2})\tau_2(\Delta(Y)^2)$.

References

1. Akhiezer, N.I.: The classical moment problem (in Russian) Moscow, 1961
2. Balian, R.: Random matrices and information theory. Nuovo Cimento, **LVII** B, No. 1, 183–193 (1968)
3. Barron, A.R.: Entropy and the central limit theorem. Ann. Prob. **14** (1), 336–342 (1986)
4. Bercovici, H., Voiculescu, D.: Levy-Hinčin type theorems for multiplicative and additive free convolution. Pacific J. Math. **153**, No. 2, 217–248 (1992)
5. Bercovici, H., Voiculescu, D.: Free convolution of measures with unbounded support. Preprint, Berkeley 1992
6. Carlen, E.A., Soffer, A.: Entropy production by block variable summation and central limit theorems. Commun. Math. Phys. **140**, 339–371 (1991)
7. Duren, P.L.: Univalent functions. Berlin, Heidelberg, New York: Springer Verlag, 1983
8. Dykema, K.J.: On certain free product factors via an extended matrix model. J. Funct. Anal. (to appear)
9. Garnett, J.B.: Bounded analytic functions. New York: Academic Press, 1981
10. Gelfand, I.M., Shilov, G.E.: Generalized functions. I (in Russian). Second edition. Moscow: Fizmatgiz, 1959
11. Kullback, S.: Information theory and statistics. New York: Dover Publications Inc., 1968
12. Landkof, N.S.: Foundations of modern potential theory. Berlin, Heidelberg, New York: Springer 1972
13. Lieb, E.: Convex Trace functions and the Wigner-Yanase-Dyson conjecture. Adv. Math. **11**, 267–288 (1973)

14. Maassen, H.: Addition of freely independent Random variables. J. Funct. Anal. **106**, 409–438 (1992)
15. Mehta, M.L.: Random matrices and the statistical theory of energy levels. Academic Press
16. Mushkhelishvili, N.I.: Singular integral equations. Groningen: Noordhoff 1953
17. Nica, A.: Asymptotically free families of Random unitaries in symmetric groups. Pacific J. Math. (to appear)
18. Popa, S.: Orthogonal pairs of ∗-subalgebras in finite von Neumann algebras. J. Operator Theory **9**, 253–268 (1983)
19. Radulescu, F.: The fundamental group of $\mathscr{L}(F_\infty)$ is $\mathbb{R}_+\backslash\{0\}$. J. Am. Math. Soc. (to appear)
20. Radulescu, F.: Stable equivalence of the weak closures of free groups convolution algebras. Preprint
21. Ruelle, D.: Statistical mechanics. New York: Benjamin 1969
22. Simon, B.: Trace ideals and their applications. Cambridge: Cambridge Univ. Press. 1979
23. Stam, A.J.: Some Inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control **2**, 101–112 (1959)
24. Tsuji, M.: Potential Theory in modern function theory. Tokyo: Maruzen 1959
25. Voiculescu, D.: Symmetries of some reduced free product C^*-algebras. In: Operator algebras and their connections with topology and ergodic theory. Lecture Notes in Math., vol. 1132. Berlin, Heidelberg, New York: Springer, pp. 556–588 (1985)
26. Voiculescu, D.: Addition of certain non-commuting random variables. J. Funct. Anal. **66**, No. 3, 323–346 (1986)
27. Voiculescu, D.: Multiplication of certain non-commuting random variables. J. Operator Theory **18**, 223–235 (1987)
28. Voiculescu, D.: Operations on certain non-commutative operator-valued random variables. INCREST Preprint No. 42/1986, Bucharest
29. Voiculescu, D.: Limit laws for random matrices and free products. Invent. Math. **104**, 201–220 (1991)
30. Voiculescu, D.: Circular and semi-circular systems and free product factors. In: Operator algebras, unitary representations, enveloping algebras and invariant theory. Progress in Math., vol. 92, Birkhäuser, Boston, pp. 45–60 (1990)
31. Voiculescu, D.: Free Non-commutative random variables, random matrices and the II_1-factors of free groups. Quantum probability and related topics. VI. Accardi, L. (ed.) Singapore: World Scientific, pp. 473–487 (1991)
32. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)
33. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**, 325–327 (1958)