From Boltzmann to random matrices and beyond

Djalil Chafaï

FROM BOLTZMANN TO RANDOM MATRICES AND BEYOND

DJALIL CHAFAÏ

Abstract. These expository notes propose to follow, across fields, some aspects of the concept of entropy. Starting from the work of Boltzmann in the kinetic theory of gases, various universes are visited, including Markov processes and their Helmholtz free energy, the Shannon monotonicity problem in the central limit theorem, the Voiculescu free probability theory and the free central limit theorem, random walks on regular trees, the circular law for the complex Ginibre ensemble of random matrices, and finally the asymptotic analysis of mean-field particle systems in arbitrary dimension, confined by an external field and experiencing singular pair repulsion. The text is written in an informal style driven by energy and entropy. It aims to be recreative and to provide to the curious readers entry points in the literature, and connections across boundaries.

Résumé. Ces notes d'exposition proposent de suivre, à travers différents domaines, quelques aspects du concept d'entropie. À partir du travail de Boltzmann en théorie cinétique des gas, plusieurs univers sont visités, incluant les processus de Markov et leur énergie libre de Helmholtz, le problème de Shannon de monotonie de l'entropie dans le théorème central limite, la théorie des probabilités libres de Voiculescu et le théorème central limite libre, les marches aléatoires sur les arbres réguliers, la loi du cercle pour l'ensemble de Ginibre complexe de matrices aléatoires, et enfin l'analyse asymptotique de systèmes de particules champ moyen en dimension arbitraire, confinées par un champ extérieur et subissant une répulsion singulière à deux corps. Le texte est écrit dans un style informel piloté par l'énergie et l'entropie. Il vise a être récréatif, à fournir aux lecteurs curieux des points d'entrée dans la littérature, et des connexions au delà des frontières.

Contents

Date: Spring 2014. Revised July 2, October 31, for Ann. Fac. Sci. Toulouse. Compiled January 13, 2015.

2010 *Mathematics Subject Classification.* 01-02; 05C80; 05C81; 15B52; 31A99; 31B99; 35Q20; 35Q83; 35Q84; 47D07; 53C44; 60J10; 60B20; 60F05; 60F10; 82C22; 46L54; 94A17.

Key words and phrases. Entropy; Fisher information; Boltzmann; Shannon; Voiculescu; Markov processes; Diffusion processes; Central limit theorem; Free probability; Random matrices; Ginibre ensemble; Circular law; Coulomb gas; Riesz kernel; Singular repulsion; Potential theory; Electrostatics; Equilibrium measure; Interacting Particle Systems; Mean-field interaction; Large Deviations Principle; Collective phenomena.

This text forms the written notes of a talk entitled "About confined particles with singular pair repulsion", given at the occasion of the workshop "Talking Across Fields" on convergence to the equilibrium of Markov chains. This workshop was organized in Toulouse from 24 to 28 March 2014 by Laurent Miclo, at the occasion of the CIMI Excellence research chair for Persi Diaconis.

Almost ten years ago, we wanted to understand by curiosity the typical global shape of the spectrum of Markov transition matrices chosen at random in the polytope of such matrices. It took us several years to make some progress [Cha1, BCC1, BCC2], in connection with the circular law phenomenon of Girko. The circular law states that the empirical measure of the eigenvalues of a random $n \times n$ matrix, with i.i.d. entries[1] of variance $1/n$, tends to the uniform law on the unit disc of the complex plane, as the dimension n tends to infinity. This universal result was proved rigorously by Tao and Vu [TV], after fifty years of contributions. The proof of this high dimensional phenomenon involves tools from potential theory, from additive combinatorics, and from asymptotic geometric analysis. The circular law phenomenon can be checked in the Gaussian case using the fact that the model is then exactly solvable. Actually, Ginibre has shown in the 1960's that if the entries are i.i.d. centered complex Gaussians then the eigenvalues of such matrices form a Coulomb gas at temperature $1/n$ in dimension 2. This in turn suggests to explore the analogue of the circular law phenomenon in dimension ≥ 3, beyond random matrices. This leads us to introduce in [CGZ] stochastic interacting particle systems in which each particle is confined by an external field, and each pair of particles is subject to a singular repulsion. Under general assumptions and suitable scaling, the empirical measure of the particles converges, as the number of particles tends to infinity, to a probability measure that minimizes a natural energy-entropy functional. In the case of quadratic confinement and Coulomb repulsion, the limiting law is uniform on a ball.

This expository text is divided into five sections, written in an informal style. The first section introduces the Boltzmann entropy and H-Theorem, and the analogous concept of Helmholtz free energy for Markov processes. The second section discusses the Shannon monotonicity problem of the Boltzmann entropy along the central limit theorem. The third section introduces Voiculescu free probability theory, the free entropy, and the free central limit theorem. The fourth section discusses the circular law for random matrices drawn from the complex Ginibre ensemble, using a large deviations rate function inspired from the Voiculescu entropy. The fifth and last section goes beyond random matrices and studies mean-field particle systems with singular interaction, for which the large deviations rate function is connected to Boltzmann and Voiculescu entropies.

Talking across fields is rarely an easy task. You may know for instance that Andreï Andreïevitch Markov (1856–1922) published his seminal article on what we call Markov chains in 1906[2], and that approximately at the same time, the theory of non-negative matrices was developed by Oskar Perron (1880–1975) and Ferdinand Georg Frobenius (1849–1917). It took several years to talk across fields, and according to Eugene Seneta [Sen], the link was made by von Mises (1883–1953). Various point of views are available on Markov chains. Beyond the concreteness of conditional construction and stochastic simulation, a Markov model can always be seen as a deterministic evolution, along the time, of a probability distribution. This mechanical view of (random) nature can be traced back to Charles Darwin (1809–1882) with his mutation-selection evolution theory,

[1] Real or complex, it does not matter, with an arbitrary mean, as soon as we look at narrow convergence.
[2] At the age of fifty, the year of the death of Ludwig Boltzmann.

and to Ludwig Boltzmann (1844–1906) with his H-Theorem in atomistic kinetic theory of gases. These two great figures of the nineteenth century can be viewed as proto-Markovian.

> "*If you ask me about my innermost conviction whether our century will be called the century of iron or the century of steam or electricity, I answer without hesitation: It will be called the century of the mechanical view of Nature, the century of Darwin.*"
>
> Ludwig Boltzmann, Vienna, 1886
> Der zweite Hauptsatz der mechanischen Wärmetheorie
> Lecture at the Festive Session, Imperial Academy of Sciences

Ludwig Boltzmann (1844–1906)

> "*Boltzmann summarized most (but not all) of his work in a two volume treatise Vorlesungen über Gastheorie. This is one of the greatest books in the history of exact sciences and the reader is strongly advised to consult it. It is tough going but the rewards are great*[3]*.*"
>
> Mark Kac, Ithaca, 1959, excerpt from [Ka]

1. Ludwig Boltzmann and his H-Theorem

1.1. **Entropy.** A simple way to introduce the Boltzmann entropy is to use the concept of combinatorial disorder. More precisely, let us consider a system of n distinguishable particles, each of them being in one of the r possible states (typically energy levels). We have $n = n_1 + \cdots + n_r$ where n_i is the number of particles in state i. The vector $(n_1, \ldots, n_r)$ encodes the macroscopic state of the system, while the microscopic state of the system is encoded by the vector $(b_1, \ldots, b_n) \in \{1, \ldots, r\}^n$ where b_i is the state of the i-th particle. The number of microscopic states compatible with a fixed macroscopic state $(n_1, \ldots, n_r)$ is given by the multinomial coefficient[4] $n!/(n_1! \cdots n_r!)$. This integer measures the microscopic degree of freedom given the macroscopic state. As a consequence, the additive degree of freedom per particle is then naturally given by $(1/n)\log(n!/(n_1! \cdots n_r!))$. But how does this behave when n is large? Let us suppose simply that n tends to ∞ while $n_i/n \to p_i$ for every $1 \leq i \leq r$. Then, thanks to the Stirling formula, we get, denoting $p := (p_1, \ldots, p_r)$,

$$\mathcal{S}(p) := \lim_{n\to\infty} \frac{1}{n} \log\left(\frac{n!}{n_1! \cdots n_r!}\right) = -\sum_{i=1}^{r} p_i \log(p_i).$$

The quantity $\mathcal{S}(p)$ is the Boltzmann entropy of the discrete probability distribution p. It appears here as an asymptotic additive degree of freedom per particle in a system with an infinite number of

[3]Personally, when I was a first year graduate student at Université Paul Sabatier in Toulouse, I borrowed this book from the university library by curiosity. I can confirm the "tough going" aspect mentioned by Mark Kac!

[4]Encoding the occurrence of each face of an r-faces dice thrown n times.

particles, each of them being in one of the r possible states, with population frequencies $p_1, \ldots, p_r$. This is nothing else but the first order asymptotic analysis of the multinomial combinatorics:

$$\frac{n!}{n_1! \cdots n_r!} \approx e^{n\mathcal{S}(n_1/n,\ldots,n_r/n)}.$$

When the disorder of the system is better described by a probability density function $f : \mathbb{R}^d \to \mathbb{R}_+$ instead of a discrete probability measure, we may introduce by analogy, or passage to the limit, the continuous Boltzmann entropy of f, denoted $\mathcal{S}(f)$, or $-H(f)$ in the terminology of Boltzmann,

$$\mathcal{S}(f) := -\int_{\mathbb{R}^d} f(x) \log(f(x))\, dx.$$

When X is a random variable, we denote by $\mathcal{S}(X)$ the entropy of its law. Here we use the notation $\mathcal{S}$, which is initially the one used by Clausius for the concept of entropy in thermodynamics.

1.2. **Maximum entropy under constraints.** The Boltzmann entropy $\mathcal{S}$ measures an average disorder. One can seek for a probability density f_* that maximizes the linear functional $f \mapsto \mathcal{S}(f)$ over a convex class $\mathcal{C}$ formed by a set of constraints on f:

$$\mathcal{S}(f_*) = \max\{\mathcal{S}(f) : f \in \mathcal{C}\}.$$

The class $\mathcal{C}$ is typically defined by linear (in f) statistics on the form $\displaystyle\int g(x) f(x)\, dx = c_g$ for $g \in \mathcal{G}$.

Following Boltzmann, suppose that the internal state of an isolated system is described by a parameter $x \in \mathbb{R}^d$ which is statistically distributed according to a probability density f, and suppose furthermore that the energy of state x is $V(x)$. Then the average energy of the system is

$$a = \int V(x)\, f(x)\, dx.$$

Let $\mathcal{C}$ be the class of probability densities f which satisfy this constraint, and let us seek for $f_* \in \mathcal{C}$ that maximizes the entropy $\mathcal{S}$ on $\mathcal{C}$, in other words such that $\mathcal{S}(f_*) = \max_{f \in \mathcal{C}} \mathcal{S}(f)$. A Lagrange variational analysis leads to $-\log f_* = \alpha + \beta V$ where α, β are Lagrange multipliers. We select $\alpha, \beta > 0$ in such a way that $f_* \in \mathcal{C}$, which gives a unique solution

$$f_* = \frac{1}{Z_\beta} e^{-\beta V} \quad \text{where} \quad Z_\beta := \int e^{-\beta V(x)}\, dx.$$

In Physics β is interpreted as an inverse temperature times a universal constant called the Boltzmann constant, selected in such a way that $f_* \in \mathcal{C}$. Indeed, by using the definition of f_*, the fact $f, f_* \in \mathcal{C}$, and the Jensen inequality for the convex function $u \geq 0 \mapsto u \log(u)$, we have

$$\mathcal{S}(f_*) - \mathcal{S}(f) = \int \frac{f}{f_*} \log\left(\frac{f}{f_*}\right) f_*\, dx \geq 0.$$

The quantity in the middle is known as the Kullback-Leibler divergence or relative entropy[5] with respect to $f_* dx$, see [KL, Ku]. The Jensen inequality and the strict convexity of $u \geq 0 \mapsto u \log(u)$ give that f_* is the unique density which achieves $\max_{\mathcal{C}} \mathcal{S}$. We write $f_* = \arg\max_{\mathcal{C}} \mathcal{S}$. The lower is the energy $V(x)$ of state x, the higher is the value $f_*(x)$ of the maximum entropy density f_*. Taking for instance $V(x) = +\infty \mathbf{1}_{K^c}(x)$ reveals that uniform laws maximize entropy under support constraint, while taking $V(x) = \|x\|_2^2$ reveals that Gaussian laws maximize entropy under second moment constraint. In particular, on $\mathbb{R}$, denoting G a Gaussian random variable,

$$\mathbb{E}(X^2) = \mathbb{E}(G^2) \Rightarrow \mathcal{S}(X) \leq \mathcal{S}(G) \quad \text{and} \quad \mathcal{S}(X) = \mathcal{S}(G) \Rightarrow X \stackrel{d}{=} G.$$

It is well known in Bayesian statistics that many other classical discrete or continuous laws are actually maximum entropy laws over classes of laws defined by natural constraints.

[5]This concept was actually introduced by Solomon Kullback (1907 – 1994) and Richard Leibler (1914 – 2003) in the 1950's as an information gain, and was inspired from the entropy in the Shannon theory of communication.

1.3. **Free energy and the law of large numbers.** Still with $f_* = Z_\beta^{-1} e^{-\beta V}$, we have

$$-\frac{1}{\beta}\log(Z_\beta) = \mathcal{A}(f_*) \quad \text{where} \quad \mathcal{A}(f) := \int V(x) f(x)\, dx - \frac{1}{\beta}\mathcal{S}(f).$$

The functional $\mathcal{A}$ is the Helmholtz[6] free energy[7]: mean energy minus temperature times entropy. The functional $\mathcal{A}$ is essentially $-\mathcal{S}$ penalized by the average energy. Also, the functional $\mathcal{A}$ admits f_* as a unique minimizer over the class of densities, without constraints. Indeed, the Helmholtz free energy is connected to the Kullback-Leibler relative entropy: for any density f,

$$\mathcal{A}(f) - \mathcal{A}(f_*) = \frac{1}{\beta}\int \frac{f}{f_*}\log\left(\frac{f}{f_*}\right) f_*\, dx \geq 0$$

with equality if and only if $f = f_*$ thanks to the strict convexity of $u \mapsto u\log(u)$. When f and f_* have same average energy, then we recover the formula for the Boltzmann entropy. As we will see later on, the Helmholtz free energy $\mathcal{A}$ plays a role for Markov processes. It emerges also from the strong law of large numbers. More precisely, let us equip the set $\mathcal{M}_1$ of probability measures on $\mathbb{R}^d$ with the narrow topology, which is the dual topology with respect to continuous and bounded functions. If $X_1, \ldots, X_N$ are i.i.d. random variables with law $\mu_* \in \mathcal{M}_1$, then their empirical distribution $\mu_N := \frac{1}{N}\sum_{k=1}^n \delta_{X_k}$ is a random variable on $\mathcal{M}_1$, and an asymptotic analysis due to Ivan Sanov (1919 – 1968) in the 1950's reveals that for every Borel set $A \subset \mathcal{M}_1$, as $N \gg 1$,

$$\mathbb{P}(\mu_N \in A) \approx \exp\left(-N \inf_A \mathcal{K}\right) \quad \text{where} \quad \mathcal{K}(\mu) := \int \frac{d\mu}{d\mu_*}\log\left(\frac{d\mu}{d\mu_*}\right) d\mu_*.$$

The rigorous version, known as the Sanov theorem, says more precisely (see [DZ] for a proof) that

$$-\inf_{\mathrm{int}(A)} \mathcal{K} \leq \liminf_{N\to\infty} \frac{\log \mathbb{P}(\mu_N \in A)}{N} \leq \limsup_{N\to\infty} \frac{\log \mathbb{P}(\mu_N \in A)}{N} \leq -\inf_{\mathrm{clo}(A)} \mathcal{K}$$

where $\mathrm{int}(A)$ and $\mathrm{clo}(A)$ are the interior and the closure of A. Using the terminology of Srinivasa Varadhan[8], $(\mu_N)_{N\geq 1}$ satisfies to a large deviations principle with speed N and rate function $\mathcal{K}$. The functional $\mathcal{K} : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ is the Kullback-Leibler relative entropy with respect to μ_*. By convention $\mathcal{K}(\mu) := +\infty$ if $\mu \not\ll \mu_*$. If $d\mu_*(x) = f_*(x)\, dx = Z_\beta^{-1} e^{-\beta V}\, dx$ and $d\mu = f\, d\mu_*$ then

$$\mathcal{K}(\mu) = \beta(\mathcal{A}(f) - \mathcal{A}(f_*)).$$

The Sanov theorem is a refinement of the strong law of large numbers, since by the first Borel-Cantelli lemma, one obtains that with probability one, $\lim_{N\to\infty} \mu_N = \mu_* = \arg\inf \mathcal{K}$.

The large deviations rate function $\mathcal{K}$ is convex and lower semicontinuous with respect to the narrow topology, which is the topology of convergence with respect to bounded and continuous test functions. This topology can be metrized by the metric $d(\mu,\nu) := \sup_{h\in\mathcal{H}} \int h\, d(\mu - \nu)$ where $\mathcal{H} := \{h : \max(\|h\|_\infty, \|h\|_{\mathrm{Lip}}) \leq 1\}$. Now for $A = A_\varepsilon = B(\mu,\varepsilon) := \{\nu : d(\mu,\nu) \leq \varepsilon\}$ we have

$$\mathbb{P}(\mu_N \in A) = \mu_*^{\otimes N}\left((x_1,\ldots,x_N) \in \mathbb{R}^N : \sup_{h\in\mathcal{H}} \left(\frac{1}{N}\sum_{i=1}^N h(x_i) - \int h\, d\mu\right) \leq \varepsilon\right).$$

and thanks to the Sanov theorem, we obtain the "volumetric" formula

$$\inf_{\varepsilon>0} \limsup_{N\to\infty} \frac{1}{N}\log \mu_*^{\otimes N}\left((x_1,\ldots,x_N) \in \mathbb{R}^N : \sup_{h\in\mathcal{H}} \left(\frac{1}{N}\sum_{i=1}^N h(x_i) - \int h\, d\mu\right) \leq \varepsilon\right) = -K(\mu).$$

1.4. **Names.** The letter $\mathcal{S}$ was chosen by Rudolf Clausius (1822 – 1888) for entropy in thermodynamics, possibly in honor of Sadi Carnot (1796 – 1832). The term *entropy* was forged by Clausius in 1865 from the Greek «$\eta\ \tau\rho o\pi\eta$». The letter H used by Boltzmann is the capital Greek letter η. The letter $\mathcal{A}$ used for the Helmholtz free energy comes from the German word "Arbeit" for work.

[6]Hermann Ludwig Ferdinand von Helmholtz (1821 – 1894), inventor, among other things, of the unified concept of energy and its conservation in physics, in competition with Julius Robert von Mayer (1814 – 1878).

[7]Should not be confused with Gibbs free energy (free enthalpy) even if they are closely related for ideal gases.

[8]Srinivasa Varadhan (1940 –) is the (main) father of modern large deviations theory.

> "I propose to name the quantity $\mathcal{S}$ the entropy of the system, after the Greek word $\eta\ \tau\rho o\pi\eta$ (en tropein), the transformation. I have deliberately chosen the word entropy to be as similar as possible to the word energy: the two quantities to be named by these words are so closely related in physical significance that a certain similarity in their names appears to be appropriate."
>
> Rudolf Clausius, 1865

1.5. **H-Theorem.** Back to the motivations of Boltzmann, let us recall that the first principle of Carnot-Clausius thermodynamics[9] states that the internal energy of an isolated system is constant, while the second principle states that there exists an extensive state variable called the entropy that can never decrease for an isolated system. Boltzmann wanted to derive the second principle from the idea (controversial, at that time) that the matter is made with atoms. Let us consider an ideal isolated gas made with particles (molecules) in a box with periodic boundary conditions (torus) to keep things as simple as possible. There are of course too many particles to write the equations of Newton for all of them. Newton is in a way beated by Avogadro! The idea of Boltzmann was to propose a statistical approach (perhaps inspired from the one of Euler in fluid mechanics, and from the work of Maxwell, see [Cro]): instead of keeping track of each particle, let $(x,v)\mapsto f_t(x,v)$ be the probability density of the distribution of position $x\in\mathbb{R}^d$ and velocity $v\in\mathbb{R}^d$ of particles at time t. Then one can write an evolution equation for $t\mapsto f_t$, that takes into account the physics of elastic collisions. It is a nonlinear partial differential equation known as the Boltzmann equation:

$$\partial_t f_t(x,v) = -v\partial_x f_t(x,v) + Q(f_t,f_t)(x,v).$$

The first term in the right hand side is a linear transport term, while the second term $Q(f_t,f_t)$ is quadratic in f_t, a double integral actually, and captures the physics of elastic collisions by averaging over all possible input and output velocities (note here a loss of microscopic information). This equation admits conservation laws. Namely, for every time $t\geq 0$, f_t is a probability density and the energy of the system is constant (first principle):

$$f_t\geq 0,\quad \partial_t\int f_t(x,v)\,dxdv = 0,\quad \partial_t\iint v^2\,f_t(x,v)\,dxdv = 0.$$

These constrains define a class of densities $\mathcal{C}$ on $\mathbb{R}^d\times\mathbb{R}^d$ over which the Boltzmann entropy $\mathcal{S}$ achieves its (Gaussian in velocity and uniform in position) maximum

$$f_* = \arg\max_{\mathcal{C}}\mathcal{S}.$$

The H-Theorem states that the entropy $\mathcal{S}=-H$ is monotonic along the Boltzmann equation:

$$\partial_t\mathcal{S}(f_t)\geq 0$$

and more precisely,

$$\mathcal{S}(f_t)\underset{t\to\infty}{\nearrow}\mathcal{S}(f_*)=\max_{\mathcal{C}}\mathcal{S}$$

where $\mathcal{C}$ is the class defined by the conservation law. A refined analysis gives that

$$f_t\underset{t\to\infty}{\longrightarrow} f_* = \arg\max_{\mathcal{C}}\mathcal{S}.$$

In the space-homogeneous simplified case, f_t depends only on the velocity variable, giving a Gaussian equilibrium for velocities by maximum entropy! In kinetic theory of gases, it is customary to call "Maxwellian law" the standard Gaussian law on velocities. We refer to [Vil2] for a discussion on the concept of irreversibility and the Boltzmann H-Theorem.

The work of Boltzmann in statistical physics (nineteenth century) echoes the works of Euler in fluid mechanics (eighteenth century), and of Newton in dynamics (seventeenth century). Before the modern formalization of probability theory and of partial differential equations with functional analysis, Boltzmann, just like Euler, was able to forge a deep concept melting the two! The Boltzmann H-Theorem had and has still a deep influence, with for instance the works of Kac, Lanford, Cercignani, Sinai, Di Perna and Lions, Desvillettes and Villani, Saint-Raymond, ...

[9] According to Vladimir Igorevitch Arnold (1937 – 2010), "Every mathematician knows it is impossible to understand an elementary course in thermodynamics.". Nevertheless, the reader may try [Fer, Z], and [Cro] for history.

> "Although Boltzmann's H-Theorem is 135 years old, present-day mathematics is unable to prove it rigorously and in satisfactory generality. The obstacle is the same as for many of the famous basic equations of mathematical physics: we don't know whether solutions of the Boltzmann equations are smooth enough, except in certain particular cases (close- to-equilibrium theory, spatially homogeneous theory, close-to-vacuum theory). For the moment we have to live with this shortcoming."
>
> Cédric Villani, 2008, excerpt from [Vil1]
> H-Theorem and beyond: Boltzmann's entropy in today's mathematics

1.6. **Keeping in mind the structure.** For our purposes, let us keep in mind this idea of evolution equation, conservation law, monotonic functional, and equilibrium as optimum (of the monotonic functional) under constraint (provided by the conservation law). It will reappear!

1.7. **Markov processes and Helmholtz free energy.** A Markov process can always be seen as a deterministic evolution equation of a probability law. By analogy with the Boltzmann equation, let us consider a Markov process $(X_t, t \in \mathbb{R}_+)$ on $\mathbb{R}^d$. Let us focus on structure and relax the rigor to keep things simple. For any $t \geq 0$ and continuous and bounded test function h, for any $x \in \mathbb{R}^d$,

$$P_t(h)(x) := \mathbb{E}(h(X_t)|X_0 = x).$$

Then $P_0(h) = h$, and, thanks to the Markov property, the one parameter family $P = (P_t, t \in \mathbb{R}_+)$ forms a semigroup of operators acting on bounded continuous test functions, with $P_0 = id$. Let us assume that the process admits an invariant measure μ_*, meaning that for every $t \geq 0$ and h,

$$\int P_t(h)\, d\mu_* = \int h\, d\mu_*.$$

The semigroup is contractive in $L^p(\mu_*)$: $\|P_t\|_{p\to p} \leq 1$ for any $p \in [1, \infty]$ and $t \geq 0$. The semigroup is Markov: $P_t(h) \geq 0$ if $h \geq 0$, and $P_t(1) = 1$. The infinitesimal generator of the semigroup is $Lh = \partial_{t=0} P_t(h)$, for any h in the domain of L (we ignore these aspects for simplicity). In particular

$$\int Lh\, d\mu_* = \partial_{t=0} \int P_t(h)\, d\mu_* = \partial_{t=0} \int h\, d\mu_* = 0.$$

When L is a second order linear differential operator without constant term, then we say that the process is a Markov diffusion. Important examples of such Markov diffusions are given in Table 1. The backward and forward Chapman-Kolmogorov equations are

$$\partial_t P_t = LP_t = P_t L.$$

Let us denote by P_t^* the adjoint of P_t in $L^2(\mu_*)$, and let us define

$$\mu_t := \mathrm{Law}(X_t) \quad \text{and} \quad g_t := \frac{d\mu_t}{d\mu_*}.$$

Then $g_t = P_t^*(g_0)$. If we denote $L^* = \partial_{t=0} P_t^*$ then we obtain the evolution equation

$$\partial_t g_t = L^* g_t.$$

The invariance of μ_* can be seen as a fixed point: if $g_0 = 1$ then $g_t = P_t^*(1) = 1$ for all $t \geq 0$, and $L^*1 = 0$. One may prefer to express the evolution of the density

$$f_t := \frac{d\mu_t}{dx} = g_t f_* \quad \text{where} \quad f_* := \frac{d\mu_*}{dx}.$$

We have then $\partial_t f_t = G^* f_t$ where $G^* h := L^*(h/f_*) f_*$. The linear evolution equation

$$\partial_t f_t = G^* f_t$$

is in a sense the Markovian analogue of the Boltzmann equation (which is nonlinear!).

Let us focus on the case where f_* is a density, meaning that μ_* is a probability measure. This excludes Brownian motion, but allows the Ornstein-Uhlenbeck process. By analogy with the Boltzmann equation, we have the first two conservation laws $f_t \geq 0$ and $\int f_t\, dx = 1$, but the

[10] The $\sqrt{2}$ factor in the S.D.E. allows to avoid a factor $1/2$ in the infinitesimal generators.

	Brownian motion	Ornstein-Uhlenbeck process	Overdamped Langevin process
S.D.E.	$dX_t = \sqrt{2}dB_t$	$dX_t = \sqrt{2}dB_t - X_t dt$	$dX_t = \sqrt{2}dB_t - \nabla V(X_t)\, dt$
μ_*	dx	$\mathcal{N}(0, I_d)$	$Z^{-1}e^{-V(x)}\, dx$
f_*	1	$(2\pi)^{-d/2}e^{-\|x\|_2^2/2}$	$Z^{-1}e^{-V}$
Lf	Δf	$\Delta f - x \cdot \nabla f$	$\Delta - \nabla V \cdot \nabla f$
$G^* f$	Δf	$\Delta f + \mathrm{div}(xf)$	$\Delta + \mathrm{div}(f \nabla V)$
μ_t	$\mathrm{Law}(X_0 + \sqrt{2t}G)$	$\mathrm{Law}(e^{-t}X_0 + \sqrt{1-e^{-2t}}G)$	Not explicit in general
P_t	Heat semigroup	O.-U. semigroup	General semigroup

TABLE 1. Two fundamental Gaussian processes on $\mathbb{R}^d$, Brownian Motion and Ornstein-Uhlenbeck, as Gaussian special cases of Markov diffusion processes[10].

average energy has no reason to be conserved for a Markov process. Indeed, the following quantity has no reason to have a constant sign (one can check this on the Ornstein-Uhlenbeck process!):

$$\partial_t \int V f_t \, dx = \partial_t P_t(V) = P_t(LV).$$

Nevertheless, if we set $f_* = Z_\beta^{-1} e^{-\beta V}$ then there exists a functional which is monotonic and which admits the invariant law μ_* as a unique optimizer: the Helmholtz free energy defined by

$$\mathcal{A}(f) := \int V(x)\, f(x)\, dx - \frac{1}{\beta}\mathcal{S}(f).$$

In order to compute $\partial_t \mathcal{A}(f_t)$, we first observe that for any test function g,

$$\int L^* g \, d\mu_* = 0.$$

Since μ_* is invariant, we have $P_t^*(1) = 1$ for every $t \geq 0$, and, since $P_t^*(g) \geq 0$ if $g \geq 0$, it follows that the linear form $g \mapsto P_t^*(g)(x)$ is a probability measure[11]. Recall that

$$\mathcal{A}(f) - \mathcal{A}(f_*) = \frac{1}{\beta}\int \Phi(g)\, d\mu_*$$

where $\Phi(u) := u \log(u)$. For every $0 \leq s \leq t$, the Jensen inequality and the invariance of μ_* give

$$\mathcal{A}(f_t) - \mathcal{A}(f_*) = \int \Phi(P_{t-s}^*(f_s))\, d\mu_* \leq \int P_{t-s}^*(\Phi(f_s))\, d\mu_* = \int \Phi(f_s)\, d\mu_* = \mathcal{A}(f_s) - \mathcal{A}(f_*),$$

which shows that the function $t \mapsto \mathcal{A}(f_t)$ is monotonic. Alternatively, the Jensen inequality gives also $\Phi(P_t^*(g)) \leq P_t^*(\Phi(g))$ and the derivative at $t = 0$ gives $\Phi'(g)L^*g \leq L^*\Phi(g)$, which provides

$$\int \Phi'(g) L^* g \, d\mu_* \leq 0.$$

Used with $g = g_t = f_t/f_*$, this gives

$$\beta \partial_t \mathcal{A}(f_t) = \partial_t \int \Phi(g_t)\, d\mu_* = \int \Phi'(g_t) L^* g_t \, d\mu_* \leq 0.$$

It follows that the Helmholtz free energy decreases along the Markov process:

$$\partial_t \mathcal{A}(f_t) \leq 0.$$

Of course we expect, possibly under more assumptions, that $\mathcal{A}(f_t) \searrow \mathcal{A}(f_*) = \min \mathcal{A}$ as $t \to \infty$. Let us assume for simplicity that $\beta = 1$ and that the process is the Markov diffusion generated by $L = \Delta - \nabla V \cdot \nabla$. In this case μ_* is symmetric for the Markov process, and $L = L^*$, which makes most aspects simpler. By the chain rule $L(\Phi(g)) = \Phi'(g)Lg + \Phi''(g)|\nabla g|^2$, and thus, by invariance,

$$\partial_t \mathcal{A}(f_t) = \int \Phi'(g_t) L g_t \, d\mu_* = -\mathcal{F}(g_t) \leq 0 \quad \text{where} \quad \mathcal{F}(g) := \int \Phi''(g)|\nabla g|^2 \, d\mu_* = \int \frac{|\nabla g|^2}{g}\, d\mu_*.$$

The functional $\mathcal{F}$ is known as the Fisher information[12]. The identity $\partial_t \mathcal{A}(f_t) = -\mathcal{F}(g_t)$ is known as the de Bruijn identity. In the degenerate case $V \equiv 0$, then $f_* \equiv 1$ is no longer a density, μ_* is

[11] Actually $P_t^*(g)(x)$ is the value at point x of the density with respect to μ_* of the law of X_t when $X_0 \sim g d\mu_*$.

[12] Named after Ronald Aylmer Fisher (1890 – 1962), father of modern statistics among other things.

the Lebesgue measure, the Markov process is Brownian motion, the infinitesimal generator in the Laplacian $L = \Delta$, the semigroup $(P_t, t \in \mathbb{R}_+)$ is the heat semigroup, and we still have a de Bruijn identity $\partial_t \mathcal{S}(f_t) = \mathcal{F}(f_t)$ where $\mathcal{S} = -\mathcal{A}$ since $V \equiv 0$.

The quantitative version of the monotonicity of the free energy along the Markov semigroup is related to Sobolev type functional inequalities. We refer to [ABC+, BGL] for more details. For instance, for every constant $\rho > 0$, the following three properties are equivalent:

- Exponential decay of free energy: $\forall f_0 \geq 0$, $\forall t \geq 0$, $\mathcal{A}(f_t) - \mathcal{A}(f_*) \leq e^{-2\rho t}(\mathcal{A}(f_0) - \mathcal{A}(f_*))$;
- Logarithmic Sobolev inequality: $\forall f \geq 0$, $2\rho(\mathcal{A}(f) - \mathcal{A}(f_*)) \leq \mathcal{F}(f/f_*)$;
- Hypercontractivity: $\forall t \geq 0$, $\|P_t\|_{q(0)\to q(t)} \leq 1$ where $q(t) := 1 + e^{2\rho t}$.

The equivalence between the first two properties follows by taking the derivative over t and by using the Grönwall lemma. The term "Logarithmic Sobolev inequality" is due to Leonard Gross, who showed in [Gro] the equivalence with hypercontractivity, via the basic fact that for any $g \geq 0$,

$$\partial_{p=1} \|g\|_p^p = \partial_{p=1} \int e^{p \log(g)} \, d\mu_* = \int g \log(g) \, d\mu_* = \mathcal{A}(g f_*) - \mathcal{A}(f_*).$$

The concept of hypercontractivity of semigroups goes back at least to Edward Nelson [N].

One may ask if $t \mapsto \mathcal{F}(f_t)$ is in turn monotonic. The answer involves a notion of curvature. Namely, using the diffusion property via the chain rule and reversibility, we get, after some algebra,

$$\partial_t^2 \mathcal{A}(f_t) = -\partial_t \mathcal{F}(g_t) = 2 \int g_t \Gamma_2(\log(g_t)) \, d\mu_*$$

where Γ_2 is the Bakry-Émery "gamma-two" functional quadratic form given by[13]

$$\Gamma_2(f) = \left\| \nabla^2 f \right\|_{\mathrm{HS}}^2 + \nabla f \cdot (\nabla^2 V) \nabla f.$$

See [ABC+, BGL] for the details. This comes from the Bochner commutation formula

$$\nabla L = L \nabla - (\nabla^2 V) \nabla.$$

If $\Gamma_2 \geq 0$ then, along the semigroup, the Fisher information is non-increasing, and the Helmholtz free energy is convex (we already know that $\partial_t \mathcal{A}(f_t) \leq 0$ and $\min \mathcal{A} = \mathcal{A}(f_*)$):

$$\partial_t \mathcal{F}(g_t) \leq 0 \quad \text{and} \quad \partial_t^2 \mathcal{A}(f_t) \geq 0.$$

This holds for instance if $\nabla^2 V \geq 0$ as quadratic forms[14]. This is the case for the Ornstein-Uhlenbeck example, for which $\nabla^2 V = I_d$. Moreover, if there exists a constant $\rho > 0$ such that for all f,

$$\Gamma_2(f) \geq \rho \Gamma(f) \quad \text{where} \quad \Gamma(f) = |\nabla f|^2$$

then, for any $t \geq 0$, $\partial_t \mathcal{F}(g_t) \leq -2\rho \mathcal{F}(g_t)$, and the Grönwall lemma gives the exponential decay of the Fisher information along the semigroup ($\rho = 1$ in the Ornstein-Uhlenbeck example):

$$\mathcal{F}(g_t) \leq e^{-2\rho t} \mathcal{F}(g_0).$$

This gives also the exponential decay of the Helmholtz free energy $\mathcal{A}$ along the semigroup with rate 2ρ, in other words, a Logarithmic Sobolev inequality with constant 2ρ: for any f_0,

$$\mathcal{A}(f_0) - \mathcal{A}(f_*) = -\int_0^\infty \partial_t \mathcal{A}(f_t) \, dt = \int_0^\infty \mathcal{F}(g_t) \, dt \leq \mathcal{F}(g_0) \int_0^\infty e^{-2\rho t} \, dt = \frac{\mathcal{F}(f_0/f_*)}{2\rho}.$$

We used here the Markov semigroup in order to interpolate between f_0 and f_*. This interpolation technique was extensively developed by Bakry and Ledoux in order to obtain functional inequalities, see [BGL] and references therein. A modest personal contribution to this topic is [Cha3]. The best possible constant ρ – which is the largest – in the inequality $\Gamma_2 \geq \rho \Gamma$ is called the Bakry-Émery curvature of the Markov semigroup. The story remains essentially the same for Markov processes on Riemannian manifolds. More precisely, in the context of Riemannian geometry, the Ricci curvature tensor contributes additively to the Γ_2, see [BGL]. Relatively recent works on this topic include extensions by Lott and Villani [LV], von Renesse and Sturm [vRS], Ollivier [O], among others. The approach can be adapted to non-elliptic hypoelliptic evolution equations, see for instance Baudoin [B]. The Boltzmannian idea of monotonicity along an evolution equation is also used in the work of Grigori Perelman on the Poincaré-Thurston conjecture, and in this case, the evolution

[13] The terminology comes from $\Gamma_2(f) := \Gamma_2(f,f) := \frac{1}{2}(L(\Gamma(f,f)) - 2\Gamma(f, Lf))$ where Γ is the "carré du champ" functional quadratic form $(f,g) \mapsto \Gamma(f,g)$ defined by $\Gamma(f) = \Gamma(f,f) := \frac{1}{2}(L(f^2) - 2fLf)$. Here $\Gamma(f) = |\nabla f|^2$.

[14] This means $y \cdot (\nabla^2 V(x)) y \geq 0$ for all $x, y \in \mathbb{R}^d$, or equivalently V is convex, or equivalently μ_* is log-concave.

equation, known as the Ricci flow of Hamilton, concerns the Ricci tensor itself, see [BBB, Li]. The exponential decay of the Boltzmann functional $H = -\mathcal{S}$ along the (nonlinear!) Boltzmann equation was conjectured by Cercignani and studied rigorously by Villani, see for instance [DMV].

Many aspects remain valid for discrete time/space Markov chains, up to the lack of chain rule if the space is discrete. For instance, if an irreducible Markov chain in discrete time and finite state space E has transition matrix P and invariant law μ_*, then $(P^n)_{n\geq 0}$ is the discrete time Markov semigroup, $L := P - I$ is the Markov generator, and one can show that for every initial law μ, the discrete Helmholtz free energy is monotonic along the evolution equation, namely

$$\sum_{x\in E} \Phi\left(\frac{\mu P^n(x)}{\mu_*(x)}\right)\mu_*(x) \underset{n\to\infty}{\searrow} 0,$$

where $\mu P^n(x) = \sum_{z\in E} \mu(z) P^n(z,x)$ and still $\Phi(u) := u\log(u)$. The details are in Thomas Liggett's book [Lig, prop. 4.2]. According to Liggett [Lig, p. 120] this observation goes back at least to Mark Kac [Ka, p. 98]. Quoting Persi Diaconis, "*the idea is that the maximum entropy Markov transition matrix with a given invariant law is clearly the matrix with all rows equal to this stationary law, taking a step in the chain increases entropy and keeps the stationary law the same.*".

Discrete versions of the logarithmic Sobolev inequality allow to refine the quantitative analysis of the convergence to the equilibrium of finite state space Markov chains. We refer for these aspects to the work of Diaconis and Saloff-Coste [DSC, SC], the work of Laurent Miclo [M], and the book [MT]. The relative entropy allows to control the total variation distance: the so-called Pinsker or Csiszár-Kullback inequality states that for any probability measures μ and ν on E with $\mu > 0$,

$$\sum_{x\in E} |\mu(x) - \nu(x)| \leq \sqrt{2\sum_{x\in E} \Phi\left(\frac{\nu(x)}{\mu(x)}\right)\mu(x)}.$$

The analysis of the convexity of the free energy along the semigroup of at most countable state space Markov chains was considered in [CDPP] and references therein. More precisely, let $(X_t)_{t\in\mathbb{R}_+}$ be a continuous time Markov chain with at most countable state space E. Let us assume that it is irreducible, positive recurrent, and aperiodic, with unique invariant probability measure μ_*, and with infinitesimal generator $L : E\times E \to \mathbb{R}$. We have, for every $x, y \in E$,

$$L(x,y) = \partial_{t=0}\mathbb{P}(X_t = y | X_0 = x).$$

We see L as matrix with non-negative off-diagonal elements and zero-sum rows: $L(x,y) \geq 0$ and $L(x,x) = -\sum_{y\neq x} L(x,y)$ for every $x, y \in E$. The invariance reads $0 = \sum_{x\in E} \mu_*(x)L(x,y)$ for every $y \in E$. The operator L acts on functions as $(Lf)(x) = \sum_{y\in E} L(x,y)f(y)$ for every $x \in E$. Since $\mu_*(x) > 0$ for every $x \in E$, the free energy at unit temperature corresponds to the energy $V(x) = -\log(\mu_*(x))$, for which we have of course $\mathcal{A}(\mu_*) = 0$. For any probability measure μ on E,

$$\mathcal{A}(\mu) - \mathcal{A}(\mu_*) = \mathcal{A}(\mu) = \sum_{x\in E} \Phi\left(\frac{\mu(x)}{\mu_*(x)}\right)\mu_*(x).$$

One can see $x \mapsto \mu(x)$ as a density with respect to the counting measure on E. For any time $t \in \mathbb{R}_+$, if $\mu_t(x) := \mathbb{P}(X_t = x)$ then $g_t(x) := \mu_t(x)/\mu_*(x)$ and $\partial_t g_t = L^* g_t$ where L^* is the adjoint of L in $\ell^2(\mu_*)$ which is given by $L^*(x,y) = L(y,x)\mu_*(y)/\mu_*(x)$. Some algebra reveals that

$$\partial_t \mathcal{A}(\mu_t) = \sum_{x\in E} [\Phi'(g_t)L^* g_t](x)\mu_*(x).$$

The right hand side is up to a sign the discrete analogue of the Fisher information. By reusing the convexity argument used before for diffusions, we get that $\partial_t \mathcal{A}(\mu_t) \leq 0$. Moreover, we get also

$$\partial_t^2 \mathcal{A}(\mu_t) = \sum_{x\in E} \left[g_t L L\log(g_t) + \frac{(L^* g_t)^2}{g_t}\right](x)\mu_*(x).$$

The right hand side is a discrete analogue of the Γ_2-based formula obtained for diffusions. It can be nicely rewritten when μ_* is reversible. The lack of chain rule in discrete spaces explains the presence of two distinct terms in the right hand side. We refer to [CDPP, JA] for discussions of examples including birth-death processes. Our modest contribution to this subject can be found in [Cha2, CJ]. We refer to [O] for some complementary geometric aspects. An analogue

on $E = \mathbb{N}$ of the Ornstein-Uhlenbeck process is given by the so-called M/M/∞ queue for which $(Lf)(x) = \lambda(f(x+1) - f(x)) + x\mu((f(x-1) - f(x))$ and $\mu_* = \text{Poisson}(\lambda/\mu)$.

2. Claude Shannon and the central limit theorem

The Boltzmann entropy plays also a fundamental role in communication theory, funded in the 1940's by Claude Elwood Shannon (1916–2001), where it is known as "Shannon entropy". It has a deep interpretation in terms of uncertainty and information in relation with coding theory [CT]. For example the discrete Boltzmann entropy $\mathcal{S}(p)$ computed with a logarithm in base 2 is the average number of bits per symbol needed to encode a random text with frequencies of symbols given by the law p. This plays an essential role in lossless coding, and the Huffman algorithm for constructing Shannon entropic codes is probably one of the most used basic algorithm (data compression is everywhere). Another example concerns the continuous Boltzmann entropy which enters the computation of the capacity of continuous telecommunication channels (e.g. DSL lines).

> "My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage."
>
> Claude E. Shannon, 1961
> Conversation with Myron Tribus, reported in [TM]

For our purposes, let us focus on the link between the Boltzmann entropy and the central limit theorem, a link suggested by Shannon when he forged information theory in the 1940's.

2.1. **The CLT as an evolution equation.** The Central Limit theorem (CLT) states that if $X_1, X_2, \dots$ are i.i.d. real random variables with mean $\mathbb{E}(X_i) = 0$ and variance $\mathbb{E}(X_i^2) = 1$, then

$$S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$$

where the convergence to the Gaussian law holds in distribution (weak sense).

2.2. **Conservation law.** The first two moments are conserved along CLT: for all $n \geq 1$,

$$\mathbb{E}(S_n) = 0 \quad \text{and} \quad \mathbb{E}(S_n^2) = 1.$$

By analogy with the H-Theorem, the CLT concerns an evolution equation of the law of S_n along the discrete time n. When the sequence $X_1, X_2, \dots$ is say bounded in L^∞ then the convergence in the CLT holds in the sense of moments, and in particular, the first two moments are constant while the remaining moments of order > 2 become universal at the limit. In other words, the first two moments is the sole information retained by the CLT from the initial data, via the conservation law. The limiting distribution in the CLT is the Gaussian law, which is the maximum of Boltzmann entropy under second moment constraint. If we denote by f^{*n} the n-th convolution power of the density f of the X_i's then the CLT writes $\text{dil}_{n^{-1/2}}(f^{*n}) \to f_*$ where f_* is the standard Gaussian and where $\text{dil}_\alpha(h) := \alpha^{-1} h(\alpha^{-1}\cdot)$ is the density of the random variable αZ when Z has density h.

2.3. **Analogy with H-Theorem.** Shannon observed [SW] that the entropy $\mathcal{S}$ is monotonic along the CLT when n is a power of 2, in other words $\mathcal{S}(S_{2^{m+1}}) \geq \mathcal{S}(S_{2^m})$ for every integer $m \geq 0$, which follows from (a rigorous proof is due to Stam [Sta])

$$\mathcal{S}\left(\frac{X_1 + X_2}{\sqrt{2}}\right) = \mathcal{S}(S_2) \geq \mathcal{S}(S_1) = \mathcal{S}(X_1).$$

By analogy with the Boltzmann H-Theorem, a conjecture attributed to Shannon (see also [Lie]) says that the Boltzmann entropy $\mathcal{S}$ is monotonic along the CLT for any n, more precisely

$$\mathcal{S}(X_1) = \mathcal{S}(S_1) \leq \cdots \leq \mathcal{S}(S_n) \leq \mathcal{S}(S_{n+1}) \leq \cdots \underset{n\to\infty}{\nearrow} \mathcal{S}(G).$$

The idea of proving the CLT using the Boltzmann entropy is very old and goes back at least to Linnik [Lin] in the 1950's, who, by the way, uses the term "Shannon entropy". But proving

convergence differs from proving monotonicity, even if these two aspects are obviously lin(ni)ked [ABBN1]. The approach of Linnik was further developed by Rényi, Csiszár, and many others, and we refer to the book of Johnson [J] for an account on this subject. The first known proof of the Shannon monotonicity conjecture is relatively recent and was published in 2004 by Artstein, Ball, Barthe, Naor [ABBN2]. The idea is to pull back the problem to the monotonicity of the Fisher information. Recall that the Fisher information of a random variable S with density g is given by

$$\mathcal{F}(S) := \int \frac{|\nabla g|^2}{g}\, dx.$$

It appears when one takes the derivative of $\mathcal{S}$ along an additive Gaussian perturbation. Namely, the de Bruijn formula states that if X, G are random variables with G standard Gaussian then

$$\partial_t \mathcal{S}(X + \sqrt{t}G) = \frac{1}{2}\mathcal{F}(X + \sqrt{t}G).$$

Indeed, if f is the density of X then the density $P_t(f)$ of $X + \sqrt{t}G$ is given by the heat kernel

$$P_t(f)(x) = (f * \mathrm{dil}_{\sqrt{t}} f_*)(x) = \int f(y) \frac{e^{-\frac{1}{2t}(y-x)^2}}{\sqrt{2\pi t}}\, dy,$$

which satisfies to $\partial_t P_t(f)(x) = \frac{1}{2}\Delta_x P_t(f)(x)$, and which gives, by integration by parts,

$$\partial_t \mathcal{S}(X + \sqrt{t}G) = -\frac{1}{2}\int (1 + \log P_t f)\Delta P_t\, dx = \frac{1}{2}\mathcal{F}(P_t f).$$

In the same spirit, we have the following integral representation, taken from [Sh2],

$$\mathcal{S}(G) - \mathcal{S}(S) = \int_0^\infty \Big(\mathcal{F}(\sqrt{e^{-2t}}S + \sqrt{1 - e^{-2t}}G) - 1\Big)\, dt.$$

This allows to deduce the monotonicity of $\mathcal{S}$ along the CLT from the one of F along the CLT,

$$\mathcal{F}(S_1) \geq \mathcal{F}(S_2) \geq \cdots \geq \mathcal{F}(S_n) \geq \mathcal{F}(S_{n+1}) \geq \cdots \searrow \mathcal{F}(G),$$

which is a tractable task [ABBN2]. The de Bruijn identity involves Brownian motion started from the random initial condition X, and the integral representation above involve the Ornstein-Uhlenbeck process started from the random initial condition X. The fact that the Fisher information $\mathcal{F}$ is non-increasing along a Markov semigroup means that the entropy $\mathcal{S}$ is concave along the Markov semigroup, a property which can be traced back to Stam (see [ABC+, Chapter 10]). As we have already mentioned before, the quantitative version of such a concavity is related to Sobolev type functional inequalities and to the notion of Bakry-Émery curvature of Markov semigroups [BGL], a concept linked with the Ricci curvature in Riemannian geometry. Recent works on this topic include extensions by Villani, Sturm, Ollivier, among others.

3. Dan-Virgil Voiculescu and the free central limit theorem

Free probability theory was forged in the 1980's by Dan-Virgil Voiculescu (1946–), while working on isomorphism problems in von Neumann operator algebras of free groups. Voiculescu discovered later in the 1990's that free probability is the algebraic structure that appears naturally in the asymptotic global spectral analysis of random matrix models as the dimension tends to infinity. Free probability theory comes among other things with algebraic analogues of the CLT and the Boltzmann entropy, see [VDN, Voi, Bia, AGZ]. The term "free" in "free probability theory" and in "free entropy" comes from the free group (see below), and has no relation with the term "free" in the Helmholtz free energy which comes from thermodynamics (available work obtainable at constant temperature). By analogy, the "free free energy" at unit temperature might be $\mathcal{A}_*(a) = \tau(V(a)) - \chi(a)$ where χ is the Voiculescu free entropy. We will see in the last sections that such a functional appears as the rate function of a large deviations principle for the empirical spectral distribution of random matrix models! This is not surprising since the Helmholtz free energy, which is nothing else but a Kullback-Leibler relative entropy, is the rate function of the large deviations principle of Sanov, which concerns the empirical measure version of the law of large numbers.

3.1. **Algebraic probability space.** Let $\mathcal{A}$ be an algebra over $\mathbb{C}$, with unity id, equipped with an involution $a \mapsto a^*$ and a normalized linear form $\tau : \mathcal{A} \to \mathbb{C}$ such that $\tau(ab) = \tau(ba)$, $\tau(id) = 1$, and $\tau(aa^*) \geq 0$. A basic non commutative example is given by the algebra of square complex matrices: $\mathcal{A} = \mathcal{M}_n(\mathbb{C})$, $id = I_n$, $a^* = \bar{a}^\top$, $\tau(a) = \frac{1}{n}\mathrm{Tr}(a)$, for which τ appears as an expectation with respect to the empirical spectral distribution: denoting $\lambda_1(a), \ldots, \lambda_n(a) \in \mathbb{C}$ the eigenvalues of a, we have

$$\tau(a) = \frac{1}{n}\sum_{k=1}^{n} \lambda_k(a) = \int x \, d\mu_a(x) \quad \text{where} \quad \mu_a := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(a)}.$$

If $a = a^*$ (we say that a is real or Hermitian) then the probability measure μ_a is supported in $\mathbb{R}$ and is fully characterized by the collection of moments $\tau(a^m)$, $m \geq 0$, which can be seen as a sort of algebraic distribution of a. Beyond this example, by analogy with classical probability theory, we may see the elements of $\mathcal{A}$ as algebraic analogues of bounded random variables, and τ as an algebraic analogue of an expectation, and $\tau(a^m)$ as the analogue of the m-th moment of a. We say that $a \in \mathcal{A}$ has mean $\tau(a)$ and variance $\tau((a-\tau(a))^2) = \tau(a^2) - \tau(a)^2$, and that a is centered when $\tau(a) = 0$. The $*$-law of a is the collection of mixed moments of a and a^* called the $*$-moments:

$$\tau(b_1 \cdots b_m) \text{ where } b_1, \ldots, b_m \in \{a, a^*\} \text{ and } m \geq 1.$$

In contrast with classical probability, the product of algebraic variables may be non commutative. When $a \in \mathcal{A}$ is real $a^* = a$, then the $*$-law of a boils down to the moments: $\tau(a^m)$, $m \geq 0$. In classical probability theory, the law of a real bounded random variable X is characterized by its moments $\mathbb{E}(X^m)$, $m \geq 0$, thanks to the (Stone-)Weierstrass theorem. When the bounded variable is not real and takes its values in $\mathbb{C}$ then we need the mixed moments $\mathbb{E}(X^m \bar{X}^n)$, $m, n \geq 0$.

One can connect $*$-law and spectrum even for non real elements. Namely, if $a \in \mathcal{M}_n(\mathbb{C})$ and if $\mu_a := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(a)}$ is its empirical spectral distribution in $\mathbb{C}$, then, for any $z \notin \{\lambda_1(a), \ldots, \lambda_n(a)\}$,

$$\begin{aligned}
\frac{1}{2}\tau(\log((a - zid)(a - zid)^*)) &= \frac{1}{n}\log|\det(a - zI_n)| \\
&= \int \log|z - \lambda| \, d\mu_a(\lambda) \\
&= (\log|\cdot| * \mu_a)(z) \\
&=: -U_{\mu_a}(z).
\end{aligned}$$

The quantity $U_{\mu_a}(z)$ is exactly the logarithmic potential at point $z \in \mathbb{C}$ of the probability measure μ_a. Since $-\frac{1}{2\pi}\log|\cdot|$ is the so-called fundamental solution of the Laplace equation in dimension 2, it follows that in the sense of Schwartz distributions,

$$\mu_a = \frac{1}{2\pi}\Delta U_{\mu_a}.$$

Following Brown, beyond the matrix case $\mathcal{A} = \mathcal{M}_n(\mathbb{C})$, this suggest to define the spectral measure of an element $a \in \mathcal{A}$ in a abstract algebra $\mathcal{A}$ as being the probability measure μ_a on $\mathbb{C}$ given by

$$\mu_a := -\frac{1}{\pi}\Delta\tau(\log((a - zid)(a - zid)^*))$$

where here again $\Delta = \partial\overline{\partial}$ is the two-dimensional Laplacian acting on z, as soon as we know how to define the operator $\log((a - zid)(a - zid)^*)$ for every z such that $a - zid$ is invertible. This makes sense for instance if $\mathcal{A}$ is a von Neumann algebra of bounded operators on a Hilbert space, since one may define $\log(b)$ if $b^* = b$ by using functional calculus. The moral of the story is that the $*$-law of a determines the $*$-law of the Hermitian element $(a - zid)(a - zid)^*$ for every $z \in \mathbb{C}$, which in turn determines the Brown spectral measure μ_a. This strategy is known as Hermitization.

The so-called Gelfand-Naimark-Segal (GNS, see [AGZ]) construction shows that any algebraic probability space can be realized as a subset of the algebra of bounded operators on a Hilbert space. Using then the spectral theorem, this shows that any compactly supported probability measure is the $*$-law of some algebraic random variable.

3.2. **Freeness.** The notion of freeness is an algebraic analogue of independence. In classical probability theory, a collection of σ-field are independent if the product of bounded random variables is centered as soon as the factors are centered and measurable with respect to different σ-fields. We say that $\mathcal{B} \subset \mathcal{A}$ is a sub-algebra of $\mathcal{A}$ when it is stable by the algebra operations, by the

involution, and contains the unity id. By analogy with classical probability, we say that the collection $(\mathcal{A}_i)_{i\in I}$ of sub-algebras of $\mathcal{A}$ are *free* when for any integer $m \geq 1$, $i_1, \ldots, i_m \in I$, and $a_1 \in \mathcal{A}_{i_1}, \ldots, a_m \in \mathcal{A}_{i_m}$, we have

$$\tau(a_1 \cdots a_m) = 0$$

as soon as $\tau(a_1) = \cdots = \tau(a_m) = 0$ and $i_1 \neq \cdots \neq i_m$ (only consecutive indices are required to be different). We say that $(a_i)_{i\in I} \subset \mathcal{A}$ are free when the sub-algebras that they generate are free. If for instance $a, b \in \mathcal{A}$ are free and centered, then $\tau(ab) = 0$, and $\tau(abab) = 0$. Note that in classical probability, the analogue of this last expression will never be zero if a and b are not zero, due to the commutation relation $abab = a^2b^2$.

Can we find examples of free matrices in $\mathcal{M}_n(\mathbb{C})$? Actually, this will not give exciting answers. Freeness is more suited for infinite dimensional operators. It turns out that the definition and the name of freeness come from a fundamental infinite dimensional example constructed from the free group. More precisely, let F_n be the free group[15] with $2n$ generators (letters and anti-letters) $g_1^{\pm 1}, \ldots, g_n^{\pm 1}$ with $n \geq 2$ (for $n = 1$, $F_n = \mathbb{Z}$ is commutative). Let $\varnothing$ be the neutral element of F_n (empty string). Let A_n be the associated free algebra identified with a sub-algebra of $\ell^2_{\mathbb{C}}(F_n)$. Each element of A_n can be seen as a finitely supported complex measure of the form $\sum_{w\in F_n} c_w \delta_w$. The collection $(\delta_w)_{w\in F_n}$ is the canonical basis: $\langle \delta_w, \delta_{w'} \rangle = \mathbf{1}_{w=w'}$. The product on A_n is the convolution of measures on F_n:

$$\left(\sum_{w\in F_n} c_w \delta_w\right)\left(\sum_{w\in F_n} c'_w \delta_w\right) = \sum_{w\in F_n}\left(\sum_{v\in F_n} c_v c'_{v^{-1}w}\right)\delta_w = \sum_{w\in F_n} (c * c')_w \delta_w.$$

Now, let $\mathcal{A}$ be the algebra over $\mathbb{C}$ of linear operators from $\ell^2_{\mathbb{C}}(F_n)$ to itself. The product in $\mathcal{A}$ is the composition of operators. The involution $*$ in $\mathcal{A}$ is the transposition-conjugacy of operators. The identity operator is denoted id. We consider the linear form $\tau : \mathcal{A} \to \mathbb{C}$ defined for every $a \in \mathcal{A}$ by

$$\tau(a) = \langle a\delta_\varnothing, \delta_\varnothing \rangle.$$

For every $w \in F_n$, let $u_w \in \mathcal{A}$ be the left translation operator defined by $u_w(\delta_v) = \delta_{wv}$. Then

$$u_w u_{w'} = u_{ww'}, \quad (u_w)^* = u_{w^{-1}}, \quad u_\varnothing = id,$$

and therefore $u_w u_w^* = u_w^* u_w = id$ (we say that u_w is unitary). We have $\tau(u_w) = \mathbf{1}_{w=\varnothing}$. Let us show that the sub-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ generated by $u_{g_1}, \ldots, u_{g_n}$ are free. Each $\mathcal{A}_i$ consists in linear combinations of $u_{g_i^r}$ with $r \in \mathbb{Z}$, and centering forces $r = 0$. For every $w_1, \ldots, w_m \in F_n$, we have

$$\tau(u_{w_1} \cdots u_{w_m}) = \tau(u_{w_1 \cdots w_m}) = \mathbf{1}_{w_1 \cdots w_m = \varnothing}.$$

Let us consider the case $w_j = g_{i_j}^{r_j} \in \mathcal{A}_{i_j}$ with $r_j \in \mathbb{Z}$, for every $1 \leq j \leq m$. Either $r_j = 0$ and $w_j = \varnothing$ or $r_j \neq 0$ and $\tau(w_j) = 0$. Let us assume that $r_1 \neq 0, \ldots, r_n \neq 0$, which implies $\tau(u_{w_1}) = \cdots = \tau(u_{w_n}) = 0$. Now since G_n is a tree, it does not have cycles, and thus if we follow a path starting from the root $\varnothing$ then we cannot go back to the root if we never go back locally along the path (this is due to the absence of cycles). Consequently, if additionally $i_1 \neq \cdots \neq i_m$ (i.e. two consecutive terms are different), then we have necessarily $w_1 \cdots w_m \neq \varnothing$, and therefore $\tau(u_{w_1} \cdots u_{w_m}) = 0$. From this observation one can conclude that $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are free.

Beyond the example of the free group: if $G_1, \ldots, G_n$ are groups and G their free product, then the algebras generated by $\{u_g : g \in G_1\}, \ldots, \{u_g : g \in G_n\}$ are always free in the one generated by $\{u_g : g \in G\}$.

3.3. **Law of free couples and free convolution.** In classical probability theory, the law of a couple of independent random variables is fully characterized by the couple of laws of the variables. In free probability theory, the $*$-law of the couple (a, b) is the collection of mixed moments in a, a^*, b, b^*. If a, b are free, then one can compute the $*$-law of the couple (a, b) by using the $*$-law of a and b, thanks to the centering trick. For instance, in order to compute $\tau(ab)$, we may write using freeness $0 = \tau((a - \tau(a))(b - \tau(b))) = \tau(ab) - \tau(a)\tau(b)$ to get $\tau(ab) = \tau(a)\tau(b)$. As a consequence, one can show that the $*$-law of a couple of free algebraic variables is fully characterized by the couple of $*$-laws of the variables. This works for arbitrary vectors of algebraic variables.

In classical probability theory, the law of the sum of two independent random variables is given by the convolution of the law of the variables. In free probability theory, the $*$-law of the sum $a + b$

[15]"Free" because it is the free product of n copies of $\mathbb{Z}$, without additional relation.

Classical probability	Free probability
Bounded r.v. X on $\mathbb{C}$	Algebra element $a \in \mathcal{A}$
$\mathbb{E}(X^m \bar{X}^n)$	$\tau(b_1 \cdots b_m)$, $b \in \{a, a^*\}^m$
Law = Moments	Law = $*$-moments
X is real	$a = a^*$
Independence	Freeness
Classical convolution $*$	Free convolution $\boxplus$
Gaussian law (with CLT)	Semicircle law (with CLT)
Boltzmann entropy $\mathcal{S}$	Voiculescu entropy χ

TABLE 2. Conceptual dictionary between classical probability and free probability. The first is commutative while the second is typically non commutative. Free probability is the algebraic structure that emerges from the asymptotic analysis, over the dimension, of the empirical spectral distribution of unitary invariant random matrices. The term free comes from the algebra of linear operators over the free algebra of the free group, an example in which the concept of freeness emerges naturally, in relation with the symmetric random walk on the infinite regular tree of even degree ≥ 4 (which is the Cayley graph of a non-commutative free group).

of two free algebraic variables $a, b \in \mathcal{A}$ is given by the so-called free convolution $\mathrm{dist}(a) \boxplus \mathrm{dist}(b)$ of the $*$-law of a and b, which can be defined using the $*$-law of the couple (a, b). Following [Bia], given an at most countable family of compactly supported probability measures on $\mathbb{R}$, one can always construct an algebraic probability space containing free algebraic variables admitting these probability measures as their $*$-distributions. The free convolution $\boxplus$ of probability measures is associative but is not distributive with respect to convex combinations (beware of mixtures!).

We have so far at hand an algebraic framework, called free probability theory, in which the concepts of algebra elements, trace, $*$-law, freeness, and free convolution are the analogue of the concepts of bounded random variables, expectation, law, independence, and convolution of classical probability theory. Do we have a CLT, and an analogue of the Gaussian? The answer is positive.

3.4. **Free CLT and semicircle law.** It is natural to define the convergence in $*$-law, denoted $\xrightarrow{*}$, as being the convergence of all $*$-moments. The Voiculescu free CLT states that if $a_1, a_2, \ldots \in \mathcal{A}$ are free, real $a_i = a_i^*$, with same $*$-law, zero mean $\tau(a_i) = 0$, unit variance $\tau(a_i^2) = 1$, then

$$s_n := \frac{a_1 + \cdots + a_n}{\sqrt{n}} \xrightarrow[n\to\infty]{*} \frac{\sqrt{4 - x^2}\mathbf{1}_{[-2,2]}}{2\pi} dx.$$

The limiting $*$-law is given by the moments of the *semicircle law*[16] on $[-2, 2]$, which are 0 for odd moments and the Catalan numbers $(C_m)_{m\geq 0}$ for even moments: for every $m \geq 0$,

$$\int_{-2}^{2} x^{2m+1} \frac{\sqrt{4 - x^2}}{2\pi} dx = 0 \quad \text{and} \quad \int_{-2}^{2} x^{2m} \frac{\sqrt{4 - x^2}}{2\pi} dx = C_m := \frac{1}{1+m}\binom{2m}{m}.$$

An algebraic variable $b \in \mathcal{A}$ has semicircle $*$-law when it is real $b = b^*$ and $\tau(b^{2m+1}) = 0$ and $\tau(b^{2m}) = C_m$ for every $m \geq 0$. The proof of the free CLT consists in computing the moments of s_n using freeness. This reveals three type of terms: terms which are zero at fixed n thanks to freeness and centering, terms having zero contribution asymptotically as $n \to \infty$, and terms which survive at the limit, and which involve only the second moment of the a_i's. See [HO][17].

As for the classical CLT, the first two moments are conserved along the free CLT: $\tau(s_n) = 0$ and $\tau(s_n^2) = 1$ for all $n \geq 1$. The semicircle $*$-law is the free analogue of the Gaussian distribution of classical probability. The semicircle $*$-law is stable by free convolution: if $a_1, \ldots, a_n$ are free with semicircle $*$-law then the $*$-law of $a_1 + \cdots + a_n$ is also semicircle and its second moment is the sum of the second moments. In particular $s_n = (a_1 + \cdots + a_n)/\sqrt{n}$ is semicircle, just like the Gaussian case in the CLT of classical probability! If $\mu_1, \ldots, \mu_n$ are semicircle laws then their free convolution $\mu_1 \boxplus \cdots \boxplus \mu_n$ is also a semicircle law and its variance is the sum of the variances.

[16] Also known as the Wigner distribution (random matrix theory) or the Sato-Tate distribution (number theory).

[17] Beyond freeness, there exists other notions of algebraic independence, allowing to compute moments, and leading to a CLT with other limiting distributions, including the Bernoulli distribution and the arcsine distribution.

3.5. **Random walks and CLT.** Let us reconsider the free group F_n. The algebraic variables

$$a_1 = \frac{u_{g_1} + u_{g_1^{-1}}}{\sqrt{2}}, \dots, a_n = \frac{u_{g_n} + u_{g_n^{-1}}}{\sqrt{2}}$$

are free, real, centered, with unit variance. Let us define

$$a = \sum_{i=1}^{n} (u_{g_i} + u_{g_i}^{-1}) = \sqrt{2} \sum_{i=1}^{n} a_i \quad \text{and} \quad p = \frac{a}{2n} = \frac{a_1 + \cdots + a_n}{\sqrt{2}n}.$$

Then a is the adjacency operator of the Cayley graph G_n of the free group F_n, which is $2n$-regular, without cycles (a tree!), rooted at $\varnothing$. Moreover, $\langle p\delta_v, \delta_w \rangle = \frac{1}{2n} \mathbf{1}_{wv^{-1} \in S}$ where $S = \{g_1^{\pm 1}, \dots, g_n^{\pm}\}$, and thus p is the transition kernel of the simple random walk on G_n. For every integer $m \geq 0$, the quantity $\tau(a^m) = \langle a^m \delta_\varnothing, \delta_\varnothing \rangle$ is the number of paths of length m in G_n starting and ending at the root $\varnothing$. From the Kesten-McKay[18] theorem [Kes, McK, HO], for every integer $m \geq 0$,

$$\tau(a^m) = \langle a^m \delta_\varnothing, \delta_\varnothing \rangle = \int x^m \, d\mu_d(x),$$

where μ_d is the Kesten-McKay distribution with parameter $d = 2n$, given by

$$d\mu_d(x) := \frac{d\sqrt{4(d-1) - x^2}}{2\pi(d^2 - x^2)} \mathbf{1}_{[-2\sqrt{d-1}, 2\sqrt{d-1}]}(x) dx.$$

By parity we have $\tau(a^{2m+1}) = 0$ for every $m \geq 0$. When $n = 1$ then $d = 2$ and G_2 is the Cayley graph of $F_1 = \mathbb{Z}$, the corresponding Kesten-McKay law μ_2 is the *arcsine law* on $[-2, 2]$,

$$\langle a^{2m} \delta_\varnothing, \delta_\varnothing \rangle = \int x^{2m} \, d\mu_2(x) = \int_{-2}^{2} \frac{x^{2m}}{\pi\sqrt{4 - x^2}} dx = \binom{2m}{m},$$

and we recover the fact that the number of paths of length $2m$ in $\mathbb{Z}$ starting and ending at the root $\varnothing$ (which is the origin 0) is given by the central binomial coefficient. The binomial combinatorics is due to the commutativity of $F_1 = \mathbb{Z}$. At the opposite side of degree, when $d = 2n \to \infty$ then μ_d, scaled by $(d-1)^{-1/2}$, tends to the semicircle law on $[-2, 2]$:

$$\lim_{d \to \infty} \frac{\langle a^{2m} \delta_\varnothing, \delta_\varnothing \rangle}{(d-1)^m} = \lim_{d \to \infty} \int \left(\frac{x}{\sqrt{d-1}} \right)^{2m} d\mu_d(x) = \int_{-2}^{2} y^{2m} \frac{\sqrt{4 - y^2}}{2\pi} dy = C_m := \frac{1}{1+m} \binom{2m}{m}.$$

As a consequence, we have, thanks to $(d-1)^m \sim_{n \to \infty} d^m = (2n)^m = (\sqrt{2n})^{2m}$,

$$\tau\left(\left(\frac{a_1 + \cdots + a_n}{\sqrt{n}} \right)^{2m} \right) = \frac{\tau(a^{2m})}{(2n)^m} = \frac{\langle a^{2m} \delta_\varnothing, \delta_\varnothing \rangle}{d^m} \underset{n \to \infty}{\longrightarrow} C_m.$$

This is nothing else but a free CLT for the triangular array $((a_1, \dots, a_n))_{n \geq 1}$! The free CLT is the algebraic structure that emerges from the asymptotic analysis, as $n \to \infty$, of the combinatorics of loop paths of the simple random walk on the Cayley graph G_n of the free group F_n, and more generally on the d-regular infinite graph without cycles (a tree!) as the degree d tends to infinity.

We have $a = b_1 + \cdots + b_n$ where $b_i = \sqrt{2} a_i = u_{g_i} + u_{g_i^{-1}}$. But for any $m \in \mathbb{N}$ we have $\tau(b_i)^{2m+1} = 0$ and $\tau(b_i^{2m}) = \sum_{r=1}^{2m} \binom{2m}{r} \tau(u_{g_i^{2(r-m)}}) = \binom{2m}{m}$, and therefore the $*$-law of b_i is the arcsine law μ_2, and thanks to the freeness of $b_1, \dots, b_n$ we obtain $\mu_d = \mu_2 \boxplus \cdots \boxplus \mu_2$ (d times).

[18] Appears for regular trees of even degree (Cayley graph of free groups) in the doctoral thesis of Harry Kesten (1931 –) published in 1959. It seems that no connections were made at that time with the contemporary works [W] of Eugene Wigner (1902 – 1995) on random matrices published in the 1950's. In 1981, Brendan McKay (1951 –) showed [McK] that these distributions appear as the limiting empirical spectral distributions of the adjacency matrices of sequences of (random) graphs which are asymptotically regular and without cycles (trees!). He does not cite Kesten and Wigner.

3.6. **Free entropy.** Inspired by the Boltzmann H-Theorem view of Shannon on the CLT of classical probability theory, one may ask if there exists, in free probability theory, a free entropy functional, maximized by the semicircle law at fixed second moment, and which is monotonic along the free CLT. We will see that the answer is positive. Let us consider a real algebraic variable $a \in \mathcal{A}$, $a^* = a$, such that there exists a probability measure μ_a on $\mathbb{R}$ such that for every integer $m \geq 0$,

$$\tau(a^m) = \int x^m \, d\mu_a(x).$$

Inspired from the micro-macro construction of the Boltzmann entropy, one may consider an approximation at the level of the moments of the algebraic variable a (which is in general infinite dimensional) by Hermitian matrices (which are finite dimensional). Namely, following Voiculescu, for every real numbers $\varepsilon > 0$ and $R > 0$ and integers $m \geq 0$ and $d \geq 1$, let $\Gamma_R(a; m; d, \varepsilon)$ be the relatively compact set of Hermitian matrices $h \in \mathcal{M}_d(\mathbb{C})$ such that $\|h\| \leq R$ and $\max_{0 \leq k \leq m} \left| \tau(a^k) - \frac{1}{d} \mathrm{Tr}(h^k) \right| \leq \varepsilon$. The volume $|\Gamma_R(a; m, d, \varepsilon)|$ measures the degree of freedom of the approximation of the algebraic variable a by matrices, and is the analogue of the cardinal (which was multinomial) in the combinatorial construction of the Boltzmann entropy. We find

$$\chi(a) := \sup_{R>0} \inf_{m \in \mathbb{N}} \inf_{\varepsilon > 0} \overline{\lim_{d \to \infty}} \left(\frac{1}{d^2} \log |\Gamma_R(a; m, d, \varepsilon)| + \frac{\log(d)}{2} \right) = \iint \log |x - y| \, d\mu_a(x) d\mu_a(y).$$

This quantity depends only on μ_a and is also denoted $\chi(\mu_a)$. It is a quadratic form in μ_a. It is the Voiculescu entropy functional [Voi]. When μ is a probability measure on $\mathbb{C}$, we will still denote

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

However, this is not necessarily the entropy of an algebraic variable when μ is not supported in $\mathbb{R}$. The Voiculescu free entropy should not be confused with the von Neumann entropy in quantum probability defined by $S(a) = -\tau(a \log(a))$, which was studied by Lieb in [Lie]. For some models of random graphs, one can imitate Voiculescu and forge some sort of graphical Boltzmann entropy, which can be related to a large deviations rate functional, see [BC1] and references therein.

The semicircle law is for the free entropy the analogue of the Gaussian law for the Boltzmann entropy. The semicircle law on $[-2, 2]$ is the unique law that maximizes the Voiculescu entropy χ among the laws on $\mathbb{R}$ with second moment equal to 1, see [AGZ]:

$$\arg\max \left\{ \chi(\mu) : \mathrm{supp}(\mu) \subset \mathbb{R}, \int x^2 \, d\mu(x) = 1 \right\} = \frac{\sqrt{4 - x^2} \mathbf{1}_{[-2,2]}(x)}{2\pi} dx.$$

How about laws on $\mathbb{C}$ instead of $\mathbb{R}$? The uniform law on the unit disc is the unique law that maximizes the functional χ among the set of laws on $\mathbb{C}$ with second moment (mean squared modulus) equal to 1, see [ST] (here $z = x + iy$ and $dz = dxdy$):

$$\arg\max \left\{ \chi(\mu) : \mathrm{supp}(\mu) \subset \mathbb{C}, \int |z|^2 \, d\mu(z) = 1 \right\} = \frac{\mathbf{1}_{\{z \in \mathbb{C} : |z| = 1\}}}{\pi} dz,$$

Under the uniform law on the unit disc, the real and the imaginary parts follow the semicircle law on $[-1, 1]$, and are not independent. If we say that an algebraic variable $c \in \mathcal{A}$ is circular when its $*$-law is the uniform law on the unit disc of $\mathbb{C}$, then, if $s_1, s_2 \in \mathcal{A}$ are free with $*$-law equal to the semicircle law on $[-2, 2]$, then $\frac{s_1 + i s_2}{\sqrt{2}}$ is circular (here $i = (0, 1) \in \mathbb{C}$ is such that $i^2 = -1$).

It turns out that the Voiculescu free entropy χ is monotonic along the Voiculescu free CLT:

$$\chi(a) = \chi(s_1) \leq \cdots \leq \chi(s_n) \leq \chi(s_{n+1}) \leq \cdots \underset{n \to \infty}{\nearrow} \max \chi(s)$$

where still $s_n = n^{-1/2}(a_1 + \cdots + a_n)$ and where s is an semicircle algebraic variable. Shlyakhtenko gave a proof of this remarkable fact, based on a free Fisher information functional (which is a Hilbert transform), that captures simultaneously the classical and the free CLT [Sh2, Sh1]. The Boltzmann-Shannon H-Theorem interpretation of the CLT is thus remarkably valid in classical probability theory, and in free probability theory.

3.7. **A theorem of Eugene Wigner.** Let H be a random $n \times n$ Hermitian matrix belonging to the Gaussian Unitary Ensemble (GUE). This means that H has density proportional to

$$e^{-\frac{n}{4}\mathrm{Tr}(H^2)} = e^{-\frac{n}{4}\sum_{1\leq j\leq 1}^{n} H_{jj}^2 - \frac{n}{2}\sum_{1\leq j<k\leq n}|H_{jk}|^2}.$$

The entries of H are Gaussian, centered, independent, with variance $2/n$ on the diagonal and $1/n$ outside. Let $\mu_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j(H)}$ be the empirical spectral distribution of H, which is a random discrete probability measure on $\mathbb{R}$. For every $m \geq 0$, the mean m-th moment of μ_n is given by

$$\mathbb{E}\int x^m\, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n}\lambda_j^m = \frac{\mathbb{E}\mathrm{Tr}(H^m)}{n} = \sum_{i_1,\ldots,i_m}\frac{\mathbb{E}(H_{i_1 i_2}\cdots H_{i_{m-1}i_m}H_{i_m i_1})}{n}.$$

In particular, the first two moments of $\mathbb{E}\mu_n$ satisfy to

$$\mathbb{E}\int x\, d\mu_n(x) = \frac{\mathbb{E}(\mathrm{Tr}(H))}{n} = \frac{\mathbb{E}\sum_{j=1}^{n} H_{jj}}{n} = 0$$

and

$$\mathbb{E}\int x^2\, d\mu_n(x) = \frac{\mathbb{E}(\mathrm{Tr}(H^2))}{n} = \mathbb{E}\frac{\sum_{j,k=1}^{n}|H_{jk}|^2}{n} = \frac{n(2/n)+(n^2-n)(1/n)}{n} \underset{n\to\infty}{\longrightarrow} 1.$$

More generally, for any $m > 2$, the computation of the limiting m-th moment of $\mathbb{E}\mu_n$ boils down to the combinatorics of the paths $i_1 \to i_2 \to \cdots \to i_m$. It can be shown that the surviving terms as $n \to \infty$ correspond to paths forming a tree and passing exactly zero or two times per each edge. This gives finally, for every integer $m \geq 0$, denoting C_m the m-th Catalan number,

$$\lim_{n\to\infty}\mathbb{E}\int x^{2m+1}\, d\mu_n(x) = 0 \quad \text{and} \quad \lim_{n\to\infty}\mathbb{E}\int x^{2m}\, d\mu_n(x) = C_m.$$

This means that $\mathbb{E}\mu_n$ tends as $n \to \infty$ in the sense of moments to the semicircle law on $[-2, 2]$ (which has unit variance). Just like the CLT, the result is actually universal, in the sense that it remains true if one drops the Gaussian distribution assumption of the entries. This is the famous Wigner theorem [W], in its modern general form, named after Eugene Paul Wigner (1902 – 1995). The GUE case is in fact exactly solvable: one can compute the density of the eigenvalues, which turns out to be proportional to

$$\prod_{j=1}^{n} e^{-\frac{n}{4}\lambda_j^2}\prod_{j<k}(\lambda_j-\lambda_k)^2 = \exp\left(-\frac{n}{4}\sum_{j=1}^{n}\lambda_j^2 - \sum_{j\neq k}\log\frac{1}{|\lambda_j-\lambda_k|}\right).$$

The logarithmic repulsion is the Coulomb repulsion in dimension 2. Also, this suggest to interpret the eigenvalues $\lambda_1, \ldots, \lambda_n$ as a Coulomb gas of two-dimensional charged particles forced to stay in a one dimensional ramp (the real line) and experiencing a confinement by a quadratic potential. These formulas allow to deduce the semicircle limit of the one-point correlation (density of $\mathbb{E}\mu_n$), by using various methods, such as orthogonal polynomials, or large deviations theory, see [AGZ].

3.8. **Asymptotic freeness of unitary invariant random matrices.** If A and B are two Hermitian $n \times n$ matrices, then the spectrum of $A + B$ depend not only on the spectrum of A and the spectrum of B, but also on the eigenvectors[19] of A and B. Now if A and B are two independent random Hermitian matrices, there is no reason to believe that the empirical spectral distribution μ_{A+B} of $A + B$ depend only on the empirical spectral distributions μ_A and μ_B of A and B. Let A be a $n \times n$ random Hermitian matrix in the GUE, normalized such that μ_A has a mean second moment equal to 1. Then the Wigner theorem says that $\mathbb{E}\mu_A$ tend in the sense of moments, as $n \to \infty$, to the semicircle law of unit variance. If B is an independent copy of A, then, thanks to the convolution of Gaussian laws, $A + B$ is identical in law to $\sqrt{2}A$, and thus $\mathbb{E}\mu_{A+B}$ tend, in the sense of moments, as $n \to \infty$, to the semicircle law of variance 2. Then, thanks to the free convolution of semicircle laws, we have

$$\mathbb{E}\mu_{A+B} - \mathbb{E}\mu_A \boxplus \mathbb{E}\mu_B \underset{n\to\infty}{\overset{*}{\longrightarrow}} 0,$$

[19] If A and B commute, then they admit the same eigenvectors and the spectrum of $A + B$ is the sum of the spectra of A and B, but this depends on the way we label the eigenvalues, which depends in turn on the eigenvectors!

where $\stackrel{*}{\to}$ denotes the convergence of moments. Voiculescu has established that this *asymptotic freeness* phenomenon remains actually true beyond the GUE case, provided that the eigenspaces of the two matrices are randomly decoupled using a random unitary conjugation. For example, let A and B two $n \times n$ Hermitian matrices such that $\mu_A \to \mu_a$ and $\mu_B \to \mu_b$ in the sense of moments as $n \to \infty$, where μ_a and μ_b are two compactly supported laws on $\mathbb{R}$. Let U and V be independent random unitary matrices uniformly distributed on the unitary group (we say Haar unitary). Then

$$\mathbb{E}\mu_{UAU^*+VBV^*} \stackrel{*}{\underset{n\to\infty}{\longrightarrow}} \mu_a \boxplus \mu_b.$$

See [AGZ]. This asymptotic freeness reveals that free probability is the algebraic structure that emerges from asymptotic analysis of large dimensional unitary invariant models of random matrices.

Since the functional χ is maximized by the uniform law on the unit disc, one may ask about an analogue of the Wigner theorem for non-Hermitian random matrices. The answer is positive. For our purposes, we will focus on a special ensemble of random matrices, introduced by Jean Ginibre[20] in the 1960's in [Gin], for which one can compute explicitly the density of the eigenvalues.

4. Jean Ginibre and his ensemble of random matrices

One of the most fascinating result in the asymptotic analysis of large dimensional random matrices is the circular law for the complex Ginibre ensemble, which can be proved using the Voiculescu functional χ (maximized at fixed second moment by uniform law on unit disc).

4.1. **Complex Ginibre ensemble.** A simple model of random matrix is the Ginibre model:

$$G = \begin{pmatrix} G_{11} & \cdots & G_{1n} \\ \vdots & \vdots & \vdots \\ G_{n1} & \cdots & G_{nn} \end{pmatrix}$$

where $(G_{jk})_{1\le j,k\le n}$ are i.i.d. random variables on $\mathbb{C}$, with $\mathfrak{Re}G_{jk}, \mathfrak{Im}G_{jk}$ of Gaussian law of mean 0 and variance $1/(2n)$. In particular, $\mathbb{E}(|G_{jk}|^2) = 1/n$. The density of G is proportional to

$$\prod_{j,k=1}^{n} e^{-n|G_{jk}|^2} = e^{-\sum_{j,k=1}^{n} n|G_{jk}|^2} = e^{-n\mathrm{Tr}(GG^*)}.$$

This shows that the law of G is unitary invariant, meaning that UGU^* and G have the same law, for every unitary matrix U. Can we compute the law of the eigenvalues of G? Let $G = UTU^*$ be the Schur unitary triangularization of G. Here $T = D + N$ where D is diagonal and N is upper triangular with null diagonal. In particular N is nilpotent, T is upper triangular, and the diagonal of D is formed with the eigenvalues $\lambda_1, \ldots, \lambda_n$ in $\mathbb{C}$ of G. The Jacobian of the change of variable $G \mapsto (U, D, N)$ is proportional to $\prod_{1\le j<k\le n} |\lambda_j - \lambda_k|^2$ (for simplicity, we neglect here delicate problems related to the non-uniqueness of the Schur unitary decomposition). On the other hand,

$$\mathrm{Tr}(GG^*) = \mathrm{Tr}(DD^*) + \mathrm{Tr}(DN^*) + \mathrm{Tr}(ND^*) + \mathrm{Tr}(NN^*) = \mathrm{Tr}(DD^*).$$

This allows to integrate out the U, N variables. The law of the eigenvalues is then proportional to

$$e^{-n\sum_{j=1}^{n} |\lambda_j|^2} \prod_{1\le j<k\le n} |\lambda_j - \lambda_k|^2.$$

This defines a determinantal process on $\mathbb{C}$: the complex Ginibre ensemble [For, KSo, HKPV2].

4.2. **Circular law for the complex Ginibre ensemble.** In order to interpret the law of the eigenvalues as a Boltzmann measure, we put the Vandermonde determinant inside the exponential:

$$e^{-n\sum_j |\lambda_j|^2 + 2\sum_{j<k} \log|\lambda_j - \lambda_k|}.$$

If we encode the eigenvalues by the empirical measure $\mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j}$, this takes the form

$$e^{-n^2 \mathcal{I}(\mu_n)}$$

[20] Jean Ginibre is also famous for FKG inequalities, and for scattering for Schrödinger operators.

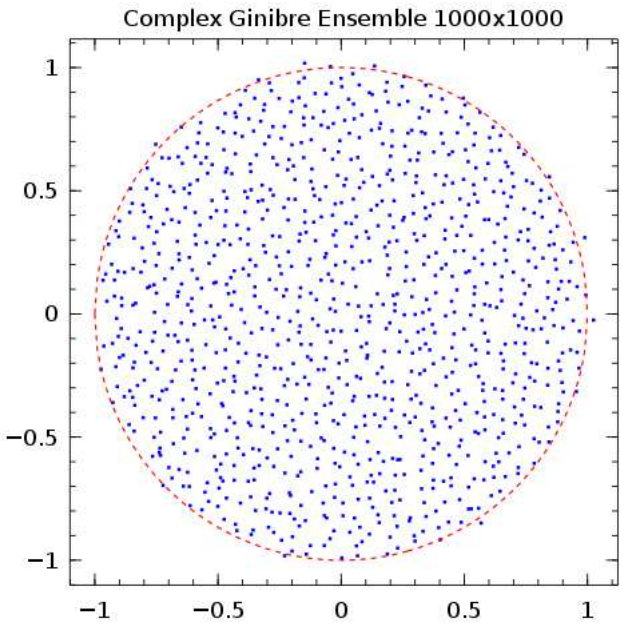


FIGURE 1. The eigenvalues of a single matrix drawn from the complex Ginibre ensemble of random matrices. The dashed line is the unit circle. This numerical experiment was performed using the promising Julia `http://julialang.org/`

```
Pkg.add("Winston"); using Winston # Pkg.add("Winston") is needed once for all.
n=1000; (D,U)=eig((randn(n,n)+im*randn(n,n))/sqrt(2*n)); I=[-1:.01:1];
J=sqrt(1-I.^2); hold(true); plot(real(D),imag(D),"b.",I,J,"r--",I,-J,"r--")
title(@sprintf("Complex Ginibre Ensemble %dx%d",n,n)); file("circ.png")
```

where the "energy" $\mathcal{I}(\mu_n)$ of the configuration μ_n is defined via

$$\mathcal{I}(\mu) := \int |z|^2 \, d\mu(z) + \iint_{\neq} \log \frac{1}{|z - z'|} \, d\mu(z) d\mu(z').$$

This suggests to interpret the eigenvalues $\lambda_1, \ldots, \lambda_n$ of G as Coulomb gas of two-dimensional charged particles, confined by a an external field (quadratic potential) and subject to pair Coulomb repulsion. Note that $-\mathcal{I}$ can also be seen as a penalized Voiculescu functional. Minimizing a penalized functional is equivalent to minimizing without penalty but under constraint (Lagrange). Presently, if $\mathcal{M}$ is the set of probability measures on $\mathbb{C}$ then $\inf_{\mathcal{M}} \mathcal{I} > -\infty$ and the infimum is achieved at a unique probability measure μ_*, which is the uniform law on the unit disc of $\mathbb{C}$.

How does the random discrete probability measure μ_n behave as $n \to \infty$? Following Hiai and Petz [PH][21], one may adopt a large deviations approach. Let $\mathcal{M}$ be the set of probability measures on $\mathbb{C}$. One may show that the functional $\mathcal{I} : \mathcal{M} \to \mathbb{R} \cup \{+\infty\}$ is lower semi continuous for the topology of narrow convergence, is strictly convex, and has compact level sets. Let us consider a distance compatible with the topology. It can be shown that for every ball B for this distance,

$$\mathbb{P}(\mu_n \in B) \approx \exp\left(-n^2 \inf_B (\mathcal{I} - \inf_{\mathcal{M}} \mathcal{I})\right)$$

Now either $\mu_* \in B$ and in this case $\mathbb{P}(\mu_n \in B) \approx 1$, or $\mu_* \notin B$ and in this case $\mathbb{P}(\mu_n \in B) \to 0$ exponentially fast. Actually, the first Borel-Cantelli lemma allows to deduce that almost surely

$$\lim_{n \to \infty} \mu_n = \mu_* = \arg\inf \mathcal{I} = \frac{\mathbf{1}_{\{z \in \mathbb{C} : |z| \leq 1\}}}{\pi} dz,$$

[21]See also Anderson, Guionnet, and Zeitouni [AGZ], Ben Arous and Zeitouni [BAZ], and Hardy [H].

where $z = x + iy$ and $dz = dxdy$. This phenomenon is known as the circular law. If one starts with a Hermitian random Gaussian matrix – the Gaussian Unitary Ensemble (GUE) – then the same analysis is available, and produces a convergence to the semicircle law on $[-2, 2]$.

The circular law is universal, in the sense that it remains valid if one drops the Gaussian assumption of the entries of the matrix, while keeping the i.i.d. structure and the $1/n$ variance. This was the subject of a long series of works by Girko, Bai, Tao and Vu, among others, see [T, BS, BC2]. Another way to go beyond the Gaussian case is to start from the Coulomb gas and to replace the quadratic confining potential $|\cdot|^2$ by a more general potential $V : \mathbb{C} \to \mathbb{R}$, not necessarily radial. This type of generalization was studied for instance by Saff and Totik, and by Hiai and Petz, among others, see for instance [ST, HP, AGZ, H].

Beyond random matrices, how about the empirical measure of random particles in $\mathbb{R}^d$ with Coulomb type singular repulsion and external field confinement? Is there an analogue of the circular law phenomenon? Does the ball replace the disc? The answer is positive.

5. Beyond random matrices

Most of the material of this section comes from our work [CGZ] with N. Gozlan and P.-A. Zitt.

5.1. **The model.** We consider a system of N particles in $\mathbb{R}^d$ at positions $x_1, \dots, x_N$, say with charge $1/N$. These particles are subject to confinement by an external field via a potential $x \in \mathbb{R}^d \mapsto V(x)$, and to internal pair interaction (typically repulsion) via a potential $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto W(x, y)$ which is symmetric: $W(x, y) = W(y, x)$. The idea is that an equilibrium may emerge as N tends to infinity. The configuration energy is

$$\begin{aligned}\mathcal{I}_N(x_1, \dots, x_N) &= \sum_{i=1}^N \frac{1}{N} V(x_i) + \sum_{1 \le i < j \le N} \frac{1}{N^2} W(x_i, x_j) \\ &= \int V(x)\, d\mu_N(x) + \frac{1}{2} \iint_{\neq} W(x, y) d\mu_N(x) d\mu_N(y)\end{aligned}$$

where μ_N is the empirical measure of the particles (global encoding of the particle system)

$$\mu_N := \frac{1}{N} \sum_{k=1}^N \delta_{x_k}.$$

The model is mean-field in the sense that each particle interacts with the others only via the empirical measure of the system. If $1 \le d \le 2$ then one can construct a random normal matrix which admits ours particles $x_1, \dots, x_N$ as eigenvalues: for any $n \times n$ unitary matrix U,

$$M = U \mathrm{Diag}(x_1, \dots, x_N) U^*,$$

which is unitary invariant if U is Haar distributed. However, we are more interested in an arbitrarily high dimension d, for which no matrix model is available. We make our particles $x_1, \dots, x_N$ random by considering the exchangeable probability measure P_N on $(\mathbb{R}^d)^N$ with density proportional to

$$e^{-\beta_N \mathcal{I}_N(x_1, \dots, x_N)}$$

where $\beta_N > 0$ is a positive parameter which may depend on N. The law P_N is a Boltzmann measure at inverse temperature β_N, and takes the form $\prod_{i=1}^N f_1(x_i) \prod_{1 \le i < j \le N} f_2(x_i, x_j)$ due to the structure and symmetries of $\mathcal{I}_N$. The law P_N on $(\mathbb{R}^d)^N$ is informally the invariant law[22] of the reversible diffusion process $(X_t)_{t \in \mathbb{R}_+}$ solution of the system of stochastic differential equations

$$dX_{t,i} = \sqrt{\frac{2}{\beta_N}} dB_{t,i} - \frac{1}{N} \nabla V(X_{t,i}) dt - \frac{1}{N^2} \sum_{j \neq i} \nabla_1 W(X_{t,i}, X_{t,j}) dt.$$

This can be seen as a special McKean-Vlasov mean-field particle system with potentially singular interaction. The infinitesimal generator of this Markov process is $L = \beta_N^{-1} \Delta - \nabla \mathcal{I}_N \cdot \nabla$. The process may explode in finite time depending on the singularity of the interaction W on the

[22] One may also view P_N as the steady state of a Fokker-Planck evolution equation with conservation laws.

diagonal (e.g. collisions of particles). The Helmholtz free energy of this Markov process is given, for every probability density f on $(\mathbb{R}^d)^N$,

$$\int_{(\mathbb{R}^d)^N} \mathcal{I}_N \, f \, dx - \frac{1}{\beta_N}\mathcal{S}(f) = \int \left(\int V \, d\mu_N + \frac{1}{2} \iint_{\neq} W \, d\mu_N^{\otimes 2} \right) f \, dx - \frac{1}{\beta_N}\mathcal{S}(f).$$

The model contains the complex Ginibre ensemble of random matrices as the special case

$$d = 2, \ \beta_N = N^2, \ V(x) = |x|^2, \ W(x,y) = 2 \log \frac{1}{|x-y|},$$

which is two-dimensional, with quadratic confinement, Coulomb repulsion, and temperature $1/N^2$. Beyond this two-dimensional example, the typical interaction potential W that we may consider is the Coulomb interaction in arbitrary dimension (we denote by $|\cdot|$ the Euclidean norm of $\mathbb{R}^d$)

$$W(x,y) = k_\Delta(x-y) \quad \text{with} \quad k_\Delta(x) = \begin{cases} -|x| & \text{if } d = 1 \\ \log \frac{1}{|x|} & \text{if } d = 2 \\ \frac{1}{|x|^{d-2}} & \text{if } d \geq 3 \end{cases}$$

and the Riesz interaction, $0 < \alpha < d$ (Coulomb if $d \geq 3$ and $\alpha = 2$) $d \geq 1$

$$W(x,y) = k_{\Delta_\alpha}(x-y) \quad \text{with} \quad k_{\Delta_\alpha}(x) = \frac{1}{|x|^{d-\alpha}}.$$

The Coulomb kernel k_Δ is the fundamental solution of the Laplace equation, while the Riesz kernel k_{Δ_α} is the fundamental solution of the fractional Laplace equation, hence the notations. In other words, in the sense of Schwartz distributions, for some constant c_d,

$$\Delta_\alpha k_{\Delta_\alpha} = c_d \delta_0.$$

If $\alpha \neq 2$ then the operator Δ_α is a non-local Fourier multiplier.

5.2. **Physical control problem.** With the Coulomb-Gauss theory of electrostatic phenomena in mind, it is tempting to consider the following physical control problem: given an internal interaction potential W and a target probability measure μ_* in $\mathbb{R}^d$, can we tune the external potential V and the cooling scheme β_N is order to force the empirical measure $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ to converge to μ_* as $N \to \infty$? One can also instead fix V and seek for W.

5.3. **Topological result: large deviations principle.** The confinement is always needed in order to produce a non degenerate equilibrium. This is done here by an external field. This can also be done by forcing a compact support, as Frostman did in his doctoral thesis [Fro], or by using a manifold as in Dragnev and Saff [DS] and Berman [Ber]. Let $\mathcal{M}_1$ be the set of probability measures on $\mathbb{R}^d$ equipped with the topology of narrow convergence, which is the dual convergence with respect to continuous and bounded test functions. From the expression of $\mathcal{I}_N$ in terms of μ_N, the natural limiting energy functional is the quadratic form with values in $\mathbb{R} \cup \{+\infty\}$ defined by

$$\mu \in \mathcal{M}_1 \mapsto \mathcal{I}(\mu) = \int V(x) \, d\mu(x) + \frac{1}{2} \iint W(x,y) \, d\mu(x) d\mu(y).$$

We make the following assumptions on V and W (fulfilled for instance when the localization potential is quadratic $V = c|\cdot|^2$ and the interaction potential W is Coulomb or Riesz).

- *Localization and repulsion.*
  - the function $V : \mathbb{R}^d \to \mathbb{R}$ continuous, $V(x) \to +\infty$ as $|x| \to +\infty$, $e^{-V} \in L^1(dx)$;
  - the function $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ continuous, finite outside diagonal, and symmetric $W(x,y) = W(y,x)$, (however W can be infinite on the diagonal!);
- *Near infinity confinement beats repulsion.* For some constants $c \in \mathbb{R}$ and $\varepsilon_o \in (0,1)$,

$$\forall x, y \in \mathbb{R}^d, \quad W(x,y) \geq c - \varepsilon_o(V(x) + V(y));$$

- *Near diagonal repulsion is integrable.* for every compact $K \subset \mathbb{R}^d$,

$$z \mapsto \sup_{x,y \in K, |x-y| \geq |z|} W(x,y) \in L^1(dz);$$

- *Regularity.* $\forall \nu \in \mathcal{M}_1(\mathbb{R}^d)$, if $\mathcal{I}(\nu) < \infty$ then

$$\exists (\nu_n) \in \mathcal{M}_1(\mathbb{R}^d), \quad \nu_n \ll dx, \quad \nu_n \to \nu, \quad \mathcal{I}(\nu_n) \to \mathcal{I}(\nu);$$

- *Cooling scheme.* $\beta_N \gg N\log(N)$ (for the Ginibre ensemble $\beta_N = N^2$).

Under these assumptions, it is proven in [CGZ] that the sequence $(\mu_N)_{N\geq 1}$ of random variables taking values in $\mathcal{M}_1$ satisfies to a Large Deviations Principle (LDP) at speed β_N with good rate function $\mathcal{I} - \inf_{\mathcal{M}_1} \mathcal{I}$. In other words:

- Rate function: $\mathcal{I}$ is lower semi-continuous with compact level sets, and $\inf_{\mathcal{M}_1} \mathcal{I} > -\infty$;
- LDP lower and upper bound: for every Borel set A in $\mathcal{M}_1$,

$$\liminf_{N\to\infty} \frac{\log P_N(\mu_N \in A)}{\beta_N} \geq - \inf_{\mu\in \mathrm{int}(A)} (\mathcal{I} - \inf \mathcal{I})(\mu)$$

and

$$\limsup_{N\to\infty} \frac{\log P_N(\mu_N \in A)}{\beta_N} \leq - \inf_{\mu\in \mathrm{clo}(A)} (\mathcal{I} - \inf \mathcal{I})(\mu);$$

- Convergence: $\arg\inf \mathcal{I}$ is not empty and almost surely $\lim_{N\to\infty} \mathrm{dist}(\mu_N, \arg\inf \mathcal{I}) = 0$ where dist is the bounded-Lipschitz dual distance (it induces the narrow topology on $\mathcal{M}_1$).

This LDP must be seen as an attractive tool in order to show the convergence of μ_N. This topological result is built on the idea that the density of P_N is proportional as $N \gg 1$ to

$$e^{-\beta_N \mathcal{I}_N(\mu_N)} \approx e^{-\beta_N \mathcal{I}(\mu_N)},$$

and thus, informally, the first order global asymptotics as $N \gg 1$ is

$$\mu_N \approx \arg\inf \mathcal{I}.$$

This generalizes the case of the complex Ginibre ensemble considered in the preceding section. At this level of generality, the set $\arg\inf \mathcal{I}$ is non-empty but is not necessarily a singleton. In the sequel, we provide a more rigid differential result in the case where W is the Riesz potential, which ensures that $\mathcal{I}$ admits a unique minimizer, which is characterized by simple properties.

The $N\log(N)$ in condition $\beta_N \gg N\log(N)$ comes from volumetric (combinatorial) estimates.

It is important to realize that W is potentially singular on the diagonal, and that this forbids the usage of certain LDP tools such as the Laplace-Varadhan lemma or the Gärtner-Ellis theorem, see [DZ]. If W is continuous and bounded on $\mathbb{R}^d \times \mathbb{R}^d$ then one may deduce our LDP from the LDP with $W \equiv 0$ by using the Laplace-Varadhan lemma. Moreover, if $W \equiv 0$ then $P_N = \eta_N^{\otimes N}$ is a product measure with $\eta_N \propto e^{-(\beta_N/N)V}$, the particles are independent, the rate functional is

$$\mathcal{I}(\mu) - \inf_{\mathcal{M}_1} \mathcal{I} = \int V\, d\mu - \inf V, \quad \text{and} \quad \arg\inf_{\mathcal{M}_1} \mathcal{I} = \mathcal{M}_V = \{\mu \in \mathcal{M}_1 : \mathrm{supp}(\mu) \subset \arg\inf V\},$$

which gives, thanks to $\beta_N \gg N\log(N) \gg N$,

$$\lim_{N\to\infty} \mathrm{dist}(\mu_N, \mathcal{M}_V) = 0.$$

5.4. **Linear inverse temperature and link with Sanov theorem.** If $\beta_N = N$ and $W \equiv 0$ then $P_N = (\mu_*)^{\otimes N}$ is a product measure, the law μ_* has density proportional to e^{-V}, the particles are i.i.d. of law μ_*, and the Sanov theorem states that $(\mu_N)_{N\geq 1}$ satisfies to an LDP in $\mathcal{M}_1$ with rate function given by the Kullback-Leibler relative entropy $\mathcal{K}$ with respect to μ_*. If $W \in \mathcal{C}_b$ then the particles are no longer independent but the Laplace-Varadhan lemma allows to deduce that $(\mu_N)_{N\geq 1}$ satisfies to an LDP in $\mathcal{M}_1$ with rate function $\mathcal{R}$ given by

$$\mathcal{R}(\mu) = \mathcal{K}(\mu) + \frac{1}{2}\iint W(x,y)\, d\mu(x) d\mu(y) = -\mathcal{S}(\mu) + \mathcal{I}(\mu)$$

where $\mathcal{S}(\mu) = \mathcal{S}(f)$ if $d\mu = f d\mu_*$ and $\mathcal{S}(\mu) = +\infty$ otherwise. We have here a contribution of the Boltzmann entropy $\mathcal{S}$ and a Voiculescu type functional χ via its penalized version $\mathcal{I}$. Various versions of this LDP was considered in the literature in various fields and in special situations, for instance in the works of Messer and Spohn [MS], Kiessling [Kie], Caglioti, Lions, Marchioro, and Pulvirenti [CLMP], Bodineau and Guionnet [BG], and Kiessling and Spohn [KSpo], among others.

5.5. **Differential result: rate function analysis.** Recall that $\mathcal{I}$ is a quadratic form on measures.

$$\frac{t\mathcal{I}(\mu)+(1-t)\mathcal{I}(\nu)-\mathcal{I}(t\mu+(1-t)\nu)}{t(1-t)}=\iint W\,d(\mu-\nu)^2.$$

This shows that $\mathcal{I}$ is convex if and only if W is weakly positive in the sense of Bochner. Term "weak" comes from the fact that we need only to check positivity on measures which are differences of probability measures. The Coulomb kernel in dimension $d=2$ is not positive, but is weakly positive. It turns out that every Coulomb or Riesz kernel is weakly positive, in any dimension. Consequently, the functional $\mathcal{I}$ is convex in the case of Coulomb or Riesz interaction, for every dimension d. One may also rewrite the quadratic form $\mathcal{I}$ as

$$\mathcal{I}(\mu)=\int V\,d\mu+\frac{1}{2}\int U_\mu\,d\mu\quad\text{where}\quad U_\mu(x):=\int W(x,y)\,d\mu(y).$$

In the Coulomb case then U_μ is the logarithmic potential in dimension $d=2$, and the Coulomb or minus the Newton potential in higher dimensions. An infinite dimensional Lagrange variational analysis gives that the gradient of the functional $\mathcal{I}$ at point μ is $V+U_\mu$ should be constant on the support of the optimum, and in the case where $W(x,y)=k_D(x-y)$ where k_D is the fundamental solution of a local (say differential) operator D, meaning $Dk_D=-\delta_0$ in the sense of distributions, we have $DU_\mu=-\mu$ which gives finally that on the support of μ_*,

$$\mu_*=DV,$$

In particular, if V is the squared Euclidean norm and if W is the Coulomb repulsion then D is the Laplacian and DV is constant, which suggests that μ_* is constant on its support (this is compatible with what we already know for dimension $d=2$, namely μ_* is uniform on the unit disc of $\mathbb{C}$).

We show in [CGZ] that if W is the Riesz interaction, then the functional $\mathcal{I}$ is strictly convex, and $\arg\inf\mathcal{I}=\{\mu_*\}$ is a singleton, and μ_* is compactly supported, and almost surely,

$$\mu_N\underset{N\to\infty}{\longrightarrow}\mu_*,$$

and moreover μ_* is characterized by the Lagrange conditions

$$U_{\mu_*}+V=C_*\text{ on }\mathrm{supp}(\mu_*)\text{ and }\geq C_*\text{ outside}$$

In the Coulomb case the constant C_* is known as the modified Robin constant. The support constraint in the Lagrange conditions make difficult the analysis of the Riesz case beyond the Coulomb case, due to the non-local nature of the fractional Laplacian. Finally, let us mention that it is shown in [CGZ] using the Lagrange conditions that one can construct V from μ_* if μ_* is smooth enough, and this gives a positive answer to the physical control problem mentioned before.

In the Coulomb case, we have $\Delta U_\mu=c_d\mu$, and an integration parts gives

$$\mathcal{I}(\mu)-\int V\,d\mu=\frac{1}{2}\int U_\mu\,d\mu=\frac{1}{2}\int U_\mu\,\Delta U_\mu\,dx=\frac{1}{2}\int|\nabla U_\mu|^2\,dx.$$

The right hand side is the integral of the squared norm of the gradient ∇U_μ of the electrostatic potential U_μ generated by μ, in other words the "squared-field" ("carré du champ" in French).

5.6. **Related problems.** In the Coulomb case, and when V is radial, then μ_* is supported in a ring and one can compute its density explicitly thanks to the Gauss averaging principle, which states that the electrostatic potential generated by a distribution of charges in a compact set is, outside the compact set, equal to the electrostatic potential generated by a unique charge at the origin, see [LG]. In particular, in the Coulomb case, and when V is the squared norm, then μ_* is the uniform law on a ball of $\mathbb{R}^d$, generalizing the circular law of random matrices.

Beyond the Coulomb case, even in the Riesz case, no Gauss averaging principle is available, and no one knows how to compute μ_*. Even in the Coulomb case, the computation of μ_* is a problem if V is not rotationally invariant. See Saff and Dragnev [DS], and Bleher and Kuijlaars [BK].

When V is weakly confining, then the LDP may be still available, but μ_* is no longer compactly supported. This was checked in dimension $d=2$ with the Coulomb potential by Hardy in [H], by using a compactification (stereographic) which can be probably used in arbitrary dimensions.

It is quite natural to ask about algorithms (exact or approximate) to simulated the law P_N. The Coulomb case in dimension $d=2$ is determinantal and this rigid structure allows exact algorithms [HKPV1]. Beyond this special situation, one may run an Euler-Langevin MCMC approximate

...	...
1660	Newton
1760	Coulomb, Euler
1800	Gauss
1820	Carnot
1850	Helmholtz
1860	Clausius
1870	Boltzmann, Gibbs, Maxwell
1900	Markov, Perron and Frobenius
1930	Fisher, Kolmogorov, Vlasov
1940	de Bruijn, Kac, von Neumann, Shannon
1950	Linnik, Kesten, Kullback, Sanov, Stam, Wigner
1960	Ginibre, McKean, Nelson
1970	Cercignani, Girko, Gross
1980	Bakry and Émery, McKay, Lions, Varadhan, Voiculescu
2000	Perelman, Tao and Vu, Villani
...	...

TABLE 3. The arrow of time and some of the main actors mentioned in the text. One may have in mind the Stigler law: "*No scientific discovery is named after its original discoverer.*", which is attributed by Stephen Stigler to Robert K. Merton.

algorithm using the McKean-Vlasov system of particles. How to do it smartly? Can we do better? Speaking about the McKean-Vlasov system of particles, one may ask about its behavior when t and/or N are large. How it depends on the initial condition, at which speed it converges, do we have a propagation of chaos, does the empirical measure converge to the expected PDE, is it a problem to have singular repulsion? Can we imagine kinetic versions in connection with recent works on Vlasov-Poisson equations for instance [ACD+]? Similar questions are still open for models with attractive interaction such as the Keller-Segel model.

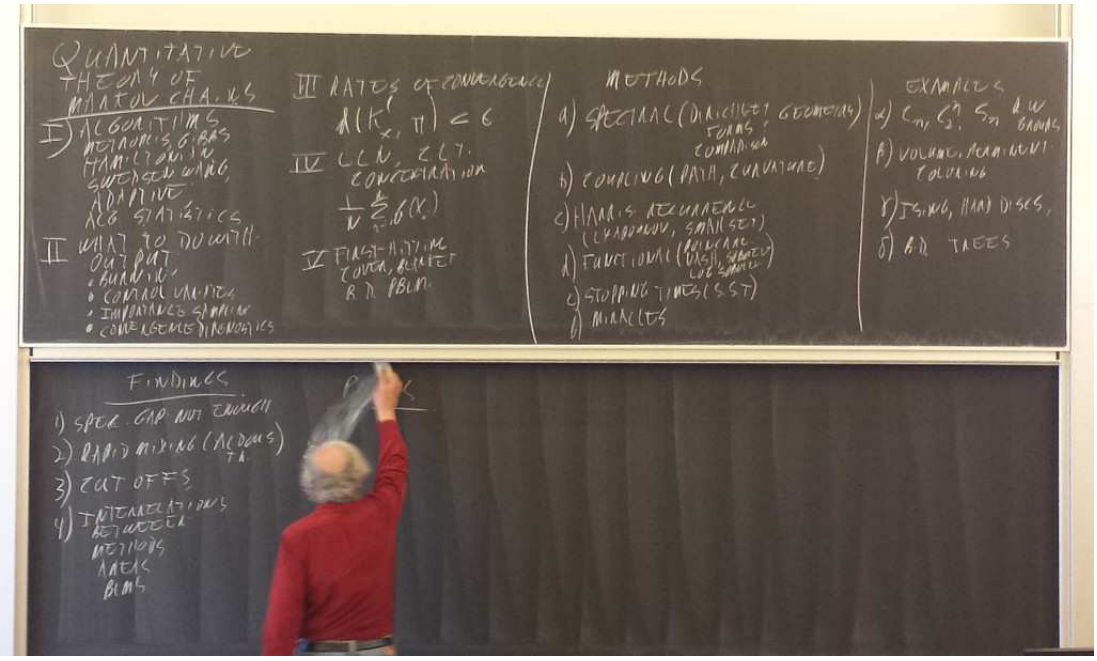
Beyond the first order global analysis, one may ask about the behavior of $\mu_N - \mu_*$. This may lead to central limit theorems in which the speed may depend on the regularity of the test function. Some answers are already available in dimension $d = 2$ in the Coulomb case by Ameur, Hedenmalm, and Makarov [AHM]. Another type of second order analysis is studied by Serfaty and her collaborators [Ser1, Ser2]. Similar infinite dimensional mean-field models are studied by Lewin and Rougerie in relation with the Hartree model for Bose-Einstein condensates [Lew].

Such interacting particle systems with Coulomb repulsion have inspired similar models in complex geometry, in which the Laplacian is replaced by the Monge-Ampere equation. The large deviations approach remains efficient in this context, and was developed by Berman [Ber].

The limit in the first order global asymptotics depend on V and W, and is thus non universal. In the case of β-ensembles of random matrix theory (particles confined on the real line $\mathbb{R}$ with Coulomb repulsion of dimension 2), the asymptotics of the local statistics are universal, and this was the subject of many contributions, with for instance the works of Ramirez, Rider, and Virág [RRV], Erdős, Schlein, and Yau [ESY], Bourgade, Erdős, and Yau [BEY], and Bekerman, Figalli, and Guionnet [BFG]. Little is known in higher dimensions or with other interaction potential W. The particle of largest norm has Tracy-Widom fluctuation in β-ensembles, and here again, little is known in higher dimensions or with other interactions, see [CP] for instance and references therein.

5.7. **Caveat.** The bibliography provided in this text is informative but incomplete. We emphasize that this bibliography should not be taken as an exhaustive reference on the history of the subject.

5.8. **Acknowledgments.** The author would like to thank Laurent Miclo and Persi Diaconis for their invitation, and François Bolley for his useful remarks on an early draft of the text. This version benefited from the constructive comments of two anonymous reviewers.



Persi Diaconis during his talk. Institut de Mathématiques de Toulouse, March 24, 2014.

References

[ABC+] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, with a preface by D. Bakry and M. Ledoux, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses 10, Société Mathématique de France (SMF), (2000) xvi+217p. MR-1845806 9, 12

[ABBN1] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, *On the rate of convergence in the entropic central limit theorem*, Probab. Theory Related Fields **129** (2004), no. 3, 381–390. MR-2128238 12

[ABBN2] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, *Solution of Shannon's problem on the monotonicity of entropy*, J. Amer. Math. Soc. **17** (2004), no. 4, 975–982 (electronic). MR-2083473 12

[ACD+] A. Arnold, J. A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P. A. Markowich, G. Toscani, C. Villani, *Entropies and equilibria of many-particle systems: an essay on recent research.* Monatsh. Math. 142 (2004), no. 1-2, 35–43. MR-2065020 25

[AGZ] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR-2760897 12, 13, 17, 18, 19, 20, 21

[AHM] Y. Ameur, H. Hedenmalm, and N. Makarov, *Fluctuations of eigenvalues of random normal matrices*, Duke Math. J. **159** (2011), no. 1, 31–81. MR-2817648 25

[B] F. Baudoin, *Bakry-Émery meet Villani*, preprint arXiv:1308.4938 9

[BAZ] G. Ben Arous and O. Zeitouni, *Large deviations from the circular law*, ESAIM Probab. Statist. **2** (1998), 123–134 (electronic). MR-1660943 20

[BBB] L. Bessières, G. Besson, and M. Boileau, *La preuve de la conjecture de Poincaré d'après G. Perelman*, Images des math., CNRS (2006) `http://images.math.cnrs.fr/La-preuve-de-la-conjecture-de.html` 10

[BC1] Ch. Bordenave and P. Caputo, *Large deviations of empirical neighborhood distribution in sparse random graphs* preprint arXiv:1308.5725 17

[BC2] Ch. Bordenave and D. Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89. MR-2908617 21

[BCC1] Ch. Bordenave, D. Chafaï, and P. Caputo, *Circular law theorem for random Markov matrices*, Probability Theory and Related Fields, 152:3-4 (2012), 751–779. MR-2892961 2

[BCC2] Ch. Bordenave, D. Chafaï, and P. Caputo, *Spectrum of Markov generators on sparse random graphs.* Comm. Pure Appl. Math. 67 (2014), no. 4, 621–669. MR-3168123 2

[BFG] F. Bekerman, A. Figalli, A. Guionnet, *Transport maps for Beta-matrix models and Universality*, preprint arXiv:1311.2315 25

[BG] Th. Bodineau and A. Guionnet, *About the stationary states of vortex systems*, Ann. Inst. H. Poincaré Probab. Statist. **35** (1999), no. 2, 205–237. MR-1678526 23

[BGL] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2014. 9, 12

[BK] P. M. Bleher and A. B. J. Kuijlaars, *Orthogonal polynomials in the normal matrix model with a cubic potential*, Adv. Math. **230** (2012), no. 3, 1272–1321. MR-2921180 24

[BS] Z.-D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR-2567175 21

[Ber] R. J. Berman, *Determinantal point processes and fermions on complex manifolds: large deviations and bosonization*, preprint arXiv:0812.4224, December 2008. 22, 25

[Bia] Ph. Biane, *Free probability for probabilists*, Quantum probability communications, Vol. XI (Grenoble, 1998), QP-PQ, XI, World Sci. Publ., River Edge, NJ, 2003, pp. 55–71. MR-2032363 12, 15

[BEY] P. Bourgade, L. Erdős, H.-T. Yau. *Bulk universality of general β-ensembles with non-convex potential*. J. Math. Phys. 53, 9, 095221, 19pp, 2012. 25

[Cha1] D. Chafaï, *Aspects of large random Markov kernels*. Stochastics 81 (2009), no. 3-4, 415–429. MR-2549497 2

[Cha2] D. Chafaï, *Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue*. ESAIM Probab. Stat. 10 (2006), 317–339. MR-2247924 10

[Cha3] D. Chafaï, *Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities*, J. Math. Kyoto Univ. 44 (2004), no. 2, 325–363. MR-MR2081075 9

[Cro] W. Cropper, *Great physicists: The life and times of leading physicists from Galileo to Hawking*, Oxford University Press, 2001. MR-1881344 6

[CDPP] P. Caputo, P. Dai Pra, and G. Posta, *Convex entropy decay via the Bochner–Bakry–Émery approach*, Ann. Inst. H. Poincaré Probab. Statist. Volume 45, Number 3 (2009), 589–886 10

[CGZ] D. Chafaï, N. Gozlan, and P.-A. Zitt, *First order global asymptotics for confined particles with singular pair repulsion*, preprint arXiv:1304.7569 to appear in The Annals of Applied Probability 2, 21, 23, 24

[CJ] D. Chafaï and A. Joulin, *Intertwining and commutation relations for birth-death processes*. Bernoulli 19 (2013), no. 5A, 1855–1879. MR-3129037 10

[CLMP] E. Caglioti, P.-L. Lions, C. Marchioro, and M. Pulvirenti, *A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description*, Comm. Math. Phys. **143** (1992), no. 3, 501–525. MR-1145596 and *II*, Comm. Math. Phys. **174** (1995), no. 2, 229–260. MR-1362165 23

[CP] D. Chafaï and S. Péché, *A note on the second order universality at the edge of Coulomb gases on the plane*, preprint arXiv:1310.0727, to appear in Journal of Statistical Physics 25

[CT] Th. M. Cover and J. A. Thomas, *Elements of information theory*. Second edition. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006. xxiv+748 pp. MR-2239987 11

[DMV] L. Desvillettes, C. Mouhot, C. Villani, *Celebrating Cercignani's conjecture for the Boltzmann equation*. Kinet. Relat. Models 4 (2011), no. 1, 277–294. MR-2765747 10

[DS] P. D. Dragnev and E. B. Saff, *Riesz spherical potentials with external fields and minimal energy points separation*, Potential Anal. **26** (2007), no. 2, 139–162. MR-2276529 22, 24

[DSC] P. Diaconis, and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. 6 (1996), no. 3, 695–750. MR-1410112 10

[DZ] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Corrected reprint of the second (1998) edition. Stochastic Modelling and Applied Probability, 38. Springer-Verlag, Berlin, 2010. xvi+396 pp. MR-2571413 5, 23

[ESY] L. Erdős, B. Schlein, H.-T. Yau. *Universality of random matrices and local relaxation flow*. Invent. Math. 185, 1, 75–119, 2011. MR-2810797 25

[Fer] E. Fermi, *Thermodynamics*, Dover, 1956, reprint of the 1936 original version. 6

[For] P. J. Forrester, *Log-gases and random matrices*, London Mathematical Society Monographs Series **34**, Princeton University Press, 2010, xiv+791 MR-2641363 19

[Fro] O. Frostman, *Potentiel d'Équilibre et Capacité des Ensembles*, Ph.D. thesis, Faculté des sciences de Lund, 1935. 22

[Gin] J. Ginibre. *Statistical ensembles of complex, quaternion, and real matrices.* J. Mathematical Phys. 6:440–449, 1965. 19

[Gro] L. Gross. *Logarithmic Sobolev inequalities*, Amer. J. Math. 97 (1975), no. 4, 1061–1083. MR-420249 9

[H] A. Hardy, *A note on large deviations for 2D Coulomb gas with weakly confining potential*, Electron. Commun. Probab. **17** (2012), no. 19, 12. MR-2926763 20, 21, 24

[HKPV1] B. J. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Determinantal processes and independence*, Probab. Surv. **3** (2006), 206–229. MR-2216966 24

[HKPV2] B. J. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian analytic functions and determinantal point processes*, University Lecture Series, vol. 51, American Mathematical Society, Providence, RI, 2009. MR-2552864 19

[HO] A. Hora and N. Obata, *Quantum probability and spectral analysis of graphs*, Theoretical and Mathematical Physics, Springer, Berlin, 2007, With a foreword by Luigi Accardi. MR-2316893 15, 16

[HP] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000. MR-1746976 21

[J] O. Johnson, *Information theory and the central limit theorem*, Imperial College Press, London, 2004. xiv+209 pp. MR-2109042 12

[JA] V. Jog and V. Anantharam, *Convex Relative Entropy Decay in Markov Chains*, to appear in the proceedings of the 48th Annual Conference on Information Sciences and Systems, CISS-2014, Princeton University, Princeton, NJ, March 19–21, 2014. 10

[KL] S. Kullback, S. and R.A. Leibler, R. A, *On information and sufficiency*. Ann. Math. Statistics 22, (1951). 79–86. 4

[KSo] B. Khoruzhenko and H.-J. Sommers, *Non-Hermitian ensembles*. The Oxford handbook of random matrix theory, 376–397, Oxford Univ. Press, Oxford, 2011. 19

[KSpo] M. K.-H. Kiessling and H. Spohn, *A note on the eigenvalue density of random matrices*, Comm. Math. Phys. **199** (1999), no. 3, 683–695. MR-1669669 23

[Ka] Mark Kac, *Probability and related topics in physical sciences*. With special lectures by G. E. Uhlenbeck, A. R. Hibbs, and B. van der Pol. Lectures in Applied Mathematics. Proceedings of the Summer Seminar, Boulder, Colo., 1957, Vol. I Interscience Publishers, London-New York 1959 xiii+266 pp. MR-0102849 3, 10

[Kes] H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354. MR-0109367 16

[Kie] M. K.-H. Kiessling, *Statistical mechanics of classical particles with logarithmic interactions*, Comm. Pure Appl. Math. **46** (1993), no. 1, 27–56. MR-1193342 23

[Ku] S. Kullback, *Information theory and statistics*, Reprint of the second (1968) edition. Dover Publications, 1997. xvi+399 pp. 4

[LG] A. López García, *Greedy energy points with external fields*, Recent trends in orthogonal polynomials and approximation theory, Contemp. Math., vol. 507, Amer. Math. Soc., Providence, RI, 2010, 189–207. MR-2647570 24

[LV] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) 169 (2009), no. 3, 903–991. MR-2480619 9

[Lew] M. Lewin, *Limite de champ moyen et condensation de Bose-Einstein*, Gazette des Mathématiciens, no. 139, 35–49 (2014). 25

[Li] X.-D. Li, *From the Boltzmann H-theorem to Perelman's W-entropy formula for the Ricci flow*, preprint arXiv:1303.5193 10

[Lie] E. H. Lieb, *Proof of an entropy conjecture of Wehrl*, Comm. Math. Phys. 62 (1978) 35–41. MR-0506364 11, 17

[Lig] Th. M. Liggett, *Interacting particle systems*, Classics in Mathematics, Springer-Verlag, Berlin, 2005, Reprint of the 1985 original. MR-2108619 10

[Lin] Ju. V. Linnik, *An information-theoretic proof of the central limit theorem with Lindeberg conditions*, Theor. Probability Appl. **4** (1959), 288–299. MR-0124081 11

[M] L. Miclo, *Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies*, Séminaire de Probabilités, XXXI, 136–167, Lecture Notes in Math., 1655, Springer, Berlin, 1997. 10

[McK] B. D. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra and its Applications **40** (1981), 203–216. 16

[MS] J. Messer and H. Spohn, *Statistical mechanics of the isothermal Lane-Emden equation*, J. Statist. Phys. **29** (1982), no. 3, 561–578. MR-704588 23

[MT] R. R. Montenegro and P. Tetali, *Mathematical Aspects of Mixing Times in Markov Chains*, Now Publishers Inc, 2006, pp. 121 10

[N] E. Nelson, *A quartic interaction in two dimensions*, Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965), 69–73, M.I.T. Press, Cambridge, Mass., 1966 MR-0210416 9

[O] Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. 256 (2009), no. 3, 810–864. MR-2484937 9, 10

[PH] D. Petz and F. Hiai, *Logarithmic energy as an entropy functional*, Advances in differential equations and mathematical physics (Atlanta, GA, 1997), Contemp. Math., vol. 217, Amer. Math. Soc., Providence, RI, 1998, pp. 205–221. MR-1606719 20

[RRV] J. A. Ramírez, B. Rider, B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*. J. Amer. Math. Soc. 24 (2011), no. 4, 919–944. MR-2813333 25

[vRS] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. 58 (2005), no. 7, 923–940. MR-2142879 9

[ST] E. B. Saff and V. Totik, *Logarithmic potentials with external fields*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 316, Springer-Verlag, Berlin, 1997, Appendix B by Thomas Bloom. MR-1485778 17, 21

[SW] C. E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill., 1949. MR-0032134 11

[SC] L. Saloff-Coste, *Lectures on finite Markov chains. Lectures on probability theory and statistics (Saint-Flour, 1996)*, 301–413, Lecture Notes in Math., 1665, Springer, Berlin, 1997. MR-1490046 10

[Sen] E. Seneta, *Markov and the Birth of Chain Dependence*, International Statistical Review **64** (1966), no. 3, 255–263. 2

[Ser1] S. Serfaty, *Coulomb Gases and Ginzburg-Landau Vortices*, preprint arXiv:1403.6860 25

[Ser2] S. Serfaty, *Ginzburg-Landau vortices, Coulomb gases, and renormalized energies*, J. Stat. Phys. 154 (2014), no. 3, 660–680. MR-3163544 25

[Sh1] D. Shlyakhtenko, *A free analogue of Shannon's problem on monotonicity of entropy*, Adv. Math. 208 (2007), no. 2, 824–833. MR-2304337 17

[Sh2] D. Shlyakhtenko, *Shannon's monotonicity problem for free and classical entropy*, Proc. Natl. Acad. Sci. USA **104** (2007), no. 39, 15254–15258 (electronic), With an appendix by Hanne Schultz. MR-2346565 12, 17

[Sta] A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), 101–112. MR-21 11

[T] T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. MR-2906465 21

[TV] T. Tao and V. Vu. *Random matrices: universality of ESDs and the circular law.* Ann. Probab. 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur. 2

[TM] M. Tribus and E. C. McIrvine, *Energy and information*, Scientific American **225** (1971), no. 3, 179–188. 11

[VDN] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992, A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. MR-1217253 12

[Vil1] C. Villani, *H-Theorem and beyond: Boltzmann's entropy in today's mathematics*, Boltzmann's legacy, ESI Lect. Math. Phys., Eur. Math. Soc., Zürich, 2008, pp. 129–143. MR-2509760 7

[Vil2] C. Villani, *Irreversibility and entropy*, Time, Prog. Math. Phys. 63, 19–79, Birkhäuser/Springer, 2013 MR-3076094 6

[Voi] D.-V. Voiculescu, Free entropy, Bull. London Math. Soc. 34 (3) (2002) 257–278. MR-1887698 12, 17

[W] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices.* Ann. of Math. (2) 67 1958 325–327. MR-0095527 16, 18

[Z] M. Zinsmeister, *Thermodynamic formalism and holomorphic dynamical systems.* Translated from the 1996 French original by C. Greg Anderson. SMF/AMS Texts and Monographs, 2. American Mathematical Society, Providence, RI; Société Mathématique de France, Paris, 2000. x+82 pp. MR-1724307 6

CEREMADE, UMR 7534, Université Paris Dauphine, PSL, IUF, France.
E-mail address: mailto:djalil(at)chafai.net
URL: http://djalil.chafai.net/