



**HAL**  
open science

# ON KAC'S CHAOS AND RELATED PROBLEMS

Maxime Hauray, Stéphane Mischler

► **To cite this version:**

Maxime Hauray, Stéphane Mischler. ON KAC'S CHAOS AND RELATED PROBLEMS. 2012. hal-00682782v1

**HAL Id: hal-00682782**

**<https://hal.science/hal-00682782v1>**

Preprint submitted on 26 Mar 2012 (v1), last revised 29 Mar 2013 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ON KAC'S CHAOS AND RELATED PROBLEMS

M. HAURAY AND S. MISCHLER

March 26, 2012

ABSTRACT. This paper is devoted to establish quantitative and qualitative estimates related to the notion of chaos as firstly formulated by M. Kac [30] in his study of mean-field limit for systems of  $N$  undistinguishable particles as  $N \rightarrow \infty$ .

First, we quantitatively liken three usual measures of *Kac's chaos*, some involving the all  $N$  variables, other involving a finite fixed number of variables. The cornerstone of the proof is a new representation of the Monge-Kantorovich-Wasserstein (MKW) distance for symmetric  $N$ -particle probabilities in terms of the distance between the law of the associated empirical measures on the one hand, and a new estimate on some MKW distance on probability spaces endowed with a suitable Hilbert norm taking advantage of the associated good algebraic structure.

Next, we define the notion of *entropy chaos* and *Fisher information chaos* in a similar way as defined by Carlen et al [13]. We show that *Fisher information chaos* is stronger than *entropy chaos*, which in turn is stronger than *Kac's chaos*. More importantly, with the help of the HWI inequality of Otto-Villani, we establish a quantitative estimate between these quantities, which in particular asserts that *Kac's chaos* plus *Fisher information bound* implies *entropy chaos*.

We then extend the above quantitative and qualitative results about chaos in two other frameworks. We first extend it to the framework of probabilities with support on the *Kac's spheres*, revisiting [13] and giving a possible answer to [13, Open problem 11]. Additionally to the above mentioned tool, we use and prove an optimal rate local CLT in  $L^\infty$  norm for distributions with finite 6-th moment and finite  $L^p$  norm,  $p > 1$ .

Last, we investigate how our techniques can be used in the context of probabilities mixtures as introduced by De Finetti, Hewitt and Savage and general sequence of probability densities of  $N$  undistinguishable particles as  $N \rightarrow \infty$ , without assuming chaos.

**Keywords:** Kac's chaos, Monge-Kantorovich-Wasserstein distance, Entropy chaos, Fisher information chaos, CLT with optimal rate, probability mixtures, De Finetti, Hewitt and Savage theorem, Mean-field limit, quantitative chaos, qualitative chaos, entropy.

## AMS Subject Classification:

26D15 Inequalities for sums, series and integrals,

94A17 Measures of information, entropy,

60F05 Central limit and other weak theorems,

82C40 Kinetic theory of gases.

43A15  $L^p$ -spaces and other function spaces on groups, semigroups

52A40 Inequalities and extremum problems

## CONTENTS

1. Introduction and main results	2
2. Kac's chaos	8
3. Entropy chaos and Fisher information chaos	23

4. Probabilities on the “Kac’s sphere”	31
5. On mixtures according to De Finetti, Hewitt and Savage	49
References	63

## 1. INTRODUCTION AND MAIN RESULTS

The Kac’s notion of chaos rigorously formalizes the intuitive idea for a family of stochastic valued vectors with  $N$  coordinates to have asymptotically independent coordinates as  $N$  goes to infinity. We refer to [54] for an introduction to that topics from a probabilistic point of view, as well as to [42] for a short survey.

**Definition 1.1.** [30, section 3] *Consider  $E \subset \mathbb{R}^d$ ,  $f \in \mathbf{P}(E)$  a probability on  $E$  and  $G^N \in \mathbf{P}_{sym}(E^N)$  a sequence of probabilities on  $E^N$ ,  $N \geq 1$ , which are invariant under coordinates permutations. We say that  $(G^N)$  is  $f$ -Kac’s chaotic (or has the “Boltzmann property”) if*

$$(1.1) \quad \forall j \geq 1, \quad G_j^N \rightharpoonup f^{\otimes j} \quad \text{weakly in } \mathbf{P}(E^j) \quad \text{as } N \rightarrow \infty,$$

where  $G_j^N$  stands for the  $j$ -th marginal of  $G^N$  defined by

$$G_j^N := \int_{E^{N-j}} G^N dx_{j+1} \dots dx_N.$$

Interacting  $N$ -indistinguishable particle systems are naturally described by exchangeable random variables (which corresponds to the fact that their associated probability law are symmetric, i.e. invariant under coordinates permutations) but they are not described by random variables with independent coordinates (which corresponds to the fact that their associated probability laws are tensor products) except for situation with no interaction! Kac’s chaos is therefore a well adapted concept to formulate and investigate the infinite number of particles limit  $N \rightarrow \infty$  for these systems as it has been illustrated by many works since the seminal article by Kac [30]. Using that definition of chaos, it is shown in [30, 38, 37, 25, 43] that if  $f(t)$  evolves according to the nonlinear space homogeneous Boltzmann equation,  $G^N(t)$  evolves according to the linear Master/Kolmogorov equation associated to the stochastic Kac-Boltzmann jumps (collisions) process and  $G^N(0)$  is  $f(0)$ -chaotic then for any time  $t > 0$  the sequence  $G^N(t)$  is also  $f(t)$ -chaotic: propagation of chaos holds for that model. As it is explained in the latest reference and using the uniqueness of statistical solutions proved by [1], some of these propagation of chaos results can be seen as a simple illustration of the “BBGKY hierarchy method” whose most famous success is the Lanford’s proof of the “Boltzmann-Grad limit”, see [32].

In order to investigate quantitative version of Kac’s chaos, the above weak convergence in (1.1) can be formulated in terms of the Monge-Kantorovich-Wasserstein (MKW) transportation distance between  $G_j^N$  and  $f^{\otimes j}$ . More precisely, given  $d_E$  a bounded distance on  $E$ , we define the correctly normalized distance  $d_{E^j}$  on  $E^j$ ,  $j \in \mathbb{N}^*$ , by setting

$$(1.2) \quad \forall X = (x_1, \dots, x_j), Y = (y_1, \dots, y_j) \in E^j \quad d_{E^j}(X, Y) := \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i),$$

and then we define  $W_1$  (without specifying the dependence on  $j$ ) the associated MKW distance in  $\mathbf{P}(E^j)$  (see the definition (2.2) below). With the notations of Definition 1.1,

$G^N$  is  $f$ -Kac's chaotic if, and only if,

$$\forall j \geq 1, \quad \Omega_j(G^N; f) := W_1(G_j^N, f^{\otimes j}) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Let us introduce now another formulation of Kac's chaos which we firstly formulate in a probabilistic language. For any  $X = (x_1, \dots, x_N) \in E^N$ , we define the associated empirical measure

$$(1.3) \quad \mu_X^N(dy) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dy) \in \mathbf{P}(E).$$

We say that an exchangeable  $E^N$ -valued random vector  $\mathcal{X}^N$  is  $f$ -chaotic if the associated  $\mathbf{P}(E)$ -valued random variable  $\mu_{\mathcal{X}^N}^N$  converges to the deterministic random variable  $f$  in law in  $\mathbf{P}(E)$ :

$$(1.4) \quad \mu_{\mathcal{X}^N}^N \Rightarrow f \quad \text{in law as } N \rightarrow \infty.$$

In the framework of Definition 1.1, the convergence (1.4) can be equivalently formulated in the following way. Introducing  $G^N := \mathcal{L}(\mathcal{X}^N)$  the law of  $\mathcal{X}^N$ , the exchangeability hypothesis means that  $G^N \in \mathbf{P}_{sym}(E^N)$ . Next the law  $\hat{G}^N := \mathcal{L}(\mu_{\mathcal{X}^N}^N)$  of  $\mu_{\mathcal{X}^N}^N$  is nothing but the (unique) measure  $\hat{G}^N \in \mathbf{P}(\mathbf{P}(E))$  such that

$$\langle \hat{G}^N, \Phi \rangle = \int_{E^N} \Phi(\mu_X^N) G^N(dX) \quad \forall \Phi \in C_b(\mathbf{P}(E)).$$

Then the convergence (1.4) just means that

$$(1.5) \quad \hat{G}^N \rightarrow \delta_f \quad \text{weakly in } \mathbf{P}(\mathbf{P}(E)) \text{ as } N \rightarrow \infty,$$

where this definition does not refer anymore to the random variables  $\mathcal{X}^N$  or  $\mu_{\mathcal{X}^N}^N$ . It is well known (see for instance [26, section 4], [30, 55, 53] and [54, Proposition 2.2]) that for a sequence  $(G^N)$  of  $\mathbf{P}_{sym}(E^N)$  and a probability  $f \in \mathbf{P}(E)$  the three following assertions are equivalent:

- (i) convergence (1.1) holds for any  $j \geq 1$ ;
- (ii) convergence (1.1) holds for some  $j \geq 2$ ;
- (iii) convergence (1.5) holds;

so that in particular (1.4) or (1.5) are indeed equivalent formulations of Kac's chaos. The chaos formulation (ii) has been used since [30], while the chaos formulation (iii) is widely used in the works by Sznitman [52], see also [53, 40, 46], where the chaos property is established by proving that the "empirical process"  $\mu_{\mathcal{X}^N}^N$  converges to a limit process with values in  $\mathbf{P}(E)$ , solution to a nonlinear martingal problem associated to the mean-field limit equation. Formulation (1.5) is also well adapted for proving quantitative propagation of chaos for deterministic dynamics associated to the Vlasov equation with regular interaction force [21] as well as singular interaction force [28, 27]. Let us briefly explain this point now, see also [42, section 1.1]. On the one hand, introducing the MKW transport distance  $\mathcal{W}_1 := \mathcal{W}_{W_1}$  on  $\mathbf{P}(\mathbf{P}(E))$  based on the MKW distance  $W_1$  on  $\mathbf{P}(E)$ , (see definition (2.6) below), the weak convergence (1.5) is nothing but the fact that

$$\Omega_0(G^N; f) := \mathcal{W}_1(\hat{G}^N, \delta_f) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

On the other hand, for the Vlasov equation with smooth and bounded force term, it is proved in [21] that

$$(1.6) \quad \forall T > 0, \forall t \in [0, T] \quad W_1(\mu_{\mathcal{X}_t^N}^N, f_t) \leq C_T W_1(\mu_{\mathcal{X}_0^N}^N, f_0),$$

where  $f_t \in \mathbf{P}(E)$  is the solution to the Vlasov equation with initial datum  $f_0$  and  $\mathcal{X}_t^N \in E^N$  is the solution to the associated system of edo with initial datum  $\mathcal{X}_0^N$ . Inequality (1.6) is a consequence of the fact that  $t \mapsto \mu_{\mathcal{X}_t^N}^N$  solves the Vlasov equation and that a local  $W_1$  stability result holds for such an equation. When  $\mathcal{X}_0$  is distributed according to an initial density  $G_0^N \in \mathbf{P}_{sym}(E^N)$  we may show that  $\mathcal{X}_t$  is distributed according to  $G_t^N \in \mathbf{P}_{sym}(E^N)$  obtained as the transported measure along the flow associated to the above mentioned system of edo or equivalently  $G_t^N$  is the solution to the associated Liouville equation. Taking the expectation in both sides of (1.6), we get

$$\begin{aligned} \int_{E^N} W_1(\mu_Y^N, f_t) G_t^N(dY) &= \mathbb{E}[W_1(\mu_{\mathcal{X}_t^N}^N, f_t)] \\ &\leq C_T \mathbb{E}[W_1(\mu_{\mathcal{X}_0^N}^N, f_0)] = C_T \int_{E^N} W_1(\mu_Y^N, f_0) G_0^N(dY), \end{aligned}$$

for any  $t \in [0, T]$ . We conclude with the following quantitative chaos propagation estimate

$$\forall t \in [0, T] \quad \Omega_0(G_t^N; f_t) \leq C_T \Omega_0(G_0^N; f_0).$$

It is worth mentioning that partially inspired from [26], it is shown in [45, 43] a similar inequality as above for more general models including drift, diffusion and collisional interactions where however the estimate may mix several chaos quantification quantities as  $\Omega_0$  and  $\Omega_2$  for instance.

There exists at least one more way to guaranty chaoticity which is very popular because that chaos formulation naturally appears in the probabilistic coupling technique, see [54], as well as [36, 9, 8] and the references therein for more recent developments. Thanks to the coupling technique we typically may show that an exchangeable  $E^N$ -valued random vector  $\mathcal{X}^N$  satisfies

$$\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N |\mathcal{X}_i^N - \mathcal{Y}_i^N|\right) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

for some  $E^N$ -valued random vector  $\mathcal{Y}^N$  with independent coordinates. Denoting by  $G^N \in \mathbf{P}_{sym}(E^N)$  the law of  $\mathcal{X}^N$ ,  $f$  the law of one coordinate  $\mathcal{Y}_i^N$ , and  $W_1$  the MKW transport distance on  $\mathbf{P}(E^N)$  based on the correctly normalized distance  $d_{E^N}$  in  $E^N$  defined by (1.2), the above convergence readily implies

$$(1.7) \quad \Omega_N(G^N; f) := W_1(G^N, f^{\otimes N}) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

which in turn guaranties that  $(G^N)$  is  $f$ -chaotic. It is generally agreed that the convergence (1.7) is a strong version of chaos, maybe because it involves the all  $N$  variables, while the Kac's original definition only involves a finite fixed number of variables.

The first natural question we consider is about the equivalence between these definitions of chaos, and more precisely the possibility to liken them in a quantitative way. The following result give a positive answer, we also refer to Theorem 1.2 - part 1, 2, 3 & 4 in section 2 where more precise versions are stated.

**Theorem 1.2** (Equivalence of measure for Kac's chaos). *For any  $k > 0$  there exist an exponent  $\gamma = \gamma(d, k) > 0$  and a constants  $C = C(d, k) \in (0, \infty)$  such that for any  $f \in \mathbf{P}(E)$ , any  $G^N \in \mathbf{P}_{sym}(E^N)$ ,  $N \geq 1$ , and any  $j, \ell \in \{0, \dots, N\}$ ,  $\ell \neq 1$ , there holds*

$$\Omega_j(G^N; f) \leq C \mathcal{M}_k^{1-2\gamma} \left( \Omega_\ell(G^N; f) + \frac{1}{N} \right)^\gamma,$$

where  $\mathcal{M}_k = M_k(f) + M_k(G_1^N)$  is the sum of the moments of order  $k$  of  $f$  and  $G_1^N$ .

It is worth emphasizing that the above inequality is definitively false in general for  $\ell = 1$ . The first outcome of our theorem is that it shows that regardless of the rate, the propagation of chaos results obtained by the coupling method is of the same nature as the propagation of chaos result obtained by the “BBGKY hierarchy method” and the “empirical measures method”.

The proof of Theorem 1.2 will be presented in section 2. Let us briefly explain the strategy. First, the fact that we may control  $\Omega_j$  by  $\Omega_\ell$  for  $1 \leq j \leq \ell \leq N$  is classical and quite easy. Next, we will establish an estimate of  $\Omega_0$  by  $\Omega_2$  following an idea introduced in [43]: we begin to prove a similar estimate where we replace  $\Omega_0$  by the MKW distance in  $\mathbf{P}(\mathbf{P}(E))$  associated to the  $H^{-s}(\mathbb{R}^d)$  norm,  $s > d/2 + 1$ , on  $\mathbf{P}(E)$  in order to take advantage of the good algebraic structure of that Hilbert norm and then we come back to  $\Omega_0$  thanks to the “uniformly topological equivalence” of metrics in  $\mathbf{P}(E)$  and the Holder inequality. Finally, and that is the other key new result, we compare  $\Omega_0$  and  $\Omega_N$ : that is direct consequence of the following identity

$$\forall F^N, G^N \in \mathbf{P}_{sym}(E^N) \quad W_1(G^N, F^N) = \mathcal{W}_1(\hat{G}^N, \hat{F}^N)$$

applied to  $F^N := f^{\otimes N}$  and a functional version of the law of large numbers.

A somewhat stronger notion of chaos can be formulated in terms of entropy functionals. Such a notion has been explicitly introduced by Carlen, Carvahlo, Loss, Leroux, Villani in [13] (in the context of probabilities with support on the “Kac’s spheres”) but it is reminiscent in the works [31, 4]. We also refer to [51, 41, 10, 11] where the  $N$  particles entropy functional below is widely used in order to identify the possible limits for a system of  $N$  particles as  $N \rightarrow \infty$ . Consider  $E \subset \mathbb{R}^d$  an open set or the adherence of a open space, in order that the gradient of a function may be well defined. For a (smooth and/or decaying enough) probability  $G^N \in \mathbf{P}_{sym}(E^N)$  we define (see also section 3 for the suitable definitions) the Boltzmann’s entropy and the Fisher information by

$$H(G^N) := \frac{1}{N} \int_{E^N} G^N \log G^N dX, \quad I(G^N) := \frac{1}{N} \int_{E^N} \frac{|\nabla G^N|^2}{G^N} dX.$$

It is worth emphasizing that contrarily to the convention adopted in [13, Definition 8] we have put the normalized factor  $1/N$  in the definitions of the entropy and the Fisher information. Moreover we use the same notation for these functionals whatever is the dimension. As a consequence, we have  $H(f^{\otimes N}) = H(f)$  and  $I(f^{\otimes N}) = I(f)$  for any probabilities  $f \in \mathbf{P}(E)$ .

**Definition 1.3.** Consider  $(G^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  such that the  $k$ -th moment  $M_k(F_1^N)$  is bounded,  $k > 0$ , and  $f \in \mathbf{P}(E)$ . We say that

(a)  $(G^N)$  is  $f$ -entropy chaotic (or  $f$ -chaotic in the sense of the Boltzmann’s entropy) if

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad H(G^N) \rightarrow H(f), \quad H(f) < \infty;$$

(b)  $(G^N)$  is  $f$ -Fisher information chaotic (or  $f$ -chaotic in the sense of the Fisher information) if

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad I(G^N) \rightarrow I(f), \quad I(f) < \infty.$$

Our second main result is the following qualitative comparison of the three above notions of chaos convergence.

**Theorem 1.4.** Assume  $E = \mathbb{R}^d$ ,  $d \geq 1$ , or  $E$  is a bi-Lipschitz volume preserving deformation of a convex set of  $\mathbb{R}^d$ ,  $d \geq 1$ . Consider  $(G^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  such that the  $k$ -th moment  $M_k(G_1^N)$  is bounded,  $k > 2$ , and  $f \in \mathbf{P}(E)$ .

In the list of assertions below, each one implies the assertion which follows:

- (i)  $(G^N)$  is  $f$ -Fisher information chaotic;
- (ii)  $(G^N)$  is  $f$ -Kac's chaotic and  $I(G^N)$  is bounded;
- (iii)  $(G^N)$  is  $f$ -entropy chaotic;
- (iv)  $(G^N)$  is  $f$ -Kac's chaotic.

More precisely, the following quantitative estimate of the implication (ii)  $\Rightarrow$  (iii) holds:

$$(1.8) \quad |H(G^N) - H(f)| \leq C_E K \Omega_N(G^N; f)^\gamma,$$

with  $\gamma := 1/2 - 1/k$ ,  $K := \sup_N I(G^N)^{1/2} \sup_N M_k(G_1^N)^{1/k}$  and  $C_E$  is a constant depending on the set  $E$  (one can choose  $C_E = 8$  when  $E = \mathbb{R}^d$ ).

The implication (ii)  $\Rightarrow$  (iii) is the most interesting part and hardest step in the proof of Theorem 1.4. It is based on estimate (1.8) which is a mere consequence of the HWI inequality of Otto and Villani proved in [48] when  $E = \mathbb{R}^d$  together with our equivalence of chaos convergences previously established. We believe that this result gives a better understanding of the different notions of chaos. Other but related notions of entropy chaos is introduced and discussed in [13, 44]. Finally also notice that our definition of entropy chaos ((a) in Definition 1.3) is slightly weaker than the notion of “entropic chaos” introduced in [13, Definition 8]. However, thanks to Theorem 1.4, they are in fact equivalent.

Subsequently, we want to illustrate our approach by generalizing our results in two other frameworks:

- when the probabilities  $G^N$  are supported by the Kac's sphere, as introduced in [30];
- when the limit of  $G^N$  is not necessarily chaotic, as it is the case in [41, 10, 11] for instance.

In section 4, we consider the framework of probabilities with support on the “Kac's sphere”  $\mathcal{KS}_N$  defined by

$$\mathcal{KS}_N := \{V = (v_1, \dots, v_N) \in E^N, v_1^2 + \dots + v_N^2 = N\}, \quad E := \mathbb{R},$$

as firstly introduced by Kac in [30]. Our aim is mainly to revisit the recent work [13] and to develop “quantitative” versions of the chaos analysis. We start with the proof of a quantified “Poincaré Lemma” stating that  $\sigma^N$  the uniform probability measure on  $\mathcal{KS}_N$  is  $\gamma$ -chaotic, with  $\gamma$  the standard gaussian, in the sense that the corresponding quantification of chaos  $\Omega_N(\sigma^N; \gamma)$  converges to 0 with rate  $1/\sqrt{N}$ . Next, we prove that for a large class of probability densities  $f \in \mathbf{P}(E)$  the corresponding sequence of “conditioned (to the Kac's sphere) product measures”  $F^N$  is  $f$ -Kac's chaotic,  $f$ -entropy chaotic as well as  $f$ -Fisher information chaotic, and we exhibit rates for these convergences. In the present context, we have to slightly modify the sense of entropy and Fisher information chaos. First, for  $f \in \mathbf{P}(E)$  smooth enough, we define the usual relative entropy and usual relative Fisher information

$$H(f|\gamma) := \int_E u \log u \gamma \, dv, \quad I(f|\gamma) := \int_E \frac{|\nabla u|^2}{u} \gamma \, dv, \quad u := f/\gamma,$$

and similarly for  $G^N \in \mathbf{P}_{sym}(\mathcal{KS}_N)$ , we define the (normalized) relative entropy and relative Fisher information

$$H(G^N|\sigma^N) := \frac{1}{N} \int_{\mathcal{KS}_N} g^N \log g^N \, d\sigma^N, \quad I(G^N|\sigma^N) := \frac{1}{N} \int_{\mathcal{KS}_N} \frac{|\nabla g^N|^2}{g^N} \, d\sigma^N,$$

where  $g^N := \frac{dG^N}{d\sigma^N}$  stands for the Radon-Nikodym derivative of  $G^N$  with respect to  $\sigma^N$ . Then for  $f \in \mathbf{P}(E)$  and  $G^N \in \mathbf{P}_{sym}(\mathcal{KS}_N)$  such that  $G_1^N \rightharpoonup f$  weakly in  $\mathbf{P}(E)$ , we say that  $(G^N)$  is

(a')  $f$ -entropy chaotic if  $H(G^N|\sigma^N) \rightarrow H(f|\gamma)$ ,  $H(f|\gamma) < \infty$ ;

(b')  $f$ -Fisher information chaotic if  $I(G^N|\sigma^N) \rightarrow I(f|\gamma)$ ,  $I(f|\gamma) < \infty$ .

The main arguments in the above result is a (maybe new)  $L^\infty$  version of the Berry-Esseen theorem, also called local central limit theorem, as well as a careful rewriting and simplification of the proofs of the similar results (given without rate) in Theorems 9, 10, 19, 20, 21 in [13].

We next generalize Theorem 1.4 to the Kac's sphere context. Additionally to the yet mentioned arguments, we use a general version of the HWI inequality proved by Lott and Villani in [35], see also [57, Theorem 30.21], and some Entropy and Fisher inequalities on the Kac's spheres established by Carlen et al. [14] and improved by Barthe et al. [2].

These results are motivated by the question of giving quantified strong version of propagation of chaos for Boltzmann-Kac jump model studied in [43] by Mouhot and the second author, where only quantitative uniform in time Kac's chaos is established. As a matter of fact, K. Carrapatoso in [15] extends the present analysis to the probabilities with support to the *Boltzmann's spheres* and proves a quantitative propagation result of entropy chaos.

Another outcome of our results is that we are able to give the following possible answer to [13, Open problem 11]:

**Theorem 1.5.** *Consider  $(G^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  with support in  $\mathcal{KS}_N$  such that*

$$(1.9) \quad I(G^N|\sigma^N) \leq C,$$

for  $C > 0$ . Also consider  $f \in \mathbf{P}(E)$ ,  $E := \mathbb{R}$ , satisfying  $M_2(f) = 1$  and

$$(1.10) \quad f \geq \exp(-\alpha|v|^k + \beta) \quad \text{on } E,$$

with  $0 < k < 2$ ,  $\alpha > 0$ ,  $\beta \in \mathbb{R}$ . If  $(G^N)$  is  $f$ -Kac's chaotic, then for any fixed  $j \geq 1$ , there holds

$$H(G_j^N|f^{\otimes j}) \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

where  $H(\cdot|\cdot)$  stands for the usual relative entropy functional defined in the flat space  $E^j$ . Remark that the condition on the second moment of  $f$  is useless if  $\sup_N M_{k'}(G_1^N) < +\infty$  for some  $k' > 2$ .

Contrarily to the conditioned tensor product assumption made in [13, Theorem 9] which can be assumed at initial time for the stochastic Kac-Boltzmann process but which is not propagated along time, our assumptions (1.9) and (1.10) in Theorem 1.5, which may seem to be stronger, are in fact more natural since they are propagated along time. We refer to [43, 15] where such problems are studied.

Last in section 5, we investigate how our techniques can be used in the context of probabilities mixtures as introduced by De Finetti, Hewitt and Savage [18, 29] and general sequences of probability densities  $G^N$  of  $N$  undistinguishable particles as  $N \rightarrow \infty$ , without assuming chaos.

In a first step, we give a new proof of De Finetti, Hewitt and Savage theorem which is based on the use of the law of the empirical measure associated to the  $j$  first coordinates like in Diaconis and Freedman's proof [19] or Lions' proof [33], but where the compactness arguments are replaced by an argument of completeness. As a back product, we give a



quantified equivalence of several notions of convergences of sequences of  $\mathbf{P}_{sym}(E^N)$  to its possible mixture limit.

In a second step, we revisit the level 3 entropy and level 3 Fisher information theory for a probabilities mixture as developed since the work by Robinson and Ruelle [51] at least. We give a comprehensive and elementary proof of the fundamental result

$$(1.11) \quad \mathcal{K}(\pi) := \int_{\mathbf{P}(E)} K(\rho) \pi(d\rho) = \lim_{j \rightarrow \infty} \frac{1}{j} \int_{E^j} K(\pi_j)$$

for any probability mixture  $\pi \in \mathbf{P}(\mathbf{P}(E))$ , where  $\pi_j$  stands for the De Finetti, Hewitt and Savage projection of  $\pi$  on the  $j$  first coordinates and  $K$  stands for the entropy or the Fisher information functional. In our last result we establish a rate of convergence for the above limit (1.11) when  $K$  is the entropy functional mainly under a boundedness of the Fisher information hypothesis and we generalize such a quantitative result establishing links between several weak notions of convergence as well as strong (entropy) notion of convergence for sequences of probability densities  $G^N \in \mathbf{P}_{sym}(E^N)$  as  $N \rightarrow \infty$ , without assuming chaos.

**Acknowledgement.** We would like to thank F. Bolley and C. Mouhot for many stimulating discussions about mean field limit and chaos. The second author also would like to acknowledge I. Gentil and C. Villani for discussions about the HWI inequality and P.-L. Lions for discussions about entropy and mollifying tricks in infinite dimension. The second author acknowledges support from the project ANR-MADCOF.

## 2. KAC'S CHAOS

In this section we show the equivalence between several ways to measure Kac's chaos as stated in Theorem 1.2. We start presenting the framework we will deal with in the sequel, and thus making precise the definitions and notations used in the introductory section.

**2.1. Definitions and notations.** In all the sequel, we denote by  $E$  a closed subset of  $\mathbb{R}^d$ ,  $d \geq 1$ , endowed with the usual topology, so that it is a locally compact Polish space. We denote by  $\mathbf{P}(E)$  the space of probability measures on the borelian  $\sigma$ -algebra  $\mathcal{B}_E$  of  $E$ .

**Monge-Kantorovitch-Wasserstein (MKW) distances.**

As they will be a cornerstone in that article, used indifferent setting, we briefly recall their definition and main properties, and refer to [56] for a very nice presentation.

On a general Polish space  $Z$ , for any distance  $D : Z \times Z \rightarrow \mathbb{R}^+$  and  $p \in [1, \infty)$ , we define  $W_{D,p}$  on  $\mathbf{P}(Z) \times \mathbf{P}(Z)$  by setting for any  $\rho_1, \rho_2 \in \mathbf{P}(Z)$

$$[W_{D,p}(\rho_1, \rho_2)]^p := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{Z \times Z} D(x, y)^p \pi(dx, dy)$$

where  $\Pi(\rho_1, \rho_2)$  is the set of probability measures  $\pi \in \mathbf{P}(Z \times Z)$  with first marginal  $\rho_1$  and second marginal  $\rho_2$ , that is  $\pi(A \times Z) = \rho_1(A)$  and  $\pi(Z \times A) = \rho_2(A)$  for any Borel set  $A \subset Z$ . It defines a distance on  $\mathbf{P}(Z)$ .

**The phase spaces  $E^N$  (its marginal's space  $E^j$ ) and  $\mathbf{P}(E)$ .**

When we study system of  $N$  particles, the natural phase space is  $E^N$ . The space of marginals  $E^j$  for  $1 \leq j \leq N$  are also important. We present here the different distances we shall use on these spaces.

- On  $E$  we will use mainly two distances :
  - the usual Euclidian distance denoted by  $|x - y|$ ;
  - a bounded version of the square distance :  $d_E(x, y) = |x - y| \wedge 1$  for any  $x, y \in E$ .

- On the space  $E^j$  for  $1 \leq j$ , we will also use the two distances
  - the *normalized* square distance  $|X - Y|_2$  defined for any  $X = (x_1, \dots, x_j) \in E^j$  and  $Y = (y_1, \dots, y_j) \in E^j$  by

$$|X - Y|_2^2 := \frac{1}{j} \sum_{i=1}^j |x_i - y_i|^2;$$

- the *normalized* bounded distance  $d_j = d_{E^j}$  defined by

$$(2.1) \quad d_{E^j}(X, Y) := \frac{1}{j} \sum_{i=1}^j d_E(x_i, y_i).$$

It is worth emphasizing that the normalizing factor  $1/j$  is important in the sequel in order to obtain formulas independent of the number  $j$  of variables.

- The introduction of the empirical measures allows to “identify” our phase space  $E^N$  to a subspace of  $\mathbf{P}(E)$ . To be more precise, we denote by  $\mathcal{P}_N(E)$  the set of empirical measures

$$\mathcal{P}_N(E) := \{ \mu_X^N, X = (x_1, \dots, x_N) \in E^N \} \subset \mathbf{P}(E),$$

where  $\mu_X^N$  stands for the empirical measure defined by (1.3) and associated to the configuration  $X = (x_1, \dots, x_N) \in E^N$ . We denote by  $p_N : E^N \rightarrow \mathcal{P}_N(E)$  the application that maps a configuration to its empirical measure :  $p_N(X) := \mu_X^N$ .

- On our phase space  $\mathbf{P}(E)$ , we will use three different distances

- The usual MKW distance of order two  $W_2$  defined as above with the choice  $D(x, y) = |x - y|^2$

$$W_2(\rho_1, \rho_2)^2 = W_{|\cdot|^2, 2}(\rho_1, \rho_2)^2 := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{E \times E} |x - y|^2 \pi(dx, dy)$$

- The MKW distance  $W_1$  associated to  $d_E$  defined by

$$(2.2) \quad W_1(\rho_1, \rho_2) = W_{d_E, 1}(\rho_1, \rho_2) := \inf_{\pi \in \Pi(\rho_1, \rho_2)} \int_{E \times E} d_E(x, y) \pi(dx, dy)$$

From the Kantorovich-Rubinstein duality theorem (see for instance [56, Theorem 1.14]) we have the following alternative characterization

$$(2.3) \quad \forall \rho_1, \rho_2 \in \mathbf{P}(E) \quad W_1(\rho_1, \rho_2) = \sup_{\|\varphi\|_{Lip} \leq 1} \int_E \varphi(x) (\rho_1(dx) - \rho_2(dx)),$$

Where  $\|\varphi\|_{Lip} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_E(x, y)}$  is the Lipschitz semi-norm relatively to the distance  $d_E$ . This semi-norm is closely related to the usual Lipschitz semi-norm since it satisfies

$$(2.4) \quad \frac{1}{2} (\|\nabla \varphi\|_\infty + \|\varphi - \varphi(0)\|_\infty) \leq \|\varphi\|_{Lip} \leq 2 (\|\nabla \varphi\|_\infty + \|\varphi\|_\infty) =: 2 \|\varphi\|_{W^{1, \infty}}.$$

It implies that  $W_1$  is equivalent to the  $(W^{1, \infty})'$ -distance denoted by  $D_{W^{1, \infty}}$

$$(2.5) \quad D_{W^{1, \infty}} := \sup_{\|\varphi\|_{W^{1, \infty}} \leq 1} \int_E \varphi(x) (\rho_1(dx) - \rho_2(dx)), \quad \frac{1}{2} D_{W^{1, \infty}} \leq W_1 \leq 2 D_{W^{1, \infty}}.$$

- The distance induced by the  $H^{-s}$  norm for  $s > \frac{d}{2}$  : for any  $\rho, \eta \in \mathbf{P}(E)$

$$\|\rho - \eta\|_{H^{-s}} := \int_{\mathbb{R}^d} |\hat{\rho}(\xi) - \hat{\eta}(\xi)|^2 \frac{d\xi}{\langle \xi \rangle^{2s}}$$

where  $\hat{\rho}$  denotes the Fourier transform of  $\rho$  (which may always be seen as a measure on the whole  $\mathbb{R}^d$ ), and  $\langle \xi \rangle = \sqrt{1 + |\xi|^2}$ .

- We will often restrict ourself to the spaces  $\mathbf{P}_k(E)$  of probabilities with finite moment of order  $k > 0$  defined by

$$\mathbf{P}_k(E) := \{\rho \in \mathbf{P}(E) \text{ s.t. } M_k(\rho) := \int_E \langle v \rangle^k \rho(dv) < +\infty\}.$$

**The probability "spaces"  $\mathbf{P}(E^N)$ , its marginals spaces  $\mathbf{P}(E^j)$ , and  $\mathbf{P}(\mathbf{P}(E))$ .** The next step is to consider probabilities on the configuration spaces.

- The space  $\mathbf{P}(E^N)$  will be endowed with two distances
  - $W_1$  the MKW distance on  $\mathbf{P}(E^N)$  associated to  $d_{E^N}$  and  $p = 1$ , which has the same properties than the one of  $\mathbf{P}(E)$  and satisfies in particular the Kantorovich-Rubinstein formulation (2.3).
  - $W_2$  the MKW distance associated to the normalized square distance  $|\cdot|_2$  defined above.

Remark that we will only work on the subspace  $\mathbf{P}_{sym}(E^N)$  of borelian probability measures which are invariant under coordinates permutations.

- On the probability space  $\mathbf{P}(\mathbf{P}(E))$ , we can define different distances thanks to the Monge-Kantorovich-Wasserstein construction. We will use three of them

–  $\mathcal{W}_1$ , the MKW distance induced by the cost function  $W_1$  on  $\mathbf{P}(E)$ . In short

$$(2.6) \quad \mathcal{W}_1(\alpha_1, \alpha_2) = W_{W_1,1}(\alpha_1, \alpha_2) := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho_1, \rho_2) \pi(d\rho_1, d\rho_2),$$

- $\mathcal{W}_2$ , the MKW distance induced by the cost function  $W_2^2$  on  $\mathbf{P}(E)$ . In short

$$\mathcal{W}_2(\alpha_1, \alpha_2)^2 = W_{W_2,2}(\alpha_1, \alpha_2)^2 := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_2^2(\rho_1, \rho_2) \pi(d\rho_1, d\rho_2),$$

- $\mathcal{W}_{H^{-s}}$ , the MKW distance induced by the cost function  $\|\cdot\|_{H^{-s}}^2$  on  $\mathbf{P}(E)$ . In short

$$\mathcal{W}_{H^{-s}}(\alpha_1, \alpha_2)^2 = \mathcal{W}_{\|\cdot\|_{H^{-s}},2}(\alpha_1, \alpha_2)^2 := \inf_{\pi \in \Pi(\alpha_1, \alpha_2)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho_1 - \rho_2\|_{H^{-s}}^2 \pi(d\rho_1, d\rho_2).$$

- Remark that the application "empirical measure"  $p_N$  allows to define by push-forward a canonical map between  $\mathbf{P}(E^N)$  and  $\mathbf{P}(\mathbf{P}(E))$ . For  $G^N \in \mathbf{P}(E^N)$  we denote its image under the application  $p_N$  by  $\hat{G}^N \in \mathbf{P}(\mathbf{P}(E))$  :  $\hat{G}^N := G^N_{\#} p_N$ . It satisfies the duality relation

$$(2.7) \quad \forall \Phi \in C_b(\mathbf{P}(E)) \quad \langle \hat{G}^N, \Phi \rangle = \int_{E^N} \Phi(\mu_X^N) G^N(dX).$$

**More properties of the space  $\mathbf{P}(\mathbf{P}(E))$ .**

- **Marginals of probabilities on  $\mathbf{P}(\mathbf{P}(E))$ .** We can define a mapping from  $\mathbf{P}(\mathbf{P}(E))$  onto  $\mathbf{P}(E^j)$  in the following way. For any  $\alpha \in \mathbf{P}(\mathbf{P}(E))$  we define the projection  $\alpha_j \in \mathbf{P}(E^j)$  thanks to the relation

$$\alpha_j := \int \rho^{\otimes j} d\alpha(\rho).$$

It may also be restated using *polynomial functions* : for any  $\varphi \in C_b(E^j)$  we define the monomial (of order  $j$ ) function  $R_\varphi \in C_b(\mathbf{P}(E))$  by

$$\forall \rho \in \mathbf{P}(E) \quad R_\varphi(\rho) := \int_{E^j} \varphi(X) \rho^{\otimes j}(dX).$$

We remark that the monomial functions of all orders generate an algebra of continuous function (for the weak convergence of measures) that are called polynomials. When  $E$  is

compact so that  $\mathbf{P}(E)$  is also compact, they form a dense subset of  $C_b(\mathbf{P}(E))$  thanks to the Stone-Weierstrass theorem.

In terms of polynomial functions, the marginal  $\alpha_j$  may be defined by

$$\forall \varphi \in C_b(E^j) \quad \langle \alpha_j, \varphi \rangle := \langle \alpha, R_\varphi \rangle.$$

• Starting from  $G^N \in \mathbf{P}_{sym}(E^N)$ , we can define its push-forward  $\hat{G}^N$  and then for any  $1 \leq j \leq N$  the marginals of the push-forward  $\hat{G}_j^N := (\hat{G}^N)_j \in \mathbf{P}_{sym}(E^j)$ . They satisfy the duality relation

$$(2.8) \quad \forall \varphi \in C_b(E^j) \quad \langle \hat{G}_j^N, \varphi \rangle := \int_{E^N} R_\varphi(\mu_X^N) G^N(dX).$$

We emphasize that it is not equals to  $G_j^N$  the  $j$ -th marginal of  $G^N$ , but we will see later that the two probabilities  $G_j^N$  and  $\hat{G}_j^N$  are close (a precise version is recalled in Lemma 2.6).

**Different quantities measuring chaoticity.** Now that everything has been defined, we introduce the quantities that we will use to quantify the chaoticity of a sequence  $G^N \in \mathbf{P}_{sym}(E^N)$  of symmetric probability with respect to a profil  $f \in \mathbf{P}(E)$ :

- The chaoticity can be mesured on  $E^j$  for  $j \geq 2$ . For any  $1 \leq j \leq N$ , we set

$$\Omega_j(G^N; f) := W_1(G_j^N, f^{\otimes j}),$$

- and also on  $\mathbf{P}(E)$  by

$$\Omega_0(G^N; f) := \mathcal{W}_1(\hat{G}^N, \delta_f) = \int_{E^N} W_1(\mu_X^N, f) G^N(dX),$$

since there is only one transference plan  $\alpha \otimes \delta_f$  in  $\Pi(\alpha, \delta_f)$ .

## 2.2. Equivalence of distances on $\mathbf{P}(E)$ , $\mathbf{P}_{sym}(E^N)$ and $\mathbf{P}(\mathbf{P}(E))$ .

To quantify the equivalence between the distances defined above on  $\mathbf{P}(E)$ , we will need some assumption on the moments. The metrics  $W_1$ ,  $W_2$  and  $\|\cdot\|_{H^{-s}}$  are uniformly topologically equivalent in  $\mathbf{P}_k(E)$  for any  $k > 0$ . More precisely, we have

**Lemma 2.1.** *Choose  $f, g \in \mathbf{P}(E)$ . For any  $k > 0$ , denotes  $\mathcal{M}_k := M_k(f) + M_k(g)$ .*

(i) *For any  $s \geq 1$  if  $d = 1$ ,  $s > d/2$  if  $d \geq 2$  there exists  $C := C(d, s) \in (0, \infty)$  such that for any  $k > 0$ , there holds*

$$(2.9) \quad W_1(f, g) \leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \|f - g\|_{H^{-s}}^{\frac{2k}{d+2ks}}.$$

(ii) *For any  $k > 2$*

$$(2.10) \quad W_2(f, g) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{1/k} W_1(f, g)^{1/2-1/k}.$$

(iii) *It also holds without moment assumptions for  $s > \frac{d+1}{2}$  and a constant  $C(s, d)$*

$$W_1(f, g) \leq W_2(f, g), \quad \|f - g\|_{H^{-s}} \leq C W_1(f, g)^{\frac{1}{2}}.$$

**PROOF OF LEMMA 2.1.** The proofs is a mere adaptation of classical results on comparison of distances in probability spaces as it can be found in [49, 16, 43] for instance. We nevertheless sketch it for the sake of completeness.

*Proof of i).*

We consider a truncation sequence  $\chi_R(x) = \chi(x/R)$ ,  $R > 0$ , with  $\chi \in C_c^\infty(\mathbb{R}^d)$ ,  $\|\nabla \chi\|_\infty \leq 1$ ,  $0 \leq \chi \leq 1$ ,  $\chi \equiv 1$  on  $B(0, 1)$ , and the sequence of mollifiers  $\omega_\varepsilon(x) = \varepsilon^{-d} \omega(x/\varepsilon)$ ,  $\varepsilon > 0$ , with  $\omega(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$ , so that  $\hat{\omega}_\varepsilon(\xi) = \exp(-\varepsilon^2 |\xi|^2/2)$ .

In view of the equivalence of distance (2.5), we choose a  $\varphi \in W^{1,\infty}(\mathbb{R}^d)$  such that  $\|\varphi\|_{W^{1,\infty}} \leq 1$ , we define  $\varphi_R := \varphi \chi_R$ ,  $\varphi_{R,\varepsilon} = \varphi_R * \omega_\varepsilon$  and we write

$$\int \varphi (df - dg) = \int \varphi_{R,\varepsilon} (df - dg) + \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) + \int (\varphi - \varphi_R) (df - dg).$$

For the last term, we have

$$\forall R > 0 \quad \left| \int (\varphi_R - \varphi) (df - dg) \right| \leq \int_{B_R^c} \|\varphi\|_\infty \frac{|x|^k}{R^k} (df + dg) \leq \frac{M_k}{R^k}.$$

For the second term, we observe that

$$\|\varphi_R - \varphi_{R,\varepsilon}\|_\infty \leq \|\nabla \varphi_R\|_\infty \int_{\mathbb{R}^d} \omega_\varepsilon(x) |x| dx \leq C(d) \varepsilon,$$

and we get

$$\left| \int (\varphi_R - \varphi_{R,\varepsilon}) (df - dg) \right| \leq C(d) \varepsilon.$$

Finally, the first term can be estimated by

$$\left| \int \varphi_{R,\varepsilon} (df - dg) \right| \leq \|\varphi_{R,\varepsilon}\|_{H^s} \|f - g\|_{H^{-s}},$$

with for any  $R \geq 1$  and  $\varepsilon \in (0, 1]$

$$\begin{aligned} \|\varphi_{R,\varepsilon}\|_{H^s} &= \left( \int \langle \xi \rangle^2 |\widehat{\varphi \chi_R}|^2 \langle \xi \rangle^{2(s-1)} |\hat{\omega}_\varepsilon|^2 d\xi \right)^{1/2} \\ &\leq \|\varphi \chi_R\|_{H^1} \|\langle \xi \rangle^{s-1} \hat{\omega}_\varepsilon(\xi)\|_{L^\infty} \leq C(s, d) R^{d/2} \varepsilon^{-(s-1)}. \end{aligned}$$

In the estimation of  $\|\langle \cdot \rangle^{s-1} \hat{\omega}_\varepsilon(\cdot)\|_\infty$  we used  $s \geq 1$ . All in all, we have

$$W_1(f, g) \leq C(s, d) \left( \varepsilon + \frac{M_k}{R^k} + R^{\frac{d}{2}} \varepsilon^{-(s-1)} \|f - g\|_{H^{-s}} \right).$$

This yields to (2.9) by optimizing the parameter  $\varepsilon$  and  $R$  with

$$R = M_k^{\frac{2s}{d+2ks}} \|f - g\|_{H^{-s}}^{-\frac{2}{d+2ks}}, \quad \text{and } \varepsilon = M_k^{\frac{d}{d+2ks}} \|f - g\|_{H^{-s}}^{\frac{2k}{d+2ks}}.$$

*Proof of ii).* We have for any  $R \geq 1$  the inequality

$$\forall x, y \in E, \quad |x - y|^2 \leq R^2 d_E(x, y) + \frac{2^k}{R^{k-2}} (|x|^k + |y|^k)$$

from which we deduce

$$\begin{aligned} W_2(f, g)^2 &\leq R^2 \inf_{\pi \in \Pi(f, g)} \int_{E \times E} d_E(x, y) \pi(dx, dy) \\ &\quad + \frac{2^k}{R^{k-2}} \sup_{\pi \in \Pi(f, g)} \int_{E \times E} (|x_i|^k + |y_i|^k) \pi(dx, dy) \\ &\leq R^2 W_1(f, g) + \frac{2^k}{R^{k-2}} (M_k(f) + M_k(g)), \end{aligned}$$

and then we get with  $(R/2)^k = \mathcal{M}_k/W_1$

$$(2.11) \quad W_2(f, g) \leq 2^{3/2} \mathcal{M}_k^{1/k} W_1(f, g)^{1/2-1/k}.$$

*Proof of iii).* The first point is classical. The second relies on the fact that  $\|\delta_x - \delta_y\|_{H^{-s}}^2 \leq C d_E(x, y)$ . □

There is also a similar result on  $E^N$ , where the  $H^{-s}$  is less useful.

**Lemma 2.2.** *Choose  $F^N, G^N \in \mathbf{P}_{\text{sym}}(E^N)$ . For any  $k > 0$ , denotes*

$$\mathcal{M}_k := M_k(F_1^N) + M_k(G_1^N).$$

For any  $k > 2$ , it holds that

$$(2.12) \quad W_2(F^N, G^N) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{1/k} W_1(F^N, G^N)^{1/2-1/k}$$

It also holds without moment assumptions that  $W_1(F^N, G^N) \leq W_2(F^N, G^N)$ .

**PROOF OF LEMMA 2.2.** The proof is a simple generalisation of (2.10) to the case of  $N$  variables. We skip it.  $\square$

The inequalities of Lemma 2.1 also sum well on  $\mathbf{P}(\mathbf{P}(E))$  in order to get

**Lemma 2.3.** *Choose  $\alpha, \beta \in \mathbf{P}(\mathbf{P}(E))$ , and define*

$$\mathcal{M}_k := M_k(\alpha) + M_k(\beta) := \int M_k(\rho) [\alpha + \beta](d\rho) = M_k(\alpha_1) + M_k(\beta_1).$$

(i) For any  $s \geq 1$  if  $d = 1$ ,  $s > d/2$  if  $d \geq 2$  there exists  $C := C(d, s) \in (0, \infty)$  such that for any  $k > 0$ ,

$$(2.13) \quad \mathcal{W}_1(\alpha, \beta) \leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\alpha, \beta)^{\frac{2k}{d+2ks}}.$$

ii) For any  $k > 2$ , it also holds that

$$(2.14) \quad (ii) \quad \mathcal{W}_2(\alpha, \beta) \leq 2^{\frac{3}{2}} \mathcal{M}_k^{\frac{1}{k}} \mathcal{W}_1(\alpha, \beta)^{\frac{1}{2}-\frac{1}{k}},$$

iii) It holds without moment assumption that  $\mathcal{W}_1 \leq \mathcal{W}_2$  and  $\mathcal{W}_{H^{-s}} \leq C \mathcal{W}_1^{\frac{1}{2}}$  for  $s > \frac{d+1}{2}$  with a constant  $C(s, d)$ .

**PROOF OF LEMMA 2.3.** All the above estimates are simple summations of the corresponding estimate of Lemma 2.1. We only prove i).

$$\begin{aligned} \mathcal{W}_1(\alpha, \beta) &= \inf_{\Pi \in \Pi(\alpha, \beta)} \int W_1(\rho, \eta) \Pi(d\rho, d\eta) \\ &\leq C \inf_{\Pi \in \Pi(\alpha, \beta)} \int [M_k(\rho) + M_k(\eta)]^{\frac{d}{d+2ks}} \|\rho - \eta\|_{H^{-s}}^{\frac{2k}{d+2ks}} \Pi(d\rho, d\eta) \\ &\leq C \left( \int M_k(\rho) [\alpha + \beta](d\rho) \right)^{\frac{d}{d+2ks}} \left( \inf_{\Pi \in \Pi(\alpha, \beta)} \int \|\rho - \eta\|_{H^{-s}}^{\frac{1}{s}} \Pi(d\rho, d\eta) \right)^{\frac{2ks}{d+2ks}} \\ &\leq C [M_k(\alpha) + M_k(\beta)]^{\frac{d}{d+2ks}} \left( \inf_{\Pi \in \Pi(\alpha, \beta)} \int \|\rho - \eta\|_{H^{-s}}^2 \Pi(d\rho, d\eta) \right)^{\frac{k}{d+2ks}} \\ &\leq C \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\alpha, \beta)^{\frac{2k}{d+2ks}} \end{aligned}$$

where we have successively used the inequality (2.9), Hölder inequality, the definition of the moment of  $\alpha$  and  $\beta$ , and Jensen inequality.  $\square$

**2.3. Quantified equivalence of chaos.** We begin with some probably well known elementary inequalities and identities concerning Monge-Kantorovich-Wasserstein in space product. For the sake of completeness we will nevertheless sketch the proofs of them. Remark that the two first formulas are particularly simple thanks to the choice of the normalization (2.1), and that they remains valid if we replace  $d_j$  by the normalized  $l^1$ -distance  $\frac{1}{j} \sum_i |x_i - y_i|$ .

**Proposition 2.4.** *a) - For any  $F^N, G^N \in \mathbf{P}_{sym}(E^N)$  and  $1 \leq j \leq N$ , there holds*

$$(2.15) \quad W_1(F_j^N, G_j^N) \leq \left( \frac{j}{N} \left[ \frac{N}{j} \right] \right)^{-1} W_1(F^N, G^N) \leq 2 W_1(F^N, G^N).$$

*b) - For any  $f, g \in P(E)$ , there holds*

$$(2.16) \quad W_1(f^{\otimes N}, g^{\otimes N}) = W_1(f, g).$$

*c) - For any  $f, g, h \in P(E)$ , there holds*

$$(2.17) \quad 2 W_1(f \otimes h, g \otimes h) = W_1(f, g).$$

As can be seen in the following proof, similar results also holds for MKW constructed with arbitrary distance  $D$  and exponents  $p$ , and therefore for the  $W_2$  distance. We do not state them precisely, but they will be useful in the proof of the next Lemma 2.5.

**PROOF OF PROPOSITION 2.4.**

*Proof of (2.15).* Consider  $\pi \in \Pi(F^N, G^N)$  an optimal transfer plan in (2.2). Introducing the Euclidian division,  $N = nj + r$ ,  $0 \leq r \leq j - 1$ , we have

$$\begin{aligned} W_1(F^N, G^N) &= \int_{E^{2N}} d_{E^N}(X, Y) \pi(dX, dY) \\ &= \frac{1}{N} \int_{E^{2N}} \left( \sum_{i=1}^n j d_{E^j}(X_i, X_i) + r d_{E^r}(X_0, Y_0) \right) \pi(dX, dY) \\ &\geq \frac{j}{N} \sum_{i=1}^n \int_{E^{2j}} d_{E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dX_i), \end{aligned}$$

with  $\tilde{\pi}_i \in \Pi(\tilde{F}_i, \tilde{G}_i)$ , where  $\tilde{F}_i$  and  $\tilde{G}_i \in \mathbf{P}(E^j)$  denote the marginal probabilities of  $F^N$  and  $G^N$  on the  $i$ -th block of variables. From the symmetry hypothesis, we have  $\tilde{F}_i = \tilde{F}_1 = F_j^N$  and  $\tilde{G}_i = \tilde{G}_1 = G_j^N$  for any  $1 \leq i \leq n$ . As a consequence, we have

$$\int_{E^{2j}} d_{E^j}(X_i, Y_i) \tilde{\pi}_i(dX_i, dX_i) \geq W_1(F_j^N, G_j^N),$$

and we then deduce the first inequality in (2.15). Since the integer portion  $n := [N/j]$  is larger than 1, we have

$$\frac{j}{N} \left[ \frac{N}{j} \right] = \frac{nj}{nj+r} \geq \frac{nj}{nj+j} \geq \frac{1}{2},$$

from which we deduce the second inequality in (2.15).

*Proof of (2.16).* We consider  $\alpha \in \Pi(f, g)$  an optimal transport plan for the  $W_1(f, g)$  distance and we define the associated transport plan  $\bar{\pi} := \alpha^{\otimes N} \in \Pi(f^{\otimes N}, g^{\otimes N})$  by

$$\forall A_i, B_i \in E \quad \bar{\pi}(A_1 \times \dots \times A_N \times B_1 \times \dots \times B_N) = \alpha(A_1 \times B_1) \times \dots \times \alpha(A_N \times B_N).$$

By definition of  $W_1(f^{\otimes N}, g^{\otimes N})$ , we then have

$$W_1(f^{\otimes N}, g^{\otimes N}) \leq \frac{1}{N} \sum_{i=1}^N \int_{E^{2N}} d(x_i, y_i) \bar{\pi}(dX, dY) = W_1(f, g),$$

Since the first inequality in (2.15) in the case  $j = 1$  implies the reverse inequality, the above inequality is an equality.

*Proof of (2.17).* On the one hand, from the definition of the distance  $W_1$  by transport plans, we have for an optimal transport plan  $\pi \in \Pi(f \otimes h, g \otimes h)$  the inequality

$$\begin{aligned} W_1(f \otimes h, g \otimes h) &= \frac{1}{2} \int_{E^4} (d_E(x_1, y_1) + d_E(x_2, y_2)) \pi(dx_1, dx_2, dy_1, dy_2) \\ &\geq \frac{1}{2} \int_{E^4} d_E(x_1, y_1) \pi_1(dx_1, dy_1) \geq \frac{1}{2} W_1(f, g), \end{aligned}$$

since the 1-marginal  $\pi_1$  defined by  $\pi_1(A \times B) = \pi(A \times E \times B \times E)$  for any  $A, B \in \mathcal{B}_E$  belongs to the transport plans set  $\Pi(f, g)$ .

On the other hand, considering an optimal transport plan  $\pi \in \Pi(f, g)$  for the  $W_1$  distance, we define the associated transport plan  $\bar{\pi}(dx, dy) := \pi(dx_1, dy_1) \otimes h(dx_2) \delta_{y_2=x_2} \in \Pi(f \otimes h, g \otimes h)$ , and we observe that

$$\begin{aligned} W_1(f \otimes h, g \otimes h) &\leq \frac{1}{2} \int_{E^4} (d_E(x_1, y_1) + d_E(x_2, y_2)) \bar{\pi}(dx_1, dx_2, dy_1, dy_2) \\ &= \frac{1}{2} \int_{E^4} d_E(x_1, y_1) \pi(dx_1, dy_1) = \frac{1}{2} W_1(f, g). \end{aligned}$$

We obtain (2.17) by gathering these two inequalities.  $\square$

As a immediate corollary of (2.15) with  $N := \ell$ ,  $F^\ell := f^{\otimes \ell}$  and  $G^\ell := G_\ell^N$ , we have

**Theorem 1.2 - part 1.** *For any  $G^N \in \mathbf{P}_{sym}(E^N)$  and  $f \in \mathbf{P}(E)$ , there holds*

$$(2.18) \quad \forall 1 \leq j \leq \ell \leq N \quad \Omega_j(G^N; f) \leq 2 \Omega_\ell(G^N; f).$$

We conclude by another lemma that allows to compare distance between measures on  $\mathbf{P}(\mathbf{P}(E))$  and distance between their marginals on  $E^j$ , and thus to compare  $\Omega_\ell$  and  $\Omega_0$ .

**Lemma 2.5.** *For any distance  $D$  on  $E$  and  $p \geq 1$ , extend  $D$  on  $E^j$  with  $D_{j,p}(V, W)^p = \frac{1}{j} \sum_i D(v_i, w_i)^p$ , and define the associated MKW distance  $W_{D_{j,p},p}$  on  $\mathbf{P}(E^j)$  and the MKW distance  $\mathcal{W}_{W_D,p}$  on  $\mathbf{P}(\mathbf{P}(E))$  associated to  $W_D$  and  $p$ . Let  $\alpha$  and  $\beta$  be two probability on  $\mathbf{P}(\mathbf{P}(E))$ . Then, for any  $j \in \mathbb{N}$ ,*

$$(2.19) \quad W_{D_{j,p},p}(\alpha_j, \beta_j) \leq \mathcal{W}_{W_D,p}(\alpha, \beta)$$

*That is in particular true for the MKW distances  $W_1$  and  $W_2$  defined in section 2.1*

$$\forall j \in \mathbb{N}, \quad W_2(\alpha_j, \beta_j) \leq \mathcal{W}_2(\alpha, \beta), \quad W_1(\alpha_j, \beta_j) \leq \mathcal{W}_1(\alpha, \beta).$$



PROOF OF LEMMA 2.5. For simplicity we denote for any  $j$ ,  $W_{D_{j,p}} = W_D$ . We choose any transference plan  $\Pi$  between  $\alpha$  and  $\beta$  and write

$$\begin{aligned} [W_D(\alpha_j, \beta_j)]^p &= \left[ W_D \left( \int \rho^{\otimes j} \alpha(d\rho), \int \rho^{\otimes j} \beta(d\rho) \right) \right]^p \\ &= \left[ W_D \left( \int \rho^{\otimes j} \pi(d\rho, d\eta), \int \eta^{\otimes j} \pi(d\rho, d\eta) \right) \right]^p \\ &\leq \left[ \int W_D(\rho^{\otimes j}, \eta^{\otimes j}) \pi(d\rho, d\eta) \right]^p \\ &\leq \int [W_D(\rho, \eta)]^p \pi(d\rho, d\eta), \end{aligned}$$

where we have used the convexity property of the Wasserstein distance, the equivalent of equality (2.16) in our general case, and Jensen inequality. By optimisation on  $\pi$  we obtain the claimed inequality.  $\square$

As a consequence of a classical combinatory trick, which goes back at least from [26], we have

**Lemma 2.6** (Quantification of the equivalence  $G_j^N \sim \hat{G}_j^N$ ). *For any  $G^N \in \mathbf{P}_{sym}(E^N)$  and any  $1 \leq j \leq N$ , we have*

$$\|G_j^N - \hat{G}_j^N\|_{TV} \leq 2 \frac{j(j-1)}{N} \quad \text{and} \quad W_1(G_j^N, \hat{G}_j^N) \leq \frac{j(j-1)}{N},$$

and in particular the first marginals are equal:  $G_1^N = \hat{G}_1^N$ .

PROOF OF LEMMA 2.6. For the first inequality, we refer to [45, Lemma 2.14 estimate (2.27)] which precisely claims that the RHS term is bounded by  $2j(j-1)/N$ .

In order to prove second inequality with bound  $\frac{j(j-1)}{N}$ , we show that for any measure  $\rho$  and  $\eta$  on  $E^j$ ,

$$W_1(\rho, \eta) \leq \frac{1}{2} \|\rho - \eta\|_{TV}.$$

To obtain that inequality, we shall only construct a transference plan that does not move the mass that  $\rho$  and  $\eta$  have in common. Precisely, the mass they share defines a positive measure which may be given by

$$\min(\rho, \eta) := \min \left( 1, \frac{d\eta}{d\rho} \right) \rho,$$

where  $\frac{d\eta}{d\rho}$  stands for the Radon-Nikodym of the part of  $\eta$  which is absolutely continuous with respect to  $\rho$ . Then we define two positive measures  $\bar{\eta} := \eta - \min(\rho, \eta)$  and  $\bar{\rho} := \rho - \min(\rho, \eta)$  which correspond to the "excess part" of the two measures. As  $\rho$  and  $\eta$  are two probability measures, we have  $\|\bar{\rho}\|_{TV} = \|\bar{\eta}\|_{TV} = \frac{1}{2} \|\rho - \eta\|_{TV}$ . Now we choose the transference plan  $\pi \in \Pi(\rho, \eta)$  defined by

$$\pi := \min(\rho, \eta)(dx) \delta_{y=x} + \frac{\bar{\rho}(dx) \otimes \bar{\eta}(dy)}{\|\bar{\rho}\|_{TV}}$$

For this plan, the transportation cost is less than

$$\int_{E^j \times E^j} d_j(x, y) d\pi \leq \|\bar{\eta}\|_{TV} = \frac{1}{2} \|\rho - \eta\|_{TV},$$

and we conclude since by definition  $W_1(\rho, \eta)$  is less than the above LHS term.  $\square$

Applying the previous lemmas 2.5 and 2.6, we can bound  $\Omega_j$  by  $\Omega_0$  and some rest. This is the second part of theorem 1.2.

**Theorem 1.2 - part 2.** *For any  $G^N \in \mathbf{P}_{sym}(E^N)$  and  $f \in \mathbf{P}(E)$ , there holds*

$$(2.20) \quad \forall 1 \leq j \leq N \quad \Omega_j(G^N; f) \leq \Omega_0(G^N; f) + \frac{j^2}{N}.$$

Of course, that part is usefull only when  $j \ll \sqrt{N}$ . It will be completed at the end of the section by a inequality valid also for large  $j$ .

PROOF OF THEOREM 1.2 - PART 2. We simply write

$$\begin{aligned} \Omega_j(G^N, f) = W_1(G_j^N, f^{\otimes j}) &\leq W_1(G_j^N, \hat{G}_j^N) + W_1(\hat{G}_j^N, f^{\otimes j}) \\ &\leq \frac{j^2}{N} + W_1(\hat{G}_j^N, \delta_f) = \frac{j^2}{N} + \Omega_0(G^N, f), \end{aligned}$$

thanks to the two previous lemmas 2.5 and 2.6 □

We establish now the key estimate which will leads to the inequalities in Theorem 1.2 where  $\Omega_0$  is controled by  $\Omega_2$ . Following [43, Lemma 4.2], the main idea is to use as an intermediate step the  $H^{-s}$  norm on  $\mathbf{P}(E)$ , rather than the Wassertsein  $W_1$  distance, because it is a monomial function of order two on  $\mathbf{P}(E)$ , and thus has a nice algebraic structure. This fact is stated in the following elementary lemma.

**Lemma 2.7.** *For  $s > d/2$ , define  $\Phi_s : \mathbb{R}^d \rightarrow \mathbb{R}$  by*

$$(2.21) \quad \forall z \in \mathbb{R}^d, \quad \Phi_s(z) := \int_{\mathbb{R}^d} e^{-iz \cdot \xi} \frac{d\xi}{\langle \xi \rangle^{2s}}.$$

*The function  $\Phi_s$  is radial, bounded, and furthermore if  $s > \frac{d+1}{2}$ , it is Lipschitz. For any  $\rho, \eta \in \mathbf{P}(E)$*

$$(2.22) \quad \|\rho - \eta\|_{H^{-s}}^2 = \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes \eta)(dx, dy) + \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\eta^{\otimes 2} - \eta \otimes \rho)(dx, dy),$$

*and in particular, for any  $\rho \in \mathbf{P}(E)$*

$$\|\rho\|_{H^{-s}}^2 = \int_{\mathbb{R}^{2d}} \Phi(x-y) \rho^{\otimes 2}(dx, dy),$$

*which means that the norm  $H^{-s}$  on  $\mathbf{P}(E)$  is the monomial function of order two associated to the function  $(x, y) \mapsto \Phi_s(x-y)$ .*

PROOF OF LEMMA 2.7. We obtain that  $\Phi_s$  is bounded from the fact that  $\int_{\mathbb{R}^d} \langle \xi \rangle^{-2s} d\xi$  is finite for  $s > d/2$ , and that it is Lipschitz from the fact that  $\int_{\mathbb{R}^d} \langle \xi \rangle^{1-2s} d\xi$  is finite when  $s > (d+1)/2$ . We now prove (2.22). Using the Fourier transform definition of the Hilbert norm of  $H^{-s}(\mathbb{R}^d)$ , we have for any  $\rho, \eta \in \mathbf{P}(E) \subset \mathbf{P}(\mathbb{R}^d) \subset H^{-s}(\mathbb{R}^d)$

$$\begin{aligned} \|\rho - \eta\|_{H^{-s}}^2 &= \int_{\mathbb{R}^d} (\hat{\rho}(\xi) - \hat{\eta}(\xi)) \overline{(\hat{\rho}(\xi) - \hat{\eta}(\xi))} \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &= \int_{\mathbb{R}^{3d}} (\rho(dx) - \eta(dx)) (\rho(dy) - \eta(dy)) e^{-i(x-y)\xi} \frac{d\xi}{\langle \xi \rangle^{2s}} \\ &= \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes \eta)(dx, dy) + \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\eta^{\otimes 2} - \eta \otimes \rho)(dx, dy). \end{aligned}$$

The last identity follows from (2.22) by choosing  $\eta = 0$ . □

Thanks to that Lemma, we will be able to obtain the following key estimate. We remark that the estimate is still true if we replace the bounded distance  $d_E$  by the usual one  $|\cdot|$  (with the constant  $C(d, s) = 2\|\nabla\Phi_s\|_\infty$  in that case).

**Proposition 2.8.** *For any  $s > \frac{d+1}{2}$  there exists a constant  $C = C(d, s) = 2\|\Phi_s\|_{Lip} \in (0, \infty)$  such that for any  $G^N \in \mathbf{P}_{sym}(E^N)$ ,  $N \geq 1$ ,  $f \in \mathbf{P}(E)$ , there holds*

$$(2.23) \quad \mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f) \leq C \left[ W_1(\hat{G}_2^N, f \otimes f) \right]^{\frac{1}{2}}.$$

PROOF OF PROPOSITION 2.8. Because  $\mathbf{P}(E) \subset \mathbf{P}(\mathbb{R}^d) \subset H^{-s}(\mathbb{R}^d)$  for  $s > \frac{d}{2}$  and  $\Pi(\hat{G}^N, \delta_f) = \{\hat{G}^N \otimes \delta_f\}$ , we have

$$\left[ \mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f) \right]^2 := \inf_{\pi \in \Pi(\hat{G}^N, \delta_f)} I[\pi] = I(\hat{G}^N \otimes \delta_f),$$

with cost functional

$$I[\pi] := \iint_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho - \eta\|_{H^{-s}}^2 \pi(d\rho, d\eta).$$

Using Lemma 2.7, we have

$$\begin{aligned} I[\hat{G}^N \otimes \delta_f] &= \int_{\mathbf{P}(E)} \left\{ \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (\rho^{\otimes 2} - \rho \otimes f)(dx, dy) \right\} \hat{G}^N(d\rho) \\ &\quad + \int_{\mathbf{P}(E)} \left\{ \int_{\mathbb{R}^{2d}} \Phi_s(x-y) (f^{\otimes 2} - f \otimes \rho)(dx, dy) \right\} \hat{G}^N(d\rho) \\ &= \iint_{E^2} \Phi_s(x-y) [\hat{G}_2^N(dx, dy) - \hat{G}_1^N(dx) f(dy)] \\ &\quad + \iint_{E^2} \Phi_s(x-y) [f(dx) f(dy) - f(dx) \hat{G}_1^N(dy)]. \end{aligned}$$

Now we may bound the cost functional as follows:

$$\begin{aligned} I[\hat{G}^N \otimes \delta_f] &\leq \|\Phi_s\|_{Lip} \left[ W_1(\hat{G}_2^N, \hat{G}_1^N \otimes f) + W_1(f \otimes f, f \otimes \hat{G}_1^N) \right] \\ &\leq \|\Phi_s\|_{Lip} \left[ W_1(\hat{G}_2^N, f \otimes f) + 2W_1(f \otimes f, \hat{G}_1^N \otimes f) \right] \\ &\leq \|\Phi_s\|_{Lip} \left[ W_1(\hat{G}_2^N, f \otimes f) + W_1(f, \hat{G}_1^N) \right] \\ &\leq 2\|\Phi_s\|_{Lip} W_1(\hat{G}_2^N, f \otimes f), \end{aligned}$$

where we have used successively the Katorovich-Rubinstein duality formula (2.3), the triangular inequality, the identity (2.17), and the first inequality in (2.15) together with the fact that  $(\hat{G}_2^N)_1 = \hat{G}_1^N$ .  $\square$

Putting together Proposition 2.8, Lemma 2.6 above and Lemma 2.3 on comparison of distance in  $\mathbf{P}(\mathbf{P}(E))$ , we have

**Theorem 1.2 - part 3.** *For any  $0 < \gamma < \frac{1}{d+1}$  and  $k > \frac{d}{\gamma^{-1}-d-1}$ , there exists a constant  $C := C(d, \gamma, k)$  such that for any  $G^N \in \mathbf{P}_{sym}(E^N)$  and  $f \in \mathbf{P}(E)$ , there holds*

$$(2.24) \quad \Omega_0(G^N; f) \leq C \mathcal{M}_k^{\frac{1}{k}} \left( \Omega_2(G^N; f) + \frac{1}{N} \right)^\gamma,$$

where as usual  $\mathcal{M}_k := M_k(f) + M_k(G_1^N)$ .

Remark that the constant  $C(d, \gamma, k)$  may be chosen decreasing in  $k$ , which is quite interesting for distribution with many moment, for instance decreasing exponentially or even with compact support.

**Remark 2.9.** Thanks to Theorem 1.2 - part 1, the  $\Omega_2$  term in the right hand side of (2.24) may be replaced by any  $\Omega_\ell$  for  $\ell \geq 2$ . But it cannot be replaced by  $\Omega_1$ , which does not measures chaoticity, as can be seen with the following example. We choose  $g$  and  $h$  two distincts probabilities on  $E$ , and take  $f := \frac{1}{2}(g + h)$ . We consider the probability  $G \in \mathbf{P}(\mathbf{P}(E))$ , and its associated sequence  $(G^N)$  of marginal probabilities on  $\mathbf{P}(E^N)$

$$G = \frac{1}{2}(\delta_g + \delta_h), \quad G^N := \frac{1}{2}g^{\otimes N} + \frac{1}{2}h^{\otimes N}.$$

As  $G_1 = f$ ,  $\Omega_1(G^N, f) = 0$  for all  $N$ , inequality (2.24) with  $\Omega_2$  replaced by  $\Omega_1$  will imply that  $\Omega_0(G^N, f)$  goes to zero. But from Theorem 1.2 - part 1 and with a little anticipation Theorem 1.2 - part 3 (See (2.30))

$$\frac{1}{2}W_1(G^2, f^{\otimes 2}) = \frac{1}{2}\Omega_2(G^N, f) \leq \Omega_N(G^N, f) \leq \Omega_0(G^N, f) + o(1).$$

There is a contradiction since  $G^2 \neq f^{\otimes 2}$  except if  $g = h$ .

PROOF OF THEOREM 1.2 - PART 3. We define  $s := \frac{1}{2\gamma} - \frac{d}{2k}$ . Notice that  $s > \frac{d+1}{2} \geq 1$  thanks to the conditions satisfied by  $\gamma$  and  $k$ . We can thus applied the point *i*) of Lemma 2.3, Proposition 2.8 and then Lemma 2.6 in order to get

$$\begin{aligned} \Omega_0(G^N; f) &:= W_1(\hat{G}^N, \delta_f) \leq C(d, s) \mathcal{M}_k^{\frac{d}{d+2ks}} \mathcal{W}_{H^{-s}}(\hat{G}^N, \delta_f)^{\frac{2k}{d+2ks}} \\ &\leq C(d, s) \mathcal{M}_k^{\frac{d}{d+2ks}} W_1(\hat{G}_2^N, f^{\otimes 2})^{\frac{k}{d+2ks}} \\ &\leq C(1 + \mathcal{M}_k)^{\frac{1}{k}} \left( W_1(G_2^N, f^{\otimes 2}) + \frac{2}{N} \right)^\gamma, \end{aligned}$$

since  $\gamma = \frac{k}{d+2ks}$ . This is the claimed inequality thanks to the definition of  $\Omega_2$ .  $\square$

**Remark 2.10.** In particular Theorem 1.2 - part 3 says that for any  $0 < \gamma < 1/(d+1)$  and  $k > d(\gamma^{-1} - d - 1)^{-1}$  there exists a constant  $C := C(d, \gamma, k)$  such that for any  $f \in \mathbf{P}(E)$ , there holds

$$(2.25) \quad \Omega_0(f^{\otimes N}; f) \leq \frac{C M_k(f)^{1/k}}{N^\gamma}.$$

It has been proved in [43, Lemma 4.2(iii)] a more accurate version of (2.25), namely that for any  $0 < \gamma < 1/d'$  and  $k > d'(\gamma^{-1} - d')^{-1}$ , with  $d' = \max(d, 2)$ , there exists a constant  $C := C(d, \gamma, k)$  such that (2.25) holds. It turns out that the above "tensorial case" is simpler and that we can apply similar technics with  $H^{-s}$  for  $s > d/2$  rather than  $s > \frac{d+1}{2}$  in the general case. This explain the gap in the exponents between the two theorems.

In the case when  $f$  is compactly supported in the ball of radius  $D$ , the bound (2.25) implies

$$\Omega_0(f^{\otimes N}; f) \leq \frac{C(d, \gamma)D}{N^\gamma}, \quad \text{for any } \gamma < \frac{1}{d}.$$

It is almost optimal since it has been shown ([20] and reference therein) that the previous inequality is valid for  $\gamma = \frac{1}{d}$  in the compact case.

We come to the proof of the last part of Theorem 1.2, which will be a consequence of the following proposition

**Proposition 2.11.** *For  $F^N, G^N \in \mathbf{P}_{sym}(E^N)$ , there holds*

$$(2.26) \quad W_1(F^N, G^N) = \mathcal{W}_1(\hat{F}^N, \hat{G}^N).$$

PROOF OF PROPOSITION 2.11. We split the proof into two steps.

*Step 1. A reformulation of the problem.* Since we are dealing with symmetric probabilities, it is natural to introduce the equivalence relation  $\sim$  in  $E^N$  by saying that  $X = (x_1, \dots, x_N), Y = (y_1, \dots, y_N) \in E^N$  are equivalent, we write  $X \sim Y$ , if there exists a permutation  $\sigma \in \mathfrak{S}_N$  such that  $Y = X_\sigma := (x_{\sigma(1)}, \dots, x_{\sigma(N)})$ .

We also introduce on  $E^N$  the "semi"-distance  $w_1$

$$(2.27) \quad w_1(X, Y) := \inf_{\sigma \in \mathfrak{S}_N} d_{E^N}(X, Y_\sigma) = \inf_{\sigma \in \mathfrak{S}_N} \frac{1}{N} \sum_{i=1}^N d_E(x_i, y_{\sigma(i)}),$$

which only satisfies  $W_1(X, Y) = 0 \Leftrightarrow X \sim Y$ . We also introduce the associated MKW functional  $W_1^\dagger$ . For  $F^N, G^N \in \mathbf{P}_{sym}(E^N)$ ,

$$W_1^\dagger(F^N, G^N) := \inf_{\pi^N \in \Pi(F^N, G^N)} \int_{E^N \times E^N} w_1(X, Y) \pi^N(dX, dY).$$

It is in fact a distance on the space of symmetric probabilities, but this point will also be a consequence of our proof. It is a classical result (see for instance [56, Introduction. Example: the discrete case]) that

$$(2.28) \quad \forall X, Y \in E^N, \quad W_1(\mu_X^N, \mu_Y^N) = w_1(X, Y),$$

(shortly, it means that we do not need to split the small Dirac masses when we try to optimize the transport between two empirical measures). We recall the notation  $p_N$  defined in section 2.1 for the application that send a configuration to the associated empirical measure :  $p_N(X) = \mu_X^N$ .

Remark that its associated push-forward mapping restricted to the symmetric probabilities

$$\tilde{p}_N : \mathbf{P}_{sym}(E^N) \rightarrow \mathbf{P}(\mathcal{P}_N(E)) \subset \mathbf{P}(\mathbf{P}(E)), \quad G^N \mapsto \hat{G}^N := G^N_{\#} p_N,$$

is a bijection. Its inverse can be simply expressed thanks to a dual formulation: for  $\alpha \in \mathbf{P}(\mathcal{P}_N(E))$ , its inverse  $\tilde{\alpha} = \tilde{p}_N^{-1} \alpha$  is the probability satisfying

$$\forall \varphi \in C_b(E^N), \quad \int_{E^N} \varphi(X) \tilde{\alpha}(dX) = \int_{\mathcal{P}_N(E)} \tilde{\varphi}(\rho) \alpha(d\rho),$$

where  $\tilde{\varphi}(\rho) := \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi(X_\sigma)$ , for any given  $X$  such that  $\mu = \mu_X^N$ . Similarly, defining  $\mathbf{P}_{s,s}(E^N \times E^N)$  the subset of  $\mathbf{P}(E^N \times E^N)$  of probabilities which are invariant under permutations on the first and second blocks of  $N$  variables separately, we have that

$$\tilde{p}_N^{\otimes 2} : \mathbf{P}_{s,s}(E^N \times E^N) \rightarrow \mathbf{P}(\mathcal{P}_N(E) \times \mathcal{P}_N(E)), \quad \pi^N \mapsto \hat{\pi}^N := \pi^N_{\#} (p_N, p_N),$$

is a bijection.

The identity (2.28) and the bijection  $\tilde{p}_N$  allows us to establish the identity

$$(2.29) \quad \forall F^N, G^N \in \mathbf{P}(E^N), \quad W^\dagger(F^N, G^N) = \mathcal{W}_1(\hat{F}^N, \hat{G}^N).$$

Indeed, denoting  $\Pi_{s,s}(F^N, G^N) = \Pi(F^N, G^N) \cap \mathbf{P}_{s,s}(E^N, E^N)$ , we have

$$\begin{aligned}
W_1^\dagger(F^N, G^N) &= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{E^N \times E^N} w_1(X, Y) \pi^N(dX, dY) \\
&= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{E^N \times E^N} W_1(p_N(X), p_N(Y)) \pi^N(dX, dY) \\
&= \inf_{\pi^N \in \Pi_{s,s}(F^N, G^N)} \int_{\mathcal{P}_N(E) \times \mathcal{P}_N(E)} W_1(\rho, \eta) \pi_{\#}^N(p_N, p_N)(d\rho, d\eta) \\
&= \inf_{\hat{\pi} \in \Pi(\hat{F}^N, \hat{G}^N)} \int_{\mathbf{P}(E) \times \mathbf{P}(E)} W_1(\rho, \eta) \hat{\pi}(d\rho, d\eta) = W_1(\hat{F}^N, \hat{G}^N),
\end{aligned}$$

where we have essentially used the invariance  $w_1(X, Y) = w_1(X_\sigma, Y_\tau)$  for any  $\sigma, \tau \in \mathfrak{S}_N$  and the fact that  $\tilde{p}_N^{\otimes 2}$  is a bijection.

*Step 2. The equality  $W_1^\dagger = W_1$ .*

The interest of the reformulation (2.29) is that we can now work on one space:  $E^N$ . Remark that since  $w_1(X, Y) \leq d_{E^N}(X, Y)$ , we always have  $W_1^\dagger \leq W_1$ , and the equality will hold only if one optimal transport for  $W_1^\dagger$  plan is concentrated on the set

$$\mathcal{C} := \left\{ (X, Y) \in E^N \times E^N \text{ s.t. } w_1(X, Y) = \inf_{\sigma \in \mathfrak{S}_N} d_{E^N}(X, Y_\sigma) = d_{E^N}(X, Y) \right\}.$$

We choose an optimal transference plan  $\pi$  for  $W_1^\dagger$ . For simplicity we will assume that  $\pi$  is symmetric, i.e. unchanged by the applications  $P_\sigma : (X, Y) \mapsto (X_\sigma, Y_\sigma)$  for any  $\sigma \in \mathfrak{S}_N$ . If not, we replace it by its symmetrization  $\frac{1}{N!} \sum_{\sigma} \pi_{\#} P_\sigma$  which will still be an optimal transference plan of  $F^N$  onto  $G^N$ . Starting from  $\pi$ , we will construct a transference plan  $\pi^* \in \Pi(F^N, G^N)$  such that

- i)  $\pi^*$  is concentrated on  $\mathcal{C}$ .
- ii)  $I_N[\pi] = \int w_1(X, Y) \pi(dX, dY) = \int w_1(X, Y) \pi^*(dX, dY) = I_N[\pi^*]$

Both properties imply then that

$$\begin{aligned}
W_1^\dagger(F^N, G^N) &= \int_{E^N \times E^N} w_1(X, Y) \pi(dX, dY) = \int_{E^N \times E^N} w_1(X, Y) \pi^*(dX, dY) \\
&= \int_{E^N \times E^N} d_{E^N}(X, Y) \pi^*(dX, dY) \geq W_1(F^N, G^N)
\end{aligned}$$

which is the desired inequality.

We define  $\pi^*$  in the following way. First, we introduce for any  $X, Y \in E^N$

$$\begin{aligned}
\mathcal{C}_{X;Y} &:= \{Z \in E^N; Z \sim Y \text{ and } d_{E^N}(X, Z) = w_1(X, Y)\} \subset E^N \\
\rho_{X;Y} &:= \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \delta_{(X,Z)}, \quad N_{X;Y} := \#\mathcal{C}_{X;Y}.
\end{aligned}$$

We note that  $Z \in \mathcal{C}_{X;Y}$  iff  $Z \sim Y$  and  $(X, Z) \in \mathcal{C}$ , so that  $\text{Supp } \rho_{X;Y} \subset \mathcal{C}$ . It can be shown that  $(X, Y) \mapsto N_{X;Y}$  is a borelian application (it takes finite values and its level set are closed) and that  $E^N \times E^N \rightarrow \mathbf{P}(E^N \times E^N)$ ,  $(X, Y) \mapsto \rho_{X;Y}$  is also borelian if  $\mathbf{P}(E^N \times E^N)$  is endowed with the weak topology of measures. This allows us to define a transference plan  $\pi^*$  by

$$\pi^* := \int_{E^N \times E^N} \rho_{X;Y} \pi(dX, dY) \in \mathbf{P}(E^N \times E^N),$$

or in other words, for any  $\psi \in C_b(E^N \times E^N)$ , we have

$$\begin{aligned} \langle \pi^*, \psi \rangle &= \int_{E^{2N}} \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \int_{E^{2N}} \psi(X', Y') \delta_{(X,Z)}(dX', dY') \pi^N(dX, dY) \\ &= \int_{E^{2N}} \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \psi(X, Z) \pi^N(dX, dY). \end{aligned}$$

It remains to proof that  $\pi^*$  satisfy the announced properties. Since  $\rho_{X;Y}$  is supported in  $\mathcal{C}$  for any  $(X, Y) \in E^N \times E^N$ , it is also the case for  $\pi^*$ . It is also not difficult to show that the transport cost for  $w_1$  is preserved. Indeed, we have

$$\begin{aligned} \int_{E^{2N}} d_{E^N}(X', Y') \pi^*(dX', dY') &= \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} d_{E^N}(X, Z) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} w_1(X, Y) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} w_1(X, Y) \pi(dX, dY). \end{aligned}$$

The fact that  $\pi^*$  has first marginal  $F^N$  is also clear since for any  $\varphi \in C_b(E^N)$

$$\begin{aligned} \int_{E^{2N}} \varphi(X') \pi^*(dX', dY') &= \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \varphi(X) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \varphi(X) \pi(dX, dY) = \int_{E^N} \varphi(X) F^N(dX). \end{aligned}$$

For the second marginal, we shall use the following properties of  $\mathcal{C}_{X;Y}$  and  $N_{X;Y}$

$$\forall \tau \in \mathfrak{S}_N, \quad Z_\tau \in \mathcal{C}_{X_\tau; Y_\tau} \Leftrightarrow Z \in \mathcal{C}_{X;Y}, \quad \text{and thus} \quad N_{X_\tau; Y_\tau} = N_{X;Y}.$$

Thanks to the invariance by symmetry of  $\pi$  and  $G^N$ , we can write for any  $\varphi \in C_b(E^N)$

$$\begin{aligned} \int_{E^{2N}} \varphi(Y) \pi^*(dX, dY) &= \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \varphi(Z) \right) \pi(dX, dY) \\ &= \frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \int_{E^{2N}} \left( \frac{1}{N_{X_\tau; Y_\tau}} \sum_{Z \in \mathcal{C}_{X_\tau; Y_\tau}} \varphi(Z) \right) \pi(dX, dY) \\ &= \frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \varphi(Z_\tau) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \left( \frac{1}{N_{X;Y}} \sum_{Z \in \mathcal{C}_{X;Y}} \tilde{\varphi}(Z) \right) \pi(dX, dY) \\ &= \int_{E^{2N}} \tilde{\varphi}(Y) \pi(dX, dY) \\ &= \int_{E^{2N}} \tilde{\varphi}(Y) G^N(dX) = \int_{E^{2N}} \varphi(Y) G^N(dX), \end{aligned}$$

where we have introduced the symmetrization of  $\varphi$  defined by  $\tilde{\varphi}(Z) := \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi(Z_\sigma)$  and we have used that  $\tilde{\varphi}(Z) = \tilde{\varphi}(Y)$  for any  $Z \in \mathcal{C}_{X;Y}$  and the fact that  $G^N$  is symmetric. This concludes the proof.  $\square$

Putting together Proposition 2.11 and (2.25), we have

**Theorem 1.2 - part 4.** *For any  $0 < \gamma < 1/d'$  and  $k > d'(\gamma^{-1} - d')$ ,  $d' = \max(d, 2)$ , there exists a constant  $C := C(d, \gamma, k)$  such that for any  $G^N \in \mathbf{P}_{\text{sym}}(E^N)$  and  $f \in \mathbf{P}_k(E)$ , there holds*

$$(2.30) \quad |\Omega_N(G^N, f) - \Omega_0(G^N, f)| \leq C \frac{M_k(f)^{1/k}}{N^\gamma}.$$

PROOF OF THEOREM 1.2 - PART 4. We have

$$\begin{aligned} |\Omega_N(G^N, f) - \Omega_0(G^N, f)| &= |W_1(G^N, f^{\otimes N}) - W_1(\hat{G}^N, \delta_f)| \\ &= |W_1(\hat{G}^N, \widehat{f^{\otimes N}}) - W_1(\hat{G}^N, \delta_f)| \\ &\leq W_1(\widehat{f^{\otimes N}}, \delta_f) = \Omega_0(f^{\otimes N}; f) \\ &\leq \frac{C M_k(f)^{1/k}}{N^\gamma}, \end{aligned}$$

where we have used the definition of  $\Omega_N$ ,  $\Omega_0$ , the triangular inequality, Proposition 2.11 and (2.25).  $\square$

### 3. ENTROPY CHAOS AND FISHER INFORMATION CHAOS

In this section  $E \subset \mathbb{R}^d$  stands for an open set or the adherence of an open space (so that the gradient of a function on  $E$  is well defined).

**3.1. Entropy chaos.** The entropy of a probability on a compact subset of  $\mathbb{R}^d$  with density  $f dx$  is well defined by the formula  $\int f \ln f$ . On a (possibly) unbounded set  $E$ , we have to be more careful because the entropy may not be defined for probability decreasing too slowly at infinity. This is a well known issue, but we present here a rigorous definition for probability  $F \in \mathbf{P}(E^j)$  having a finite moment  $M_k$  for some  $k > 0$ . It will be useful in the section 5 where we define level 3 entropy on  $\mathbf{P}(\mathbf{P}(E))$ .

We emphasize that in the sequel we shall use the same notation  $F$  for a probability and its density  $F dx$  with respect to the Lebesgue measure, when the last quantity exists. For any  $k > 0$  and  $F \in \mathbf{P}_k(E^j) \cap L^1$ , we define the (opposite of the Boltzmann's) entropy

$$(3.1) \quad H_j(\rho) = H_j(F) := \int_{E^j} F \log F = \int_E h(F/G_k^j) G_k^j + \int_E F \log G_k^j \quad (=: H_j^{(1)}(F))$$

with  $G_k^j(V) := c^j \exp(-|v_1|^k - \dots - |v_j|^k) \in \mathbf{P}(E^j)$ ,  $k > 0$  and  $c = c(k)$  chosen so that  $G_k^j$  is a probability, and  $h(s) := s \log s - s + 1$ . The RHS term is well defined in  $\mathbb{R} \cup \{+\infty\}$  as the sum of a nonnegative term and a finite real number, and it can be checked that it is equal to the middle term, which has thus a sense. Next, we extend the entropy functional to any  $F \in \mathbf{P}_k(E^j)$  by setting

$$(3.2) \quad H_j(F) := \sup_{\phi_j \in \mathcal{C}_b(E^j)} \left\{ \langle F, \phi_j \rangle - H^*(\phi_j) \right\} + \int_E F \log G_k^j \quad (=: H^{(2)}(\rho))$$

where

$$H^*(\phi_j) := \int_{E^j} h^*(\phi_j) G_j$$



and where  $h^*(t) := e^t - 1$  is the Legendre transform of  $h$ . Finally, we define the normalized entropy functional  $H$  by

$$(3.3) \quad \forall F \in \mathbf{P}_k(E^j) \quad H(F) := \frac{1}{j} H_j(F).$$

We start recalling without proof a very classical result concerning the entropy.

**Lemma 3.1.** *Let us fix  $k > 0$ . The entropy functional  $\mathbf{P}_k(E) \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $\rho \mapsto H_j(\rho)$  is well defined by the expression (3.2) is convex and is lsc for the following notion of converging sequences:  $\rho_n \rightarrow \rho$  in the weak sense of measures in  $\mathbf{P}(E)$  and  $\langle \rho_n, |v|^m \rangle$  is bounded for some  $m > k$  (the same holds of course for  $H$ ). Moreover,  $H_j(F)$  does not depend on the choice of  $k$  used in the expression (3.2),*

$$H(F) \geq \log c(k) - M_k(f) \quad \forall F \in \mathbf{P}_k(E),$$

and  $H(F) < \infty$  iff  $F \in L^1$ ,  $F \log F \in L^1(E)$ , and then  $H(F) = H^{(1)}(F)$ .

We also recall the definition of the (non-normalized) relative entropy between two probabilities  $\rho$  and  $\eta$  of  $\mathbf{P}(E^j)$  :

$$(3.4) \quad H_j(\rho|\eta) := \int_{E^j} \ln \left( \frac{d\rho}{d\eta} \right) d\rho = \int_{E^j} (g \ln g + 1 - g) d\eta$$

with  $g = \frac{d\rho}{d\eta}$  if  $\rho$  is absolutely continuous with respect to  $\eta$ . If  $g$  is not defined, then  $H_j(\rho|\eta) := +\infty$ . The associated normalized quantity is simply  $H(\rho|\eta) := \frac{1}{j} H_j(\rho|\eta)$ . The relative entropy can also be defined using a dual formula similar to (3.2). Let us remark that the relative entropy is defined without moment assumption since the quantity under the last integral is positive. For a fixed  $\eta$  it has the same properties than the entropy.

We now give two elementary and well known results which are fundamental for the analysis of the entropy defined on space product.

**Lemma 3.2.** *On  $\mathbf{P}_m(E^j)$ ,  $m > 0$ , the entropy satisfies the identity*

$$(3.5) \quad \forall f \in \mathbf{P}_m(E) \quad H(f^{\otimes j}) = H(f).$$

PROOF OF LEMMA 3.2. If  $f \in \mathbf{P}_m(E)$  is a function such that  $H(f) < \infty$ , then we may use (3.1) as a definition and

$$H(f^{\otimes j}) = \frac{1}{j} \int_{E^j} f^{\otimes j} \log f^{\otimes j} = \int_{E^j} f(v_1) \log f^{\otimes j}(v_1, \dots, v_j) = H_1(f).$$

In the contrary,  $H_1(f) = \infty$  and  $H_j(f^{\otimes j}) = \infty$ . □

**Lemma 3.3.** (i) *For any functions  $f, g \in L_m^1(E) \cap \mathbf{P}(E)$ ,  $m > 0$ , there holds*

$$(3.6) \quad H(f) := \int_E f \log f \geq \int_E f \log g, \quad \text{or} \quad H(f|g) := \int_E f \log(f/g) \geq 0,$$

with equality only if  $f = g$  a.e..

(ii) *More generally, for any nonnegative functions  $f, g \in L_m^1(E)$ ,  $m > 0$ , there holds*

$$\int_E f \log \frac{f}{g} \geq F \log \frac{F}{G}, \quad \text{avec} \quad F := \int_E f, \quad G := \int_E g.$$

(iii) *A consequence of (i) is that if  $F \in \mathbf{P}(E^j)$  has first marginal  $f$  with  $H(f) < +\infty$ , then*

$$H(F) \geq H(f) \quad \text{with equality only if } F = f^{\otimes j} \text{ a.e..}$$

(iv) The entropy is superadditive: for any  $F \in \mathbf{P}_m(E^{i+j}) \cap \mathbf{P}_{sym}(E^{i+j})$ ,  $i, j \in \mathbb{N}^*$ ,  $m > 0$ , the following inequality holds

$$(3.7) \quad H_{i+j}(F_{i+j}) \geq H_i(F_i) + H_j(F_j), \quad (\text{non-normalized entropy}),$$

where  $F_k$  as usual stands for the  $k$ -th marginal of  $F$ .

PROOF OF LEMMA 3.3. (i) To obtain the inequality, write  $H(f|g) = \int h(f/g)f$  and use the fact that  $h(s) = s \log s - s + 1$  is a non-negative function. Next there is equality only if  $h(f/g) = 0$  a.e. on  $\{f > 0\}$ . Since  $h$  vanishes only at  $s = 1$ , it means that  $f = g$  a.e. on  $\{f > 0\}$ . And from the fact that  $\int f = \int g = 1$ , we obtain the claimed equality.

(ii) We write

$$\int_E f \log \frac{f}{g} = F \int_E f/F \log \frac{f/F}{g/G} + \int_E f \log \frac{F}{G},$$

the first term is nonnegative thanks to (3.6) and the second term is the one which appears on the RHS of the claimed inequality.

(iii) We use the first inequality (3.6) on  $E^j$  with  $F$  and  $f^{\otimes j}$

$$H(F) = \frac{1}{j} \int_{E^j} F \log F \geq \frac{1}{j} \int_{E^j} F \log f^{\otimes j} = \int_{E^j} F(V) \log f(v_1) dV = H(f).$$

Using again the point *i*), we see that equality can occur only if  $F = f^{\otimes j}$  a.e..

(iv) Denote  $h_\ell := H_\ell(F_\ell)$ . If  $h_{i+j} = +\infty$  there is nothing to prove. Otherwise, we have  $h_{i+j} < \infty$  which in turn implies  $F \in L^1(E^{i+j})$ , then  $F_i \in L^1(E^i)$ ,  $F_j \in L^1(E^j)$ , so that the entropy may be defined thanks to (3.1). In  $\mathbb{R} \cup \{-\infty\}$ , we compute

$$\begin{aligned} h_{i+j} - h_i - h_j &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} \\ &\quad - \int_{E^{i+j}} F_{i+j} \log F_i(v_1, \dots, v_i) - \int_{E^{i+j}} F_{i+j} \log F_j(v_{i+1}, \dots, v_{i+j}) \\ &= \int_{E^{i+j}} F_{i+j} \log F_{i+j} - \int_{E^{i+j}} F_{i+j} \log F_i \otimes F_j \geq 0, \end{aligned}$$

thanks to (3.6). □

Our first result shows that entropy chaos is a stronger notion than Kac's chaos.

**Theorem 3.4** (Entropy and chaos). *Consider  $(G^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  such that  $\langle G_1^N, |v|^m \rangle \leq a$  for some  $m, a > 0$  and  $f \in \mathbf{P}(E)$ .*

1) *If  $G_j^N \rightharpoonup F_j$  weakly in  $\mathbf{P}(E^j)$  for some given  $j \geq 1$ , then*

$$(3.8) \quad H(F_j) \leq \liminf H(G^N).$$

*In particular, when  $(G^N)$  is  $f$ -Kac's chaotic, (3.8) holds for any  $j \geq 1$  with  $F_j := f^{\otimes j}$ .*

2) *On the other way round, if  $(G^N)$  is  $f$ -entropy chaotic, then  $(G^N)$  is  $f$ -Kac's chaotic.*

PROOF OF THEOREM 3.4. *Step 1.* For any  $N \geq j$  we introduce the Euclidian decomposition  $N = nj + r$ ,  $0 \leq r \leq j - 1$ . Iterating  $n$  times the superadditivity inequality (3.7) we have

$$H_N(F^N) \geq n H_j(F_j^N) + H(F_r^N),$$

with the convention  $H(F_r^N) = 0$  when  $r = 0$ . We get (3.8) by passing to the limit in that inequality using that  $H$  is lsc and that  $H(F_r^N)$  is bounded by below thanks to Lemma 3.1 and the condition on the moment.

Step 2. We assume that  $(G^N)$  is  $f$ -entropy chaotic, that is

$$G_1^N \rightharpoonup f \text{ weakly in } \mathbf{P}(E) \quad \text{and} \quad H(G^N) \rightarrow H(f) < \infty.$$

Let us fix  $j \geq 1$ . The sequence  $(G_j^N)$  being bounded in  $\mathbf{P}_m(E^j)$ , there exists  $F_j \in \mathbf{P}(E^j)$  and a subsequence  $(G_j^{N'})$  such that  $G_j^{N'} \rightharpoonup F_j$  weakly in  $\mathbf{P}(E^j)$ . Thanks to step 1, we have

$$H(F_j) \leq \liminf H(G_j^{N'}) \leq \liminf H(G^N) = H(f) = H(f^{\otimes j}).$$

Since the first marginal of  $F_j$  is  $(F_j)_1 = \lim_{N \rightarrow +\infty} G_1^N = f$ , the third point of Lemma 3.3 gives that  $F_j = f^{\otimes j}$  a.e.. As a conclusion and because we have identified the limit, we have proved that the all sequence  $(G_j^N)$  weakly converges to  $f^{\otimes j}$ .  $\square$

**3.2. Fisher chaos.** We now establish similar results for the Fisher information functional. For an arbitrary probability  $G \in \mathbf{P}(E^j)$ , we define the normalized Fisher information by

$$(3.9) \quad I_j^{(1)}(G) := \begin{cases} \int_{E^j} \frac{|\nabla G|^2}{G} = \int_{E^j} |\nabla \ln G|^2 G \in \mathbb{R} \cup \{+\infty\} & \text{if } G \in W^{1,1}(E^j), \\ +\infty & \text{if } G \notin W^{1,1}(E^j), \end{cases}$$

For  $G \in \mathbf{P}(E^j)$ , we also give an alternative definition

$$(3.10) \quad I_j^{(2)}(G) := \sup_{\psi \in C_b^1(E^j)^d} \langle G, -\frac{|\psi|^2}{4} - \operatorname{div} \psi \rangle \in \mathbb{R} \cup \{+\infty\}.$$

**Lemma 3.5.** *for all  $j \in \mathbb{N}$ , The identity  $I_j^{(1)} = I_j^{(2)}$  holds on  $\mathbf{P}(E^j)$ , and we simply denoted by  $I_j$  the usual (non-normalized) Fisher information and by  $I = j^{-1} I_j$  the normalized Fisher information on the all set  $\mathbf{P}(E^j)$ . The functionals  $I_j$  and  $I$  are proper, convex, lsc (in the sense of the weak convergence of measures) on  $\mathbf{P}(E^j)$ .*

PROOF OF LEMMA 3.5. For the sake of simplicity, we only deal with the case  $j = 1$ . We split the proof into two steps.

Step 1. Assume that  $f \in W^{1,1}$ . Since for all  $\psi \in C_b^1(E)^d$

$$|\nabla \ln f|^2 - \nabla \ln f \cdot \psi + \frac{|\psi|^2}{4} = \left| \nabla \ln f - \frac{\psi}{2} \right|^2 \geq 0,$$

we have

$$I^{(1)}(f) = \int_E |\nabla \ln f|^2 f \geq \int_E \left( \nabla \ln f \cdot \psi - \frac{|\psi|^2}{4} \right) f.$$

For any sequence  $(\psi_n)$  of smooth functions approximating  $\nabla \ln f = \frac{\nabla f}{f}$ , we obtain that

$$(3.11) \quad \begin{aligned} I^{(1)}(f) &= \sup_{\psi \in C_b^1(E)^d} \int_E \left( \nabla \ln f \cdot \psi - \frac{|\psi|^2}{4} \right) f \\ &= \sup_{\psi \in C_b^1(E)^d} \int_E \left[ \nabla f \cdot \psi - f \frac{|\psi|^2}{4} \right] =: I^{(3)}(f). \end{aligned}$$

The remaining equality  $I^{(3)} = I^{(2)}$  is just a simple integration by parts. Remark that maximizing sequences  $(\psi_n)$  must converge (up to some subsequence) pointwise to  $2 \nabla \ln f$  a.e. on  $\{f \neq 0\}$ . We shall use that point in the sequel.

We also remark that this reformulation  $I^{(2)}$  is also exactly the one obtained when using the general Fenchel-Moreau theorem on the convex function  $(a, b) \rightarrow \frac{|b|^2}{a}$  (that is used in the integral defining  $I^{(1)}$ ).

*Step 2.* It remains to check that the equality  $I^{(1)} = I^{(2)}$  is also true on  $\mathbf{P}(E) \setminus W^{1,1}(E)$ . In other words that if  $f \notin W^{1,1}(E)$  then  $I^{(2)}(f) = +\infty$ . In what follows, we prove the contraposition :  $I^{(2)}(f) < +\infty$  implies  $f \in W^{1,1}(E)$ . Once it will be done, we will have  $I^{(1)} = I^{(2)}$  everywhere, from which follows that  $I$  is lsc in the sense of the weak convergence of measures.

Consider  $f \in \mathbf{P}(E)$  and assume  $I^{(2)}(f) < \infty$ . We deduce that for any  $\psi \in C_b^1(\mathbb{R}^d)$  and any  $t \in \mathbb{R}$

$$\int_E f \left[ -t^2 \frac{\psi^2}{4} - t \operatorname{div} \psi \right] \leq I^{(2)}(f),$$

so that by optimizing in  $t \in \mathbb{R}$  and using that  $f \in \mathbf{P}(E)$ , we get

$$\forall \psi \in C_b^1(\mathbb{R}^d) \quad \left| \int_E f \operatorname{div} \psi \right|^2 \leq 4 I^{(2)}(f) \int_E f \frac{\psi^2}{4} \leq I^{(2)}(f) \|\psi\|_{L^\infty}^2.$$

That inequality implies  $f \in BV(E)$  and  $\|\nabla f\|_{TV} \leq \sqrt{I^{(2)}(f)}$ . Using that  $f \in BV(E)$  and making an integration by part in the definition of  $I^{(2)}(f)$ , we find

$$I^{(2)}(f) = \sup_{\psi \in C_b^1(E)^d} \int_E [\nabla f \cdot \psi - f \frac{\psi^2}{4}] = I^{(3)}(f).$$

Now, for any compact subset  $K \subset E$  with zero Lebesgue measure, we may find a sequence  $\psi_\varepsilon \in C_c^1(\mathbb{R}^d)$  such that  $0 \leq \psi_\varepsilon \leq 1$ ,  $\psi_\varepsilon = 1$  on  $K$  and  $\psi_\varepsilon \rightarrow \mathbf{1}_K$  a.e., so that for any  $t > 0$  and using that  $f \in BV(E) \subset L^1(E)$ , we get

$$t \int_K |\nabla f| \leq \liminf_{\varepsilon \rightarrow 0} \int_E [|\nabla f| t \psi_\varepsilon - f t^2 \frac{\psi_\varepsilon^2}{4}] \leq I^{(3)}(f).$$

Passing to the limit  $t \rightarrow \infty$ , we deduce that  $\nabla f$  vanishes on  $K$ , which precisely means that  $\nabla f$  is a measurable function. We have proved  $f \in W^{1,1}(\mathbb{R}^d)$ .  $\square$

Similarly, we define for two measures  $\rho$  and  $\eta$  on  $E^j$  their (non-normalized) relative Fisher information  $I(\rho|\eta)$  by

$$(3.12) \quad I_j(\rho|\eta) := \int_{E^j} \frac{|\nabla g|^2}{g} d\eta = \int_{E^j} \left| \nabla \ln \frac{d\rho}{d\eta} \right|^2 d\rho,$$

where  $g = \frac{d\rho}{d\eta}$  if  $\rho$  is absolutely continuous with respect to  $\eta$ . If not,  $I_j(\rho|\eta) := +\infty$ . The associated normalized quantity is simply  $I(\rho|\eta) := \frac{1}{j} I_j(\rho|\eta)$ . For a fixed  $\eta$ , the relative Fisher information has roughly the same properties than the Fisher information. In particular, if  $\eta$  as a derivable density, we have the equality

$$(3.13) \quad I_j(\rho|\eta) = \sup_{\varphi \in C_b^1(\mathbb{R}^j)^j} \int_{\mathbb{R}^j} \left( -\varphi \cdot \frac{\nabla \eta}{\eta} - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) d\rho.$$

**Lemma 3.6.** *For any  $f \in \mathbf{P}(E)$  there holds  $I(f^{\otimes j}) = I(f)$ .*

**PROOF OF LEMMA 3.6.** If  $I(f) < \infty$  then  $f \in W^{1,1}(E)$  and also  $f^{\otimes j} \in W^{1,1}(E^j)$ . The following computation is then meaningful

$$I(f^{\otimes j}) = \frac{1}{j} \int_{E^j} \frac{|\nabla_{E^j} f^{\otimes j}|^2}{f^{\otimes j}} = \int_{E^j} \frac{|\nabla_E f|^2}{f} \otimes f^{\otimes(j-1)} = I(f).$$

Since  $I_j(f^{\otimes j}) < \infty$  implies  $f^{\otimes j} \in W^{1,1}(E^j)$  and then  $f \in W^{1,1}(E)$ , we also have  $I_j(f^{\otimes j}) = j I_1(f)$  if  $I(f) = \infty$ .  $\square$

**Lemma 3.7.** *For any  $F \in P_{\text{sym}}(E^j)$  and  $1 \leq \ell \leq j$ , then holds*

(i)  $I(F_\ell) \leq I(F)$ .

(ii) *The Fisher information is super-additive. It means that*

$$(3.14) \quad I_j(F) \geq I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell}), \quad (\text{non-normalized Fisher information}),$$

*with in the case  $I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell}) < +\infty$  equality only if  $F = F_\ell \otimes F_{j-\ell}$ .*

(iii) *If  $I(F_1) < +\infty$ , the equality  $I(F_1) = I(F)$  holds only if  $F = (F_1)^{\otimes j}$ .*

**PROOF OF LEMMA 3.7.** We give two proofs of (i), each one laying on one of the two equivalent definitions of  $I$ .

*Proof 1.* If  $I(F) = +\infty$  the conclusion is clear. Otherwise, we have  $I(F) < \infty$  and then  $F \in W^{1,1}(E^j)$ . We are able to perform the following computation

$$I(F) = \frac{1}{j} \int_{E^j} \frac{|\nabla F|^2}{F} = \int_{E^j} \frac{|\nabla_1 F|^2}{F} = \int_{E^j} u(F, \nabla_1 F).$$

Thanks to the equivalent definition  $I^{(3)}$  of the Fisher information, we have

$$\begin{aligned} I(F) &= \sup_{\psi \in C_b(E^j)} \int_{E^j} \left( \psi \cdot \nabla_1 F - F \frac{|\psi|^2}{4} \right) \\ &\geq \sup_{\psi \in C_b(E^\ell)} \int_{E^j} \left( \psi \otimes \mathbf{1}^{j-\ell} \cdot \nabla_1 F - F \frac{|\psi \otimes \mathbf{1}^{j-\ell}|^2}{4} \right) \\ &= \sup_{\psi \in C_b(E^\ell)} \int_{E^\ell} \left( \psi \cdot \nabla_1 F_\ell - F_\ell \frac{|\psi|^2}{4} \right) = \int_{E^\ell} u(F_\ell, \nabla_1 F_\ell), \end{aligned}$$

and the last term is nothing but  $I(F_\ell)$ .

*The superadditivity.* The first proof of that result seems to be the one by Carlen in [12, Theorem 3]. We sketch now another proof that uses the third formulation  $I^{(3)}$ . In fact, we recall that

$$I_j(F) = I_j^{(3)}(F) = \sup_{\psi \in C_b^1(E^j)^{jd}} \int_E [\nabla f \cdot \psi - f \frac{\psi^2}{4}]$$

where the sup is taken on all  $\psi = (\psi_1, \dots, \psi_j)$ , with all  $\psi_i : E^j \rightarrow \mathbb{R}^d$ . We will restrict the supremum over the  $\psi$  such that :

- The  $\ell$  first  $\psi_i$  depend only on  $(x_1, \dots, x_\ell)$ , with the notation  $\psi^\ell = (\psi_1, \dots, \psi_\ell)$ .
- The  $(j-\ell)$  last  $\psi_i$  depend only on  $(x_{\ell+1}, \dots, x_j)$ , with the notation  $\psi^{j-\ell} = (\psi_{\ell+1}, \dots, \psi_j)$ .

We then have the inequality

$$\begin{aligned} I_j(F) &\geq \sup_{\psi^\ell, \psi^{j-\ell}} \int_{E^j} [\nabla_\ell f \cdot \psi^\ell + \nabla_{j-\ell} f \cdot \psi^{j-\ell} - f \frac{|\psi^\ell|^2 + |\psi^{j-\ell}|^2}{4}] \\ &= \sup_{\psi^\ell \in C_b^1(E^\ell)^{\ell d}} \int_{E^\ell} [\nabla_\ell f \cdot \psi^\ell - f_\ell \frac{|\psi^\ell|^2}{4}] \\ &\quad + \sup_{\psi^{j-\ell} \in C_b^1(E^{j-\ell})^{(j-\ell)d}} \int_{E^{j-\ell}} [\nabla_{j-\ell} f \cdot \psi^{j-\ell} - f_{j-\ell} \frac{|\psi^{j-\ell}|^2}{4}] \\ &= I_\ell(F_\ell) + I_{j-\ell}(F_{j-\ell}) \end{aligned}$$

If the inequality is an equality, we use the remark made at the end of Step 1 in the proof of Lemma 3.5 : Maximizing sequences  $\psi_n^\ell$  and  $\psi_n^{j-\ell}$  for respectively  $I_\ell$  (resp.  $I_{j-\ell}$ ) should converge pointwise towards  $2 \nabla \ln f_\ell$  (resp.  $2 \nabla \ln f_{j-\ell}$ ) up to some subsequence, a.e. on

$\{f_\ell \neq 0\}$  (resp.  $\{f_{j-\ell} \neq 0\}$ ). If we have equality, we also must have  $(\psi_n^\ell, \psi_n^{j-\ell}) \rightarrow 2 \nabla \ln f$  on  $\{f \neq 0\}$ , a set that is included in  $\{f_\ell \neq 0\} \times \{f_{j-\ell} \neq 0\}$  and thus

$$\nabla \ln f = (\nabla \ln f_\ell, \nabla \ln f_{j-\ell}) = \nabla \ln(f_\ell \otimes f_{j-\ell}),$$

which implies the claimed equality since  $f$  and  $f_\ell \otimes f_{j-\ell}$  are probabilities.

*The case of equality iii).* Using recursively the superadditivity in that particular case, we get with the notation  $F_1 = f$

$$I(f) = I(F) \geq \frac{j-1}{j} I(F_{j-1}) + \frac{1}{j} I(f) \geq \frac{j-2}{j} I(F_{j-2}) + \frac{2}{j} I(f) \geq \dots \geq I(f).$$

Therefore, all the inequalities are equalities. We obtain that

$$F = F_{j-1} \otimes f = F_{j-2} \otimes f \otimes f = \dots = f^{\otimes j},$$

by applying recursively the case of equality in (3.14).  $\square$

It is well-known that for  $(f_n)$  a sequence of  $\mathbf{P}(E)$ , the conditions

$$f_n \rightharpoonup f \text{ weakly in } \mathbf{P}(E), \quad M_k(f_n) \text{ bounded, } k > 0, \quad \text{and} \quad I(f_n) \leq C$$

imply that  $H(f_n) \rightarrow H(f)$ . A natural question is whether a similar result holds for a sequence  $(F^N)$  in  $\mathbf{P}(E^N)$ . Before answering affirmatively to that question, we establish a normalized non-relative HWI inequality for a large class of sets  $E \subset \mathbb{R}^d$ . It is a variant of the famous HWI inequality of Otto-Villani that will be the cornerstone of the argument. Let us mention that its good behaviour in any dimension is of particular importance here.

**Proposition 3.8.** *Assume that  $E \subset \mathbb{R}^d$  is a bi-Lipschitz volume preserving deformation of a convex set of  $\mathbb{R}^d$ ,  $d \geq 1$ : there exists a convex subset  $E_1 \subset \mathbb{R}^d$  and a bi-lipschitz diffeomorphism  $T : E_1 \rightarrow E$  which preserves the volume (i.e. its Jacobian is always equals to 1). Then, the normalized non relative HWI inequality holds in  $E$ : there exists a constant  $C_E \in [1, \infty)$  such that*

$$(3.15) \quad \forall F^N, G^N \in \mathbf{P}(E^N) \quad H(F^N) \leq H(G^N) + C_E W_2(F^N, G^N) \sqrt{I(F^N)}.$$

*More precisely, the above inequality holds with  $C_E := \|\nabla T\|_\infty \|\nabla T^{-1}\|_\infty$  where  $\|\nabla T\|_\infty := \sup_{v \in E} \sup_{|h|_2 \leq 1} |\nabla T(v) h|_2$ .*

**PROOF OF PROPOSITION 3.8.** We proceed in three steps.

*Step 1.  $E = \mathbb{R}^d$ .* Let us first recall the famous HWI inequality of Otto-Villani. Consider  $\rho = e^{-V(x)} dx$  a probability measure on  $\mathbb{R}^D$  such that  $D^2V \geq 0$ . For any probability measures  $f_0, f_1 \in \mathbf{P}_2(\mathbb{R}^D)$ , there holds

$$(3.16) \quad H_D(f_0|\rho) \leq H_D(f_1|\rho) + \tilde{W}_2(f_0, f_1) \sqrt{I_D(f_0|\rho)},$$

where  $H_D$  and  $I_D$  stand for the (non normalized) relative entropy and relative Fisher information defined in (3.4) and (3.12) respectively, and  $\tilde{W}_2$  stands for the non normalized quadratic MKW distance in  $\mathbb{R}^D$  based on the usual Euclidian norm  $|V| = (\sum_{i=1}^D |v_i|^2)^{1/2}$  for any  $V = (v_1, \dots, v_D) \in \mathbb{R}^D$ . Inequality (??) has been proved in [48], see also [56, 57, 47, 7, 17]. We easily deduce the “non relative” inequality (3.15) from the “relative” inequality (3.16). In order to do so, we simply apply the HWI inequality (3.16) in  $\mathbb{R}^D$ ,  $D = dN$ , with respect to the Gaussian  $\gamma_\lambda(v) := (2\pi\lambda)^{-D/2} e^{-|v|^2/2\lambda}$ , and we get

$$H_D(F^N|\gamma_\lambda) \leq H_D(G^N|\gamma_\lambda) + \tilde{W}_2(F^N, G^N) \sqrt{I_D(F^N|\gamma_\lambda)}.$$

We write the relative entropy and the relative Fisher information in terms of the non-relative ones, and we get

$$\begin{aligned} H_D(F^N|\gamma_\lambda) &= H_D(F^N) - \int F^N \ln(\gamma_\lambda) = H_D(F^N) + \frac{D}{2} \log(2\pi\lambda) + \frac{M_2(F^N)}{2\lambda}, \\ I_D(F^N|\gamma_\lambda) &= \int F^N \left| \nabla \ln F^N + \frac{v}{\lambda} \right|^2 = I_D(f_0) + \frac{2}{\lambda} \int v \cdot \nabla f_0 + \frac{M_2(f_0)}{\lambda^2} \\ &= I_D(f_0) - \frac{2D}{\lambda} + \frac{M_2(f_0)}{\lambda^2}. \end{aligned}$$

Inserting this in the relative HWI inequality, simplifying the terms involving  $\log(2\pi\lambda)$ , letting  $\lambda \rightarrow +\infty$  and dividing the resulting limit by  $N$ , we obtain the claimed result.

*Step 2.  $E \subset \mathbb{R}^d$  is convex.* The proof is the same as in the case  $E = \mathbb{R}^d$  using that the HWI inequality (3.16) holds in a convex set. We have no precise reference for that last result but all the necessary arguments can be find in [57]. More precisely, [57, Chapter 20] explains that the HWI inequality (3.16) holds when the entropy is displacement convex, while it is proved in [57, Chapters 16 and 17] that the entropy on a convex set  $E$  is displacement convex, exactly as on  $\mathbb{R}^d$ .

*Step 3. General case.* We choose two absolutely continuous probabilities  $F^N$  and  $G^N$  on  $E^N$ , and defined the corresponding probabilities  $F_1^N$  and  $G_1^N$  on  $E_1^N$  by

$$F_1^N(v_1, \dots, v_N) := F^N(T(v_1), \dots, T(v_N)) = F^N \circ T^{\otimes N}(V),$$

and the same formula for  $G_1^N$ . It can be checked that  $\nabla_{v_j} F_1^N = {}^t \nabla T(v_j) \nabla_{v_j} F^N \circ T^{\otimes N}$ , so that  $|\nabla_{v_j} F_1^N| \leq \|\nabla T\|_\infty |\nabla_{v_j} F^N \circ T^{\otimes N}|$ . Turning to Fisher information, it comes

$$I(F_1^N) := \int_{E_1^N} \frac{|\nabla F_1^N|^2}{F_1^N} dV \leq \|\nabla T\|_\infty^2 \int_{E_1^N} \frac{|\nabla F^N \circ T^{\otimes N}|^2}{F^N \circ T^{\otimes N}} dV = \|\nabla T\|_\infty^2 I(F^N),$$

where we have used the fact that  $T$  preserves the volume.

For the MKW distance, remark that  $|(T^{-1})^{\otimes N}(V) - (T^{-1})^{\otimes N}(V')| \leq \|\nabla T^{-1}\|_\infty |V - V'|$ . Therefore,

$$\begin{aligned} W_2(F_1^N, G_1^N)^2 &= \inf_{\pi_1 \in \Pi(F_1^N, G_1^N)} \int |V - V'|^2 \pi_1(dV, dV') \\ &= \inf_{\pi \in \Pi(F^N, G^N)} \int |(T^{-1})^{\otimes N}(V) - (T^{-1})^{\otimes N}(V')|^2 \pi(dV, dV') \\ &\leq \|\nabla T^{-1}\|_\infty^2 \inf_{\pi \in \Pi(F^N, G^N)} \int |V - V'| \pi(dV, dV') \\ &= \|\nabla T^{-1}\|_\infty^2 W_2(F^N, G^N)^2. \end{aligned}$$

For the entropy, the preservation of volume ensures the equality  $H(F_1^N) = H(F^N)$ , and a similar one for  $G^N$ . Finally, using the HWI inequality in  $E_1$  proved in step 2 and the above properties, we get

$$\begin{aligned} H(F^N) &= H(F_1^N) \leq H(G_1^N) + \sqrt{I(F_1^N)} W_2(F_1^N, G_1^N) \\ &\leq H(G^N) + \|\nabla T\|_\infty \|\nabla T^{-1}\|_\infty \sqrt{I(F^N)} W_2(F^N, G^N), \end{aligned}$$

which is exactly the claimed result.  $\square$

Let us finally prove now our main result Theorem 1.4 which is a consequence of the characterization of the Kac's chaos in Theorem 1.2 together with Proposition 3.8.

PROOF OF THEOREM 1.4. We split the proof into two steps.

*Step 1. (i)  $\Rightarrow$  (ii).* Fix a  $j \in \mathbb{N}$ , there exists a subsequence of  $(G^N)$ , still denoted by  $(G^N)$ , and some compatible and symmetric probabilities  $F_j \in \mathbf{P}(E^j)$ , such that  $G_j^N \rightarrow F_j$  weakly in  $\mathbf{P}(E^j)$ . In particular  $F_1 = f$ . As a consequence of Lemma 3.5 and Lemma 3.7 point (i), we have

$$I(f) \leq I(F_j) \leq \liminf I(G_j^N) \leq \liminf I(G^N) = I(f).$$

Using now the third point of Lemma 3.7 we deduce  $F_j = f^{\otimes j}$ . The unicity of the limit implies that the whole sequence  $G^N$  is in fact  $f$ -Kac's chaotic.

*Step 2. (ii)  $\Rightarrow$  (iii).* We write twice the normalized non relative HWI inequality of Proposition 3.8, and get

$$|H(G^N) - H(f^{\otimes N})| \leq C_E W_2(G^N, f^{\otimes N}) \left( \sqrt{I(G^N)} + \sqrt{I(f^{\otimes N})} \right).$$

Using the previous inequalities together with the inequality of the Lemma 2.2

$$W_2(G^N, f^{\otimes N}) \leq C_E 2^{\frac{3}{2}} [M_k(G_1^N) + M_k(f)]^{1/k} W_1(G^N, f^{\otimes N})^{1/2-1/k}$$

we get (1.8) since  $M_k(f) \leq \sup M_k(G_1^N)$  and  $I(f) \leq \sup I(G^N)$ .  $\square$

#### 4. PROBABILITIES ON THE "KAC'S SPHERE"

We generalize the preceding two sections to the important case of probability measures on the "Kac's spheres". We refer to [15] where similar results are obtained to the (even more important) case of probability measures on the "Boltzmann's spheres".

##### 4.1. On uniform probability measures on the Kac's sphere as $N \rightarrow \infty$ .

**Definition 4.1.** For any  $N \in \mathbb{N}^*$  and  $r > 0$ , we denote by  $\sigma^{N,r}$  the uniform probability measure of  $\mathbb{R}^N$  carried by the sphere  $S_r^{N-1}$  defined by

$$S_r^{N-1} := \{V \in \mathbb{R}^N; |V|^2 = r^2\}.$$

We define  $\sigma^N \in \mathbf{P}(E^N)$ ,  $E = \mathbb{R}$ , the sequence  $\sigma^N := \sigma^{N,\sqrt{N}}$  of probability measures uniform on the sphere of Kac's collisions

$$\mathcal{KS}_N := S_{\sqrt{N}}^{N-1} := \{V \in \mathbb{R}^N; |V|^2 = N\}.$$

We begin with a classical and elementary lemma that we will use several time in the sequel.

**Lemma 4.2.** (i) For any  $1 \leq \ell \leq N-1$ , there holds

$$\sigma_\ell^N(V) = \left(1 - \frac{|V|^2}{N}\right)_+^{\frac{N-\ell-2}{2}} \frac{|S_1^{N-\ell-1}|}{N^{\ell/2} |S_1^{N-1}|},$$

where we recall that  $|S_1^{k-1}| = 2\pi^{k/2}/\Gamma(k/2)$ .

(ii) For any fixed  $\ell$ , the sequence  $(\sigma_\ell^N)_{N \geq N_\ell}$  is bounded in  $L^\infty$  (with  $N_\ell = \ell + 4$ ), in  $L_k^1$  for any  $k \geq 0$  (with  $N_\ell = \ell + 1$ ), and in  $H^s$  for any  $s \geq 0$  (with  $N_\ell = N(\ell, k)$  large enough).

(iii) For any function  $\varphi \in C_b(\mathbb{R}^N)$ , any  $r > 0$  and  $1 \leq \ell \leq N-1$ , there holds

$$\int_{S_r^{N-1}} \varphi(V, V') d\sigma_r^N(V, V') = \int_{B^\ell(r)} \frac{|S_{\sqrt{r^2-V^2}}^{N-\ell-1}|}{|S_r^{N-1}|} \left\{ \int_{S_{\sqrt{r^2-V^2}}^{N-\ell-1}} \varphi(V, V') d\sigma_{\sqrt{r^2-V^2}}^{N-\ell}(V') \right\} dV,$$



where  $V \in \mathbb{R}^\ell$  and  $V' \in \mathbb{R}^{N-\ell}$ . This precisely means that

$$\sigma^N(dV, dV') = \sigma_\ell^N(dV) \sigma_{\sqrt{N-|V|^2}}^{N-\ell}(dV').$$

PROOF OF LEMMA 4.2. (i) One possible definition of  $\sigma^{N,r}$  is

$$\sigma^{N,r} := \frac{1}{r^{N-1} |S_1^{N-1}|} \lim_{h \rightarrow 0} \frac{1}{h} \left( \mathbf{1}_{B^N(r+h)} - \mathbf{1}_{B^N(r)} \right), \quad B^N(\rho) := \{V \in \mathbb{R}^N; |V| \leq \rho\},$$

where the surface  $r^{N-1} |S_1^{N-1}|$  of the Sphere  $S_r^{N-1}$  stands for the normalization constant such that  $\sigma^{N,r}$  is a probability measure. For any  $\varphi \in C_b(E^\ell)$ ,  $1 \leq \ell \leq N-1$ , we compute

$$\begin{aligned} \langle \mathbf{1}_{B(\rho)}, \varphi \otimes \mathbf{1}^{N-\ell} \rangle &= \int_{\mathbb{R}^\ell} \mathbf{1}_{|V|^2 \leq \rho^2} \varphi(V) \left\{ \int_{\mathbb{R}^{N-\ell}} \mathbf{1}_{x_{\ell+1}^2 + \dots + x_N^2 \leq \rho^2 - |V|^2} dx_{\ell+1} \dots dx_N \right\} dV \\ &= \int_{\mathbb{R}^\ell} \varphi(V) \omega^{N-\ell} (\rho^2 - |V|^2)_+^{\frac{N-\ell}{2}} dV, \end{aligned}$$

where  $\omega^k = |B^k(1)|$  is the volume of the unit ball of  $\mathbb{R}^k$ . We deduce

$$\sigma_\ell^N(r) = \frac{1}{Z_{N,r}} \frac{d}{dr} \left[ \omega^{N-\ell} (r^2 - |V|^2)_+^{\frac{N-\ell}{2}} \right] = \frac{\omega^{N-\ell} (N-\ell)}{r^{N-1} |S_1^{N-1}|} r (r^2 - |V|^2)_+^{\frac{N-\ell}{2}-1}.$$

We conclude using the relation  $|S_1^{k-1}| = k \omega^k$ .

(ii) The estimates on  $\sigma_\ell^N$  are deduced from its explicit expression after some tedious but easy calculations.

(iii) We come back to the proof of (i). We set  $m = \ell$  and  $n = N - \ell$  and we write

$$\begin{aligned} \langle \sigma^{N,r}, \varphi \rangle &= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[ \int_{B^N(r+h)} \varphi - \int_{B^N(r)} \varphi \right] \\ &= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[ \int_{|v| \leq r+h} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi - \int_{|v| \leq r} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi \right] \\ &\quad + \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \left[ \int_{|v| \leq r} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi - \int_{|v| \leq r} \int_{|v'| \leq \sqrt{r^2 - |v|^2}} \varphi \right] \\ &= \frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \int_{r \leq |v| \leq r+h} \int_{|v'| \leq \sqrt{(r+h)^2 - |v|^2}} \varphi \\ &\quad + \frac{1}{Z_{N,r}} \int_{B_r^m} \lim_{h \rightarrow 0} \frac{1}{h} \left[ \int_{B^n(\sqrt{(r+h)^2 - |v|^2})} \varphi - \int_{B^n(\sqrt{r^2 - |v|^2})} \varphi \right]. \end{aligned}$$

We invert the integral and the limit on the last line using dominated convergence, since the integral on  $v'$  are bounded by  $\|\varphi\|_\infty / \sqrt{r^2 - |v|^2}$ . The first term is bounded (for any  $0 < h \leq r$ ) by

$$\frac{1}{Z_{N,r}} \lim_{h \rightarrow 0} \frac{1}{h} \int_{r \leq |v| \leq r+h} \int_{|v'| \leq \sqrt{3rh}} |\varphi| \leq C_{N,r} \|\varphi\|_{L^\infty} \lim_{h \rightarrow 0} \sqrt{h} = 0,$$

and the second term converges to

$$\int_{B^m(r)} \frac{Z_{n, \sqrt{r^2 - |v|^2}}}{Z_{m+n,r}} \left\{ \int_{S_{\sqrt{r^2 - |v|^2}}^{n-1}} \varphi(v, v') d\sigma^n_{\sqrt{r^2 - |v|^2}}(v') \right\} dv,$$

which is exactly the claimed identity.  $\square$

Let us recall the following classical result.

**Theorem 4.3.** *The sequence  $\sigma^N$  is  $\gamma$ -chaotic, where  $\gamma$  still stands for the gaussian distribution  $\gamma(dx) = (2\pi)^{-1/2} e^{-x^2/2} dx$  on  $\mathbb{R}$ , and more precisely*

$$(4.1) \quad \|\sigma_\ell^N - \gamma^{\otimes \ell}\|_{L^1} \leq 2 \frac{\ell + 3}{N - \ell - 3} \quad \text{pour tout } 1 \leq \ell \leq N - 4.$$

The fact that  $\sigma^N$  is  $\gamma$ -chaotic is sometime called ‘‘Poincaré’s Lemma’’. In fact, it should go back to Mehler [39] in 1866. Anyway, we refer to [19, 13] for a bibliographic discussion about this important result, and to [19] for a proof of estimate (4.1). We also may prove the following quantitative version of the ‘‘Poincaré’s Lemma’’.

**Theorem 4.4.** *There exists a numerical constant  $C \in (0, \infty)$  such that*

$$(4.2) \quad W_1(\sigma^N, \gamma^{\otimes N}) \leq \frac{C}{\sqrt{N}}.$$

**Remark 4.5.** *It is worth observing that it is not clear that one can deduce (4.2) from (4.1) or that the reverse implication holds. In particular, using (4.1) and Theorem 1.2 we obtain an estimate on  $W_1(\sigma^N, \gamma^{\otimes N})$  which is weaker than (4.2).*

**PROOF OF THEOREM 4.4.** There is a simple transport map from  $\gamma^{\otimes N}$  onto  $\sigma^N$  which is given by the radial projection  $P : V \mapsto \frac{V}{|V|_2}$  with the notation  $|V|_k = (N^{-1} \sum_i |v_i|^k)^{1/k}$  for any  $k > 0$  for the normalized distance of order  $k$ . The fact it is an admissible map comes from the invariance by rotation of  $\gamma^{\otimes N}$  and  $\sigma^N$ . Is it optimal? It is not obvious because  $P(V)$  is not necessary the point of  $\mathcal{KS}_N$  which is the closest to  $V \in \mathbb{R}^N$ , for the  $|\cdot|_1$  distance (for which it costs less to displace in the direction of the axis). However, it may still be optimal for rotationnal symmetry reasons, but it is less obvious. Nevertheless, it will be sufficient for our estimate. Since,

$$|P(V) - V|_1 = \left| \frac{1}{|V|_2} - 1 \right| |V|_1$$

we get as all our distances are normalized

$$\begin{aligned} W_1(\gamma^{\otimes N}, \sigma^N) &\leq \int_{\mathbb{R}^N} |P(V) - V|_1 \gamma^{\otimes N}(dV) \\ &= \int_{\mathbb{R}^N} \left| \frac{1}{|V|_2} - 1 \right| |V|_1 \gamma^{\otimes N}(dV) \\ &= \left( \int_0^{+\infty} \left| \frac{\sqrt{N}}{R} - 1 \right| R^N e^{-R^2/2} dR \right) \frac{|S^{N-1}|}{(2\pi)^{N/2}} \left( \int_{S_1^{N-1}} |V|_1 d\sigma^{N,1} \right). \end{aligned}$$

Using that  $|V|_1 \leq |V|_2$  because of the normalization, we may bound the last integral by

$$\int_{S_1^{N-1}} |V|_1 d\sigma^N \leq \int_{S_1^{N-1}} |V|_2 d\sigma^N = N^{-1/2}.$$

Remark that this integral is also equal to  $\frac{1}{\sqrt{N}} M_1(\sigma^N)$  which can be explicitated thanks to the formula for  $\sigma_1^N$  of Lemma 4.2. Using this in the previous inequality and performing the change of variable  $R = \sqrt{N}R'$ , we get

$$\begin{aligned} W_1(\gamma^{\otimes N}, \sigma^N) &\leq \frac{|S^{N-1}|}{\sqrt{N}(2\pi)^{N/2}} \int_0^{+\infty} |\sqrt{N} - R| R^{N-1} e^{-R^2/2} dR \\ &\leq \frac{|S^{N-1}| N^{\frac{N}{2}}}{(2\pi)^{N/2}} \int_0^{+\infty} |1 - R'| (R')^{N-1} e^{-NR'^2/2} dR'. \end{aligned}$$

We can simplify the prefactor, using the formula for  $|S^{N-1}|$  and Stirling's formula

$$\begin{aligned} \frac{|S^{N-1}| N^{\frac{N}{2}}}{(2\pi)^{N/2}} &= \frac{N^{\frac{N}{2}}}{\Gamma(\frac{N}{2}) 2^{N/2-1}} \\ &= \frac{\sqrt{N} e^{N/2}}{\sqrt{\pi}} [1 + O(1/N)]. \end{aligned}$$

Turning back to the transportation cost, we get

$$W_1(\gamma^{\otimes N}, \sigma^N) \leq \frac{e\sqrt{N}}{\sqrt{\pi}} [1 + O(1/N)] \int_0^{+\infty} \left( R e^{(1-R^2)/2} \right)^{N-1} e^{-R^2/2} |1-R| dR.$$

After studying the function  $g(r) = r e^{(1-r^2)/2}$ , we remark that it is strictly increasing from 0 to 1, then strictly decreasing from 1 to  $+\infty$ , that its maximum in 1 is 1, and that  $g(1+\varepsilon) = 1 - \varepsilon^2 + O(\varepsilon^3)$ . We shall also use the less sharp but exact bound

$$g(1+\varepsilon) \leq 1 - \frac{\varepsilon^2}{4}, \quad \text{for } \varepsilon \in [-\frac{1}{2}, \sqrt{2}-1].$$

We can now cut the previous integral in three parts  $\int_0^{1/2} + \int_{1/2}^{\sqrt{2}} + \int_{\sqrt{2}}^{+\infty}$ . We bound the first part by

$$\int_0^{1/2} \dots \leq \frac{1}{2} g\left(\frac{1}{2}\right)^{N-1},$$

and the third part by

$$\int_{\sqrt{2}}^{+\infty} \dots \leq g(\sqrt{2})^{N-1} \int_{\sqrt{2}}^{+\infty} e^{-r^2/2} r dr = g(\sqrt{2})^{N-1} e^{-1}.$$

For the last part, we perform the change of variable  $r = 1 + u/\sqrt{N}$ . It comes

$$\begin{aligned} \int_{1/2}^{\sqrt{2}} \dots &= \frac{1}{N} \int_{-\sqrt{N}/2}^{(\sqrt{2}-1)\sqrt{N}} g\left(1 + \frac{u}{\sqrt{N}}\right)^{N-1} |u| e^{-u/\sqrt{N} - u^2/2N} du \\ &\leq \frac{1}{N} \int_{-\sqrt{N}/2}^{(\sqrt{2}-1)\sqrt{N}} \left(1 - \frac{u^2}{4N}\right)^{N-1} |u| du \\ &\leq \frac{1}{N} (1 + O(N^{-1})) \int_{-\infty}^{+\infty} e^{-\frac{u^2}{4}} |u| du \\ &\leq \frac{4}{N} (1 + O(N^{-1})) \end{aligned}$$

Putting all together, we finally get

$$W_1(\gamma^{\otimes N}, \sigma^N) \leq \frac{C}{\sqrt{N}} (1 + O(N^{-1})) + C\sqrt{N}\lambda^N,$$

with  $\lambda = \max(g(\sqrt{2}), g(1/2)) < 0.86$ . This implies the claimed inequality.  $\square$

**4.2. Conditioned tensor products on the Kac's sphere.** We begin with a sharp version of local central limit theorem (local CLT) or Berry-Esseen type theorem which will be the cornerstone argument in this section.

**Theorem 4.6.** Consider  $g \in \mathbf{P}_3(\mathbb{R}^D) \cap L^p(\mathbb{R}^D)$ ,  $p \in (1, \infty]$ , such that

$$(4.3) \quad \int_{\mathbb{R}^D} x g(x) dx = 0, \quad \int_{\mathbb{R}^D} x \otimes x g(x) dx = Id, \quad \int_{\mathbb{R}^D} |x|^3 g(x) dx =: M_3.$$

We define the iterated and renormalized convolution by

$$(4.4) \quad g_N(x) := \sqrt{N} g^{(*N)}(\sqrt{N} x).$$

There exists an integer  $N(p)$  and a constant  $C_{BE} = C(p, k, M_3(g), \|g\|_{L^p})$  such that

$$(4.5) \quad \forall N \geq N(p) \quad \|g_N - \gamma\|_{L^\infty} \leq \frac{C_{BE}}{\sqrt{N}}.$$

**Remark 4.7.** Theorem 4.6 is a sharper but less general version of [13, Proposition 26]. The proof follows the proof of [13, Proposition 26] and uses an argument from [13, Proposition 26], see also [34]. The first local CLT have been established in the pioneer works by A. C. Berry [5] and C.-G. Esseen [22] who proved the convergence in  $\mathcal{O}(1/\sqrt{N})$  uniformly on the distribution function in dimension  $D = 1$ , see for instance [23, Theorem 5.1, Chapter XVI]. Since that time, many variants of the local CLT have been established corresponding to different regularity assumption made on the probability  $g$ , we refer the interested reader to the recent works [50], [6], [2] and the references therein.

The proof of Theorem 4.6 use the following technical lemma which proof is postponed after the proof of the Theorem.

**Lemma 4.8.** (i) Consider  $g \in \mathbf{P}_3(\mathbb{R}^D)$  satisfying (4.3). There exists  $\delta \in (0, 1)$  such that

$$\forall \xi \in B(0, \delta) \quad |\hat{g}(\xi)| \leq e^{-|\xi|^2/4}.$$

(ii) Consider  $g \in \mathbf{P}(\mathbb{R}^D) \cap L^p(\mathbb{R}^D)$ ,  $p \in (1, \infty]$ . For any  $\delta > 0$  there exists  $\kappa = \kappa(M_3(g), \|g\|_{L^p}, \delta) \in (0, 1)$  such that

$$(4.6) \quad \sup_{|\xi| \geq \delta} |\hat{g}(\xi)| \leq \kappa(\delta).$$

**PROOF OF THEOREM 4.6.** We follow closely the proof of [13, Theorem 27] which is more general but less precise, and we use a trick that we found in the proof of [24, Theorem 1]. We observe that

$$\hat{g}_N(\xi) = (\hat{g}(\xi/\sqrt{N}))^N, \quad \hat{\gamma}(\xi) = (\hat{\gamma}(\xi/\sqrt{N}))^N.$$

Because  $g \in L^1 \cap L^p$ , the Hausdorff-Young inequality implies  $\hat{g} \in L^{p'} \cap L^\infty$  with  $p' \in [1, \infty)$ , and then  $\hat{g}_N(\xi) = (\hat{g}(\xi/\sqrt{N}))^N \in L^1$  for any  $N \geq p'$ . As a consequence we may write

$$|g_N(x) - \gamma(x)| = (2\pi)^D \left| \int_{\mathbb{R}^D} (\hat{g}_N(\xi) - \hat{\gamma}(\xi)) e^{i\xi \cdot x} d\xi \right| \leq (2\pi)^D \int_{\mathbb{R}^D} |\hat{g}_N - \hat{\gamma}| d\xi.$$

We split the above integral between low and high frequencies

$$\begin{aligned} \|g_N - \gamma\|_{L^\infty} &\leq \int_{|\xi| \geq \sqrt{N}\delta} |\hat{g}_N| d\xi + \int_{|\xi| \geq \sqrt{N}\delta} |\hat{\gamma}| d\xi \\ &\quad + \int_{|\xi| < \sqrt{N}\delta} |\hat{g}_N - \hat{\gamma}| d\xi \quad (=: T_1 + T_2 + T_3). \end{aligned}$$

For the first term, we have

$$\begin{aligned} T_1 &\leq \int_{|\xi| \geq \sqrt{N}\delta} \left| \hat{g} \left( \frac{\xi}{\sqrt{N}} \right) \right|^N d\xi = N^{d/2} \int_{|\eta| \geq \delta} |\hat{g}(\eta)|^N d\eta \\ &\leq \left( \sup_{|\eta| \geq \delta} |\hat{g}(\eta)| \right)^{N-p'} N^{d/2} \int_{|\eta| \geq \delta} |\hat{g}(\eta)|^{p'} d\eta \\ &\leq \kappa(\delta)^{N-p'} N^{d/2} C_p \|g\|_{L^p}^p \end{aligned}$$

with  $\delta \in (0, 1)$  given by point (i) of Lemma 4.8,  $\kappa(\delta)$  given by point (ii) of Lemma 4.8 and  $N \geq p'$ . The second term may be estimated in the same way, and we clearly obtain that there exists a constant  $C_1 = C_1(D, p, \|g\|_{L^p})$  such that

$$(4.7) \quad T_1 + T_2 \leq \frac{C_1}{\sqrt{N}}.$$

Concerning the third term, we write

$$T_3 = \int_{|\xi| \leq \sqrt{N}\delta} \frac{|\hat{g}_N(\xi) - \hat{\gamma}_N(\xi)|}{|\xi|^3} |\xi|^3 d\xi,$$

with

$$\begin{aligned} \frac{|\hat{g}_N(\xi) - \hat{\gamma}_N(\xi)|}{|\xi|^3} &= \frac{1}{N^{3/2}} \frac{|\hat{g}(\xi/\sqrt{N})^N - \hat{\gamma}(\xi/\sqrt{N})^N|}{|\xi/\sqrt{N}|^3} \\ &= \frac{1}{N^{3/2}} \frac{|\hat{g}(\xi/\sqrt{N}) - \hat{\gamma}(\xi/\sqrt{N})|}{|\xi/\sqrt{N}|^3} \\ &\quad \times \left| \sum_{k=0}^{N-1} \hat{g}(\xi/\sqrt{N})^k \hat{\gamma}(\xi/\sqrt{N})^{N-k-1} \right|. \end{aligned}$$

Estimate (i) of Lemma 4.8 implies

$$\begin{aligned} &\left| \sum_{k=0}^{N-1} \hat{g}(\xi/\sqrt{N})^k \hat{\gamma}(\xi/\sqrt{N})^{N-k-1} \right| \\ &\leq \sum_{k=0}^{N-1} e^{-\frac{|\xi|^2}{4N}k} e^{-\frac{|\xi|^2}{2N}(N-k-1)} \leq N e^{-\frac{|\xi|^2}{4} \frac{N-1}{N}} \leq N e^{-\frac{|\xi|^2}{8}}. \end{aligned}$$

We deduce

$$\begin{aligned} T_3 &= \frac{1}{N^{3/2}} \left( \sup_{\eta} \frac{|\hat{g}(\eta) - \hat{\gamma}(\eta)|}{|\eta|^3} \right) \int_{\mathbb{R}^D} N e^{-\frac{|\xi|^2}{8}} |\xi|^3 d\xi \\ &\leq \frac{1}{N^{1/2}} (M_3(g) + M_3(\gamma)) C_{k,d}. \end{aligned}$$

We conclude by gathering the estimates on each term. □

PROOF OF LEMMA 4.8. Thanks to a Taylor expansion, we have

$$\begin{aligned} \hat{g}(\xi) &= 1 - \frac{\xi^2}{2} + \mathcal{O}(M_3(g) |\xi|^3) \\ \hat{\omega}(\xi) &= 1 - \frac{\xi^2}{4} + \mathcal{O}(|\xi|^3), \quad \omega(x) := \frac{1}{\sqrt{\pi}} e^{-x^2}, \end{aligned}$$

from which we deduce that there exists  $\delta = \delta(M_3(g)) \in (0, 1)$  small enough such that

$$\forall \xi \in B_\delta \quad |\hat{g}(\xi)| \leq 1 - \frac{3}{8} \xi^2 \leq \hat{\omega}(\xi), \quad \hat{\omega}(\xi) := e^{-\xi^2/4}.$$

That is nothing but (i). On the other hand, (ii) is a consequence of [13, Proposition 26, (iii)].  $\square$

For a given ‘‘smooth enough’’ probability  $f \in \mathbf{P}(E)$ ,  $E = \mathbb{R}$ , we define

$$Z_N(r) := \int_{S^{N-1}(r)} f^{\otimes N} d\sigma^{N,r}, \quad Z'_N(r) := \int_{S^{N-1}(r)} \frac{f^{\otimes N}}{\gamma^{\otimes N}} d\sigma^{N,r} = \frac{Z_N(r)}{\gamma^{\otimes N}(r)}.$$

We give a sharp estimate on the asymptotic behavior of  $Z'_N$  as  $N \rightarrow \infty$ .

**Theorem 4.9.** *Consider  $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$ ,  $p \in (1, \infty]$ , satisfying*

$$(4.8) \quad \int_{\mathbb{R}} f v dv = 0,$$

and define

$$(4.9) \quad E := \int_{\mathbb{R}} f |v|^2 dv, \quad \Sigma := \left( \int_{\mathbb{R}} (v^2 - E)^2 f(v) dv \right)^{1/2}.$$

Then  $Z_N(r), Z'_N(r)$  are well defined for all  $r > 0$  and there holds with the above notations

$$(4.10) \quad Z'_N(r) \alpha_N(r^2) = \frac{\sqrt{2}}{\Sigma} \alpha_N(N) \left( \exp \left\{ - \left( \frac{r^2 - N E}{\sqrt{N} \Sigma} \right)^2 \right\} + \frac{R_N(r)}{\sqrt{N}} \right)$$

where

$$\alpha_N(s) = s^{\frac{N}{2}-1} e^{-\frac{s}{2}} \quad \text{and} \quad \|R_N\|_\infty \leq C(p, \|f\|_p, M_6(f))$$

As a particular case, there holds

$$(4.11) \quad Z'_N := Z'_N(\sqrt{E N}) = \frac{\sqrt{2}}{\Sigma} \left( 1 + \mathcal{O}\left(N^{-\frac{1}{2}}\right) \right).$$

**PROOF OF THEOREM 4.9.** We follow the proof of [13, Theorem 14] but using the sharper (but less general) estimate proved in Theorem 4.6 instead of [13, Theorem 27].

Before going on, let us remark that it is not obvious that  $Z_N(f; r)$  is well defined for all  $r > 0$  under our assumption on  $f$  which is not necessarily continuous, since we are restricting  $f^{\otimes N}$  to surfaces of  $\mathbb{R}^N$ . But, in fact the product structure of  $f^{\otimes N}$  make it possible. To see this, take  $f$  and  $g$  two measurable functions equal almost everywhere, and call  $N$  the negligible set on which they differ. Then the tensor products  $f^{\otimes N}$  and  $g^{\otimes N}$  differs only on the negligible set  $\bar{N} = \cup_i \mathbb{R}^{\otimes(i-1)} \times N \times \mathbb{R}^{\otimes(N-i)}$ . It is not difficult to see that because of the particular structure of  $\bar{N}$ , the intersection of  $\bar{N} \cap S_r^{N-1}$  is also  $\sigma_r^N$ -negligible for all  $r > 0$ . Therefore  $f^{\otimes N}$  and  $g^{\otimes N}$  are equals  $\sigma_r^N$ -almost everywhere on  $S_r^{N-1}$ , and there is no ambiguity in the definition of  $Z_N(f, r)$  for all  $r > 0$ .

We now define the law  $g$  of  $v^2$  under  $f$

$$(4.12) \quad h(u) := \frac{1}{2\sqrt{u}} (f(\sqrt{u}) + f(-\sqrt{u})) \mathbf{1}_{u>0},$$

remarking that  $h \in \mathbf{P}_3(\mathbb{R}) \cap L^q(\mathbb{R})$  with  $q > 1$  as it has been shown in the proof of [13, Theorem 14]. Consider  $(\mathcal{V}_j)$  is a sequence of random variables which is i.i.d. according to  $f$ . On the one hand, the law  $s_N(du)$  of the random variable

$$S_N := \sum_{j=1}^N |\mathcal{V}_j|^2$$

can be computed by writing

$$\begin{aligned} \mathbb{E}(\varphi(S_N)) &= \int_0^\infty \varphi(r^2) |S_1^{N-1}| r^{N-1} \left( \int_{S_r^{N-1}} f^{\otimes N}(V) \sigma^{N,r}(dV) \right) dr \\ &= \int_0^\infty \varphi(u) |S_1^{N-1}| u^{\frac{N-1}{2}} \left( \int_{S_{\sqrt{u}}^{N-1}} f^{\otimes N}(V) \sigma^{N,\sqrt{u}}(dV) \right) \frac{du}{2\sqrt{u}}, \end{aligned}$$

which implies

$$s_N(du) = \frac{1}{2} |S_1^{N-1}| u^{\frac{N}{2}-1} Z_N(\sqrt{u}).$$

On the other hand, we have  $s_N = h^{(*N)}$ . Gathering these two identities, we get

$$\begin{aligned} h^{(*N)}(r^2) &= \frac{1}{2} |S_1^{N-1}| r^{N-2} Z_N(r) = \frac{\pi^{N/2}}{\Gamma(N/2)} r^{N-2} Z'_N(r) \frac{e^{-r^2/2}}{(2\pi)^{N/2}} \\ (4.13) \quad &= \frac{\alpha_N(r^2)}{\Gamma(N/2)} \frac{Z'_N(r)}{2^{N/2}}. \end{aligned}$$

Let us define  $g(u) := \Sigma h(E + \Sigma u)$ , so that  $g \in \mathbf{P}_3(\mathbb{R}) \cap L^q(\mathbb{R})$  and

$$\int_{\mathbb{R}} g(y) y dy = 0, \quad \int_{\mathbb{R}} g(y) |y|^2 dy = 1.$$

Applying Theorem 4.6 to  $g$  and using the identity  $g^{(*N)}(u) = \Sigma h^{(*N)}(N E + \Sigma u)$ , we obtain

$$(4.14) \quad \sup_{r \geq 0} \left| h^{(*N)}(r^2) - \frac{1}{\sqrt{N} \Sigma} \gamma \left( \frac{r^2 - N E}{\sqrt{N} \Sigma} \right) \right| \leq \frac{C_{BE}}{N \Sigma},$$

where  $C_{BE}$  is the constant given in Theorem 4.6 and associated to  $g$ . Gathering the Stirling formula

$$(4.15) \quad \Gamma(N/2) = \sqrt{\pi N} \alpha_N(N) 2^{-\frac{N}{2}+1} \left( 1 + \mathcal{O}(N^{-1/2}) \right),$$

with (4.13), (4.14), we obtain

$$\forall r > 0 \quad \left| \frac{\alpha_N(r^2) Z'_N(r)}{\sqrt{\pi N} \alpha_N(N) 2 (1 + \mathcal{O}(N^{-1/2}))} - \frac{1}{\sqrt{N} \Sigma \sqrt{2\pi}} \exp \left( \left( \frac{r^2 - N E}{\sqrt{N} \Sigma} \right)^2 / 2 \right) \right| \leq \frac{C_{BE}}{N \Sigma}.$$

Estimate (4.10) readily follows.  $\square$

For a given  $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$ ,  $p > 1$ , we define the corresponding sequence of “conditioned product measures” (according to the Kac’s sphere  $\mathcal{KS}_N$ ), we write  $F^N := [f^{\otimes N}]_{\mathcal{KS}_N}$ , by

$$(4.16) \quad F^N := \frac{1}{Z_N(f; \sqrt{N})} f^{\otimes N} \sigma^N.$$

We show that  $(F^N)$  is well defined for  $N$  large enough and is  $f$ -chaotic.

**Theorem 4.10.** Consider  $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$ ,  $p > 1$ , satisfying

$$(4.17) \quad \int_{\mathbb{R}} f v dv = 0 \quad \text{and} \quad \int_{\mathbb{R}} f v^2 dv = 1.$$

The sequence  $(F^N)$  of corresponding conditioned product measure is  $f$ -chaotic, more precisely

$$\Omega_\ell(F^N, f) := W_1(F_\ell^N, f^{\otimes \ell}) \leq \frac{1}{2} \|F_\ell^N - f^{\otimes \ell}\|_1 \leq \frac{C \ell^2}{\sqrt{N}}.$$

**Remark 4.11.** The  $f$ -Kac's chaoticity property of the sequence  $F^N = [f^{\otimes N}]_{\mathcal{KS}_N}$  is stated and proved for smooth densities  $f$  in the seminal article by M. Kac [30]. Next, the same chaoticity property is proved with large generality (on  $f$ ) in [13]. Theorem 4.10 is a "quantified" version of [13, Theorem 4, 9] and [30, paragraph 5].

**PROOF OF THEOREM 4.10.** As in Theorem 4.9, it is not obvious that  $F^N$  is well defined under our assumption on  $f$  which is not necessarily continuous, since we are restricting  $f^{\otimes N}$  to a surface of  $\mathbb{R}^N$ . But the argument given at the beginning of the proof of Theorem 4.9 shows in fact that the restriction of  $f^{\otimes N}$  to  $\mathcal{KS}_N$  is unambiguously defined. Since, Theorem 4.9 implies that  $Z_N(f, \sqrt{N})$  is finite and non zero for  $N$  large enough, we deduce that  $F^N$  is well defined for  $N$  large enough.

Let us fix  $\ell \geq 1$  and  $N \geq \ell + 1$ . Denoting  $V = (V_\ell, V_{\ell, N})$ , with  $V_\ell = (v_j)_{1 \leq j \leq \ell}$ ,  $V_{\ell, N} = (v_j)_{\ell+1 \leq j \leq N}$ , we write thanks to the equality (iii) of Lemma 4.2

$$F^N(dV) = \left(\frac{f}{\gamma}\right)^{\otimes \ell} (V_\ell) \frac{1}{Z'_N(\sqrt{N})} \left(\frac{f}{\gamma}\right)^{\otimes N-\ell} (V_{\ell, N}) \sigma^{N-\ell, \sqrt{N-|V|^2}}(dV_{\ell, N}) \sigma_\ell^N(V_\ell) dV_\ell,$$

so that, coming back to the notation  $V = V_\ell = (v_j)_{1 \leq j \leq \ell} \in \mathbb{R}^\ell$ , we have

$$F_\ell^N(V) = \left(\prod_{j=1}^{\ell} \frac{f(v_j)}{\gamma(v_j)}\right) \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \sigma_\ell^N(V) = \left(\prod_{j=1}^{\ell} f(v_j)\right) \theta_{N, \ell}(V),$$

if we define the quantity  $\theta_{N, \ell}$  by

$$(4.18) \quad \theta_{N, \ell}(V) := (2\pi)^{\frac{\ell}{2}} e^{-\frac{|V|^2}{2}} \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \sigma_\ell^N(V).$$

The key point is now to prove that  $\theta_{N, \ell}$  goes to 1. Recalling the Stirling formula  $\Gamma(k) = \sqrt{\frac{2\pi}{k}} \left(\frac{k}{e}\right)^k (1 + \mathcal{O}(k^{-1}))$ , we write  $\sigma_\ell^N$  as

$$\begin{aligned} \sigma_\ell^N(V) &= \frac{|S_1^{N-\ell-1}|}{|S_1^{N-1}|} \frac{(N-|V|^2)_+^{\frac{N-\ell-2}{2}}}{N^{\frac{N-2}{2}}} \\ &= \frac{\alpha_{N-\ell}(N-|V|^2)}{N^{-\frac{1}{2}} \alpha_N(N)} \frac{e^{-\frac{|V|^2}{2}}}{(2\pi)^{\frac{1}{2}}} \mathbf{1}_{|V| \leq \sqrt{N}} (1 + \mathcal{O}\left(\frac{\ell^2}{N}\right)), \end{aligned}$$



from which we deduce

$$\begin{aligned}
\theta_{N,\ell}(V) &= \frac{Z'_{N-\ell}(\sqrt{N-|V|^2})}{Z'_N(\sqrt{N})} \frac{\alpha_{N-\ell}(N-|V|^2)}{N^{-\frac{1}{2}}\alpha_N(N)} \mathbf{1}_{|V|\leq\sqrt{N}} (1 + \mathcal{O}(\frac{\ell^2}{N})) \\
&= \frac{\alpha_{N-\ell}(N-\ell)}{N^{-\frac{1}{2}}\alpha_N(N)} \frac{e^{-\left(\frac{\ell-|V|^2}{\sqrt{N-\ell}\Sigma}\right)^2} + \mathcal{O}((N-\ell)^{-1/2})}{1 + \mathcal{O}(N^{-1/2})} \mathbf{1}_{|V|\leq\sqrt{N}} (1 + \mathcal{O}(\frac{\ell^2}{N})) \\
(4.19) \quad &= \underbrace{\left( e^{-\left(\frac{\ell-|V|^2}{\sqrt{N-\ell}\Sigma}\right)^2} + \mathcal{O}((N-\ell)^{-1/2}) \right)}_{\theta_{N,\ell}^1(V)} \underbrace{\left( 1 + \mathcal{O}(\frac{\ell^2}{N}) \right) \mathbf{1}_{|V|\leq\sqrt{N}}}_{\theta_{N,\ell}^2}
\end{aligned}$$

where we have successively used (4.10), (4.11) the definition of  $\alpha_{N-\ell}(N-\ell)$ , and a calculation yielding

$$\frac{\alpha_{N-\ell}(N-\ell)}{N^{-\frac{1}{2}}\alpha_N(N)} = 1 + \mathcal{O}(\ell^2/N).$$

It implies in particular the two following estimates on  $\theta_{N,\ell}$  which will also be very useful in the proof of the next theorems

$$(4.20) \quad \theta_{N,\ell}(v) \leq C \mathbf{1}_{|V|\leq\sqrt{N}}, \quad |\theta_{N,\ell}(V) - 1| \leq \frac{C\ell^2}{N^{1/2}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V|\geq N^{1/8}}.$$

Once they are proven, the conclusion follows since from the second one

$$\begin{aligned}
\|F_\ell^N - f^{\otimes \ell}\|_1 &= \|(\theta_{N,\ell} - 1) f^{\otimes \ell}\|_1 \\
&\leq \frac{C\ell^2}{N} \|f\|_1 + \frac{C}{N^{1/2}} \|v^6 f\|_1.
\end{aligned}$$

It only remains to prove the estimates (4.20). The uniform estimate is clear from 4.19. In fact  $\|\theta_{N,\ell}^1\|_\infty$  and  $\theta_{N,\ell}^2$  are also uniformly bounded. For the second estimate, we first control

$$\begin{aligned}
|\theta_{N,\ell}^1(V) - 1| &= |\theta_{N,\ell}^1(V) - 1| \mathbf{1}_{|V|\leq N^{1/8}} + |\theta_{N,\ell}^1(V) - 1| \mathbf{1}_{|V|\geq N^{1/8}} \\
&\leq \left| 2 \left( \frac{\ell - |V|^2}{\sqrt{N-\ell}\Sigma} \right)^2 + \mathcal{O}(N^{-1/2}) \right| \mathbf{1}_{|V|\leq N^{1/8}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V|\geq N^{1/8}} \\
&\leq \frac{C\ell^2}{N^{1/2}} \mathbf{1}_{|V|\leq N^{1/8}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V|\geq N^{1/8}},
\end{aligned}$$

which implies a similar bound for  $\theta_{N,\ell}$  since

$$\begin{aligned}
|\theta_{N,\ell}(V) - 1| &\leq |\theta_{N,\ell}^2(V)| |\theta_{N,\ell}^1(V) - 1| + |\theta_{N,\ell}^2(V) - 1| \\
&\leq C |\theta_{N,\ell}^1(V) - 1| + C \frac{\ell^2}{N} \\
&\leq \frac{C\ell^2}{N^{1/2}} + C \frac{|V|^4}{N^{1/2}} \mathbf{1}_{|V|\geq N^{1/8}}.
\end{aligned}$$

This concludes the proof.  $\square$

**4.3. Improved chaos for conditioned tensor products on the Kac's sphere.** We define the notion of Entropy and Fisher information chaos for sequences. We present the relationship between chaos and entropic chaos in the Kac's sphere framework. We first recall the definition of these two notions which involves the normalized entropy and Fisher information with respect to  $\sigma^N$ .

**Definition 4.12.** *We say that a sequence  $(G^N)$  of  $\mathbf{P}(\mathcal{KS}_N)$  is*

*i)  $f$ -entropy chaotic if  $G_1^N \rightharpoonup f$  and*

$$H(G^N|\sigma^N) \rightarrow H(f|\gamma),$$

*ii)  $f$ -Fisher (information) chaotic if  $G_1^N \rightharpoonup f$  and*

$$I(G^N|\sigma^N) \rightarrow I(f|\gamma).$$

It is worth emphasizing again that our definition is slightly different (weaker) than the corresponding definition in [13]. But they are in fact equivalent as we shall see in the next section (Theorem 4.19).

**Theorem 4.13.** *For any  $f \in \mathbf{P}_6(\mathbb{R}) \cap L^p(\mathbb{R})$ ,  $p > 1$ , satisfying the moment assumptions (4.17) of Theorem 4.10, the corresponding conditioned product sequence of measures  $(F^N)$  defined by (4.16) is  $f$ -entropy chaotic. More precisely, there exists  $C = C(p, \|f\|_{L^p}, M_6(f))$  such that*

$$(4.21) \quad |H(F^N|\sigma^N) - H(f|\gamma)| \leq \frac{C}{\sqrt{N}}.$$

PROOF OF THEOREM 4.13. With the notation  $F^N := [f^{\otimes N}]_{\mathcal{KS}_N}$ , we write for any  $N \geq 1$

$$\begin{aligned} H(F^N|\sigma^N) &= \frac{1}{N} \int_{\mathcal{KS}_N} \left( \log \frac{f^{\otimes N}}{Z'_N(f) \gamma^{\otimes N}} \right) dF^N \\ &= \int_{\mathbb{R}} \left( \log \frac{f}{\gamma} \right) F_1^N - \frac{1}{N} \log Z'_N(f). \end{aligned}$$

Thanks to the bound (4.10) on  $Z'_N(f)$  which implies that  $(Z'_N(f))$  is bounded, we deduce

$$H(F^N|\sigma^N) = \int_{\mathbb{R}} F_1^N \left( \log \frac{f}{\gamma} \right) + \mathcal{O}(1/N).$$

Recalling the notation  $\theta_N := \theta_{N,1}$  defined in (4.18) and the estimates (4.20) it satisfies, we may then write

$$H(F^N|\sigma^N) = H(f|\gamma) + \underbrace{\int_{\mathbb{R}} (\theta_N - 1) f \left( \log \frac{f}{\gamma} \right)}_{=:T} + \mathcal{O}(1/N),$$

with

$$|T| \leq C \underbrace{\int_{\mathbb{R}} |\theta_N - 1| f (1 + |v|^2) dv}_{=:T_1} + \underbrace{\int_{\mathbb{R}} |\theta_N - 1| f |\log f| dv}_{=:T_2}.$$

In order to deal with  $T_1$ , we use the second estimate of (4.20) and get

$$T_1 \leq \frac{C}{N^{1/2}} \int_{\mathbb{R}^d} \langle v \rangle^2 f dv + \frac{C}{N^{1/2}} \int_{\mathbb{R}^d} \langle v \rangle^6 f dv = \frac{C}{N^{1/2}}.$$

In order to deal with  $T_2$ , we make the more sophisticated (but standard) splitting: for any  $N, R, M \geq 1$ , we write

$$\begin{aligned} T_2 &\leq \int_{B_R} |\theta_N - 1| f |\log f| + C_\theta \int_{B_R^c} f |\log f| \\ &\leq \sup_{B_R} |\theta_N - 1| C_f + C_\theta \int_{B_R^c} f (\log f)_+ \mathbf{1}_{f \geq M} + C_\theta \int_{B_R^c} f (\log f)_+ \mathbf{1}_{M \geq f \geq 1} \\ &\quad + C_\theta \int_{B_R^c} f (\log f)_- \mathbf{1}_{1 \geq f \geq e^{-|v|^2}} + C_\theta \int_{B_R^c} f (\log f)_- \mathbf{1}_{e^{-|v|^2} \geq f \geq 0}. \end{aligned}$$

For the second term, we write  $f (\log f)_+ \leq f^{(1+p)/2} \leq f^p / M^{(p-1)/2}$  on  $\{f \geq M\}$ . For the third term, we write  $f (\log f)_+ \leq f \log M \leq f (\log M) |v|^6 / R^6$  on  $\{f \leq M, |v| \geq R\}$ . For the fourth term, we write  $\log f \geq -|v|^2$  on  $\{f \geq \exp(-|v|^2)\}$ , and thus  $f (\log f)_- \leq f |v|^2 \leq f |v|^6 / R^4$  on  $\{1 \geq f \geq e^{-|v|^2}, |v| \geq R\}$ . For the last term, we write  $f (\log f)_- \leq 4\sqrt{f}$  on  $\{0 \leq f \leq 1\}$ , and thus  $f (\log f)_- \leq 4e^{-|v|^2/2}$  on  $\{e^{-|v|^2} \geq f \geq 0, |v| \geq R\}$ . We deduce

$$\begin{aligned} T_2 &\leq C_f \sup_{B_R} |\theta_N - 1| + C_\theta \left( \frac{1}{M^{(p-1)/2}} + \frac{(\log M)_+}{R^6} + \frac{1}{R^4} + e^{-R} \right) \\ &\leq \frac{C(\|f\|_p, M_6(f))}{N^{1/2}}, \end{aligned}$$

with the choice  $R = N^{1/8}$  (which allows to use the second estimate of (4.20)), and then  $M^{(p-1)/2} = R^6$ .  $\square$

Before stating a similar result with the Fisher information, we introduce a notation : the gradient on the Kac's sphere  $\mathcal{KS}_N$  will be denoted by  $\nabla_\sigma$

$$\nabla_\sigma F(V) := P_{V^\perp} \nabla F(V) = \left( Id - \frac{V \otimes V}{|V|^2} \right) \nabla F(V) = \nabla F(V) - \frac{V \cdot \nabla F(V)}{N} V,$$

if  $F$  is a smooth function on  $\mathbb{R}^N$ .  $P_{V^\perp}$  stands for the projection on the hyperplan perpendicular to  $V$ . We will use many times that

$$(4.22) \quad \nabla \left[ F \left( \frac{V}{|V|} \right) \right] = \frac{1}{|V|} P_{V^\perp} \nabla F \left( \frac{V}{|V|} \right) = \frac{1}{|V|} \nabla_\sigma F \left( \frac{V}{|V|} \right).$$

**Theorem 4.14.** *For any  $f \in \mathbf{P}_6(\mathbb{R})$ , satisfying the moment assumptions (4.17) of Theorem 4.10, the corresponding conditioned product sequence of measures  $(F^N)$  defined by (4.16) satisfies*

$$\sup_{N \in \mathbb{N}} I(F^N | \sigma_N) < +\infty$$

if  $I(f) < +\infty$ . If moreover

$$\int_{\mathbb{R}} \frac{f'(v)^2}{f(v)} \langle v \rangle^2 dv < +\infty,$$

the sequence  $F^N$  is Fisher information chaotic.

**PROOF OF THEOREM 4.14.** We only proof the second point. The first point (boundness of the Fisher information) can be deduced from the above proof. It suffices in fact to use the simple bound  $|\nabla_\sigma G| \leq |\nabla G|$  instead of equality (4.23).

Remark also since  $E = \mathbb{R}$ , the bound on the Fisher information implies that  $f$  is bounded. Therefore, the  $L^p$  (for  $p > 1$ ) assumption which is necessary in theorem 4.10 is implied by our bounds on Fisher information. We can therefore apply the estimates (4.20)

on the quantity  $\theta_{N,i}$  for  $i = 1, 2$  defined in (4.18). They implies in particular that  $\|\theta_{N,i}\|_\infty$  is uniformly bounded and that  $\theta_{N,i}$  converges point-wise towards 1. We start with the formula

$$I(F^N | \sigma^N) = \frac{1}{N} \int_{\mathcal{K}_{S_N}} |\nabla_\sigma \ln F^N|^2 F^N(dV) = \frac{1}{N} \int_{\mathcal{K}_{S_N}} \left| \nabla_\sigma \ln \left( \frac{f^{\otimes N}}{\gamma^{\otimes N}} \right) \right|^2 F^N(dV),$$

As  $\nabla_\sigma$  is the projection on the Kac's sphere of the usual gradient, we have from (4.22) for any function  $G$  on  $\mathbb{R}^N$

$$(4.23) \quad |\nabla_\sigma G(V)|^2 = |\nabla G(V)|^2 - \frac{1}{N} |V \cdot \nabla G(V)|^2.$$

Using this with  $G = \ln \left( \frac{f^{\otimes N}}{\gamma^{\otimes N}} \right)$  in the previous calculation, it comes

$$I(F^N | \sigma^N) = \frac{1}{N} \int_{\mathcal{K}_{S_N}} \left| \nabla \ln \left( \frac{f^{\otimes N}}{\gamma^{\otimes N}} \right) \right|^2 F^N(dV) - \frac{1}{N^2} \int_{\mathcal{K}_{S_N}} \left| V \cdot \nabla \ln \left( \frac{f^{\otimes N}}{\gamma^{\otimes N}} \right) \right|^2 F^N(dV).$$

Recalling that  $F_1^N = f \theta_{N,1}$  from (4.18), by symmetry, the first term in the right hand side is equals to

$$(4.24) \quad \int_{\mathbb{R}} \left| \partial_v \ln \frac{f(v)}{\gamma(v)} \right|^2 F_1^N(dv) = I(f|\gamma) + \int_{\mathbb{R}} \left| \frac{\nabla f(v)}{f(v)} + v \right|^2 (\theta_{N,1}(v) - 1) f(v) dv.$$

The last term goes to zero from the hypothesis on  $f$ , the uniform bound  $|\theta_{N,1}| \leq C$  and the pointwise convergence of  $\theta_{N,1}$  to 1. To handle the second term in the RHS of (4.24), we compute

$$\begin{aligned} \frac{1}{N^2} \left| V \cdot \nabla \ln \left( \frac{f}{\gamma} \right)^{\otimes N} \right|^2 &= \frac{1}{N^2} \left( \sum_{i=1}^N v_i \left[ \ln \frac{f}{\gamma} \right]'(v_i) \right)^2 \\ &= \frac{1}{N^2} \sum_{i=1}^N v_i^2 \left( \left[ \ln \frac{f}{\gamma} \right]'(v_i) \right)^2 + \frac{1}{N^2} \sum_{i \neq j} v_i v_j \left[ \ln \frac{f}{\gamma} \right]'(v_i) \left[ \ln \frac{f}{\gamma} \right]'(v_j). \end{aligned}$$

After integration, it comes thanks to the symmetry of  $F^N$

$$\begin{aligned} \frac{1}{N^2} \int_{\mathcal{K}_{S_N}} \left| V \cdot \nabla \ln \left( \frac{f}{\gamma} \right)^{\otimes N} \right|^2 F^N(dV) &= \frac{1}{N} \int_{\mathbb{R}} v^2 \left( \left[ \ln \frac{f}{\gamma} \right]'(v) \right)^2 F_1^N(dv) \\ &\quad + \frac{N-1}{N} \int_{\mathbb{R}} v_1 v_2 \left[ \ln \frac{f}{\gamma} \right]'(v_1) \left[ \ln \frac{f}{\gamma} \right]'(v_2) F_2^N(dv_1, dv_2). \end{aligned}$$

Using the uniform bound  $F_1^N(v) = \theta_{N,1}(v) f(v) \leq C f(v)$ , and the hypothesis on  $f$ , we obtain that the first term of the r.h.s. is bounded by  $\frac{C}{N}$ . The second denoted by  $R_2(N)$  is equals to

$$\begin{aligned} R_2(N) &= \frac{N-1}{N} \int_{\mathbb{R}^2} v_1 v_2 \left[ \ln \frac{f}{\gamma} \right]'(v_1) \left[ \ln \frac{f}{\gamma} \right]'(v_2) f(v_1) f(v_2) dv_1 dv_2 \\ &\quad + \frac{N-1}{N} \int_{\mathbb{R}^2} v_1 v_2 \left[ \ln \frac{f}{\gamma} \right]'(v_1) \left[ \ln \frac{f}{\gamma} \right]'(v_2) (\theta_{N,2}(v_1, v_2) - 1) f(v_1) f(v_2) dv_1 dv_2 \\ &= \frac{N-1}{N} \left( \int_{\mathbb{R}} (v f'(v) + v^2 f(v)) dv \right)^2 + R_3(N) = R_3(N), \end{aligned}$$

after an integration by parts and because of the equality  $\int v^2 df = 1$ . The term  $R_3(N)$  goes to zero by dominated convergence since

$$\begin{aligned} \int_{\mathbb{R}^2} v_1 v_2 \left| \left[ \ln \frac{f}{\gamma} \right]'(v_1) \left[ \ln \frac{f}{\gamma} \right]'(v_2) \right| f(v_1) f(v_2) dv_1 dv_2 &= \left( \int_{\mathbb{R}} v \left| \left[ \ln \frac{f}{\gamma} \right]'(v) \right| f(v) dv \right)^2 \\ &\leq I(f|\gamma) \int_{\mathbb{R}} v^2 df = I(f|\gamma) = I(f) - 1, \end{aligned}$$

where the last equality follows from a simple integration by parts and the moment properties of  $f$ . This concludes the proof.  $\square$

**4.4. Chaos for arbitrary sequence of probabilities on the Kac's sphere.** We begin that section with a result which is the analog for probabilities on the Kac's sphere of the lower semi continuity of the Entropy and Fisher information.

**Theorem 4.15.** *For any sequence  $(G^N)$  of  $\mathbf{P}(\mathcal{K}\mathcal{S}_N)$  such that  $G_j^N \rightharpoonup G_j$  weakly in  $\mathbf{P}(E^j)$ , there holds*

$$H(G_j|\gamma^{\otimes j}) \leq \liminf H(G^N|\sigma^N), \quad I(G_j|\gamma^{\otimes j}) \leq \liminf I(G^N|\sigma^N).$$

For the proof, we shall need the following integration by parts formula on the Kac' sphere, which proof is postponed to the end the proof of Theorem 4.15 .

**Lemma 4.16.** *Assume that  $F$  (resp.  $\Phi$ ) is a function (resp. vector field in  $\mathbb{R}^N$ ) on on the Kac's sphere  $\mathcal{K}\mathcal{S}_N$  with integrable gradient. Then the following integration by part formula holds*

$$(4.25) \quad \int_{\mathcal{K}\mathcal{S}_N} \left[ \nabla_{\sigma} F(V) \cdot \Phi(V) + F(V) \operatorname{div}_{\sigma} \Phi(V) - \frac{N-1}{N} F(V) \Phi(V) \cdot V \right] d\sigma^N(V) = 0$$

where  $\operatorname{div}_{\sigma}$  stands for the divergence on the sphere, given by

$$\operatorname{div}_{\sigma} \Phi(V) := \sum_{i=1}^N \nabla_{\sigma} \Phi_i(V) \cdot e_i = \operatorname{div} \Phi(V) - \sum_{i=1}^N \frac{V \cdot \nabla \Phi_i(V)}{|V|^2} v_i$$

where the last formula is useful only if  $\Phi$  is defined on a neighborhood of the sphere.

PROOF OF THEOREM 4.15.

We refer to [13, Theorem 17] for a proof of the inequality involving the entropy and give only the proof of the second inequality, which in fact relies on the characterization  $I^{(3)}$  of the Fisher information. Precisely, the previous Lemma 4.16 can be used to get a reformulation of the Fisher information on the sphere information relative to  $\sigma^N$

$$\begin{aligned} I_N(G^N|\sigma^N) &:= \int_{\mathcal{K}\mathcal{S}_N} |\nabla_{\sigma} \ln G^N|^2 G^N(dV) = \sup_{\Phi \in C_b^1(\mathbb{R}^N)^N} \int_{\mathcal{K}\mathcal{S}_N} \left( \nabla \ln G^N \cdot \Phi - \frac{|\Phi|^2}{4} \right) G^N \\ (4.26) \quad &= \sup_{\Phi \in C_b^1(\mathbb{R}^N)^N} \int_{\mathcal{K}\mathcal{S}_N} \left( \frac{N-1}{N} \Phi(V) \cdot V - \operatorname{div}_{\sigma} \Phi(V) - \frac{|\Phi(V)|^2}{4} \right) G^N(dV). \end{aligned}$$

Next applying the equality (3.13) to the probability  $\gamma^{\otimes j}$ , we get that for or any  $\varepsilon > 0$ , we can choose a  $\varphi \in C_b^1(\mathbb{R}^j)^j$  such that

$$\frac{1}{j} I_j(F^j|\gamma^{\otimes j}) - \varepsilon \leq \frac{1}{j} \int_{\mathbb{R}^j} \left( \varphi \cdot V_j - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) F^j(dV_j).$$

Remark that the r.h.s. is quite similar to (4.26). With the notation  $N = nj + r$ ,  $0 \leq r < j$  and  $V_N = (V_{j,1}, \dots, V_{j,n}, V_r)$ , we define

$$\Phi(V_N) := (\varphi(V_{j,1}, \dots, \varphi(V_{j,n}), 0) \in C_b^1(\mathbb{R}^N)^N,$$

and use it in the equality (4.26). We get

$$\begin{aligned} \frac{1}{N} I(G^N | \sigma^N) &\geq \frac{1}{N} \int_{\mathcal{K}_{S_N}} \left( \frac{N-1}{N} \Phi(V_N) \cdot V_N - \operatorname{div}_\sigma \Phi(V_N) - \frac{|\Phi(V_N)|^2}{4} \right) G^N(dV_N) \\ &\geq \frac{n}{N} \int_{\mathbb{R}^j} \left( \frac{N-1}{N} \varphi(V_j) \cdot V_j - \operatorname{div} \varphi(V_j) - \frac{|\varphi(V_j)|^2}{4} \right) G_j^N(dV_j) + \frac{R_\varphi(N)}{N}, \end{aligned}$$

where

$$\begin{aligned} R_\varphi(N) &= \frac{1}{N} \int \left( \sum_{i=1}^N [V \cdot \nabla \Phi_i(V)] v_i \right) G^N(dV_N) = \frac{1}{N} \int \left( \sum_{i,\ell}^N \frac{\partial \Phi_i}{\partial v_\ell} v_i v_\ell \right) G^N(dV_N) \\ &= \frac{n}{N} \int \left( \sum_{i,\ell}^j \frac{\partial \varphi_i}{\partial v_\ell} v_i v_\ell \right) G_j^N(dV_j) = O(1), \end{aligned}$$

if  $\nabla \varphi$  decrease sufficiently quickly at infinity. Passing to the limit, we get

$$\liminf_{N \rightarrow +\infty} I(G^N | \sigma^N) \geq \frac{1}{j} \int_{\mathbb{R}^j} \left( \varphi \cdot V_j - \operatorname{div} \varphi - \frac{|\varphi|^2}{4} \right) F^j(dV_j) \geq I(F^j | \gamma^{\otimes j}) - \varepsilon$$

which concludes the proof.  $\square$

PROOF OF LEMMA 4.16 As before, we will use the normalized norm  $|V|_2 := \sqrt{\frac{1}{N} \sum v_i^2}$ . Choosing any smooth function  $q$  on  $(0, +\infty)$  with compact support, we define

$$w(V) := q(|V|_2) F \left( \frac{V}{|V|_2} \right) \Phi \left( \frac{V}{|V|_2} \right).$$

Its divergence is given by

$$\begin{aligned} \operatorname{div} w &= \frac{q'(|V|_2)}{N} F \left( \frac{V}{|V|_2} \right) \Phi \left( \frac{V}{|V|_2} \right) \cdot \frac{V}{|V|_2} + \frac{q(|V|_2)}{|V|_2} \nabla_\sigma F \left( \frac{V}{|V|_2} \right) \Phi \left( \frac{V}{|V|_2} \right) \cdot \frac{V}{|V|_2} \\ &\quad + \frac{q(|V|_2)}{|V|_2} F \left( \frac{V}{|V|_2} \right) \operatorname{div}_\sigma \Phi \left( \frac{V}{|V|_2} \right). \end{aligned}$$

Integrating this equality, and using polar coordinate, we get

$$\begin{aligned} 0 &= \left( \int_{\mathcal{K}_{S_N}} [\nabla_\sigma F(V) \cdot \Phi(V) + F(V) \operatorname{div}_\sigma \Phi(V)] \sigma^N(dV) \right) \left( \int_0^\infty q(r) r^{N-2} dr \right) \\ &\quad + \frac{1}{N} \left( \int_{\mathcal{K}_{S_N}} F(V) \Phi(V) \cdot V \sigma^N(dV) \right) \left( \int_0^\infty q'(r) r^{N-1} dr \right). \end{aligned}$$

Since  $\int_0^\infty q'(r) r^{N-1} dr = -(N-1) \int_0^\infty q(r) r^{N-2} dr$ , we obtain

$$\int_{\mathcal{K}_{S_N}} \left[ \nabla_\sigma F(V) \cdot \Phi(V) + F(V) \operatorname{div}_\sigma \Phi(V) - \frac{N-1}{N} F(V) \cdot \Phi(V) \cdot V \right] d\sigma^N(V) = 0,$$

which is the claimed result.  $\square$

The next theorem will be the key estimate in the proof of the analog of Theorem 1.4 on the Kac's sphere. It relies on the HWI inequality on the Kac's sphere, which allows to quantify the convergence of the relative entropy.

**Theorem 4.17.** *Consider  $(G^N)$  a sequence of  $\mathbf{P}(\mathcal{KS}_N)$  which is  $f$ -chaotic,  $f \in \mathbf{P}(E)$ . Assume furthermore that*

$$M_k(G^N)^{\frac{1}{k}} \leq K \text{ for } k \geq 6, \quad \text{and} \quad I(G^N|\sigma^N) \leq K.$$

*Then  $f$  satisfies  $M_k(f) < \infty$ ,  $I(f) < \infty$ , and  $(G^N)$  is  $f$ -entropy chaotic. More precisely, there exists  $C_1 := C_1(K)$  and for any  $\gamma_2 < \frac{1}{8} \frac{k-2}{k+1}$  a constant  $C_2(\gamma_2)$  such that*

$$|H(G^N|\sigma^N) - H(f|\gamma)| \leq C_1 \left( W_1(G^N, f^{\otimes N})^{\gamma_1} + C_2 N^{-\gamma_2} \right),$$

with  $\gamma_1 := 1/2 - 1/k$ .

The proof use the following estimate

**Theorem 4.18.** ([14, Theorem 2], [3, Theorem 2]). *For any sequence  $(G^N)$  of  $\mathbf{P}(\mathcal{KS}_N)$ , there holds*

$$\forall 1 \leq k \leq N, \quad H(G_k^N|\sigma_k^N) \leq 2H(G^N|\sigma^N) \quad \text{and} \quad I(G_k^N|\sigma_k^N) \leq 2I(G^N|\sigma^N).$$

PROOF OF THEOREM 4.17. STEP 1. Thanks to Theorem 4.18, we have

$$I(G_1^N|\sigma_1^N) \leq 2K.$$

Using the strong convergence of  $\sigma_1^N$  to  $\gamma$  stated in 4.2, we pass to the (inferior) limit and get

$$I(f|\gamma) \leq 2K \quad \text{and then} \quad I(f) \leq 2K.$$

Introducing the restriction  $F^N = f^{\otimes N}/Z(\sqrt{N})\sigma^N$  of  $f^{\otimes N}$  to  $\mathcal{KS}_N$  defined in (4.16) and using point  $i)$  of Theorem 4.14, we get

$$\sup_N I(F^N|\sigma^N) \leq C_2.$$

*Step 2.* Because the Ricci curvature of the metric space  $\mathcal{KS}_N$  is positive (it is  $K := (N - 1)/N$ ) we may use the HWI inequality in weak  $CD(K, \infty)$  geodesic space (see [57, Theorem 30.21]) which generalizes the standard HWI inequality (3.16) quoted in Proposition 3.8. However, we have to be careful, because it is no valid with  $\tilde{W}_2$  replaced by the MKW distance constructed with the geodesic distance on the sphere, and not the distance induced by the square norm of  $\mathbb{R}^N$ . Fortunately, both distance are equivalent, and if we add a constant  $\frac{\pi}{2}$  in the right hand side, we can still write the HWI inequality with our usual distance  $\tilde{W}_2$ . We then have

$$H(F^N|\sigma^N) - H(G^N|\sigma^N) \leq \frac{\pi}{2} \sqrt{I(F^N|\sigma^N)} W_2(F^N, G^N),$$

and

$$H(G^N|\sigma^N) - H(F^N|\sigma^N) \leq \frac{\pi}{2} \sqrt{I(G^N|\sigma^N)} W_2(F^N, G^N),$$

so that

$$|H(F^N|\sigma^N) - H(G^N|\sigma^N)| \leq C_2 W_2(F^N, G^N).$$

We rewrite it under the form

$$|H(G^N|\sigma^N) - H(f|\gamma)| \leq C_3 [W_2(G^N, f^{\otimes N}) + W_2(F^N, f^{\otimes N})] + |H(F^N|\sigma^N) - H(f|\gamma)|.$$

For the first term, we have using inequality of Lemma 2.2

$$W_2(G^N, f^{\otimes N}) \leq 4K W_1(G^N, f^{\otimes N})^{1/2-1/k}.$$

For the second term, we have for any  $\varepsilon > 0$

$$\begin{aligned} W_2(F^N, f^{\otimes N}) &\leq 4K \Omega_N(F^N; f)^{1/2-1/k} \\ &\leq 4K \left( \Omega_0(F^N; f) + C_\varepsilon N^{-\frac{1}{2+\varepsilon+2/k}} \right)^{1/2-1/k} \\ &\leq C_\varepsilon \left( \Omega_2(F^N; f)^{\frac{1}{2+\varepsilon+1/k}} + C_\varepsilon N^{-\frac{1}{2+\varepsilon+2/k}} \right)^{1/2-1/k} \\ &\leq C_\varepsilon N^{-\frac{1/4-1/2k}{2+\varepsilon+2/k}}, \end{aligned}$$

where we have successively used Lemma 2.2, the inequality (2.30), (2.24) and Theorem 4.10 in the case  $d = 1$  (and then  $d' = \max(d, 2) = 2$ ). The third and last term is bounded by  $C N^{-1/2}$  thanks to Theorem 4.13.  $\square$

The lower semi continuity properties of Theorem 4.15 and Theorem 4.17 allow us to give a variant of Theorem 1.4 in the framework of probabilities with support on the Kac's spheres.

**Theorem 4.19.** *Consider  $(G^N)$  a sequence of  $\mathbf{P}_{sym}(\mathcal{KS}_N)$  such that  $M_6(G_1^N)$  is bounded and  $G_1^N \rightharpoonup f$  weakly in  $\mathbf{P}(\mathbb{R})$ .*

*In the list of assertions below, each one implies the assertion which follows:*

- (i)  $(G^N)$  is  $f$ -Fisher information chaotic, i.e.  $I(G^N|\sigma^N) \rightarrow I(f|\gamma)$ ,  $I(f) < \infty$ ;
- (ii)  $(G^N)$  is  $f$ -Kac's chaotic and  $I(G^N|\sigma^N)$  is bounded;
- (iii)  $(G^N)$  is  $f$ -entropy chaotic, that is  $H(G^N|\sigma^N) \rightarrow H(f|\gamma)$ ,  $H(f) < \infty$ ;
- (iv)  $(G^N)$  is  $f$ -Kac's chaotic.

PROOF OF THEOREM 4.19. The proof is very similar to the one of Theorem 1.4.  $i) \Leftrightarrow ii)$  and  $iii) \Leftrightarrow iv)$  relies on the l.s.c. properties of Theorem 4.15. And  $ii) \Leftrightarrow iii)$  uses Theorem 4.17. We omit the details.  $\square$

We finally conclude this section with the proof of Theorem 1.5.

PROOF OF THEOREM 1.5. We only deal with the case  $j = 1$ , but the general case  $j \geq 1$  can be managed in a very similar way because we already know that  $G_j^N \rightharpoonup f^{\otimes j}$  weakly in  $\mathbf{P}(E^j)$  thanks to Theorem 4.17 and Theorem 4.19. With the notations of Theorem 1.5, we have to prove

$$H(G_1^N|f) = \int_E \log(G_1^N/f) G_1^N \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

First, we observe that since  $G^N$  is symmetric and has support on the Kac's sphere,  $M_2(G^N) = 1$ . Moreover,

$$\begin{aligned} I(G_1^N|\sigma_1^N) &= \int_E |\nabla \log G_1^N - \nabla \log \sigma_1^N|^2 G_1^N \\ &= I(G_1^N) + \int_E [2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2] G_1^N, \end{aligned}$$

so that

$$I(G_1^N) \leq I(G_1^N|\sigma_1^N) + \int_E (2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2)_- G_1^N.$$

We easily compute

$$\begin{aligned} 2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2 &= \\ &= \frac{N-3}{2} \left\{ 2 \frac{(2v)^2/N^2}{(1-v^2/N)^2} - 2 \frac{2/N}{(1-v^2/N)} + \frac{(2v/N)^2}{(1-v^2/N)^2} \right\} \mathbf{1}_{v^2 \leq N} \end{aligned}$$



and then

$$\begin{aligned} (2 \Delta \log \sigma_1^N + |\nabla \log \sigma_1^N|^2)_- &= 2 \frac{N-3}{N} \frac{(4v^2/N-1)_-}{(1-v^2/N)^2} \mathbf{1}_{v^2 \leq N/4} \\ &\leq 2 \frac{1}{(1-1/4)^2} = \frac{32}{9}. \end{aligned}$$

Thanks to the boundedness assumption (1.9) we get that  $I(G_1^N) \leq C$  for some constant  $C \in (0, \infty)$ , and then  $I(G_1^N | \gamma) \leq 2[I(G^N) + M_2(G^N)] \leq C$ .

Next, we introduce the splitting

$$H(G_1^N | f) = \underbrace{H(G_1^N | \gamma) - H(f | \gamma)}_{=: T_1} + \underbrace{\int_E (f - G_1^N) \log \frac{f}{\gamma}}_{=: T_2}$$

and we show that  $T_i \rightarrow 0$  for any  $i = 1, 2, 3$ . For the first term, using twice the HWI inequality we have

$$|T_1| \leq \left( \sqrt{I(G_1^N | \gamma)} + \sqrt{I(f | \gamma)} \right) W_2(G_1^N, f) \rightarrow 0$$

because of the uniform bounds on the Fisher information and on the  $k$ -th moment,  $k > 2$ , together with the weak convergence  $G_1^N \rightharpoonup f$  as  $N \rightarrow \infty$ .

Before dealing with the last term, we remark that the bound on the Fisher information of  $f$  implies some regularity, precisely that  $\sqrt{f}$  and then  $f$  are  $\frac{1}{2}$ -Hölder. Therefore  $\ln \frac{f}{\gamma}$  is continuous and satisfies from the assumption (1.10) the bound

$$\left| \ln \frac{f}{\gamma} \right| \leq \ln \|f\|_\infty + \alpha |v|^k + |\beta| + \frac{v^2}{2} \leq C \langle v \rangle^2.$$

Using, then the fact that  $M_2(f) = 1$ , the regularity of  $f$  and the weak convergence of  $G_1^N$  towards  $f$ , we can conclude that  $T_2 \rightarrow 0$ . Precisely, we choose  $\varepsilon > 0$ , and  $A > 1$  sufficiently large so that  $\int_{|v| \geq A} (1 + v^2) f(v) dv \leq \varepsilon$ . We split

$$\ln \frac{f}{\gamma} := h_1 + h_2, \quad \text{with} \quad h_1 = \ln \frac{f}{\gamma} \chi_A, \quad h_2 = \ln \frac{f}{\gamma} (1 - \chi_A),$$

where  $\chi_A$  is a smooth function such that  $\chi_A(v) = 1$  on  $[-A, A]$  and 0 outside  $[-2A, 2A]$ . Since  $h_1$  is smooth and compactly supported, it is clear that  $\int (G_1^N - f) h_1 \rightarrow 0$ . Next

$$\begin{aligned} \left| \int (G_1^N - f) h_2 \right| &\leq C \int (1 - \chi_A(v)) (1 + v^2) [G_1^N(v) + f(v)] dv \\ &\leq C\varepsilon + 2C - C \int \chi_A(v) (1 + v^2) G_1^N(v) dv \\ \liminf_{N \rightarrow \infty} \left| \int (G_1^N - f) h_2 \right| &\leq C\varepsilon + 2C - C \int \chi_A(v) (1 + v^2) f(v) dv \\ &\leq C\varepsilon + C \int (1 - \chi_A(v)) (1 + v^2) f(v) dv \leq 2C\varepsilon. \end{aligned}$$

□

## 5. ON MIXTURES ACCORDING TO DE FINETTI, HEWITT AND SAVAGE

In this section we develop a quantitative and qualitative approach concerning the sequence of probabilities of  $\mathbf{P}_{sym}(E^N)$ ,  $E \subset \mathbb{R}^d$ , in the general framework of convergence to “mixture probability” (here we do not assume chaos property).

Depending of the result, we will need some hypothesis on the set  $E$  that we will make precise in each statement. While in the first and second sections the results hold with great generality only assuming that

- $E$  is a Borel set of  $\mathbb{R}^d$ ;

we shall assume in the third section that

- $E = \mathbb{R}^d$  or  $E$  is a open set of  $\mathbb{R}^d$  with smooth boundary in order that the strong maximum principle and the Hopf lemma hold;

and we shall also assume in the third and fourth sections that

- the normalized non relative HWI inequality (3.15) holds in  $E$ .

**5.1. The De Finetti, Hewitt and Savage theorem and convergence in  $\mathbf{P}(E^N)$ .** We begin by recalling the famous De Finetti, Hewitt and Savage theorem [18, 29] for which we state a quantified version that is maybe new.

**Theorem 5.1.** *Assume  $E \subset \mathbb{R}^d$  is a Borel set. Consider a sequence  $(\pi^j)$  of symmetric and compatibles probabilities of  $\mathbf{P}(E^j)$ , that is  $\pi^j \in \mathbf{P}_{sym}(E^j)$  and  $(\pi^j)|_{E^\ell} = \pi^\ell$  for any  $1 \leq \ell \leq j$ , and consider  $(\hat{\pi}^j)$  the associated sequence of empirical distribution in  $\mathbf{P}(\mathbf{P}(E))$  defined according to (2.7). For any  $s > \frac{d}{2}$ , the sequence  $(\hat{\pi}^j)$  is a Cauchy sequence for the distance  $\mathcal{W}_{H^{-s}}$ , and precisely*

$$(5.1) \quad [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 \leq 2\|\Phi_s\|_\infty \left( \frac{1}{M} + \frac{1}{N} \right),$$

where  $\Phi_s$  is the fonction introduced in Lemma 2.7. In particular, the sequence  $(\hat{\pi}^j)$  converges towards some  $\pi \in \mathbf{P}(\mathbf{P}(E))$  with the speed  $\mathcal{W}_{H^{-s}}(\hat{\pi}^j, \pi) \leq \frac{C}{\sqrt{j}}$ . The limit  $\pi$  is characterized by the relations

$$(5.2) \quad \forall j \geq 1, \quad \pi^j = \pi_j := \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \quad \text{in } \mathbf{P}_{sym}(E^j),$$

or in other words, with the notations of section 2.1

$$(5.3) \quad \forall \varphi \in C_b(E^j) \quad \langle \pi^j, \varphi \rangle = \int_{\mathbf{P}(E)} R_\varphi(\rho) \pi(d\rho).$$

Reciprocally, for any mixture probability  $\pi \in \mathbf{P}(\mathbf{P}(E))$ , the sequence  $(\pi_j)$  of probabilities in  $\mathbf{P}(E^j)$  defined by the second identity in (5.2) is such that the  $\pi_j$  are symmetric and compatible.

PROOF OF THEOREM 5.1. We split the proof into two steps.

*Step 1.* As in proof of Proposition 2.8, we shall use the fact that  $\|\cdot\|_{H^{-s}}^2$  is a polynomial on  $\mathbf{P}(E)$ , but we have to choose a good transference plan. Fortunately, there is at least a simple choice. The compatibility and symmetry conditions on  $(\pi^N)$  tells us that  $\pi^{N+M}$  is a admissible transference between  $\pi^N$  and  $\pi^M$ . Using the symmetry of  $\pi^{N+M}$  and the isometry between  $(E^N/\mathfrak{S}_N, w_1)$  and  $(\mathbf{P}_N(E), W_1)$  stated in step 1 in the proof of Proposition 2.11, we will interpret it as a transference plan  $\tilde{\pi}^{N+M}$  on  $\mathcal{P}_N(E) \times \mathcal{P}_M(E)$  between  $\hat{\pi}^N$  and  $\hat{\pi}^M$ . More precisely,  $\tilde{\pi}^{N+M} \in \mathbf{P}(\mathbf{P}(E) \times \mathbf{P}(E))$  is defined as the probability satisfying

$$\forall \Phi \in C_b(\mathbf{P}(E) \times \mathbf{P}(E)) \quad \langle \tilde{\pi}^{N+M}, \Phi \rangle = \int_{E^N \times E^M} \Phi(\mu_X^N, \mu_Y^N) \pi^{N+M}(dX, dY).$$

We have

$$\begin{aligned} [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \|\rho - \eta\|_{H^{-s}}^2 \tilde{\pi}^{N+M}(d\rho, d\eta) \\ &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \left( \int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\rho^{\otimes 2} - \rho \otimes \eta) + (\eta^{\otimes 2} - \eta \otimes \rho)](dx, dy) \right) \tilde{\pi}^{N+M}(d\rho, d\eta), \end{aligned}$$

with the help of (2.22). We can then compute

$$\begin{aligned} &[\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 \leq \\ &\leq \int \left( \int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\mu_X^N)^{\otimes 2} - \mu_X^N \otimes \mu_Y^M](dx, dy) \right) \pi^{N+M}(dX, dY) \\ &\quad + \int \left( \int_{\mathbb{R}^{2d}} \Phi_s(x-y) [(\mu_Y^M)^{\otimes 2} - \mu_Y^M \otimes \mu_X^N](dx, dy) \right) \pi^{N+M}(dX, dY) \\ &\leq \int \left( \frac{1}{N^2} \sum_{i,j=1}^N \Phi_s(x_i - x_j) - \frac{1}{NM} \sum_{i,j'=1}^M \Phi_s(x_i - y_{j'}) \right) \pi^{N+M}(dX, dY) \\ &\quad + \int \left( \frac{1}{M^2} \sum_{i',j'=1}^M \Phi_s(y_{i'} - y_{j'}) - \frac{1}{NM} \sum_{i',j=1}^M \Phi_s(x_{i'} - y_j) \right) \pi^{N+M}(dX, dY) \\ &\leq \frac{\Phi_s(0)}{N} + \frac{N-1}{N} \int \Phi_s(x-y) \pi^2(dx, dy) - \int \Phi_s(x-y) \pi^2(dx, dy) \\ &\quad + \frac{\Phi_s(0)}{M} + \frac{M-1}{M} \int \Phi_s(x-y) \pi^2(dx, dy) - \int \Phi_s(x-y) \pi^2(dx, dy), \end{aligned}$$

and we conclude with

$$\begin{aligned} [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \hat{\pi}^M)]^2 &\leq \left( \frac{1}{M} + \frac{1}{N} \right) \left( \Phi_s(0) - \int \Phi_s(x-y) \pi^2(dx, dy) \right) \\ &\leq 2\|\Phi_s\|_\infty \left( \frac{1}{M} + \frac{1}{N} \right). \end{aligned}$$

The existence of the limit  $\pi$  is due to the completeness of  $\mathbf{P}(\mathbf{P}(E))$ .

*Step 2.* Now it remains to characterize the limit  $\pi$ . We fix  $j \in \mathbb{N}$ , we denote by  $\pi_j$  its  $j$ -th marginal defined thanks to the second identity in (5.2) and by  $\hat{\pi}_j^N = (\hat{\pi}^N)_j$  the  $j$ -th marginal of the empirical probability  $\hat{\pi}^N$  as defined in (2.8). We easily compute

$$\begin{aligned} \|\hat{\pi}_j^N - \pi_j\|_{H^{-s}}^2 &= \left\| \int_{\mathbf{P}(E)} \rho^{\otimes j} \hat{\pi}^N(d\rho) - \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \right\|_{H^{-s}}^2 \\ &= \inf_{\Pi \in \Pi(\hat{\pi}^N, \pi)} \left\| \int_{\mathbf{P}(E)} [\rho^{\otimes j} - \eta^{\otimes j}] \Pi(d\rho, d\eta) \right\|_{H^{-s}}^2 \\ &\leq \inf_{\Pi \in \Pi(\hat{\pi}^N, \pi)} \int_{\mathbf{P}(E)} \|\rho^{\otimes j} - \eta^{\otimes j}\|_{H^{-s}}^2 \Pi(d\rho, d\eta) \\ &= [\mathcal{W}_{H^{-s}}(\hat{\pi}^N, \pi)]^2 \leq \frac{C}{N}. \end{aligned}$$

Next we fix  $s > \frac{jd}{2}$ , so that using Sobolev embeddings on  $\mathbb{R}^{jd}$ ,  $\|\varphi\|_\infty \leq C\|\varphi\|_{H^s}$  for any  $\varphi \in H^s(\mathbb{R}^{jd})$ , which implies by duality that  $\|\rho\|_{H^{-s}} \leq C\|\rho\|_{TV}$  for any  $\rho \in \mathbf{P}(\mathbb{R}^{jd})$ .

Using the Grunbaum lemma 2.6 and the compatibility assumption  $\pi_j^N = \pi^j$ , we get the inequality

$$\|\pi^j - \hat{\pi}_j^N\|_{H^{-s}} = \|\pi_j^N - \hat{\pi}_j^N\|_{H^{-s}} \leq C \|\pi_j^N - \hat{\pi}_j^N\|_{TV} \leq \frac{Cj^2}{N}.$$

Combining the two previous inequalities leads to

$$\|\pi^j - \pi_j\|_{H^{-s}} \leq \|\pi^j - \hat{\pi}_j^N\|_{H^{-s}} + \|\hat{\pi}_j^N - \pi_j\|_{H^{-s}} \leq \frac{C}{\sqrt{N}} + \frac{Cj^2}{N},$$

which implies the claimed equality in the limit  $N \rightarrow +\infty$ .  $\square$

Let us now introduce some definitions. For  $k > 0$ , we define

$$\mathbf{P}_k(\mathbf{P}(E)) := \{\pi \in \mathbf{P}(\mathbf{P}(E)); M_k(\pi) := M_k(\pi_1) < \infty\}.$$

We adopt the following convention. For given sequences  $(F^N)$  of  $\mathbf{P}_{sym}(E^N)$ ,  $\pi^N$  of  $\mathbf{P}(\mathbf{P}(E))$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$ , we say

- $(F^N)$  is bounded in  $\mathbf{P}_k(E^N)$  if there exists a constant  $a > 0$  such that  $M_k(F_1^N) \leq a$ ;
- $(\pi_n)$  is bounded in  $\mathbf{P}_k(\mathbf{P}(E))$  if there exists a constant  $a > 0$  such that  $M_k(\pi_{n,1}) \leq a$ ;
- $(F^N)$  weakly converges to  $\pi$  in  $\mathbf{P}_k(E^j)_{\forall j}$  if  $(F^N)$  is bounded in  $\mathbf{P}_k(E^N)$  and  $F_j^N \rightharpoonup \pi_j$  weakly in  $\mathbf{P}(E^j)$  for any  $j \geq 1$ ;
- $(\pi_n)$  weakly converges to  $\pi$  in  $\mathbf{P}_k(\mathbf{P}(E))$  if  $(\pi_n)$  is bounded in  $\mathbf{P}_k(\mathbf{P}(E))$  and  $\pi_n \rightharpoonup \pi$  weakly in  $\mathbf{P}(\mathbf{P}(E))$ .

With that (not conventional) definitions, any bounded sequence in  $\mathbf{P}_k(\mathbf{P}(E))$  is weakly compact in  $\mathbf{P}_k(\mathbf{P}(E))$ , and for any sequence  $(F^N)$  of probabilities of  $\mathbf{P}_{sym}(E^N)$  which is bounded in  $\mathbf{P}_k(E^N)$ ,  $k > 0$ , there exists a subsequence  $(F^{N'})$  and a mixture probability  $\pi \in \mathbf{P}(\mathbf{P}(E))$  such that  $F^{N'} \rightharpoonup \pi$  in  $\mathbf{P}(E^j)_{\forall j}$ .

We now present a result about the equivalence of convergences for sequence of  $\mathbf{P}_{sym}(E^N)$ ,  $N \rightarrow \infty$ , without any chaos hypothesis.

**Theorem 5.2.** *Assume  $E \subset \mathbb{R}^d$  is a Borel set.*

**(1)** *Consider  $(F^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$ . The three following assertions are equivalent:*

- (i)  $F^N \rightharpoonup \pi$  in  $\mathbf{P}(E^j)_{\forall j}$ ;
- (ii)  $\hat{F}^N \rightharpoonup \pi$  weakly in  $\mathbf{P}(\mathbf{P}(E))$ ;
- (iii)  $W_1(F^N, \pi_N) \rightarrow 0$ .

**(2)** *For any  $\gamma \in [\frac{1}{2d}, \frac{1}{d})$  (recall that  $d' = \max(d, 2)$ ), and any  $k > \frac{d'}{\gamma^{1-d'}} \geq 1$ , there exists a constant  $C = C(\gamma, d, k)$  such that the following estimate holds*

$$(5.4) \quad \forall N \geq 1 \quad |W_1(F^N, \pi_N) - \mathcal{W}_1(\hat{F}^N, \pi)| \leq \frac{C M_k(\pi_1)^{1/k}}{N^\gamma}.$$

**(3)** *With the same notations as in the second point, we have for any mixture probabilities  $\alpha, \beta \in \mathbf{P}(\mathbf{P}(E))$*

$$(5.5) \quad \mathcal{W}_1(\hat{\alpha}_j, \alpha) \leq \frac{C M_k(\alpha_1)^{1/k}}{j^\gamma},$$

where  $\hat{\alpha}_j$  is empirical probability distribution in  $\mathbf{P}(\mathbf{P}(E))$  associated to the  $j$ -th marginal  $\alpha_j \in \mathbf{P}(E^j)$ , as well as

$$(5.6) \quad \mathcal{W}_1(\alpha, \beta) - \frac{C(M_k(\alpha_1)^{\frac{1}{k}} + M_k(\beta_1)^{\frac{1}{k}})}{j^\gamma} \leq \mathcal{W}_1(\alpha_j, \beta_j) \leq \mathcal{W}_1(\alpha, \beta).$$

PROOF OF THEOREM 5.2. *Step 1.* Equivalence between (i) and (ii) is classical. Let us just sketch the proof. For any  $\varphi \in C_b(E^j)$  we have from the Grunbaum lemma recalled in Lemma 2.6 that

$$\begin{aligned} \langle \hat{F}^N, R_\varphi \rangle &= \langle F^N, \varphi \otimes \widetilde{\mathbf{1}^{\otimes N-j}} \rangle + \mathcal{O}(j^2/N) \\ &= \langle F_j^N, \varphi \rangle + \mathcal{O}(j^2/N). \end{aligned}$$

We deduce that the convergence  $\langle \hat{F}^N, R_\varphi \rangle \rightarrow \langle \pi, R_\varphi \rangle$  is equivalent to the convergence  $\langle F_j^N, \varphi \rangle \rightarrow \langle \pi_j, \varphi \rangle$  since that  $\langle \pi, R_\varphi \rangle = \langle \pi_j, \varphi \rangle$  thanks to Theorem 5.1.

Therefore, *i*) is equivalent to the convergence  $\langle \hat{F}^N, \Phi \rangle \rightarrow \langle \pi, \Phi \rangle$  for any polynomial function  $\Phi \in C_b(\mathbf{P}(E))$ . But now, the family of probability  $\hat{F}^N$  (and  $\pi$ ) belongs to the compact subset of  $\mathbf{P}(\mathbf{P}(E))$

$$\mathcal{K} := \{\alpha \in \mathbf{P}(\mathbf{P}(E)), \text{ s.t. } \alpha_1 = F_1\},$$

and also any converging subsequence  $\hat{F}^{N'}$  should converge weakly towards a probability  $\tilde{\pi}$  having the same marginals than  $\pi$ . Since by Theorem 5.1 marginals uniquely characterize a probability on  $\mathbf{P}(\mathbf{P}(E))$ , it implies  $\tilde{\pi} = \pi$  and then weak convergence again polynomial function implies the standard weak convergence of probability ii).

It is classical that the MKW distance is a metrization of the weak convergence of measures. Even in that "abstract" case, ii) is equivalent to  $\mathcal{W}_1(\hat{F}^N, \pi) \rightarrow 0$ . Thus, for sequence having a bounded moment  $M_k(F_1^N)$  for some  $k > 0$ , the equivalence between ii) and iii) will be a consequence of (5.4). For sequence which do not possess any moment  $M_k$ , it is still true. The correct argument still relies on a version of inequality (5.4), with a slower and less explicit rate of convergence, which can be obtained from an adaptation of Lemma 2.1.

*Step 2.* We now prove (5.4). For  $\hat{\pi}_N$  we have the following representation:

$$(5.7) \quad \hat{\pi}_N = \int \widehat{\rho^{\otimes N}} \pi(d\rho) = \int \widehat{\rho^{\otimes N}} \pi(d\rho).$$

Thanks to Proposition 2.11, we may compute

$$\begin{aligned} |\mathcal{W}_1(F^N, \pi_N) - \mathcal{W}_1(\hat{F}^N, \pi)| &= |\mathcal{W}_1(\hat{F}^N, \hat{\pi}_N) - \mathcal{W}_1(\hat{F}^N, \pi)| \\ &\leq \mathcal{W}_1(\hat{\pi}_N, \pi) = \mathcal{W}_1\left(\int_{\mathbf{P}(E)} \widehat{\rho^{\otimes N}} \pi(d\rho), \int_{\mathbf{P}(E)} \delta_\rho \pi(d\rho)\right) \\ &\leq \int_{\mathbf{P}(E)} \mathcal{W}_1(\widehat{\rho^{\otimes N}}, \delta_\rho) \pi(d\rho) = \int_{\mathbf{P}(E)} \Omega_0(\rho) \pi(d\rho), \\ &\leq \frac{C(d, \gamma, k)}{N^\gamma} \int_{\mathbf{P}(E)} M_k(\rho)^{1/k} \pi(d\rho) \leq \frac{C(d, \gamma, k)}{N^\gamma} M_k(\pi_1)^{1/k}, \end{aligned}$$

where we have successively used the triangular inequality for the  $\mathcal{W}_1$  distance, the relation (5.7), the convexity property of the  $\mathcal{W}_1$  distance and the definition of the chaos measure  $\Omega_0$ . We also used the bound (2.25) and the Jensen inequality (recall that  $1/k \in (0, 1]$ ) in the last line.

*Step 3.* We now prove the third point. For the first inequality, choose  $s = \frac{1}{2\gamma} - \frac{d}{2k}$ . Then by our assumptions,  $s > \max(1, \frac{d}{2})$  and we can apply Lemma 2.3 on the comparison of distance in  $\mathbf{P}(\mathbf{P}(E))$  and Theorem 5.1 to get

$$\mathcal{W}_1(\hat{\alpha}_j, \alpha) \leq C M_k(\alpha_1)^{\frac{1}{k}} \mathcal{W}_{H^{-s}}(\hat{\alpha}_j, \alpha)^{\frac{2k}{d+2ks}} \leq \frac{C M_k(\alpha_1)^{\frac{1}{k}}}{j^\gamma}$$

For the first part of the second inequality (5.6) we write

$$\mathcal{W}_1(\alpha, \beta) \leq \mathcal{W}_1(\alpha, \hat{\alpha}_j) + \mathcal{W}_1(\hat{\alpha}_j, \hat{\beta}_j) + \mathcal{W}_1(\hat{\beta}_j, \beta),$$

we use the inequality just proved above and the identity (2.11). The second part of the second inequality (5.6) is a mere application of Lemma 2.5.  $\square$

**5.2. Level 3 functional: an abstract setting.** We present a general result which state how we may defined a functional on  $\mathbf{P}(\mathbf{P}(E))$  from a family of compatible functionals on  $\mathbf{P}_{sym}(E^j)$ . The above theorems will be applied to the Boltzmann entropy functional, recovering classical results, as well as on the Fisher information.

**Proposition 5.3.** *Consider  $K_j$  a sequence of functionals on  $\mathbf{P}_m(E^j)$ ,  $m \geq 0$ , such that*

- (i)  $j^{-1}K_j(f^{\otimes j}) = K_1(f) \forall j \geq 1, \forall f \in \mathbf{P}_m(E)$ ;
- (ii) *any  $K_j : \mathbf{P}_m(E^j) \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, convex and lsc in the sense of the weak convergence. More precisely, we assume that there exists a class of smooth functions  $\mathcal{C}$  and a functional  $K_j^* : \mathcal{C} \rightarrow \mathbb{R} \cup \{+\infty\}$  such that*

$$\forall F \in \mathbf{P}_m(E^j) \quad K_j(F) = \sup_{\varphi \in \mathcal{C}} \{ \langle F, \psi \rangle - K_j^*(\varphi) \},$$

where  $\psi := \psi[\varphi] \in C(E^j)$  satisfies  $\psi/\langle v \rangle^m \rightarrow 0$  when  $|v| \rightarrow \infty$ ;

- (iii) *there exists  $(K_1^t)_{t>0}$  such that  $K_1^t : \mathbf{P}_m(E) \rightarrow [-C_t, C_t]$ ,  $C_t \in (0, \infty)$ , is continuous in the sense of the weak convergence for any  $t > 0$ , and*

$$\forall \rho \in \mathbf{P}_m(E) \quad K_1^t(\rho) \nearrow K_1(\rho).$$

Then for any  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ , we have

$$(5.8) \quad \mathcal{K}_1(\pi) := \sup_{j \in \mathbb{N}^*} j^{-1}K_j(\pi_j) = \mathcal{K}_2(\pi) := \int_{\mathbf{P}(E)} K_1(\rho) \pi(d\rho),$$

where  $\pi_j$  denotes  $j$ -th marginal defined thanks to Theorem 5.1, and we simply denote  $\mathcal{K}(\pi) := \mathcal{K}_1(\pi) = \mathcal{K}_2(\pi)$  that quantity. Then  $\mathcal{K} : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, linear, l.s.c. for the weak converge of  $\mathbf{P}_m(\mathbf{P}(E))$ .

**Proposition 5.4.** *In addition to the assumptions of Proposition 5.3, we assume*

- (iv) *for any  $j \geq 1$  and  $a \in (0, \infty)$  there exists  $\theta_j(N)$  and  $\varepsilon_{j,a}(N)$  such that  $\theta_j(N) \rightarrow 1$ ,  $\varepsilon_{j,a}(N) \rightarrow 0$  when  $N \rightarrow \infty$  and*

$$\forall F \in \mathcal{B}\mathbf{P}_{m,a}(E^N) \quad N^{-1}K_N(F) \geq \theta_j(N) j^{-1}K_j(F_j) - \varepsilon_{j,a}(N),$$

where  $F_j$  stands for the  $j$ -th marginal of  $F$ .

Then we have the following additional characterization of  $\mathcal{K}$

$$(5.9) \quad \mathcal{K}(\pi) = \lim_{j \rightarrow \infty} j^{-1}K_j(\pi_j).$$

Moreover, for any sequence  $(F^N)$  in  $\mathbf{P}_{sym}(E^N) \cap \mathbf{P}_m(E^N)$  and  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$  such that  $F^N \rightharpoonup \pi$  in  $\mathbf{P}_m(E^j)_{\forall j}$ , we have

$$(5.10) \quad \mathcal{K}(\pi) \leq \liminf_{N \rightarrow \infty} N^{-1}K_N(F^N).$$

**PROOF OF PROPOSITION 5.3.** We split the proof into five steps. We fix  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$  and we establish the identity (5.8).

*Step 1.* For  $j \geq 1$  and  $\varepsilon > 0$  fixed, there exists  $\varphi_j \in C_b(E^j)$  and  $\psi_j = \psi[\varphi_j]$  such that

$$K_j(\pi_j) \leq \langle \pi_j, \psi_j \rangle - K_j^*(\varphi_j) + \varepsilon,$$

so that from the definition of  $\mathcal{K}_2$ , we get

$$\begin{aligned}
\mathcal{K}_2(\pi) &= j^{-1} \int_{\mathbf{P}(E)} K_j(\rho^{\otimes j}) \pi(d\rho) \\
&= j^{-1} \int_{\mathbf{P}(E)} \sup_{\varphi_j \in C_b(E^j)} \{ \langle \rho^{\otimes j}, \psi_j \rangle - K_j^*(\varphi_j) \} \pi(d\rho) \\
&\geq j^{-1} \int_{\mathbf{P}(E)} \{ \langle \rho^{\otimes j}, \psi_j \rangle - K_j^*(\varphi_j) \} \pi(d\rho) \\
&= j^{-1} \left( \langle \pi_j, \psi_j \rangle - K_j^*(\varphi_j) \right) \geq j^{-1} K_j(\pi_j) - \varepsilon.
\end{aligned}$$

Taking the supremum over  $j$  in this inequality, and then passing to the limit  $\varepsilon \rightarrow 0$ , we get the first inequality

$$(5.11) \quad \mathcal{K}_1(\pi) = \sup_{j \geq 1} j^{-1} K_j(\pi_j) \leq \mathcal{K}_2(\pi).$$

*Step 2.* For given  $\omega_i$ ,  $1 \leq i \leq N$ , a partition of  $\mathbf{P}_m(E)$ , we introduce

$$\pi = \alpha_1 \gamma^1 + \dots + \alpha_N \gamma^N, \quad \gamma^i := \frac{1}{\alpha_i} \mathbf{1}_{\omega_i} \pi, \quad \alpha_i := \int_{\omega_i} \pi(d\rho),$$

and

$$\pi^N := \sum_{i=1}^N \alpha_i \delta_{f_i}, \quad f_i := \gamma_1^i = \int_{\mathbf{P}(E)} \rho \gamma^i(d\rho).$$

For any  $1 \leq i \leq N$ , we have

$$\mathcal{K}_1(\gamma^i) := \sup_{j \geq 1} j^{-1} K_j(\gamma_j^i) \geq K_1(\gamma_1^i) = K_1(f_i).$$

Using the convexity of  $\mathcal{K}_1$  (consequence of the convexity of each  $K_j$ ), the above inequality and the definitions of  $\pi^N$  and  $\mathcal{K}_2$ , we get

$$\begin{aligned}
\mathcal{K}_1(\pi) &\geq \alpha_1 \mathcal{K}_1(\gamma^1) + \dots + \alpha_N \mathcal{K}_1(\gamma^N) \\
&\geq \alpha_1 K_1(f_1) + \dots + \alpha_N K_1(f_N) \\
(5.12) \quad &= \int_{\mathbf{P}(E)} K_1(\rho) \pi^N(d\rho) = \mathcal{K}_2(\pi^N).
\end{aligned}$$

*Step 3.* Clearly

$$\mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R}, \quad \pi \mapsto \mathcal{K}_2^t(\pi) := \int_{\mathbf{P}(E)} K_1^t(\rho) \pi(d\rho) = \langle \pi, K_1^t \rangle$$

is continuous because  $K_1^t \in C_b(\mathbf{P}_m(E); \mathbb{R})$ . The Beppo-Lévi monotone convergence theorem implies that  $\mathcal{K}_2^t(\pi) \nearrow \mathcal{K}_2(\pi)$  for any  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ , and therefore  $\mathcal{K}_2$  is lsc.

Next, for any fixed  $\varepsilon > 0$ , we may cover  $\mathcal{B}\mathbf{P}_{m,1/\varepsilon}$  by a finite number of small balls  $B_i := \{f \in \mathcal{B}\mathbf{P}_{m,1/\varepsilon}; W_1(f, f_i) < \varepsilon\}$ ,  $1 \leq i \leq N-1$  because of the compactness of  $\mathcal{B}\mathbf{P}_{m,1/\varepsilon}$ . We define a partition of  $\mathbf{P}_m(E)$  by the family  $(\omega_k)_{1 \leq k \leq N}$  by setting  $\omega_1 := B_1$ , ...,  $\omega_k := B_k \setminus (B_1 \cup \dots \cup B_{k-1})$  for any  $1 \leq k \leq N-1$  and  $\omega_N := \mathbf{P}_m(E) \setminus (B_1 \cup \dots \cup B_{N-1})$ . We consider now a sequence  $\varepsilon \rightarrow 0$  and the sequence  $(\pi^N)$  associated to the above partition thanks to the construction made in step 2. We also define  $T^N : \mathbf{P}(E) \rightarrow \{\gamma^1, \dots, \gamma^N\}$  by  $T^N(\rho) = \gamma^i$  for any  $\rho \in \omega_i$ . We notice that  $\pi^N = (T^N)_\# \pi$  and then  $\mathcal{W}_1(\pi, \pi^N) \leq \langle (id \otimes T^N)_\# \pi, W_1(\cdot, \cdot) \rangle \leq \varepsilon$ . From  $\pi_1^N = \pi_1$  and  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ , we get

$$\langle \pi_1^N, |v|^m \rangle = \langle \pi_1, |v|^m \rangle = M_m(\pi) < \infty.$$

All together, we have proved  $\pi^N \rightharpoonup \pi$  weakly in  $\mathbf{P}_m(\mathbf{P}(E))$ .

As a conclusion, inequality (5.12), the above convergence and the lsc property of  $\mathcal{K}_2$  imply the second (and reverse) inequality

$$(5.13) \quad \mathcal{K}_2(\pi) \leq \liminf_{N \rightarrow \infty} \mathcal{K}_2(\pi^N) \leq \mathcal{K}_1(\pi).$$

The other properties of  $\mathcal{K}$  follow straightforwardly from those of  $\mathcal{K}_1$  and  $\mathcal{K}_2$ .  $\square$

**PROOF OF PROPOSITION 5.3.** On the one hand, for any  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ , we have from (iv)

$$N^{-1}K_N(\pi_N) \geq \theta_j(N)j^{-1}K_j(\pi_j) - \varepsilon_{j,a}(N) \quad \forall N \geq j \geq 1,$$

since  $M_m((\pi_N)_1) = M_m(\pi_1) =: a < \infty$ . Passing to the limit  $N \rightarrow \infty$ , we get

$$\liminf_{N \rightarrow \infty} N^{-1}K_N(\pi_N) \geq j^{-1}K_j(\pi_j) \quad \forall j \geq 1,$$

which in turn implies

$$\liminf_{N \rightarrow \infty} N^{-1}K_N(\pi_N) \geq \limsup_{j \rightarrow \infty} j^{-1}K_j(\pi_j),$$

and finally (5.9).

On the other hand, let us consider  $(F^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$  such that  $F^N \rightharpoonup \pi$  weakly in  $\mathbf{P}_m(E^j)_{\forall j}$ , in particular  $\langle F_1^N, |v|^m \rangle \leq a$  for some  $a \in (0, \infty)$ . For any  $j \geq 1$ , we have

$$\begin{aligned} j^{-1}K_j(\pi_j) &\leq \liminf_{N \rightarrow \infty} j^{-1}K_j(F_j^N) \\ &\leq \liminf_{N \rightarrow \infty} \theta_j(N)^{-1} \{N^{-1}K_N(F^N) + \varepsilon_{j,a}(N)\} \\ &= \liminf_{N \rightarrow \infty} N^{-1}K_N(F^N), \end{aligned}$$

where we have used successively the lsc nature of  $K_j$  and the inequality (iv). We deduce (5.10) thanks to the definition of  $\mathcal{K}_1$ .  $\square$

**5.3. Boltzmann entropy and Fisher information for mixtures.** We first recover some well known results on the Boltzmann entropy for mixture probability as stated in [1] and proved in [51]. It is obtained as a direct consequence of Propositions 5.3 and 5.4 and the properties of the entropy recalled in section 3.

**Theorem 5.5.** *Assume  $E \subset \mathbb{R}^d$  is a connected open set with smooth boundary. Let us fix a real number  $m > 0$ .*

(1) *For any  $\pi \in \mathbf{P}_m(\mathbf{P}(E))$ , there holds*

$$(5.14) \quad \mathcal{H}(\pi) := \int_{\mathbf{P}(E)} H(\rho) \pi(d\rho) = \sup_{j \in \mathbb{N}^*} H(\pi_j) = \lim_{j \rightarrow \infty} H(\pi_j),$$

where  $\pi_j$  is the  $j$ -th marginal of  $\pi$  defined in Theorem 5.1 and  $H$  is the normalized Boltzmann's entropy defined on  $\mathbf{P}_m(E^j)$  for any  $j \geq 1$ . Moreover, the functional  $\mathcal{H} : \mathbf{P}_m(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, linear and lsc.

(2) *Consider  $(F^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$  such that  $F^N \rightharpoonup \pi$  weakly in  $\mathbf{P}_m(E^j)_{\forall j}$ . Then*

$$(5.15) \quad \mathcal{H}(\pi) \leq \liminf_{N \rightarrow \infty} H(F^N).$$



PROOF OF THEOREM 5.5. The claims are straightforward consequence of Propositions 5.3 and 5.4 since (3.5) implies (i), (3.2) implies (ii), (3.7) and the lower bound in (3.1) imply (iv), and we only have to verify that (iii) is fulfilled. In order to do so, we introduce the family  $H^t : \mathbf{P}_m(E) \rightarrow \mathbb{R}$ ,  $t > 0$ , defined by  $\rho \in \mathbf{P}(E) \mapsto H^t(\rho) = H(\rho_t)$ , where  $\rho_t$  is the solution to the heat equation

$$(5.16) \quad \partial_t \rho_t - \Delta \rho_t = 0, \quad \rho_0 = \rho \in \mathbf{P}(E),$$

with Neumann condition when  $E \neq \mathbb{R}^d$ . We easily compute

$$\partial_t \int_E \rho_t \langle v \rangle^k = \int_E \rho_t \Delta \langle v \rangle^k \leq k^2 \int_E \rho_t \langle v \rangle^k,$$

so that  $\rho_t \in \mathbf{P}_k(\mathbf{P}(E))$  for any  $t \geq 0$ . That bound allows to define the entropy  $H(\rho_t)$  and we may then compute

$$\partial_t H(\rho_t) = \int_E (1 + \log \rho_t) \Delta \rho_t = - \int_E \frac{|\nabla \rho_t|^2}{\rho_t} \leq 0.$$

We deduce that  $H(\rho_t) \leq H(\rho)$  for any  $t \geq 0$ . Moreover,  $\rho_t \rightharpoonup \rho$  weakly in  $\mathbf{P}(E)$  when  $t \rightarrow 0$ , so that  $H(\rho_t) \rightarrow H(\rho)$  when  $t \rightarrow 0$ , because  $H$  is l.s.c.. Also, we clearly have  $\rho \rightarrow H(\rho_t)$  is continuous from  $\mathbf{P}(E)$  into  $\mathbb{R}$  for any  $t > 0$  thanks to the smoothing effect of the heat equation. All in all, we have proved that

$$H^t \text{ is continuous; } \quad H^t \nearrow H \text{ as } t \searrow 0,$$

which is nothing but (iii). □

We state now a similar result for the Fisher information of mixing probabilities.

**Theorem 5.6.** *Assume  $E \subset \mathbb{R}^d$  is a connected open set with smooth boundary.*

(1) *For any  $\pi \in \mathbf{P}(\mathbf{P}(E))$ , there holds*

$$(5.17) \quad \mathcal{I}(\pi) := \int_{\mathbf{P}(E)} I(\rho) \pi(d\rho) = \sup_{j \in \mathbb{N}^*} I(\pi_j) = \lim_{j \rightarrow \infty} I(\pi_j),$$

where  $I$  stands for the normalized Fisher information defined in  $\mathbf{P}(E^j)$  for any  $j \geq 1$ . The functional  $\mathcal{I} : \mathbf{P}(\mathbf{P}(E)) \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, linear and lsc for the weak convergence.

(2) *Moreover, the entropy  $\mathcal{H}$  is continuous on bounded sets relatively to  $\mathcal{I}$ . In other and more precise words, if  $(\pi_n)$  is a bounded sequence of  $\mathbf{P}_m(\mathbf{P}(E))$ ,  $m > 0$ , such that*

$$\pi_n \rightarrow \pi \text{ weakly in } \mathbf{P}(\mathbf{P}(E)) \quad \text{and} \quad \mathcal{I}(\pi_n) \leq C,$$

then  $\mathcal{H}(\pi_n) \rightarrow \mathcal{H}(\pi)$ .

(3) *Consider  $(F^N)$  a sequence of  $\mathbf{P}_{sym}(E^N)$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$  such that  $F^N \rightharpoonup \pi$  weakly in  $\mathbf{P}(E^j)_{\forall j}$ . Then*

$$(5.18) \quad \mathcal{I}(\pi) \leq \liminf I(F^N).$$

PROOF OF THEOREM 5.6. *Step 1.* Since the assumption (i) in Propositions 5.3 is a consequence of Lemma 3.6, the assumption (ii) in Propositions 5.3 is a consequence of (3.10) and Lemma 3.5, and the assumption (i) in Propositions 5.4 is a consequence of Lemma 3.7, we only have to verify the assumption (iii) in Propositions 5.3, and that is the aim of the next lemma. That will end the proof of statements (1) and (3).

*Step 2.* We prove **(2)**. We come back to the proof of Theorem 5.5, considering again  $\rho_t$  the solution to the heat equation (5.16) for any given  $\rho \in \mathbf{P}_m(E)$ , and the associated mollified entropy functional

$$\mathcal{H}_t(\pi) := \int_{\mathbf{P}(E)} H(\rho_t) \pi(d\rho).$$

On the one hand, because  $\rho \mapsto H(\rho_t) \in C(\mathbf{P}_m(E); \mathbb{R})$  for any  $t > 0$ , we have

$$|\mathcal{H}_t(\pi_n) - \mathcal{H}_t(\pi)| \rightarrow 0 \quad \text{when } n \rightarrow \infty.$$

On the other hand, because  $0 \leq I(\rho_s) \leq I(\rho)$  for any  $s \geq 0$ , we have

$$|H(\rho_t) - H(\rho)| = \left| \int_0^t I(\rho_s) ds \right| \leq t I(\rho),$$

from what we deduce that for any  $\alpha \in \mathbf{P}(\mathbf{P}(E))$

$$|\mathcal{H}_t(\alpha) - \mathcal{H}(\alpha)| \leq \int_{\mathbf{P}(E)} |H(\rho_t) - H(\rho)| \alpha(d\rho) \leq t \mathcal{I}(\alpha).$$

We finally write

$$\mathcal{H}(\pi_n) - \mathcal{H}(\pi) = (\mathcal{H}(\pi_n) - \mathcal{H}_t(\pi_n)) + (\mathcal{H}_t(\pi_n) - \mathcal{H}_t(\pi)) + (\mathcal{H}_t(\pi) - \mathcal{H}(\pi)),$$

and we conclude gathering the two preceding estimates on  $\mathcal{H}_t$  and the fact that  $\mathcal{I}(\pi) \leq C$  thanks to its l.s.c. property.  $\square$

**Lemma 5.7.** *There exists  $I^t : \mathbf{P}_m(E) \rightarrow \mathbb{R}$ ,  $m > 0$ , continuous, such that  $I^t(f) \nearrow I(f)$  for any  $f \in \mathbf{P}(E)$ . As a consequence, condition (iii) of Proposition 5.3 is fulfilled.*

**PROOF OF LEMMA 5.7.** *Step 1.* We start with two elementary and classical computations. The Gâteaux derivative of  $I$  writes

$$I'(f) \cdot h = 2 \int \frac{\nabla f}{f} \nabla h - \frac{|\nabla f|^2}{f^2} h.$$

Then, thanks to an integration by part on the  $v_i$  variable, we get

$$\begin{aligned} \frac{1}{2} I'(f) \cdot \Delta f &= \int \frac{1}{f} \partial_j f \partial_{ij} f - \int \frac{1}{2f^2} \partial_{ii} f (\partial_j f)^2 \\ &= \int \frac{\partial_i f}{f^2} \partial_j f \partial_{ij} f - \frac{1}{f} \partial_{ij} f \partial_{ij} f + \int \frac{1}{f^2} \partial_i f \partial_j f \partial_{ij} f - \frac{\partial_i f}{f^3} \partial_i f (\partial_j f)^2 \\ &= - \sum_{ij} \int \left( \frac{1}{f^2} \partial_i f \partial_j f - \frac{1}{f} \partial_{ij} f \right)^2 f =: -\frac{1}{2} J(f). \end{aligned}$$

We come back to the mollifying argument already used in the proof of Theorem 5.5. We denote again  $\rho_t$  the solution to the heat equation (5.16) for any given  $\rho \in \mathcal{BP}_{m,a}(E)$ ,  $a > 0$ . The smoothing properties of the heat equation imply that  $\rho_t$  is a smooth function for any  $t > 0$ . First, for any  $t > 0$ , we have  $\|\rho_t\|_{L_m^1 \cap H^2} \leq C_1(t, a) < \infty$ . That implies  $H(\rho_t) < \infty$  and the following equation is licit

$$\frac{d}{dt} H(\rho_t) = -I(\rho_t).$$

As a consequence,  $H(\rho_t)$  is decreasing and also  $\int_0^t I(\rho_s) ds < \infty$ , so that  $I(\rho_t) < \infty$  a.e.  $t > 0$ . We then may compute

$$\frac{d}{dt} I(\rho_t) = I'(\rho_t) \cdot \Delta \rho_t = -J(\rho_t) \leq 0.$$

Again, that implies  $I(\rho_t) \leq I(\rho) \in \mathbb{R} \cup \{+\infty\}$ ,  $I(\rho_t) \nearrow$  when  $t \searrow 0$ . Since furthermore  $I(\rho) \leq \liminf I(\rho_t)$  from the lsc of Fisher information, we get  $I^t(\rho) := I(\rho_t) \nearrow I(\rho)$  for any  $\rho \in \mathbf{P}_m(E)$ .

*Step 2.* We show that  $I^t : \mathbf{P}_m(E) \rightarrow \mathbb{R}$ ,  $\rho \mapsto I^t(\rho)$  is continuous. For any  $k \in (0, m)$ ,  $k \leq 2$ , we define

$$\tilde{H}(\rho) := \int_E \rho \log \rho \langle v \rangle^k, \quad \tilde{I}(\rho) := \int_E \frac{|\nabla \rho|^2}{\rho} \langle v \rangle^k, \quad \tilde{J}(\rho) := \int_E |D^2 \log \rho|^2 \rho \langle v \rangle^k,$$

and we remark that  $\tilde{H}(\rho_t) \leq C_2(t, a) < \infty$  for any  $t > 0$  and  $\rho \in \mathcal{BP}_{m,a}(E)$ . Next, we compute

$$\frac{d}{dt} \tilde{H}(\rho_t) \leq -\tilde{I}(\rho_t) + C \tilde{H}(\rho_t).$$

That inequality implies  $\inf_{s \in [0, t]} \tilde{I}(\rho_s) \leq t^{-1} \int_0^t I(\rho_s) ds \leq C_3(t, a)$  for any  $t > 0$ . The evolution of the weight Fisher information is

$$\frac{d}{dt} \tilde{I}(\rho_t) = -\tilde{J}(\rho_t) + R(\rho_t)$$

where, with the notation  $\varphi := \langle v \rangle^k$ ,

$$R(f) := -2 \int \frac{1}{f} \partial_j f \partial_{ij} f \partial_i \varphi + \int \frac{1}{f^2} \partial_i f (\partial_j f)^2 \partial_i \varphi.$$

Observing that  $\partial_{ij}^2 f = f \partial_{ij}^2 \log f + f^{-1} \partial_i f \partial_j f$  and  $f^{-1} \partial_i f = \partial_i \log f$ , we may estimate the last term as follows

$$\begin{aligned} R(f) &= -2 \int f \partial_j \log f \partial_{ij}^2 \log f \partial_i \varphi - \int \partial_i f (\partial_j \log f)^2 \partial_i \varphi \\ &= -2 \int f \partial_j \log f \partial_{ij}^2 \log f \partial_i \varphi + \int f \partial_i [(\partial_j \log f)^2] \partial_i \varphi + \int f (\partial_j \log f)^2 \partial_{ii}^2 \varphi \\ &= \int f (\partial_j \log f)^2 \partial_{ii}^2 \varphi \leq C_k I(f). \end{aligned}$$

We deduce from the resulting differential inequality and the first bound on  $\tilde{I}(\rho_t)$  that  $\tilde{I}(\rho_t) \leq C_4(t, a)$  for any  $t > 0$ . Finally, from  $\|\rho_t\|_{L_m^1} \leq C_1(t, a)$  we deduce when  $E = \mathbb{R}^d$

$$\begin{aligned} \rho_t(x) &\geq c_t \int_{B_R} f(y) e^{-\frac{|x-y|^2}{2t}} \geq c_t e^{-\frac{R^2}{t}} e^{-\frac{|x|^2}{t}} \int_{B_R} f(y) dy \\ &\geq c_t e^{-\frac{R^2}{t}} e^{-\frac{|x|^2}{t}} \left(1 - \frac{a}{R^m}\right) = K_t e^{-\frac{|x|^2}{t}}. \end{aligned}$$

In the general case, we also have thanks to the strong maximal principle, the Hopf lemma and the Neumann condition :

$$\rho_t(x) \geq r_{t,R} > 0 \quad \forall t > 0, \forall x \in E \cap B_R.$$

As a conclusion, for  $k \in (0, \min(2, m))$  fixed, we write

$$I^t(\rho) = \int_{B_R} \frac{|\nabla \rho_t|^2}{\rho_t} + \int_{B_R^c} \frac{|\nabla \rho_t|^2}{\rho_t}.$$

The first term is clearly continuous with respect to  $\rho \in \mathcal{BP}_{m,a}(E)$  for any fixed  $R > 0$  thanks to the above uniform inferior estimate on  $B_R$  and the  $H^2$  bound, and the second term is smaller than  $C_4(t, a)/R^k$ .  $\square$

We can now use lemma 2.5 to get an HWI inequality on  $\mathbf{P}(\mathbf{P}_2(E))$ . It is stated in the following proposition.

**Proposition 5.8.** *Assume  $E = \mathbb{R}^d$  or more generally that (3.15) holds for  $N = 1$ . For any  $\alpha, \beta \in \mathbf{P}(\mathbf{P}_2(E))$ , we have*

$$(5.19) \quad \mathcal{H}(\alpha) \leq \mathcal{H}(\beta) + \sqrt{\mathcal{I}(\alpha)} \mathcal{W}_2(\alpha, \beta).$$

PROOF OF PROPOSITION 5.8. A first way in order to prove (5.19) is just to pass in the limit in the HWI inequality (3.15) for  $\alpha_N$  and  $\beta_N$  and use the inequality stated in lemma 2.5 for the quadratic cost, and the result of the previous section about level 3 entropy and Fisher information 5.14 et 5.17.

Another possibility is to sum up the HWI inequality (3.16) for  $\rho \in \mathbf{P}(E)$ . Choosing an optimal transference plan  $\Pi$  for  $\mathcal{W}_2$  between  $\alpha$  and  $\beta$ , we have

$$\int_{\mathbf{P}(E)} H(\rho) \Pi(d\rho, d\eta) \leq \int_{\mathbf{P}(E)} H(\eta) \Pi(d\rho, d\eta) + \int_{\mathbf{P}(E)} \sqrt{\mathcal{I}(\rho)} \mathcal{W}_2(\rho, \eta) \Pi(d\rho, d\eta),$$

so that

$$H(\alpha) \leq H(\beta) + \left( \int_{\mathbf{P}(E)} \mathcal{I}(\rho) \Pi(d\rho, d\eta) \right)^{\frac{1}{2}} \left( \int_{\mathbf{P}(E)} \mathcal{W}_2(\rho, \eta)^2 \Pi(d\rho, d\eta) \right)^{\frac{1}{2}},$$

thanks to Cauchy-Schwarz inequality. It leads to the desired inequality.  $\square$

**Proposition 5.9.** *Consider  $\pi \in \mathbf{P}(\mathbf{P}(E))$  and  $(\pi_j)$  the associated family of compatible and symmetric probabilities in  $\mathbf{P}(E^j)$  defined as in the De Finetti, Hewitt & Savage theorem. For any  $p \in [1, +\infty]$ , the following equality holds*

$$(5.20) \quad \pi - \text{Suppess} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \} = \sup_{j \in \mathbb{N}} \|\pi_j\|_p^{\frac{1}{j}} = \lim_{j \rightarrow +\infty} \|\pi_j\|_p^{\frac{1}{j}}.$$

It is part of the result that the limit exists. In particular, it implies the equivalence

$$\forall j \in \mathbb{N}, \|\pi_j\|_{L^p(E^j)} \leq C^j \iff \pi - \text{Suppess} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \} \leq C.$$

PROOF OF PROPOSITION 5.9. First remark that there is nothing to prove for  $p = 1$  since we are dealing with probabilities. Now, one inequality is a simple consequence of the De Finetti, Hewitt & Savage theorem. In fact, using the definition of  $\pi_j$ , we get

$$\|\pi_j\|_p = \left\| \int_{\mathbf{P}(E)} \rho^{\otimes j} \pi(d\rho) \right\|_p \leq \int_{\mathbf{P}(E)} \|\rho^{\otimes j}\|_p \pi(d\rho) = \int_{\mathbf{P}(E)} \|\rho\|_p^j \pi(d\rho),$$

and the last quantity is clearly bounded by  $M = \pi - \text{Suppess} \{ \|\rho\|_p, \rho \in \mathbf{P}(E) \}$ .

For the reverse inequality, we denote by  $q \in [1, +\infty)$  the real conjugate to  $p$ . Because  $L^q(E) = (L^p(E))'$ , the Hahn-Banach separation theorem infers that for any  $\lambda < M$  there exists  $f$  in the unit ball of  $L^q(E)$  so that the set

$$\mathcal{B} := \{ \rho \in \mathbf{P}(E) \text{ s.t. } \int f(x) \rho(dx) \geq \lambda \}$$

is of  $\pi$ -measure positive :  $\delta := \int_{\mathcal{B}} \pi(d\rho) > 0$ . Now for any  $j \in \mathbb{N}$

$$\|\pi_j\|_p \geq \int_{E^j} f^{\otimes j} d\pi_j = \int_{\mathbf{P}(E)} \left( \int_{E^j} f^{\otimes j} \rho^{\otimes j} \right) d\pi(\rho) \geq \delta \lambda^j,$$

which implies the reserve inequality  $M \leq \lim_{j \rightarrow +\infty} \|\pi_j\|_p^{\frac{1}{j}}$ .  $\square$

**5.4. Strong version of De Finetti, Hewitt and Savage theorem.** The preceding results together with some smoothing techniques and the HWI inequality makes possible to compare different senses of convergence for sequences of  $\mathbf{P}(E^N)$ ,  $N \rightarrow \infty$ , without any assumption of chaos.

**Theorem 5.10.** *Assume  $E = \mathbb{R}^d$  or more generally assume that  $E \subset \mathbb{R}^d$  is a connected open subset with smooth boundary and that (3.15) holds. Consider  $(F^N)$  a sequence of  $\mathbf{P}_{\text{sym}}(E^N)$  and  $\pi \in \mathbf{P}(\mathbf{P}(E))$  such that  $F^N \rightarrow \pi$  weakly in  $\mathbf{P}_k(E^j)_{\forall j}$ ,  $k > 2$ .*

(1) *In the list of assertions below, each assertion implies the one which follows:*

- (i)  $I(F^N) \rightarrow \mathcal{I}(\pi)$ ,  $\mathcal{I}(\pi) < \infty$ ;
- (ii)  $I(F^N)$  is bounded;
- (iii)  $H(F^N) \rightarrow \mathcal{H}(\pi)$ ,  $\mathcal{H}(\pi) < \infty$ .

(2) *More precisely, the following version of the implication (ii)  $\Rightarrow$  (iii) holds. For any  $k > 2$  and  $K > 0$ , there exists a constant  $C$  such that for any any sequence  $(F^N)$  of  $\mathbf{P}_{\text{sym}}(E^N)$  satisfying*

$$\forall N \quad M_k(F_1^N) \leq K, \quad I(F^N) \leq K,$$

there holds

$$(5.21) \quad \forall N \geq 2 \quad |H(F^N) - \mathcal{H}(\pi)| \leq C K^2 \left[ W_2(F^N, \pi_N) + \frac{\ln N}{N^\gamma} \right],$$

with  $\gamma := \frac{k-2}{k-2+2d(k+2)}$ .

(3) *In particular, for any sequence  $(\pi_j)$  of symmetric and compatible probability measures of  $\mathbf{P}(E^j)$  satisfying*

$$M_k(\pi_1) \leq K, \quad \forall j \geq 1 \quad I(\pi_j) \leq K,$$

there holds

$$(5.22) \quad \forall j \geq 2, \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq C \frac{\ln j}{j^\gamma},$$

for the same value of  $\gamma$ . In other words, (5.22) gives a rate of convergence for the limit (5.14).

**Remark 5.11.** *In the case where  $E$  is compact, the choice " $k = +\infty$ " is possible and the previous inequalities are true with  $\gamma = \frac{1}{2d+1}$ .*

**PROOF OF THEOREM 5.10.** We split the proof into four steps.

*Step 1.* i) implies ii) is clear. For ii) implies iii), we use the HWI inequality (3.15) and we write

$$\begin{aligned} |H(F^N) - \mathcal{H}(\pi)| &= |H(F^N) - H(\pi_N) + H(\pi_N) - \mathcal{H}(\pi)| \\ &\leq C_E \left( \sqrt{I(F^N)} + \sqrt{I(\pi_N)} \right) W_2(F^N, \pi_N) + |H(\pi_N) - \mathcal{H}(\pi)|. \end{aligned}$$

We know from (5.14) that  $\mathcal{H}(\pi) = \lim H(\pi_N)$  and from (5.17) and (5.18) that  $I(\pi_N) \leq \mathcal{I}(\pi) \leq \liminf I(F^N) \leq K$ , from which we conclude that there exist a sequence  $\varepsilon_\pi(N) \rightarrow 0$  such that

$$|H(F^N) - \mathcal{H}(\pi)| \leq 2 C_E K W_2(F^N, \pi_N) + \varepsilon(N).$$

We now aim to estimate  $\varepsilon(N)$  more explicitly as claimed in point (3). Then (2) will be a direct consequence of (3) and the above estimate.

From now on, we only consider the case  $E = \mathbb{R}^d$  since the general case is similar (and the case when  $E$  is compact is even simpler).

*Step 2.* From [9, Theorem A.1] we know that for any  $R, \delta > 0$  we may cover  $\mathbf{P}(B_R)$  by  $\mathcal{N}(R, \delta/2)$  balls of radius  $\delta/2$  in  $W_1$  distance (which is less accurate than the one considered in the above quoted result) with

$$\mathcal{N}(R, \delta) \leq \left( \frac{C'_1 R}{\delta} \right)^{C'_2 (R/\delta)^d}.$$

Let us fix  $a \geq 1$  and recall that we define  $\mathcal{BP}_{k,a}(E) := \{\rho \in \mathbf{P}(E) \text{ s.t. } M_k(\rho) \leq a\}$ . Next, for any  $\rho \in \mathcal{BP}_{k,a}(E)$ , we define  $\rho_R \in \mathbf{P}(B_R)$  by  $\rho_R = \rho(B_R)^{-1} \rho \mathbf{1}_{B_R}$  for  $R$  large enough (so that it defines a probability), and we observe that for any  $f \in \mathbf{P}(E)$  we have

$$W_1(\rho, f) \leq W_1(\rho_R, f) + W_1(\rho_R, \rho),$$

and that for any  $R \geq a^{1/k}$

$$\begin{aligned} W_1(\rho_R, \rho) &\leq \|\rho_R - \rho\|_{TV} \leq \left| 1 - \frac{1}{\rho(B_R)} \right| + \rho(B_R^c) \\ &\leq \left( 1 + \frac{1}{\rho(B_R)} \right) \rho(B_R^c) \leq 3 \frac{a}{R^k}, \end{aligned}$$

since then  $\rho(B_R) \geq 1 - \frac{a}{R^k} \geq \frac{1}{2}$ .

As a consequence, for any  $\delta \leq 1$  and  $a \geq 1$ , choosing  $R$  such that  $3a/R^k = \delta/2$  in the two preceding estimates, we may cover  $\mathcal{BP}_{k,a}(E)$  by  $\mathcal{N}_a(\delta) = \mathcal{N}(R, \delta/2)$  balls of radius  $\delta$  in  $W_1$  distance, with

$$\frac{1}{\delta} \leq \mathcal{N}_a(\delta) \leq \left( C_1 a^{\frac{1}{k}} \delta^{-1-\frac{1}{k}} \right)^{C_2 a^{\frac{d}{k}} \delta^{-d-\frac{d}{k}}}.$$

The above lower bound on  $\mathcal{N}_a(\delta)$  is straightforwardly obtained by considering balls centered on Dirac masses distributed on a line. In the sequel, we shall often use the shortcut  $\mathcal{N} = \mathcal{N}_a(\delta)$ . Let us then introduce a family  $\omega_i^\delta \subset \mathcal{BP}_{k,a}(E)$ ,  $1 \leq i \leq \mathcal{N}_a(\delta)$ , such that

$$\sup_{\rho, \eta \in \omega_i^\delta} W_1(\rho, \eta) \leq 2\delta, \quad \omega_i^\delta \cap \omega_j^\delta = \emptyset \text{ if } i \neq j, \quad \mathcal{BP}_{k,a}(E) \subset \bigcup_{i=1}^{\mathcal{N}_a(\delta)} \omega_i^\delta,$$

as well as the masses and centers of mass

$$\alpha_i^\delta := \int_{\omega_i^\delta} \pi, \quad f_i^\delta := \frac{1}{\alpha_i^\delta} \int_{\omega_i^\delta} \rho \pi(d\rho).$$

We also denote  $\omega_0 := \left( \bigcup_{i=1}^{\mathcal{N}_a(\delta)} \omega_i^\delta \right)^c$  and  $\alpha_0^\delta := \int_{\omega_0^\delta} \pi$ , so that  $\sum_{i=0}^{\mathcal{N}_a(\delta)} \alpha_i^\delta = 1$ . Denoting  $\mathcal{Z} := \{i = 1, \dots, \mathcal{N}_a(\delta); \alpha_i^\delta \geq \mathcal{N}_a(\delta)^{-2}\}$ , we finally defined

$$\pi^\delta := \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta \delta_{f_i^\delta}, \quad \text{with } \beta_i^\delta := \frac{\alpha_i^\delta}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \text{ if } i \in \mathcal{Z} \text{ and } \beta_i^\delta := 0 \text{ if } i \notin \mathcal{Z}.$$

Remark that by our moment assumption

$$\alpha_0^\delta \leq \int_{(\mathcal{BP}_{k,a}(E))^c} \pi(d\rho) \leq \int_{\mathbf{P}(E)} \frac{M_k(\rho)}{a} \pi(d\rho) = \frac{M_k(\pi_1)}{a} \leq \frac{K}{a}.$$

Since  $\sum_{i \notin \mathcal{Z}} \alpha_i^\delta \leq \mathcal{N}^{-1} \leq \delta$ , we necessarily have  $\mathcal{Z} \neq \emptyset$  if  $\delta + \frac{K}{a} \leq \frac{1}{2} < 1$ , an assumption that we will make in the sequel. In that case, we have

$$(5.23) \quad 1 \geq \sum_{j \in \mathcal{Z}} \alpha_j^\delta = \sum_{j=1}^{\mathcal{N}_a(\delta)} \alpha_j^\delta - \sum_{j \notin \mathcal{Z}} \alpha_j^\delta \geq 1 - \frac{M_k(\pi_1)}{a} - \delta \geq \frac{1}{2}.$$

Now, by convexity of the Fisher information

$$I(f_i^\delta) \leq \frac{1}{\alpha_i^\delta} \int_{\omega_i^\delta} I(\rho) \pi(d\rho),$$

which in turns implies that

$$\mathcal{I}(\pi^\delta) = \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta I(f_i^\delta) \leq \frac{1}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \sum_{i \in \mathcal{Z}} \int_{\omega_i^\delta} I(\rho) \pi(d\rho) \leq 2\mathcal{I}(\pi).$$

Similarly, for the moment of order  $k$  :

$$M_k(\pi_1^\delta) = \sum_{i=1}^{\mathcal{N}_a(\delta)} \beta_i^\delta M_k(f_i^\delta) \leq \frac{1}{\sum_{j \in \mathcal{Z}} \alpha_j^\delta} \sum_{i \in \mathcal{Z}} \int_{\omega_i^\delta} M_k(\rho) \pi(d\rho) \leq 2M_k(\pi_1).$$

In order to prove (5.22), we introduce the splitting

$$(5.24) \quad |H(\pi_j) - \mathcal{H}(\pi)| \leq |H(\pi_j) - H(\pi_j^\delta)| \\ + |H(\pi_j^\delta) - \mathcal{H}(\pi^\delta)| + |\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)|,$$

where we have written  $\pi_j^\delta := (\pi^\delta)_j$ . We now estimate each term separately.

*Step 3.* On the one hand, defining  $T^\delta : \mathbf{P}(E) \rightarrow \{f_0^\delta, \dots, f_N^\delta\}$ ,  $T^\delta(\rho) = f_i^\delta$  if  $\rho \in \omega_i^\delta$ ,  $T^\delta(\rho) = f_0^\delta = \delta_0$  if  $\rho \in \omega_0^\delta$  and  $\beta_0^\delta := 0$ , we compute

$$\begin{aligned} \mathcal{W}_1(\pi, \pi^\delta) &\leq \mathcal{W}_1\left(\pi, \sum_{i=0}^{\mathcal{N}} \alpha_i^\delta \delta_{f_i^\delta}\right) + \mathcal{W}_1\left(\sum_{i=0}^{\mathcal{N}} \alpha_i^\delta \delta_{f_i^\delta}, \sum_{i=0}^{\mathcal{N}} \beta_i^\delta \delta_{f_i^\delta}\right) \\ &\leq \int_{\mathbf{P}(E) \times \mathbf{P}(E)} \mathcal{W}_1(\rho, \eta) (Id \otimes T^\delta) \# \pi + \left\| \sum_{i=0}^{\mathcal{N}} (\alpha_i^\delta - \beta_i^\delta) \delta_{f_i^\delta} \right\|_{TV} \\ &\leq \int_{\mathbf{P}(E)} \mathcal{W}_1(\rho, T^\delta(\rho)) \pi(d\rho) + \sum_{i=1}^{\mathcal{N}} |\alpha_i^\delta - \beta_i^\delta| + |\alpha_0^\delta| \\ &\leq 5 \left( \delta + \frac{K}{a} \right), \end{aligned}$$

where we have used several times estimation (5.23), in particular in order to get the inequality

$$\sum_{i=1}^{\mathcal{N}} |\alpha_i^\delta - \beta_i^\delta| = \left( 1 - \frac{1}{\sum_{i \in \mathcal{Z}} \alpha_i^\delta} \right) \sum_{i \in \mathcal{Z}} \alpha_i^\delta + \sum_{i \notin \mathcal{Z}} \alpha_i^\delta \leq 3 \left( \delta + \frac{M_k(\pi_1)}{a} \right).$$

Using the lemma 2.3 and the bound on  $M_k(\pi_1^\delta)$ , we obtain a bound on  $\mathcal{W}_2(\pi^\delta, \pi)$  as follows (we recall that the constant  $C$  that appears is numerical :  $C = 2^{3/2}$ )

$$\mathcal{W}_2(\pi^\delta, \pi) \leq C M_k(\pi_1)^{1/k} \mathcal{W}_1(\pi^\delta, \pi)^{1/2-1/k} \leq C K^{1/k} \left( \delta + \frac{K}{a} \right)^{1/2-1/k}.$$

Now, we use the HWI inequality on  $\mathbf{P}(E)$  stated in Proposition 5.8 and we bound the first term in (5.24) by

$$|\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)| \leq \left[ \sqrt{\mathcal{I}(\pi^\delta)} + \sqrt{\mathcal{I}(\pi)} \right] \mathcal{W}_2(\pi^\delta, \pi) \leq C \mathcal{W}_2(\pi^\delta, \pi),$$

and the third term in (5.24) very similarly

$$\begin{aligned} |H(\pi_j^\delta) - \mathcal{H}(\pi_j)| &\leq (\sqrt{I(\pi_j^\delta)} + \sqrt{I(\pi_j)}) W_2(\pi_j^\delta, \pi_j) \\ &\leq (\sqrt{\mathcal{I}(\pi^\delta)} + \sqrt{\mathcal{I}(\pi)}) \mathcal{W}_2(\pi^\delta, \pi) \leq 2C\mathcal{W}_2(\pi^\delta, \pi), \end{aligned}$$

where we have used the properties (5.17) of the level 3 Fisher information and Lemma 2.5 in order to bound  $W_2$  by  $\mathcal{W}_2$ . All together, we have proved

$$(5.25) \quad |\mathcal{H}(\pi^\delta) - \mathcal{H}(\pi)| + |H(\pi_j^\delta) - \mathcal{H}(\pi_j)| \leq C \left( \delta + \frac{1}{a} \right)^{1/2-1/k},$$

for some constant  $C$  only depending on the bound  $K$ .

*Step 4.* We estimate the second and last term in (5.24). Using that  $\pi_j^\delta = \beta_1^\delta (f_1^\delta)^{\otimes j} + \dots + \beta_{\mathcal{N}}^\delta (f_{\mathcal{N}}^\delta)^{\otimes j}$ , we write

$$\begin{aligned} H(\pi_j^\delta) &= \frac{1}{j} \int_{E^j} \pi_j^\delta \log \pi_j^\delta \\ &= \sum_{i=1}^{\mathcal{N}} \beta_i^\delta H(f_i^\delta) + \frac{1}{j} \int_{E^j} \pi_j^\delta \Lambda \left( \frac{\beta_1^\delta (f_1^\delta)^{\otimes j}}{\pi_j^\delta}, \dots, \frac{\beta_{\mathcal{N}}^\delta (f_{\mathcal{N}}^\delta)^{\otimes j}}{\pi_j^\delta} \right), \end{aligned}$$

with  $\Lambda : \{U = (u_i) \in \mathbb{R}_+^{\mathcal{N}}, \sum_i u_i = 1\} \rightarrow \mathbb{R}$  defined by

$$\Lambda(U) := u_1 \log \left( \frac{\beta_1^\delta}{u_1} \right) + \dots + u_{\mathcal{N}} \log \left( \frac{\beta_{\mathcal{N}}^\delta}{u_{\mathcal{N}}} \right).$$

Observing that  $\Lambda$  is in fact (the opposite of) a discrete relative entropy, we have for any  $U \in \mathbb{R}_+^{\mathcal{N}}$  with  $\sum_i u_i = 1$

$$-\log(\mathcal{N}^2) \leq \log(\min \beta_i^\delta) \leq \Lambda(U) \leq 0,$$

we deduce

$$|H(\pi_j^\delta) - \mathcal{H}(\pi^\delta)| \leq \frac{2}{j} \log \mathcal{N}_a(\delta).$$

*Step 5.* All in all, observing that  $\log \mathcal{N}_a(\delta) \leq C \delta^{-d(1+\frac{1}{k})} a^{\frac{d}{k}} \ln(a^{1/k} \delta^{-1})$ , we have

$$C^{-1} |H(\pi_j) - \mathcal{H}(\pi)| \leq \left( \delta + \frac{1}{a} \right)^{1/2-1/k} + \frac{1}{j} \frac{a^{\frac{d}{k}}}{\delta^{d(1+\frac{1}{k})}} \ln(a^{1/k} \delta^{-1}).$$

We can now (almost) optimize by choosing  $\delta = a^{-1}$ ,  $a = j^r$ , with  $r^{-1} := \frac{1}{2} - \frac{1}{k} + d(1 + \frac{2}{k})$  we obtain (5.22) for the integers  $j$  large enough so that the condition  $\delta + \frac{K}{a} \leq \frac{1}{2}$  is fulfilled in order to ensure that  $\mathcal{Z} \neq \emptyset$ . For smaller values of  $j$  we may choose a (larger) constant  $C$  so that (5.22) also holds, and that concludes the proof.  $\square$

## REFERENCES

- [1] ARKERYD, L., CAPRINO, S., AND IANIRO, N. The homogeneous Boltzmann hierarchy and statistical solutions to the homogeneous Boltzmann equation. *J. Statist. Phys.* 63, 1-2 (1991), 345–361.
- [2] ARTSTEIN, S., BALL, K. M., BARTHE, F., AND NAOR, A. On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields* 129, 3 (2004), 381–390.
- [3] BARTHE, F., CORDERO-ERAUSQUIN, D., AND MAUREY, B. Entropy of spherical marginals and related inequalities. *J. Math. Pures Appl. (9)* 86, 2 (2006), 89–99.
- [4] BEN AROUS, G., AND ZEITOUNI, O. Increasing propagation of chaos for mean field models. *Ann. Inst. H. Poincaré Probab. Statist.* 35, 1 (1999), 85–102.
- [5] BERRY, A. C. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* 49 (1941), 122–136.



- [6] BOBKOV, S. G., CHISTYAKOW, G., AND GÖTZE, F. Berry-essen bounds in the entropic central limit theorem. arXiv:1104.3994v1.
- [7] BOBKOV, S. G., GENTIL, I., AND LEDOUX, M. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)* 80, 7 (2001), 669–696.
- [8] BOLLEY, F., GUILLIN, A., AND MALRIEU, F. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.* 44, 5 (2010), 867–884.
- [9] BOLLEY, F., GUILLIN, A., AND VILLANI, C. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* 137, 3-4 (2007), 541–593.
- [10] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. *Comm. Math. Phys.* 143, 3 (1992), 501–525.
- [11] CAGLIOTI, E., LIONS, P.-L., MARCHIORO, C., AND PULVIRENTI, M. A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. II. *Comm. Math. Phys.* 174, 2 (1995), 229–260.
- [12] CARLEN, E. A. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.* 101, 1 (1991), 194–211.
- [13] CARLEN, E. A., CARVALHO, M. C., LE ROUX, J., LOSS, M., AND VILLANI, C. Entropy and chaos in the Kac model. *Kinet. Relat. Models* 3, 1 (2010), 85–122.
- [14] CARLEN, E. A., LIEB, E. H., AND LOSS, M. A sharp analog of Young’s inequality on  $S^N$  and related entropy inequalities. *J. Geom. Anal.* 14, 3 (2004), 487–520.
- [15] CARRAPATOSO, K. Quantitative and qualitative Kac’s chaos on the Boltzmanns sphere. in preparation.
- [16] CARRILLO, J., AND TOSCANI, G. Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Rivista Matematica di Parma* 6 (2007), 75–198.
- [17] CORDERO-ERAUSQUIN, D. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.* 161, 3 (2002), 257–269.
- [18] DE FINETTI, B. La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* 7, 1 (1937), 1–68.
- [19] DIACONIS, P., AND FREEDMAN, D. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.* 23, 2, suppl. (1987), 397–423.
- [20] DOBRIĆ, V., AND YUKICH, J. E. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* 8, 1 (1995), 97–118.
- [21] DOBRUŠIN, R. L. Vlasov equations. *Funktsional. Anal. i Prilozhen.* 13, 2 (1979), 48–58, 96.
- [22] ESSEEN, C.-G. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* 28A, 9 (1942), 19.
- [23] FELLER, W. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.
- [24] GOUDON, T., JUNCA, S., AND TOSCANI, G. Fourier-based distances and Berry-Esseen like inequalities for smooth densities. *Monatsh. Math.* 135, 2 (2002), 115–136.
- [25] GRAHAM, C., AND MÉLÉARD, S. Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *The Annals of Probability* 25 (1997), 115–132.
- [26] GRÜNBAUM, F. A. Propagation of chaos for the Boltzmann equation. *Arch. Rational Mech. Anal.* 42 (1971), 323–345.
- [27] HAURAY, M., AND JABIN, P.-E. Particles approximations of vlasov equations with singular forces : Part. 2. hal-00609453.
- [28] HAURAY, M., AND JABIN, P.-E.  $N$ -particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.* 183, 3 (2007), 489–524.
- [29] HEWITT, E., AND SAVAGE, L. J. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* 80 (1955), 470–501.
- [30] KAC, M. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III* (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.
- [31] KOSYGINA, E. The behavior of the specific entropy in the hydrodynamic scaling limit. *Ann. Probab.* 29, 3 (2001), 1086–1110.
- [32] LANFORD, III, O. E. Time evolution of large classical systems. In *Dynamical systems, theory and applications (Recontres, Battelle Res. Inst., Seattle, Wash., 1974)*. Springer, Berlin, 1975, pp. 1–111. Lecture Notes in Phys., Vol. 38.

- [33] LIONS, P.-L. Théorie des jeux de champ moyen et applications (mean field games). In *Cours du Collège de France*. [http://www.college-de-france.fr/default/EN/all/equ\\_der/audio\\_video.jsp](http://www.college-de-france.fr/default/EN/all/equ_der/audio_video.jsp), 2007–2009.
- [34] LIONS, P.-L., AND TOSCANI, G. A strengthened central limit theorem for smooth densities. *J. Funct. Anal.* *129*, 1 (1995), 148–167.
- [35] LOTT, J., AND VILLANI, C. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)* *169*, 3 (2009), 903–991.
- [36] MALRIEU, F. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.* *95*, 1 (2001), 109–132.
- [37] MCKEAN, H. P. The central limit theorem for Carleman's equation. *Israel J. Math.* *21*, 1 (1975), 54–92.
- [38] MCKEAN, JR., H. P. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*. Air Force Office Sci. Res., Arlington, Va., 1967, pp. 41–57.
- [39] MEHLER, F. G. Ueber die entwicklung einer function von beliebig vielen variablen nach laplaschen functionen höherer ordnung. *Crelle's Journal* *66* (1866), 161–176.
- [40] MÉLÉARD, S. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, vol. 1627 of *Lecture Notes in Math*. Springer, Berlin, 1996, pp. 42–95.
- [41] MESSER, J., AND SPOHN, H. Statistical mechanics of the isothermal Lane-Emden equation. *J. Statist. Phys.* *29*, 3 (1982), 561–578.
- [42] MISCHLER, S. Le programme de Kac sur les limites de champ moyen. Séminaire EDP-X, Décembre 2010.
- [43] MISCHLER, S., AND MOUHOT. Kac's program in kinetic theory. hal-00447988.
- [44] MISCHLER, S., AND MOUHOT, C. On chaos properties of a Boltzmann-Kac system for inelastic collisions. Work in progress.
- [45] MISCHLER, S., MOUHOT, C., AND WENNBURG, B. A new approach to quantitative chaos propagation for drift, diffusion and jump processes. <http://hal.archives-ouvertes.fr/ccsd-00559132>.
- [46] OSADA, H. Propagation of chaos for the two-dimensional Navier-Stokes equation. *Proc. Japan Acad. Ser. A Math. Sci.* *62*, 1 (1986), 8–11.
- [47] OTTO, F. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* *26*, 1-2 (2001), 101–174.
- [48] OTTO, F., AND VILLANI, C. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* *173*, 2 (2000), 361–400.
- [49] RACHEV, S. T., AND RÜSCHENDORF, L. *Mass transportation problems. Vol. II. Probability and its Applications* (New York). Springer-Verlag, New York, 1998. Applications.
- [50] RIO, E. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* *45*, 3 (2009), 802–817.
- [51] ROBINSON, D. W., AND RUELLE, D. Mean entropy of states in classical statistical mechanics. *Comm. Math. Phys.* *5* (1967), 288–300.
- [52] SZNITMAN, A.-S. Équations de type de Boltzmann, spatialement homogènes. *Z. Wahrsch. Verw. Gebiete* *66*, 4 (1984), 559–592.
- [53] SZNITMAN, A.-S. A propagation of chaos result for Burgers' equation. In *Hydrodynamic behavior and interacting particle systems (Minneapolis, Minn., 1986)*, vol. 9 of *IMA Vol. Math. Appl.* Springer, New York, 1987, pp. 181–188.
- [54] SZNITMAN, A.-S. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math*. Springer, Berlin, 1991, pp. 165–251.
- [55] TANAKA, H. Some probabilistic problems in the spatially homogeneous Boltzmann equation. In *Theory and application of random fields (Bangalore, 1982)*, vol. 49 of *Lecture Notes in Control and Inform. Sci.* Springer, Berlin, 1983, pp. 258–267.
- [56] VILLANI, C. *Topics in Optimal Transportation*, vol. 58 of *Graduate Studies in Mathematics series*. American Mathematical Society, 2003.
- [57] VILLANI, C. *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

MAXIME HAURAY

UNIVERSITÉ D'AIX-MARSEILLE

LATP, UMR CNRS 7353  
CENTRE DE MATHÉMATIQUES ET INFORMATIQUE (CMI)  
39, RUE F. JOLIOT CURIE 13453 MARSEILLE CEDEX 13  
FRANCE

E-MAIL: [auray@cmi.univ-mrs.fr](mailto:auray@cmi.univ-mrs.fr)

STÉPHANE MISCHLER

UNIVERSITÉ PARIS-DAUPHINE & IUF  
CEREMADE, UMR CNRS 7534  
PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY 75775 PARIS CEDEX 16  
FRANCE

E-MAIL: [mischler@ceremade.dauphine.fr](mailto:mischler@ceremade.dauphine.fr)