Large Deviations for Young Measures and Statistical Mechanics of Infinite Dimensional Dynamical Systems with Conservation Law

Julien Michel, Raoul Robert

CNRS, Laboratoire d'analyse numérique, Université Lyon 1, 43 Bd du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Abstract: We describe an approach, based on Baldi's large deviation theorem, to carry out the statistical mechanics of a class of infinite dimensional dynamical systems.

I. Introduction

In previous works [32–34] we showed that large deviation theory (Sanov's theorem), used in a Young measure framework, provides an efficient tool to carry out the thermodynamic limit yielding the equilibrium states for two-dimensional Euler equations. The interest of this approach was reinforced by the confirmation both experimentally [10, 38] and by numerical simulations [40] of the theoretical predictions in the case of a vortex patch.

One aim of this paper is to provide a complete proof of the large deviation estimates used in [33]. Our proof is an application of Baldi's large deviation theorem [3] (in a slightly modified version). More generally, we show that Baldi's theorem is an elegant and powerful tool to carry out thermodynamic limits in various functional frameworks.

Another aim is to describe a class of infinite dimensional dynamical systems to which the theory can be applied.

It is out of the scope of this paper to review the main contributions to statistical hydrodynamics, nevertheless some comments and references are given in Sect. IV. It is also worth noticing that a work close to ours, although expounded from a more physical point of view, is developed by Miller et al. [26].

II. Baldi's Large Deviation Theorem and Thermodynamic Limits

Baldi's theorem gives general conditions under which a family of probability measures on a locally convex topological vector space has the large deviation property.

As we will see, it provides a powerful tool to carry out thermodynamic limits for infinite dimensional systems.

The Large Deviation Property. Let E be a locally convex Hausdorff topological vector space. We consider a family μ_h, $h > 0$, of Borel probability measures on E.

We will say (see for example Varadhan [42] or Ellis [12]) that the family μ_h has the large deviation property with constants $\lambda(h)$ and rate function L iff:

(i) $\lambda(h)$ is > 0 and $\lim\limits_{h \to +\infty} \lambda(h) = +\infty$.

(ii) $L: E \to [0, +\infty]$ is a lower semi-continuous functional on E (not identical to $+\infty$). Moreover L is inf-compact, that is: the set $\{\nu \mid L(\nu) \leq b\}$ is compact for all real number b.

(iii) For every Borel subset A of E, we have:

$$-\Lambda(\overset{\circ}{A}) \leq \liminf_{h \to \infty} \frac{1}{\lambda(h)} \operatorname{Log} \mu_h(A) \quad \text{and}$$

$$\limsup_{h \to \infty} \frac{1}{\lambda(h)} \operatorname{Log} \mu_h(A) \leq -\Lambda(\bar{A}),$$

where $\Lambda(A) = \inf\limits_{\nu \in A} L(\nu)$.

The functional L is also usually called the information functional, and $-L$ the entropy functional.

Let E' be the topological dual of E, endowed with the weak-star topology $\sigma(E', E)$. For a Borel probability measure μ on E, we define its Laplace transform:

$$\hat{\mu}(\varphi) = \int_E \exp(\langle \varphi, \nu \rangle) d\mu(\nu), \quad \text{for } \varphi \in E'.$$

As it is well known, $\hat{\mu}$ is a convex, lower semi-continuous and proper functional on E'. The same is true for the functional $\operatorname{Log} \hat{\mu}(\varphi)$.

Baldi's Theorem 2.1. *Let μ_h be a family of Borel probability measures on E, satisfying the following assumptions:*
(1) There is a function $\lambda(h)$ as in (i) *such that*

$$\lim_{h \to \infty} \frac{1}{\lambda(h)} \operatorname{Log} \hat{\mu}_h(\lambda(h)\varphi) = F(\varphi),$$

where F is a convex, lower semi-continuous and proper functional on E' which is finite on a neighborhood of the origin.
(2) Compacity assumption:
    For every $R > 0$, there is a compact set $K_R \subset E$ such that

$$\limsup_{h \to +\infty} \frac{1}{\lambda(h)} \operatorname{Log} \mu_h(K_R^c) \leq -R.$$

Let us denote by L the Young-Fenchel transform of F, that is:

$$L(\nu) = \sup_{\varphi \in E'} (\langle \varphi, \nu \rangle - F(\varphi)), \quad \text{for } \nu \in E.$$

L is a convex, lower semi-continuous, and proper functional on E.

    Baldi's theorem states that under the assumptions (1) *and* (2) *the upper bound in* (iii) *holds.*

    If we suppose that L has some additional strict-convexity property, we can also derive the lower bound. We will suppose that L satisfies the following condition.

(3) *For every real number r, the set $A_r = \{\nu \mid L(\nu) \le r\}$ is the closure of the subset of the points ν of A_r, where the subdifferential $\partial L(\nu)$ (see [12]) is non-empty and contains an element φ such that:*

$$L(\nu') > L(\nu) + \langle \varphi, \nu' - \nu \rangle, \quad \text{for all } \nu' \ne \nu.$$

Then Baldi's theorem asserts that under the hypothesis (1), (2), (3) *the two bounds in* (iii) *hold. So, we see that the family μ_h has the large deviation property with constants $\lambda(h)$ and rate function L. Indeed, one easily checks that the functional L is inf-compact on E: for every real number b, the set A_b is closed and the lower bound applied to the open set K_{b+1}^c yields $A_b \subset K_{b+1}$ [with the notation of (2)].*

Comments. 1) In practice it may be difficult to check that the hypothesis (3) is satisfied. In fact, Baldi's proof works as well with the following weaker hypothesis (3').

(3') For every ν such that $L(\nu) < +\infty$, for every open set O containing ν and every $\varepsilon > 0$, there is $\nu_1 \in O$ such that $L(\nu_1) \le L(\nu) + \varepsilon$ and L is strictly convex at ν_1, that is: $\exists \varphi \in \partial L(\nu_1)$ such that

$$L(\nu') > L(\nu_1) + \langle \varphi, \nu' - \nu_1 \rangle, \quad \text{for all } \nu' \ne \nu_1.$$

2) L is strictly convex at ν if, for example, $\partial L(\nu)$ is non-empty and

$$L(t\nu + (1-t)\nu') < tL(\nu) + (1-t)L(\nu')$$

for all $0 < t < 1$, $\nu' \in \mathrm{dom}\, L$, $\nu' \ne \nu$.

3) In the case where only the hypotheses (1), (2) are satisfied, as we have seen, Baldi's theorem gives an upper bound. But the functional L may fail to be inf-compact in that case. Nevertheless we can see that the set

$$A_0 = \{\nu \in E \mid L(\nu) = 0\} \quad \text{is non-empty}.$$

Notice first that we obviously have $F(0) = 0$ and since F is also the Young-Fenchel transform of L, we get:

$$\inf_{\nu \in E} L(\nu) = 0.$$

Furthermore, we have $\mu_h(K_1) + \mu_h(K_1^c) = 1$, and from (2) we know that $\mu_h(K_1^c) \to 0$ (when $h \to \infty$).

Now, if A_0 were empty we should have $\Lambda(K_1) > 0$. Then, applying Baldi's theorem we should have $\mu_h(K_1) \to 0$; this would yield a contradiction. Moreover, one can easily deduce that for any open set U containing A_0 there is a number $\alpha > 0$ such that:

$$\mu_h(U^c) \le \exp(-\lambda(h)\alpha), \quad \text{for } h \text{ large enough}.$$

We shall say that the family μ_h concentrates about the set A_0.

Thermodynamic Limits and the Concentration Property. When dealing with thermodynamic limits one usually encounters the following situation, which we summarize here in an abstract form. Let δ_h be a family of random variables with values in a Hausdorff locally convex topological vector space E. δ_h generally comes from some finite dimensional approximation of an infinite dimensional system. If we can prove that, for h large, with a high probability, δ_h remains in a neighborhood of some points ν^* of E, then ν^* is the equilibrium state of our system, and the thermodynamic limit is performed.

Large deviation theory shows us that such a situation is very common. We will assume in the sequel that the family δ_h (or the associated probability distributions μ_h on E) has the large deviation property with constants $\lambda(h)$ and rate function L.

Since $\mathrm{Prob}(\delta_h \in E) = 1$, for all h, we have $\inf_{\nu \in E} L(\nu) = 0$ and the set A_0 is a non-empty compact subset of E. And, as in the above comment 3, for every open set U containing A_0, there is some $\alpha > 0$, such that:

$$\mathrm{Prob}(\delta_h \in U^c) \leq \exp(-\lambda(h)\alpha), \quad \text{for large } h.$$

That is, the family δ_h concentrates about the set A_0 which is the equilibrium set of the system.

In our "microcanonical" approach, we will study now the situation where δ_h satisfies some given constraints (it would be more correct to say that we introduce some conditioning on the random variables δ_h). These constraints will be given, for example, by the constants of the motion of an infinite dimensional dynamical system. We introduce the constraints in the general form $\delta_h \in \mathscr{E}$, where $\mathscr{E}$ is some subset of E. Of course, since δ_h comes from a finite dimensional approximation, the ideal constraints $\delta_h \in \mathscr{E}$ will not be exactly satisfied, but only up to some approximation given by an open neighborhood of 0 in E, W. Let us denote $\mathscr{E}_W = \mathscr{E} + W$. We shall then consider $\delta_h \in \mathscr{E}_W$.

Let us now give a definition.

Definition. Let $\mathscr{E}, \mathscr{E}^*$ be subsets of E, we will say that δ_h concentrates about $\mathscr{E}^*$ conditionally to $\mathscr{E}$ iff:

(i) $$\forall W', \liminf_{h \to \infty} \frac{1}{\lambda(h)} \mathrm{Log}\, \mathrm{Prob}(\delta_h \in \mathscr{E}_{W'}) > -\infty,$$

(ii) $$\forall W^*, \ \exists \alpha > 0, \ \exists W, \ \forall W'$$
$$\frac{\mathrm{Prob}(\delta_h \in \mathscr{E}_W \backslash \mathscr{E}^*_{W^*})}{\mathrm{Prob}(\delta_h \in \mathscr{E}_{W'})} \leq \exp(-\lambda(h)\alpha), \quad \text{for } h \text{ large enough}.$$

Here W^*, W, W' denote open neighborhoods of 0 in E.

Remarks. 1) Heuristically, this definition means that if we know that δ_h takes its values in a neighborhood of $\mathscr{E}$, then it will be in a neighborhood of $\mathscr{E}^*$ with a high probability.
2) As previously noticed, we have to widen the sets $\mathscr{E}, \mathscr{E}^*$ into open neighborhoods. In fact $\mathrm{Prob}(\delta_h \in \mathscr{E})$ is not defined for an arbitrary subset $\mathscr{E}$; and even if $\mathscr{E}$ is a Borel subset, it can be zero.
3) The condition (i) ensures that, when $h \to \infty$, $\mathrm{Prob}(\delta_h \in \mathscr{E}_{W'})$ cannot be too small.

Now we derive the following concentration result which will be useful to carry out thermodynamic limits.

Concentration Theorem 2.2. *We suppose that δ_h has the large deviation property with constants $\lambda(h)$ and rate function L. Let $\mathscr{E}$ be a non-empty closed subset of E and $\mathscr{E}^*$ the subset of $\mathscr{E}$, where L achieves its minimum value on $\mathscr{E}$. Then δ_h concentrates about $\mathscr{E}^*$ conditionally to $\mathscr{E}$.*

Proof. L is inf-compact on E, thus $\mathscr{E}^*$ is always non-empty. In the case where $\inf_{\nu \in \mathscr{E}} L(\nu) = +\infty$, we obviously have $\mathscr{E}^* = \mathscr{E}$. In this case, we may say, by

convention, that δ_h concentrates about $\mathscr{E}^*$ cond. to $\mathscr{E}$. Otherwise, by the l.d.p. (large deviation property), we have

$$\liminf \frac{1}{\lambda(h)} \mathrm{Log\,Prob}(\delta_h \in \mathscr{E}_{W'}) \geq - \inf_{\nu \in \mathscr{E}_{W'}} L(\nu) \geq - \inf_{\nu \in \mathscr{E}} L(\nu) \ ,$$

and (i) is satisfied.

Now, let W^* be given, and consider a real number $\alpha_1 > 0$ such that

$$\inf_{\nu \in \mathscr{E} \backslash \mathscr{E}_{W^*}^*} L(\nu) - \inf_{\nu \in \mathscr{E}} L(\nu) > \alpha_1 \ .$$

We shall use the following result.

Lemma 2.3. *Let $\mathscr{E}, \mathscr{F}$ be closed subsets of E, then we have*

$$\inf_{\nu \in \mathscr{F} \cap \overline{\mathscr{E}}_W} L(\nu) \to \inf_{\nu \in \mathscr{F} \cap \mathscr{E}} L(\nu) \quad (when \ W \to 0) \ .$$

Proof. Let us denote by $\mathscr{W}$ the filter of the neighborhoods of 0 in E. For $W \in \mathscr{W}$, we define

$$l(W) = \inf_{\nu \in \mathscr{F} \cap \overline{\mathscr{E}}_W} L(\nu) \ , \quad l^* = \sup_{W \in \mathscr{W}} l(W) \ , \quad l^{**} = \inf_{\nu \in \mathscr{F} \cap \mathscr{E}} L(\nu) \ .$$

We obviously have $l(W) \to l^*$ (when $W \to 0$), and $l^* \leq l^{**}$.

If $l^* = +\infty$, there is nothing to prove, let us suppose $l^* < +\infty$.

Let us take $\varepsilon > 0$ and, for each W, we choose $\nu_W \in \mathscr{F} \cap \overline{\mathscr{E}}_W$ such that $l(W) \leq L(\nu_W) \leq l^* + \varepsilon$.

The set $\{\nu \in E \mid L(\nu) \leq l^* + \varepsilon\}$ is a compact subset of E, and there is a subfilter $\mathscr{W}'$ of $\mathscr{W}$ for which ν_W converges towards ν^*. We have $\nu^* \in \mathscr{F} \cap \mathscr{E}$, and the lower semi-continuity of L implies:

$$L(\nu^*) \leq \liminf_{\mathscr{W}'} L(\nu_W) \leq l^* + \varepsilon \ ,$$

from where $l^{**} \leq l^* + \varepsilon$, and the lemma is proved.

From the lemma we deduce that there is a number $\alpha > 0$ such that, for W small enough, we have:

$$\inf_{\nu \in \overline{\mathscr{E}}_W \backslash \mathscr{E}_{W^*}^*} L(\nu) - \inf_{\nu \in \mathscr{E}_{W'}} L(\nu) > \alpha \ , \quad \text{for all } W' \ .$$

Applying the l.d.p., we get:

$$\limsup \frac{1}{\lambda(h)} \mathrm{Log\,Prob}(\delta_h \in \mathscr{E}_W \backslash \mathscr{E}_{W^*}^*) \leq -\Lambda(\overline{\mathscr{E}_W \backslash \mathscr{E}_{W^*}^*})$$

$$\leq -\Lambda(\overline{\mathscr{E}}_W \backslash \mathscr{E}_{W^*}^*) \ ,$$

and

$$\liminf \frac{1}{\lambda(h)} \mathrm{Log\,Prob}(\delta_h \in \mathscr{E}_{W'}) \geq -\Lambda(\mathscr{E}_{W'}) \ .$$

Now, we have

$$\limsup \frac{1}{\lambda(h)} \operatorname{Log} \frac{\operatorname{Prob}(\delta_h \in \mathscr{E}_W \backslash \mathscr{E}_{W^*}^*)}{\operatorname{Prob}(\delta_h \in \mathscr{E}_{W'})}$$

$$\leq \limsup \frac{1}{\lambda(h)} \operatorname{Log} \operatorname{Prob}(\delta_h \in \mathscr{E}_W \backslash \mathscr{E}_{W^*}^*)$$

$$- \liminf \frac{1}{\lambda(h)} \operatorname{Log} Prob(\delta_h \in \mathscr{E}_{W'})$$

$$\leq -\Lambda(\overline{\mathscr{E}}_W \backslash \mathscr{E}_{W^*}^*) + \Lambda(\mathscr{E}_{W'}) < -\alpha \,,$$

from which we get (ii) and the theorem is proved.

Remark. The set $\mathscr{E}^*$ in which δ_h approximately remains with a high probability is the equilibrium set of the system. If $\mathscr{E}^*$ does not reduce to a point (the equilibrium state) we are in a phase transition situation. We shall see in the following section how we can use this concentration result to derive a maximum-entropy principle for Young measures. We shall use this principle in Sect. 4 to get the equilibrium states of some infinite dimensional dynamical systems.

III. A Maximum-Entropy Principle for Young Measures

It currently happens when dealing with a limit process for a sequence of bounded measurable functions that the sequence does not converge and shows an oscillating limit behavior, whereas some estimates and conservation laws hold. In such a case, the concept of Young measure has been found relevant to describe the behavior of the sequence (examples can be found in hyperbolic systems of conservation laws, homogenization, hydrodynamics...). Young measures can be viewed as giving a macroscopic description of the system, whereas the bounded measurable functions are all the microscopic states.

We use the results of Sect. 2 to derive a maximum entropy principle for Young measures. That is: the macrostate (Young measure) which realizes the maximum of an entropy functional has a natural concentration property (a large majority of the microstates satisfying a given set of constraints are in a neighborhood of that macrostate). It turns out that this entropy functional is the classical Kullback entropy (see Sanov's theorem).

Young Measures. Throughout this section X, Y will denote two locally compact separable and metrizable topological spaces. Let us suppose that a positive Borel measure dx is given on X.

Let us recall that Young measures [44] are a natural way to generalize the notion of measurable mapping from X to Y: at any point $x \in X$, we no longer have a well determined value, but only some probability distribution on Y. In other words, a Young measure ν is a measurable mapping $x \to \nu_x$ from X to the set $M_1(Y)$ of the Borel probability measures on Y endowed with the narrow topology.

Clearly, ν defines a positive Borel measure on $Y \times Y$ (that we will also denote by ν) by:

$$\langle \nu, f \rangle = \int_X \langle \nu_x, f(x, \cdot) \rangle dx \,,$$

for every real function $f(x,y)$, continuous and compactly supported on $X \times Y$ ($f \in C_c(X \times Y)$). Moreover, for $f(x) \in C_c(X)$, we have

$$\langle \nu, f \rangle = \int\limits_X f(x)\, dx\,,$$

that is, the projection of ν on X is dx.

It is well known [19] that this property gives an equivalent definition of Young measures. That is, for any positive Borel measure ν on $X \times Y$ whose projection on X is dx, there is a measurable mapping $x \to \nu_x$ such that the above formula holds. The mapping $x \to \nu_x$ is unique up to the dx-almost everywhere equality.

To any measurable mapping $f : X \to Y$, we associate the Young measure $\delta_f : x \to \delta_{f(x)}$, Dirac mass at $f(x)$.

We shall make two additional assumptions:

(*) The measure dx is diffuse and of finite total mass $dx(X) = |X|$.

(**) There is a distance function $d(x, x')$ giving the topology of X, such that: for all $\varepsilon > 0$, there is a finite partition of X into measurable subsets $\mathscr{X} = \{X^i \mid i = 1, \ldots, n(\mathscr{X})\}$ with $|X^i| = |X^j|$ for all i, j (we shall say that $\mathscr{X}$ is an equipartition of X), and satisfying $d(\mathscr{X}) \leq \varepsilon$, where $d(\mathscr{X}) = \sup\limits_i \sup\limits_{x, x' \in X^i} d(x, x')$ is the diameter of $\mathscr{X}$.

Notice that (*) and (**) implies that $|A|$ goes to zero when the diameter of a measurable set A goes to zero.

Hypotheses (*), (**) are satisfied, for example, if X is an open convex and bounded subset of $\mathbb{R}^n$ with $dx =$ Lebesgue's measure, and also if we consider any image of $\bar{X}$ by a dx-preserving homeomorphism.

We shall denote by M the convex set of Young measures on $X \times Y$, and we recall some useful properties.

– M is closed in the space $M_b(X \times Y)$ of all bounded Radon measures on $X \times Y$ (with the narrow topology), the narrow topology is equal on M to the vague topology (weak topology associated to the continuous compactly supported functions) and it is metrizable. Furthermore if Y is compact then M is compact. In the sequel M will be endowed with the narrow topology.

– $\{\delta_f \mid f : X \to Y \text{ measurable}\}$ is a dense subset of M. The proof of this property can be found in [5] for the case Y compact. The general case follows by approximation. Approximate first (for the vague topology) a given Young measure ν by $\nu_{\mathscr{X}}$ (as in the proof of Theorem 3.1 below) which is constant, equal to ν^i, on each set X^i of an equipartition $\mathscr{X}$, and then approximate each ν^i by a probability measure with compact support.

A Large Deviation Property. Suppose now that we are given a basic Borel probability measure π_0 on Y. Then to any equipartition $\mathscr{X}$ of X we can associate a Borel probability measure $\mu_{\mathscr{X}}$ on M in the following way. We take $\mathscr{Y}_1, \ldots, \mathscr{Y}_n$, $n(\mathscr{X})$ Y-valued independent random variables with the same distribution π_0. We consider the random function

$$f_{\mathscr{X}} = \sum_i \mathscr{Y}_i \mathbf{1}_{X^i}\,,$$

where $\mathbf{1}_{X^i}$ is the characteristic function of the set X^i. We denote $\delta_{\mathscr{X}}$ the Young measure associated to $f_{\mathscr{X}}$ and $\mu_{\mathscr{X}}$ the probability distribution on M of the random variable $\delta_{\mathscr{X}}$. Now, we can state the main result of this section.

Theorem 3.1. *When $d(\mathscr{X}) \to 0$, the family $\mu_{\mathscr{X}}$ has the large deviation property with constants $n(\mathscr{X})/|X|$ and rate function $I_\pi(\nu)$, where $\pi = dx \otimes \pi_0$ and $I_\pi(\nu)$ is the classical Kullback information functional (see Varadhan [42]), defined on M by:*

$$I_\pi(\nu) = \int\limits_{X \times Y} \text{Log}\, \frac{d\nu}{d\pi}\, d\nu, \text{ is } \nu \text{ is absolutely continuous with respect to } \pi\,,$$

$$I_\pi(\nu) = +\infty \text{ otherwise}\,.$$

Proof. We shall show that Baldi's theorem applies to that case. We take for E the space $M_b(X \times Y)$ endowed with the vague topology. Taking the vague instead of the narrow topology on E makes no change on M but gives the compacity property (2) for $\mu_{\mathscr{X}}$, since the closure $\bar{M}$ of M in M_b for the vague topology is compact. Then we have $E' = C_c(X \times Y)$.

We check now that the assumption (1) is satisfied. We take $\varphi(x, y) \in C_c(X \times Y)$ and compute the limit in (1),

$$\hat{\mu}_{\mathscr{X}}\left(\frac{n(\mathscr{X})}{|X|}\,\varphi\right) = \int\limits_E \exp\left(\left\langle \nu, \frac{n}{|X|}\,\varphi \right\rangle\right) d\mu_{\mathscr{X}}(\nu)$$

$$= \prod_i \int\limits_Y \exp\left(\frac{1}{|X^i|} \int\limits_{X^i} \varphi(x, y)\, dx\right) d\pi_0(y)\,.$$

Let us define

$$\varphi_{\mathscr{X}}(x, y) = \frac{1}{|X^i|} \int\limits_{X^i} \varphi(x, y) dx\,, \quad \text{if } x \in X^i\,,$$

then we have

$$\frac{|X|}{n}\, \text{Log}\, \hat{\mu}_{\mathscr{X}}\left(\frac{n}{|X|}\,\varphi\right) = \int\limits_X dx\, \text{Log}\left(\int\limits_Y \exp(\varphi_{\mathscr{X}}(x, y)) d\pi_0(y)\right)\,.$$

When $d(\mathscr{X})$ goes to zero, $\varphi_{\mathscr{X}}(x, y)$ converges towards $\varphi(x, y)$ and by Lebesgue's theorem, we get:

$$F(\varphi) = \int\limits_X dx\, \text{Log}\left(\int\limits_Y \exp(\varphi(x, y)) d\pi_0(y)\right)\,.$$

The functional F is everywhere finite on $C_c(X \times Y)$, it is convex since it is a limit of convex functions, and it is obviously continuous for the norm topology of C_c. Thus it is also lower semi-continuous for the weak topology $\sigma(C_c, M_b)$.

Now, to check that the hypothesis (3') is satisfied, we have to compute the Young-Fenchel transform L of F. For $\nu \in M_b(X \times Y)$, we have:

$$L(\nu) = \sup_{\varphi \in C_c} (\langle \nu, \varphi \rangle - F(\varphi))\,. \tag{4}$$

First Point. We have $L(\nu) = +\infty$ if $\nu \notin M$. We can easily see that $L(\nu) = +\infty$ unless ν is positive. Indeed, if ν is not positive, there is some $\varphi \in C_c$, $\varphi \leq 0$, such that $\langle \nu, \varphi \rangle > 0$. Then we have $F(\varphi) \leq 0$ and $\langle \nu, \varphi \rangle - F(\varphi) > 0$. Taking the supremum over the $\lambda\varphi$, $\lambda > 0$, gives the result.

Now we prove that $L(\nu) < +\infty$ implies $\nu \in M$. Let us notice that the supremum in (4) can be extended to any $\varphi \in C_b(X \times Y)$. Thus for any $\varphi(x)$ continuous and bounded on X, we have

$$+\infty > L(\nu) \geq \langle \nu, \lambda\varphi \rangle - \lambda \int_X \varphi(x)dx \,, \quad \text{for all real number } \lambda \,,$$

hence $\langle \nu, \varphi \rangle = \int_X \varphi(x)dx$, that is $\nu \in M$.

Second Point. On M we have $L(\nu) = I_\pi(\nu)$. Let us first prove that $L(\nu) \leq I_\pi(\nu)$. Of course we may suppose that $I_\pi(\nu) < +\infty$, in which case ν is absolutely continuous with respect to $\pi : \nu = \varrho(x, y)\pi$, where ϱ is a π-integrable Borel function. Then we follow the method given by (Varadhan [42], Theorem 4.1) and use the inequality

$$ab \leq b \operatorname{Log} b + \exp(a - 1) \,, \quad \text{for all real number } a, \text{ and for } b \geq 0 \,.$$

For any $\varphi \in C_c(X \times Y)$, and any $c(x)$, we have:

$$\int_Y (\varphi - c)\varrho \, d\pi_0(y) \leq \int_Y \varrho \operatorname{Log} \varrho \, d\pi_0(y) + \int_Y \exp(\varphi - c - 1) \, d\pi_0(y) \,,$$

taking $c(x) = \operatorname{Log}\left(\int_Y \exp(\varphi) \, d\pi_0 \right) - 1$, and using $\int_Y \varrho(x, y) \, d\pi_0(y) = 1$, we get:

$$\int_Y \varphi\varrho \, d\pi_0 \leq \int_Y \varrho \operatorname{Log} \varrho \, d\pi_0 + \operatorname{Log}\left(\int_Y \exp(\varphi) \, d\pi_0 \right) \,,$$

integrating this over X yields $L(\nu) \leq I_\pi(\nu)$.

Let us now prove the converse inequality $L(\nu) \geq I_\pi(\nu)$. It is well known that:

$$I_\pi(\nu) = \sup_{\varphi \in C_c} \left(\langle \nu, \varphi \rangle - |X| \operatorname{Log}\left(\int \exp\varphi \, \frac{dx}{|X|} \, d\pi_0 \right) \right) \,,$$

the result thus follows by applying Jensen's inequality to the convex function $- \operatorname{Log}$.

It remains to check that the strict-convexity condition of Baldi's theorem is satisfied. This is not obvious since if ν is such that $L(\nu) < +\infty$, $\partial L(\nu)$ may be empty. In fact, we shall see that $(3')$ is satisfied. At first we prove that L is strictly convex at the Young measures $\nu = \exp(\varphi)\pi$, for $\varphi \in C_c(X \times Y)$. Since we know that $I_\pi(\nu)$ is strictly convex on its domain, we only have to prove that for such Young measures $\varphi \in \partial L(\nu)$. Let us consider the functional

$$I(\varrho) = \int_{X \times Y} \varrho \operatorname{Log} \varrho \, d\pi \,.$$

One can easily check that for any Borel function ϱ, $0 < c \leq \varrho \leq C < +\infty$, the functional I is Fréchet-derivable at ϱ in the space $L^\infty(\pi)$. From the convexity of I, it comes:

$$I(\varrho') \geq I(\varrho) + \int_{X \times Y} (1 + \operatorname{Log} \varrho)(\varrho' - \varrho) \, d\pi \,, \quad \text{for all } \varrho' \geq 0, \, \varrho' \in L^\infty(\pi) \,.$$

It follows that for all ν in $M_b(X \times Y)$ we have

$$L(\nu) \geq L(\varrho\pi) + \langle \nu - \varrho\pi, \operatorname{Log} \varrho \rangle \,.$$

Hence for $\varrho = \exp(\varphi)$, we get $\varphi \in \partial L(\exp(\varphi)\pi)$.

Now let ν be given such that $L(\nu) < +\infty$. We have $\nu = \varrho\pi$. For any equipartition $\mathscr{X}$ of X, we define the densities

$$\varrho^i(y) = \frac{1}{|X^i|} \int\limits_{X^i} \varrho(x, y)\, dx \,,$$

$$\varrho_{\mathscr{X}}(x, y) = \varrho^i(y), \text{ for } x \in X^i; \text{ and } \nu_{\mathscr{X}} = \varrho_{\mathscr{X}}\pi \,.$$

One easily checks that $\nu_{\mathscr{X}}$ converges towards ν, for the vague topology, when $d(\mathscr{X}) \to 0$.

Furthermore, we have

$$L(\nu_{\mathscr{X}}) = \sum_i |X^i| I_{\pi_0}(\varrho^i \pi_0) \,.$$

By Jensen's inequality, we have

$$\varrho^i(y) \operatorname{Log} \varrho^i(y) \leq \int\limits_{X^i} \varrho(x, y) \operatorname{Log} \varrho(x, y) \frac{dx}{|X^i|} \,,$$

from which $L(\nu_{\mathscr{X}}) \leq L(\nu)$.

Now we have to approximate the measure $\nu_{\mathscr{X}}$. The following lemma, whose proof is an easy exercise, will be useful.

Lemma 3.2. *For any given density $\varrho(y)$ in $L^1(\pi_0)$, there is a sequence $\varrho_n(y)$ of continuous functions satisfying:*
 (i) $\varrho_n(y) > 0$, $\int \varrho_n(y)\, d\pi_0(y) = 1$,
 (ii) $\varrho_n = 1$ *out of a compact set*,
 (iii) $\int |\varrho_n - \varrho|\, d\pi_0 \to 0$ *when $n \to \infty$*,
 (iv) $I_{\pi_0}(\varrho_n \pi_0) \to I_{\pi_0}(\varrho \pi_0)$.
 Furthermore if $|\varrho|_\infty \leq m$, we can choose ϱ_n such that $|\varrho_n|_\infty \leq 2m + 1$.

Then, for each $\varrho^i(y)$, we consider the sequence ϱ_n^i given by the lemma. One can also find sequences $\alpha_n^i(x)$ of continuous functions with compact support on X, such that

$$0 \leq \alpha_n^i(x) \leq 1 \,, \quad \text{and} \quad \int\limits_X |1_{X^i} - \alpha_n^i|\, dx \leq \frac{1}{n} \,.$$

Now let us consider the sequence of continuous functions

$$\varrho_n(x, y) = \sum_i \alpha_n^i(x) \varrho_n^i(y) + 1 - \sum_i \alpha_n^i(x) \,;$$

we easily check that $\varrho_n > 0$, $\varrho_n \pi$ is a Young measure, $\varrho_n = 1$ out of a compact subset of $X \times Y$ and

$$\int\limits_{X \times Y} |\varrho_n - \varrho_{\mathscr{X}}|\, d\pi \to 0 \quad \text{when } n \to \infty \,.$$

It follows that $\varrho_n \pi$ converges towards $\nu_{\mathscr{X}}$ for the vague topology and $L(\varrho_n \pi) \to L(\nu_{\mathscr{X}})$ (it is a straightforward application of Lebesgue's theorem if $\varrho_{\mathscr{X}}$ is bounded, the general case can be reduced to that case). Since $\varphi_n = \operatorname{Log} \varrho_n \in C_c(X \times Y)$, the hypothesis $(3')$ is satisfied, Baldi's theorem applies and the proof is complete.

In this Young measure framework, Theorem 2.2 yields the following result.

Corollary 3.3. *Let $\mathscr{E}$ be a non-empty closed subset of M, $\mathscr{E}^*$ the subset of $\mathscr{E}$ where the functional I_π achieves its minimum value on $\mathscr{E}$. Then $\delta_{\hat{x}}$ concentrates about $\mathscr{E}^*$ conditionally to $\mathscr{E}$.*

Remark. Theorem 3.1 appears as a generalization of the well-known Sanov's theorem [36]. Indeed, apply the contraction principle to the mapping $\nu \rightarrow \int \nu_x \, dx$.

We refer to [34] for a direct proof of Corollary 3.3 using Sanov's theorem.

IV. Statistical Equilibrium States
for a Class of Infinite Dimensional Dynamical Systems

1) A Class of Infinite Dimensional Dynamical Systems. Now we apply the tools of the previous sections to evolution equations of the form:

$$\text{(I)} \quad \left\{ \begin{array}{l} q_t + \operatorname{div}(q\mathbf{u}) = 0 \,, \\ \mathbf{u} = \mathscr{L}(q), \ \operatorname{div}(\mathbf{u}) = 0 \end{array} \right\},$$

where $q(t, \mathbf{x})$ is some scalar density function defined on $\mathbb{R} \times \Omega$ (Ω is a bounded connected smooth domain of $\mathbb{R}^d$), $\mathbf{u}(t, \mathbf{x})$ is an incompressible velocity field taking its values in $\mathbb{R}^d$, which can be recovered from q by solving a P.D.E. system. Thus $\mathscr{L}$ denotes a (not necessarily linear) integro-differential operator.

Let us give some well-known examples of such systems.
1. The simplest example of (I) is the linear transport equation, where $\mathbf{u}(\mathbf{x})$ is a given incompressible velocity field on Ω.
2. 2D incompressible Euler equations in the usual velocity-vorticity formulation are clearly of the form (I). Take for q the vorticity: $q = \operatorname{curl} \mathbf{u}$, then $\mathbf{u}$ is given by

$$\left\{ \begin{array}{l} \operatorname{curl} \mathbf{u} = q \,, \\ \operatorname{div} \mathbf{u} = 0 \,, \\ \mathbf{u} \cdot \mathbf{n} = 0 \ \text{on} \ \partial\Omega \,. \end{array} \right\}$$

This is a particular case of the quasi-geostrophic model used in geophysical fluid dynamics [24], in which q is the potential vorticity, and $\mathbf{u}$ is given by:

$$\left\{ \begin{array}{l} \mathbf{u} = \operatorname{curl} \psi \,, \\ -\Delta\psi + r^2\psi = q + f \,, \\ \psi = 0 \ \text{on} \ \partial\Omega \,. \end{array} \right\}$$

Here $f(\mathbf{x})$ is a given function (topography) on Ω, ψ the usual scalar stream function and r^2 a non-negative constant.
3. Collisionless kinetic equations such as the Boltzmann-Poisson equation of stellar dynamics and Vlasov-Maxwell equations of plasmas can also be written in the form (I).

The first step in our program is to define a flow associated to (I) on the phase space $L^\infty(\Omega)$. Unfortunately, to our knowledge, there is no general existence-uniqueness result for the Cauchy problem for system like (I). Examples 1 and 2 are well known, but for kinetic equations, although some existence results are available [2, 11], it seems that the uniqueness problem is not yet solved.

To proceed further, we shall make the following hypothesis.

(H1) The system (I) defines a flow $\Phi_t : L^\infty(\Omega) \to L^\infty(\Omega)$, whose restriction to any ball $L_r^\infty = \{q : |q|_\infty \leq r\}$ is continuous for the strong L^2 topology.

So, for any given initial datum $q(\mathbf{x})$, we have a velocity field $\mathbf{u}(t, \mathbf{x})$; to this field we associate a Lagrangian flow $\varphi_t(\mathbf{x})$, defined by:

$$\left\{ \begin{array}{c} \dfrac{d\varphi_t(\mathbf{x})}{dt} = \mathbf{u}(t, \varphi_t(\mathbf{x})), \\ \varphi_0(\mathbf{x}) = \mathbf{x}. \end{array} \right\}$$

We shall suppose:

(H2) The Lagrangian flow φ_t is globally (in time) defined on Ω. The mappings φ_t are volume-preserving homeomorphisms of Ω, satisfying an estimate:

$$|\varphi_t(\mathbf{x}) - \varphi_t(\mathbf{y})| \leq C(T, r) |\mathbf{x} - \mathbf{y}|^{\alpha(T, r)}, \qquad \alpha(T, r) > 0,$$

for $|q|_\infty \leq r$, $\mathbf{x}, \mathbf{y} \in \Omega$, $t \in [0, T]$.

And we have $\Phi_t q(\mathbf{x}) = q(\varphi_t^{-1}(\mathbf{x}))$.

For any given $r > 0$, let us denote M_r the space of Young measures on $\Omega \times [-r, r]$. A natural extension of Φ_t to M_r is given by:

$$(\bar{\Phi}_t \nu)_{\mathbf{x}} = \nu_{\varphi_t^{-1}(\mathbf{x})},$$

where φ_t is the Lagrangian flow associated to the initial density

$$q(\mathbf{x}) = \bar{\nu}(\mathbf{x}) = \int a \, d\nu_{\mathbf{x}}(a).$$

Heuristically, we may say that the oscillations of the density function are merely frozen and convected by the velocity field associated to the mean density.

(H3) $\bar{\Phi}_t : M_r \to M_r$ is continuous for the narrow topology.

Notice that the density of L_r^∞ in M_r implies that such an extension is unique. Notice also that the topology on L_r^∞ induced by the narrow topology of M is equal to the strong L^2-topology (whereas the corresponding uniform structures are different).

Hypothesis H1, H2, H3 are obviously satisfied in Example 1 if $\mathbf{u}$ is a C^1 velocity field on $\bar{\Omega}$ which is tangent to the boundary. They are also satisfied in Example 2: this is the classical Youdovich's theorem in the case of Euler equations [43]; its generalization to the quasi-geostrophic model will be found in [23].

Constants of the Motion. For systems of the form (I), there is a family of constants of the motion which will play a crucial role. These are the functionals

$$C_f(q) = \int_\Omega f(q(x)) dx,$$

for any given continuous function f on $\mathbb{R}$. Let us define the distribution measure of q, π_q by $\langle \pi_q, f \rangle = C_f(q)$. Then π_q is conserved by the flow.

Notice that, in the case of Example 2, the quasi-geostrophic model has in a natural way a Hamiltonian structure [24], and these invariants are the well known Casimir invariants associated to the degeneracy of the Poisson brackets. According to each particular case, we will have also to take into account the classical constants of the motion of the system (energy, angular momentum, . . .).

2) Motivation to Define Equilibrium States. Observations, experiments, or numerical simulations in 2D fluid dynamics display in many cases the appearance of well

organized structures. This is a striking feature of 2D fluid turbulence. We observe that after a complex evolution the system converges towards some stationary state which is usually called a coherent structure. Since in 2D turbulence the energy of the flow is conserved Onsager [30] suggested that such structures were equilibrium states and it would be possible to predict them by a relevant statistical mechanics approach. The same idea was later rediscovered by Lynden-Bell [22] in the context of stellar dynamics. Lynden-Bell tried to explain the particular distribution of light observed in elliptical galaxies. We do not intend to discuss further these enlightening and pioneering works. The interested reader can find some information in the papers [9, 13, 21, 26, 27, 35] (this list is of course far from exhaustive). To summarize our view, let us say that these works have clearly shown the need for an appropriate statistical mechanics approach working in a sufficiently broad frame.

3) Previous Attempts to Define Equilibrium States. A natural way to define equilibrium states is to construct invariant Gibbs measures on the phase space. Although this can be done for some infinite dimensional systems (see for example [15]), we do not know how to construct such measures on the natural phase space $L^\infty(\Omega)$ for systems like (I). In the case of Euler equations some work has been devoted to the study of Gibbs measures with formal densities given by the enstrophy and the energy [4], and also to Gibbs measures associated to the law of vorticity conservation along the trajectories of the fluid particles [6]. Unfortunately all these measures are supported by very "large" functional spaces so that not only the mean energy and enstrophy of these states are infinite but the phase space $L^\infty(\Omega)$ is of null measure. So, it is only at a formal level that this makes sense. Moreover this approach fails to give any prediction on the long time dynamics corresponding to a given initial vorticity function.

The most common approach is to use a convenient finite dimensional approximation of the system, possessing an invariant Liouville measure. Then one can consider the canonical measures associated to the constants of the motion and try to perform a thermodynamic limit in the space of generalized functions when the number of degrees of freedom goes to infinity. For example, for 2D Euler equations one can consider the N Fourier-mode approximation or the point-vortex approximation. Two difficulties arise in this approach. The first is to choose a relevant scaling to perform the limit; an interesting comment on this point can be found in [26], see also [16, 4]. The second is even more fundamental: generally, the approximate system will have less constants of the motion than the continuous one (I), so that the long time dynamics of that system may be very different from that of the continuous one. For more comments and references on these previous attempts see for example [33, 35, 26].

4) Our Approach. As expounded in [33] for 2D Euler equations, our approach is based on the following points.

1. Identify the long time limits of the dynamical system as Young measures. M_r is a suitable compactification of L_r^∞ since the narrow convergence (when t goes to infinity) of $\delta\Phi_t q$ towards some Young measure ν preserves the information given by the constants of the motion, that is, for all function f

$$\int_\Omega f(\Phi_t q(\mathbf{x}))\, d\mathbf{x} \to \int_\Omega \langle \nu_\mathbf{x}, f \rangle\, d\mathbf{x},$$

but the left-hand side is constant and equal to $\langle \pi_q, f \rangle$, so that:

$$\int_\Omega \nu_{\mathbf{x}} \, d\mathbf{x} = \pi_q \, .$$

The same kind of arguments applies to the other invariants. For example, in the case of Euler equations, since $\Phi_t q$ converges weakly towards $\bar{\nu}(\mathbf{x})$, we have for the energy $E(\Phi_t q) \to E(\bar{\nu})$, which is the energy of the Young measure ν, see [33].

Thus we see that the constants of the motion of (I) bring constraints on the possible long time limits.

2. For a given initial state q, the corresponding equilibrium state is the Young measure ν^* which minimizes the Kullback information functional $I_\pi(\nu)$ ($\pi = d\mathbf{x} \otimes \pi_0$, $|\Omega|\pi_0 = \pi_q$) under the constraints given by the constants of the motion. Thus ν^* (or more generally the set $\mathscr{E}^*$) will have the concentration property given by Corollary 3.3.

3. In other words, to get ν^*, we have computed the thermodynamic limit in the space M of the "microcanonical" measures $\mu_{\mathscr{X}}$ (defined in Sect. III) with the conditioning given by the constants of the motion. But the probability measure $\pi_{\mathscr{X}} = \overset{n}{\bigotimes} \pi_0$, on the space $E_{\mathscr{X}}$ of the density functions which are constant on each element of the equipartition $\mathscr{X}$, is not conserved by the flow Φ_t; so that the compatibility with the dynamics of the system is not a priori ensured.

In [33] the method was justified by an invariance theorem. That is, the concentration property is conserved by the extended flow $\bar{\Phi}_t$. The proof given in [33] works equally well for systems like (I) satisfying hypothesis H1, H2, H3. Indeed, the crucial point is the (obvious) fact that the functional $I_\pi(\nu)$ is conserved by the extended flow. Another straightforward consequence is that the equilibrium set $\mathscr{E}^*$ is conserved by $\bar{\Phi}_t$ ($\bar{\Phi}_t(\mathscr{E}^*) = \mathscr{E}^*$). So, if $\mathscr{E}^* = \{\nu^*\}$ then $\bar{\Phi}_t(\nu^*) = \nu^*$ and the mean density $\overline{\nu^*}(\mathbf{x})$ is a stationary solution of (I).

It was argued by Eyink and Spohn [13] that the above invariance property was not enough to provide a complete justification of the method. Their interesting comments lead us to develop here some further arguments.

Let us apply the contraction principle to the mapping $\nu \to \bar{\nu}$ from M_r into L_r^∞ endowed with the weak topology $\sigma(L^\infty, L^1)$ (we suppose that π_0 is supported by the interval $[-r, r]$). Then we deduce from Theorem 3.1 that the family of measures $\pi_{\mathscr{X}}$ has the large deviation property with rate function

$$J(f) = \inf_{\bar{\nu} = f} I_\pi(\nu) \, .$$

Obviously, the functional J is conserved by Φ_t.

Now, for any Borel subset A of L_r^∞, we have (using the notation $\approx$ as a convenient shortcut):

$$\pi_{\mathscr{X}}(\Phi_t(A)) \approx \exp\left(-\frac{n(\mathscr{X})}{|\Omega|} J(\Phi_t(A)) \right) = \exp\left(-\frac{n(\mathscr{X})}{|\Omega|} J(A) \right) ,$$

where $J(A) = \inf_{f \in A} J(f)$, so that

$$\pi_{\mathscr{X}}(\Phi_t(A)) \approx \pi_{\mathscr{X}}(A) \, , \quad \text{when } d(\mathscr{X}) \to 0 \, .$$

It means that the measures $\pi_{\mathscr{X}}$ are, in some approximate sense, conserved by the flow Φ_t. It appears that this argument can be sharpened to give something looking like a classical justification.

5) Towards a More Complete Justification. Let us start with some heuristical considerations. Let us suppose that the flow Φ_t is "well approximated" by some N-dimensional dynamical system:

$$(\Phi_N)_t : E_N \rightarrow E_N \,,$$

that is, for any given q, $(\Phi_N)_t q_N$ is close to $\Phi_t q$ in the L^2 norm (here q_N denotes an approximation of q in the L^2-sense); and this approximation is uniform for t in the interval $[0, T]$, for N large enough (depending on T).

Let us suppose now that $(\Phi_N)_t$ has an invariant Liouville measure μ_N on E_N. Of course, we cannot expect that $(\Phi_N)_t$ will have the same constants of the motion as Φ_t, but since the constants of the motion of Φ_t are uniformly continuous functionals on the compact space M_r, any finite subset of these functionals will be approximately conserved by the flow $(\Phi_N)_t$ during the time interval of interest (for N large enough).

Therefore, since we are interested in the long time dynamics of (I), say for example at time T, it is natural to consider the microcanonical measures $\tilde{\mu}_N$ obtained by conditioning μ_N by the constants of the motion of (I); and then to perform the thermodynamic limit of $\tilde{\mu}_N$ in the space M, which is the good candidate for the long time limits.

Of course this program is not easy to carry through in general. Nevertheless it is exactly what is done in Corollary 3.3 with the "Liouville" measure $\pi_{\mathscr{X}}$.

It remains to show that $\pi_{\mathscr{X}}$ is actually conserved by a flow on $E_{\mathscr{X}}$ which approximate Φ_t.

An Approximation of the Flow Φ_t. Let us begin with $t = 1$, let $\mathscr{X}$ be an equipartition of Ω and $E_{\mathscr{X}}$ the subset of the functions of L_r^∞ which are constant on the sets Ω^i of $\mathscr{X}$. We can construct an approximation of Φ_1 in the following way: For a given integer $p \geq 1$, we write

$$E_{\mathscr{X}} = \bigcup_{s=1}^{(2p)^{n(\mathscr{X})}} C^s \,,$$

where each C^s is a box which is a product of intervals of the form

$$I_k = \left] r\,\frac{k}{p}, r\,\frac{k+1}{p} \right], \qquad k = -p, \dots, p-1 \,.$$

First, we construct a mapping $\Psi_p : E_{\mathscr{X}} \rightarrow L_r^\infty$. For $q \in C^s$, $q = \sum q_i \mathbf{1}_{\Omega^i}$, $\Psi_p(q) = \sum q_i \mathbf{1}_{\varphi_1^s(\Omega^i)}$, where φ_1^s denotes the Lagrangian flow associated to q^s, the center of the box C^s. Now, we approximate this mapping Ψ_p by a mapping $\Psi_{p,\mathscr{X}}$ taking its values in $E_{\mathscr{X}}$. This can be done by means of the following lemma whose proof is given in the appendix.

Lemma 4.1. *Let Ω be an open bounded subset of $\mathbb{R}^d$, $\mathscr{X} = \{\Omega^i\}$, $\mathscr{O} = \{O^i\}$, two equipartitions of Ω with the same number n of elements. Then, there is a permutation σ of $\{1, \dots, n\}$ such that $\mathrm{vol}(\Omega^i \cap O^{\sigma(i)}) > 0$, for all i.*

We apply the lemma for each s and choose a permutation σ_s such that

$$\mathrm{vol}(\varphi_1^s(\Omega^i) \cap \Omega^{\sigma_s(i)}) > 0, \quad \text{for all } i \,.$$

Then we define the mapping $\Psi_{p,\mathscr{X}}$ by:

$$\Psi_{p,\mathscr{X}}\left(\sum q_i \mathbf{1}_{\Omega^i}\right) = \sum q_i \mathbf{1}_{\Omega}\sigma_s(i), \quad \text{if } q = \sum q_i \mathbf{1}_{\Omega^i} \in C^s.$$

Then we have the following result.

Proposition 4.2. *For $d(\mathscr{X})$ small enough and p large enough (depending on $d(\mathscr{X})$), $\Psi_{p,\mathscr{X}}$ uniformly approximates Φ_1 on $E_{\mathscr{X}}$. That is: $\Psi_{p,\mathscr{X}}(q)$ is close to $\Phi_1(q)$ in L_r^∞ endowed with the narrow uniform structure of M (uniformly with respect to $q \in E_{\mathscr{X}}$).*

Proof. At first, it is an easy exercise to check that for $d(\mathscr{X})$ small enough $\Psi_{p,\mathscr{X}}(q)$ is close to $\Psi_p(q)$ for the narrow uniform structure of M (uniformly for $q \in E_{\mathscr{X}}$). This follows from the fact that the Lagrangian flow φ_1 satisfies a Hölder estimate which is uniform for q in L_r^∞.

Now, let $\mathscr{X}$ fixed. For p large enough, $\Psi_p(q)$ will be close to $\Phi_1(q)$ in the L^2 norm (uniformly for $q \in E_{\mathscr{X}}$). Indeed, by (H1), Φ_1 is continuous from $E_{\mathscr{X}}$ into L_r^∞ (for the L^2 norm) and since $E_{\mathscr{X}}$ is compact, it is uniformly continuous on $E_{\mathscr{X}}$. The result follows by noticing that on L_r^∞ the uniform structure of the L^2 norm is stronger than the one induced by the narrow uniform structure of M.

It is clear that $\Psi_{p,\mathscr{X}}$ conserves the distribution π_q. For any probability measure π_0 on $[-r, r]$, the measure $\overset{n}{\bigotimes}\pi_0$ on $E_{\mathscr{X}}$ is also obviously conserved on each box C^s. But it is not globally conserved, since two different boxes C^s and $C^{s'}$ can have the same image by $\Psi_{p,\mathscr{X}}$.

With a slight change, one can get a one to one mapping $\Theta_{p,\mathscr{X}}$ from $E_{\mathscr{X}}$ onto $E_{\mathscr{X}}$. We proceed as follows. M_r is a compact metrizable space for the narrow topology, let us denote $d(\cdot, \cdot)$ a distance function defining the uniform structure. Now, we define a one to one mapping $q^s \to q^{\xi(s)}$, where ξ is a permutation of the indices of the boxes, in the following way:

$$\max_s d(q^{\xi(s)}, \Phi_1 q^s) = \min_k \max_s d(q^{k(s)}, \Phi_1 q^s),$$

where k runs over the permutations of the indices of the boxes satisfying $\pi_{q^{k(s)}} = \pi_{q^s}$. This mapping is straightforwardly extended (as above for Ψ_p) from C^s onto $C^{\xi(s)}$ to get finally a one to one mapping $\Theta_{p,\mathscr{X}}$ from $E_{\mathscr{X}}$ onto $E_{\mathscr{X}}$. Obviously, π_q and the measure $\pi_{\mathscr{X}}$ are conserved by $\Theta_{p,\mathscr{X}}$.

Of course, it remains to prove that $\Theta_{p,\mathscr{X}}$ is also a good approximation of Φ_1 (i.e. an analog of Proposition 4.2 holds for $\Theta_{p,\mathscr{X}}$). Despite the fact that this property follows common intuition, it is not easy to prove rigorously. We can only conjecture that it is true at present.

Finally, if we suppose that such a result is true, at long time (say $t = T$, T integer), for $d(\mathscr{X})$ small enough and p large enough, the flow Φ_T will be well approximated by $(\Theta_{p,\mathscr{X}})^T$, so that the classical appeal to ergodic theory holds.

6) Resolution of the Variational Problem. As we have previously seen, the equilibrium set $\mathscr{E}^*$ corresponding to a given initial datum q is the set of the solutions ν^* of the variational problem,

$$\text{(V.P.)} \quad I_\pi(\nu^*) = \inf\{I_\pi(\nu) \mid \nu \in \mathscr{E}\},$$

where $\mathscr{E}$ is the closed set of the Young measures ν satisfying

(*)
$$\int_{\Omega} \nu_{\mathbf{x}} \, d\mathbf{x} = \pi_q \, ,$$

(**)   other constraints (energy. . .) .

Using Lagrange multipliers, we can get easily the equation satisfied by the critical points of (V.P.) (Gibbs states' equation). We refer to [33] for Euler equations and to [24] for the quasi-geostrophic model.

In this section we shall make two comments. The first comment is relative to the choice of the basic probability measure π_0. Indeed, it is natural to examine how the equilibrium set defined by (V.P.) depends on π_0. As previously, denote $\pi_0 = \dfrac{1}{|\Omega|} \pi_q$, and suppose that we choose another probability distribution π_1 on $[-r, r]$ to define the rate function $I_{\pi'}$, $\pi' = d\mathbf{x} \otimes \pi_1$. The condition (*) implies that, for almost all $\mathbf{x}, \nu_{\mathbf{x}}$ is absolutely continuous with respect to π_0. On the other hand, one easily gets for $\nu \in \mathscr{E}$:

$$I_{\pi'}(\nu) = I_{\pi}(\nu) + |\Omega| I_{\pi_1}(\pi_0) \, .$$

Thus we see that, provided that $I_{\pi_1}(\pi_0) < +\infty$, minimizing I_{π} or $I_{\pi'}$ on $\mathscr{E}$ gives the same equilibrium set $\mathscr{E}^*$. In the degenerate case $I_{\pi_1}(\pi_0) = +\infty$, $I_{\pi'}$ is identical to $+\infty$, and the corresponding equilibrium set is equal to $\mathscr{E}$.

To summarize: in the non-degenerate case, the equilibrium set does not depend on the choice of π_1. It is then very natural to choose $\pi_1 = \pi_0$.

We come now to the resolution of (V.P.) in the particular case of Example 1. We have:

$$\mathscr{E} = \left\{ \nu \in M \; \middle| \; \int_{\Omega} \nu_{\mathbf{x}} \, d\mathbf{x} = \pi_q \right\} \, ,$$

and the unique solution of (V.P.) is $\nu^* = d\mathbf{x} \otimes \pi_0$. Let us suppose that the Lagrangian flow φ_t associated to $\mathbf{u}$ is mixing in the usual sense of ergodic theory, that is: for any measurable subsets A, B of Ω, we have

$$\lim_{t \to \infty} \mathrm{vol}(\varphi_t^{-1}(A) \cap B) = \mathrm{vol}(A) \cdot \mathrm{vol}(B) \, .$$

In that case, we can easily prove that when $t \to \infty$,

$$\delta_{\Phi_t q} \to \nu^* \, , \quad \text{for the narrow topology} \, .$$

This simple result linking the long time dynamics with the equilibrium state leads to the following important comment.

For a given q, since $\Phi_t q = q(\varphi_t^{-1})$, many other topological invariants are conserved by the flow: for example, if q is a patch with n connected components, the same will be true of $\Phi_t q$ for all t. So we may a priori think that an equilibrium state will depend not only on the constants of the motion that we have considered but on q itself. Example 1 shows that in fact these invariants play no role. We see that, when t goes to infinity, the Dirac Young measure $\delta_{\Phi_t q}$ approximates better and better in the narrow topology the limit ν^*, while the supplementary invariants that $\Phi_t q$ conserves bring no constraints at the limit on the state ν^*. This is due to the fact that these invariants are not uniformly continuous for the uniform structure induced on L_r^∞ by the narrow

uniform structure of M_r. Indeed, any uniformly continuous invariant has a unique continuous extension F on M_r. Then, at the limit, ν^* must satisfy the constraint $F(\nu^*) = \lim_{t \to \infty} F(\Phi_t q) = F(q)$.

Let us notice also that, for Example 1, the mapping $\Phi_{\mathscr{X}} : E_{\mathscr{X}} \to E_{\mathscr{X}}$ defined by

$$\Phi_{\mathscr{X}}\left(\sum q_i \mathbf{1}_{\Omega^i}\right) = \sum q_i \mathbf{1}_{\Omega^{\sigma(i)}},$$

where the permutation σ is chosen such that $\mathrm{vol}(\varphi_1(\Omega^i) \cap \Omega^{\sigma(i)}) > 0$, is obviously one to one and conserves the distribution π_q and the measure $\pi_{\mathscr{X}}$.

V. Comments

1. The idea to consider such equilibrium states for the Boltzmann-Poisson equation was first introduced by Lynden-Bell [22]. As noticed in [41], since the region occupied by the stars is the whole space $\mathbb{R}^3$, equilibrium states do not exist in that case. Moreover, even if the particles are constrained to move in a bounded domain of the space, their speed is distributed over all of $\mathbb{R}^3$; so that the domain is not naturally bounded in the case of kinetic equations. This of course needs some supplementary technical work to define rigorously the equilibrium states.

2. In the case of Euler equations and more generally for the quasi-geostrophic model the same equilibrium states can be obtained by using the classical point-vortex approximation, see [23, 24].

3. In a recent paper [26], Miller et al. exhibit a N-dimensional approximation of 2D Euler equations on the torus, which preserves the main features of the Hamiltonian structure; that is, the degeneracy of the Poisson brackets yielding $O(\sqrt{N})$ Casimir invariants. This approximation was previously published by Zeitlin [46], see also [14, 45]. While it is likely that this equation provides a good approximation of Euler flow, the thermodynamic limit in the space of Young measures of the associated canonical Gibbs measures is far from obvious. The following simpler problem is not even solved.

Let us consider $e_1, \ldots, e_N$, the usual real orthonormal Fourier basis, $\omega^N = \sum \omega_i e_i$, and let ω^N evolve according to the usual spectral approximation scheme. Then it is well known that the measure $d\omega_1 \ldots d\omega_N$ is conserved, so are the energy and the enstrophy $\sum \omega_i^2$. Let us consider the Gibbs measures

$$\mu_N = \frac{1}{Z} \exp\left(-N\alpha \sum \omega_i^2\right) d\omega_1 \ldots d\omega_N, \quad \text{for } \alpha > 0 \text{ fixed}.$$

The following question arises naturally:

Do the images, by the mapping $\omega \to \delta_\omega$, of the measures μ_N have the large deviation property with constants N and rate function $I_\pi(\nu)$ $\left(\pi = d\mathbf{x} \otimes \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha\omega^2) d\omega\right)$?

To conclude this comment, let us notice that the large deviation results of Sects. II and III provide an easy and rigorous proof of the thermodynamic limit result of Miller et al., see [33] Comment 7.

4. We left no place in this paper for the discussion on the physical relevance of these equilibrium states. Such a discussion is necessarily based on an analysis of the macroscopic spatial scale at which the flow is observed, of the viscous dissipation scale

(if some viscous dissipation process occurs), and of some characteristic relaxation time of the system. This greatly varies from one system to another (there is no viscous dissipation for collisionless kinetic equations) and must be carefully done specifically in each case. This is clearly out of the scope of our paper which is an attempt to give some clarification on the mathematical basis of the theory. In our opinion we need to distinguish clearly between the mathematical properties of ideal models and the variable conditions under which such equilibrium states might be physically relevant. Nevertheless, some elements for such an analysis can be found in [35, 40, 26, 13] for 2D turbulence, in [39] for Jupiter's Great Red Spot, and in [22, 41] for stellar systems.

Appendix

Proof of Lemma 4.1. We proceed by induction. More precisely, we prove that for all k, $1 \leq k \leq n$, and any family $\Omega^1, \ldots, \Omega^k$ of k distinct elements of $\mathscr{X}$, we can find $O^1, \ldots, O^k$ distinct elements of $\mathscr{O}$ such that $\mathrm{vol}(\Omega^i \cap O^i) > 0$.

This is obviously true for $k = 1$. Now, let us suppose that the property is true up to $k < n$, and consider a family $\Omega^1, \ldots, \Omega^{k+1}$.

By the induction hypothesis, there is $O^1, \ldots, O^k$, in $\mathscr{O}$ such that $\mathrm{vol}(\Omega^i \cap O^i) > 0$ for $i = 1, \ldots, k$. Let us consider Ω^{k+1}. If Ω^{k+1} intersects [in the sense $\mathrm{vol}(\Omega^{k+1} \cap O^*) > 0$] an element O^* of $\mathscr{O}$ different from $O^1, \ldots, O^k$, the result is proved.

Otherwise, we have $\Omega^{k+1} \subset O^1 \cup \ldots \cup O^k$ [in the sense $\mathrm{vol}(\Omega^{k+1} \cap (\ldots)) = \mathrm{vol}(\Omega^{k+1})$], then Ω^{k+1} intersects at least two different sets $O^1, \ldots, O^k$, let us denote them $O^1, \ldots, O^p$ (renumber), $2 \leq p \leq k$. Now we have $\Omega^{k+1} \subset O^1 \cup \ldots \cup O^p$. If one of the $\Omega^1, \ldots, \Omega^p$, say Ω^i, intersects $\Omega \backslash (O^1 \cup \ldots \cup O^k)$ the result clearly follows: since Ω^i intersects a set O^* of $\mathscr{O}$ different from $O^1, \ldots, O^k$, we only have to associate O^* to Ω^i and then O^i to Ω^{k+1}.

So, we may suppose that $\Omega^{k+1} \cup \Omega^1 \cup \ldots \cup \Omega^p \subset O^1 \cup \ldots \cup O^k$. Since the measure of the first set is equal to $(p + 1) \mathrm{vol}(\Omega^1)$, it cannot be included in $O^1 \cup \ldots \cup O^p$, so that there is a set O^{p+1} (renumber) which intersects a set Ω^i for $i \leq p$. If Ω^{p+1} intersects $\Omega \backslash (O^1 \cup \ldots \cup O^k)$, the result follows; otherwise, we iterate the construction and it yields a sequence O^i, $i = 1 \ldots m$, $p < m \leq k$, such that Ω^m intersects $\Omega \backslash (O^1 \cup \ldots \cup O^k)$ and for each O^i, $p < i \leq m$, there is a Ω^j with $j < i$ which intersects O^i.

Now, to conclude, we associate to Ω^m a set O^* of $\mathscr{O}$ different from $O^1, \ldots, O^k$. Then we associate O^m to Ω^i for $i < m$, O^i to Ω^j for $j < i$, and so on till we get an index $j \leq p$. We associate this O^j to Ω^{k+1} and the proof is complete.

References

1. Arnold, V.I.: Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier **16**, 319–361 (1966)
2. Arsenev, A.: Global existence of a weak solution of Vlasov's system of equations. U.S.S.R. Comp. Math. and Math. Phys. **15**, 131–143 (1975)

3. Baldi, P.: Large deviations and stochastic homogenization. Ann. Mat. Pura Appl. **4**, **151**, 161–177 (1988)
4. Benfatto, G., Pico, P., Pulvirenti, M.: J. Stat. Phys. **46**, 729 (1987)
5. Berliocchi, H., Lasry, J.M.: Intégrandes normales et mesures paramétrées en calcul des variations. Bull. Soc. Math. France **101**, 129–184 (1973)
6. Boldrighini, C., Frigio, S.: Equilibrium states for a plane incompressible perfect fluid. Commun. Math. Phys. **72**, 55–76 (1980)
7. Caglioti, E., Lions, P.L., Marchioro, C., Pulvirenti, M.: A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. Commun. Math. Phys. **143**, 501–525 (1992)
8. Campbell, L.J., O'Neil, K.: Statistics of two-dimensional point vortices and high energy vortex states. J. Stat. Phys. **65**, n° 3/4, 495–529 (1991)
9. Chorin, A.J.: Lectures on turbulence theory. Publish or Perish, 1975
10. Denoix, M.A.: Thesis, Institut de mécanique de Grenoble, 1992
11. Di Perna, R., Lions, P.L.: Global weak solutions of Vlasov-Maxwell system. Commun. Pure Appl. Math. **6**, 729–757 (1989)
12. Ellis, R.S.: Entropy, large deviations and statistical mechanics. Berlin, Heidelberg, New York: Springer 1985
13. Eyink, G.L., Spohn, H.: Negative temperature states and large-scale long-lived vortices in two-dimensional turbulence (to appear)
14. Fairlie, D.B., Fletcher, P., Zachos, C.K.: Infinite-dimensional algebras and a trigonometric basis for the classical algebras. J. Math. Phys. **31** (5), 1088–1094 (1990)
15. Friedlander, L.: An invariant measure for the equation $u_{tt} - u_{xx} + u^3 = 0$. Commun. Math. Phys. **98**, 1–16 (1985)
16. Fröhlich, J., Ruelle, D.: Statistical mechanics of vortices in an inviscid two-dimensional fluid. Commun. Math. Phys. **87**, 1–36 (1982)
17. Hopfinger, E.J.: Turbulence and vortices in rotating fluids. Proceedings of the XVIIth International Congress of Theoretical and Applied Mechanics. Amsterdam: North-Holland 1989
18. Ingersoll, A.P., Cuong, P.G.: J. Atm. Sci. **38**, 2067–2076 (1981)
19. Jirina, M.: On regular conditional probabilities. Czech. Math. J. **9**, n° 3, 445 (1959)
20. Kiessling, M.K.H.: Commun. Pure Appl. Math. **46**, 27–56 (1993)
21. Kraichnan, R.H., Montgomery, D.: Two-dimensional turbulence. Rep. Prog. Phys. **43**, 547–617 (1980)
22. Lynden-Bell, D.: Statistical mechanics of violent relaxation in stellar systems. Mon. Not. R. Astr. Soc. **136**, 101–121 (1967)
23. Michel, J.: Thesis (to appear)
24. Michel, J., Robert, R., Sommeria, J.: Statistical mechanical theory of the Great Red Spot of Jupiter (to appear)
25. Miller, J.: Statistical mechanics of Euler equations in two dimensions. Phys. Rev. Lett. **65**, 17, 2137–2140 (1990)
26. Miller, J., Weichman, P.B., Cross, M.C.: Statistical mechanics, Euler's equation, and Jupiter's Red Spot. Phys. Rev. A **45**, 2328–2359 (1992)
27. Montgomery, D., Joyce, G.: Statistical mechanics of negative temperature states. Phys. Fluids **17**, 1139–1145 (1974)
28. Novikov, E.A.: Dynamics and statistics of a system of vortices. Sov. Phys. JETP. **41**, 937–943 (1976)
29. Olver, P.J.: Applications of Lie groups to differential equations. Berlin, Heidelberg, New York: Springer 1986
30. Onsager, L.: Statistical hydrodynamics. Nuovo Cimento suppl. **6**, 279 (1949)
31. Pointin, Y.B., Lundgren, T.S.: Statistical mechanics of two dimensional vortices in a bounded container. Phys. Fluids **10**, 1459–1470 (1976)
32. Robert, R.: Concentration et entropie pour les mesures d'Young. C.R. Acad. Sci. Paris, t. **309**, série I, 757–760 (1989)
33. Robert, R.: A maximum entropy principle for two-dimensional Euler equations. J. Stat. Phys. **65**, 3/4, 531–553 (1991)
34. Robert, R.: Concentration and entropy for Young measures. Preprint, Université de Lyon, 1989
35. Robert, R., Sommeria, J.: Statistical equilibrium states for two-dimensional flows. J. Fluid Mech. **229**, 291–310 (1991)

36. Sanov, I.N.: Mat. Sb. **42**, 11 (1975)
37. Sommeria, J., Meyers, S.D., Swinney, H.L.: Laboratory simulation of Jupiter's Great Red Spot. Nature **331**, 1 (1988)
38. Sommeria, J., Denoix, M.A.: To appear
39. Sommeria, J., Nore, C., Dumont, T., Robert, R.: Théorie statistique de la tache rouge de Jupiter. C.R. Acad. Sci. Paris, t. **312**, Série II, 999–1005 (1991)
40. Sommeria, J., Staquet, C., Robert, R.: Final equilibrium state of a two-dimensional shear layer. J. Fluid Mech. **233**, 661–689 (1991)
41. Tremaine, S., Henon, M., Lynden-Bell, D.: H-functions and mixing in violent relaxation. Mon. Not. R. astr. Soc. **219**, 285–297 (1986)
42. Varadhan, S.R.S.: Large deviations and applications. Ecole d'été de probabilités de Saint-Flour XV–XVII, 1985–1987
43. Youdovitch, V.I.: Non-stationary flow of an ideal incompressible liquid. Zh. Vych. Mat. **3**, 1032–1066 (1963)
44. Young, L.C.: Generalized surfaces in the calculus of variations. Ann. Math. t. **43**, 84–103 (1942)
45. Zachos, C.K.: Hamiltonian flows, $SU(\infty)$, $SO(\infty)$, $USp(\infty)$, and strings. In: Differential geometric methods in theoretical physics. Chau, L.L., Nahm, W. (eds.). New York: Plenum Press 1990
46. Zeitlin, V.: Finite-mode analogs of 2D ideal hydrodynamics: Coadjoint orbits and local canonical structure. Physica D **49**, 353–362 (1991)

Communicated by J.L. Lebowitz