2000

The Large Deviation Principle for Coarse-Grained Processes

Richard S Ellis, *University of Massachusetts - Amherst*
Kyle Haven
Bruce Turkington

The Large Deviation Principle for Coarse-Grained Processes

Richard S. Ellis*, Kyle Haven†, and Bruce Turkington‡
rsellis@math.umass.edu, haven@math.umass.edu, turk@math.umass.edu
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003

January 1, 2014

Abstract

The large deviation principle is proved for a class of L^2-valued processes that arise from the coarse-graining of a random field. Coarse-grained processes of this kind form the basis of the analysis of local mean-field models in statistical mechanics by exploiting the long-range nature of the interaction function defining such models. In particular, the large deviation principle is used in a companion paper [8] to derive the variational principles that characterize equilibrium macrostates in statistical models of two-dimensional and quasi-geostrophic turbulence. Such macrostates correspond to large-scale, long-lived flow structures, the description of which is the goal of the statistical equilibrium theory of turbulence. The large deviation bounds for the coarse-grained process under consideration are shown to hold with respect to the strong L^2 topology, while the associated rate function is proved to have compact level sets with respect to the weak topology. This compactness property is nevertheless sufficient to establish the existence of equilibrium macrostates for both the microcanonical and canonical ensembles.

Key words and phrases: Large deviation principle, Cramér's Theorem, coarse-graining, statistical models of turbulence

1 Introduction

In many statistical mechanical models coarse-graining is a fundamental construction that mediates between a microscopic scale on which the model is defined and a macroscopic

*This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376) and by a grant from the National Science Foundation (NSF-DMS-9700852).

†This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376).

‡This research was supported by a grant from the Department of Energy (DE-FG02-99ER25376) and by a grant from the National Science Foundation (NSF-DMS-9971204).

scale on which the model is analyzed in a thermodynamic or continuum limit. In this general context, the coarse-graining is determined by an averaging procedure on an intermediate scale. Its usefulness relies on the property that the functions defining the interactions of the model's variables can be well approximated by corresponding functions of the averaged variables. Typically, this situation is met in models that have long-range interactions and therefore have the character of local mean-field theories. For models of this kind, a complete and rigorous analysis of the thermodynamic or continuum limit can be carried out once the asymptotic behavior of an appropriate coarse-grained process is characterized. Such a characterization is provided by a large deviation principle, which expresses in a sharp form the statistical effect of the averaging procedure.

The well-known models of two-dimensional turbulence [13, 15, 16], or more generally quasi-geostrophic turbulence [4, 9], are prime examples of this general class of local mean-field theories. As we explain later, a coarse-grained process for these models is defined by taking a certain local average of the underlying microscopic vorticity field. The dynamical invariants, which in these models include the energy and circulation, are then expressed as functions of this coarse-grained process via approximations that become exact in the continuum limit. With this representation in hand, rigorous large deviation techniques allow one to deduce, in an essentially intuitive way, the asymptotic behavior of the vorticity field in the continuum limit. In fluid dynamical terms, the large deviation principle distinguishes certain mean flow structures, which may take the form of jets or vortices, as the most probable macrostates against a background of fluctuating, filamentary microscopic vorticity. In this way the large deviation analysis plays a pivotal role in realizing the main goal of such models: to explain the emergence and persistence of coherent structures within the turbulent flow.

For the sake of definiteness, let us consider two-dimensional turbulence in the unit torus $T^2 \doteq [0,1)\times[0,1)$. The underlying Hamiltonian system is governed by the Euler equations for an inviscid, incompressible fluid with periodic boundary conditions on the velocity and pressure fields. This system is most conveniently described as an evolution equation for the vorticity $\omega(x,t) \doteq \partial u_2/\partial x_1 - \partial u_1/\partial x_2$, which is the perpendicular component of the curl of the velocity field $u = (u_1(x,t), u_2(x,t)), \quad x = (x_1, x_2) \in T^2$. With respect to this dynamics the vorticity ω is advected, or rearranged, by the incompressible velocity field u that it induces instantaneously. Generically, this self-straining motion produces a fine-grained vorticity field ω that exhibits complex fluctuations on the small spatial scales. Statistical equilibrium models are introduced to capture the essential features of the flow without resolving the small-scale dynamics.

In order to construct a statistical equilibrium model, one discretizes the dynamics and replaces the fine-grained vorticity field ω by a microstate ζ defined on a suitable lattice $\mathcal{L}_n$ in T^2. Specifically, for each $n \in I\!N$ let $\mathcal{L}_n$ be a uniform lattice of $a_n \doteq 2^{2n}$ sites s in T^2. The intersite spacing in each coordinate direction is 2^{-n}. Each such lattice of a_n sites induces a dyadic partition of T^2 into a_n squares called microcells, each having area $1/a_n$. For each $s \in \mathcal{L}_n$ we denote by $M(s)$ the unique microcell containing the site s in its lower left corner. The configuration spaces for the model are $\mathcal{Y}^{a_n}$, where $\mathcal{Y}$ is a given closed subset of $I\!R$. The elements of $\mathcal{Y}^{a_n}$ are the microstates $\zeta = \{\zeta(s), s \in \mathcal{L}_n\}$, which we can identify with piecewise-constant vorticity fields ζ relative to $\mathcal{L}_n$; that is, $\zeta(x) = \zeta(s)$ for

all $x \in M(s)$, $s \in \mathcal{L}_n$.

The probabilistic structure of the discretized microstate ζ is chosen to be consistent with the postulated behavior of the fine-grained vorticity field ω. Specifically, it is determined by a probability measure P_n defined as follows. Let ρ be a probability measure on $I\!R$ with support $\mathcal{Y}$. If $\mathcal{Y}$ is unbounded, assume that

$$\int_{I\!R} e^{\alpha y}\, \rho(dy) < \infty \quad \text{for all } \alpha \in I\!R. \tag{1.1}$$

We then define P_n to be the product measure on $\mathcal{Y}^{a_n}$ with identical one-dimensional marginals ρ. With respect to P_n, the collection $\{\zeta(s), s \in \mathcal{L}_n\}$ is a finite family of independent, identically distributed (i.i.d.) random variables having common distribution ρ and common range $\mathcal{Y}$. Under a suitable ergodic hypothesis, the given measure ρ incorporates the family of invariants associated with incompressible rearrangement of vorticity. Details are discussed in [2, 8, 16].

The basic dynamical invariant of the Euler equations is the kinetic energy, which is expressible as the following functional of the vorticity:

$$H(\omega) \doteq \frac{1}{2} \int_{T^2 \times T^2} g(x - x')\, \omega(x)\, \omega(x')\, dx dx'. \tag{1.2}$$

In this formula $g(x - x')$ is the generalized Green's function for $-\triangle$ on T^2. In the lattice model on $\mathcal{L}_n$, $H(\omega)$ is replaced by a lattice Hamiltonian H_n. This is defined for $\zeta \in \mathcal{Y}^{a_n}$ by

$$H_n(\zeta) \doteq \frac{1}{2n^2} \sum_{s,s' \in \mathcal{L}} g_n(s - s')\zeta(s)\zeta(s')\,, \tag{1.3}$$

where $g_n(s - s')$ is a certain lattice approximation to the generalized Green's function $g(x - x')$. For instance, $g_n(s - s')$ may be determined from a finite-difference discretization or a spectral truncation.

The formalism of equilibrium statistical mechanics provides two joint probability distributions for microstates $\zeta \in \mathcal{Y}^{a_n}$: the microcanonical ensemble and the canonical ensemble. In probabilistic terms, the microcanonical ensemble expresses the conditioning of P_n on the energy shell $\{\zeta \in \mathcal{Y}^{a_n} : H_n(\zeta) = E\}$, where $E \in I\!R$ is a specified energy value. In order to avoid technical problems with the existence of regular conditional probability distributions, we shall condition P_n on the thickened energy shell $\{H_n(\zeta) \in [E - \varepsilon, E + \varepsilon]\}$, where $\varepsilon > 0$. Thus, the microcanonical ensemble is the measure defined for Borel subsets B of $\mathcal{Y}^{a_n}$ by

$$P_n^{E,\varepsilon}\{B\} = P_n\{B \,|\, H_n \in [E - \varepsilon, E + \varepsilon]\} = \frac{P_n\{B \cap \{H_n \in [E - \varepsilon, E + \varepsilon]\}\}}{P_n\{H_n \in [E - \varepsilon, E + \varepsilon]\}}.$$

Correspondingly, the canonical ensemble is defined for Borel subsets B of $\mathcal{Y}^{a_n}$ by

$$P_{n,\beta}\{B\} \doteq \frac{1}{Z(n,\beta)} \int_B \exp[-\beta H_n]\, dP_n.$$

Here β is a real number denoting the inverse temperature, and $Z(n,\beta)$ is the partition function $\int_{\mathcal{Y}^{a_n}} \exp[-\beta H_n]\, dP_n$, which normalizes the probability measure $P_{n,\beta}$.

Models of this kind were originally proposed independently by Miller et. al. [12, 13] and Robert et. al. [14, 15] in the context of two-dimensional Euler flow. Subsequently, they were extended to geophysical fluid dynamics, such as barotropic quasi-geostrophic flow [4, 8]. A model of spin systems on a circle that exhibits an interesting phase transition has a similar, but simpler structure [6].

In the setting of models of two-dimensional and quasi-geostrophic turbulence, there are two basic goals of the equilibrium statistical theory: first, to predict the formation of stable, coherent flow structures from either the microcanonical ensemble or the canonical ensemble; second, to deduce whether the two ensembles yield equivalent results. In order to achieve these two goals, which depend on deriving properties of the two ensembles in the continuum limit, the crucial innovation is to introduce a two-parameter stochastic process that is defined by a coarse-graining, or local averaging, of the microscopic vorticity field over an intermediate scale. We now present this key construction.

Given $n \in I\!N$ and a positive integer $r < n$, we consider a dyadic partition of the lattice $\mathcal{L}_n$ into $\gamma_r \doteq 2^{2r}$ blocks, each block containing a_n/γ_r lattice sites. In correspondence with this partition we have a dyadic partition $\{D_{r,k}, k = 1, \ldots, \gamma_r\}$ of T^2 into macrocells. Each macrocell is the union of a_n/γ_r microcells $M(s)$. This partition of $\mathcal{L}_n$ into γ_r blocks represents a coarse-graining of the lattice $\mathcal{L}_n$. With respect to this partition, we define the following coarse-grained process, obtained by a local averaging over the sites of the macrocells $D_{r,k}$:

$$W_{n,r}(x) = W_{n,r}(\zeta, x) \doteq \sum_{k=1}^{\gamma_r} 1_{D_{r,k}}(x)\, S_{n,r,k}(\zeta), \tag{1.4}$$

where

$$S_{n,r,k}(\zeta) \doteq \frac{1}{a_n/\gamma_r} \sum_{s \in D_{r,k}} \zeta(s). \tag{1.5}$$

The doubly indexed process $W_{n,r}$ takes values in the space $L^2(T^2)$.

The process $W_{n,r}$ has the following two properties, which will allow us to evaluate its continuum limit with respect to either the microcanonical or canonical ensemble.

1. In the double limit $n \to \infty$, $r \to \infty$, with respect to the product measures P_n, $W_{n,r}$ satisfies the large deviation principle (LDP) on $L^2(T^2)$ with scaling constants a_n and an explicitly determined rate function I.

2. In the double limit $n \to \infty$, $r \to \infty$, the Hamiltonian $H_n(\zeta)$ is asymptotic to $H(W_{n,r}(\zeta))$ uniformly over microstates, where the functional H mapping $L^2(T^2)$ into $I\!R$ is defined in (1.2); in symbols,

$$\lim_{r\to\infty} \lim_{n\to\infty} \sup_{\zeta \in \mathcal{Y}^{a_n}} |H_n(\zeta) - H(W_{n,r}(\zeta))| = 0. \tag{1.6}$$

The proof of the two-parameter LDP in item 1 is the main task of this paper. We will give a heuristic proof later in this section. In Section 2 we will formulate the LDP for a natural generalization of $W_{n,r}$ and will prove this in Section 3.

The verification of (1.6) in item 2 can be carried out as in [1, §4.2], where a similar approximation is verified. Essentially, this approximation depends on the fact that the

vortex interactions governed by H_n are long-range, being determined essentially by the Green's function $g(x-x')$. For this reason, H_n is not sensitive to the small-scale structure of the vorticity field and depends only on the local mean vorticity in the continuum limit. In other words, H_n is well approximated by a function of the coarse-grained process $W_{n,r}$. This kind of behavior is typical of local mean-field theories. The turbulence models under consideration here have the property that their local mean-field approximations are asymptotically exact [13].

In the next part of this section, we motivate the two-parameter LDP in item 1. Later in Section 5 we indicate how items 1 and 2 together allow one to evaluate the continuum limit of $W_{n,r}$ with respect to the microcanonical ensemble and the canonical ensemble. These limits are expressed in terms of variational formulas whose solutions correspond to coherent structures for the two ensembles. There we also discuss the question of equivalence and nonequivalence of ensembles.

In order to motivate the LDP in item 1, we note that with respect to P_n the normalized sums $S_{n,r,k}$ are sample means of the a_n/γ_r i.i.d. random variables $\zeta(s), s \in D_{r,k}$. Cramér's Theorem therefore implies that, for each $k = 1, \ldots \gamma_r$, $\{S_{n,r,k}, n \in I\!N\}$ satisfies the LDP with respect to P_n with scaling constants a_n/γ_r and with the rate function

$$i(z) \doteq \sup_{\alpha \in I\!R} \{\alpha z - c(\alpha)\} \quad \text{for } z \in I\!R.$$

This function is the lower semicontinuous, convex function conjugate to the cumulant generating function

$$c(\alpha) \doteq \log \int_{I\!R} e^{\alpha y}\, \rho(dy), \qquad \alpha \in I\!R\,.$$

For $y_k \in I\!R$ we summarize the LDP for $S_{n,r,k}$ by the heuristic notation

$$\lim_{n\to\infty} \frac{1}{a_n/\gamma_r} \log P_n\{S_{n,r,k} \sim y_k\} \approx -i(y_k).$$

This basic LDP makes use of the fact that $c(\alpha)$ is finite for all $\alpha \in I\!R$, which follows from the assumed property (1.1) of ρ. The Gärtner-Ellis Theorem allows one to extend Cramér's Theorem to measures ρ for which $c(\alpha)$ is finite for α in a subset A of $I\!R$ that contains 0 in its interior and for which $\lim_{n\to\infty} |c'(\alpha_n)| = \infty$ whenever α_n is a sequence in $\mathrm{int}(A)$ converging to a boundary point of $\mathrm{int}(A)$ [3]. This extension is useful because it applies to measures ρ having exponential tails that arise in certain turbulence models [8].

For each ζ, $W_{n,r}(\zeta)$ is piecewise constant on the macrocells $D_{r,k}$. To give a heuristic derivation of the LDP for $W_{n,r}$, we approximate a general $f \in L^2(T^2)$ by a piecewise constant function of the form

$$\varphi(x) = \sum_{k=1}^{\gamma_r} \varphi_k 1_{D_{r,k}}(x).$$

Then using Cramér's Theorem for each $k = 1, \ldots, \gamma_r$ and the independence of $S_{n,r,1}, \ldots, S_{n,r,\gamma_r}$, we have for all sufficiently large r

$$\lim_{n\to\infty} \frac{1}{a_n} \log P_n\{W_{n,r} \sim f\} \tag{1.7}$$

$$
\begin{aligned}
&\approx \lim_{n\to\infty} \frac{1}{a_n} \log P_n\{W_{n,r} \sim \varphi\} \\
&= \frac{1}{\gamma_r} \lim_{n\to\infty} \frac{1}{a_n/\gamma_r} \log P_n\{S_{n,r,1} \sim \varphi_1, \dots S_{n,r,\gamma_r} \sim \varphi_{\gamma_r}\} \\
&= \frac{1}{\gamma_r} \sum_{k=1}^{\gamma_r} \lim_{n\to\infty} \frac{1}{a_n/\gamma_r} \log P_n\{S_{n,r,k} \sim \varphi_k\} \\
&\approx -\frac{1}{\gamma_r} \sum_{k=1}^{\gamma_r} i(\varphi_k) \;=\; -\int_{T^2} i(\varphi(x))\, dx \;\approx\; -\int_{T^2} i(f(x))\, dx.
\end{aligned}
$$

This calculation makes it reasonable to expect that $W_{n,r}$ satisfies a two-parameter LDP on $L^2(T^2)$ with the rate function

$$
I(f) \doteq \int_{T^2} i(f(x))\, dx\,. \tag{1.8}
$$

Here and throughout the paper the term "rate function on a space $\mathcal{X}$" denotes a lower semicontinuous function mapping $\mathcal{X}$ into $[0, \infty]$. A rate function need not have compact level sets.

In Section 2 we consider a natural generalization of the doubly indexed process $W_{n,r}$ taking values in an L^2 space and formulate an LDP in the strong topology on that space [Thm. 2.3]. The LDP is proved in Section 3. In Section 4 we show that the rate function in this LDP has compact level sets with respect to the weak topology on the L^2 space, though not in general with respect to the strong topology. The results in Sections 2, 3, and 4 are derived from basic principles using relatively straightforward proofs. These techniques are related to those introduced in [2], which establishes an analogous two-parameter LDP for a class of spatialized random measures. Those random measures also arise in the analysis of the continuum limit for turbulence models [1, 11]. The results in the present paper, however, are more elementary, both conceptually and technically. At the end of Section 4 we comment on the paper [2] in the light of the present paper. We also point out, in the next to last paragraph of that section, an oversight in a proof in [2].

In Section 5 we summarize typical applications of our main LDP stated in Theorem 2.3. In particular, we state the variational principles for the microcanonical and canonical ensembles defined in this introduction, and we demonstrate how to obtain these principles from a large deviation analysis of the corresponding ensembles. The solutions of the variational principles are called equilibrium macrostates. In Section 5 we also point out how the existence of equilibrium macrostates makes use of the weak compactness of the level sets of the rate function. With these results in hand, we then comment on the equivalence and nonequivalence of ensembles for the microcanoncial and canonical ensembles for which definitive results are given in [7]. A complete discussion of the physical applications of these results is contained in [8], where families of stable, steady mean flows for a general class of geophysical fluid dynamical models are characterized and computed.

2 Statement of the LDP

In this section we formulate the LDP for a natural generalization of the random functions $W_{n,r}$ defined in (1.4)–(1.5). Let $(\Omega, \mathcal{F}, P)$ be a probability space and d a positive integer. For each $r \in I\!N$, let $\{S_{n,r}, n \in I\!N\}$ be a sequence of random vectors mapping Ω into $I\!R^d$ and satisfying the LDP as $n \to \infty$ with positive scaling constants $c_{n,r}$ and convex rate function i independent of r. In other words, for each r $c_{n,r} \to \infty$ as $n \to \infty$; for any closed subset F of $I\!R^d$

$$\limsup_{n\to\infty} \frac{1}{c_{n,r}} \log P\{S_{n,r} \in F\} \le -i(F);$$

and for any open subset G of $I\!R^d$

$$\liminf_{n\to\infty} \frac{1}{c_{n,r}} \log P\{S_{n,r} \in G\} \ge -i(G).$$

Here $i(B)$ denotes the infimum of i over the set B. We assume throughout that i is lower semicontinuous and convex on $I\!R^d$. We do not assume that i has compact level sets, even though this extra property is satisfied in many applications. The setup in Section 1 corresponds to choosing $S_{n,r}$ as in (1.5) and $c_{n,r} = a_n/\gamma_r = 2^{2(n-r)}$ whenever $1 \le r < n$; otherwise, $S_{n,r}$ and $c_{n,r}$ equal 0.

In order to give a general formulation of the LDP, we consider a Polish space Λ with metric b (a complete separable metric space) and let θ be a probability measure on Λ. $L^2(\Lambda, \theta)$ denotes the set of functions f mapping Λ into $I\!R^d$ and satisfying

$$\|f\|_2^2 \doteq \int_\Lambda |f|^2 d\theta < \infty,$$

where $|\cdot|$ denotes the Euclidean norm on $I\!R^d$. Let γ_r be a sequence of positive integers tending ∞. For each positive integer n and r we introduce $S_{n,r,1}, \ldots, S_{n,r,\gamma_r}$, which are i.i.d. copies of $S_{n,r}$ mapping Ω into $I\!R^d$. For each $r \in I\!N$ we assume that Λ is partitioned into γ_r subsets $D_{r,1}, \ldots, D_{r,\gamma_r}$ and that these sets satisfy the following condition.

Condition 2.1. *For each $r \in I\!N$*

(i) $\theta\{D_{r,k}\} = 1/\gamma_r$, $k = 1, \ldots, \gamma_r$,

(ii) $\lim_{r\to\infty} \max_{k\in\{1,\ldots,\gamma_r\}} \{\mathrm{diam}(D_{r,k})\} = 0$, *where* $\mathrm{diam}(D_{r,k}) \doteq \sup_{x,y\in D_{r,k}} b(x,y)$.

In Section 1 we worked with Lebesgue measure on the unit torus $\Lambda = T^2$, which was partitioned into 2^{2r} macrocells $D_{r,k}$.

The process whose asymptotics we wish to analyze is the doubly indexed sequence of random functions defined for $\zeta \in \Omega$ and $x \in \Lambda$ by

$$W_{n,r}(x) = W_{n,r}(\zeta, x) \doteq \sum_{k=1}^{\gamma_r} 1_{D_{r,k}}(x)\, S_{n,r,k}(\zeta). \tag{2.1}$$

Clearly, $W_{n,r}$ maps Ω into $L^2(\Lambda,\theta)$. In Theorem 2.3 we formulate the large deviation bounds for $W_{n,r}$ with respect to the strong topology on that space. The definition of $W_{n,r}$ in (2.1) is more general than in (1.4) because here we do not assume that $S_{n,r,k}$ has the form (1.5).

The rate function that appears in the LDP for $W_{n,r}$ is defined next.

Definition 2.2. *Let i denote the rate function in the LDP for $\{S_{n,r}, n \in \mathbb{N}\}$ on $\mathbb{R}^d$. Given $f \in L^2(\Lambda,\theta)$ define*

$$I(f) \doteq \int_\Lambda i \circ f \, d\theta.$$

Since i is nonnegative and convex, it follows that I is well-defined, nonnegative and convex. At the end of Section 3 we prove that I is lower semicontinuous with respect to the strong topology, which is the topology generated by the open balls $B(f,\varepsilon) \doteq \{g \in L^2(\Lambda,\theta) : \|f-g\|_2 < \varepsilon\}$ for $f \in L^2(\Lambda,\theta)$ and $\varepsilon > 0$. In general, however, I does not have compact level sets with respect to the strong topology. This is easily seen by returning to the setup of Section 1, in which $S_{n,r,k}$ is a normalized sum of i.i.d. random variables each distributed by ρ. If we choose ρ to be a Gaussian measure on $\mathbb{R}$ having mean 0 and variance 1, then $i(z) = \frac{1}{2}z^2$ and $I(f) = \frac{1}{2}\|f\|_2^2$. In this case, level sets of I coincide with closed balls in $L^2(\Lambda,\theta)$ centered at the origin, and these sets are not compact with respect to the strong topology. In Section 4 we prove, in a setting midway between those of Sections 1 and 2, that I has compact level sets with respect to the weak topology on $L^2(\Lambda,\theta)$ under the assumption that ρ decays at infinity at least as fast as a Gaussian.

We now state the two-parameter LDP for $W_{n,r}$, which we prove in Section 3.

Theorem 2.3. *We assume Condition* 2.1 *and consider $L^2(\Lambda,\theta)$ with the strong topology. Then the function I given in Definition* 2.2 *maps $L^2(\Lambda,\theta)$ into $[0,\infty]$ and is lower semicontinuous, but in general I does not have compact level sets. In addition, the sequence $W_{n,r}$ satisfies the two-parameter LDP on $L^2(\Lambda,\theta)$ with rate function I in the following sense. For any strongly closed subset F of $L^2(\Lambda,\theta)$*

$$\limsup_{r\to\infty} \limsup_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{W_{n,r} \in F\} \le -I(F),$$

and for any strongly open subset G of $L^2(\Lambda,\theta)$

$$\liminf_{r\to\infty} \liminf_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{W_{n,r} \in G\} \ge -I(G).$$

In our companion paper [8] we need the special case of Theorem 2.3 discussed in Section 1, in which the coarse-grained process $W_{n,r}$ is defined by (1.4)–(1.5).

3 Proof of Theorem 2.3

In this section we prove the upper and lower large deviation bounds for $W_{n,r}$ in separate, but elementary steps. The following lemma is used in both steps.

Lemma 3.1. *For each r the sequence $\{(S_{n,r,1},\ldots,S_{n,r,\gamma_r}), n\in I\!N\}$ satisfies the LDP on $(I\!R^d)^{\gamma_r}$ with scaling constants $c_{n,r}$ and the rate function*

$$(\nu_1,\ldots,\nu_{\gamma_r})\mapsto\sum_{k=1}^{\gamma_r} i(\nu_k).$$

Proof. The LDP is an immediate consequence of Lemmas 2.5–2.8 in [10], since $S_{n,r,1},\ldots,S_{n,r,\gamma_r}$ are i.i.d. copies of $S_{n,r}$ and each sequence $\{S_{n,r}, n\in I\!N\}$ satisfies the LDP on $I\!R^d$ with scaling constants $c_{n,r}$ and rate function i. ■

Proof of the Large Deviation Upper Bound

Let F be a strongly closed subset of $L^2(\Lambda,\theta)$. For $r\in I\!N$ we define the closed set

$$F_r \doteq \left\{(\nu_1,\ldots,\nu_{\gamma_r})\in (I\!R^d)^{\gamma_r} : \sum_{k=1}^{\gamma_r}\nu_k 1_{D_{r,k}}(x)\in F\right\}.$$

We also define L^2_r to be the set of $f\in L^2(\Lambda,\theta)$ of the form $f(x)=\sum_{k=1}^{\gamma_r}\nu_k 1_{D_{r,k}}(x)$ for some $\nu_1,\ldots,\nu_{\gamma_r}\in I\!R^d$. By Lemma 3.1, since $\theta\{D_{r,k}\}=1/\gamma_r$,

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{c_{n,r}}\log P\{W_{n,r}\in F\}\\
&\quad=\limsup_{n\to\infty}\frac{1}{c_{n,r}}\log P\{(S_{n,r,1},\ldots,S_{n,r,\gamma_r})\in F_r\}\\
&\quad\le -\gamma_r\inf\left\{\frac{1}{\gamma_r}\sum_{k=1}^{\gamma_r} i(\nu_k) : (\nu_1,\ldots,\nu_{\gamma_r})\in F_r\right\}\\
&\quad=-\gamma_r\inf\left\{\int_\Lambda i\circ f\,d\theta : f\in F\cap L^2_r\right\}\\
&\quad\le -\gamma_r\inf\left\{\int_\Lambda i\circ f\,d\theta : f\in F\right\}\\
&\quad=-\gamma_r I(F).
\end{aligned}$$

Dividing by γ_r and sending $r\to\infty$ gives the desired large deviation upper bound. ■

Proof of the Large Deviation Lower Bound

In order to prove this bound, we need to approximate arbitrary functions in $L^2(\Lambda,\theta)$ by functions that are piecewise-constant relative to the partition $D_{r,k}, k=1,\ldots,\gamma_r$. This is carried out in the next lemma.

Lemma 3.2. *We assume Condition* 2.1. *Let f be any function in $L^2(\Lambda,\theta)$. For $r\in I\!N$ define*

$$f^r(x)\doteq\sum_{k=1}^{\gamma_r} f^r_k 1_{D_{r,k}}(x),\quad \textit{where}\quad f^r_k\doteq\gamma_r\int_{D_{r,k}} f\,d\theta.$$

Then as $r\to\infty$, $\|f-f^r\|_2\to 0$.

Proof. For any given $\varepsilon > 0$ there exists a bounded Lipschitz function φ mapping Λ into $I\!R^d$ and satisfying $\|f - \varphi\|_2 < \varepsilon$ [5]. Since the operator mapping $f \in L^2(\Lambda, \theta) \mapsto f^r$ is an orthogonal projection,

$$\|f^r - \varphi^r\|_2 = \|(f - \varphi)^r\|_2 \leq \|f - \varphi\|_2 < \varepsilon. \tag{3.1}$$

Hence it suffices to estimate $\|\varphi - \varphi^r\|_2$, where φ is a Lipschitz function with constant $M < \infty$; that is, $|\varphi(x) - \varphi(y)| \leq M b(y, x)$ for all $x, y \in \Lambda$. Because the disjoint sets $D_{r,k}$ have measure $\theta\{D_{r,k}\} = 1/\gamma_r$ [Cond. 2.1(i)], a straightforward calculation gives

$$\begin{aligned}
\|\varphi - \varphi^r\|_2^2 &= \sum_{k=1}^{\gamma_r} \int_\Lambda 1_{D_{r,k}} |\varphi - \varphi_k^r|^2 \, d\theta \\
&= \sum_{k=1}^{\gamma_r} \int_{D_{r,k}} \left| \gamma_r \int_{D_{r,k}} [\varphi(y) - \varphi(x)] \, \theta(dx) \right|^2 \theta(dy) \\
&\leq \sum_{k=1}^{\gamma_r} \int_{D_{r,k}} \left(\gamma_r \int_{D_{r,k}} M \, b(y, x) \, \theta(dx) \right)^2 \theta(dy) \\
&\leq \left(M \cdot \max_{k=1,\ldots,\gamma_r} \{\mathrm{diam}(D_{r,k})\} \right)^2 .
\end{aligned}$$

Sending $r \to \infty$ and using Cond. 2.1(ii), we complete the proof. ■

Given a strongly open subset G of $L^2(\Lambda, \theta)$, let f be any function in G and choose $\varepsilon > 0$ so that $B(f, \varepsilon) \subset G$. Also choose $N \in I\!N$ such that for all $r \geq N$ $B(f^r, \varepsilon/2) \subset B(f, \varepsilon)$, where f^r is the function defined in Lemma 3.2. Such an N exists because of the $L^2(\Lambda, \theta)$-convergence of f^r to f proved in that lemma. Define the open set

$$G_{r,\varepsilon} \doteq \left\{ (\nu_1, \ldots, \nu_{\gamma_r}) \in (I\!R^d)^{\gamma_r} : \sum_{k=1}^{\gamma_r} \nu_k 1_{D_{r,k}}(x) \in B(f^r, \varepsilon/2) \right\}.$$

Then for all $r \geq N$ Lemma 3.1 yields

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{W_{n,r} \in G\} \\
&\quad \geq \liminf_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{W_{n,r} \in B(f^r, \varepsilon/2)\} \\
&\quad = \liminf_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{(S_{n,r,1}, \ldots, S_{n,r,\gamma_r}) \in G_{r,\varepsilon}\} \\
&\quad \geq -\frac{1}{\gamma_r} \inf \left\{ \sum_{k=1}^{\gamma_r} i(\nu_k) : (\nu_1, \ldots, \nu_{\gamma_r}) \in G_{r,\varepsilon} \right\} \\
&\quad \geq -\frac{1}{\gamma_r} \sum_{k=1}^{\gamma_r} i(f_k^r) \;=\; -\frac{1}{\gamma_r} \sum_{k=1}^{\gamma_r} i\left(\gamma_r \int_{D_{r,k}} f \, d\theta \right) \\
&\quad \geq -\sum_{k=1}^{\gamma_r} \int_{D_{r,k}} i \circ f \, d\theta \;=\; -\int_\Lambda i \circ f \, d\theta \;=\; -I(f).
\end{aligned}$$

The last inequality in this display follows from Jensen's inequality. The preceding display states that

$$\liminf_{r\to\infty} \liminf_{n\to\infty} \frac{1}{\gamma_r c_{n,r}} \log P\{W_{n,r} \in G\} \geq -I(f)$$

for arbitrary $f \in G$. Taking the supremum of $-I(f)$ over $f \in G$ yields the desired large deviation lower bound. ■

Proof that I Is Strongly Lower Semicontinuous

We end this section by proving that I is lower semicontinuous with respect to the strong topology on $L^2(\Lambda, \theta)$. Namely, we show that $\liminf_{n\to\infty} I(f_n) \geq I(f)$ for any strongly convergent sequence $f_n \to f$ in $L^2(\Lambda, \theta)$.

There exists a subsequence $\{f_{n_k}\}$ such that $\lim_{n_k\to\infty} I(f_{n_k}) = \liminf_{n\to\infty} I(f_n)$ and $f_{n_k} \to f$ θ-a.s. Fatou's lemma and the lower semicontinuity of i on $\mathbb{R}^d$ then yield

$$\begin{aligned} \liminf_{n\to\infty} I(f_n) &= \liminf_{n_k\to\infty} I(f_{n_k}) = \liminf_{n_k\to\infty} \int_\Lambda i \circ f_{n_k}\, d\theta \\ &\geq \int_\Lambda \liminf_{n_k\to\infty} i \circ f_{n_k}\, d\theta \geq \int_\Lambda i \circ f\, d\theta = I(f). \end{aligned}$$

This completes the proof of strong lower semicontinuity. ■

4 Weak versus Strong Topology

In the first half of this section we prove, in the general setting of Section 2, that the function I in Definition 2.2 is lower semicontinuous with respect to the weak topology on $L^2(\Lambda, \theta)$ [Thm. 4.1]. We also show, in a setting midway between those of Sections 1 and 2, that I has compact level sets with respect to the weak topology [Thm. 4.2].

The fact that I is weakly lower semicontinuous and has weakly compact level sets is used to establish the existence of equilibrium macrostates for both the microcanonical and canonical ensembles introduced in Section 1. We recall that in Section 1 Λ equals T^2 and θ equals Lebesgue measure. In the microcanonical case the equilibrium macrostates are characterized as solutions of the following constrained minimization problem: minimize $I(f)$ over $f \in L^2(T^2)$ subject to $H(f) \in [E - \varepsilon, E + \varepsilon]$. Direct methods in the calculus of variations then assure that a minimizer exists since the functional H is weakly continuous on $L^2(T^2)$. Similarly, in the canonical case the equilibrium macrostates are characterized as solutions of the following minimization problem: minimize $I(f) + \beta H(f)$ over $f \in L^2(T^2)$. Again, a minimizer exists by virtue of the properties of I and H with respect to the weak topology on $L^2(T^2)$. These applications are further discussed in Section 5 and in [8].

A related issue arises in the Miller-Robert theory of coherent structures in two-dimensional turbulence [13, 14]. In that theory the generalized enstrophy invariants

$$A(\omega) \doteq \int_\Lambda a(\omega)\, d\theta$$

are included together with the energy invariant H. A family of moment functions a parameterize these extra constraints; these functions may be chosen arbitrarily provided

certain regularity and growth conditions are satisfied [2, 16]. Unlike the Hamiltonian H, the functionals A are generally not continuous with respect to the weak topology on $L^2(\Lambda, \theta)$. The classical, quadratic enstrophy $A(\omega) = \int_\Lambda \frac{1}{2}\omega^2 \, d\theta$ provides a counterexample. Moreover, as the same example shows, the crucial approximation property (1.6) of H is not shared by the functionals A. For this reason, in the case of the Miller-Robert theory one must rely on a coarse-graining process at the level of empirical measures rather than at the level of sample means. Details are given in [1, 2]. After the proof of Theorem 4.2, we comment on the connection between the main results in the present paper and those in [2] and then point out an oversight in a proof in [2].

I Is Weakly Lower Semicontinuous

We work in the general setting of Section 2. Let $\langle \cdot, \cdot \rangle$ denote the inner product on $L^2(\Lambda, \theta)$. The weak topology on $L^2(\Lambda, \theta)$ is generated by the neighborhoods

$$B(f; f_1, \ldots, f_p, \varepsilon) \doteq \{g \in L^2(\Lambda, \theta) : |\langle f, f_i \rangle_2 - \langle g, f_i \rangle_2| < \varepsilon, i = 1, \ldots, p\}$$

for $f \in L^2(\Lambda, \theta), p \in I\!N, f_1, \ldots, f_p \in L^2(\Lambda, \theta)$, and $\varepsilon > 0$.

Theorem 4.1. *I is lower semicontinuous with respect to the weak topology on $L^2(\Lambda, \theta)$.*

Proof. Our strategy is to approximate I by a sequence of functionals for which the lower semicontinuity is almost immediate. For $r \in I\!N$ and $f \in L^2(\Lambda, \theta)$ define $I^r(f) \doteq I(f^r)$, where f^r is the approximating function given in Lemma 3.2. We claim that $I(f) = \sup_{r \in I\!N} I^r(f)$. On the one hand, Jensen's inequality ensures that

$$I^r(f) = \int_\Lambda i \circ f^r \, d\theta = \frac{1}{\gamma_r} \sum_{k=1}^{\gamma_r} i\left(\gamma_r \int_{D_{r,k}} f \, d\theta\right) \leq I(f).$$

On the other hand, since $f^r \to f$ strongly in $L^2(\Lambda, \theta)$ [Lem. 3.2], the lower limit

$$\liminf_{r \to \infty} I^r(f) \geq I(f)$$

is valid by the strong lower semicontinuity of I proved at the end of Section 3. Hence if we prove that each I^r is weakly lower semicontinuous, the weak lower semicontinuity of I follows.

Since the sum of weakly lower semicontinuous functions is weakly lower semicontinuous, it suffices to prove that for each k the function mapping $f \mapsto i(\gamma_r \int_{D_{r,k}} f \, d\theta)$ is weakly lower semicontinuous. This is an immediate consequence of the facts that the linear mapping $f \mapsto \gamma_r \int_{D_{r,k}} f \, d\theta$ is weakly continuous and that the extended real-valued function i is lower semicontinuous. ■

I Has Weakly Compact Level Sets

We carry out this analysis in a setting generalizes that of Section 1, but is more special than that of Section 2. Let Λ be a Polish space, θ a probability measure on Λ, and a_n and γ_r two sequences of positive integers tending to ∞. Consider a subset $\mathcal{L}_n$ of Λ consisting of

a_n points. For each r, Λ is assumed to be partitioned into γ_r sets $D_{r,1}, \ldots, D_{r,\gamma_r}$ satisfying Condition 2.1. We then define $W_{n,r}$ as in (1.4)–(1.5). We also impose the additional condition that the measure ρ that defines the common distribution of $\zeta(s), s \in \mathcal{L}_n$, decays at infinity at least as fast as a Gaussian.

Theorem 4.2. *We assume that there exists $\delta > 0$ such that*

$$\int_{I\!R^d} \exp\left(\frac{\delta}{2}|y|^2\right) \rho(dy) < \infty. \tag{4.1}$$

Then the level sets of I are compact with respect to the weak topology on $L^2(\Lambda, \theta)$.

Proof. First, we claim that there exists $D \in (0, \infty)$ such that for all $z \in I\!R^d$

$$i(z) \geq \frac{\delta}{2}|z|^2 - D.$$

The inequality $\langle \alpha, y \rangle \leq \frac{1}{2\delta}|\alpha|^2 + \frac{\delta}{2}|y|^2$ for α and y in $I\!R^d$ implies that

$$c(\alpha) \leq \frac{|\alpha|^2}{2\delta} + \log \int_{I\!R^d} \exp\left(\frac{\delta}{2}|y|^2\right) \rho(dy) = \frac{|\alpha|^2}{2\delta} + D.$$

Hence

$$i(z) \doteq \sup_{\alpha \in I\!R^d} \{\langle \alpha, z \rangle - c(\alpha)\} \geq \sup_{\alpha \in I\!R^d} \left\{\langle \alpha, z \rangle - \frac{|\alpha|^2}{2\delta}\right\} - D = \frac{\delta}{2}|z|^2 - D.$$

This establishes the claim.

Using this estimate, we see that for any $f \in L^2(\Lambda, \theta)$

$$I(f) \doteq \int_\Lambda i \circ f \, d\theta \geq \frac{\delta}{2}\|f\|_2^2 - D, \tag{4.2}$$

from which it follows that for $M < \infty$

$$\{f \in L^2(\Lambda, \theta) : I(f) \leq M\} \subset \left\{f \in L^2(\Lambda, \theta) : \|f\|_2^2 \leq \frac{2}{\delta}(M + D)\right\}.$$

Since the closed balls $\{f \in L^2(\Lambda, \theta) : \|f\|_2^2 \leq \frac{2}{\delta}(M + D)\}$ are weakly compact, the weak compactness of the level sets of I follows from the fact that the level sets are weakly closed. This is a consequence of the weak lower semicontinuity of I proved in Theorem 4.1 ■

Comments on Coarse-Grained Empirical Measures

Here we comment briefly on the paper [2] in the light of the present work. In that paper we proved the LDP for a general class of doubly indexed random measures. For the purpose of comparison with the present paper, we consider only a special case of those

random measures having a form analogous to (1.4)–(1.5). In a notation analogous to that in the present paper, define the coarse-grained empirical measures

$$\tilde{W}_{n,r}(dx \times dy) = \tilde{W}_{n,r}(\zeta, dx \times dy) \doteq \theta(dx) \otimes \sum_{k=1}^{\gamma_r} 1_{D_{r,k}}(x)\, L_{n,r,k}(\zeta, dy), \tag{4.3}$$

where

$$L_{n,r,k}(\zeta, dy) \doteq \frac{1}{a_n/\gamma_r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(dy). \tag{4.4}$$

The LDP for $\tilde{W}_{n,r}$ follows from Sanov's Theorem [3], just as the LDP for $W_{n,r}$ in the present paper follows from Cramer's Theorem. The rate function associated with $\tilde{W}_{n,r}$ is the relative entropy $R(\cdot|\theta \times \rho)$. As is well-known, the relative entropy is weakly lower semicontinuous and has compact level sets with respect to the weak topology on the space of probability measures on $\Lambda \times \mathcal{Y}$.

The doubly indexed process $W_{n,r}(\zeta, x)$ considered in the present paper equals the density, with respect to $\theta(dx)$, of the mean with respect to dy of $\tilde{W}_{n,r}(\zeta, dx \times dy)$. However, since the mapping taking a measure in two variables to the density, with respect to $\theta(dx)$, of the mean with respect to dy is not continuous, it is not efficient to try to prove the LDP for $W_{n,r}$ from that for $\tilde{W}_{n,r}$. Instead, it is much better to construct the self-contained proof of the LDP for $W_{n,r}$ given in Section 3 of the present paper, using the ideas introduced in [2] for the analysis of $\tilde{W}_{n,r}$.

An abstract setting analogous to the setup in Section 2 of this paper is introduced in Section 2 of [2]. In that generality, the purported proof in [2] that the rate function J given in Definition 2.3 has compact level sets involves the following circular reasoning (see pages 318–319). While μ in the last display on page 318 depends on r, the r appearing in the first display on page 319 depends on N, which in turn depends on μ. The conclusion is that the proof is invalid.

Although we cannot conclude in general that the quantity J in [2] has compact level sets, this need not be a serious hindrance. Indeed, in many cases one can prove directly from the form of J that it has compact level sets; a number of examples are given in Example 2.7 in [2]. The most basic of these examples is given in part (a), where the rate function equals the relative entropy; in this case the compactness of the level sets is automatic. It is this particular example that is used in the applications paper [1].

5 Discussion of Applications

In this final section we return to the setting of Section 1 and indicate how Theorem 2.3 is applied to characterize equilibrium macrostates for the turbulence models. As in Section 1, we consider both the microcanoncial ensemble and the canonical ensemble, taking into account the energy constraint. After summarizing how Theorems 4.1 and 4.2 are used to prove the existence of equilibrium macrostates, we point out another interesting property of these macrostates; namely, the stability of the steady mean flows that the macrostates determine.

The theoretical tools needed to carry out this analysis are fully developed in [7] for a general class of statistical equilibrium models of local mean-field type. In that paper we derive, from underlying LDP's, variational principles for the equilibrium macrostates in both the microcanonical and canonical ensembles. Moreover, we give complete and definitive results concerning the equivalance of these ensembles at the level of their equilibrium macrostates, emphasizing the possibility of nonequivalence.

In an applied companion paper [8] these general results are applied to a widely used geophysical model; namely, barotropic, quasi-geostrophic turbulence in a zonal channel. The results in [8] rely on the LDP stated in Theorem 2.3 in the present paper for the coarse-grained process $W_{n,r}$. For the geophysical model we find that nonequivalence of ensembles occurs over a wide range of the physical parameters. This surprising result is then shown to be related to stability conditions on the steady mean flows determined by the equilibrium macrostates. The main conclusions of the paper are that when nonequivalence prevails, equilibrium macrostates corresponding to the microcanonical ensemble are richer than those corresponding to the canonical ensemble and furthermore that the microcanonical equilibrium macrostates express the essential features of the coherent structures that form in geophysical fluid flows.

We now summarize some of these results, stressing their connections to the probabilistic questions addressed in the preceding sections of this paper.

Variational Principles for Equilibrium Macrostates

We first consider the microcanonical ensemble $P_n^{E,\varepsilon}$, where E is an admissible energy value; i.e., one for which the constraint set $\{f \in L^2(T^2) : H(f) = E\}$ is nonempty. For admissible E, we prove in Theorem 3.2 of [7] that the coarse-grained vorticity field $W_{n,r}$ satisfies the LDP on $L^2(T^2)$ with scaling constants $a_n \doteq 2^{2n}$ in the continuum limit $n \to \infty$, $r \to \infty$, $\varepsilon \to 0$. The rate function I^E for this LDP is given explicitly in terms of the rate function $I(f)$ defined in (1.8); namely,

$$I^E(f) \doteq \begin{cases} I(f) + S(E) & \text{if } H(f) = E, \\ \infty & \text{otherwise,} \end{cases} \tag{5.1}$$

where

$$S(E) \doteq -\inf_{g \in L^2(T^2)} \{I(g) : H(g) = E\}. \tag{5.2}$$

The quantity $S(E)$ is called the microcanonical entropy. The LDP with respect to the microcanonical ensemble follows readily from the LDP, given in Theorem 2.3, for $W_{n,r}$ with respect to the product measures P_n.

For $f \in L^2(T^2)$ satisfying $H(f) = E$, we summarize the LDP with respect to $P_n^{E,\varepsilon}$ by the formal asymptotic statement

$$P_n^{E,\varepsilon}\{W_{n,r} \sim f\} \approx \exp[-a_n I^E(f)] \quad \text{as } n \to \infty, r \to \infty, \varepsilon \to 0. \tag{5.3}$$

In this setting, $L^2(T^2)$ is the state space of the coarse-grained process, and its elements f are the macrostates or coarse-grained vorticity fields.

The set of microcanonical equilibrium macrostates is defined to be

$$\mathcal{E}^E \doteq \{f \in L^2(T^2) : I^E(f) = 0\}.$$

This set plays a central role in the theory. For any $f \in L^2(T^2) \setminus \mathcal{E}^E$ we have $I^E(f) > 0$. The formal statement (5.3) suggests that the macrostate f has an exponentially small probability of being observed as a coarse-grained vorticity field in the continuum limit of the microcanonical ensemble. As a consequence, (5.3) suggests, and the LDP for $W_{n,r}$ with respect to the microcanonical ensemble allows one to prove, that the macrostates $f \in \mathcal{E}^E$ are the overwhelmingly most probable coarse-grained vorticity fields compatible with the microcanonical constraint $H(f) = E$. For this reason, the macrostates in $\mathcal{E}^E$ determine long-lived, large-scale coherent structures in the turbulent vorticity field, the prediction of which is the goal of the statistical equilibrium theory.

Analogous results apply to the canonical ensemble $P_{n,\beta}$ for any $\beta \in I\!R$. In order to obtain the LDP, the inverse temperature β must be scaled with a_n; the physical reason for this is given in [1, 16]. With respect to $P_{n,a_n\beta}$, the same coarse-grained process $W_{n,r}$ satisfies the LDP with scaling constants a_n in the continuum limit $n \to \infty$, $r \to \infty$ [7, Thm. 2.4]. The rate function I_β is given by

$$I_\beta(f) \doteq I(f) + \beta H(f) - \varphi(\beta), \tag{5.4}$$

where

$$\varphi(\beta) \doteq \inf_{g \in L^2(T^2)} \{I(g) + \beta H(g)\}.$$

Again, this LDP for $W_{n,r}$ follows readily from the LDP given in Theorem 2.3.

The set of canonical equilibrium macrostates is defined by

$$\mathcal{E}_\beta \doteq \{f \in L^2(T^2) : I_\beta(f) = 0\}.$$

The relationship between the set $\mathcal{E}^E$ and the microcanonical measures $P_n^{E,\varepsilon}$ is mirrored by the relationship between $\mathcal{E}_\beta$ and the scaled canonical measures $P_{n,a_n\beta}$. Namely, with respect to the latter measures, $\mathcal{E}_\beta$ consists of the overwhelmingly most probable coarse-grained states. They correspond to long-lived, large-scale coherent structures within a turbulent vorticity field at a given β. It is difficult, however, to justify on physical grounds prescribing a "turbulent temperature" $1/\beta$, especially when $\beta < 0$. This negative temperature regime is nevertheless the one of most physical interest in real applications.

Existence of Equilibrium Macrostates

Throughout this discussion of the existence of equilibrium macrostates, we assume that ρ satisfies the decay condition given in (4.1). We first ask whether there exist microcanonical equilibrium macrostates for each admissible energy value E. In other words, for each admissible E is the set $\mathcal{E}^E$ nonempty? Because of (5.1), determining the equilibrium macrostates in $\mathcal{E}^E$ is equivalent to solving the constrained minimization problem

$$\text{minimize } I(f) \text{ over } \{f \in L^2(T^2) : H(f) = E\}.$$

Analogously, we ask whether there exist canonical equilibrium macrostates for each given inverse temperature β. In other words, is the set $\mathcal{E}_\beta$ nonempty? Because of (5.4), determining the equilibrium macrostates in $\mathcal{E}_\beta$ is equivalent to solving the unconstrained minimization problem

$$\text{minimize } I(f) + \beta H(f) \text{ over } f \in L^2(T^2).$$

These variational problems are dual in the sense that the Lagrange multiplier for the constraint $H(f) = E$ in the microcanonical problem is the prescribed parameter β in the canonical problem.

In both variational problems, the existence of solutions is assured by Theorems 4.1 and 4.2. With respect to the weak topology on $L^2(T^2)$, I is lower semicontinuous, its level sets are compact, and H is continuous. Consequently, the direct methods of the calculus of variations apply to the microcanonical and canonical problems, yielding the existence of minimizers in both cases.

The following mean-field equation is satisfied by a solution f of either the microcanonical or canonical variational principle:

$$i'(f) \;=\; -\beta \int_{T^2} g(x - x') f(x')\, dx' \,, \tag{5.5}$$

where, as in (1.2), $g(x - x')$ is the generalized Green's function. This equation shows that the most probable, coarse-grained vorticity fields f in both ensembles correspond to steady, deterministic flows [1, 16]. These first-order conditions are identical for the two ensembles, except that β is specified in the canonical ensemble while β is determined along with the solution f in the microcanonical ensemble.

Stability of Equilibrium Macrostates

Even though the mean-field equations satisfied by equilibria in $\mathcal{E}^E$ and in $\mathcal{E}_\beta$ are identical, the correspondence between these sets of minimizers is subtle. This issue is commonly called the equivalence of ensembles; it investigates the relationships between the set of solutions of the constrained minimization problem that characterizes microcanonical equilibrium macrostates and the set of solutions of the unconstrained minimization problem that characterizes canonical equilibrium macrostates. This topic is discussed in great detail in [7] for a wide range of models that includes as a special case the model of barotropic quasi-geostrophic turbulence studied in [8].

As we show in [7], the equivalence or nonequivalence of ensembles depends entirely on concavity properties of the microcanonical entropy S, which is defined in (5.2). In general, the microcanonical equilibrium macrostates are richer than the canonical equilibrium macrostates. This assertion is a consequence of the following results. Every $f \in \mathcal{E}_\beta$ lies in $\mathcal{E}^E$ for some E. If S is strictly concave and smooth, then the usual thermodynamic relation $\beta = \partial S/\partial E$ defines a one-to-one correspondence between the two families of equilibrium sets. If, on the other hand, S is not concave at a value $E = E^*$, then the ensembles are nonequivalent in the sense that $\mathcal{E}^{E^*}$ is disjoint from the sets $\mathcal{E}_\beta$ for all values of β. A precise and general formulation of this striking behavior is given in Theorem 4.4 in [7]. Moreover, in Section 2 of [7] and in Section 6 of [8], a number of examples of models of turbulence are given in which the microcanonical entropy is not concave over a substantial subset of its domain, and so the ensembles are nonequivalent.

The question of equivalence between the microcanonical and canonical ensembles is intimately related to stability conditions for the equilibrium macrostates f. Such stability criteria are derived from the second-order conditions satisfied by a minimizer f, and these conditions are different for the constrained and unconstrained variational problems.

With respect to the canonical ensemble, the functional $I + \beta H$ itself provides a Lyapunov functional at f whenever f is a nondegenerate minimizer. This construction, which is known as Arnold nonlinear stability analysis in the deterministic context, proves that a perturbation of a canonical equilibrium macrostate f that is small in the L^2 norm remains close to f in the L^2 norm for all time. Interestingly, the strong L^2 topology for this stability theorem coincides with the topology for which the LDP holds in the statistical equilibrium theory. In this setting, the LDP can be viewed as a weak form of a stability statement for microstates; under an ergodic evolution of the microstates, the coarse-grained process remains close to the equilibrium macrostate in the L^2 norm.

The familiar Arnold construction, however, is not adequate to prove the stability of microcanonical equilibrium macrostates when the ensembles are not equivalent. Nevertheless, a refined argument based on penalizing the functional $I + \beta H$ furnishes the needed Lyapunov functional for the microcanonical ensemble. Then an L^2 stability result analogous to the one mentioned for the canonical ensemble is valid. This refined stability analysis fills an important gap in the known stability criteria for two-dimensional flows and their geophysical counterparts. The reader is referred to [8] for a complete discussion.

Summary

Section 5 of this paper shows the importance, in applications, of Theorem 2.3, which proves the LDP for the doubly indexed process $W_{n,r}$ with respect to the product measures P_n. The special case of this process given in (1.4)–(1.5) is needed in our companion paper [8], which gives a rather complete analysis of the equilibrium macrostates for a basic geophysical model. For this model Theorem 2.3 allows one to prove LDP's for the coarse-grained process $W_{n,r}$ with respect to both the microcanonical ensemble and the canonical ensemble. In turn, these LDP's allow one to characterize equilibrium macrostates with respect to both ensembles, the equivalence and nonequivalence of which are determined by concavity properties of the microcanonical entropy. In addition, in [8] stability properties of the equilibrium macrostates are derived, using both the familiar Arnold construction and an extension of this construction based on penalizing the Lyapunov functional used in the Arnold construction. The fundamental innovation in this work is coarse-graining, which, via the doubly indexed process $W_{n,r}$, allows one to mediate between a microscopic scale on which the model is defined and a macroscopic scale on which the equilibrium macrostates are defined and analyzed. The LDP for $W_{n,r}$ given in Theorem 2.3 is the basic mathematical result that makes all the other analysis possible.

References

[1] C. Boucher, R. S. Ellis, and B. Turkington. "Derivation of maximum entropy principles in two-dimensional turbulence via large deviations." *J. Statist. Phys.* 98:1235-1278, 2000.

[2] C. Boucher, R. S. Ellis, and B. Turkington. "Spatializing random measures: doubly indexed processes and the large deviation principle." *Ann. Prob.* 27:297-324, 1999.

[3] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Second edition. New York: Springer-Verlag, 1998.

[4] M. DiBattista, A. Majda and B. Turkington, "Prototype geophysical vortex structures via large-scale statistical theory." *Geophys. Astrophys. Fluid Dyn.* 89:235-283, 1998.

[5] R. M. Dudley. *Real Analysis and Probability*. New York: Chapman & Hall, 1989.

[6] T. Eisele and R. S. Ellis. "Symmetry breaking and random waves for magnetic systems on a circle." *Z. f. Wahrsch. verw. Geb.* 63:297-348, 1983.

[7] R. S. Ellis, K. Haven, and B. Turkington. "Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles." To appear in *J. Stat. Phys.* 101:999-1064 (2000).

[8] R. S. Ellis, K. Haven, and B. Turkington. "Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows." Preprint submitted for publication.

[9] G. Holloway. "Eddies, waves, circulation and mixing: statistical geofluid mechanics." *Annual Rev. Fluid Mech.* 18:91-147, 1986.

[10] J. Lynch and J. Sethuraman. "Large deviations for processes with independent increments." *Ann. Probab.* 15:610-627, 1987.

[11] J. Michel and R. Robert. "Large deviations for Young measures and statistical mechanics of infinite dimensional dynamical systems with conservation law." *Commun. Math. Phys.* 159:195-215, 1994.

[12] J. Miller. "Statistical mechanics of Euler equations in two dimensions." *Phys. Rev. Lett.* 65:2137-2140, 1990.

[13] J. Miller, P. Weichman and M. C. Cross. "Statistical mechanics, Euler's equations, and Jupiter's red spot." *Phys. Rev. A* 45:2328-2359, 1992.

[14] R. Robert. "A maximum-entropy principle for two-dimensional perfect fluid dynamics." *J. Statist. Phys.* 65:531-553, 1991.

[15] R. Robert and J. Sommeria. "Statistical equilibrium states for two-dimensional flows." *J. Fluid Mech.* 229:291-310, 1991.

[16] B. Turkington. "Statistical equilibrium measures and coherent states in two-dimensional turbulence." *Commun. Pure Appl Math.* 52: 781-809, 1999.