Probability, Geometry and Integrable Systems
MSRI Publications
Volume **55**, 2007

Global optimization, the Gaussian ensemble, and universal ensemble equivalence

MARIUS COSTENIUC, RICHARD S. ELLIS, HUGO TOUCHETTE, AND BRUCE TURKINGTON

With great affection this paper is dedicated to Henry McKean on the occasion of his 75th birthday.

ABSTRACT. Given a constrained minimization problem, under what conditions does there exist a related, unconstrained problem having the same minimum points? This basic question in global optimization motivates this paper, which answers it from the viewpoint of statistical mechanics. In this context, it reduces to the fundamental question of the equivalence and nonequivalence of ensembles, which is analyzed using the theory of large deviations and the theory of convex functions.

In a 2000 paper appearing in the *Journal of Statistical Physics*, we gave necessary and sufficient conditions for ensemble equivalence and nonequivalence in terms of support and concavity properties of the microcanonical entropy. In later research we significantly extended those results by introducing a class of Gaussian ensembles, which are obtained from the canonical ensemble by adding an exponential factor involving a quadratic function of the Hamiltonian. The present paper is an overview of our work on this topic. Our most important discovery is that even when the microcanonical and canonical ensembles are not equivalent, one can often find a Gaussian ensemble that satisfies a strong form of equivalence with the microcanonical ensemble known as universal equivalence. When translated back into optimization theory, this implies that an unconstrained minimization problem involving a Lagrange multiplier and a quadratic penalty function has the same minimum points as the original constrained problem.

The results on ensemble equivalence discussed in this paper are illustrated in the context of the Curie–Weiss–Potts lattice-spin model.

Keywords: Equivalence of ensembles, Gaussian ensemble, microcanonical entropy, large deviation principle, Curie–Weiss–Potts model.

1. Introduction

Oscar Lanford, at the beginning of his groundbreaking paper [Lanford 1973], describes the underlying program of statistical mechanics:

> The objective of statistical mechanics is to explain the macroscopic properties of matter on the basis of the behavior of the atoms and molecules of which it is composed. One of the most striking facts about macroscopic matter is that in spite of being fantastically complicated on the atomic level — to specify the positions and velocities of all molecules in a glass of water would mean specifying something of the order of 10^{25} parameters — its macroscopic behavior is describable in terms of a very small number of parameters; e.g., the temperature and density for a system containing only one kind of molecule.

Lanford shows how the theory of large deviations enables this objective to be realized. In statistical mechanics one determines the macroscopic behavior of physical systems not from the deterministic laws of Newtonian mechanics, but from a probability distribution that expresses both the behavior of the system on the microscopic level and the intrinsic inability to describe precisely what is happening on that level. Using the theory of large deviations, one shows that, with probability converging to 1 exponentially fast as the number of particles tends to ∞, the macroscopic behavior is describable in terms of a very small number of parameters

The success of this program depends on the correct choice of probability distribution, also known as an ensemble. One starts with a prior measure on configuration space, which, as an expression of the lack of information concerning the behavior of the system on the atomic level, is often taken to be the uniform measure. As Boltzmann recognized, the most natural choice of ensemble is the microcanonical ensemble, obtained by conditioning the prior measure on the set of configurations for which the Hamiltonian per particle equals a constant energy u. Boltzmann also introduced a mathematically more tractable probability distribution known as the canonical ensemble, in which the conditioning that defines the microcanonical ensemble is replaced by an exponential factor involving the Hamiltonian and the inverse temperature β, a parameter dual to the energy parameter u [Gibbs 1902].

Among other reasons, the canonical ensemble was introduced in the hope that in the limit $n \to \infty$ the two ensembles are equivalent; i.e., all macroscopic properties of the model obtained via the microcanonical ensemble could be realized as macroscopic properties obtained via the canonical ensemble. While ensemble equivalence is valid for many standard and important models, ensemble equivalence does not hold in general, as numerous studies cited later in this

introduction show. There are many examples of statistical mechanical models for which nonequivalence of ensembles holds over a wide range of model parameters and for which physically interesting microcanonical equilibria are often omitted by the canonical ensemble.

The present paper is an overview of our work on this topic. One of the beautiful aspects of the theory is that it elucidates a fundamental issue in global optimization, which in fact motivated our work on the Gaussian ensemble. Given a constrained minimization problem, under what conditions does there exist a related, unconstrained minimization problem having the same minimum points?

In order to explain the connection between ensemble equivalence and global optimization and in order to outline the contributions of this paper, we introduce some notation. Let $\mathcal{X}$ be a space, I a function mapping $\mathcal{X}$ into $[0, \infty]$, and $\tilde{H}$ a function mapping $\mathcal{X}$ into $\mathbb{R}$. For $u \in \mathbb{R}$ we consider the following constrained minimization problem:

$$\text{minimize } I(x) \text{ over } x \in \mathcal{X} \text{ subject to the contraint } \tilde{H}(x) = u. \tag{1-1}$$

A partial answer to the question posed at the end of the preceding paragraph can be found by introducing the following related, unconstrained minimization problem for $\beta \in \mathbb{R}$:

$$\text{minimize } I(x) + \beta \tilde{H}(x) \text{ over } x \in \mathcal{X}. \tag{1-2}$$

The theory of Lagrange multipliers outlines suitable conditions under which the solutions of the constrained problem (1-1) lie among the critical points of $I + \beta \tilde{H}$. However, it does not give, as we will do in Theorems 3.1 and 3.3, necessary and sufficient conditions for the solutions of (1-1) to coincide with the solutions of the unconstrained minimization problem (1-2) and with the solutions of the unconstrained minimization problem appearing in (1-5).

We denote by $\mathcal{E}^u$ and $\mathcal{E}_\beta$ the respective sets of solutions of the minimization problems (1-1) and (1-2). These problems arise in a natural way in the context of equilibrium statistical mechanics [Ellis et al. 2000], where u denotes the energy and β the inverse temperature. As we will outline in Section 2, the theory of large deviations allows one to identify the solutions of these problems as the respective sets of equilibrium macrostates for the microcanonical ensemble and the canonical ensemble.

The paper [Ellis et al. 2000] analyzes equivalence of ensembles in terms of relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$. In turn, these relationships are expressed in terms of support and concavity properties of the microcanonical entropy

$$s(u) = -\inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}. \tag{1-3}$$

The main results in [Ellis et al. 2000] are summarized in Theorem 3.1. Part (a) of that theorem states that if s has a strictly supporting line at an energy

value u, then full equivalence of ensembles holds in the sense that there exists a β such that $\mathcal{E}^u = \mathcal{E}_\beta$. In particular, if s is strictly concave on $\operatorname{dom} s$, then s has a strictly supporting line at all $u \in \operatorname{dom} s$ except possibly boundary points [Theorem 3.2(a)] and thus full equivalence of ensembles holds at all such u. In this case we say that the microcanonical and canonical ensembles are universally equivalent.

The most surprising result, given in part (c), is that if s does not have a supporting line at u, then nonequivalence of ensembles holds in the strong sense that $\mathcal{E}^u \cap \mathcal{E}_\beta = \varnothing$ for all $\beta \in \mathbb{R}^\sigma$. That is, if s does not have a supporting line at u — equivalently, if s is not concave at u — then microcanonical equilibrium macrostates cannot be realized canonically. This is to be contrasted with part (d), which states that for any $x \in \mathcal{E}_\beta$ there exists u such that $x \in \mathcal{E}^u$; i.e., canonical equilibrium macrostates can always be realized microcanonically. Thus of the two ensembles, in general the microcanonical is the richer.

The paper [Costeniuc et al. 2005b] addresses the natural question suggested by part (c) of Theorem 3.1. If the microcanonical ensemble is not equivalent with the canonical ensemble on a subset of energy values u, then is it possible to replace the canonical ensemble with another ensemble that is universally equivalent with the microcanonical ensemble? We answered this question by introducing a penalty function $\gamma[\tilde{H}(x)-u]^2$ into the unconstrained minimization problem (1-2), obtaining the following:

$$\text{minimize } I(x) + \beta \tilde{H}(x) + \gamma[\tilde{H}(x) - u]^2 \text{ over } x \in \mathcal{X}. \tag{1-4}$$

Since for each $x \in \mathcal{X}$

$$\lim_{\gamma\to\infty} \gamma[\tilde{H}(x) - u]^2 = \begin{cases} 0 & \text{if } \tilde{H}(x) = u \\ \infty & \text{if } \tilde{H}(x) \neq u, \end{cases}$$

it is plausible that for all sufficiently large γ minimum points of the penalized problem (1-4) are also minimum points of the constrained problem (1-1). Since β can be adjusted, (1-4) is equivalent to the following:

$$\text{minimize } I(x) + \beta \tilde{H}(x) + \gamma[\tilde{H}(x)]^2 \text{ over } x \in \mathcal{X}. \tag{1-5}$$

The theory of large deviations allows one to identify the solution of this problem as the set of equilibrium macrostates for the so-called Gaussian ensemble. It is obtained from the canonical ensemble by adding an exponential factor involving γh_n^2, where h_n denotes the Hamiltonian energy per particle. The utility of the Gaussian ensemble rests on the simplicity with which the quadratic function γu^2 defining this ensemble enters the formulation of ensemble equivalence. Essentially all the results in [Ellis et al. 2000] concerning ensemble equivalence, including Theorem 3.1, generalize to the setting of the Gaussian ensemble by

replacing the microcanonical entropy $s(u)$ by the generalized microcanonical entropy

$$s_\gamma(u) = s(u) - \gamma u^2. \tag{1-6}$$

The generalization of Theorem 3.1 is stated in Theorem 3.3, which gives all possible relationships between the set $\mathcal{E}^u$ of equilibrium macrostates for the microcanonical ensemble and the set $\mathcal{E}_{\beta,\gamma}$ of equilibrium macrostates for the Gaussian ensemble. These relationships are expressed in terms of support and concavity properties of s_γ.

For the purpose of applications the most important consequence of Theorem 3.3 is given in part (a), which states that if s_γ has a strictly supporting line at an energy value u, then full equivalence of ensembles holds in the sense that there exists a β such that $\mathcal{E}^u = \mathcal{E}_{\beta,\gamma}$. In particular, if s_γ is strictly concave on $\operatorname{dom} s$, then s_γ has a strictly supporting line at all $u \in \operatorname{dom} s$ except possibly boundary points [Theorem 3.4(a)] and thus full equivalence of ensembles holds at all such u. In this case we say that the microcanonical and Gaussian ensembles are universally equivalent.

In the case in which s is C^2 and s'' is bounded above on the interior of $\operatorname{dom} s$, then the strict concavity of s_γ is easy to show. In fact, the strict concavity is a consequence of

$$s''_\gamma(u) = s''(u) - 2\gamma < 0 \quad \text{for all } u \in \operatorname{int}(\operatorname{dom} s),$$

and this in turn is valid for all sufficiently large γ [Theorem 4.2]. For such γ it follows, therefore, that the microcanonical and Gaussian ensembles are universally equivalent.

Defined in (2.6), the Gaussian ensemble is mathematically much more tractable than the microcanonical ensemble, which is defined in terms of conditioning. The simpler form of the Gaussian ensemble is reflected in the simpler form of the unconstrained minimization problem (1-5) defining the set $\mathcal{E}_{\beta,\gamma}$ of Gaussian equilibrium macrostates. In (1-5) the constraint appearing in the minimization problem (1-1) defining the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates is replaced by the linear and quadratic terms involving $\tilde{H}(x)$. The virtue of the Gaussian formulation should be clear. When the microcanonical and Gaussian ensembles are universally equivalent, then from a numerical point of view, it is better to use the Gaussian ensemble because in contrast to the microcanonical one, the Gaussian ensemble does not involve an equality constraint, which is difficult to implement numerically. Furthermore, within the context of the Gaussian ensemble, it is possible to use Monte Carlo techniques without any constraint on the sampling [Challa and Hetherington 1988a; Challa and Hetherington 1988b].

By giving necessary and sufficient conditions for the equivalence of the three ensembles in Theorems 3.1 and 3.3, we make contact with the duality theory of global optimization and the method of augmented Lagrangians [Bertsekas 1982, § 2.2], [Minoux 1986, § 6.4]. In the context of global optimization the primal function and the dual function play the same roles that the microcanonical entropy (resp., generalized microcanonical entropy) and the canonical free energy (resp., Gaussian free energy) play in statistical mechanics. Similarly, the replacement of the Lagrangian by the augmented Lagrangian in global optimization is paralleled by our replacement of the canonical ensemble by the Gaussian ensemble.

The Gaussian ensemble is a special case of the generalized canonical ensemble, which is obtained from the canonical ensemble by adding an exponential factor involving $g(h_n)$, where g is a continuous function that is bounded below. Our paper [Costeniuc et al. 2005b] gives all possible relationships between the sets of equilibrium macrostates for the microcanonical and generalized canonical ensembles in terms of support and concavity properties of an appropriate entropy function. Our paper [Touchette et al. 2006] shows that the generalized canonical ensemble can be used to transform metastable or unstable nonequilibrium macrostates for the standard canonical ensemble into stable equilibrium macrostates for the generalized canonical ensemble.

Equivalence and nonequivalence of ensembles is the subject of a large literature. An overview is given in the introduction of [Lewis et al. 1995]. A number of theoretical papers on this topic, including [Deuschel et al. 1991; Ellis et al. 2000; Eyink and Spohn 1993; Georgii 1993; Lewis et al. 1994; Lewis et al. 1995; Roelly and Zessin 1993], investigate equivalence of ensembles using the theory of large deviations. In [Lewis et al. 1994, § 7] and [Lewis et al. 1995, § 7.3] there is a discussion of nonequivalence of ensembles for the simplest mean-field model in statistical mechanics; namely, the Curie–Weiss model of a ferromagnet. However, despite the mathematical sophistication of these and other studies, none of them except for our papers [Costeniuc et al. 2005b; Ellis et al. 2000] explicitly addresses the general issue of the nonequivalence of ensembles.

Nonequivalence of ensembles has been observed in a wide range of systems that involve long-range interactions and that can be studied by the methods of [Costeniuc et al. 2005b; Ellis et al. 2000]. In all of these cases the microcanonical formulation gives rise to a richer set of equilibrium macrostates. For example, it has been shown computationally that the strongly reversing zonal-jet structures on Jupiter as well as the Great Red Spot fall into the nonequivalent range of an appropriate microcanonical ensemble [Turkington et al. 2001]. Other models for which ensemble nonequivalence has been observed include a

number of long-range, mean-field spin models including the Hamiltonian mean-field model [Dauxois et al. 2002], the mean-field X-Y model [Dauxois et al. 2000], and the mean-field Blume–Emery–Griffiths model [Barré et al. 2002; 2001; Ellis et al. 2004b]. For a mean-field version of the Potts model called the Curie–Weiss–Potts model, equivalence and nonequivalence of ensembles is analyzed in detail in [Costeniuc et al. 2005a; Costeniuc et al. 2006a]. Ensemble nonequivalence has also been observed in models of turbulent vorticity dynamics [DiBattista et al. 2001; Dibattista et al. 1998; Ellis et al. 2002a; Eyink and Spohn 1993; Kiessling and Lebowitz 1997; Robert and Sommeria 1991], models of plasmas [Kiessling and Neukirch 2003; Smith and O'Neil 1990], gravitational systems [Gross 1997; Hertel and Thirring 1971; Lynden-Bell and Wood 1968; Thirring 1970], and models of the Lennard–Jones gas [Borges and Tsallis 2002; Kiessling and Percus 1995]. A detailed discussion of ensemble nonequivalence for models of coherent structures in two dimensional turbulence is given in [Ellis et al. 2000, § 1.4].

Gaussian ensembles were introduced in [Hetherington 1987] and studied further in [Challa and Hetherington 1988a; Challa and Hetherington 1988b; Hetherington and Stump 1987; Johal et al. 2003; Stump and Hetherington 1987]. As these papers discuss, an important feature of Gaussian ensembles is that they allow one to account for ensemble-dependent effects in finite systems. Although not referred to by name, the Gaussian ensemble also plays a key role in [Kiessling and Lebowitz 1997], where it is used to address equivalence-of-ensemble questions for a point-vortex model of fluid turbulence.

Another seed out of which the research summarized in the present paper germinated is the paper [Ellis et al. 2002a]. There we study the equivalence of the microcanonical and canonical ensembles for statistical equilibrium models of coherent structures in two-dimensional and quasigeostrophic turbulence. Numerical computations demonstrate that, as in other cases, nonequivalence of ensembles occurs over a wide range of model parameters and that physically interesting microcanonical equilibria are often omitted by the canonical ensemble. In addition, in Section 5 of [Ellis et al. 2002a], we establish the nonlinear stability of the steady mean flows corresponding to microcanonical equilibria via a new Lyapunov argument. The associated stability theorem refines the well-known Arnold stability theorems, which do not apply when the microcanonical and canonical ensembles are not equivalent. The Lyapunov functional appearing in this new stability theorem is defined in terms of a generalized thermodynamic potential similar in form to $I(x) + \beta \tilde{H}(x) + \gamma[\tilde{H}(x)]^2$, the minimum points of which define the set of equilibrium macrostates for the Gaussian ensemble [see (2.14)].

Our goal in this paper is to give an overview of our theoretical work on ensemble equivalence presented in [Costeniuc et al. 2005b; Ellis et al. 2000]. The paper [Costeniuc et al. 2006b] investigates the physical principles underlying this theory. In Section 2 of the present paper, we first state the assumptions on the statistical mechanical models to which the theory of the present paper applies. We then define the three ensembles — microcanonical, canonical, and Gaussian — and specify the three associated sets of equilibrium macrostates in terms of large deviation principles. In Section 3 we state two sets of results on ensemble equivalence. The first involves the equivalence of the microcanonical and canonical ensembles, necessary and sufficient conditions for which are given in terms of support properties of the microcanonical entropy s defined in (1-3). The second involves the equivalence of the microcanonical and Gaussian ensembles, necessary and sufficient conditions for which are given in terms of support properties of the generalized microcanonical entropy s_γ defined in (1-6). Section 4 addresses a basic foundational issue in statistical mechanics. There we show that when the canonical ensemble is nonequivalent to the microcanonical ensemble on a subset of energy values u, it can often be replaced by a Gaussian ensemble that is universally equivalent to the microcanonical ensemble. In Section 5 the results on ensemble equivalence discussed in this paper are illustrated in the context of the Curie–Weiss–Potts lattice-spin model, a mean-field approximation to the nearest-neighbor Potts model. Several of the results presented near the end of this section are new.

2. Definitions of models and ensembles

One of the objectives of this paper is to show that when the canonical ensemble is nonequivalent to the microcanonical ensemble on a subset of energy values u, it can often be replaced by a Gaussian ensemble that is equivalent to the microcanonical ensemble for all u. Before introducing the various ensembles as well as the methodology for proving this result, we first specify the class of statistical mechanical models under consideration. The models are defined in terms of the following quantities.

1. A sequence of probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ indexed by $n \in \mathbb{N}$, which typically represents a sequence of finite dimensional systems. The Ω_n are the configuration spaces, $\omega \in \Omega_n$ are the microstates, and the P_n are the prior measures on the σ fields $\mathcal{F}_n$.
2. A sequence of positive scaling constant $a_n \to \infty$ as $n \to \infty$. In general a_n equals the total number of degrees of freedom in the model. In many cases a_n equals the number of particles.

3. For each $n \in \mathbb{N}$ a measurable functions H_n mapping Ω_n into $\mathbb{R}$. For $\omega \in \Omega_n$ we define the energy per degree of freedom by

$$h_n(\omega) = \frac{1}{a_n} H_n(\omega).$$

Typically, H_n in item 3 equals the Hamiltonian, which is associated with energy conservation in the model. The theory is easily generalized by replacing H_n by a vector of appropriate functions representing additional dynamical invariants associated with the model [Costeniuc et al. 2005b; Ellis et al. 2000].

A large deviation analysis of the general model is possible provided that there exist a space of macrostates, macroscopic variables, and an interaction representation function and provided that the macroscopic variables satisfy the large deviation principle (LDP) on the space of macrostates. These concepts are explained next.

4. **Space of macrostates**. This is a complete, separable metric space $\mathcal{X}$, which represents the set of all possible macrostates.
5. **Macroscopic variables.** These are a sequence of random variables Y_n mapping Ω_n into $\mathcal{X}$. These functions associate a macrostate in $\mathcal{X}$ with each microstate $\omega \in \Omega_n$.
6. **Interaction representation function.** This is a bounded, continuous functions $\tilde{H}$ mapping $\mathcal{X}$ into $\mathbb{R}$ such that as $n \to \infty$

$$h_n(\omega) = \tilde{H}(Y_n(\omega)) + \mathrm{o}(1) \quad \text{uniformly for } \omega \in \Omega_n; \tag{2.1}$$

i.e.,

$$\lim_{n\to\infty} \sup_{\omega\in\Omega_n} |h_n(\omega) - \tilde{H}(Y_n(\omega))| = 0.$$

The function $\tilde{H}$ enable us to write h_n, either exactly or asymptotically, as a function of the macrostate via the macroscopic variables Y_n.

7. **LDP for the macroscopic variables.** There exists a function I mapping $\mathcal{X}$ into $[0, \infty]$ and having compact level sets such that with respect to P_n the sequence Y_n satisfies the LDP on $\mathcal{X}$ with rate function I and scaling constants a_n. In other words, for any closed subset F of $\mathcal{X}$

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in F\} \le -\inf_{x\in F} I(x),$$

and for any open subset G of $\mathcal{X}$

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n\{Y_n \in G\} \ge -\inf_{x\in G} I(x).$$

It is helpful to summarize the LDP by the formal notation $P_n\{Y_n \in dx\} \asymp \exp[-a_n I(x)]$. This notation expresses the fact that, to a first degree of

approximation, $P_n\{Y_n \in dx\}$ behaves like an exponential that decays to 0 whenever $I(x) > 0$.

The assumptions on the statistical mechanical models just stated, as well as a number of definitions to appear later, follow the presentation in [Costeniuc et al. 2005b], which is adapted for applications to lattice spin systems and related models. These assumptions and definitions differ slightly from those in [Ellis et al. 2000], where they are adapted for applications to statistical mechanical models of coherent structures in turbulence. The major difference is that in the asymptotic relationship (2.1) and in the definition (2.3) of the microcanonical ensemble $P_n^{u,r}$, h_n is replaced by H_n in [Ellis et al. 2000]. In addition, in the definition (2.4) of the canonical ensemble $P_{n,\beta}$, $a_n h_n$ is replaced by H_n in [Ellis et al. 2000]. Similarly, in the definition (2.6) of the Gaussian ensemble $P_{n,\beta,\gamma}$, $a_n h_n$ and $a_n h_n^2$ are replaced by H_n and H_n^2 to yield the Gaussian ensemble used to study models of coherent structures in turbulence. Finally, in the present paper the LDP for Y_n is derived with respect to $P_{n,\beta}$ and $P_{n,\beta,\gamma}$ while in models of coherent structures in turbulence the LDP for Y_n is derived with respect to $P_{n,a_n\beta}$ and $P_{n,a_n\beta,a_n\gamma}$, in which β and γ are both scaled by a_n. With only these minor changes in notation, all the results stated here are applicable to models of coherent structures in turbulence and in turn, all the results derived in [Ellis et al. 2000] for models of coherent structures in turbulence are applicable here.

A wide variety of statistical mechanical models satisfy the assumptions listed in items 1–7 or the modifications just discussed. Hence they can be studied by the methods of [Costeniuc et al. 2005b; Ellis et al. 2000]. We next give six examples. The first two are long-range spin systems, the third a class of short-range spin systems, the fourth a model of two-dimensional turbulence, the fifth a model of quasigeostrophic turbulence, and the sixth a model of dispersive wave turbulence.

1. The mean-field Blume–Emery–Griffiths model [Blume et al. 1971] is one of the simplest lattice-spin models known to exhibit both a continuous, second-order phase transition and a discontinuous, first-order phase transition. The space of macrostates for this model is the set of probability measures on a certain finite set, the macroscopic variables are the empirical measures associated with the spin configurations, and the associated LDP is Sanov's Theorem, for which the rate function is a relative entropy. Various features of this model are studied in [Barré et al. 2002; Barré et al. 2001; Ellis et al. 2005; Ellis et al. 2004b].
2. The Curie–Weiss–Potts model is a mean-field approximation to the nearest-neighbor Potts model [Wu 1982]. For the Curie–Weiss–Potts model, the space of macrostates, the macroscopic variables, and the associated LDP are similar to those in the mean-field Blume–Emery–Griffiths model. The

Curie–Weiss–Potts model nicely illustrates the general results on ensemble equivalence discussed in this paper and is discussed in Section 5.

3. Short-range spin systems such as the Ising model on $\mathbb{Z}^d$ and numerous generalizations can also be handled by the methods of this paper. The large deviation techniques required to analyze these models are much more subtle than in the case of the long-range, mean-field models considered in items 1 and 2. For the Ising model the space of macrostates is the space of translation-invariant probability measures on $\mathbb{Z}^d$, the macroscopic variables are the empirical processes associated with the spin configurations, and the rate function in the associated LDP is the mean relative entropy [Ellis 1985; Föllmer and Orey 1988; Olla 1988].
4. The Miller–Robert model is a model of coherent structures in an ideal, two-dimensional fluid that includes all the exact invariants of the vorticity transport equation [Miller 1990; Robert 1991]. The space of macrostates is the space of Young measures on the vorticity field. The large deviation analysis of this model developed first in [Robert 1991] and more recently in [Boucher et al. 2000] gives a rigorous derivation of maximum entropy principles governing the equilibrium behavior of the ideal fluid.
5. In geophysical applications, another version of the model in item 4 is preferred, in which the enstrophy integrals are treated canonically and the energy and circulation are treated microcanonically [Ellis et al. 2002a]. In those formulations, the space of macrostates is $L^2(\Lambda)$ or $L^\infty(\Lambda)$ depending on the constraints on the vorticity field. The large deviation analysis is carried out in [Ellis et al. 2002b]. The paper [Ellis et al. 2002a] shows how the nonlinear stability of the steady mean flows arising as equilibrium macrostates can be established by utilizing the appropriate generalized thermodynamic potentials.
6. A statistical equilibrium model of solitary wave structures in dispersive wave turbulence governed by a nonlinear Schrödinger equation is studied in [Ellis et al. 2004a]. The large deviation analysis given in [Ellis et al. 2004a] derives rigorously the concentration phenomenon observed in long-time numerical simulations and predicted by mean-field approximations [Jordan et al. 2000; Lebowitz et al. 1989]. The space of macrostates is $L^2(\Lambda)$, where Λ is a bounded interval or more generally a bounded domain in $\mathbb{R}^d$. The macroscopic variables are certain Gaussian processes.

We now return to the general theory, first introducing the function whose support and concavity properties completely determine all aspects of ensemble equivalence and nonequivalence. This function is the microcanonical entropy, defined for $u \in \mathbb{R}$ by

$$s(u) = -\inf\{I(x) : x \in \mathcal{X}, \tilde{H}(x) = u\}. \tag{2.2}$$

Since I maps $\mathcal{X}$ into $[0,\infty]$, s maps $\mathbb{R}^\sigma$ into $[-\infty,0]$. Moreover, since I is lower semicontinuous and $\tilde{H}$ is continuous on $\mathcal{X}$, s is upper semicontinuous on $\mathbb{R}^\sigma$. We define $\operatorname{dom} s$ to be the set of $u \in \mathbb{R}^\sigma$ for which $s(u) > -\infty$. In general, $\operatorname{dom} s$ is nonempty since $-s$ is a rate function [Ellis et al. 2000, Prop. 3.1(a)]. For each $u \in \operatorname{dom} s$, $r > 0$, $n \in \mathbb{N}$, and set $B \in \mathcal{F}_n$ the microcanonical ensemble is defined to be the conditioned measure

$$P_n^{u,r}\{B\} = P_n\{B \mid h_n \in [u-r, u+r]\}. \tag{2.3}$$

As shown in [Ellis et al. 2000, p. 1027], if $u \in \operatorname{dom} s$, then for all sufficiently large n, $P_n\{h_n \in [u-r, u+r]\} > 0$; thus the conditioned measures $P_n^{u,r}$ are well defined.

A mathematically more tractable probability measure is the canonical ensemble. For each $n \in \mathbb{N}$, $\beta \in \mathbb{R}$, and set $B \in \mathcal{F}_n$ we define the partition function

$$Z_n(\beta) = \int_{\Omega_n} \exp[-a_n \beta h_n]\, dP_n,$$

which is well defined and finite, and the probability measure

$$P_{n,\beta}\{B\} = \frac{1}{Z_n(\beta)} \cdot \int_B \exp[-a_n \beta h_n]\, dP_n. \tag{2.4}$$

The measures $P_{n,\beta}$ are Gibbs states that define the canonical ensemble for the given model.

The Gaussian ensemble is a natural perturbation of the canonical ensemble. For each $n \in \mathbb{N}$, $\beta \in \mathbb{R}$, and $\gamma \in [0,\infty)$ we define the Gaussian partition function

$$Z_n(\beta,\gamma) = \int_{\Omega_n} \exp[-a_n \beta h_n - a_n \gamma h_n^2]\, dP_n. \tag{2.5}$$

This is well defined and finite because the h_n are bounded. For $B \in \mathcal{F}_n$ we also define the probability measure

$$P_{n,\beta,\gamma}\{B\} = \frac{1}{Z_n(\beta,\gamma)} \cdot \int_B \exp[-a_n \beta h_n - a_n \gamma h_n^2]\, dP_n, \tag{2.6}$$

which we call the Gaussian ensemble. One can generalize this by replacing the quadratic function by a continuous function g that is bounded below. This gives rise to the generalized canonical ensemble, which the theory developed in [Costeniuc et al. 2005b] allows one to treat.

Using the theory of large deviations, one introduces the sets of equilibrium macrostates for each ensemble. It is proved in [Ellis et al. 2000, Theorem 3.2]

that with respect to the microcanonical ensemble $P_n^{u,r}$, Y_n satisfies the LDP on $\mathcal{X}$, in the double limit $n \to \infty$ and $r \to 0$, with rate function

$$I^u(x) = \begin{cases} I(x) + s(u) & \text{if } \tilde{H}(x) = u \\ \infty & \text{otherwise .} \end{cases} \tag{2.7}$$

I^u is nonnegative on $\mathcal{X}$, and for $u \in \operatorname{dom} s$, I^u attains its infimum of 0 on the set

$$\begin{aligned} \mathcal{E}^u &= \{x \in \mathcal{X} : I^u(x) = 0\} \\ &= \{x \in \mathcal{X} : I(x) \text{ is minimized subject to } \tilde{H}(x) = u\}. \end{aligned} \tag{2.8}$$

This set is precisely the set of solutions of the constrained minimization problem (1-1).

In order to state the LDPs for the other two ensembles, we bring in the canonical free energy, defined for $\beta \in \mathbb{R}$ by

$$\varphi(\beta) = -\lim_{n\to\infty} \frac{1}{a_n} \log Z_n(\beta),$$

and the Gaussian free energy, defined for $\beta \in \mathbb{R}$ and $\gamma \geq 0$ by

$$\varphi(\beta, \gamma) = -\lim_{n\to\infty} \frac{1}{a_n} \log Z_n(\beta, \gamma).$$

It is proved in [Ellis et al. 2000, Theorem 2.4] that the limit defining $\varphi(\beta)$ exists and is given by

$$\varphi(\beta) = \inf_{y\in\mathcal{X}} \{I(y) + \beta \tilde{H}(y)\} \tag{2.9}$$

and that with respect to $P_{n,\beta}$, Y_n satisfies the LDP on $\mathcal{X}$ with rate function

$$I_\beta(x) = I(x) + \beta \tilde{H}(x) - \varphi(\beta). \tag{2.10}$$

I_β is nonnegative on $\mathcal{X}$ and attains its infimum of 0 on the set

$$\begin{aligned} \mathcal{E}_\beta &= \{x \in \mathcal{X} : I_\beta(x) = 0\} \\ &= \{x \in \mathcal{X} : I(x) + \langle \beta, \tilde{H}(x)\rangle \text{ is minimized}\}. \end{aligned} \tag{2.11}$$

This set is precisely the set of solutions of the unconstrained minimization problem (1-2).

A straightforward extension of these results shows that the limit defining $\varphi(\beta, \gamma)$ exists and is given by

$$\varphi(\beta, \gamma) = \inf_{y\in\mathcal{X}} \{I(y) + \beta \tilde{H}(y) + \gamma[\tilde{H}(y)]^2\} \tag{2.12}$$

and that with respect to $P_{n,\beta,g}$, Y_n satisfies the LDP on $\mathcal{X}$ with rate function

$$I_{\beta,\gamma}(x) = I(x) + \beta \tilde{H}(x) + \gamma[\tilde{H}(x)]^2 - \varphi(\beta, \gamma). \tag{2.13}$$

$I_{\beta,\gamma}$ is nonnegative on $\mathcal{X}$ and attains its infimum of 0 on the set

$$\begin{aligned}\mathcal{E}_{\beta,\gamma} &= \{x \in \mathcal{X} : I_{\beta,\gamma}(x) = 0\} \\ &= \{x \in \mathcal{X} : I(x) + \langle \beta, \tilde{H}(x)\rangle + \gamma[\tilde{H}(x)]^2 \text{ is minimized}\}.\end{aligned} \tag{2.14}$$

This set is precisely the set of solutions of the penalized minimization problem (1-5).

For $u \in \operatorname{dom} s$, let x be any element of $\mathcal{X}$ satisfying $I^u(x) > 0$. The formal notation

$$P_n^{u,r}\{Y_n \in dx\} \asymp e^{-a_n I^u(x)}$$

suggests that x has an exponentially small probability of being observed in the limit $n \to \infty$, $r \to 0$. Hence it makes sense to identify $\mathcal{E}^u$ with the set of microcanonical equilibrium macrostates. In the same way we identify with $\mathcal{E}_\beta$ the set of canonical equilibrium macrostates and with $\mathcal{E}_{\beta,\gamma}$ the set of generalized canonical equilibrium macrostates. A rigorous justification is given in [Ellis et al. 2000, Theorem 2.4(d)].

3. Equivalence and nonequivalence of the three ensembles

Having defined the sets of equilibrium macrostates $\mathcal{E}^u$, $\mathcal{E}_\beta$, and $\mathcal{E}_{\beta,\gamma}$ for the microcanonical, canonical and Gaussian ensembles, we now show how these sets are related to one another. In Theorem 3.1 we state the results proved in [Ellis et al. 2000] concerning equivalence and nonequivalence of the microcanonical and canonical ensembles. Then in Theorem 3.3 we extend these results to the Gaussian ensemble [Costeniuc et al. 2005b].

Parts (a)–(c) of Theorem 3.1 give necessary and sufficient conditions, in terms of support properties of s, for equivalence and nonequivalence of $\mathcal{E}^u$ and $\mathcal{E}_\beta$. These assertions are proved in Theorems 4.4 and 4.8 in [Ellis et al. 2000]. Part (a) states that s has a strictly supporting line at u if and only if full equivalence of ensembles holds; i.e., if and only if there exists a β such that $\mathcal{E}^u = \mathcal{E}_\beta$. The most surprising result, given in part (c), is that s has no supporting line at u if and only if nonequivalence of ensembles holds in the strong sense that $\mathcal{E}^u \cap \mathcal{E}_\beta = \varnothing$ for all β. Part (c) is to be contrasted with part (d), which states that for any β canonical equilibrium macrostates can always be realized microcanonically. Part (d) is proved in Theorem 4.6 in [Ellis et al. 2000]. Thus one conclusion of this theorem is that at the level of equilibrium macrostates, in general the microcanonical ensemble is the richer of the two ensembles.

THEOREM 3.1. *In parts* (a), (b), *and* (c), *u denotes any point in* $\operatorname{dom} s$.

(a) ***Full equivalence***. *There exists β such that $\mathcal{E}^u = \mathcal{E}_\beta$ if and only if s has a strictly supporting line at u with slope β; i.e.,*

$$s(v) < s(u) + \beta(v-u) \text{ for all } v \neq u\,.$$

(b) ***Partial equivalence***. *There exists β such that $\mathcal{E}^u \subset \mathcal{E}_\beta$ but $\mathcal{E}^u \neq \mathcal{E}_\beta$ if and only if s has a nonstrictly supporting line at u with slope β; i.e.,*

$$s(v) \leq s(u) + \beta(v-u) \text{ for all } v \text{ with equality for some } v \neq u.$$

(c) ***Nonequivalence***. *For all β, $\mathcal{E}^u \cap \mathcal{E}_\beta = \varnothing$ if and only if s has no supporting line at u; i.e.,*

$$\text{for all } \beta \text{ there exists } v \text{ such that } s(v) > s(u) + \beta(v-u).$$

(d) ***Canonical is always realized microcanonically***. *For any $\beta \in \mathbb{R}$ we have $\tilde{H}(\mathcal{E}_\beta) \subset \operatorname{dom} s$ and*

$$\mathcal{E}_\beta = \bigcup_{u \in \tilde{H}(\mathcal{E}_\beta)} \mathcal{E}^u.$$

We highlight several features of the theorem in order to illuminate their physical content. In part (a) let us add the assumption that for a given $u \in \operatorname{dom} s$ there exists a unique β such that $\mathcal{E}^u = \mathcal{E}_\beta$. If s is differentiable at u and s and the double Legendre–Fenchel transform s^{**} are equal in a neighborhood of u, then β is given by the standard thermodynamic formula $\beta = s'(u)$ [Costeniuc et al. 2005b, Theorem A.4(b)]. The inverse relationship can be obtained from part (d) of the theorem under the added assumption that $\mathcal{E}_\beta$ consists of a unique macrostate or more generally that for all $x \in \mathcal{E}_\beta$ the values $\tilde{H}(x)$ are equal. Then $\mathcal{E}_\beta = \mathcal{E}^{u(\beta)}$, where $u(\beta) = \tilde{H}(x)$ for any $x \in \mathcal{E}_\beta$; $u(\beta)$ denotes the mean energy realized at equilibrium in the canonical ensemble. The relationship $u = u(\beta)$ inverts the relationship $\beta = s'(u)$. Partial ensemble equivalence can be seen in part (d) under the added assumption that for a given β, $\mathcal{E}_\beta$ can be partitioned into at least two sets $\mathcal{E}_{\beta,i}$ such that for all $x \in \mathcal{E}_{\beta,i}$ the values $\tilde{H}(x)$ are equal but $\tilde{H}(x) \neq \tilde{H}(y)$ whenever $x \in \mathcal{E}_{\beta,i}$ and $y \in \mathcal{E}_{\beta,j}$ for $i \neq j$. Then $\mathcal{E}_\beta = \bigcup_i \mathcal{E}^{u_i(\beta)}$, where $u_i(\beta) = \tilde{H}(x)$, $x \in \mathcal{E}_{\beta,i}$. Clearly, for each i, $\mathcal{E}^{u_i(\beta)} \subset \mathcal{E}_\beta$ but $\mathcal{E}^{u_i(\beta)} \neq \mathcal{E}_\beta$. Physically, this corresponds to a situation of coexisting phases that normally takes place at a first-order phase transition [Touchette et al. 2004].

Before continuing with our analysis of ensemble equivalence, we make a number of basic definitions. A function f on $\mathbb{R}$ is said to be concave on $\mathbb{R}$ if f maps $\mathbb{R}$ into $\mathbb{R} \cup \{-\infty\}$, $f \not\equiv -\infty$, and for all u and v in $\mathbb{R}$ and all $\lambda \in (0,1)$

$$f(\lambda u + (1-\lambda)v) \geq \lambda f(u) + (1-\lambda) f(v).$$

Let $f \not\equiv -\infty$ be a function mapping $\mathbb{R}$ into $\mathbb{R} \cup \{-\infty\}$. We define dom f to be the set of u for which $f(u) > -\infty$. For β and u in $\mathbb{R}$ the Legendre–Fenchel transforms f^* and f^{**} are defined by

$$f^*(\beta) = \inf_{u \in \mathbb{R}} \{\langle \beta, u \rangle - f(u)\} \quad \text{and} \quad f^{**}(u) = \inf_{\beta \in \mathbb{R}} \{\langle \beta, u \rangle - f^*(\beta)\}.$$

The function f^* is concave and upper semicontinuous on $\mathbb{R}$ and for all u we have $f^{**}(u) = f(u)$ if and only if f is concave and upper semicontinuous on $\mathbb{R}$ [Ellis 1985, Theorem VI.5.3]. When f is not concave and upper semicontinuous, then f^{**} is the smallest concave, upper semicontinuous function on $\mathbb{R}$ that satisfies $f^{**}(u) \geq f(u)$ for all u [Costeniuc et al. 2005b, Prop. A.2]. In particular, if for some u, $f(u) \neq f^{**}(u)$, then $f(u) < f^{**}(u)$.

Let $f \not\equiv -\infty$ be a function mapping $\mathbb{R}$ into $\mathbb{R} \cup \{-\infty\}$, u a point in dom f, and K a convex subset of dom f. We have the following four additional definitions: f is concave at u if $f(u) = f^{**}(u)$; f is not concave at u if $f(u) < f^{**}(u)$; f is concave on K if f is concave at all $u \in K$; and f is strictly concave on K if for all $u \neq v$ in K and all $\lambda \in (0, 1)$

$$f(\lambda u + (1-\lambda)v) > \lambda f(u) + (1-\lambda) f(v).$$

We also introduce two sets that play a central role in the theory. Let f be a concave function on $\mathbb{R}$ whose domain is an interval having nonempty interior. For $u \in \mathbb{R}$ the superdifferential of f at u, denoted by $\partial f(u)$, is defined to be the set of β such that β is the slope of a supporting line of f at u. Any such β is called a supergradient of f at u. Thus, if f is differentiable at $u \in \operatorname{int}(\operatorname{dom} f)$, then $\partial f(u)$ consists of the unique point $\beta = f'(u)$. If f is not differentiable at $u \in \operatorname{int}(\operatorname{dom} f)$, then $\operatorname{dom} \partial f$ consists of all β satisfying the inequalities

$$(f')^+(u) \leq \beta \leq (f')^-(u),$$

where $(f')^-(u)$ and $(f')^+(u)$ denote the left-hand and right-hand derivatives of f at u. The domain of ∂f, denoted by $\operatorname{dom} \partial f$, is then defined to be the set of u for which $\partial f(u) \neq \varnothing$.

Complications arise because $\operatorname{dom} \partial f$ can be a proper subset of dom f, as simple examples clearly show. Let b be a boundary point of dom f for which $f(b) > -\infty$. Then b is in $\operatorname{dom} \partial f$ if and only if the one-sided derivative of f at b is finite. For example, if b is a left hand boundary point of dom f and $(f')^+(b)$ is finite, then $\partial f(b) = [(f')^+(b), \infty)$; any $\beta \in \partial f(b)$ is the slope of a supporting line at b. The possible discrepancy between $\operatorname{dom} \partial f$ and dom f introduces unavoidable technicalities in the statements of several results concerning the existence of supporting lines.

One of our goals is to find concavity and support conditions on the microcanonical entropy guaranteeing that the microcanonical and canonical ensembles are fully equivalent at all points $u \in \operatorname{dom} s$ except possibly boundary points. If this is the case, then we say that the ensembles are universally equivalent. Here is a basic result in that direction. The universal equivalence stated in part (b) follows from part (a) and from part (a) of Theorem 3.1. The rest of the theorem depends on facts concerning concave functions [Costeniuc et al. 2005b, p. 1305].

THEOREM 3.2. *Assume that* $\operatorname{dom} s$ *is an interval having nonempty interior and that* s *is strictly concave on* $\operatorname{int}(\operatorname{dom} s)$ *and continuous on* $\operatorname{dom} s$. *The following conclusions hold.*

(a) s *has a strictly supporting line at all* $u \in \operatorname{dom} s$ *except possibly boundary points.*

(b) *The microcanonical and canonical ensembles are universally equivalent; i.e., fully equivalent at all* $u \in \operatorname{dom} s$ *except possibly boundary points.*

(c) s *is concave on* $\mathbb{R}$, *and for each* u *in part* (b) *the corresponding* β *in the statement of full equivalence is any element of* $\partial s(u)$.

(d) *If* s *is differentiable at some* $u \in \operatorname{dom} s$, *then the corresponding* β *in part* (b) *is unique and is given by the standard thermodynamic formula* $\beta = s'(u)$.

The next theorem extends Theorem 3.1 by giving equivalence and nonequivalence results involving $\mathcal{E}^u$ and $\mathcal{E}_{\beta,\gamma}$, the sets of equilibrium macrostates with respect to the microcanonical and Gaussian ensembles. The chief innovation is that $s(u)$ in Theorem 3.1 is replaced here by the generalized microcanonical entropy $s(u) - \gamma u^2$. As we point out after the statement of Theorem 3.3, for the purpose of applications part (a) is its most important contribution. The usefulness of Theorem 3.3 is matched by the simplicity with which it follows from Theorem 3.1. Theorem 3.3 is a special case of Theorem 3.4 in [Costeniuc et al. 2005b], obtained by specializing the generalized canonical ensemble and the associated set of equilibrium macrostates to the Gaussian ensemble and the set $\mathcal{E}_{\beta,\gamma}$ of Gaussian equilibrium macrostates.

THEOREM 3.3. *Given* $\gamma \geq 0$, *define* $s_\gamma(u) = s(u) - \gamma u^2$. *In parts* (a), (b), *and* (c), u *denotes any point in* $\operatorname{dom} s$.

(a) ***Full equivalence***. *There exists* β *such that* $\mathcal{E}^u = \mathcal{E}_{\beta,\gamma}$ *if and only if* s_γ *has a strictly supporting line at* u *with slope* β.

(b) ***Partial equivalence***. *There exists* β *such that* $\mathcal{E}^u \subset \mathcal{E}_{\beta,\gamma}$ *but* $\mathcal{E}^u \neq \mathcal{E}_{\beta,\gamma}$ *if and only if* s_γ *has a nonstrictly supporting line at* u *with slope* β.

(c) ***Nonequivalence***. *For all* β, $\mathcal{E}^u \cap \mathcal{E}_{\beta,\gamma} = \varnothing$ *if and only if* s_γ *has no supporting line at* u.

(d) ***Gaussian is always realized microcanonically***. *For any β we have*

$$\tilde{H}(\mathcal{E}_{\beta,\gamma}) \subset \operatorname{dom} s, \quad \mathcal{E}_{\beta,\gamma} = \bigcup_{u \in \tilde{H}(\mathcal{E}_{\beta,\gamma})} \mathcal{E}^u.$$

PROOF. For $\gamma \geq 0$ and $B \in \mathcal{F}_n$ we define a new probability measure

$$P_{n,\gamma}\{B\} = \frac{1}{\displaystyle\int_{\Omega_n} \exp[-a_n \gamma h_n^2]\, dP_n} \cdot \int_B \exp[-a_n \gamma h_n^2]\, dP_n.$$

With respect to $P_{n,\gamma}$, Y_n satisfies the LDP on $\mathcal{X}$ with rate function

$$I_\gamma(x) = I(x) + \gamma[\tilde{H}(x)]^2 - \psi(\gamma),$$

where $\psi(\gamma) = \inf_{y \in \mathcal{X}}\{I(y) + \gamma[\tilde{H}(y)]^2\}$. Replacing the prior measure P_n in the canonical ensemble with $P_{n,\gamma}$ gives the Gaussian ensemble $P_{n,\beta,\gamma}$, which has $\mathcal{E}_{\beta,\gamma}$ as the associated set of equilibrium macrostates. On the other hand, replacing the prior measure P_n in the microcanonical ensemble with $P_{n,\gamma}$ gives

$$P_{n,\gamma}^{u,r}\{B\} = P_{n,\gamma}\{B \mid h_n \in [u-r, u+r]\},$$

By continuity, for ω satisfying $h_n(\omega) \in [u-r, u+r]$, $[h_n(\omega)]^2$ converges to u^2 uniformly in ω and n as $r \to 0$. It follows that with respect to $P_{n,\gamma}^{u,r}$, Y_n satisfies the LDP on $\mathcal{X}$, in the double limit $n \to \infty$ and $r \to 0$, with the same rate function I^u as in the LDP for Y_n with respect to $P_n^{u,r}$. As a result, the set of equilibrium macrostates corresponding to $P_{n,\gamma}^{u,r}$ coincides with the set $\mathcal{E}^u$ of microcanonical equilibrium macrostates.

It follows from parts (a)–(c) of Theorem 3.1 that all equivalence and nonequivalence relationships between $\mathcal{E}^u$ and $\mathcal{E}_{\beta,\gamma}$ are expressed in terms of support properties of the function $\tilde{s}_\gamma$ obtained from s by replacing the rate function I by the new rate function I_γ. The function $\tilde{s}_\gamma$ is given by

$$\begin{aligned}\tilde{s}_\gamma(u) &= -\inf\{I_\gamma(x) : x \in \mathcal{X}, \tilde{H}(x) = u\} \\ &= -\inf\{I(x) + \gamma \tilde{H}(x)^2 : x \in \mathcal{X}, \tilde{H}(x) = u\} + \psi(\gamma) \\ &= s(u) - \gamma u^2 + \psi(\gamma).\end{aligned}$$

Since $\tilde{s}_\gamma(u)$ differs from $s_\gamma(u) = s(u) - \gamma u^2$ by the constant $\psi(\gamma)$, we conclude that all equivalence and nonequivalence relationships between $\mathscr{E}^u$ and $\mathcal{E}_{\beta,\gamma}$ are expressed in terms of the same support properties of s_γ. This completes the derivation of parts (a)–(c) of Theorem 3.3 from parts (a)–(c) of Theorem 3.1. Similarly, part (d) of Theorem 3.3 follows from part (d) of Theorem 3.1. □

The importance of part (a) of Theorem 3.3 in applications is emphasized by the following theorem, which will be applied in the sequel. This theorem is the analogue of Theorem 3.2 for the Gaussian ensemble, s in that theorem being replaced by s_γ. The functions s and s_γ have the same domains. The universal equivalence stated in part (b) of the next theorem follows from part (a) and from part (a) of Theorem 3.3.

THEOREM 3.4. *For $\gamma \geq 0$, define $s_\gamma(u) = s(u) - \gamma u^2$. Assume that* $\operatorname{dom} s$ *is an interval having nonempty interior and that s_γ is strictly concave on* $\operatorname{int}(\operatorname{dom} s)$ *and continuous on dom s. The following conclusions hold.*

(a) *s_γ has a strictly supporting line at all $u \in \operatorname{dom} s$ except possibly boundary points.*

(b) *The microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are universally equivalent; i.e., fully equivalent at all $u \in \operatorname{dom} s$ except possibly boundary points.*

(c) *s_γ is concave on $\mathbb{R}$, and for each u in part* (b) *the corresponding β in the statement of full equivalence is any element of $\partial s_\gamma(u)$.*

(d) *If s_γ is differentiable at some $u \in \operatorname{dom} s$, then the corresponding β in part* (b) *is unique and is given by the thermodynamic formula $\beta = s'_\gamma(u)$.*

The most important repercussion of Theorem 3.4 is the ease with which one can prove that the microcanonical and Gaussian ensembles are universally equivalent in those cases in which the microcanonical and canonical ensembles are not fully or partially equivalent. This rests mainly on part (b) of Theorem 3.4, which states that universal equivalence of ensembles holds if there exists a $\gamma \geq 0$ such that s_γ is strictly concave on $\operatorname{int}(\operatorname{dom} s)$. The existence of such a γ follows from a natural set of hypotheses on s stated in Theorem 4.2 in the next section.

4. Universal equivalence via the Gaussian ensemble

This section addresses a basic foundational issue in statistical mechanics. Under the assumption that the microcanonical entropy is C^2 and s'' is bounded above, we show in Theorem 4.2 that when the canonical ensemble is nonequivalent to the microcanonical ensemble on a subset of energy values u, it can often be replaced by a Gaussian ensemble that is universally equivalent to the microcanonical ensemble; i.e., fully equivalent at all $u \in \operatorname{dom} s$ except possibly boundary points. Theorem 4.3 is a weaker version that can often be applied when s'' is not bounded above. In the last section of the paper, these results will be illustrated in the context of the Curie–Weiss–Potts lattice-spin model.

In Theorem 4.2 the strategy is to find a quadratic function γu^2 such that $s_\gamma(u) = s(u) - \gamma u^2$ is strictly concave on $\operatorname{int}(\operatorname{dom} s)$ and continuous on $\operatorname{dom} s$. Parts (a) and (b) of Theorem 3.4 then yields the universal equivalence. As

the next proposition shows, an advantage of working with quadratic functions is that support properties of s_γ involving a supporting line are equivalent to support properties of s involving a supporting parabola defined in terms of γ. This observation gives a geometrically intuitive way to find a quadratic function guaranteeing universal ensemble equivalence.

In order to state the proposition, we need a definition. Let f be a function mapping $\mathbb{R}$ into $\mathbb{R} \cup \{-\infty\}$, u and β points in $\mathbb{R}$, and $\gamma \geq 0$. We say that f has a supporting parabola at u with parameters (β, γ) if

$$f(v) \leq f(u) + \beta(v-u) + \gamma(v-u)^2 \quad \text{for all } v. \tag{4.1}$$

The parabola is said to be strictly supporting if the inequality is strict for all $v \neq u$.

PROPOSITION 4.1. *f has a (strictly) supporting parabola at u with parameters (β, γ) if and only if $f - \gamma(\cdot)^2$ has a (strictly) supporting line at u with slope $\tilde{\beta}$. The quantities β and $\tilde{\beta}$ are related by $\tilde{\beta} = \beta - 2\gamma u$.*

PROOF. We use the identity $(v-u)^2 = v^2 - 2u(v-u) - u^2$. If f has a strictly supporting parabola at u with parameters (β, γ), then for all $v \neq u$

$$f(v) - \gamma v^2 < f(u) - \gamma u^2 + \tilde{\beta}(v-u),$$

where $\tilde{\beta} = \beta - 2\gamma u$. Thus $f - \gamma(\cdot)^2$ has a strictly supporting line at u with slope $\tilde{\beta}$. The converse is proved similarly, as is the case in which the supporting line or parabola is supporting but not strictly supporting. □

The first application of Theorem 3.4 is Theorem 4.2, which gives a criterion guaranteeing the existence of a quadratic function γu^2 such that $s_\gamma(u) = s(u) - \gamma u^2$ is strictly concave on $\operatorname{dom} s$. The criterion — that s'' is bounded above on the interior of $\operatorname{dom} s$ — is essentially optimal for the existence of a fixed quadratic function γu^2 guaranteeing the strict concavity of s_γ. The situation in which s'' is not bounded above on the interior of $\operatorname{dom} s$ can often be handled by Theorem 4.3, which is a local version of Theorem 4.2.

THEOREM 4.2. *Assume that $\operatorname{dom} s$ is an interval having nonempty interior. Assume also that s is continuous on $\operatorname{dom} s$, s is twice continuously differentiable on $\operatorname{int}(\operatorname{dom} s)$, and s'' is bounded above on $\operatorname{int}(\operatorname{dom} s)$. Then for all sufficiently large $\gamma \geq 0$, conclusions* (a)–(c) *hold. Specifically, if s is strictly concave on $\operatorname{dom} s$, then we choose any $\gamma \geq 0$, and otherwise we choose*

$$\gamma > \gamma_0 = \tfrac{1}{2} \cdot \sup_{u \in \operatorname{int}(\operatorname{dom} s)} s''(u). \tag{4.2}$$

(a) *$s_\gamma(u) = s(u) - \gamma u^2$ is strictly concave and continuous on $\operatorname{dom} s$.*

(b) *s_γ has a strictly supporting line, and s has a strictly supporting parabola, at all $u \in \operatorname{dom} s$ except possibly boundary points. At a boundary point s_γ has a strictly supporting line, and s has a strictly supporting parabola, if and only if the one-sided derivative of s_γ is finite at that boundary point.*

(c) *The microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are universally equivalent; i.e., fully equivalent at all $u \in \operatorname{dom} s$ except possibly boundary points. For all $u \in \operatorname{int}(\operatorname{dom} s)$ the value of β defining the universally equivalent Gaussian ensemble is unique and equals $\beta = s'(u) - 2\gamma u$.*

PROOF. (a) If s is strictly concave on $\operatorname{dom} s$, then s_γ is also strictly concave on this set for any $\gamma \geq 0$. We now consider the case in which s is not strictly concave on $\operatorname{dom} s$. For any $\gamma \geq 0$, s_γ is continuous on $\operatorname{dom} s$. If, in addition, we choose $\gamma > \gamma_0$ in accordance with (4.2), then for all $u \in \operatorname{int}(\operatorname{dom} s)$

$$s''_\gamma(u) = s''(u) - 2\gamma < 0.$$

A straightforward extension of the proof of Theorem 4.4 in [Rockafellar 1970], in which the inequalities in the first two displays are replaced by strict inequalities, shows that $-s_\gamma$ is strictly convex on $\operatorname{int}(\operatorname{dom} s)$ and thus that s_γ is strictly concave on $\operatorname{int}(\operatorname{dom} s)$. If s_γ is not strictly concave on $\operatorname{dom} s$, then s_γ must be affine on an interval. Since this violates the strict concavity on $\operatorname{int}(\operatorname{dom} s)$, part (a) is proved.

(b) The first assertion follows from part (a) of the present theorem, part (a) of Theorem 3.4, and Proposition 4.1. Concerning the second assertion about boundary points, the reader is referred to the discussion before Theorem 3.2.

(c) The universal equivalence of the two ensembles is a consequence of part (a) of the present theorem and part (b) of Theorem 3.4. The full equivalence of the ensembles at all $u \in \operatorname{int}(\operatorname{dom} s)$ is equivalent to the existence of a strictly supporting line at each $u \in \operatorname{int}(\operatorname{dom} s)$ [Theorem 3.3(a)]. Since $s_\gamma(u)$ is differentiable at all $u \in \operatorname{int}(\operatorname{dom} s)$, for each u the slope of the strictly supporting line at u is unique and equals $s'_\gamma(u)$ [Costeniuc et al. 2005b, Theorem A.1(b)]. □

Suppose that s is C^2 on the interior of $\operatorname{dom} s$ but the second-order partial derivatives of s are not bounded above. This arises, for example, in the Curie–Weiss–Potts model, in which $\operatorname{dom} s$ is a closed, bounded interval of $\mathbb{R}$ and $s''(u) \to \infty$ as u approaches the right hand endpoint of $\operatorname{dom} s$ [see § 5]. In such cases one cannot expect that the conclusions of Theorems 4.2 will be satisfied; in particular, that there exists $\gamma \geq 0$ such that $s_\gamma(u) = s(u) - \gamma u^2$ has a strictly supporting line at each point of the interior of $\operatorname{dom} s$ and thus that the ensembles are universally equivalent.

In order to overcome this difficulty, we introduce Theorem 4.3, a local version of Theorem 4.2. Theorem 4.3 handles the case in which s is C^2 on an open set K but either K is not all of $\operatorname{int}(\operatorname{dom} s)$ or $K = \operatorname{int}(\operatorname{dom} s)$ and s'' is not bounded

above on K. In neither of these situations are the hypotheses of Theorem 4.2 satisfied.

In Theorem 4.3 other hypotheses are given guaranteeing that for each $u \in K$ there exists γ such that s_γ has a strictly supporting line at u; in general, γ depends on u. However, with the same γ, s_γ might also have a strictly supporting line at other values of u. In general, as one increases γ, the set of u at which s_γ has a strictly supporting line cannot decrease. Because of part (a) of Theorem 3.3, this can be restated in terms of ensemble equivalence involving the set $\mathcal{E}_{\beta,\gamma}$ of Gaussian equilibrium macrostates. Defining

$$F_\gamma = \{u \in K : \text{there exists } \beta \text{ such that } \mathcal{E}_{\beta,\gamma} = \mathcal{E}^u\},$$

we have $F_{\gamma_1} \subset F_{\gamma_2}$ whenever $\gamma_2 > \gamma_1$ and because of Theorem 4.3, $\bigcup_{\gamma>0} F_\gamma = K$. This phenomenon is investigated in Section 5 for the Curie–Weiss–Potts model.

In order to state Theorem 4.3, we define for $u \in K$ and $\lambda \geq 0$

$$D(u, s'(u), \lambda) = \left\{v \in \operatorname{dom} s : s(v) \geq s(u) + s'(u)(v-u) + \lambda(v-u)^2\right\}.$$

Geometrically, this set contains all points for which the parabola with parameters $(s'(u), \lambda)$ passing through $(u, s(u))$ lies below the graph of s. Clearly, since $\lambda \geq 0$, we have $D(u, s'(u), \lambda) \subset D(u, s'(u), 0)$; the set $D(u, s'(u), 0)$ contains all points for which the graph of the line with slope $s'(u)$ passing through $(u, s(u))$ lies below the graph of s. Thus, in the next theorem the hypothesis that for each $u \in K$ the set $D(u, s'(u), \lambda)$ is bounded for some $\lambda \geq 0$ is satisfied if $\operatorname{dom} s$ is bounded or, more generally, if $D(u, s'(u), 0)$ is bounded. The latter set is bounded if, for example, $-s$ is superlinear; i.e.,

$$\lim_{|v|\to\infty} s(v)/|v| = -\infty.$$

The quantity $\gamma_0(u)$ appearing in the next theorem is defined in equation (5.7) in [Costeniuc et al. 2005b].

THEOREM 4.3. *Let K an open subset of* $\operatorname{dom} s$ *and assume that s is twice continuously differentiable on K. Assume also that* $\operatorname{dom} s$ *is bounded or, more generally, that for every $u \in \operatorname{int} K$ there exists $\lambda \geq 0$ such that $D(u, s'(u), \lambda)$ is bounded. Then for each $u \in K$ there exists $\gamma_0(u) \geq 0$ with the following properties.*

(a) *For each $u \in K$ and any $\gamma > \gamma_0(u)$, s has a strictly supporting parabola at u with parameters $(s'(u), \gamma)$.*

(b) *For each $u \in K$ and any $\gamma > \gamma_0(u)$, $s_\gamma = s - \gamma(\cdot)^2$ has a strictly supporting line at u with slope $s'(u) - 2\gamma u$.*

(c) *For each $u \in K$ and any $\gamma > \gamma_0(u)$, the microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are fully equivalent at u. The value of β defining the Gaussian ensemble is unique and is given by $\beta = s'(u) - 2\gamma u$.*

COMMENTS ON THE PROOF. (a) We first choose a parabola that is strictly supporting in a neighborhood of u and then adjust γ so that the parabola becomes strictly supporting on all $\mathbb{R}$. Proposition 4.1 guarantees that $s - \gamma(\cdot)^2$ has a strictly supporting line at u. Details are given in [Costeniuc et al. 2005b, pp. 1319–1321].

(b) This follows from part (a) of the present theorem and Proposition 4.1.

(c) For $u \in K$ the full equivalence of the ensembles follows from part (b) of the present theorem and part (a) of Theorem 3.3. The value of β defining the fully equivalent Gaussian ensemble is determined by a routine argument given in [Costeniuc et al. 2005b, p. 1321]. □

Theorem 4.3 suggests an extended form of the notion of universal equivalence of ensembles. In Theorem 4.2 we are able to achieve full equivalence of ensembles for all $u \in \operatorname{dom} s$, except possibly boundary points, by choosing an appropriate γ that is valid for all u. This leads to the observation that the microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are universally equivalent. In Theorem 4.3 we can also achieve full equivalence of ensembles for all $u \in K$. However, in contrast to Theorem 4.2, the choice of γ for which the two ensembles are fully equivalent depends on u. We summarize the ensemble equivalence property articulated in part (c) of Theorem 4.3 by saying that relative to the set of quadratic functions, the microcanonical and Gaussian ensembles are universally equivalent on the open set K of energy values.

We complete our discussion of the generalized canonical ensemble and its equivalence with the microcanonical ensemble by noting that the smoothness hypothesis on s in Theorem 4.3 is essentially satisfied whenever the microcanonical ensemble exhibits no phase transition at any $u \in K$. In order to see this, we recall that a point u_c at which s is not differentiable represents a first-order, microcanonical phase transition [Ellis et al. 2004b, Figure 3]. In addition, a point u_c at which s is differentiable but not twice differentiable represents a second-order, microcanonical phase transition [Ellis et al. 2004b, Figure 4]. It follows that s is smooth on any open set K not containing such phase-transition points. Hence, if the other hypotheses in Theorem 4.3 are valid, then the microcanonical and Gaussian ensembles are universally equivalent on K relative to the set of quadratic functions. In particular, if the microcanonical ensemble exhibits no phase transitions, then s is smooth on all of $\operatorname{int}(\operatorname{dom} s)$. This implies the universal equivalence of the two ensembles provided that the other conditions are valid in Theorem 4.2.

In the next section we apply the results in this paper to the Curie–Weiss–Potts model.

5. Applications to the Curie–Weiss–Potts model

The Curie–Weiss–Potts model is a mean-field approximation to the nearest-neighbor Potts model, which takes its place next to the Ising model as one of the most versatile models in equilibrium statistical mechanics [Wu 1982]. Although the Curie–Weiss–Potts model is considerably simpler to analyze, it is an excellent model to illustrate the general theory presented in this paper, lying at the boundary of the set of models for which a complete analysis involving explicit formulas is available. As we will see, there exists an interval N such that for any $u \in N$ the microcanonical ensemble is nonequivalent to the canonical ensemble. The main result, stated in Theorem 5.2, is that for any $u \in N$ there exists $\gamma \geq 0$ such that the microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are fully equivalent for all $v \leq u$. While not as strong as universal equivalence, the ensemble equivalence proved in Theorem 5.2 is considerably stronger than the local equivalence stated in Theorem 4.3.

Let $q \geq 3$ be a fixed integer and define $\Lambda = \{\theta^1, \theta^2, \dots, \theta^q\}$, where the θ^i are any q distinct vectors in $\mathbb{R}^q$. In the definition of the Curie–Weiss–Potts model, the precise values of these vectors is immaterial. For each $n \in \mathbb{N}$ the model is defined by spin random variables $\omega_1, \omega_2, \dots, \omega_n$ that take values in Λ. The ensembles for the model are defined in terms of probability measures on the configuration spaces Λ^n, which consist of the microstates $\omega = (\omega_1, \omega_2, \dots, \omega_n)$. We also introduce the n-fold product measure P_n on Λ^n with identical one-dimensional marginals

$$\bar{\rho} = \frac{1}{q} \sum_{i=1}^{q} \delta_{\theta^i}.$$

Thus for all $\omega \in \Lambda^n$, $P_n(\omega) = \frac{1}{q^n}$. For $n \in \mathbb{N}$ and $\omega \in \Lambda^n$ the Hamiltonian for the q-state Curie–Weiss–Potts model is defined by

$$H_n(\omega) = -\frac{1}{2n} \sum_{j,k=1}^{n} \delta(\omega_j, \omega_k),$$

where $\delta(\omega_j, \omega_k)$ equals 1 if $\omega_j = \omega_k$ and equals 0 otherwise. The energy per particle is defined by $h_n(\omega) = \frac{1}{n} H_n(\omega)$.

With this choice of h_n and with $a_n = n$, the microcanonical, canonical, and Gaussian ensembles for the model are the probability measures on Λ^n defined as in (2.3), (2.4), and (2.6). The key to our analysis of the Curie–Weiss–Potts

model is to express h_n in terms of the macroscopic variables

$$L_n = L_n(\omega) = (L_{n,1}(\omega), L_{n,2}(\omega), \dots, L_{n,q}(\omega)),$$

the ith component of which is defined by

$$L_{n,i}(\omega) = \frac{1}{n}\sum_{j=1}^{n} \delta(\omega_j, \theta^i).$$

This quantity equals the relative frequency with which ω_j, $j \in \{1, \dots, n\}$, equals θ^i. The empirical vectors L_n take values in the set of probability vectors

$$\mathcal{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q),\ \text{each } \nu_i \geq 0, \sum_{i=1}^{q} \nu_i = 1 \right\}.$$

Each probability vector in $\mathcal{P}$ represents a possible equilibrium macrostate for the model.

There is a one-to-one correspondence between $\mathcal{P}$ and the set $\mathcal{P}(\Lambda)$ of probability measures on Λ, $\nu \in \mathcal{P}$ corresponding to the probability measure $\sum_{i=1}^{q} \nu_i \delta_{\theta^i}$. The element $\rho \in \mathcal{P}$ corresponding to the one-dimensional marginal $\bar{\rho}$ of the prior measures P_n is the uniform vector having equal components $\frac{1}{q}$. For $\omega \in \Lambda^n$ the element of $\mathcal{P}$ corresponding to the empirical vector $L_n(\omega)$ is the empirical measure of the spin random variables $\omega_1, \omega_2, \dots, \omega_n$.

We denote by $\langle \cdot, \cdot \rangle$ the inner product on $\mathbb{R}^q$. Since

$$\sum_{i=1}^{q}\sum_{j=1}^{n} \delta(\omega_j, \xi^i) \cdot \sum_{k=1}^{n} \delta(\omega_k, \xi^i) = \sum_{j,k=1}^{n} \delta(\omega_j, \omega_k),$$

it follows that the energy per particle can be rewritten as

$$h_n(\omega) = -\frac{1}{2n^2} \sum_{j,k=1}^{n} \delta(\omega_j, \omega_k) = -\tfrac{1}{2}\langle L_n(\omega), L_n(\omega)\rangle,$$

i.e.,

$$h_n(\omega) = \tilde{H}(L_n(\omega)),\ \text{where } \tilde{H}(\nu) = -\tfrac{1}{2}\langle \nu, \nu \rangle \text{ for } \nu \in \mathcal{P}.$$

$\tilde{H}$ is the energy representation function for the model.

In order to define the sets of equilibrium macrostates with respect to the three ensembles, we appeal to Sanov's Theorem. This states that with respect to the product measures P_n, the empirical vectors L_n satisfy the LDP on $\mathcal{P}$ with rate function given by the relative entropy $R(\cdot \mid \rho)$ [Ellis 1985, Theorem VIII.2.1]. For $\nu \in \mathcal{P}$ this is defined by

$$R(\nu|\rho) = \sum_{i=1}^{q} \nu_i \log(q\nu_i).$$

With the choices $I = R(\cdot\,|\,\rho)$, $\tilde{H} = -\frac{1}{2}\langle\cdot,\cdot\rangle$, and $a_n = n$, L_n satisfies the LDP on $\mathcal{P}$ with respect to each of the three ensembles with the rate functions given by (2.7), (2.10), and (2.13). In turn, the corresponding sets of equilibrium macrostates are given by

$$\mathcal{E}^u = \big\{\nu \in \mathcal{P} : R(\nu|\rho) \text{ is minimized subject to } \tilde{H}(\nu) = u\big\},$$
$$\mathcal{E}_\beta = \big\{\nu \in \mathcal{P} : R(\nu|\rho) + \beta\tilde{H}(\nu) \text{ is minimized }\big\},$$
$$\mathcal{E}_{\beta,\gamma} = \big\{\nu \in \mathcal{P} : R(\nu|\rho) + \beta\tilde{H}(\nu) + \gamma[\tilde{H}(\nu)]^2 \text{ is minimized }\big\}.$$

Each element ν in $\mathcal{E}^u$, $\mathcal{E}_\beta$, and $\mathcal{E}_{\beta,\gamma}$ describes an equilibrium configuration of the model with respect to the corresponding ensemble in the thermodynamic limit. The ith component ν_i gives the asymptotic relative frequency of spins taking the value θ^i.

As in (2.2), the microcanonical entropy is defined by

$$s(u) = -\inf\{R(\nu|\rho) \,:\, \nu \in \mathcal{P}, \tilde{H}(\nu) = u\}.$$

Since $R(\nu|\rho) < \infty$ for all $\nu \in \mathcal{P}$, $\operatorname{dom} s$ equals the range of $\tilde{H}(\nu) = -\frac{1}{2}\langle\nu,\nu\rangle$ on $\mathcal{P}$, which is the closed interval $[-\frac{1}{2}, -\frac{1}{2q}]$. The set $\mathcal{E}^u$ of microcanonical equilibrium macrostates is nonempty precisely for $u \in \operatorname{dom} s$. For $q = 3$, the microcanonical entropy can be determined explicitly. For all $q \geq 4$ the microcanonical entropy can also be determined explicitly provided Conjecture 4.1 in [Costeniuc et al. 2005a] is valid; this conjecture has been verified numerically for all $q \in \{4, 5, \ldots, 10^4\}$. The formulas for the microcanonical entropy are given in Theorem 4.3 in [Costeniuc et al. 2005a].

We first consider the relationships between $\mathcal{E}^u$ and $\mathcal{E}_\beta$, which according to Theorem 3.1 are determined by support properties of s. These properties can be seen in Figure 1. The quantity u_0 appearing in this figure equals $[-q^2 + 3q - 3]/[2q(q-1)]$ [Costeniuc et al. 2005a, Lem. 6.1]. Figure 1 is not the actual graph of s but a schematic graph that accentuates the shape of the graph of s together with the intervals of strict concavity and nonconcavity of this function.

These and other details of the graph of s are also crucial in analyzing the relationships between $\mathcal{E}^u$ and $\mathcal{E}_{\beta,\gamma}$. Denote $\operatorname{dom} s$ by $[u_\ell, u_r]$, where $u_\ell = -\frac{1}{2}$ and $u_r = -\frac{1}{2q}$. These details include the observation that there exists $w_0 \in (u_0, u_r)$ such that s is a concave-convex function with break point w_0; i.e., the restriction of s to (u_ℓ, w_0) is strictly concave and the restriction of s to (w_0, u_r) is strictly convex. A difficulty in validating this observation is that for certain values of q, including $q = 3$, the intervals of strict concavity and strict convexity are shallow and therefore difficult to discern. Furthermore, what seem to be strictly concave and strictly convex portions of this function on the scale of the entire graph might reveal themselves to be much less regular on a finer scale. Conjecture 5.1 gives a set of properties of s implying there exists $w_0 \in (u_0, u_r)$

such that s is a concave-convex function with break point w_0. In particular, this property of s guarantees that s has the support properties stated in the three items appearing in the next paragraph. Conjecture 5.1 has been verified numerically for all $q \in \{4, 5, \ldots, 10^4\}$.

We define the sets

$$F = (u_\ell, u_0) \cup \{u_r\},\ P = \{u_0\},\ and\ N = (u_0, u_r).$$

Figure 1 and Theorem 3.1 then show that these sets are respectively the sets of full equivalence, partial equivalence, and nonequivalence of the microcanonical and canonical ensembles. The details are given in the next three items. In Theorem 6.2 in [Costeniuc et al. 2005a] all these conclusions concerning ensemble equivalence and nonequivalence are proved analytically without reference to the form of s given in Figure 1.

1. s is strictly concave on the interval (u_ℓ, u_0) and has a strictly supporting line at each $u \in (u_\ell, u_0)$ and at u_r. Hence for $u \in F = (u_\ell, u_0) \cup \{u_r\}$ the ensembles are fully equivalent in the sense that there exists β such that $\mathcal{E}^u = \mathcal{E}_\beta$ [Theorem 3.1(a)].
2. s is concave but not strictly concave at u_0 and has a nonstrictly supporting line at u_0 that also touches the graph of s over the right hand endpoint u_r. Hence for $u \in P = \{u_0\}$ the ensembles are partially equivalent in the sense that there exists β such that $\mathcal{E}^u \subset \mathcal{E}_\beta$ but $\mathcal{E}^u \neq \mathcal{E}_\beta$ [Theorem 3.1(b)].

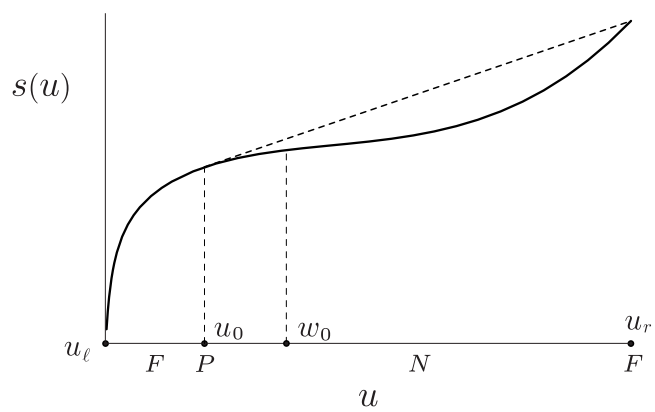

Figure 1. Schematic graph of $s(u)$, showing the set $F = (u_\ell, u_0) \cup \{u_r\}$ of full ensemble equivalence, the singleton set $P = \{u_0\}$ of partial equivalence, and the set $N = (u_0, u_r)$ of nonequivalence, where $u_\ell = -\frac{1}{2}$ and $u_r = -\frac{1}{2q}$. For $u \in F \cup P = (u_\ell, u_0] \cup \{u_r\}$, $s(u) = s^{**}(u)$; for $u \in N$, $s(u) < s^{**}(u)$ and the graph of s^{**} consists of the dotted line segment with slope β_c. The slope of s at u_ℓ is ∞. The quantity w_0 is discussed after Conjecture 5.1.

3. s is not concave on $N = (u_0, u_r)$ and has no supporting line at any $u \in N$. Hence for $u \in N$ the ensembles are nonequivalent in the sense that for all β, $\mathcal{E}^u \cap \mathcal{E}_\beta = \varnothing$ [Theorem 3.1(c)].

The explicit calculation of the elements of $\mathcal{E}_\beta$ and $\mathcal{E}^u$ given in [Costeniuc et al. 2005a] shows different continuity properties of these two sets. $\mathcal{E}_\beta$ undergoes a discontinuous phase transition as β increases through the critical inverse temperature $\beta_c = \frac{2(q-1)}{q-2}\log(q-1)$, the unique macrostate ρ for $\beta < \beta_c$ bifurcating discontinuously into the q distinct macrostates for $\beta > \beta_c$. By contrast, $\mathcal{E}^u$ undergoes a continuous phase transition as u decreases from the maximum value $u_r = -\frac{1}{2q}$, the unique macrostate ρ for $u = u_r$ bifurcating continuously into the q distinct macrostates for $u < u_r$. The different continuity properties of these phase transitions shows already that the canonical and microcanonical ensembles are nonequivalent.

For u in the interval N of ensemble nonequivalence, the graph of s^{**} is affine; this is depicted by the dotted line segment in Figure 1. One can show that the slope of the affine portion of the graph of s^{**} equals the critical inverse temperature β_c.

This completes the discussion of the equivalence and nonequivalence of the microcanonical and canonical ensembles. The equivalence and nonequivalence of the microcanonical and Gaussian ensembles depends on the relationships between the sets $\mathcal{E}^u$ and $\mathcal{E}_{\beta,\gamma}$ of corresponding equilibrium macrostates, which in turn are determined by support properties of the generalized microcanonical entropy $s_\gamma(u) = s(u) - \gamma u^2$. As we just saw, for each $u \in N = (u_0, u_r)$, the microcanonical and canonical ensembles are nonequivalent. For $u \in N$ we would like to recover equivalence by replacing the canonical ensemble by an appropriate Gaussian ensemble.

Theorem 4.2 is not applicable. Although the first three of the hypotheses are valid, unfortunately s'' is not bounded above on the interior of $\operatorname{dom} s$. Indeed, using the explicit formula for s given in Theorem 4.3 in [Costeniuc et al. 2005a], one verifies that $\lim_{u\to(u_r)^-} s''(u) = \infty$. However, we can appeal to Theorem 4.3 in the present paper, which is applicable since s is twice continuously differentiable on N. We conclude that for each $u \in N$ and all sufficiently large γ there exists a corresponding Gaussian ensemble that is equivalent to the microcanonical ensemble for that u.

By using other conjectured properties of the microcanonical entropy, we are able to deduce the stronger result on the equivalence of the microcanonical and Gaussian ensembles stated in Theorem 5.2. As before, we denote $\operatorname{dom} s$ by $[u_\ell, u_r]$, where $u_\ell = -\frac{1}{2}$ and $u_r = -\frac{1}{2q}$, and write

$$s'(u_\ell) = \lim_{u\to(u_\ell)^+} s'(u) \quad\text{and}\quad s'(u_r) = \lim_{u\to(u_r)^-} s'(u)$$

with a similar notation for $s''(u_\ell)$ and $s''(u_r)$. Using the explicit but complicated formula for s given in Theorem 4.2 in [Costeniuc et al. 2005a], the following conjecture was verified numerically for all $q \in \{4, 5, \dots, 10^4\}$ and all $u \in (u_\ell, u_r)$ of the form $u = u_\ell + 0.02k$, where k is a positive integer.

CONJECTURE 5.1. *For all $q \geq 3$ the microcanonical entropy s has the following two properties.*

(a) $s'''(u) > 0$ *for all* $u \in (u_\ell, u_r)$.

(b) $s'(u_\ell) = \infty,\ 0 < s'(u_r) < \infty,\ s''(u_\ell) = -\infty,$ *and* $s''(u_r) = \infty$.

The conjecture implies that s'' is an increasing bijection of (u_ℓ, u_r) onto $\mathbb{R}$. Therefore, there exists a unique point $w_0 \in (u_\ell, u_r)$ such that $s''(u) < 0$ for all $u \in (u_\ell, w_0)$, $s''(w_0) = 0$, and $s''(u) > 0$ for all $u \in (w_0, u_r)$. It follows that the restriction of s to $[u_\ell, w_0]$ is strictly concave and the restriction of s to $[w_0, u_r]$ is strictly convex. These properties, which can be seen in Figure 1, are summarized by saying that s is a concave-convex function with break point w_0.

The interval $N = (u_0, u_r)$ exhibited in Figure 1 contains all energy values u for which there exists no canonical ensemble that is equivalent with the microcanonical ensemble. Assuming the truth of Conjecture 5.1, we now show that for each $u \in N$ there exists $\gamma \geq 0$ and an associated Gaussian ensemble that is equivalent with the microcanonical ensemble for all $v \leq u$. In order to do this, for $\gamma \geq 0$ we bring in the generalized microcanonical entropy

$$s_\gamma(u) = s(u) - \gamma u^2$$

and note that the properties of s stated in Conjecture 5.1 are invariant under the addition of the quadratic $-\gamma u^2$. Hence, if Conjecture 5.1 is valid, then s_γ satisfies the same properties as s. In particular, s_γ must be a concave-convex function with some break point w_γ, which is the unique point in (u_ℓ, u_r) such that $s''_\gamma(u) < 0$ for all $u \in (u_\ell, w_\gamma)$, $s''_\gamma(w_\gamma) = 0$, and $s''_\gamma(u) > 0$ for all $u \in (w_\gamma, u_r)$. A straightforward argument, which we omit, and an appeal to Theorem 3.3 show that there exists a unique point $u_\gamma \in (u_\ell, w_\gamma)$ having the properties listed in the next three items. These properties show that u_γ plays the same role for ensemble equivalence involving the Gaussian ensemble that the point u_0 plays for ensemble equivalence involving the canonical ensemble.

1. For $\gamma \geq 0$, s_γ is strictly concave on the interval (u_ℓ, u_γ) and has a strictly supporting line at each $u \in (u_\ell, u_\gamma)$ and at u_r. Hence for $u \in F_\gamma = (u_\ell, u_\gamma) \cup \{u_r\}$ the ensembles are fully equivalent in the sense that there exists β such that $\mathcal{E}^u = \mathcal{E}_{\beta,\gamma}$ [Theorem 3.3(a)].
2. For $\gamma \geq 0$, s_γ is concave but not strictly concave at u_γ and has a nonstrictly supporting line at u_γ that also touches the graph of s over the right hand endpoint u_r. Hence for $u \in P_\gamma = \{u_\gamma\}$ the ensembles are partially equivalent

in the sense that there exists β such that $\mathcal{E}^u \subset \mathcal{E}_{\beta,\gamma}$ but $\mathcal{E}^u \neq \mathcal{E}_{\beta,\gamma}$ [Theorem 3.3(b)].

3. For $\gamma \geq 0$, s_γ is not concave on the interval $N = (u_\gamma, u_r)$ and has no supporting line at any $u \in N$. Hence for $u \in N_\gamma$ the ensembles are nonequivalent in the sense that for all β, $\mathcal{E}^u \cap \mathcal{E}_{\beta,\gamma} = \varnothing$ [Theorem 3.3(c)].

We now state our main result.

THEOREM 5.2. *We assume that Conjecture* 5.1 *is valid. Then as a function of* $\gamma \geq 0$, $F_\gamma = (u_\ell, u_\gamma) \cup \{u_r\}$ *is strictly increasing, and as* $\gamma \to \infty$, $F_\gamma \uparrow (u_\ell, u_r]$. *It follows that for any* $u \in N = (u_0, u_r)$, *there exists* $\gamma \geq 0$ *such that the microcanonical ensemble and the Gaussian ensemble defined in terms of this* γ *are fully equivalent for all* $v \in (u_\ell, u_r)$ *satisfying* $v \leq u$. *The value of* β *defining the Gaussian ensemble is unique and is given by* $\beta = s'(v) - 2\gamma v$.

The proof of the theorem relies on the next lemma, part (a) of which uses Proposition 4.1. When applied to s_γ, this proposition states that s_γ has a strictly supporting line at a point if and only if s has a strictly supporting parabola at that point. Proposition 4.1 illustrates why one can achieve full equivalence with the Gaussian ensemble when full equivalence with the canonical ensemble fails. Namely, even when s does not have a supporting line at a point, it might have a supporting parabola at that point; in this case the supporting parabola can be made strictly supporting by increasing γ. The proofs of parts (b)–(d) of the next lemma rely on Theorem 4.3 and on the properties of the sets F_γ, P_γ, and N_γ stated in the three items appearing just before Theorem 5.2.

LEMMA 5.3. *We assume that Conjecture* 5.1 *is valid. Then*:

(a) *If for some* $\gamma \geq 0$, s_γ *has a supporting line at a point* u, *then for any* $\tilde{\gamma} > \gamma$, $s_{\tilde{\gamma}}$ *has a strictly supporting line at* u.

(b) *For any* $0 \leq \gamma < \tilde{\gamma}$, $F_\gamma \cup P_\gamma \subset F_{\tilde{\gamma}}$.

(c) u_γ *is a strictly increasing function of* $\gamma \geq 0$ *and* $\lim_{\gamma\to\infty} u_\gamma = u_r$.

(d) *As a function of* $\gamma \geq 0$, F_γ *is strictly increasing*.

PROOF. (a) Suppose that s_γ has a supporting line at u with slope $\bar{\beta}$. Then by Proposition 4.1 s has a supporting parabola at u with parameters (β, γ), where $\beta = \bar{\beta} + 2\gamma u$. As the definition (4.1) makes clear, replacing γ by any $\tilde{\gamma} > \gamma$ makes the supporting parabola at u strictly supporting. Again by Proposition 4.1 $s_{\tilde{\gamma}}$ has a strictly supporting line at u.

(b) If $u \in F_\gamma \cup P_\gamma$, then s_γ has a supporting line at u. Since $0 \leq \gamma < \tilde{\gamma}$, part (a) implies that $s_{\tilde{\gamma}}$ has a strictly supporting line at u. Hence u must be an element of $F_{\tilde{\gamma}}$.

(c) If $0 \leq \gamma < \tilde{\gamma}$, then by part (a) of the present lemma $u_\gamma \in P_\gamma \subset F_{\tilde{\gamma}}$. Since $F_{\tilde{\gamma}} = (u_\ell, u_{\tilde{\gamma}}) \cup \{u_r\}$ and since $u_\gamma < u_r$, it follows that $u_\gamma < u_{\tilde{\gamma}}$. Thus u_γ is

a strictly increasing function of $\gamma \geq 0$. We now prove that $\lim_{\gamma\to\infty} u_\gamma = u_r$. For any $u \in (u_\ell, u_r)$, part (b) of Theorem 4.3 states that there exists $\gamma_u > 0$ such that $s_{\gamma_u}(u)$ has a strictly supporting line at u. It follows that $u \in F_{\gamma_u} = (u_\ell, u_{\gamma_u}) \cup \{u_r\}$ and thus that $u < u_{\gamma_u} < u_r$. Since u_γ is a strictly increasing function of γ, it follows that for all $\gamma > \gamma_u$, we have $u_\gamma > u_{\gamma_u} > u$. We have shown that for any $u \in (u_\ell, u_r)$, there exists $\gamma_u > 0$ such that for all $\gamma > \gamma_u$, we have $u_\gamma > u$. This completes the proof that $\lim_{\gamma\to\infty} u_\gamma = u_r$.

(d) Since $F_\gamma = (u_\ell, u_\gamma) \cup \{u_r\}$, this follows immediately from the first property of u_γ in part (c). The proof of the lemma is complete. □

We are now ready to prove Theorem 5.2. The properties of F_γ stated there follow immediately from Lemma 5.3. Indeed, since u_γ is a strictly increasing function of $\gamma \geq 0$, F_γ is also strictly increasing. In addition, since $\lim_{\gamma\to\infty} u_\gamma = u_r$ it follows that as $\gamma \to \infty$, $F_\gamma \uparrow (u_\ell, u_r]$. Since F_γ is the set of full ensemble equivalence, we conclude that for any $u \in N = (u_0, u_r)$, there exists $\gamma > 0$ such that the microcanonical ensemble and the Gaussian ensemble defined in terms of this γ are fully equivalent for all $v \in (u_\ell, u_r)$ satisfying $v \leq u$. The last statement concerning β is a consequence of part (c) of Theorem 4.3. The proof of Theorem 5.2 is complete.

Acknowledgements

The research of Marius Costeniuc was supported by a grant from the National Science Foundation (DMS-0202309), the research of Richard S. Ellis was supported by grants from the National Science Foundation (DMS-0202309 and DMS-0604071), the research of Bruce Turkington was supported by grants from the National Science Foundation (DMS-0207064 and DMS-0604071), and the research of Hugo Touchette was supported by the Natural Sciences and Engineering Research Council of Canada and the Royal Society of London (Canada–UK Millennium Fellowship).

References

[Barré et al. 2001] J. Barré, D. Mukamel, and S. Ruffo, "Inequivalence of ensembles in a system with long-range interactions", *Phys. Rev. Lett.* **87** (2001), #030601.

[Barré et al. 2002] J. Barré, D. Mukamel, and S. Ruffo, "Ensemble inequivalence in mean-field models of magnetism", pp. 45–67 in *Dynamics and thermodynamics of systems with long-range interactions* (Les Houches, 2002), edited by T. Dauxois et al., Lecture Notes in Phys. **602**, Springer, Berlin, 2002.

[Bertsekas 1982] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, New York, 1982.

[Blume et al. 1971] M. Blume, V. J. Emery, and R. B. Griffiths, "Ising model for the λ transition and phase separation in ^{3}He-^{4}He mixtures", *Phys. Rev. A* **4** (1971), 1071–1077.

[Borges and Tsallis 2002] E. P. Borges and C. Tsallis, "Negative specific heat in a Lennard–Jones-like gas with long-range interactions", *Physica A* **305** (2002), 148–151.

[Boucher et al. 2000] C. Boucher, R. S. Ellis, and B. Turkington, "Derivation of maximum entropy principles in two-dimensional turbulence via large deviations", *J. Statist. Phys.* **98**:5-6 (2000), 1235–1278.

[Challa and Hetherington 1988a] M. S. S. Challa and J. H. Hetherington, "Gaussian ensemble: an alternate Monte-Carlo scheme", *Phys. Rev. A* **38** (1988), 6324–6337.

[Challa and Hetherington 1988b] M. S. S. Challa and J. H. Hetherington, "Gaussian ensemble as an interpolating ensemble", *Phys. Rev. Lett.* **60** (1988), 77–80.

[Costeniuc et al. 2005a] M. Costeniuc, R. S. Ellis, and H. Touchette, "Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model", *J. Math. Phys.* **46**:6 (2005), #063301.

[Costeniuc et al. 2005b] M. Costeniuc, R. S. Ellis, H. Touchette, and B. Turkington, "The generalized canonical ensemble and its universal equivalence with the microcanonical ensemble", *J. Statist. Phys.* **119**:5-6 (2005), 1283–1329.

[Costeniuc et al. 2006a] M. Costeniuc, R. S. Ellis, and H. Touchette, "Nonconcave entropies from generalized canonical ensembles", *Phys. Rev. E* **74**:1 (2006), #010105.

[Costeniuc et al. 2006b] M. Costeniuc, R. S. Ellis, H. Touchette, and B. Turkington, "Generalized canonical ensembles and ensemble equivalence", *Phys. Rev. E* **73**:2 (2006), #026105.

[Dauxois et al. 2000] T. Dauxois, P. Holdsworth, and S. Ruffo, "Violation of ensemble equivalence in the antiferromagnetic mean-field XY model", *Eur. Phys. J. B* **16** (2000), 659–667.

[Dauxois et al. 2002] T. Dauxois, V. Latora, A. Rapisarda, S. Ruffo, and A. Torcini, "Dynamics and thermodynamics of systems with long-range interactions", pp. 458–487 in *Dynamics and thermodynamics of systems with long-range interactions* (Les Houches), edited by T. Dauxois et al., Lecture Notes in Physics **602**, Springer, Berlin, 2002.

[Deuschel et al. 1991] J.-D. Deuschel, D. W. Stroock, and H. Zessin, "Microcanonical distributions for lattice gases", *Comm. Math. Phys.* **139**:1 (1991), 83–101.

[Dibattista et al. 1998] M. T. Dibattista, A. J. Majda, and B. Turkington, "Prototype geophysical vortex structures via large-scale statistical theory", *Geophys. Astrophys. Fluid Dynam.* **89**:3-4 (1998), 235–283.

[DiBattista et al. 2001] M. T. DiBattista, A. J. Majda, and M. J. Grote, "Meta-stability of equilibrium statistical structures for prototype geophysical flows with damping and driving", *Physica D* **151**:2-4 (2001), 271–304.

[Ellis 1985] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, Grundlehren der math. Wissenschaften **271**, Springer, Berlin, 1985. Reprinted 2006.

[Ellis et al. 2000] R. S. Ellis, K. Haven, and B. Turkington, "Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles", *J. Statist. Phys.* **101**:5-6 (2000), 999–1064.

[Ellis et al. 2002a] R. S. Ellis, K. Haven, and B. Turkington, "Nonequivalent statistical equilibrium ensembles and refined stability theorems for most probable flows", *Nonlinearity* **15**:2 (2002), 239–255.

[Ellis et al. 2002b] R. S. Ellis, B. Turkington, and K. Haven, "Analysis of statistical equilibrium models of geostrophic turbulence", *J. Appl. Math. Stochastic Anal.* **15**:4 (2002), 341–361.

[Ellis et al. 2004a] R. S. Ellis, R. Jordan, P. Otto, and B. Turkington, "A statistical approach to the asymptotic behavior of a class of generalized nonlinear Schrödinger equations", *Comm. Math. Phys.* **244**:1 (2004), 187–208.

[Ellis et al. 2004b] R. S. Ellis, H. Touchette, and B. Turkington, "Thermodynamic versus statistical nonequivalence of ensembles for the mean-field Blume–Emery–Griffiths model", *Physica A* **335**:3-4 (2004), 518–538.

[Ellis et al. 2005] R. S. Ellis, P. T. Otto, and H. Touchette, "Analysis of phase transitions in the mean-field Blume–Emery–Griffiths model", *Ann. Appl. Probab.* **15**:3 (2005), 2203–2254.

[Eyink and Spohn 1993] G. L. Eyink and H. Spohn, "Negative-temperature states and large-scale, long-lived vortices in two-dimensional turbulence", *J. Statist. Phys.* **70**:3-4 (1993), 833–886.

[Föllmer and Orey 1988] H. Föllmer and S. Orey, "Large deviations for the empirical field of a Gibbs measure", *Ann. Probab.* **16**:3 (1988), 961–977.

[Georgii 1993] H.-O. Georgii, "Large deviations and maximum entropy principle for interacting random fields on $\mathbf{Z}^d$", *Ann. Probab.* **21**:4 (1993), 1845–1875.

[Gibbs 1902] J. W. Gibbs, *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics*, Yale University Press, New Haven, 1902. Reprinted Dover, New York, 1960.

[Gross 1997] D. H. E. Gross, "Microcanonical thermodynamics and statistical fragmentation of dissipative systems: the topological structure of the n-body phase space", *Phys. Rep.* **279** (1997), 119–202.

[Hertel and Thirring 1971] P. Hertel and W. Thirring, "A soluble model for a system with negative specific heat", *Ann. Phys. (NY)* **63** (1971), 520.

[Hetherington 1987] J. H. Hetherington, "Solid ^{3}He magnetism in the classical approximation", *J. Low Temp. Phys.* **66** (1987), 145–154.

[Hetherington and Stump 1987] J. H. Hetherington and D. R. Stump, "Sampling a Gaussian energy distribution to study phase transitions of the $Z(2)$ and $U(1)$ lattice gauge theories", *Phys. Rev. D* **35** (1987), 1972–1978.

[Johal et al. 2003] R. S. Johal, A. Planes, and E. Vives, "Statistical mechanics in the extended Gaussian ensemble", *Phys. Rev. E* **68** (2003), #056113.

[Jordan et al. 2000] R. Jordan, B. Turkington, and C. L. Zirbel, "A mean-field statistical theory for the nonlinear Schrödinger equation", *Physica D* **137**:3-4 (2000), 353–378.

[Kiessling and Lebowitz 1997] M. K.-H. Kiessling and J. L. Lebowitz, "The microcanonical point vortex ensemble: beyond equivalence", *Lett. Math. Phys.* **42** (1997), 43–58.

[Kiessling and Neukirch 2003] M. K.-H. Kiessling and T. Neukirch, "Negative specific heat of a magnetically self-confined plasma torus", *Proc. Natl. Acad. Sci. USA* **100**:4 (2003), 1510–1514.

[Kiessling and Percus 1995] M. K.-H. Kiessling and J. K. Percus, "Nonuniform van der Waals theory", *J. Statist. Phys.* **78**:5-6 (1995), 1337–1376.

[Lanford 1973] O. E. Lanford, "Entropy and equilibrium states in classical statistical mechanics", pp. 1–113 in *Statistical mechanics and mathematical problems: Battelle Rencontres* (Seattle, WA, 1971), edited by A. Lenard, Lecture Notes in Physics **20**, Springer, Berlin, 1973.

[Lebowitz et al. 1989] J. L. Lebowitz, H. A. Rose, and E. R. Speer, "Statistical mechanics of the nonlinear Schrödinger equation. II. Mean field approximation", *J. Statist. Phys.* **54**:1-2 (1989), 17–56.

[Lewis et al. 1994] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, "The equivalence of ensembles for lattice systems: some examples and a counterexample", *J. Statist. Phys.* **77**:1-2 (1994), 397–419.

[Lewis et al. 1995] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, "Entropy, concentration of probability and conditional limit theorems", *Markov Process. Related Fields* **1**:3 (1995), 319–386.

[Lynden-Bell and Wood 1968] D. Lynden-Bell and R. Wood, "The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems", *Mon. Notic. Roy. Astron. Soc.* **138** (1968), 495.

[Miller 1990] J. Miller, "Statistical mechanics of Euler equations in two dimensions", *Phys. Rev. Lett.* **65**:17 (1990), 2137–2140.

[Minoux 1986] M. Minoux, *Mathematical programming: theory and algorithms*, Wiley, Chichester, 1986.

[Olla 1988] S. Olla, "Large deviations for Gibbs random fields", *Probab. Theory Related Fields* **77**:3 (1988), 343–357.

[Robert 1991] R. Robert, "A maximum-entropy principle for two-dimensional perfect fluid dynamics", *J. Statist. Phys.* **65**:3-4 (1991), 531–553.

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

[Rockafellar 1970] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series **28**, Princeton Univ. Press, Princeton, NJ, 1970.

[Roelly and Zessin 1993] S. Roelly and H. Zessin, "The equivalence of equilibrium principles in statistical mechanics and some applications to large particle systems", *Exposition. Math.* **11**:5 (1993), 385–405.

[Smith and O'Neil 1990] R. A. Smith and T. M. O'Neil, "Nonaxisymmetric thermal equilibria of a cylindrically bounded guiding center plasma or discrete vortex system", *Phys. Fluids B* **2** (1990), 2961–2975.

[Stump and Hetherington 1987] D. R. Stump and J. H. Hetherington, "Remarks on the use of a microcanonical ensemble to study phase transitions in the lattice gauge theory", *Phys. Lett. B* **188** (1987), 359–363.

[Thirring 1970] W. Thirring, "Systems with negative specific heat", *Z. Physik* **235** (1970), 339–352.

[Touchette et al. 2004] H. Touchette, R. S. Ellis, and B. Turkington, "An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles", *Physica A* **340**:1-3 (2004), 138–146.

[Touchette et al. 2006] H. Touchette, M. Costeniuc, R. S. Ellis, , and B. Turkington, "Metastability within the generalized canonical ensemble", *Physica A* **365** (2006), 132–137.

[Turkington et al. 2001] B. Turkington, A. Majda, K. Haven, and M. DiBattista, "Statistical equilibrium predictions of jets and spots on Jupiter", *Proc. Natl. Acad. Sci. USA* **98**:22 (2001), 12346–12350.

[Wu 1982] F. Y. Wu, "The Potts model", *Rev. Modern Phys.* **54**:1 (1982), 235–268.

MARIUS COSTENIUC
MAX PLANCK INSTITUTE FOR MATHEMATICS IN THE SCIENCES
INSELSTRASSE 22–26
D-04103 LEIPZIG
GERMANY
marius.costeniuc@gmail.com

RICHARD S. ELLIS
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MA 01003
UNITED STATES
rsellis@math.umass.edu

HUGO TOUCHETTE
SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY, UNIVERSITY OF LONDON
LONDON E1 4NS
UNITED KINGDOM
htouchet@alum.mit.edu

BRUCE TURKINGTON
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MA 01003
UNITED STATES
turk@math.umass.edu