LARGE DEVIATIONS AND THE THERMODYNAMIC FORMALISM: A NEW PROOF OF THE EQUIVALENCE OF ENSEMBLES[a]

J.T. Lewis[b], C.–E. Pfister[c], W.G. Sullivan [b],[d]

[b]Dublin Institute for Advanced Studies
10 Burlington Road
Dublin 4, Ireland
[c]Ecole Polytechnique Fédérale de Lausanne
Département de Mathématiques
CH–1015 Lausanne Switzerland
[d]University College
Department of Mathematics
Belfield, Dublin 4, Ireland

1 The Equivalence of Ensembles

In statistical mechanics the problem of the equivalence of ensembles goes back to Boltzmann and Gibbs. Here it is the problem of proving that, in the thermodynamic limit, the microcanonical measures and the grand canonical measures are equivalent; making precise the meaning of "equivalent" is part of the problem. It is commonly believed that in good statistical mechanical models such an equivalence holds, even in the presence of a phase–transition. On the other hand, it is believed that equivalence of ensembles fails in mean–field models such as the Curie–Weiss model.

There is a second statement which is also known as the equivalence of ensembles: in the thermodynamic limit, the negative of the entropy and the pressure are conjugate functions in the sense of convexity theory. In statistical mechanics, the entropy function is defined directly in the microcanonical setting and the pressure in the grand canonical setting. We refer to this statement as the equivalence of ensembles at the level of thermodynamic functions. This form of the equivalence of ensembles is known to hold for good statistical models and to fail for mean–field models. One version of our main result may be stated roughly as: *equivalence of ensembles holds at the level of measures whenever it holds at the level of thermodynamic functions.*

The problem of the equivalence of ensembles is not confined to statistical mechanics; it can be found in other areas of applied probability theory – in information theory, for example. Here the problem is to prove that a sequence of conditioned measures is

[a]Lecture delivered by J.T. Lewis

equivalent, in an appropriate sense, to a sequence of "tilted" measures. Our choice of setting is sufficiently general to cover such applications.

Probabilistic methods have been used for at least fifty years to prove results about the equivalence of ensembles: Khinchin (1943) used a local central limit theorem to prove it for a classical ideal (non–interacting) gas; Dobrushin and Tirozzi (1977) proved it for lattice gas models for which they were able to establish a local central limit theorem – a restriction which ruled–out models which exhibit first–order phase transitions. Typically, local central limit theorems hold on the scale of the square–root of the volume. The right scale for the investigation of the equivalence of ensembles, however, turns out to be that of the volume itself; this is the scale on which a large deviation principles hold. Deuschel et al. (1991) and Georgii (1993) used a large deviation principle for empirical measures to prove the equivalence of ensembles. One draw–back with this approach is that it is technically difficult: since it involves measures on a space of measures, there are subtle points to be settled. Another is that the connection with thermodynamic functions is obscured. Our approach is more elementary and direct: we go back to the common origin of large deviation theory and statistical mechanics, the Principle of the Largest Term, and prove a result about the specific information gain of a sequence of conditioned measures with respect to a sequence of tilted measures. This is a "soft" theorem – it uses nothing deeper than the order–completeness of the reals, but it has a wide applicability. For non–interacting systems, the equivalence of ensembles for measures then follows from an inequality relating the information gain $\mathcal{H}(\mu|\nu)$ of μ with respect to ν to the total variation norm $\|\cdot\|_{TV}$ of the difference of the two measures:

$$2\mathcal{H}(\mu|\nu) \geq \|\mu - \nu\|_{TV}^2. \tag{1.1}$$

For interacting systems, our "soft" theorem has to be supplemented by a "hard" theorem, proved using the combinational devices introduced in Sullivan (1973) and perfected by Preston (1976); using it, we prove the equivalence of ensembles at the level of measures for a lattice gas with translation invariant summable potentials. In order to state this result precisely, we have to describe this setting in detail; this we do in § 2. In § 3 we discuss the Principle of the Largest Term and its consequences, sketching the proof of our "soft" theorem. In § 4, we give an application to the non–interacting case. In § 5, we state precisely the general result for the lattice gas. Detailed proofs will be published elsewhere.

2 Conditioning and Tilting

Let $\{(\Omega_n, \mathcal{F}_n, \rho_n)\}_{n\geq 1}$ be a sequence of measure spaces; here ρ_n is a positive measure referred to as the *reference measure*, which may or may not be normalized. Let $V_\circ := \{V_n \in (0, \infty)\}_{n\geq 1}$ be a *scale*, a sequence of positive numbers diverging to $+\infty$ as $n \to \infty$. Typically, in the applications to statistical mechanics, V_n will be the volume of a region Λ_n in a Euclidean space $\mathbb{R}^d$ or the number of lattice sites in a box Λ_n in an integer lattice $\mathbb{Z}^d$, and Ω_n will be a configuration space associated with Λ_n. Let $T_\circ := \{T_n : \Omega_n \to X\}_{n\geq 1}$ be a sequence of random variables taking values in X, a closed convex subset of E, a locally convex topological vector space; we denote the Borel subsets of X by $\mathcal{B}(X)$ and the topological dual of E by E^*. In this exposition we will assume that X is compact and that $E = \mathbb{R}^k$ $(k \geq 1)$. These assumptions are not necessary (for the general case, see Lewis et al. (1993)) but they simplify the proofs and yet are adequate to cover the applications we make to the lattice gas.

For $C \in \mathcal{B}(X)$ such that $0 < \rho_n[T_n^{-1}C] < \infty$ for all n sufficiently large, we define the *conditioned measures* on $\mathcal{F}_n$ by

$$\nu_n^C[d\omega] := \frac{1_{T_n^{-1}C}(\omega)\rho_n[d\omega]}{\rho_n[T_n^{-1}C]} ; \tag{2.1}$$

for $t \in E^*$ such that $0 < \int_{\Omega_n} \exp(V_n\langle t, T_n(\omega)\rangle)\rho_n[d\omega] < \infty$ for all n sufficiently large, we define the *tilted measures* on $\mathcal{F}_n$ by

$$\gamma_n^t[d\omega] := \frac{\exp(V_n\langle t, T_n(\omega)\rangle)\rho_n[d\omega]}{\int_{\Omega_n} \exp(V_n\langle t, T_n(\omega')\rangle)\rho_n[d\omega']} . \tag{2.2}$$

We shall compute the specific information gain $\lim_{n\to\infty} \frac{1}{V_n}\mathcal{H}(\nu_n^C|\gamma_n^t)$; recall that $\mathcal{H}(\lambda_1|\lambda_2)$, the information gain of λ_1 with respect to λ_2, is defined by

$$\mathcal{H}(\lambda_1|\lambda_2) := \begin{cases} \int_{\Omega_n} \ln \dfrac{d\lambda_1}{d\lambda_2}(\omega)\lambda_1[d\omega], & \lambda_1 \ll \lambda_2, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.3}$$

In the statistical mechanical applications, the T_n are k–tuples of functions such as energy–per–site and magnetization–per–site; then ν_n^C is the microcanonical measure conditioned on T_n taking values in C and γ_n^t is the grand canonical measure at generalized chemical potential t. Notice that both ν_n^C and γ_n^t are absolutely continuous with respect to the reference measure ρ_n and their densities are both functions of T_n; we exploit this by using the change of variable formula in computing the specific information gain. Define the distribution $\mathbb{M}_n$ of T_n under ρ_n by $\mathbb{M}_n := \rho_n \circ T_n^{-1}$; we have

$$\begin{aligned} \nu_n^C \circ T_n^{-1} &= \mathbb{M}_n[\,\cdot\,|C] := \frac{\mathbb{M}_n[\,\cdot\, \cap C]}{\mathbb{M}_n[C]} , && (2.4) \\ \gamma_n^t \circ T_n^{-1} &= \mathbb{M}_n^t[\,\cdot\,|X] := \frac{\mathbb{M}_n^t[\,\cdot\,]}{\mathbb{M}_n^t[X]} , && (2.5) \end{aligned}$$

where $\mathbb{M}_n^t[dx] := \exp(V_n\langle t, x\rangle)\mathbb{M}_n[dx]$. Thus we have

$$\mathcal{H}(\nu_n^C|\gamma_n^t) = \mathcal{H}(\mathbb{M}_n[\,\cdot\,|C]|\mathbb{M}_n^t[\,\cdot\,|X]) . \tag{2.6}$$

We shall see that this formula is the basic manoeuvre in our treatement; it reduces an integral over Ω_n to an integral over X and relates the information gain $\mathcal{H}(\nu_n^C|\gamma_n^t)$ to the thermodynamic functions which we are about to define in this setting.

3 The Principle of the Largest Term

We need to examine the behaviour as $n \to \infty$ of the measures on X defined in § 2. Since the spaces $(\Omega_n, \mathcal{F}_n, \rho_n)$ and the random variables T_n play no part in the considerations of this section it is best to start afresh. Let $\mathbb{M}_\circ := \{\mathbb{M}_n\}_{n\geq 1}$ be a sequence of locally finite positive measures on $\mathcal{B}(X)$, the Borel subsets of X, a compact convex subset of $E = \mathbb{R}^k$. Let $V_\circ$ be a scale; define set–functions m_n, $\underline{m}$, $\overline{m}$ on $\mathcal{B}(X)$:

$$\begin{aligned} m_n[B] &:= \frac{1}{V_n} \ln \mathbb{M}_n[B] , && (3.1) \\ \underline{m} &:= \liminf_{n\to\infty} m_n[B] , && (3.2) \\ \overline{m} &:= \limsup_{n\to\infty} m_n[B] . && (3.3) \end{aligned}$$

The following properties are straightforward consequences of the definitions

$$\underline{m}[B] \leq \overline{m}[B] \ \text{ for all } B \in \mathcal{B}(X) \ ; \tag{3.4}$$
$$\underline{m} \ \text{ and } \ \overline{m} \ \text{ are increasing on } \ \mathcal{B}(X) \ . \tag{3.5}$$

The next property is an abstract version of the Principle of the Largest Term, well–known in traditional accounts of statistical mechanics (see, for example, Huang (1963)). Since it is central to our development, we give a proof. (For a, $b \in \mathbb{R}$, we denote the maximum of a and b by $a \vee b$.)

Lemma 3.1 *On $\mathcal{B}(X)$, we have*

$$\overline{m}[B_1 \cup B_2] = \overline{m}[B_1] \vee \overline{m}[B_2] \ . \tag{3.6}$$

Proof:
For $j = 1, 2$, we have

$$\mathbb{M}_n[B_j] \leq \mathbb{M}_n[B_1 \cup B_2] \leq \mathbb{M}_n[B_1] + \mathbb{M}_n[B_2] \tag{3.7}$$

so that

$$\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2] \leq \mathbb{M}_n[B_1 \cup B_2] \leq 2\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2] \ ; \tag{3.8}$$

it follows that

$$\overline{m}[B_1 \cup B_2] = \limsup_{n \to \infty} (m_n[B_1] \vee m_n[B_2]) \ . \tag{3.9}$$

But for each pair $\{a_n\}_{n \geq 1}$, $\{b_n\}_{n \geq 1}$ of sequences of real numbers, we have

$$\limsup_{n \to \infty} (a_n \vee b_n) = (\limsup_{n \to \infty} a_n) \vee (\limsup_{n \to \infty} b_n) \ . \tag{3.10}$$

Thus (3.6) follows from (3.9) and (3.10). □

Define functions $\underline{\mu}$, $\overline{\mu}$ on X as follows:

$$\underline{\mu}(x) \ := \ \inf_{G \ni x} \underline{m}[G] \ \ , \ \ G \text{ open} \ , \tag{3.11}$$
$$\overline{\mu}(x) \ := \ \inf_{G \ni x} \overline{m}[G] \ \ , \ \ G \text{ open} \ . \tag{3.12}$$

The following properties are direct consequences of the definitions:

$$\underline{\mu} \ \text{ and } \ \overline{\mu} \ \text{ are upper semicontinuous functions} \ ; \tag{3.13}$$
$$\overline{m}[G] \geq \sup_{x \in G} \overline{\mu}(x) \ , \ \ G \text{ open} \ , \tag{3.14}$$
$$\underline{m}[G] \geq \sup_{x \in G} \underline{\mu}(x) \ , \ \ G \text{ open} \ . \tag{3.15}$$

The lower bound (3.14) for $\overline{m}$ on open sets is rarely used; of greater importance is the following upper bound for $\overline{m}$ on compact sets, a consequence of the Principle of the Largest Term (3.6)

$$\overline{m}[K] \leq \sup_{x \in K} \overline{\mu}(x) \ , \ \ K \text{ compact} \ . \tag{3.16}$$

Our first application of (3.16) is to the *concentration of measures*. Let $\mathbb{M}_\circ$ be a sequence of probability measures on $\mathcal{B}(X)$; if $\mathbb{M}_\circ$ converges weakly to a Dirac measure δ_x at some point $x \in X$, we say $\mathbb{M}_\circ$ obeys a weak law of large numbers (WLLN). In the absence of a first–order phase transition, a WLLN holds in the grand canonical ensemble. We require a substitute for a WLLN which holds regardless of phase transitions.

We say that a sequence $\mathbb{M}_\circ$ of probability measures on $\mathcal{B}(X)$ is *eventually concentrated on a set* A if, for each open neighbourhood G of A, we have

$$\lim_{n\to\infty} \mathbb{M}_n[G] = 1 . \tag{3.17}$$

[If $A = \{x\}$ and $\mathbb{M}_\circ$ is eventually concentrated on A, then $\mathbb{M}_\circ$ converges weakly to the Dirac measure δ_x.] We shall need the following

Lemma 3.2 *Let $\mathbb{M}_\circ$ be a sequence of probability measures on $\mathcal{B}(X)$ which is eventually concentrated on a set A; if $f : X \to \mathbb{R}$ is lower semicontinuous and bounded below on X, then*

$$\inf_{x\in A} f(x) \leq \liminf_{n\to\infty} \int_X f(x)\mathbb{M}_n[dx] . \tag{3.18}$$

[There is an obvious complementary upper bound; together they yields the usual characterization of the WLLN in terms of bounded continuous functions when A reduces to a single point.]

The function $\overline{\mu}$, defined at (3.12) for the pair $(\mathbb{M}_\circ, V_\circ)$, enables us to determine a concentration–set for the sequence $\mathbb{M}_\circ$. (How useful it is depends on how well we have chosen the scale $V_\circ$.) Notice that, for probability measures, the function $\overline{\mu}$ is bounded above by zero; in fact, it always attains this bound and the set on which it attains it is a concentration–set for $\mathbb{M}_\circ$. Let $N_{\overline{\mu}}$ be the set defined by

$$N_{\overline{\mu}} := \{x \in X : \overline{\mu}(x) = 0\} \tag{3.19}$$

Lemma 3.3 *Let $\mathbb{M}_\circ$ be a sequence of probability measures and $V_\circ$ a scale. Then*
(a) $N_{\overline{\mu}}$ is compact and non–empty;
(b) the sequence $\mathbb{M}_\circ$ is eventually concentrated on $N_{\overline{\mu}}$.

The proofs of both (a) and (b) make use of the bound (3.16)

Let $\overline{\mu}^t$, $\underline{\mu}^t$ be the upper and lower functions determined by the pair $(\mathbb{M}^t_\circ, V_\circ)$; they are related to $\overline{\mu}$ and $\underline{\mu}$ as follows:

$$\overline{\mu}^t(x) = \overline{\mu}(x) + \langle t, x\rangle , \tag{3.20}$$
$$\underline{\mu}^t(x) = \underline{\mu}(x) + \langle t, x\rangle . \tag{3.21}$$

These relations are a consequence of the continuity of the function $x \mapsto \langle t, x\rangle$. We are now ready for our third application of the bound (3.16): we prove a special case of Varadhan's Theorem (see Varadhan (1966)). If $\overline{\mu}(x) = \underline{\mu}(x)$ for all $x \in X$, we say the *Ruelle–Lanford function (RL–function) μ exists* for the pair $(\mathbb{M}_\circ, V_\circ)$ and is given by

$$\mu(x) := \underline{\mu}(x) = \overline{\mu}(x) . \tag{3.22}$$

When the RL–function exists, the bounds (3.15) and (3.16) can be restated as

$$\overline{m}[K] \leq \sup_{x\in K} \mu(x) , \quad K \text{ compact} , \tag{3.23}$$
$$\underline{m}[G] \geq \sup_{x\in G} \mu(x) , \quad G \text{ open} ; \tag{3.24}$$

When (3.23) and (3.24) hold, we say (following Varadhan (1966)) that a large deviation principle (LDP) holds with rate–function $I = -\mu$ for the pair $(\mathbb{M}_\circ, V_\circ)$. This means that the sequence $m_\circ$ of set–functions m_n, defined at (3.1), converges to the set–function

$$B \mapsto \sup_{x\in B} \mu(x) \tag{3.25}$$

in *exactly the same sense* that a sequence of probability measures $\mathbb{M}_\circ$ converges to a measure δ_x in a WLLN (remember that X is assumed to be compact). [We have given μ the name "Ruelle–Lanford function" because, in the setting of a lattice gas with translation–invariant summable potentials, our definition coincides with the definition of entropy given by Ruelle (1965) and Lanford (1973). Ruelle and Lanford understood that giving precise meaning to Boltzmann's formula

$$S = k \ln W \;, \tag{3.26}$$

relating the entropy S of a macroscopic equilibrium state to the number W of corresponding microscopic states is the *same problem* as that of making sense of the convergence of the sequence $m_\circ$ to the set–function (3.25); by so doing, they introduced a new technique to the theory of large deviations (compare Bahadur and Zabel (1979)).]

We are now ready to begin the calculation of the specific information gain using (2.6). First we have a result which is proved using (3.23) and (3.24):

Lemma 3.4 *Suppose the RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$ and the set $C \in \mathcal{B}(X)$ is such that*

$$-\infty < \sup_{x\in C} \mu(x) = \underline{m}[C] = \overline{m}[C] = \sup_{x\in \overline{C}} \mu(x) \;; \tag{3.27}$$

then the sequence $\mathbb{M}_\circ[\,\cdot\,|C]$ of probability measures is eventually concentrated on the set

$$X_{\overline{C}} := \{x \in \overline{C} : \mu(x) = \sup_{y\in\overline{C}} \mu(y)\} \;. \tag{3.28}$$

Lemma 3.5 *Suppose that the RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$; then*
(a) the RL–function μ^t exists for the pair $(\mathbb{M}^t_\circ, V_\circ)$;
(b) the pair $(\mathbb{M}^t_\circ, V_\circ)$ obeys an LDP:

$$\overline{m}^t[K] \;\le\; \sup_{x\in K} \mu^t(x) \;,\quad K \text{ compact} \;, \tag{3.29}$$

$$\underline{m}^t[G] \;\ge\; \sup_{x\in G} \mu^t(x) \;,\quad G \text{ open} \;; \tag{3.30}$$

(c) μ^t is given by

$$\mu^t(x) = \langle t, x\rangle + \mu(x) \;. \tag{3.31}$$

If $\overline{m}^t[X] = \underline{m}^t[X]$ for all $t \in E^*$, we say that the *scaled generating function* p exists for the pair $(\mathbb{M}_\circ, V_\circ)$ and is given by

$$p(t) := \overline{m}^t[X] = \underline{m}^t[X] \;. \tag{3.32}$$

(In the statistical mechanical setting, p is called the *grand canonical pressure.*) Recall that if $f : X \to \overline{\mathbb{R}}$, then $f^* : E^* \to \overline{\mathbb{R}}$ is defined by

$$f^*(t) := \sup_{x\in X}\{\langle t, x\rangle - f(x)\} \;. \tag{3.33}$$

Corollary 3.1 *Suppose the RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$; then the scaled generating function p exists and is given by*

$$p(t) = (-\mu)^*(t) \tag{3.34}$$

Proof:
Since X is both compact and open (as a topological space), we have

$$\sup_{x\in X}\mu^t(x) \le \underline{m}^t[X] \le \overline{m}^t[X] \le \sup_{x\in X}\mu^t(x)\ . \tag{3.35}$$

□

We define the set X^t for $t \in E^*$ by

$$X^t := \{x \in X : p(t) = \langle t, x\rangle + \mu(x)\}\ . \tag{3.36}$$

Theorem 3.1 *Suppose the RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$ and condition (3.27) holds; if $X_{\overline{C}} \subset X^t$, then the specific information gain is zero:*

$$\lim_{n\to\infty}\frac{1}{V_n}\mathcal{H}(\nu_n^C|\gamma_n^t) = 0\ . \tag{3.37}$$

Proof:
By (2.6), we have

$$\begin{aligned}\frac{1}{V_n}\mathcal{H}(\nu_n^C|\gamma_n^t) &= \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n[\,\cdot\,|C]|\mathbb{M}_n^t[\,\cdot\,|X]) \\ &= -\int\langle t,x\rangle\mathbb{M}_n[dy|C] + m_n^t[X] - m_n[C]\ .\end{aligned} \tag{3.38}$$

By Lemmas 3.4, 3.2, Corollary 3.1 and condition (3.27), we have

$$\begin{aligned}0 \le \limsup_{n\to\infty}\frac{1}{V_n}\mathcal{H}(\nu_n^C|\gamma_n^t) &\le -\inf_{y\in X_{\overline{C}}}\langle t,x\rangle + p(t) - \sup_{y\in X_{\overline{C}}}\mu(y) \\ &= \sup_{y\in X_{\overline{C}}}\{p(t) - \langle t,y\rangle - \mu(y)\} \\ &= 0\end{aligned} \tag{3.39}$$

if $X_{\overline{C}} \subset X^t$. □

4 An Application

To illustrate how Theorem 3.1 may be applied, we consider a case of sums of independent identically distributed random variables. We set $\Lambda_n := \{1,\ldots,n\}$, and in this example $V_n := |\Lambda_n| = n$, $\Omega_n := \{0,1\}^{\Lambda_n}$, $\mathcal{F}_n := \mathcal{P}(\Omega_n)$. For $\omega \in \Omega_n$, put $\xi_j(\omega) := \omega(j)$, $j \in \Lambda_n$, and set $\rho_n[\xi_j = 0] = \frac{1}{2} = \rho_n[\xi_j = 1]$. Then $T_n := V_n^{-1}\sum_{j\in\Lambda_n}\xi_j$, $X := [0,1]$, $E := \mathbb{R} = E^*$. Define $s : X \to [0,1]$ by

$$s(x) = -x\ln x - (1-x)\ln(1-x)\ ,\ x \in (0,1)\ ,\ s(0) = s(1) = 0\ . \tag{4.1}$$

Choose $C = (c_1, c_2) \subset [0,1]$; the RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$ and is given by

$$\mu(x) = s(x) - \ln 2\ ; \tag{4.2}$$

the set $X_{\overline{C}} = \{x^*\}$ where

$$x^* = \begin{cases} c_1, & \frac{1}{2} \le c_1, \\ \frac{1}{2}, & c_1 < \frac{1}{2} < c_2, \\ c_2, & c_2 \le \frac{1}{2}; \end{cases} \tag{4.3}$$

p is given by

$$p(t) = \ln(1 + \mathrm{e}^t) - \ln 2\ ; \tag{4.4}$$

and the set $X^t = \{x_t\}$ where

$$x_t = p'(t) = \frac{\mathrm{e}^t}{1 + \mathrm{e}^t} . \tag{4.5}$$

Given C, we can find t^* such that $X_{\overline{C}} = X^{t^*}$; thus we have

$$\lim_{n\to\infty} \frac{1}{V_n} \mathcal{H}(\nu_n^C | \gamma_n^{t^*}) = 0 \tag{4.6}$$

We can use (4.6) to obtain a result on the limit of the sequence $\{\nu_{n,\Delta}^C\}_{n\geq 1}$, where $\nu_{n,\Delta}^C$ is the restriction to a finite subset Δ of $\mathbb{N}$. Notice that γ_n^t is a product measure; this has two important consequences:

1. the restriction of γ_n^t to $\Delta \subset \{1, \ldots, n\}$ is independent of n and we denote it by γ_Δ^t;
2. if Δ_1 and Δ_2 are disjoint copies of Δ such that $\Delta_1 \cup \Delta_2 \subset \{1, \ldots, n\}$, then

$$\mathcal{H}(\nu_{n,\Delta_1\cup\Delta_2}^C | \gamma_{\Delta_1\cup\Delta_2}^t) \geq \mathcal{H}(\nu_{n,\Delta_1}^C | \gamma_{\Delta_1}^t) + \mathcal{H}(\nu_{n,\Delta_2}^C | \gamma_{\Delta_2}^t) . \tag{4.7}$$

But

$$\mathcal{H}(\nu_{n,\Delta_1}^C | \gamma_{\Delta_1}^t) = \mathcal{H}(\nu_{n,\Delta_2}^C | \gamma_{\Delta_2}^t) , \tag{4.8}$$

so that

$$\mathcal{H}(\nu_n^C | \gamma_n^t) \geq \left\lfloor \frac{V_n}{|\Delta|} \right\rfloor \mathcal{H}(\nu_{n,\Delta}^C | \gamma_\Delta^t) ; \tag{4.9}$$

hence (4.6) implies that

$$\lim_{n\to\infty} \mathcal{H}(\nu_{n,\Delta}^C | \gamma_\Delta^{t^*}) = 0 . \tag{4.10}$$

It now follows from (1.1) that $\{\nu_{n,\Delta}^C\}_{n\geq 1}$ converges in total variation norm to the product measure $\gamma_\Delta^{t^*}$.

5 The Lattice Gas

We consider the lattice gas model: let $\mathbb{Z}^d$ $(d \leq 1)$ be an integer–lattice, let $\{\Lambda_n\}_{n\geq 1}$ be an increasing sequence of cubes in $\mathbb{Z}^d$ with $V_n := |\Lambda_n| \to \infty$ as $n \to \infty$; at each site $j \in \Lambda_n$ we have a configuration space S_j which is a copy of some fixed compact Hausdorff space S. For each $n \geq 1$, the configuration space Ω_n is the space $\Omega_n = \prod_{j\in\Lambda_n} S_j$ which we regard as a subspace of the product space $\Omega = \prod_{j\in\mathbb{Z}^d} S_j$ equipped with the product topology, hence which is compact; the σ–field $\mathcal{F}_n$ is the σ–field of Borel subsets of Ω generated by the coordinate projections $\Omega \to S_j$. For each $j \in \mathbb{Z}^d$ we have the action of $\mathbb{Z}^d$ on itself given by $i \mapsto i + j$, $i \in \mathbb{Z}^d$; this lifts to $\theta_j : \Omega \to \Omega$ given by $(\theta_j\omega)(i) = \omega(i - j)$ for each configuration $\omega \in \Omega$. On each S_j we define a reference measure ρ^j, a copy of a fixed positive measure on S with $\rho^j(S_j) = 1$; on Ω we define the product measure $\rho = \prod_{j\in\mathbb{Z}^d} \rho^j$ and we take ρ_n to be the restriction of ρ to $\mathcal{F}_n$. The interaction in the model is given by a k–dimensional vector of translation–invariant absolutely summable potentials with either free or fixed boundary conditions. Using these potentials, we define mappings $T_n : \Omega_n \to X$ which give the energy per site of a configuration; here X is a compact convex subset of $E = \mathbb{R}^k$. We now define the conditioned measures ν_n^C and the tilted measures γ_n^t as in § 2; in this setting, the measure ν_n^C is the microcanonical measure on the cube Λ_n condition on T_n taking values in C (if C is an open neighbourhood of a point in X, then $T_n^{-1}C$ is what is sometimes called a "thin energy–shell" in Ω_n) and γ_n^t is a Gibbs measure on Λ_n with generalized chemical potential $t \in E^* = \mathbb{R}^k$. Using standard methods, we prove that $\overline{\mu}$ and $\underline{\mu}$ are independent of boundary conditions. Let $B_\varepsilon(x)$ be an open ball of radius ε and centre x in X; we prove, in the case of free boundary conditions, the following result.

Lemma 5.1 *Let x_0, x_1, $x_2 \in X$ satisfy $x_0 + x_1 = 2x_2$ and let $0 < \varepsilon' < \varepsilon$; then*

$$2\underline{m}[B_\varepsilon(x_2)] \geq \overline{m}[B_{\varepsilon'}(x_0)] + \overline{m}[B_{\varepsilon'}(x_1)] \, . \tag{5.1}$$

¿From this and the independence of $\overline{\mu}$ and $\underline{\mu}$ on the boundary conditions, we deduce the

Corollary 5.1 *The RL–function μ exists for the pair $(\mathbb{M}_\circ, V_\circ)$ and is concave on X.*

We have reserved the name "entropy" for the RL–functions which are concave; henceforth in this section, we refer to μ as the entropy of the pair $(\mathbb{M}_\circ, V_\circ)$ and to p, given by $p(t) = (-\mu)^*(t)$, as the grand canonical pressure. We now choose C to be an open convex subset of X; using convexity theory, we prove

Lemma 5.2 *Let C be an open convex subset of X; if μ is concave, then*
(a) $\sup_{x\in C} \mu(x) = \underline{m}[C] = \overline{m}[\overline{C}] = \sup_{x\in\overline{C}} \mu(x)$;
(b) the entropy μ_C of the pair $(\mathbb{M}_\circ[\,\cdot\,|C], V_\circ)$ is given by

$$\mu_C(x) = \begin{cases} \mu(x) - \sup_{y\in\overline{C}} \mu(y) \, , & y \in \overline{C}, \\ -\infty \, , & y \in X\backslash\overline{C}. \end{cases} \tag{5.2}$$

We see from (a) that, provided C is chosen so that it contains a point at which μ is finite, condition (3.27) is satisfied. Part (b) gives an interpretation of $X_{\overline{C}}$ in this case: $X_{\overline{C}} = N_{\mu_C}$, the set on which the entropy attains its supremum. There is also an interpretation of the set X^t which follows from the concavity of μ: using convexity theory we can show that

$$X^t = \partial p(t) \, , \tag{5.3}$$

(∂f denotes the subgradients to a convex function f; when $\dim X = 1$, the interval $\partial p(t)$ is "a phase–transition segment" in the grand canonical ensemble; it reduces to a point in the absence of a first order transition.) We see that Theorem 3.1 now yields

Theorem 5.1 *Let μ be the entropy of a lattice gas with translation invariant summable potential. Let C be an open convex neighbourhood of a point at which μ is finite. Then there exists t^* such that*

$$\lim_{n\to\infty} \frac{1}{V_n} \mathcal{H}(\nu_n^C | \gamma_n^{t^*}) = 0 \, . \tag{5.4}$$

Because, in the presence of a non–trivial interaction, the Gibbs measures γ_n^t are not product measures, the subadditivity argument used in § 4 fails. There is a second difficulty: in § 4 we exploited permutation–invariance (exchangeability) at (4.8); here we must replace it by translation–invariance, but the measures ν_n^C associated with the cubes Λ_n are not translation–invariant. The way–out is to introduce translation–averages: define

$$\overline{\nu}_n^C := \frac{1}{V_n} \sum_{j\in\Lambda_n} \nu_n^C \circ \theta_j^{-1} \, , \tag{5.5}$$

where ν_n^C is extended to Ω in the usual way. We are able to prove

Theorem 5.2 *Suppose that (5.4) holds; then any weak limit point of the sequence $\{\overline{\nu}_n^C\}_{n\geq 1}$ is a Gibbs state with respect to the specification associated with $\{\gamma_n^{t^*}\}_{n\geq 1}$.*

The statement of this theorem make precise the sense in which the measures $\overline{\nu}_n^C$ and $\gamma_n^{t^*}$ are "equivalent" in the thermodynamic limit – something we said in § 1 was part of the problem.

Putting Theorems 5.1 and 5.2 together, we see that the entropy μ can be used to find a value t^* of the chemical potential such that any weak limit of the sequence $\{\overline{\nu}_n^C\}_{n\geq 1}$ is a Gibbs state with respect to the specification determined by $\{\gamma_n^{t^*}\}_{n\geq 1}$. This is possible because, as a consequence of the concavity of μ, we have $\mu(x) = -p^*(x)$ as well as $p(t) = (-\mu)^*(t)$; but these statements together constitute the equivalence of ensembles at the level of thermodynamic functions. It is in this sense that *equivalence of ensemble holds at the level of measures whenever it holds at the level of thermodynamic functions.*

REFERENCES

1. R.R. Bahadur and S.L. Zabell, Large deviations of the sample mean in general vector spaces, *Ann. Prob.* 7: 587 (1979).

2. J.–D. Deuschel, D.W. Stroock and H. Zessin, Microcanonical distribution for lattice gases, *Commun. Math. Phys.* 139: 83 (1991).

3. R.L. Dobrushin and B. Tirozzi, The central limit theorem and the problem of equivalence of ensembles, *Commun. Math. Phys.* 54: 173 (1973).

4. H.–O. Georgii, Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$, *Ann. Prob.* to appear (1993).

5. K. Huang, Statistical Mechanics, Wiley, New–York (1963).

6. A.Ya. Khinchin, Matematicheskie Osnovaniya Statisticheskoi Mekhaniki, Gostekhizdat, Moscow–Leningrad (1943). Translation: Mathematical Foundations of Statistical Mechanics, Dover, New–York (1949).

7. O.E. Lanford, Entropy and equilibrium states in classical mechanics, in Statistical Mechanics and Mathematical Problems, A. Lenard, ed., Lecture Notes in Physics 20, Springer (1973).

8. J.T. Lewis, C.–E. Pfister and W.G. Sullivan, DIAS preprint (1993).

9. C.J. Preston, Random Fields. Lecture Notes in Mathematics 534, Spinger, Berlin (1976).

10. D. Ruelle, Correlation functionals, *J. Math. Physics* 6: 201 (1965).

11. W.G. Sullivan, Potentials for almost markovian random fields, *Commun. Math. Phys.* 33: 61 (1973).

12. S.R.S. Varadhan, Asymptotic probabilities and differential equations, *Comm. Pure App. Math.* 19: 261 (1966).