TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 185, November 1973

STATISTICAL MECHANICS ON A COMPACT SET WITH $\mathbf{Z}^\nu$ ACTION SATISFYING EXPANSIVENESS AND SPECIFICATION

BY
DAVID RUELLE

ABSTRACT. We consider a compact set Ω with a homeomorphism (or more generally a $\mathbf{Z}^\nu$ action) such that expansiveness and Bowen's specification condition hold. The entropy is a function on invariant probability measures. The pressure (a concept borrowed from statistical mechanics) is defined as function on $\mathcal{C}(\Omega)$—the real continuous functions on Ω. The entropy and pressure are shown to be dual in a certain sense, and this duality is investigated.

0. Introduction. Invariant measures for an Anosov diffeomorphism have been studied by Sinai [16], [17]. More generally, Bowen [2], [3] has considered invariant measures on basic sets for an Axiom A diffeomorphism. The problems encountered are strongly reminiscent of those of statistical mechanics (for a classical lattice system—see [14, Chapter 7]). In fact Sinai [18] has explicitly used techniques of statistical mechanics to show that an Anosov diffeomorphism does not in general have a smooth invariant measure.

In this paper, we rewrite a part of the general theory of statistical mechanics for the case of a compact set Ω satisfying expansiveness and the specification property of Bowen [2]. Instead of a $\mathbf{Z}$ action we consider a $\mathbf{Z}^\nu$ action as is usual in lattice statistical mechanics, where $\Omega = F^{\mathbf{Z}^\nu}$ (F: a finite set). This rewriting gives a more general and intrinsic formulation of (part of) statistical mechanics; it presents a number of technical problems, but the basic ideas are contained in the papers of Gallavotti, Lanford, Miracle, Robinson, and Ruelle [7], [11], [12], [13], etc. The ideas of Bowen [2] and Goodwyn [8] on the relation between topological and measure-theoretical entropy are also used.

We describe now some of our results in the case of a homeomorphism T of a metrizable compact set Ω satisfying expansiveness and specification (see §1).

Let $\Pi_a = \{x \in \Omega: T^a x = \{x\}\}$, and let $\mathcal{C}(\Omega)$ be the Banach space of real continuous functions on Ω. The *pressure* P is a continuous convex function on $\mathcal{C}(\Omega)$ defined by

$$P(\varphi) = \lim_{a\to\infty} \frac{1}{a} \log Z(\varphi, a), \qquad Z(\varphi, a) = \sum_{x\in\Pi_a} \exp \sum_{m=1}^{a} \varphi(T^m x)$$

(§2). Let I be the set of probability measures on Ω, invariant under T with the vague topology. The (measure theoretic) *entropy* s is an affine upper semicontinuous function on I defined in the usual way (§4). The following variational principle holds (§5)

Received by the editors June 20, 1972.
AMS (MOS) subject classifications (1970). Primary 28A65, 54H20; Secondary 58F99, 82A30.

(0.1) $$P(\varphi) = \max_{\mu \in I}[s(\mu) + \mu(\varphi)], \qquad s(\mu) = \inf_{\varphi \in \mathcal{C}(\Omega)}[P(\varphi) - \mu(\varphi)].$$

Those μ for which the maximum is reached in (0.1) form a nonempty set I_φ. I_φ is a Choquet simplex and consists of precisely those $\mu \in I$ such that

$$P(\varphi + \psi) - P(\varphi) \geq \mu(\psi), \quad \text{all } \psi \in \mathcal{C}(\Omega).$$

Let $\mu_{\varphi,a}$ be the measure on Ω which is carried by Π_a and gives $x \in \Pi_a$ the mass

$$\mu_{\varphi,a}(\{x\}) = Z(\varphi, a)^{-1} \exp \sum_{m=1}^{a} \varphi(T^m x).$$

Then, any limit point of $\mu_{\varphi,a}$ as $a \to \infty$ is in I_φ (§3). There is a residual subset D of $\mathcal{C}(\Omega)$ such that I_φ consists of one single point μ_φ if $\varphi \in D$. In that case $\lim_{a\to\infty} \mu_{\varphi,a} = \mu_\varphi$.

Miscellaneous properties of invariant states are reviewed in §6.

I am indebted to J. Robbin for acquainting me with Bowen's papers, starting the present work.

1. **Notation and assumptions.** We denote by $|S|$ the cardinal of the set S. If $m = (m_1, \ldots, m_\nu) \in \mathbf{Z}^\nu$, $\nu \geq 1$, we let $\|m\| = \sup_i |m_i|$. Given integers $a_1, \ldots, a_\nu > 0$, we define $\Lambda(a) = \{m \in \mathbf{Z}^\nu : 0 \leq m_i < a_i\}$. If (Λ_α) is a directed family of finite subsets of $\mathbf{Z}^\nu$, $\Lambda_\alpha \uparrow \infty$ means $|\Lambda_\alpha| \to \infty$ and $|\Lambda_\alpha + F|/|\Lambda_\alpha| \to 1$ for every finite $F \subset \mathbf{Z}^\nu$. In particular $\Lambda(a) \uparrow \infty$ when $a \to \infty$ (i.e. when $a_1, \ldots, a_\nu \to \infty$).

Let $\mathbf{Z}^\nu$ act by homeomorphisms on the compact set Ω. We suppose that Ω is metrizable with metric d. $\mathcal{C}(\Omega)$ is the space of real continuous functions on Ω with the sup norm. On the space $\mathcal{C}(\Omega)^*$ of real measures on Ω, we put the vague topology. We denote by δ_x the unit mass at x.

The following assumptions are made.[1]

1.1. *Expansiveness.* There exists $\delta^* > 0$ such that

$$(d(mx, my) \leq \delta^* \text{ for all } m \in \mathbf{Z}^\nu) \Rightarrow (x = y).$$

1.2. *Weak specification.* Given $\delta > 0$ there exists $p(\delta) > 0$ such that for any families $(\Lambda_i)_{i \in \mathcal{J}}$, $(x_i)_{i \in \mathcal{J}}$ satisfying

(i) if $i \neq j$, the distance of Λ_i, Λ_j (as subsets of $\mathbf{Z}^\nu$, with the distance $\|\cdot\|$) is $> p(\delta)$,

there is $x \in X$ such that

$$d(m_i x, m_i x_i) < \delta, \quad \text{all } i \in \mathcal{J}, \text{ all } m_i \in \Lambda_i.$$

1.3. *Strong specification.* Let $\mathbf{Z}^\nu(a)$ be the subgroup of $\mathbf{Z}^\nu$ with generators $(a_1, 0, \ldots, 0), \ldots, (0, \ldots, a_\nu)$, and let $\Pi_a = \{x \in \Omega : \mathbf{Z}^\nu(a)x = \{x\}\}$. For any

[1] Cf. Bowen [2].

families $(\Lambda_i)_{i\in\mathcal{I}}$, $(x_i)_{i\in\mathcal{I}}$ satisfying

$$\text{(ii)} \qquad \begin{array}{l} \Lambda_i \subset \Lambda(a) \text{ for all } i \text{ and, if } i \neq j, \\ \text{the distance of } \Lambda_i + \mathbf{Z}^\nu(a) \text{ and } \Lambda_j \text{ is } > p(\delta), \end{array}$$

there is $x \in \Pi_a$ such that

$$d(m_i x, m_i x_i) < \delta, \quad \text{all } i \in \mathcal{I}, \text{ all } m_i \in \Lambda_i.$$

It is easily seen that strong specification implies weak specification. If Ω is a basic set for an Axiom A diffeomorphism ($\nu = 1$), it is known that expansiveness [19] holds, and that (strong) specification [2] holds for some iterate of the diffeomorphism.

We note that expansiveness has the following easy consequence.

1.4. **Proposition** [9]. *If* $0 < \delta$ *there exists* $q(\delta)$ *such that* $(d(mx,my) \leq \delta^*$ *if* $|m| < q(\delta)) \Rightarrow (d(x,y) < \delta)$.

2. Partition functions and pressure.

2.1. **Definitions.** Let $\delta > 0$; $E \subset \Omega$ is (δ, Λ)-separated if $(x, y \in E$, and $d(mx,my) < \delta$ for all $m \in \Lambda) \Rightarrow (x = y)$. Let $\varphi \in \mathcal{C}(\Omega)$. Given $\delta > 0$ and a finite $\Lambda \subset \mathbf{Z}^\nu$, or given $a = (a_1, \ldots, a_\nu)$ we introduce the *partition functions*

$$Z(\varphi, \delta, \Lambda) = \max_E \sum_{x\in E} \exp \sum_{m\in\Lambda} \varphi(mx) \tag{2.1}$$

where the max is taken over all (δ, Λ)-separated sets, or

$$Z(\varphi, a) = \sum_{x\in\Pi_a} \exp \sum_{m\in\Lambda(a)} \varphi(mx). \tag{2.2}$$

We write

$$P(\varphi, \delta, \Lambda) = (1/|\Lambda|)\log Z(\varphi, \delta, \Lambda), \tag{2.3}$$

$$P(\varphi, a) = (1/|\Lambda(a)|)\log Z(\varphi, a). \tag{2.4}$$

2.2. **Theorem.** *If* $0 < \delta < \delta^*$, *the following limits exist*:

$$\lim_{\Lambda\uparrow\infty} P(\varphi, \delta, \Lambda) = P(\varphi), \tag{2.5}$$

$$\lim_{a\to\infty} P(\varphi, a) = P(\varphi), \tag{2.6}$$

and define a finite-valued convex function P *on* $\mathcal{C}(X)$. *Furthermore*

$$|P(\varphi) - P(\psi)| \leq \|\varphi - \psi\| \tag{2.7}$$

and if $\tau_m\psi(x) = \psi(mx)$, $t \in \mathbf{R}$,

$$P(\varphi + \tau_m\psi - \psi + t) = P(\varphi) + t. \tag{2.8}$$

P is called the pressure.

Let $\epsilon > 0$; we choose $\delta' > 0$ so small that $\delta + 2\delta' \leq \delta^*$ and

$$(2.9) \qquad (d(x,y) < \delta') \Rightarrow |\varphi(x) - \varphi(y)| < \epsilon;$$

then take $p(\delta')$ according to 1.2. Given a, write $b = (a_1 + p(\delta'), \ldots, a_\nu + p(\delta'))$. We consider the partition $(\Lambda(b) + r)_{r \in \mathbf{Z}^\nu(b)}$ of $\mathbf{Z}^\nu$. For a finite $\Lambda \subset \mathbf{Z}^\nu$, let $R = \{r: \Lambda(b) + r \subset \Lambda\}$. Using specification we obtain

$$\begin{aligned} &Z(\varphi, \delta, \Lambda) \\ (2.10) \qquad &\geq [Z(\varphi, \delta + 2\delta', \Lambda(a)) \exp(-|\Lambda(a)|\epsilon) \exp(-(|\Lambda(b)| - |\Lambda(a)|)\|\varphi\|)]^{|R|} \\ &\quad \cdot \exp(-(|\Lambda| - |R|\,|\Lambda(b)|)\|\varphi\|). \end{aligned}$$

Since Π_a is $(\delta^*, \Lambda(a))$-separated by expansiveness, we have also

$$(2.11) \qquad Z(\varphi, \delta^*, \Lambda(a)) \geq Z(\varphi, a).$$

If $\Lambda \uparrow \infty$ we have $|R|\,|\Lambda(b)|/|\Lambda| \to 1$, and therefore (2.10) and (2.11) yield

$$(2.12) \qquad \liminf_{\Lambda \uparrow \infty} P(\varphi, \delta, \Lambda) \geq \frac{|\Lambda(a)|}{|\Lambda(b)|} \cdot [P(\varphi, a) - \epsilon] - \left(1 - \frac{|\Lambda(a)|}{|\Lambda(b)|}\right)\|\varphi\|.$$

Suppose now that $\delta' < \frac{1}{2}\delta$, and let N be the cardinal of a finite cover of Ω by sets of diameter $< \delta$. Let F be a $(\delta', \Lambda(b))$ separated set such that

$$Z(\varphi, \delta', \Lambda(b)) = \sum_{y \in F} \exp \sum_{m \in \Lambda(b)} \varphi(my).$$

Given $x \in E$ and $r \in R$ we choose $y \in F$ such that $d((r + m)x, my) < \delta'$, for all $m \in \Lambda(b)$. The mapping $(x, r) \to y$ defines an injection $E \to F^R$, and therefore

$$(2.13) \qquad Z(\varphi, \delta, \Lambda) \leq [Z(\varphi, \delta', \Lambda(b)) \exp(|\Lambda(b)|\epsilon)]^{|R|} (N \exp \|\varphi\|)^{|\Lambda| - |R|\,|\Lambda(b)|}.$$

Taking $c = (b_1 + p(\delta'), \ldots, b_\nu + p(\delta'))$, strong specification gives

$$(2.14) \qquad Z(\varphi, \delta', \Lambda(b)) \exp(-|\Lambda(b)|\epsilon) \exp(-(|\Lambda(c)| - |\Lambda(b)|)\|\varphi\|) \leq Z(\varphi, c).$$

From (2.13) and (2.14) we obtain

$$(2.15) \qquad \limsup_{\Lambda \uparrow \infty} P(\varphi, \delta, \Lambda) \leq \frac{|\Lambda(c)|}{|\Lambda(b)|} P(\varphi, c) + 2\epsilon + \left(\frac{|\Lambda(c)|}{|\Lambda(b)|} - 1\right)\|\varphi\|.$$

Letting $a \to \infty$ in (2.12) and (2.15) we obtain (2.5) and (2.6).

The finiteness of $P(\varphi)$ follows from $\exp(-|\Lambda|\,\|\varphi\|) \leq Z(\varphi, \delta, \Lambda) \leq N^{|\Lambda|} \exp(|\Lambda|\,\|\varphi\|)$. The other properties follow from Lemma 2.3 below.

2.3. **Lemma.** *$P(\varphi, \delta, \Lambda)$ is a convex function of φ. Furthermore $|P(\varphi, \delta, \Lambda) - P(\psi, \delta, \Lambda)| \leq \|\varphi - \psi\|$ and $P(\varphi + t, \delta, \Lambda) = P(\varphi, \delta, \Lambda) + t$, if $t \in \mathbf{R}$. Similar* *properties hold for $P(\varphi, a)$, and also $P(\varphi + \tau_m \psi - \psi, a) = P(\varphi, a)$.*

We have $P(\varphi, \delta, \Lambda) = \max_E p(\varphi)$ where

$$p(\varphi) = (1/|\Lambda|)\log Z(\varphi), \qquad Z(\varphi) = \sum_{x \in E} \exp \sum_{m \in \Lambda} \varphi(mx),$$

$$\frac{d}{dt} p(\varphi + t\psi) = \frac{1}{Z(\varphi + t\psi)} \sum_x \frac{1}{|\Lambda|}\left[\sum_{m'} \psi(m'x)\right] \exp \sum_m [\varphi(mx) + t\psi(mx)].$$

Therefore

$$|\Lambda| \frac{d^2}{dt^2} p(\varphi + t\psi)|_{t=0}$$
$$= \frac{1}{Z^2} \sum_x \sum_y \frac{1}{2}\left[\sum_m \psi(mx) - \sum_m \psi(my)\right]^2 \exp \sum_m [\varphi(mx) + \varphi(my)] \geq 0.$$

On the other hand $|dp(\varphi + t\psi)/dt| \leq \|\psi\|$; hence

$$|p(\varphi) - p(\psi)| \leq \sup_{0 \leq t \leq 1} \left|\frac{d}{dt} p(\varphi + t(\psi - \varphi))\right| \leq \|\psi - \varphi\|.$$

Finally $Z(\varphi + t) = e^{|\Lambda| t} Z(\varphi)$, $Z(\varphi + \tau_m \psi - \psi, a) = Z(\varphi, a)$.

2.4. **Remark.** Let Σ be the subgroup of $\mathbf{Z}^\nu$ with linearly independent generators $s_1, \ldots, s_\nu$, and define $\Lambda(\Sigma) = \{m \in \mathbf{Z}^\nu \colon m = \sum_1^\nu t_i s_i \text{ with } t_i \text{ real}, 0 \leq t_i < 1\}$. If a suitable extension of the strong specification property holds, one can prove

$$P(\varphi) = \lim_{\Lambda(\Sigma) \uparrow \infty} \frac{1}{|\Lambda(\Sigma)|} \log \sum_{x \in \Pi_\Sigma} \exp \sum_{m \in \Lambda(\Sigma)} \varphi(mx),$$

where $\Pi_\Sigma = \{x \colon \Sigma x = \{x\}\}$.

On the other hand, except for (2.6), Theorem 2.2 can be proved without the strong specification property (but assuming expansiveness and weak specification).

3. Equilibrium states.

3.1. **Definition.** Let $\mu_{\varphi,a}$ be the measure on Ω which is carried by Π_a and gives $x \in \Pi_a$ the mass

$$\mu_{\varphi,a}(\{x\}) = Z(\varphi, a)^{-1} \exp \sum_{m \in \Lambda(a)} \varphi(mx). \tag{3.1}$$

3.2. **Theorem.** (a) *Let $I_\varphi \subset \mathcal{C}(\Omega)^*$ be the set of measures μ such that*

$$P(\varphi + \psi) \geq P(\varphi) + \mu(\psi) \tag{3.2}$$

for all ψ (equilibrium states for φ). Then I_φ is nonempty and there is a residual (2) *set $D \subset \mathcal{C}(\Omega)$ such that I_φ consists of a single point μ_φ if $\varphi \in D$.*

(2) I.e. D is a countable intersection of dense open subsets of $\mathcal{C}(\Omega)$; in particular D is dense in $\mathcal{C}(\Omega)$ by Baire's theorem.

(b) I_φ *is convex, (vaguely) compact, and consists of* $\mathbf{Z}^\nu$ *invariant probability measures.*

(c) *The probability measure* $\mu_{\varphi,a}$ *is* $\mathbf{Z}^\nu$ *invariant, and*

$$\mu_{\varphi,a}(\psi) = dP(\varphi + t\psi, a)/dt\,|_{t=0}\,. \tag{3.3}$$

(d) *If* μ *is a (vague) limit point of the* $(\mu_{\varphi,a})$ *when* $a \to \infty$, *then* $\mu \in I_\varphi$. *In particular, if* $\varphi \in D$,

$$\lim_{a\to\infty} \mu_{\varphi,a} = \mu_\varphi\,. \tag{3.4}$$

(e) *If* $\mathcal{B}$ *is dense in* $\mathcal{C}(\Omega)$ *and is a separable Banach space with respect to a norm* $|||\cdot||| \geq \|\cdot\|$, *then* $D \cap \mathcal{B}$ *is residual in* $\mathcal{B}$.

(a) holds for any convex continuous function P on a separable Banach space (see Dunford-Schwartz [6, Theorem V.9.8]). This proves also (e).

Let μ satisfy (3.2). Then by (2.8),

$$0 = P(\varphi + \tau_m\psi - \psi) - P(\varphi) \geq \mu(\tau_m\psi - \psi) \geq -[P(\varphi - \tau_m\psi + \psi) - P(\varphi)] = 0$$

so that μ is $\mathbf{Z}^\nu$ invariant. Using (2.7) and (2.8) we obtain also $\pm\mu(\psi) \leq P(\varphi \pm \psi) - P(\varphi) \leq \|\psi\|$ and $\mu(1) = -\mu(-1) \geq -[P(\varphi - 1) - P(\varphi)] = 1$. Therefore $\|\mu\| \leq 1$, $\mu(1) \geq 1$ which implies that $\mu \geq 0$, $\|\mu\| = 1$, i.e. μ is a probability measure. Clearly, I_φ is convex and compact, and (b) is thus proved.

(c) follows readily from the definitions. From (3.3) and the convexity of $P(\cdot, a)$ (Lemma 2.3), we obtain

$$P(\varphi + \psi, a) \geq P(\varphi, a) + \mu_{\varphi,a}(\psi).$$

If $\mu_{\varphi,a} \to \mu$ this yields (3.2), proving (d).

4. Entropy.(3)

4.1. **Definitions.** Let $\mathcal{A} = (A_i)_{i\in\mathcal{I}}$ be a finite Borel partition of Ω, and Λ a finite subset of $\mathbf{Z}^\nu$. We denote by $\mathcal{A}^\Lambda$ the partition of Ω consisting of the sets $A(k) = \bigcap_{m\in\Lambda}(-m)A_{k(m)}$ indexed by maps $k\colon \Lambda \to \mathcal{I}$. We write

$$S(\mu, \mathcal{A}) = -\sum_i \mu(A_i)\log \mu(A_i). \tag{4.1}$$

Let I be the (convex compact) set of $\mathbf{Z}^\nu$-invariant probability measures on Ω.

4.2. **Theorem.** *If* $\mathcal{A}$ *consists of sets with diameter* $\leq \delta^*$, *and* $\mu \in I$, *then*

$$\lim_{\Lambda\uparrow\infty} \frac{1}{|\Lambda|} S(\mu, \mathcal{A}^\Lambda) = \inf_\Lambda \frac{1}{|\Lambda|} S(\mu, \mathcal{A}^\Lambda) = s(\mu). \tag{4.2}$$

(3) See also J.-P. Conze, *Entropie d'un groupe abélien de transformations* [Z. Wahrscheinlichkeitstheorie Verw. Gebiete **25** (1972), 11–30].

This limit is finite ≥ 0, and independent of $\mathcal{A}$. Furthermore, s is affine upper semicontinuous on I; s is called the entropy.

$S(\mu, \mathcal{A}^\Lambda)$ is an increasing function of Λ, and satisfies the strong subadditivity property

$$(4.3) \qquad S(\mu, \mathcal{A}^{\Lambda_1 \cup \Lambda_2}) + S(\mu, \mathcal{A}^{\Lambda_1 \cap \Lambda_2}) \leq S(\mu, \mathcal{A}^{\Lambda_1}) + S(\mu, \mathcal{A}^{\Lambda_2}).$$

[These are well-known properties. The increase follows from increase of the logarithm. To prove strong subadditivity we write $S(\mu, \mathcal{A}^\Lambda) = S_\Lambda$, and use the inequality $-\log(1/t) \leq t - 1$, then

$$\begin{aligned} &S_{\Lambda_1 \cup \Lambda_2} + S_{\Lambda_1 \cap \Lambda_2} - S_{\Lambda_1} - S_{\Lambda_2} \\ &\quad = -\sum_{k:\Lambda_1 \cap \Lambda_2 \to \mathcal{O}} \ \sum_{k':\Lambda_1 \backslash \Lambda_2 \to \mathcal{O}} \ \sum_{k'':\Lambda_2 \backslash \Lambda_1 \to \mathcal{O}} \mu(A(k,k',k''))\log \frac{\mu(A(k,k',k''))\mu(A(k))}{\mu(A(k,k'))\mu(A(k,k''))} \\ &\quad \leq \sum_{kk'k''} \mu(A(k,k',k''))\left[\frac{\mu(A(k,k'))\mu(A(k,k''))}{\mu(A(k,k',k''))\mu(A(k))} - 1\right] \\ &\quad = \sum_{kk'} \frac{\mu(A(k,k'))}{\mu(A(k))} \sum_{k''} \mu(A(k,k'')) - \sum_{kk'k''} \mu(A(k,k',k'')) \\ &\quad = \sum_{kk'} \mu(A(k,k')) - 1 = 0.] \end{aligned}$$

If $\Lambda_1 \cap \Lambda_2 = \varnothing$, (4.3) becomes subadditivity: $S(\mu, \mathcal{A}^{\Lambda_1 \cup \Lambda_2}) \leq S(\mu, \mathcal{A}^{\Lambda_1}) + S(\mu, \mathcal{A}^{\Lambda_2})$. Since $\mu \in I$ we have also $S(\mu, \mathcal{A}^\Lambda) = S(\mu, \mathcal{A}^{\Lambda+m})$ and therefore(4)

$$(4.4) \qquad \lim_{a\to\infty} \frac{1}{|\Lambda(a)|} S(\mu, \mathcal{A}^{\Lambda(a)}) = \inf_a \frac{1}{|\Lambda(a)|} S(\mu, \mathcal{A}^{\Lambda(a)}) = s.$$

Given $\epsilon > 0$, choose a such that $|\Lambda(a)|^{-1} S(\mu, \mathcal{A}^{\Lambda(a)}) \leq s + \epsilon$. Consider the partition $(\Lambda(a) + r)_{r \in \mathbf{Z}^\nu(a)}$ of $\mathbf{Z}^\nu$, and let $R = \{r \in \mathbf{Z}^\nu(a)\colon (\Lambda(a) + r) \cap \Lambda \neq \varnothing\}$. If $\Lambda_+ = \bigcup_{r \in R}(\Lambda(a) + r)$ we have by increase and subadditivity

$$S(\mu, \mathcal{A}^\Lambda) \leq S(\mu, \mathcal{A}^{\Lambda_+}) \leq |R| S(\mu, \mathcal{A}^{\Lambda(a)}) \leq |R|\,|\Lambda(a)|(s + \epsilon).$$

But $|R|\,|\Lambda(a)|/|\Lambda| \to 1$ when $\Lambda \uparrow \infty$, and therefore

$$(4.5) \qquad \limsup_{\Lambda \uparrow \infty} |\Lambda|^{-1} S(\mu, \mathcal{A}^\Lambda) \leq s + \epsilon.$$

Strong subadditivity shows that

$$(4.6) \qquad S(\mu, \mathcal{A}^{\Lambda \cup \{m\}}) - S(\mu, \mathcal{A}^\Lambda) \geq S(\mu, \mathcal{A}^{\Lambda' \cup \{m\}}) - S(\mu, \mathcal{A}^{\Lambda'})$$

when $m \notin \Lambda' \supset \Lambda$. This permits an estimate of the increase in the entropy for a set Λ to which points are added successively in lexicographic order. In

(4) See for instance [14, Proposition 7.2.4].

particular if Λ is fixed and one takes for Λ' the sets successively obtained in the lexicographic construction of a large $\Lambda(a)$, (4.6) holds for most Λ'. Therefore

$$S(\mu, \mathscr{A}^{\Lambda\cup\{m\}}) - S(\mu, \mathscr{A}^{\Lambda}) \geq \lim_{a\to\infty} |\Lambda(a)|^{-1} S(\mu, \mathscr{A}^{\Lambda(a)}) = s$$

and hence

$$S(\mu, \mathscr{A}^{\Lambda}) \geq |\Lambda| s \tag{4.7}$$

for all Λ; (4.2) follows from (4.5) and (4.7).

Let $x \in \Omega$ and for each $m \in \Lambda$, let B_m be the union of those A_i which contain x in their closure. Then $B_\Lambda = \bigcap_{m\in\Lambda}(-m)B_m$ contains x in its interior and is a union of elements of $\mathscr{A}^\Lambda$. If $y \in B_\Lambda$ and $\Lambda = \{m: |m| < q(\delta)\}$, then $d(x,y) < \delta$ (see (1.4)). Therefore the σ-field generated by the $\mathscr{A}^\Lambda$ is the Borel σ-field. The Kolmogorov-Sinai theorem (see [20, 5.5]) holds for the group $\mathbf{Z}^\nu$ and implies that the limit (4.2) is independent of $\mathscr{A}$ (it is clearly finite ≥ 0).

If $\mu, \mu' \in I$, and $0 < \alpha < 1$, the following inequalities are standard:

$$\begin{aligned} \alpha S(\mu, \mathscr{A}) + (1-\alpha) S(\mu', \mathscr{A}) &\leq S(\alpha\mu + (1-\alpha)\mu', \mathscr{A}) \\ &\leq \alpha S(\mu, \mathscr{A}) + (1-\alpha) S(\mu', \mathscr{A}) + \log 2. \end{aligned} \tag{4.8}$$

[Writing $\mu_i = \mu(A_i)$, $\mu'_i = \mu(A'_i)$ we have indeed, using the convexity of $t \log t$ and the increase of $\log t$,

$$\begin{aligned} &-\sum_i [\alpha\mu_i \log \mu_i + (1-\alpha)\mu'_i \log \mu'_i] \\ &\quad\leq -\sum_i [\alpha\mu_i + (1-\alpha)\mu'_i]\log[\alpha\mu_i + (1-\alpha)\mu'_i] \\ &\quad\leq -\sum_i [\alpha\mu_i \log \alpha\mu_i + (1-\alpha)\mu'_i \log(1-\alpha)\mu'_i] \\ &\quad= -\sum_i [\alpha\mu_i \log \mu_i + (1-\alpha)\mu'_i \log \mu'_i] - \alpha\log\alpha - (1-\alpha)\log(1-\alpha) \\ &\quad\leq -\sum_i [\alpha\mu_i \log \mu_i + (1-\alpha)\mu'_i \log \mu'_i] + \log 2.] \end{aligned}$$

(4.8) implies that s is affine.

To prove that s is upper semicontinuous at μ, choose $\mathscr{A}$ such that the boundaries of the A_i have μ-measure zero. [If $x \in \Omega$ one can choose $\delta \leq \frac{1}{2}\delta^*$ such that the boundary of the sphere of radius δ centered at x has μ-measure 0. Take a finite covering of Ω by such spheres and let $\mathscr{A}$ be generated by this covering.] The boundaries of the $A(k) \in \mathscr{A}^\Lambda$ have also measure 0, hence

$$\lim_{\mu'\to\mu} \mu'(A(k)) = \mu(A(k)), \qquad \lim_{\mu'\to\mu} S(\mu', \mathscr{A}^\Lambda) = S(\mu, \mathscr{A}^\Lambda),$$

and s is upper semicontinuous as inf of continuous functions.

4.3. Remarks. (a) Theorem 4.2 reduces to the usual definition of the measure theoretic entropy for $\nu = 1$.

(b) The condition that the diameters of the A_i are $\leq \delta^*$ can be replaced by the weaker condition that the $\mathcal{A}^\Lambda$ generate the Borel σ-field (see the proof).

(c) The proof of Theorem 4.2 assumes expansiveness, but specification is not used.

5. Variational principle.

5.1. **Theorem.** *For all* $\varphi \in \mathcal{C}(\Omega)$,

$$P(\varphi) = \max_{\mu \in I}[s(\mu) + \mu(\varphi)] \tag{5.1}$$

and the maximum is reached precisely on I_φ. *For all* $\mu \in I$,

$$s(\mu) = \inf_\varphi [P(\varphi) - \mu(\varphi)].$$

Let $\varphi \in \mathcal{C}(\Omega)$ and $\mu \in I$ be given. Since Ω is metrizable compact, there exists a finite set $\{\psi_1, \ldots, \psi_t\}$ of elements of $\mathcal{C}(\Omega)$ such that if $|\psi_l(x) - \psi_l(y)| < 1$ for $l = 1, \ldots, t$, then $d(x,y) < \delta^*$. Given $\epsilon > 0$ and a we construct a partition $\mathcal{B} = (B_i)_{i \in \mathcal{J}}$ consisting of sets of the form $B_i = \{x: u_{ilm} \leq \psi_l(mx) < v_{ilm}$ and $u'_{im} \leq \varphi(mx) < v'_{im}$ for all i, l, and $m \in \Lambda(a)\}$. By suitable choice of the u_{ilm}, v_{ilm}, u'_{im}, v'_{im} we can achieve that

(a) the diameter of each set $(-m)B_i$, for $m \in \Lambda(a)$, is $\leq \delta^*$;

(b) if B_i, B_j are adjacent (i.e. $\bar{B}_i \cap \bar{B}_j \neq \emptyset$) and $x \in B_i$, $y \in B_j$, then

$$|\varphi(mx) - \varphi(my)| < \epsilon/2 \text{ for all } m \in \Lambda(a);$$

(c) each $x \in X$ is contained in the closure of at most $(t+1)|\Lambda(a)| + 1$ sets B_i. (5)

Because of (c) there exists δ, $0 < \delta < \delta^*$, such that for each x there are at most $(t+1)|\Lambda(a)| + 1$ sets B_i with distance $< \delta$ to x, and these sets are all adjacent to that containing x.

Let R be a subset of $\mathbf{Z}^\nu(a)$, then

$$\inf_R \frac{1}{|R|} S(\mu, \mathcal{B}^R) = |\Lambda(a)| s(\mu). \tag{5.3}$$

To see this notice that the $\mathcal{B}^R$ generate the Borel σ-field (by (a) above), and apply Remark 4.3(b) with $\mathbf{Z}^\nu$ replaced by $\mathbf{Z}^\nu(a)$. It follows that the left-hand side of (5.3) is not changed if $\mathcal{B}$ is replaced by $\mathcal{A}^{\Lambda(a)}$, and (5.3) follows. If E is a maximal (δ, R)-separated set, for each $k: R \to \mathcal{J}$ such that $B(k) \neq \emptyset$, one can choose $x \in B(k)$ and then $x_k \in E$ such that $d(rx_k, rx) < \delta$, all $r \in R$. By the choice of δ, rx_k is in a set B_i adjacent to $B_{k(r)}$. Therefore, by (b),

$$\left| \sum_{m \in \Lambda(a)} \varphi((r+m)x_k) - \sum_{m \in \Lambda(a)} \varphi(my) \right| < |\Lambda(a)|\epsilon/2$$

(5) The B_i may be viewed as $(t+1)|\Lambda(a)|$-dimensional rectangles and they can be adjusted so that at most $(t+1)|\Lambda(a)| + 1$ meet at a corner. This idea is used by Goodwyn [8].

for all $y \in B_{k(r)}$. Choose $y_i \in B_i$ for each $i \in \mathcal{J}$, then

$$\begin{aligned}
&\frac{1}{|R|}\sum_{k:R\to\mathcal{J}}\mu(B(k))\sum_{r\in R}\sum_{m\in\Lambda(a)}\varphi((r+m)x_k)\\
&\qquad\geq \frac{1}{|R|}\sum_{r\in R}\sum_{i\in\mathcal{J}}\sum_{k:k(r)=i}\mu(B(k))\Big[\sum_{m\in\Lambda(a)}\varphi(my_i)-|\Lambda(a)|\epsilon/2\Big]\\
&\qquad= \frac{1}{|R|}\sum_{r\in R}\sum_{i\in\mathcal{J}}\mu(B_i)\sum_{m\in\Lambda(a)}\varphi(my_i)-|\Lambda(a)|\epsilon/2\\
&\qquad= \sum_{i\in\mathcal{J}}\mu(B_i)\sum_{m\in\Lambda(a)}\varphi(my_i)-|\Lambda(a)|\epsilon/2\\
&\qquad\geq |\Lambda(a)|(\mu(\varphi)-\epsilon).
\end{aligned}\tag{5.4}$$

Notice that each $x_k \in E$ comes from at most $[(t+1)|\Lambda(a)|+1]^{|R|}$ different k's. Using this, and also (5.3), (5.4) and the concavity of the log, we obtain

$$\begin{aligned}
&|\Lambda(a)|(s(\mu)+\mu(\varphi)-\epsilon)\\
&\qquad\leq \frac{1}{|R|}\sum_k \mu(B(k))\Big[-\log\mu(B(k))+\sum_{r\in R}\sum_{m\in\Lambda(a)}\varphi((r+m)x_k)\Big]\\
&\qquad= \frac{1}{|R|}\sum_k \mu(B(k))\log\Big(\exp\Big(\sum_r\sum_m\varphi((r+m)x_k)\Big)\Big/\mu(B(k))\Big)\\
&\qquad\leq \frac{1}{|R|}\log\sum_k\exp\sum_r\sum_m\varphi((r+m)x_k)\\
&\qquad\leq \frac{1}{|R|}\log[(t+1)|\Lambda(a)|+1]^{|R|}\sum_{x\in E}\exp\sum_r\sum_m\varphi((r+m)x).
\end{aligned}$$

If $\Lambda = \bigcup_{r\in R}(\Lambda(a)+r)$ then E is (δ,Λ)-separated; therefore

$$\frac{1}{|R|}\log\sum_{x\in E}\exp\sum_r\sum_m\varphi((r+m)x)\leq|\Lambda(a)|P(\varphi,\delta,\Lambda).$$

so that

$$s(\mu)+\mu(\varphi)-\epsilon\leq P(\varphi,\delta,\Lambda)+(1/|\Lambda(a)|)\log[(t+1)|\Lambda(a)|+1].$$

By taking $|\Lambda(a)|$ large then letting $\Lambda\uparrow\infty$, this yields

$$s(\mu)+\mu(\varphi)\leq P(\varphi).\tag{5.5}$$

We show now that equality holds in (5.5) for some μ. Let $\langle u\rangle = (2^u,\dots,2^u)$ and let μ be a limit of the sequence $\mu_{\varphi,\langle u\rangle}$. Choose now a partition of $\mathcal{A}$ consisting of sets with diameter $< \delta^*$, and with boundaries of μ-measure 0. Given $\epsilon > 0$, there exists u such that $s(\mu)+\epsilon/2 > (1/|\Lambda(\langle u\rangle)|)S(\mu,\mathcal{A}^{\Lambda(\langle u\rangle)})$ and since $\mu_{\varphi,\langle v\rangle}(A(k))\to\mu(A(k))$ when $v\to\infty$, one can choose $V\geq u$ such that if $v\geq V$,

$$\begin{aligned} s(\mu) + \epsilon &> (1/|\Lambda(\langle u\rangle)|)S(\mu_{\varphi,\langle v\rangle}, \mathcal{A}^{\Lambda(\langle u\rangle)}) \\ &\geq (1/|\Lambda(\langle v\rangle)|)S(\mu_{\varphi,\langle v\rangle}, \mathcal{A}^{\Lambda(\langle v\rangle)}) \\ &\geq (1/|\Lambda(\langle v\rangle)|) \sum_{x\in\Pi_{\langle v\rangle}} \mu_{\varphi,\langle v\rangle}(\{x\}) \log \mu_{\varphi,\langle v\rangle}(\{x\}) \end{aligned}$$

where we have used the subadditivity of $\Lambda \to S(\mu, \mathcal{A}^\Lambda)$, and then expansiveness. Using the definition of $\mu_{\varphi,\langle v\rangle}$ we obtain

$$\begin{aligned} s(\mu) + \epsilon &> -\frac{1}{|\Lambda(\langle v\rangle)|} \sum_{x\in\Pi_{\langle v\rangle}} \mu_{\varphi,\langle v\rangle}(\{x\}) \left[\sum_{m\in\Lambda(\langle v\rangle)} \varphi(mx) - \log Z(\varphi, \langle v\rangle) \right] \\ &= -\mu_{\varphi,\langle v\rangle}(\varphi) + (1/|\Lambda(\langle v\rangle)|) \log Z(\varphi, \langle v\rangle) \end{aligned}$$

and the desired result follows by letting $\mu_{\varphi,\langle v\rangle} \to \mu$. We have thus proved (5.1).

Let $J_\varphi = \{\mu \in I: s(\mu) + \mu(\varphi) = P(\varphi)\}$; J_φ is the set where the affine upper semicontinuous function $\mu \to s(\mu) + \mu(\varphi)$ reaches its maximum; hence J_φ is convex and compact. If $\mu \in J_\varphi$, we have

$$\begin{aligned} P(\varphi + \psi) \geq s(\mu) + \mu(\varphi + \psi) &= s(\mu) + \mu(\varphi) + \mu(\psi) \\ &= P(\varphi) + \mu(\psi); \end{aligned}$$

hence $\mu \in I_\varphi$. Therefore $J_\varphi \subset I_\varphi$. If J_φ were different from I_φ one could find $\psi \in \mathcal{C}(\Omega)$ such that

(5.6) $$\sup_{\mu\in I_\varphi} \mu(\psi) > \sup_{\mu\in J_\varphi} \mu(\psi).$$

Let $\mu_n \in J_{\varphi+\psi/n}$ and $\mu \in I_\varphi$, we have

$$\begin{aligned} \mu(\psi) = n\mu(\psi/n) &\leq n[P(\varphi + \psi/n) - P(\varphi)] \\ &\leq n[P(\varphi + \psi/n) - s(\mu_n) - \mu_n(\varphi)] \\ &= n[\mu_n(\varphi + \psi/n) - \mu_n(\varphi)] = \mu_n(\psi). \end{aligned}$$

If μ^* is a limit point of the sequence (μ_n), then $\mu^* \in J_\varphi$ (by upper semicontinuity of s), and therefore $\mu(\psi) \leq \mu^*(\psi)$ for all $\mu \in I_\varphi$, in contradiction with (5.6). We have thus shown that $J_\varphi = I_\varphi$.

We want now to prove (5.2). We already know by (5.5) that $s(\mu) \leq P(\varphi) - \mu(\varphi)$ and it remains to show that by proper choice of φ the right-hand side becomes as close as desired to $s(\mu)$. Let $C = \{(\mu, t) \in \mathcal{C}(\Omega)^* \times \mathbf{R}: \mu \in I$ and $0 \leq t \leq s(\mu)\}$. Since s is affine upper semicontinuous, C is convex and compact. Given $\mu^* \in I$ and $u > s(\mu^*)$ there exist (because C is convex and compact) $\varphi \in \mathcal{C}(\Omega)$ and $c \in \mathbf{R}$ such that

$$-\mu^*(\varphi) + c = u, \qquad -\mu(\varphi) + c > s(\mu), \quad \text{for all } \mu \in I;$$

hence $-\mu(\varphi) + u + \mu^*(\varphi) > s(\mu)$ and we have, if $\mu \in I_\varphi$,

$$\begin{aligned} 0 &\leq P(\varphi) - s(\mu^*) - \mu^*(\varphi) \\ &= s(\mu) + \mu(\varphi) - s(\mu^*) - \mu^*(\varphi) \\ &< u - s(\mu^*). \end{aligned}$$

The right-hand side is arbitrarily small and (5.2) follows.

5.2. **Remark.** If Ω is a basic set for an Axiom A diffeomorphism it is known [3] that $0 \in D$, i.e., the maximum of $s(\mu)$ is reached for just one $\mu \in I$. Further results on D have been obtained for Anosov diffeomorphisms using methods of statistical mechanics [18].

6. **The sets of invariant states.** In this section we study the set I of all $\mathbf{Z}^\nu$-invariant probability measures and its relations with the I_φ.

6.1. **Proposition.** *For each $\varphi \in \mathcal{C}(\Omega)$, I_φ is a Choquet simplex, and a face (see* [4]) *of the simplex I.*

It is well known that the set I of invariant probability measures is a simplex.(6) If $\mu \in I_\varphi$, let m_μ be the unique probability measure on I, carried by the extremal points of I, and with resultant μ. Writing $\hat{\varphi}(\nu) = \nu(\varphi)$, we have (see [4])

$$m_\mu(s + \hat{\varphi}) = s(\mu) + \mu(\varphi) = P(\varphi);$$

hence the support of m_μ is contained in $\{\nu \in I: s(\nu) + \nu(\varphi) = P(\varphi)\} = I_\varphi$. This shows that I_φ is a simplex and a face of I.

6.2. **Proposition.** *Suppose that $\mathcal{B}$ is dense in $\mathcal{C}(\Omega)$ and is a separable Banach space with respect to a norm $|||\cdot||| \geq \|\cdot\|$. If $\varphi \in \mathcal{B}$, then I_φ is the closed convex hull of the set of μ such that*

$$\mu = \lim_{n\to\infty} \mu_{\varphi(n)}, \qquad \lim_{n\to\infty} |||\varphi(n) - \varphi||| = 0, \qquad \varphi(n) \in D \cap \mathcal{B},$$

where D is defined in Theorem 3.2(a). *This applies in particular with $\mathcal{B} = \mathcal{C}(\Omega)$.*

We have $P(\varphi(n) + \psi) \geq P(\varphi(n)) + \mu_{\varphi(n)}(\psi)$ for all ψ, hence $P(\varphi + \psi) \geq P(\varphi) + \mu(\psi)$ so that $\mu \in I_\varphi$ if μ is of the above form.

Suppose now that I_φ were not in the closed convex hull of those μ. There would then exist $\psi \in \mathcal{B}$ such that

$$\sup_{\nu \in I_\varphi} \nu(\psi) > \sup_\mu \mu(\psi). \tag{6.1}$$

Let $\varphi(n) = \varphi + \psi/n + \chi_n \in D \cap \mathcal{B}$; then, by convexity of P, if $\nu \in I_\varphi$,

$$\nu(\psi/n + \chi_n) \leq \mu_{\varphi(n)}(\psi/n + \chi_n).$$

(6) See for instance Jacobs [10, p. 162].

Using Theorem 3.2(e) we may take $|||\chi_n||| < 1/n^2$; we have thus

$$\nu(\psi) - 1/n \leq \mu_{\varphi(n)}(\psi) + 1/n,$$

and if μ^* is a limit point of $(\mu_{\varphi(n)})$, $\nu(\psi) \leq \mu^*(\psi)$ in contradiction with (6.1).

6.3. **Proposition.** *The set of measures μ on Ω such that*

$$\mu(\varphi) \leq P(\varphi) \quad \text{for all } \varphi \in \mathcal{C}(\Omega) \tag{6.2}$$

is I.

If $\mu \in I$ we have $\mu(\varphi) \leq P(\varphi) - s(\mu) \leq P(\varphi)$ because $s \geq 0$. Let now (6.2) hold for some $\mu \in \mathcal{C}(\Omega)^*$. By (2.8) we have

$$\mu(\varphi) - \mu(\tau_m\varphi) = t^{-1}\mu(t\varphi - t\tau_m\varphi) \leq t^{-1}P(t\varphi - t\tau_m\varphi) = t^{-1}P(0).$$

Letting $t \to \infty$ gives $\mu(\varphi) - \mu(\tau_m\varphi) \leq 0$. Replacing φ by $-\varphi$ yields $\mu(\varphi) = \mu(\tau_m\varphi)$. Therefore μ is $\mathbf{Z}^\nu$ invariant. Using now (2.7) and (2.8) we find

$$\pm\mu(\varphi) = \lim_{t\to\pm\infty} \frac{1}{|t|}\mu(t\varphi) \leq \lim_{t\to\pm\infty} \frac{1}{|t|}P(t\varphi)$$
$$\leq \lim_{t\to\pm\infty} \frac{1}{|t|}[P(0) + \|t\varphi\|] = \|\varphi\|$$

so that $\|\mu\| \leq 1$. Furthermore (2.8) shows that, for all t, $t\mu(1) = \mu(t) \leq P(0) + t$, so that $\mu(1) = 1$. Since $\|\mu\| = 1$ and $\mu(1) = 1$, μ is a probability measure.

6.4. **Proposition.**(7) *The set*

$$\mathcal{M}_p = \cup_a \left\{ \frac{1}{|\Lambda(a)|} \sum_{m\in\Lambda(a)} \delta_{mx} : x \in \Pi_a \right\}$$

is dense in I.

A vague neighbourhood of $\mu \in I$ is given by $\{\nu \in I: \|\nu - \mu\|_{\varphi_i} < \epsilon$ for $i = 1, \ldots, n\}$ where $\|\nu - \mu\|_{\varphi_i} = |\nu(\varphi_i) - \mu(\varphi_i)|$ and $\varphi_1, \ldots, \varphi_n \in \mathcal{C}(\Omega)$, $\epsilon > 0$. We assume without loss of generality that $\|\varphi_i\| \leq 1$ for $i = 1, \ldots, n$.

Given $\epsilon > 0$, we choose $\delta > 0$ such that $d(x,y) < \delta$ implies $|\varphi_i(x) - \varphi_i(y)| < \epsilon$ for $i = 1, \ldots, n$.

Let $p(\delta)$ be given by 1.2, $N > p(\delta)/\epsilon$ and $a = (N, \ldots, N)$, $b = (N + p(\delta), \ldots, N + p(\delta))$. By the density of measures with finite support we can choose $c_\alpha > 0$, $x_\alpha \in \Omega$ such that

$$\sum_\alpha c_\alpha = 1, \qquad \left\| \sum_\alpha c_\alpha \delta_{x_\alpha} - \mu \right\|_{\tau_m\varphi_i} < \epsilon,$$

(7) Sigmund [15] has proved this result by a somewhat different method for $\nu = 1$.

for $i = 1, \ldots, n$, and $m \in \Lambda(b)$. We have thus

$$\left\| \sum_\alpha c_\alpha \delta_{mx_\alpha} - \mu \right\|_{\varphi_i} < \epsilon \quad \text{for } m \in \Lambda(b);$$

hence

$$\left\| \frac{1}{|\Lambda(b)|} \sum_{m \in \Lambda(b)} \sum_\alpha c_\alpha \delta_{mx_\alpha} - \mu \right\|_{\varphi_i} < \epsilon. \tag{6.3}$$

By 1.3, we can choose $y_\alpha \in \Pi_b$ such that $|\varphi_i(mx_\alpha) - \varphi_i(my_\alpha)| < \epsilon$ for $m \in \Lambda(a)$, and we have $|\varphi_i(mx_\alpha) - \varphi_i(my_\alpha)| \leq 2$ for $m \in \Lambda(b)\backslash\Lambda(a)$; hence

$$\begin{aligned} &\left\| \sum_{m \in \Lambda(b)} \sum_\alpha \frac{c_\alpha}{|\Lambda(b)|} \delta_{my_\alpha} - \sum_{m \in \Lambda(b)} \sum_\alpha \frac{c_\alpha}{|\Lambda(b)|} \delta_{mx_\alpha} \right\|_{\varphi_i} \\ &\quad < \epsilon \frac{|\Lambda(a)|}{|\Lambda(b)|} + 2 \frac{|\Lambda(b)| - |\Lambda(a)|}{|\Lambda(b)|} < \epsilon + 2(1+\epsilon)^\nu - 2. \end{aligned} \tag{6.4}$$

We can now find integers P, $M_\alpha > 0$ such that $\sum_\alpha M_\alpha = P^\nu$ and

$$\left\| \sum_\alpha \frac{M_\alpha}{|\Lambda(b)| P^\nu} \sum_{m \in \Lambda(b)} \delta_{my_\alpha} - \sum_{m \in \Lambda(b)} \sum_\alpha \frac{c_\alpha}{|\Lambda(b)|} \delta_{my_\alpha} \right\|_{\varphi_i} < \epsilon. \tag{6.5}$$

Let $c = ((N + p(\delta))P, \ldots, (N + p(\delta))P)$. By application of (1.3), there exists $y \in \Pi_c$ such that when $\tilde{m}$ varies over $\Lambda(c)$, my takes M_α times a value close to my_α for each α and each $m \in \Lambda(a)$. Close means $d(\tilde{m}y, my_\alpha) < \delta$. Then

$$\begin{aligned} &\left\| \frac{1}{|\Lambda(c)|} \sum_{\tilde{m} \in \Lambda(c)} \delta_{\tilde{m}y} - \frac{1}{|\Lambda(b)| P^\nu} \sum_\alpha M_\alpha \sum_{m \in \Lambda(b)} \delta_{my_\alpha} \right\|_{\varphi_i} \\ &\quad \leq \epsilon \frac{|\Lambda(a)|}{|\Lambda(b)|} + 2 \frac{|\Lambda(b)| - |\Lambda(a)|}{|\Lambda(b)|} < \epsilon + 2(1+\epsilon)^\nu - 2. \end{aligned} \tag{6.6}$$

Finally, (6.3), (6.4), (6.5), (6.6) give

$$\left\| \frac{1}{|\Lambda(c)|} \sum_{\tilde{m} \in \Lambda(c)} \delta_{\tilde{m}y} - \mu \right\|_{\varphi_i} < 4\epsilon + 4(1+\epsilon)^\nu - 4,$$

proving the proposition.

6.5. **Proposition.**(8) (a) *The set of ergodic measures* (*extremal points of I*) *is residual in I.*

(b) *The set of measures with zero entropy is residual in I.*

Since $\mathcal{M}_p$ is dense (Proposition 6.4) and consists of ergodic measures with zero entropy, it suffices to show that the set of ergodic measures and the set of measures with zero entropy are G_δ (i.e. countable intersections of open sets). For ergodic measures this is well known (see [4]); for measures with zero entropy, it follows from the fact that the entropy is upper semicontinuous.

(8) See Sigmund [15] where other residual sets are also discussed.

Added in proof. A proof of the variational principle (0, 1) has been obtained without the expansiveness and specification assumptions by P. Walters (preprint).

REFERENCES

1. R. L. Adler, A. G. Konheim and M. H. McAndrew, *Topological entropy*, Trans. Amer. Math. Soc. **114** (1965), 309–319. MR **30** #5291.

2. R. Bowen, *Periodic points and measures for Axiom* A *diffeomorphisms*, Trans. Amer. Math. Soc. **154** (1971), 377–397. MR **43** #8084.

3.———, *Markov partitions for Axiom* A *diffeomorphisms*, Amer. J. Math. **92** (1970), 725–747. MR **43** #2740.

4. G. Choquet and P. A. Meyer, *Existence et unicité des représentations intégrales dans les convexes compacts quelconques*, Ann. Inst. Fourier (Grenoble) **13** (1963), fasc. 1, 139–154. MR **26** #6748; MR **30** #1203.

5. E. I. Dinaburg, *A correlation between topological entropy and metric entropy*, Dokl. Akad. Nauk SSSR **190** (1970), 19–22 = Soviet Math. Dokl. **11** (1970), 13–16. MR **41** #425.

6. N. Dunford and J. T. Schwartz, *Linear operators*. I. *General theory*, Pure and Appl. Math., vol. 7, Interscience, New York, 1958. MR **22** #8302.

7. G. Gallavotti and S. Miracle-Sole, *Statistical mechanics of lattice systems*, Comm. Math. Phys. **5** (1967), 317–323. MR **36** #1173.

8. L. Goodwyn, *Topological entropy bounds measure-theoretic entropy*, Proc. Amer. Math. Soc. **23** (1969), 679–688. MR **40** #299.

9. W. H. Gottschalk and G. A. Hedlund, *Topological dynamics*, Amer. Math. Soc. Colloq. Publ., vol 36, Amer. Math. Soc., Providence, R.I., 1955. MR **17**, 650.

10. K. Jacobs, *Lecture notes on ergodic theory*. I, II, Mat. Inst., Aarhus Univ., Aarhus, 1963, pp. 1–207, 208–505. MR **28** #1247; #3138.

11. O. E. Lanford and D. W. Robinson, *Statistical mechanics of quantum spin systems*. III, Comm. Math. Phys. **9** (1968), 327–338. MR **38** #3012.

12. D. W. Robinson and D. Ruelle, *Mean entropy of states in classical statistical mechanics*, Comm. Math. Phys. **5** (1967), 288–300. MR **37** #1146.

13. D. Ruelle, *A variational formulation of equilibrium statistical mechanics and the Gibbs phase rule*, Comm. Math. Phys. **5** (1967), 324–329. MR **36** #699.

14.———, *Statistical mechanics. Rigorous results*, Benjamin, New York, 1969. MR **44** #6279.

15. K. Sigmund, *Generic properties of invariant measures for Axiom* A *diffeomorphisms*, Invent. Math. **11** (1970), 99–109. MR **44** #3349.

16. Ja. G. Sinaĭ, *Markov partitions and Y-diffeomorphisms*, Funkcional Anal. i Priložen. **2** (1968), no.1, 64–89 = Functional Anal. Appl. **2** (1968), 61–82. MR **38** #1361.

17.———, *Construction of Markov partitionings*, Funkcional. Anal. i Priložen. **2** (1968), no. 3, 70–80. (Russian) MR **40** #3591.

18.———, *Invariant measures for Anosov's dynamical systems*, Proc. Internat. Congress Math. (Nice, 1970), vol. 2, Gauthier-Villars, Paris, 1971, pp. 929–940.

19. S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817. MR **37** #3598.

20. M. Smorodinsky, *Ergodic theory, entropy*, Lecture Notes in Math., vol. 214, Springer-Verlag, Berlin, 1971.

INSTITUT DES HAUTES ETUDES SCIENTIFIQUES, BURES-SUR-YVETTE, FRANCE