CHAPTER 2
ERGODIC THEORY ON COMPACT METRIC SPACES

Dec.5, 1999

In the previous Chapter a measure preserved by a measurable map T was given *a priori*. Here a continuous mapping T of a topological compact space is given and we look for various measures preserved by T. Given a real continuous function ϕ on X we try to maximize the functional *measure theoretical entropy* $+$*integral*, i.e. $\mathrm{h}_\mu(T) + \int \phi \, d\mu$. Supremum over all probability measures on the Borel σ-algebra happens to be *topological pressure*, similar to P in the prototype lemma on the finite space or $P(\alpha)$ for ϕ_α on the Cantor set discussed in Introduction. We discuss equilibria, namely measures on which supremum is attained. This Chapter provides an introduction to the theory called *thermodynamical formalism*, which will be the main technical tool in this book. We shall continue to introduce thermodynamical formalism in more specific situations in Chapter 4.

§2.1 INVARIANT MEASURES FOR CONTINUOUS MAPPINGS

We recall in this Section basic facts from functional analysis to study the space of measures and invariant measures. We recall Riesz representation theorem, weak* topology, Schauder fixed point theorem. We recall also Krein-Milman theorem on extremal points and its stronger form: Choquet representation theorem. This gives a variant of Ergodic Decomposition Theorem from Chapter 1.

Let X be a topological space. The *Borel* σ-algebra $\mathcal{B}$ in subsets of X is defined as generated by open subsets of X. We call every probability measure on Borel σ-algebra for X, a *Borel probability measure* on X. We denote the set of all such measures by $M(X)$.

Denote by $C(X)$ the Banach space of real continuous functions on X with the supremum norm: $|\phi| := \sup_{x \in X} |\phi(x)|$. Sometimes we shall use the notation $||\phi||_\infty$, introduced in Ch.1.1 in $L^\infty(\mu)$, though it is compatible only if μ is positive on open sets.

Note that each Borel probability measure μ on X induces a bounded linear functional F_μ on $C(X)$ defined by the formula

$$F_\mu(\phi) = \int \phi \, d\mu. \tag{2.1.1}$$

One can extend the notion of measure and consider σ-additive set function, another name : *signed measure*. Just in definition of measure in Ch.1.1 consider $\mu : \mathcal{F} \to [-\infty, \infty)$

or $\mu : \mathcal{F} \to (-\infty, \infty]$ and keep the notation $(X, \mathcal{F}, \mu)$ from Ch.1. The set of signed measures is a linear space. On the set of finite signed measures, namely with the range $I\!R$, one can introduce the following total variation norm:

$$v(\mu) := \sup \sum_{i=1}^{n} |\mu(A_i)|$$

supremum taken over all finite sequences of disjoint sets in $\mathcal{F}$.

It is easy to prove that every finite signed measure is bounded and that it has finite total variation. It is also not difficult to prove the following

Theorem 2.1.1. (Hahn-Jordan decomposition). For every signed measure on a σ-algebra $\mathcal{F}$ there exists $A_\mu \in \mathcal{F}$ and two measures μ^+ and μ^- such that $\mu = \mu^+ - \mu^-$, μ^- is zero on all measurable subsets of A_μ, μ^- is zero on all measurable subsets of $X \setminus A_\mu$.

Notice that $v(\mu) = \mu^+(X) + \mu^-(X)$.)

A measure (or signed measure) is called *regular*, if for every $A \in \mathcal{F}$ and $\varepsilon > 0$ there exist $E_1, E_2 \in \mathcal{F}$ such that $\mathrm{cl}E_1 \subset A \subset \mathrm{int}E_2$ and for every $C \in \mathcal{F}$, $C \subset E_2 \setminus E_1$ we have $|\mu(C)| < \varepsilon$.

If X is a topological space, denote the space of all regular finite signed measures as above with the total variation norm by $\mathrm{rca}(X)$. The abbreviation rca replaces *regular countably additive*.

If $\mathcal{F} = \mathcal{B}$ Borel σ-algebra and X is metrizable, regularity holds for every finite signed measure. It can be proved by Carathéodory's outer measure argument, compare Proof of Corollary 1.8.7"'.

Denote by $C(X)^*$ the space of all bounded linear functionals on $C(X)$. This is called the *dual space*. *Bounded* means here bounded on the unit ball in $C(X)$, which is equivalent to *continuous*. The space $C(X)^*$ is equipped with the norm $||F|| = \sup\{F(\phi) : \phi \in C(X), |\phi| \leq 1\}$, in which it is a Banach space.

There is a natural order in $\mathrm{rca}(X)$: $\nu_1 \leq \nu_2$ iff $\nu_2 - \nu_1$ is a measure.

Also in the space $C(X)^*$ one can distinguish positive functionals, similarly to measures in signed measures, as those which are non-negative on the set of functions $C^+(X) := \{\phi \in C(X) : \phi(x) \geq 0$ for every $x \in X\}$. This gives the order: $F \leq G$ for $F, G \in C(X)^*$ iff $G - F$ is positive.

Remark that $F \in C(X)^*$ is positive iff $||F|| = F(1\!\!1)$, where $1\!\!1$ is the function on X identically equal to 1. Also for every bounded linear operator $F : C\!\!\!C(X) \to C(X)$ which is positive, namely $F(C^+(X)) \subset C^+(X)$, we have $||F|| = |F(1\!\!1)|$.

Remark that (2.1.1) transforms measures to positive linear functionals.

The following fundamental theorem of F. Riesz says more about the transformation $\mu \mapsto F_\mu$ in (2.1.1) (see [DS, pp. 373,380] for the history of this theorem):

Theorem 2.1.2 (Riesz representation theorem). If X is a compact Hausdorff space, the transformation $\mu \mapsto F_\mu$ defined by (2.1.1) is an isometric isomorphism between the Banach space $C(X)^*$ and $\mathrm{rca}(X)$. Furthermore this isomorphism preserves order.

In the sequel we shall often write μ instead of F_μ and *vice versa* and $\mu(\phi)$ or $\mu\phi$ instead of $F_\mu(\phi)$ or $\int \phi \, d\mu$.

Notice that in Theorem 2.1.2 the hard part is the existence, i.e. that for every $F \in C(X)^*$ there exists $\mu \in \mathrm{rca}(X)$ such that $F = F_\mu$. The uniqueness is just the following:

Lemma 2.1.3. If μ and ν are finite regular Borel signed measures on a compact Hausdorff space X, such that $\int \phi \, d\mu = \int \phi \, d\nu$ for each $\phi \in C(X)$, then $\mu = \nu$.

Proof. This is an exercise on the use the regularity of μ and ν. Let $\eta := \mu - \nu = \eta^+ - \eta^-$ in Hahn-Jordan decomposition. Suppose that that $\mu \neq \nu$. Then η^+ (or η^-) is non-zero, say $\eta^+(X) = \eta^+(A_\eta) = \varepsilon > 0$, where A_η is the set defined in Th.2.1.1. Let E_1 be a closed set and E_2 an open set, such that $E_1 \subset A_\eta \subset E_2$ and $\eta^-(E_2 \backslash A_\eta) < \varepsilon/3$, $\eta^+(A_\eta \backslash E_1) < \varepsilon/3$. There exists $\phi \in C(X)$ with values in $[0,1]$ identically 1 on E_1 and 0 on $X \setminus E_2$. Then $\int \phi \, d\eta = \int_{E_1} \phi \, d\eta + \int_{E_2 \backslash A_\eta} \phi \, d\eta + \int_{A_\eta \backslash E_1} \phi \, d\eta \geq \eta^+(E_1) - \varepsilon/3 \geq \eta^+(A_\eta) - 2\varepsilon/3 > 0$. ♣

The space $C(X)^*$ can be equipped with weak* topology. In the case X is *metrizable*, i.e. if there exists a metric on X such that the topology induced by this metric is the original topology on X, weak* topology is characterized by the property that a sequence $\{F_n : n = 1, 2, \ldots\}$ of functionals in $C(X)^*$ converges to a functional $F \in C(X)^*$ if and only if

$$\lim_{n \to \infty} F_n(\phi) = F(\phi) \tag{2.1.2}$$

for every function $\phi \in C(X)$.

If we do not assume X is metrizable, weak* topology is defined as the smallest one in which all elements of $C(X)$ are continuous on $C(X)^*$ (recall that $\phi \in C(X)$ acts on $F \in C(X)^*$ by $F(f)$). One says *weak** to distinguish this topology from the *weak topology* where one considers all continuous functionals on $C(X)^*$, not only those represented by $f \in C(X)$. This discussion of topologies concerns of course every Banach space B and its dual B^*.

Using the bijection established by Riesz representation theorem we can move the weak* topology from $C(X)^*$ to $\mathrm{rca}(X)$ and restrict it to $M(X)$. The topology on $M(X)$ obtained in this way is usually called *weak* topology* on the space of probability measures (sometimes one omits * to simplify the language and notation but one still has in mind weak*, unless stated otherwise). In view of (2.1.2) if X is metrizable this topology is characterized by the property that a sequence $\{\mu_n : n = 1, 2, \ldots\}$ of measures in $M(X)$ converges to a measure $\mu \in M(X)$ if and only if

$$\lim_{n\to\infty} \mu_n(\phi) = \mu(\phi) \tag{2.1.3}$$

for every function $\phi \in C(X)$. Such convergence of measures will be called *weak* convergence* or *weak convergence* and can be also characterized as follows.

Theorem 2.1.4 Suppose that X is metrizable (we do not assume compactness here). A sequence $\{\mu_n : n = 1, 2, \ldots\}$, of Borel probability measures on X converges weakly to a measure μ if and only if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for every Borel set A such that $\mu(\partial A) = 0$.

Proof. By $\mu(\partial A) = 0$, there exist sets $E_1 \subset \mathrm{int}A, E_2 \supset \mathrm{cl}A$ such that $\mu(E_2 \setminus E_1) = \varepsilon$ is arbitrarily small. Indeed metrizability of X implies that every open set, in particular $\mathrm{int}A$, is union of a sequence of closed sets and every closed set is an intersection of a sequence of open sets. For example $\mathrm{int}A = \bigcup_{n=1}^{\infty}\{x \in X : \inf_{z\notin \mathrm{int}A} \rho(x,z) \geq 1/n\}$ for a metric ρ.

Next, there exist $f, g \in C(X)$ such that f is identically 1 on E_1, 0 on $X \setminus \mathrm{int}A$, g identically 1 on $X \setminus \mathrm{cl}A$ and 0 on $\mathrm{cl}A$. Then $\mu_n(f) \to \mu(f)$ and $\mu_n(g) \to \mu(g)$ As $\mu(E_1) \leq \mu(f) \leq \mu(g) \leq \mu(E_2)$ and $\mu_n(f) \leq \mu_n(A) \leq \mu_n(g)$ we obtain

$$\mu(E_1) \leq \lim_{n\to\infty} \mu_n(f) \leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \leq \lim_{n\to\infty} \mu_n(g) \leq \mu(E_2).$$

As also $\mu(E_1) \leq \mu(A) \leq \mu(E_2)$ we obtain, letting $\varepsilon \to 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

Proof in the opposite direction follows from the definition of integral: approximate an arbitrary continuous function f by simple functions $\sum_{i=1}^{k} \alpha_i 1\!\!1_{E_i}$ where $E_i = \{x \in X : \varepsilon_i \leq f(x) < \varepsilon_{i+1}\}$, for an increasing sequence $\varepsilon_i, i = 1, ..., k$ such that $\varepsilon_i - \varepsilon_{i-1} < \varepsilon$ and $\mu(f^{-1}(\{\varepsilon_i\})) = 0$, with $\varepsilon \to 0$. This is possible to find such numbers ε_i because only countably many sets $f^{-1}(a)$ for $a \in I\!\!R$ can have non-zero measure. ♣

Example 2.1.5 The assumption $\mu(\partial A) = 0$ is substantial. Denote by δ_x the Dirac measure concentrated at the point x, which is defined by the following formula:

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases} \quad A \in \mathcal{B}.$$

Consider non-atomic measures μ_n each supported on the ball $B(x, \frac{1}{n})$, (X is a metric space here). μ_n converge weakly to δ_x but do not converge on $\{x\}$.

Of particular importance is the following
Theorem 2.1.6. The space $M(X)$ is compact in weak* topology.

This theorem follows immediately from compactness in weak* topology of any subset of $C(X)^*$ closed in weak* topology, which is bounded in the standard norm of the dual space

$C(X)^*$ (compare for example [DS, V.4.3], where this result is proved for all spaces dual to Banach spaces) and from the way we introduced the weak topology on $M(X)$.

It turns out (see [DS, V.5.1]) that if X is compact metrizable, the space $C(X)^*$ with weak* topology is metrizable, hence in particular $M(X)$ is metrizable.

Let now $T : X \to X$ be a continuous transformation of X. The mapping T is measurable with respect to the Borel σ-algebra. In the very begining of Chap.1.2 we have defined T-invariant meaures μ to satisfy the condition $\mu = \mu \circ T^{-1}$. It means that Borel probability T-invariant meaures are exactly fixed points of the transformation $T_* : M(X) \to M(X)$ defined by the formula $T_*(\mu) = \mu \circ T^{-1}$. It easily follows from the definitions that T_* is continuous.

We denote the set of all T-invariant measures in $M(X)$ by $M(X,T)$. This notation is consistent with the notation from Chapter 1.2. We omit here σ-algebra $\mathcal{F}$ because it is always Borel $\mathcal{B}$.

Noting that $\int \phi \, d(\mu \circ T^{-1}) = \int \phi \circ T \, d\mu$ for any $\mu \in M(X)$ and any integrable function ϕ (Prop. 1.2.0), it follows from Lemma 2.1.3 that a Borel probability measure μ is T-invariant if and only if for every continuous function $\phi : X \to I\!R$

$$\int \phi \, d\mu = \int \phi \circ T \, d\mu. \tag{2.1.4}$$

In order to look for fixed points for T_* one can apply the following very general result whose proof (and the definition of locally convex topological vector spaces, abbreviation: LCTVS) can be found for example in [DS] or [Edwards].

Theorem 2.1.7. (Schauder-Tychonoff theorem [DS, V.10.5) If K is a non-empty compact convex subset of an LCTVS then any continuous transformation $H : K \to K$ has a fixed point.

Assume from now on that X is compact, metrizable. To apply Schauder-Tychonoff theorem consider the LCTVS $C(X)^*$ with weak* topology and $K \subset C(X)^*$, being the image of $M(X)$ under the identification between measures and functionals, given by Riesz representation theorem. Move also T_* to K with the use of this identification. Note that the resulting continuous linear operator, denote it also by T_*, conjugate to $\phi \mapsto \phi \circ T$, restricted to K, is continuous also in the weak* topology. This is an easy fact about conjugate operators. We obtain

Theorem 2.1.8. (Bogolyubov–Krylov theorem [Walters, 6.9.1]) There exists a Borel probability measure μ invariant under T.

Thus, our $M(X,T)$ is non-empty. It is also weak* compact, since it is closed as the set of fixed points for a continuous transformation.

As an immediate consequence of this theorem and Theorem 1.8.8 (Ergodic Decomposition Theorem), we get the following:

Corollary 2.1.9. There exists a Borel ergodic probability measure μ invariant under T.

We shall use the notation $M_e(X,T)$ for the set of all ergodic measures in $M(X,T)$. Write also $\mathcal{E}(M(X,T))$ for the set of extreme points in $M(X,T)$.

Thus, in view of Theorem 1.2.5 and Corollary 2.1.9, we know that $M_e(X,T) = \mathcal{E}(M(X,T)) \neq \emptyset$.

In fact Corollary 2.1.9 can be obtained in a more elementery way without using Theorem 1.8.8. Namely it now immediately follows from Theorem 1.2.5 and the following

Theorem 2.1.10. (Krein–Milman theorem on extremal points [DS, V.8.4]) If K is a nonempty compact convex subset of an LCTVS then the set $\mathcal{E}(K)$ of extreme points of K is nonempty and moreover K is the closure of the convex hull of $\mathcal{E}(K)$.

Below we state Choquet representation theorem which is stronger than Krein-Milman theorem. It corresponds to the Ergodic Decomposition Theorem (Th. 1.8.8). We formulate it in $C(X)^*$ with weak* topology as in [Walters, p.153]. The reader can find a general LCTVS version for example in [Edwards]. We rely here also on [Ruelle, Appendix A.5], where the reader can find further references.

Theorem 2.1.11. Choquet representation theorem. Let K be a nonempty compact convex set in $M(X)$ with weak* topology. Then for every $\mu \in K$ there exists a "mass distribution" i.e. measure $\alpha_\mu \in M(\mathcal{E}(K))$ such that

$$\mu = \int m \, d\alpha_\mu(m).$$

This integral converges in weak* topology which means that for every $f \in C(X)$

$$\mu(f) = \int m(f) \, d\alpha_\mu(m). \tag{2.1.5}$$

Notice that we have had already the formula analogous to (2.1.5) in Theorem 2.1.6.

Notice that Krein-Milman theorem follows from Choquet representation theorem because one can weakly approximate α_μ by measures on $\mathcal{E}(K)$ with finite support (finite linear combinations of Dirac measures).

Example. 2.1.12. For $K = M(X)$ we have $\mathcal{E}(K) = \{\text{Dirac measures on } X\}$. Then $\alpha_\mu\{\delta_x : x \in A\} = \mu(A)$ for every $A \in \mathcal{B}$ defines a Choquet representation for every $\mu \in M(X)$. (Exercise)

Choquet theorem asserts the existence of α_μ satisfying (2.1.5) but not uniqueness, which is usually not true. A compact closed set K with the uniqueness of α_μ satisfying (2.1.5), for every $\mu \in M(K)$ is called *symplex.*

Theorem 2.1.13. $K = M(X)$ or $K = M(X,T)$ for every continuous $T : X \to X$ is a symplex.

Proof in the case of $K = M(X)$ is very easy, see Example 2.1.12. A proof for $K = M(X,T)$ is also not hard. The reader can look in [Ruelle, A.5.5]. Proof relies on the fact that two different measures $\mu_1, \mu_2 \in \mathcal{E}(M(X,T))$ are singular (see Theorem 1.2.3). Observe that $||\mu_1 - \mu_2|| = 2$.

One proves in fact that for every $\mu_1, \mu_2 \in M(X,T)$, $||\alpha_{\mu_1} - \alpha_{\mu_2}|| = ||\mu_1 - \mu_2||$.

Let us go back to Schauder-Tychonoff theorem (Th 2.1.7). We shall use it in this book later, in Chapter 4 Sec.2, for maps different from T_*. Just Bogolyubov-Krylov theorem proved above with the help of Theorem 2.1.7, has a different much simpler proof due to the fact that T_* is affine. A general theorem on the existence of a fixed point for a family of commuting continuous affine maps on K is called Markov-Kakutani theorem , [DS, V.10.6], [Walters, 6.9]).

2.1.14. An alternative proof of Theorem 2.1.8. Take an arbitrary $\nu \in M(X)$ and consider the sequence

$$\mu_n = \mu_n(\nu) = \frac{1}{n} \sum_{j=0}^{n-1} T_*^j(\mu)$$

In view of Theorem 2.1.4 it has a weakly convergent subsequence, say $\{\mu_{n_k} : k = 1, 2, \ldots\}$. Denote its limit by μ. We shall show that μ is T-invariant.

We have

$$T_*(\mu_{n_k}) = T_*(\frac{1}{n_k} \sum_{j=0}^{n_k-1} T_*^j(\nu)) = (\frac{1}{n_k} \sum_{j=0}^{n_k-1} T_*^{j+1}(\nu))$$

So for every $\phi \in C(X)$ we have

$$|\mu(\phi) - T_*(\mu(\phi))| = |\lim_{k\to\infty} \mu_{n_k}(\phi) - T_*(\mu_{n_k})(\phi)| \le$$

$$\lim_{k\to\infty} \frac{1}{n_k} |\nu(\phi) - T_*^{n_k}(\nu)(\phi)| \le \lim_{k\to\infty} \frac{2}{n_k} |\phi| = 0.$$

This in view of Lemma 2.1.3 finishes the proof. ♣

Remark. If in the above proof we consider $\nu = \delta_x$, a Dirac measure, then $T_*^j(\delta_x) = \delta_{T^j(x)}$ and $\mu_n(\phi) = \frac{1}{n}\sum_{j=0}^{n-1} \phi(T^j(x))$. If we have *a priori* $\mu \in M(X,T)$ then

$$\mu_n(\delta_x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j(x)}$$

is weakly convergent for μ-a.e.$x \in X$ by Birkhoff ergodic theorem.

Remark. Recall that in Birkhof ergodic theorem (Chapter 1), for $\mu \in M(X,T)$ for every integrable f one considers $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \phi(T^j(x))$ for a.e. x. This "almost every" depends on f. If X is compact, as in this Chapter, one can reverse the order of quantificators for continuous functions.

Namely there exists $\Lambda \in \mathcal{B}$ such that $\mu(\Lambda) = 1$ and for every $f \in C(X)$ and $x \in \Lambda$ the limit $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \phi(T^j(x))$ exists.

Remark. We could take in 2.1.14 an arbitrary sequence $\nu_n \in M(X)$ and take $\mu_n := \mu_n(\nu_n)$. This gives a general method of constructing measures in $M(X,T)$, see for example Proof of Variational principle in Section 4. (This point of view is taken from [Walters]).

We end this Section with the following useful later

Lemma 2.1.15. For every finite partition P of the space $(X, \mathcal{B}, \mu)$ where X is as above a metrizable compact space $\mathcal{B}$ is Borel σ-algebra and $\mu \in M(X,T)$, such that $\sum_{A\in\mathcal{A}} \mu(\partial A) = 0$, then the entropy $\mathrm{H}_\nu(\mathrm{P})$ is a continuous function of $\nu \in M(X,T)$ at μ. The entropy $\mathrm{h}_\nu(T,\mathrm{P})$ is upper semicontinuous at μ.

Proof. The continuity of $H_\nu(\mathrm{P})$ follows immediately from Theorem 2.1.4. This applied to the partitions $\bigvee_{i=1}^{n-1} T^{-i}\mathrm{P}$ gives the upper semicontinuity of $\mathrm{h}_\nu(T,\mathrm{P})$ as limit of a decreasing sequence of continuous functions $\frac{1}{n}H_\nu(\bigvee_{i=1}^{n-1} T^{-i}\mathrm{P})$. See Lemma 1.4.3. ♣

§2.2 TOPOLOGICAL PRESSURE AND TOPOLOGICAL ENTROPY

This section is of topological character and no measure is involved. We introduce and examine here some basic topological invariants coming from thermodynamic formalism of statistical physics.

Let $\mathcal{U} = \{A_i\}_{i\in I}$ and $\mathcal{V} = \{B_j\}_{j\in J}$ be two covers of the compact metric space X considered in the previous section. We define the new cover $\mathcal{U} \vee \mathcal{V}$ putting

$$\mathcal{U} \vee \mathcal{V} = \{A_i \cap B_j : \ i \in I,\ j \in J\} \tag{2.2.1}$$

and we write

$$\mathcal{U} \prec \mathcal{V} \iff \forall_{j\in J}\ \exists_{i\in I} \quad B_j \subset A_i \tag{2.2.2}$$

Let, as in the previous section, $T : X \to X$ be a continuous transformation of X. Let $\phi : X \to I\!R$ be a continuous function, frequently called it potential and let $\mathcal{U}$ be a finite, open cover of X. For every integer $n \ge 1$ we set

$$\mathcal{U}^n = \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee \ldots \vee T^{-(n-1)}(\mathcal{U}),$$

for every set $Y \subset X$

$$S_n\phi(Y) = \sup\{\sum_{k=0}^{n-1} \phi \circ T^k(x) : x \in Y\}$$

and for every $n \ge 1$

$$Z_n(\phi, \mathcal{U}) = \inf_{\mathcal{V}}\{\sum_{U\in\mathcal{V}} \exp S_n\phi(U)\} \tag{2.2.3}$$

where $\mathcal{V}$ ranges over all covers of X contained (in the sense of inclusion) in $\mathcal{U}^n$. The quantity $Z_n(\phi, \mathcal{U})$ is sometimes called the partition function.

Lemma 2.2.1. The limit $\mathrm{P}(\phi, \mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\phi, \mathcal{U})$ exists and moreover it is finite. In fact $\mathrm{P}(\phi, \mathcal{U}) \ge -||\phi||_\infty$.

Proof. Fix $m, n \ge 1$ and consider arbitrary covers $\mathcal{V} \subset \mathcal{U}^m$, $\mathcal{G} \subset \mathcal{U}^{\backslash}$ of X. If $U \in \mathcal{V}$ and $V \in \mathcal{G}$ then

$$S_{m+n}\phi(U \cap T^{-m}(V)) \le S_m\phi(U) + S_n\phi(V)$$

and thus

$$\exp\big(S_{m+n}\phi(U \cap T^{-m}(V))\big) \le \exp S_m\phi(U) \exp S_n\phi(V)$$

Since $U \cap T^{-m}(V) \in \mathcal{V} \vee T^{-m}(\mathcal{G}) \subset \mathcal{U}^{\Updownarrow} \vee \mathcal{T}^{-\Updownarrow}(\mathcal{U}^{\backslash}) = \mathcal{U}^{\Updownarrow+\backslash}$, we therefore obtain

$$Z_{m+n}(\phi, \mathcal{U}) \le \sum_{U\in\mathcal{V}} \sum_{V\in\mathcal{G}} \exp\big(S_{m+n} f(U \cap T^{-m}(V))\big) \le \sum_{U\in\mathcal{V}} \sum_{V\in\mathcal{G}} \exp S_m\phi(U) \exp S_n\phi(V)$$

$$= \sum_{U\in\mathcal{V}} \exp S_m\phi(U) \times \sum_{V\in\mathcal{G}} \exp S_n\phi(V) \tag{2.2.4}$$

Ranging now over all $\mathcal{V}$ and $\mathcal{G}$ as specified in definition (2.2.3) we get $Z_{m+n}(\phi,\mathcal{U}) \le Z_m(\phi,\mathcal{U}) \cdot Z_n(\phi,\mathcal{U})$ which implies that

$$\log Z_{m+n}(\phi,\mathcal{U}) \le \log Z_m(\phi,\mathcal{U}) + \log Z_n(\phi,\mathcal{U}).$$

Moreover, $Z_n(\phi,\mathcal{U}) \ge \exp(-n||\phi||_\infty)$. So, $\log Z_n(\phi,\mathcal{U}) \ge -n||\phi||_\infty$ and applying now Lemma 1.4.1 finishes the proof. ♣

Notice that, although in the notation $\mathrm{P}(\phi,\mathcal{U})$, the transformation T does not directly appear, however this quantity depends obviously also on T. If we want to indicate this dependence we write $\mathrm{P}(T,\phi,\mathcal{U})$ and similarly $Z_n(T,\phi,\mathcal{U})$ for $Z_n(\phi,\mathcal{U})$. Given an open cover $\mathcal{V}$ of X let

$$\mathrm{osc}(\phi,\mathcal{V}) = \sup_{V\in\mathcal{V}}\big(\sup\{|\phi(x)-\phi(y)| : x,y\in V\}\big).$$

Lemma 2.2.2. If $\mathcal{U}$ and $\mathcal{V}$ are finite open covers of X such that $\mathcal{U} \succ \mathcal{V}$, then $\mathrm{P}(\phi,\mathcal{U}) \ge \mathrm{P}(\phi,\mathcal{V}) - \mathrm{osc}$.

Proof. Take $U \in \mathcal{U}^n$. Then there exists $V = i(U) \in \mathcal{V}^n$ such that $U \subset V$. For every $x, y \in V$ we have $|S_n\phi(x) - S_n\phi(y)| \le \mathrm{osc} n$ and therefore

$$S_n\phi(U) \ge S_n\phi(V) - \mathrm{osc}(\phi,\mathcal{V})n \tag{2.2.5}$$

Let now $\mathcal{G} \subset \mathcal{U}_\setminus$ be a cover of X and let $\tilde{\mathcal{G}} = \{i(U) : U \in \mathcal{U}_n\}$. The family $\tilde{\mathcal{G}}$ is also an open finite cover of X and $\tilde{\mathcal{G}} \subset \mathcal{V}_n$. In view of (2.2.5) and (2.2.3)

$$\sum_{U\in\mathcal{G}} \exp S_n\phi(U) \ge \sum_{V\in\tilde{\mathcal{G}}} \exp S_n\phi(V)e^{-\mathrm{osc} n} \ge e^{-\mathrm{osc}(\phi,\mathcal{V})n} Z_n(\phi,\mathcal{V})$$

Therefore applying (2.2.3) again, we get $Z_n(\phi,\mathcal{U}) \ge \exp(-\mathrm{osc}(\phi,\mathcal{V})n)Z_n(\phi,\mathcal{V})$. Hence $\mathrm{P}(\phi,\mathcal{U}) \ge \mathrm{P}(\phi,\mathcal{V}) - \mathrm{osc}(\phi,\mathcal{V})$.

Definition 2.2.3. Consider now the family of all sequences $\{\mathcal{V}_n : n = 1,2,\dots\}$ of open finite covers of X such that

$$\lim_{n\to\infty} \mathrm{diam}(\mathcal{V}_n) = 0 \tag{2.2.6}$$

and define the topological pressure $\mathrm{P}(T,\phi)$ as the supremum of upper limits

$$\limsup_{n\to\infty} \mathrm{P}(\phi,\mathcal{V}_n)$$

taken over all such sequences. Notice that by Lemma 2.2.1, $\mathrm{P}(T,\phi) \ge -||\phi||_\infty$.

The following lemma gives us a simpler way to calculate topological pressure showing that in fact in its definition we do not have to take the supremum.

Lemma 2.2.4. If $\{\mathcal{U}_n : n = 1,2,\dots\}$ is a sequence of open finite covers of X such that $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}_n) = 0$, then the limit $\lim_{n\to\infty} \mathrm{P}(\phi,\mathcal{U}_n)$ exists and equals $\mathrm{P}(T,\phi)$.

Proof. Assume first that $\mathrm{P}(T,\phi)$ is finite and fix $\varepsilon > 0$. By the definition of pressure and uniform continuity of ϕ there exists W, an open cover of X, such that

$$\operatorname{osc}(\phi,\mathcal{W}) \le \frac{\varepsilon}{2} \text{ and } \mathrm{P}(\phi,\mathcal{W}) \ge \mathrm{P}(T,\phi) - \frac{\varepsilon}{2}. \tag{2.2.7}$$

Fix now $q \ge 1$ so large that for all $n \ge q$, $\operatorname{diam}(\mathcal{U}_n)$ does not exceed a Lebesgue number of the cover $\mathcal{W}$. Take $n \ge q$. Then $\mathcal{U}_n \succ \mathcal{W}$ and applying (2.2.7) and Lemma 2.2.2 we get

$$\mathrm{P}(\phi,\mathcal{U}_n) \ge \mathrm{P}(\phi,\mathcal{W}) - \frac{\varepsilon}{2} \ge \mathrm{P}(T,\phi) - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = \mathrm{P}(T,\phi) - \varepsilon. \tag{2.2.8}$$

Hence, letting $\varepsilon \to 0$, $\liminf_{n\to\infty} \mathrm{P}(\phi,\mathcal{U}_n) \ge \mathrm{P}(T,\phi)$. This finishes the proof in the case of finite pressure $\mathrm{P}(T,\phi)$. Notice also that actually the same proof goes through in the infinite case. ♣

Since in the definition of numbers $\mathrm{P}(\phi,\mathcal{U})$ no metric is involved, they do not depend on a compatible metric under consideration. And since also the convergence to zero of diameters of a sequence of subsets of X does not depend on a compatible metric, we come to the conclusion that the topological pressure $\mathrm{P}(T,\phi)$ is independent of any compatible metric (depends of course on topology).

The reader familiar with directed sets will notice easily that the family of all finite open covers $\mathcal{U}$ of X equipped with the relation "$\prec$" is a directed set and topological pressure $\mathrm{P}(T,\phi)$ is the limit of the generalized sequence $\mathrm{P}(\phi,\mathcal{U})$. However we can assure him/her that this remark will not be used anywhere in this book book.

If the funcion ϕ is identically zero, the pressure $\mathrm{P}(T,\phi)$ is usually called *topological entropy* of the map T and is denoted by $\mathrm{h_{top}}(T)$.

In the rest of this section we establish basic elementary properties of pressure and provide its more effective characterizations. Applying Lemma 2.2.2 we obtain

Corollary 2.2.5. If $\mathcal{U}$ is a finite, open cover of X, then $\mathrm{P}(T,\phi) \ge \mathrm{P}(\phi,\mathcal{U}) - \varepsilon(\phi,\mathcal{U})$.

Lemma 2.2.6. $\mathrm{P}(T^n, S_n\phi) = n\mathrm{P}(T,\phi)$ for every $n \ge 1$. In particular $\mathrm{h_{top}}(T^n) = n\mathrm{h_{top}}(T)$.

Proof. Put $g = S_n\phi$. Take $\mathcal{U}$, a finite open cover of X. Let $\overline{\mathcal{U}} = \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee \ldots \vee T^{-(n-1)}(\mathcal{U})$. Since now we actually deal with two separate transformations T and T^n, we do not use the symbol $\mathcal{U}^n$ just to avoid possible misunderstandings. For any $m \ge 1$ consider an open set $U \in \mathcal{U} \vee T^{-1}(\mathcal{U}) \vee \ldots \vee T^{-(nm-1)}(\mathcal{U}) = \overline{\mathcal{U}} \vee T^{-n}(\overline{\mathcal{U}}) \vee \ldots \vee T^{-n(m-1)}(\overline{\mathcal{U}})$. Then for every $x \in U$ we have

$$\sum_{k=0}^{mn-1} \phi \circ T^k(x) = \sum_{k=0}^{m-1} g \circ T^{nk}(x)$$

and therefore $S_{mn}\phi(U) = S_m g(U)$, where the symbol S_m is considered with respect to the map T^n. Hence $Z_{mn}(T,\phi,\mathcal{U}) = Z_m(T^n, g, \overline{\mathcal{U}})$ and this implies that $\mathrm{P}(T^n, g, \overline{\mathcal{U}}) =$

$n\mathrm{P}(T, \phi, \mathcal{U})$. Since given a sequence $\mathcal{U}_k$ of open covers of X whose diameters converge to zero, the diameters of the sequence of its refinements $\overline{\mathcal{U}_k}$ also converge to zero, applying now Lemma 2.2.4 finishes the proof. ♣

Lemma 2.2.7. If $T : X \to X$ and $S : Y \to Y$ are continuous mappings of compact metric spaces and $\pi : X \to Y$ is a continuous surjection such that $S \circ \pi = \pi \circ T$, then for every continuous function $\phi : Y \to I\!\!R$ we have $\mathrm{P}(S, \phi) \le \mathrm{P}(T, \phi \circ \pi)$.

Proof. For every finite, open cover $\mathcal{U}$ of Y we get

$$\mathrm{P}(S, \phi, \mathcal{U}) = \mathrm{P}(T, \phi \circ \pi, \pi^{-1}(\mathcal{U})). \tag{2.2.9}$$

In view of Corollary 2.2.5 we have

$$\mathrm{P}(T, \phi \circ \pi) \ge \mathrm{P}(T, \phi \circ \pi, \pi^{-1}(\mathcal{U})) - \mathrm{osc}(\phi \circ \pi, \pi^{-1}(\mathcal{U})) = \mathrm{P}(T, \phi \circ \pi, \pi^{-1}(\mathcal{U})) - \mathrm{osc}(\phi, \mathcal{U}) \tag{2.2.10}$$

Let $\{\mathcal{U}_n : n = 1, 2, \ldots\}$, be a sequence of open finite covers of Y whose diameters converge to 0. Then also $\lim_{n\to\infty} \mathrm{osc}(\phi, \mathcal{U}_n)) = 0$ and therefore, using Lemma 2.2.4, (2.2.9) and (2.2.10) we obtain

$$\mathrm{P}(S, \phi) = \lim_{n\to\infty} \mathrm{P}(S, \phi, \mathcal{U}_n) = \lim_{n\to\infty} \mathrm{P}(T, \phi \circ \pi, \pi^{-1}(\mathcal{U}_n)) \le \mathrm{P}(T, \phi \circ \pi)$$

The proof is finished. ♣

In the sequel we will need the following technical result.

Lemma 2.2.8. If $\mathcal{U}$ is a finite open cover of X then $\mathrm{P}(\phi, \mathcal{U}^k) = \mathrm{P}(\phi, \mathcal{U})$ for every $k \ge 1$.

Proof. Fix $k \ge 1$ and let $\gamma = \sup\{|S_{k-1}\phi(x)| : x \in X\}$. Since $S_{k+n-1}\phi(x) = S_n\phi(x) + S_{k-1}\phi(T^n(x))$ for every $n \ge 1$ and $x \in X$ we get

$$S_n\phi(x) - \gamma \le S_{k+n-1}\phi(x) \le S_n\phi(x) + \gamma$$

and therefore for every $n \ge 1$ and every $U \in \mathcal{U}^{k+n-1}$

$$S_n\phi(U) - \gamma \le S_{k+n-1}\phi(U) \le S_n\phi(U) + \gamma$$

Since $(\mathcal{U}^k)^n = \mathcal{U}^{k+n-1}$, these inequalities imply that

$$e^{-\gamma} Z_n(\phi, \mathcal{U}^k) \le Z_{n+k-1}(\phi, \mathcal{U}) \le e^{\gamma} Z_n(\phi, \mathcal{U}^k).$$

Letting now $n \to \infty$, we get the result required. ♣

Recall that ρ is a metric on X. For every $n \ge 1$ we define on X the new metric ρ_n by putting

$$\rho_n(x, y) = \max\{\rho(T^j(x), T^j(y)) : j = 0, 1, \ldots, n-1\}$$

Given $r > 0$ and $x \in X$ by $B_n(x,r)$ we denote the open ball in the metric ρ_n centered at x and of radius r. Let $\varepsilon > 0$ and let $n \geq 1$ be an integer. A set $F \subset X$ ia said to be (n,ε)-*spanning* if and only if the family of balls $\{B_n(x,\varepsilon) : x \in F\}$ covers the space X. A set $S \subset X$ is said to be (n,ε)-*separated* if and only if $\rho_n(x,y) \geq \varepsilon$ for any pair x, y of different points in S. The following fact is obvious.

Lemma 2.2.9. Every maximal in the sense of inclusion (n,ε)-separated set forms an (n,ε)-spanning set.

We would like to emphasize here that the world maximal refering to separated sets will be in this book always understood in the sense of inclusion and not in the sense of cardinality. We finish this section with the following characterization of pressure.

Theorem 2.2.10. For every $\varepsilon > 0$ and every $n \geq 1$ let $F_n(\varepsilon)$ be a maximal (n,ε) separated set in X. Then

$$\mathrm{P}(T,\phi) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in F_n(\varepsilon)}\exp S_n\phi(x) = \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\sum_{x\in F_n(\varepsilon)}\exp S_n\phi(x)$$

Proof. Fix $\varepsilon > 0$ and let $\mathcal{U}(\varepsilon)$ be a finite cover of X by open balls of radii $\varepsilon/2$. For any $n \geq 1$ consider $\underline{\mathcal{U}}^n$, a subcover of $\mathcal{U}(\varepsilon)n$ such that

$$Z_n(\phi,\mathcal{U}(\varepsilon)) = \sum_{U\in\underline{\mathcal{U}}}\exp S_n\phi(U),$$

where $Z_n(\phi,\mathcal{U}(\varepsilon))$ was defined by formula (2.2.3). For every $x \in F_n(\varepsilon)$ let $U(x)$ be an element of $\underline{\mathcal{U}}$ containing x. Since $F_n(\varepsilon)$ is an (n,ε)-separated set, we deduce that the function $x \mapsto U(x)$ is injective. Therefore

$$Z_n(\phi,\mathcal{U}(\varepsilon)) = \sum_{U\in\underline{\mathcal{U}}}\exp S_n\phi(U) \geq \sum_{x\in F_n(\varepsilon)}\exp S_n\phi(U(x)) \geq \sum_{x\in F_n(\varepsilon)}\exp S_n\phi(x).$$

Thus by Lemma 2.2.1

$$\mathrm{P}(\phi,\mathcal{U}(\varepsilon)) \geq \limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in F_n(\varepsilon)}\exp S_n\phi(x).$$

Hence, letting $\varepsilon \to 0$ and applying Corollary 2.2.4 we get

$$\mathrm{P}(T,\phi) \geq \limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\sum_{x\in F_n(\varepsilon)}\exp S_n\phi(x). \tag{2.2.11}$$

Let now $\mathcal{V}$ be an arbitrary finite open cover of X and let $\delta > 0$ be a Lebesgue number of $\mathcal{V}$. Take $\varepsilon < \delta/2$. Since for any $k = 0, 1, \ldots, n-1$ and for every $x \in F_n(\varepsilon)$

$$\mathrm{diam}\big(T^k(B_n(x,\varepsilon))\big) \leq 2\varepsilon < \delta,$$

we conclude that for some $U_k(x) \in \mathcal{V}$

$$T^k(B_n(x,\varepsilon)) \subset U_k(x)$$

Since the family $\{B_n(x,\varepsilon) : x \in F_n(\varepsilon)\}$ covers X (by Lemma 2.2.9), it implies that the family $\{U(x) : x \in F_n(\varepsilon)\} \subset \mathcal{V}^n$ also covers X, where $U(x) = U_0(x) \cap T^{-1}(U_1(x)) \cap \ldots \cap T^{-(n-1)}(U_{n-1}(x))$. Therefore

$$Z_n(\phi,\mathcal{V}) \le \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(U(x)) \le \exp\bigl(n\mathrm{osc}(\phi,\mathcal{V})\bigr) \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x)$$

Hence

$$\mathrm{P}(\phi,\mathcal{V}) \le \mathrm{osc}(\phi,\mathcal{V}) + \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x)$$

and consequently

$$\mathrm{P}(\phi,\mathcal{V}) - \mathrm{osc}(\phi,\mathcal{V}) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x).$$

Letting $\mathrm{diam}(\mathcal{V}) \to 0$ and applying Corollary 2.2.4 we get

$$\mathrm{P}(T,\phi) \le \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x).$$

Combining this and (2.2.11) finishes the proof. ♣

Frequently the limits

$$\limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x)$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \sum_{x\in F_n(\varepsilon)} \exp S_n\phi(x)$$

will be denoted respectively by $\overline{\mathrm{P}}(T,\phi,\varepsilon)$ and $\underline{\mathrm{P}}(T,\phi,\varepsilon)$. Actually these depend also on the sequence $\{F_n(\varepsilon) : n = 1,2,\ldots\}$ of maximal (n,ε)-separated sets under consideration. However it will be always clear from the context which such sequence is meant.

§2.3 VARIATIONAL PRINCIPLE

In this section we shall prove the following theorem, called *variational principle*, which has a long story and which establishes an interesting relationship between measure-theoretic dynamics and topological dynamics.

Theorem 2.3.1. (Variational principle) If $T : X \to X$ is a continuous transformation of a compact metric space X and $\phi : X \to I\!R$ is a continuous function then

$$\mathrm{P}(T,\phi) = \sup\left\{\mathrm{h}_\mu(T) + \int \phi \, d\mu : \mu \in M(T)\right\},$$

where $M(T)$ denotes the set of all Borel probability T-invariant measures on X.

The proof of this theorem consists of two parts. In the Part I we show that $\mathrm{h}_\mu(T)+\int \phi\, d\mu \le \mathrm{P}(T,\phi)$ for every measure $\mu \in M(T)$ and the Part II is devoted to proving inequality $\sup\{\mathrm{h}_\mu(T) + \int \phi\, d\mu : \mu \in M(T)\} \ge \mathrm{P}(T,\phi)$.

Proof of Part I. Let $\mu \in M(T)$. Fix $\varepsilon > 0$ and consider a finite partition $\mathcal{U} = \{A_1, \dots, A_s\}$ of X into Borel sets. One can find compact sets $B_i \subset A_i$, $i = 1, 2, \dots, s$, such that for the partition $\mathcal{V} = \{B_1, \dots, B_s, X \setminus (B_1 \cup \ldots \cup B_s)\}$ we have

$$\mathrm{H}_\mu(\mathcal{U}|\mathcal{V}) \le \varepsilon,$$

where the conditional entropy $\mathrm{H}_\mu(\mathcal{U}|\mathcal{V})$ has been defined in (1.3.3). Therefore by Theorem 1.3.1(e),(d) and (1.4.1) we get for every $n \ge 1$

$$\mathrm{H}_\mu(\mathcal{U}^n|\mathcal{V}^n) \le \sum_{k=0}^{n-1} \mathrm{H}_\mu(T^{-k}\mathcal{U}|\mathcal{V}^n) \le \sum_{k=0}^{n-1} \mathrm{H}_\mu(T^{-k}\mathcal{U}|T^{-k}\mathcal{V}) \le \varepsilon n.$$

Thus, applying Theorem 1.3.1(b),

$$\mathrm{H}_\mu(\mathcal{U}^n) \le \mathrm{H}_\mu(\mathcal{U}^n \vee \mathcal{V}^n) = \mathrm{H}_\mu(\mathcal{V}^n) + \mathrm{H}_\mu(\mathcal{U}^n|\mathcal{V}^n) \le \mathrm{H}_\mu(\mathcal{V}^n) + \varepsilon n \tag{2.3.1}$$

Our first aim is to estimate from above the number $\mathrm{H}_\mu(\mathcal{V}^n) + \int S_n\phi \, d\mu$. Putting $b_n = \sum_{B\in\mathcal{V}^n} \exp S_n\phi(B)$, keeping notation $k(x) = -x\log x$ and using concavity of h we obtain

$$\begin{aligned}
\mathrm{H}_\mu(\mathcal{V}^n) + \int S_n\phi \, d\mu &\le \sum_{B\in\mathcal{V}^n} \mu(B)\big(S_n\phi(B) - \log\mu(B)\big) \\
&= b_n \sum_{B\in\mathcal{V}^n} b_n^{-1} e^{S_n\phi(B)} k\big(\mu(B)e^{-S_n\phi(B)}\big) \\
&\le b_n k\big(\sum_{B\in\mathcal{V}^n} b_n^{-1} e^{S_n\phi(B)} \mu(B) e^{-S_n\phi(B)}\big) = b_n k(b_n^{-1}) \\
&= \log\big(\sum_{B\in\mathcal{V}^n} \exp S_n\phi(B)\big) \qquad (2.3.2)
\end{aligned}$$

Take now $0 < \delta < \frac{1}{2}\inf\{\rho(B_i, B_j) : 1 \le i \ne j \le s\} > 0$ so small that

$$|\phi(x) - \phi(y)| < \varepsilon \tag{2.3.3}$$

if $\rho(x, y) < \delta$. Consider an arbitrary maximal (n, δ)-separated set $E_n(\delta)$. Fix $B \in \mathcal{V}^n$. Then, by Lemma 2.2.9, for every $x \in B$ there exists $y \in E_n(\delta)$ such that $x \in B_n(y, \delta)$, whence $|S_n\phi(x) - S_n\phi(y)| \le \varepsilon n$ by (2.3.3). Therefore, using finiteness of the set $E_n(\delta)$, we see that there exists $y(B) \in E_n(\delta)$ such that

$$S_n\phi(B) \le S_n\phi(y(B)) + \varepsilon n \tag{2.3.4}$$

and

$$B \cap B_n(y(B), \delta) \neq \emptyset.$$

The definitions of δ and partition $\mathcal{V}$ imply that for every $z \in X$

$$\#\{B \in \mathcal{V} : B \cap B(z, \delta) \neq \emptyset\} \le 2$$

Thus

$$\#\{B \in \mathcal{V}^n : B \cap B_n(z, \delta) \neq \emptyset\} \le 2^n$$

Therefore the function $\mathcal{V}^n \ni B \mapsto y(B) \in E_n(\delta)$ is at most 2^n to 1. Hence, using (2.3.4),

$$2^n \sum_{y \in E_n(\delta)} \exp S_n\phi(y) \ge \sum_{B \in \mathcal{V}^n} \exp\bigl(S_n\phi(B) - \varepsilon n\bigr) = e^{-\varepsilon n} \sum_{B \in \mathcal{V}^n} \exp S_n\phi(B)$$

Taking now the logarithms of both sides of this inequality, dividing them by n and applying (2.3.2), we get

$$\begin{aligned} \log 2 + \frac{1}{n}\log\bigl(\sum_{y \in E_n(\delta)} \exp S_n\phi(y)\bigr) &\ge -\varepsilon + \frac{1}{n}\log\bigl(\sum_{B \in \mathcal{V}^n} \exp S_n\phi(B)\bigr) \\ &\ge \frac{1}{n}\mathrm{H}_\mu(\mathcal{V}^n) + \frac{1}{n}\int S_n\phi \, d\mu - \varepsilon. \end{aligned}$$

So, by (2.3.1),

$$\frac{1}{n}\log\bigl(\sum_{y \in E_n(\delta)} \exp S_n\phi(y)\bigr) \ge \frac{1}{n}\mathrm{H}_\mu(\mathcal{U}^n) + \int \phi \, d\mu - (2\varepsilon + \log 2).$$

In view of the definition of entropy $\mathrm{h}_\mu(T, \mathcal{U})$ presented just after Lemma 1.4.2, by letting $n \to \infty$, we get

$$\underline{\mathrm{P}}(T, \phi, \delta) \ge \mathrm{h}_\mu(T, \mathcal{U}) + \int \phi \, d\mu - (2\varepsilon + \log 2).$$

Applying now Theorem 2.2.10 with $\delta \to 0$ and next letting $\varepsilon \to 0$ and taking supremum over all Borel partitions $\mathcal{U}$ lead us to the following

$$\mathrm{P}(T, \phi) \ge \mathrm{h}_\mu(T) + \int \phi \, d\mu - \log 2.$$

And applying with every $n \geq 1$ this estimate to the transformation T^n and the function $S_n\phi$ we obtain

$$\mathrm{P}(T^n, \phi) \geq \mathrm{h}_\mu(T^n) + \int S_n\phi \, d\mu - \log 2$$

or equivalently, by Lemma 2.2.6 and Theorem 1.4.6(a)

$$n\mathrm{P}(T, \phi) \geq n\mathrm{h}_\mu(T) + n\int \phi \, d\mu - \log 2$$

Dividing both sides of this inequality by n and letting then $n \to \infty$, the proof of Part I follows. ♣

In the proof of part II we will need the following two lemmas.

Lemma 2.3.2. If μ is a Borel probability measure on X, then for every $\varepsilon > 0$ there exists a finite partition $\mathcal{A}$ such that $\mathrm{diam}(\mathcal{A}) \leq \varepsilon$ and $\mu(\partial A) = 0$ for every $A \in \mathcal{A}$.

Proof. Let $E = \{x_1, \ldots, x_s\}$ be an $\varepsilon/4$-spanning set (that is with respect to the metric $\rho = \rho_0$) of X. Since for every $i \in \{1, \ldots, s\}$ the sets $\{x : \rho(x, x_i) = r\}$, $\varepsilon/4 < r < \varepsilon/2$, are closed and mutually disjoint, only countably many of them can have positive measure μ. Hence, there exists $\varepsilon/4 < t < \varepsilon/2$ such that for every $i \in \{1, \ldots, s\}$

$$\mu(\{x : \rho(x, x_i) = t\}) = 0 \tag{2.3.5}$$

Define inductively the sets $A_1, A_2, \ldots, A_s$ putting $A_1 = \{x : \rho(x, x_1) \leq t\}$ and for every $i = 2, 3, \ldots, s$

$$A_i = \{x : \rho(x, x_i) \leq t\} \setminus (A_1 \cup A_2 \cup \ldots \cup A_{i-1})$$

The family $\mathcal{U} = \{A_1, \ldots, A_s\}$ is a partition of X with diameter not exceeding ε. Using (2.3.5) and noting that generally $\partial(A \setminus B) \subset \partial A \cup \partial B$, we conclude by induction that $\mu(\partial A_i) = 0$ for every $i = 1, 2, \ldots, s$. ♣

Lemma 2.3.3. If μ is a Borel probability measure on X, if μ_n, $n = 1, 2, \ldots$, is a sequnce of probability measures converging weakly to μ and if $\mathcal{U}$ is a finite partition of X such that $\mu(\partial A) = 0$ for every $A \in \mathcal{U}$, then $\lim_{n\to\infty} \mathrm{H}_{\mu_n}(\mathcal{U}) = \mathrm{H}_\mu(\mathcal{U})$.

Proof. In view of Theorem 2.1.1, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for every $A \in \mathcal{U}$, whence the continuity of the function $k(x) = -x \log x$ implies that

$$\lim_{n\to\infty} k(\mu_n(A)) = k(\mu(A)) \tag{2.3.6}$$

Since the partition $\mathcal{U}$ is finite and since, by definition, $\mathrm{H}_\nu(\mathcal{U}) = \sum_{A \in \mathcal{U}} h(\nu(A))$ for any measure ν, the lemma follows by applying (2.3.6). ♣

Proof of Part II. Fix $\varepsilon > 0$ and let $E_n(\varepsilon)$, $n = 1, 2, \ldots$, be a sequence of maximal (n, ε)-separated set in X. For every $n \geq 1$ define measures

$$\mu_n = \frac{\sum_{x \in E_n(\varepsilon)} \delta_x \exp S_n\phi(x)}{\sum_{x \in E_n(\varepsilon)} \exp S_n\phi(x)}, \quad m_n = \frac{1}{n}\sum_{k=0}^{n-1} \mu_n \circ T^{-k}$$

where δ_x denotes the Dirac measure concentrated at the point x (see (2.1.2)). Let $\{n_i, i = 1, 2, \ldots\}$ be an increasing sequence such that m_{n_i} converges weakly, say to m and

$$\lim_{i\to\infty} \frac{1}{n_i} \log \sum_{x\in E_{n_i}(\varepsilon)} \exp S_n\phi(x) = \limsup_{n\to\infty} \frac{1}{n} \log \sum_{x\in E_n(\varepsilon)} \exp S_n\phi(x) \tag{2.3.9}$$

Clearly $m \in M(T)$. In view of Lemma 2.3.2 there exists a finite partition γ such that $\mathrm{diam}(\gamma) \le \varepsilon$ and $\mu(\partial G) = 0$ for every $G \in \gamma$. For any $n \ge 1$ put $g_n = \sum_{x\in E_n(\varepsilon)} \exp S_n\phi(x)$. Since $\#(G \cap E_n(\varepsilon)) \le 1$ for every $G \in \gamma^n$, we obtain

$$\begin{aligned}
\mathrm{H}_{\mu_n}(\gamma^n) + \int S_n\phi\, d\mu_n &= \sum_{x\in E_n(\varepsilon)} \big(-\log\mu_n(x) + S_n\phi(x)\big)\mu_n(x) \\
&= \sum_{x\in E_n(\varepsilon)} \frac{\exp S_n\phi(x)}{g_n}\big(S_n\phi(x) - \log\big(\frac{\exp S_n\phi(x)}{g_n}\big)\big) \\
&= g_n^{-1} \sum_{x\in E_n(\varepsilon)} \exp S_n\phi(x)\big(S_n\phi(x) - S_n\phi(x) + \log g_n\big) = \log g_n \qquad (2.3.10)
\end{aligned}$$

Fix now $M \in I\!N$ and $n \ge 2M$. For $j = 0, 1, \ldots, M-1$ let $s(j) = \mathrm{E}(\frac{n-j}{M}) - 1$, where $\mathrm{E}(x)$ denotes the integer part of x. Note that

$$\bigvee_{k=0}^{s(j)} T^{-(kM+j)}\gamma^M = T^{-j}\gamma \vee \ldots \vee T^{-(s(j)M+j)-(M-1)}\gamma = T^{-j}\gamma \vee \ldots \vee T^{-((s(j)+1)M+j-1)}\gamma$$

and

$$(s(j)+1)M + j - 1 \le n - j + j - 1 = n - 1$$

Therefore, setting $R_j = \{0, 1, \ldots, j-1, (s(j)+1)M + j, \ldots, n-1\}$, we can write

$$\gamma^n = \bigvee_{k=0}^{s(j)} T^{-(kM+j)}\gamma^M \vee \bigvee_{i\in R_j} T^{-i}\gamma.$$

Hence

$$\begin{aligned}
\mathrm{H}_{\mu_n}(\gamma^n) &\le \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n}\big(T^{-(kM+j)}\gamma^M\big) + \mathrm{H}_{\mu_n}\big(\bigvee_{i\in R_j} T^{-i}\gamma\big) \\
&\le \sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n\circ T^{-(kM+j)}}(\gamma^M) + \log\big(\#\big(\bigvee_{i\in R_j} T^{-i}\gamma\big)\big).
\end{aligned}$$

Summing now over all $j = 0, 1, \ldots, M-1$ we then get

$$\begin{aligned}
M\mathrm{H}_{\mu_n}(\gamma^n) &\le \sum_{j=0}^{M-1}\sum_{k=0}^{s(j)} \mathrm{H}_{\mu_n\circ T^{-(kM+j)}}(\gamma^M) + \sum_{j=0}^{M-1} \log\big(\#\gamma^{\# R_j}\big) \\
&\le \sum_{l=0}^{n-1} \mathrm{H}_{\mu_n\circ T^{-l}}(\gamma^M) + 2M^2\log\#\gamma \le n\mathrm{H}_{\frac{1}{n}\sum_{l=0}^{n-1}\mu_n\circ T^{-l}}(\gamma^M) + 2M^2\log\#\gamma.
\end{aligned}$$

And applying (2.3.10) we obtain

$$M\log\Big(\sum_{x\in E_n(\varepsilon)}\exp S_n\phi(x)\Big)\le n\mathrm{H}_{m_n}(\gamma^M)+M\int S_n\phi\,d\mu_n+2M^2\log\#\gamma$$

Dividing both sides of this inequality by Mn, we get

$$\frac{1}{n}\log\Big(\sum_{x\in E_n(\varepsilon)}\exp S_n\phi(x)\Big)\le\frac{1}{M}\mathrm{H}_{m_n}(\gamma^M)+\int\phi\,dm_n+2\frac{M}{n}\log\#\gamma.$$

Since $\partial T^{-1}(A)\subset T^{-1}(\partial A)$ for every set $A|sbtX$, the measure m of the boundaries of the partition γ^M is equal to 0. Letting therefore $n\to\infty$ along the subsequence $\{n_i\}$ we conclude from this inequality, (2.3.7) and Lemma 2.3.3 that

$$\overline{\mathrm{P}}(T,\phi,\varepsilon)\le\frac{1}{M}\mathrm{H}_m(\gamma^M)+\int\phi\,dm.$$

Now letting $M\to\infty$ we get

$$\overline{\mathrm{P}}(T,\phi,\varepsilon)\le\mathrm{h}_m(T,\gamma)+\int\phi\,dm\le\sup\left\{h_\mu(T)+\int\phi\,d\mu:\mu\in M(T)\right\}.$$

Applying finally Theorem 2.2.10 and letting $\varepsilon\searrow 0$, we get the desired inequality. ♣

Corollary 2.3.4. Under assumptions of Theorem 2.3.1

$$\mathrm{P}(T,\phi)=\sup\{\mathrm{h}_\mu(T)+\int\phi\,d\mu:\mu\in M_e(T)\},$$

where $M_e(T)$ denotes the set of all Borel ergodic probability T-invariant measures on X.

Proof. Let $\mu\in M(T)$ and let $\{\mu_x:x\in X\}$ be the ergodic decomposition of μ. Then $\mathrm{h}_\mu=\int\mathrm{h}_{\mu_x}\,d\mu(x)$ and $\int\phi\,d\mu=\int(\int\phi\,d\mu_x)\,d\mu(x)$. Therefore

$$\mathrm{h}_\mu+\int\phi\,d\mu=\int\left(\mathrm{h}_{\mu_x}+\int\phi\,d\mu_x\right)\,d\mu(x)$$

and consequently there exists $x\in X$ such that $\mathrm{h}_{\mu_x}+\int\phi\,d\mu_x\ge\mathrm{h}_\mu+\int\phi\,d\mu$ which finishes the proof. ♣

Corollary 2.3.5. If $T:X\to X$ is a continuous transformation of a compact metric space X, $\phi:X\to I\!\!R$ is a continuous function and Y is a forward invariant subset of X (i.e. $T(Y)\subset Y$), then $\mathrm{P}(T|_Y,\phi|_Y)\le\mathrm{P}(T,\phi)$.

Proof. The proof follows immediatly from Theorem 2.3.1 by the remark that each $T|_Y$-invariant measure on Y can be treated as a measure on X and is T-invariant.

§2.4 EQUILIBRIUM STATES AND EXPANSIVE MAPS

We keep in this section the notation of the previous one. A measure $\mu \in M(T)$ is called an *equilibrium state* for the transformation T and function ϕ if $\mathrm{P}(T,\phi) = \mathrm{h}_\mu(T) + \int \phi\, d\mu$. The set of all those measures will be denoted by $E(\phi)$. In the case $\phi = 0$ the equilibrium states are also called as *maximal measures*. Similarly (in fact even easier) as Corollary 2.3.5 one can prove the following

Proposition 2.4.1 If $E(\phi) \neq \emptyset$ then $E(\phi)$ contains ergodic measures.

As shows the following example there exist transformations and functions which admit no equilibrium states.

Example 2.4.2. Let $\{T_n : X_n \to X_n\}_{n \geq 1}$ be a sequence of continuous mappings of compact metric spaces X_n such that for every $n \geq 1$

$$\mathrm{h}_{\mathrm{top}}(T_n) < \mathrm{h}_{\mathrm{top}}(T_{n+1}) \text{ and } \sup_n \mathrm{h}_{\mathrm{top}}(T_n) < \infty \tag{2.4.1}$$

The disjoint union $\oplus_{n=1}^{\infty} X_n$ of the spaces X_n is a locally compact space and let $X = \{\omega\} \cup \oplus_{n=1}^{\infty} X_n$ be its one-point (Alexandroff) compactification. Define the map $T : X \to X$ by $T|_{X_n} = T_n$ and $T(\omega) = \omega$. The reader will check easily that T is continuous. Suppose that μ is an ergodic maximal measure for T. Then $\mu(X_n) = 1$ for some $n \geq 1$ and therefore $\mathrm{h}_{\mathrm{top}}(T) = \mathrm{h}_\mu(T_n) \leq \mathrm{h}_{\mathrm{top}}(T_n)$ which contradicts formula (2.4.1) and Corollary 2.3.5. In view of Proposition 2.4.1 this shows that T has no maximal measure.

A more difficult problem is to find a transitive and smooth example (see for instance [Mi, 1973]).

The remaining part of this section is devoted to provide sufficient conditions for the existence of equilibrium states and we start with the following simple general criterion which will be the base to obtain all others.

Lemma 2.4.3. If the function $M(T) \ni \mu \to \mathrm{h}_\mu(T)$ is upper semi-continuous then each continuous function $\phi : X \to \mathbb{R}$ has an equilibrium state.

Proof. By the definition of weak topology the function $M(T) \ni \mu \to \int \phi\, d\mu$ is continuous. Therefore the lemma follows from the assumption, the sequential compactness of the space $M(T)$ and Theorem 2.1.1. ♣

As an immediate consequence of this lemma and Theorem 2.3.1 (variational principle) we obtain the following.

Corollary 2.4.4. If $\mathrm{h_{top}}(T) = 0$ then each continuous function on X has an equilibrium state.

A continuous transformation $T : X \to X$ of a compact metric space X equipped with a metric ρ is said to be *(positively) expansive* if and only if

$$\exists \delta > 0 \, [\forall n \geq 0 \; \rho(T^n(x), T^n(y)) \leq \delta] \Longrightarrow \; x = y$$

and the number δ which has appeared in this definition is called an *expansive constant.*

Although at the end of this section we will introduce a related but different notion of expansiveness of homeomorphisms we will frequently omit the word "positively". Note that the property of being expansive does not depend on the choice of a metric compatible with the topology. From now on in this chapter the transformation T will be assumed to be positively expansive, unless stated otherwise. The following lemma is an immediate consequence of expansiveness.

Lemma 2.4.5. If $\mathcal{A}$ is a finite Borel partition of X with diameter not exceeding an expansive constant then $\mathcal{A}$ is a generator for every Borel probability T-measure μ on X.

The main result concerning expansive maps is the following.

Theorem 2.4.6. If $T : X \to X$ is positively expansive then the function $M(T) \ni \mu \to \mathrm{h}_\mu(T)$ is upper semi-continuous and consequently (by Lemma 2.4.3) each continuous function on X has an equilibrium state.

Proof. Let $\delta > 0$ be an expansive constant of T and let $\mu \in M(T)$. By Lemma 2.3.2 there exists a finite partition $\mathcal{A}$ of X such that $\mathrm{diam}(\mathcal{A}) \leq \delta$ and $\mu(\partial A) = 0$ for every $A \in \mathcal{A}$. Thus in view of Lemma 2.4.5 and Theorem 1.8.7(b) $\mathrm{h}_\mu(T) = \mathrm{h}_\mu(T, \mathcal{A})$ whence by the definition of the entropy $\mathrm{h}_\mu(T, \mathcal{A})$ (cf. Lemma 1.4.2) there exists $m \geq 1$ such that

$$\frac{1}{m+1} \mathrm{H}_\mu(\mathcal{A}^m) \leq \mathrm{h}_\mu(T) + \frac{\varepsilon}{2} \tag{2.4.2}$$

Consider now a sequence $\{\mu_n : n = 1, 2, \ldots\}$ of invariant measures converging weakly to μ. By the definition of the entropy of partition, by Theorem 2.3.3 and by the choice of partition $\mathcal{A}$, $\lim_{n\to\infty} \mathrm{H}_{\mu_n}(\mathcal{A}^m) = \mathrm{H}_\mu(\mathcal{A}^m)$. Therefore there exists $n_0 \geq 1$ such that for every $n \geq n_0$

$$\frac{1}{m+1} \left|\mathrm{H}_{\mu_n}(\mathcal{A}^m) - \mathrm{H}_\mu(\mathcal{A}^m)\right| \leq \frac{\varepsilon}{2}$$

Combining this and 2.4.2, and using Lemma 1.4.3 we get for every $n \geq n_0$

$$\mathrm{h}_{\mu_n}(T) = \mathrm{h}_{\mu_n}(T, \mathcal{A}) \leq \frac{1}{m+1} \mathrm{H}_{\mu_n}(\mathcal{A}^m) \leq \frac{1}{m+1} \mathrm{H}_\mu(\mathcal{A}^m) + \frac{\varepsilon}{2} \leq \mathrm{h}_\mu(T) + \varepsilon$$

The proof is finished. ♣

Below we prove three additional interesting results about expansive maps.

Lemma 2.4.7. If $\mathcal{U}$ is a finite open cover of X with diameter not exceeding an expansive constant of an expansive map $T : X \to X$, then $\lim_{n\to\infty} \mathrm{diam}(\mathcal{U}^n) = 0$.

Proof. Let $\mathcal{U} = \{U_1, U_2, \dots, U_s\}$. By expansiveness for every sequence $\{a_n : n = 0, 2, \dots\}$ of elements of the set $\{1, 2, \dots, s\}$

$$\#\Big(\bigcap_{n=0}^{\infty} T^{-n}(\overline{U_{a_n}}\Big) \le 1$$

and hence

$$\lim_{k\to\infty} \mathrm{diam}\Big(\bigcap_{n=0}^{k} T^{-n}(\overline{U_{a_n}})\Big) = 0$$

Therefore, given a fixed $\varepsilon > 0$ there exists a minimal finite $k = k(\{a_n\})$ such that

$$\mathrm{diam}\Big(\bigcap_{n=0}^{k} T^{-n}(\overline{U_{a_n}})\Big) < \varepsilon$$

Note now that the function $\{1, 2, \dots, s\}^{I\!N} \ni \{a_n\} \mapsto k(\{a_n\})$ is continuous, even more, it is locally constant. Thus, by compactness of the space $\{1, 2, \dots, s\}^{I\!N}$, it is bounded, say by t, and therefore

$$\mathrm{diam}(\mathcal{U}^n) < \varepsilon$$

for every $n \ge t$. The proof is finished. ♣

Combining now Lemma 2.2.4, Lemma 2.4.7 and Lemma 2.2.8 we get the following.

Proposition 2.4.8. If $\mathcal{U}$ is a finite open cover of X with diameter not exceeding an expansive constant then $\mathrm{P}(T, \phi) = \mathrm{P}(T, \phi, \mathcal{U})$.

As the last result of this section we shall prove the following.

Proposition 2.4.9. There exists a constant $\eta > 0$ such that $\forall\, 0 < \varepsilon < \eta\ \exists\, n(\varepsilon) \ge 1$

$$\rho(x, y) \ge \varepsilon \implies \rho_{n(\varepsilon)}(x, y) > \eta$$

Proof. Let $\mathcal{U} = \{U_1, U_2, \dots, U_s\}$ be a finite open cover of X with diameter not exceeding an expansive constant δ and let η be a Lebesgue number of $\mathcal{U}$. Fix $\varepsilon > 0$. In view of Lemma 2.4.7 there exists an $n(\varepsilon) \ge 1$ such that

$$\mathrm{diam}(\mathcal{U}^{n(\varepsilon)}) < \varepsilon. \tag{2.4.3}$$

Let $\rho(x, y) \ge \varepsilon$ and suppose that $\rho_{n(\varepsilon)}(x, y) \le \eta$. Then

$$\forall\, (0 \le j \le n(\varepsilon) - 1)\ \exists\, (U_{i_j} \in \mathcal{U})\ \ T^j(x), T^j(y) \in U_{i_j}$$

and therefore

$$x, y \in \bigcap_{j=0}^{n(\varepsilon)-1} T^{-j}(U_{i_j}) \in \mathcal{U}^{n(\varepsilon)}$$

Hence $\operatorname{diam}(\mathcal{U}^{n(\varepsilon)}) \geq \rho(x,y) \geq \varepsilon$ which contradicts (2.4.3). The proof is finished. ♣

As we mentioned at the begining of this section there is a notion related to positive expansiveness which makes sense only for homeomorphisms. Namely we say that a homeomorphism $T : X \to X$ is *expansive* if and only if

$$\exists \delta > 0 \ [\forall n \in \boldsymbol{Z} \ \rho(T^n(x), T^n(y)) \leq \delta] \Longrightarrow \ x = y$$

We will not explore this notion in our book – we only want to emphasize that for expansive homeomorphisms analogous results (with obvious modifications) can be proved (in the same way) as for positively expansive mappings. Of course each positively expansive homeomorphism is expansive.

§2.5 TOPOLOGICAL PRESSURE AS A FUNCTION ON THE BANACH SPACE OF CONTINUOUS FUNCTIONS. THE ISSUE OF UNIQUENESS OF EQUILIBRIUM STATES

Let $T : X \to X$ be a continuous mapping of a compact topological space X. We shall discuss here the topological pressure function $\mathrm{P} : C(X) \to I\!R$, $\mathrm{P}(\phi) = \mathrm{P}(T, \phi)$. Assume that the topological entropy is finite, $\mathrm{h_{top}}(T) < \infty$. Hence the pressure P is also finite, for example

$$\mathrm{P}(\phi) \leq \mathrm{h_{top}}(T) + ||\phi|||_\infty. \tag{2.5.1}$$

This estimate follows directly from the definitions, see Section 2. It is also an immediate consequence of Theorem 2.3.1 (Variational Principle) in case X is metrizable.

Let us start with the following easy

Theorem 2.5.1. The pressure function P is Lipschitz continuous with the Lipschitz constant 1.

Proof. Let $\phi \in C(X)$. Recall from Section 2.2 that in the definition of pressure we have considered the following partition function

$$Z_n(\phi, \mathcal{U}) = \inf_{\mathcal{V}} \{ \sum_{U \in \mathcal{V}} \exp S_n\phi(U) \},$$

where $\mathcal{V}$ ranges over all covers of X contained in $\mathcal{U}^n$. Now if also $\psi \in C(X)$, then we obtain for every open cover $\mathcal{U}$ and positive integer n

$$Z_n(\psi, \mathcal{U}) e^{-||\phi-\psi||_\infty n} \leq Z_n(\phi, \mathcal{U}) \leq Z_n(\psi, \mathcal{U}) e^{||\phi-\psi||_\infty n}$$

Taking limits if $n \nearrow \infty$ we get $\mathrm{P}(\psi) - ||\phi - \psi||_\infty \le \mathrm{P}(\phi) \le \mathrm{P}(\psi) + ||\phi - \psi||_\infty$, hence $|P(\psi) - P(\phi)| \le ||\psi - \phi||_\infty$. ♣

Theorem 2.5.2. If X is metrizable, then the topological pressure function $\mathrm{P} : C(X) \to I\!\!R$ is convex.

we provide two different proofs of this important theorem. One elementary, the second relying on the variational principle (Theorem 2.3.1).

Proof 1. By Hölder inequality applied with the exponents $a = 1/\alpha, b = 1/(1-\alpha)$, so that $1/a + 1/b = \alpha + 1 - \alpha = 1$ we obtain for an arbitrary finite set $E \subset X$

$$\frac{1}{n}\log\sum_E e^{S_n(\alpha\phi)+S_n(1-\alpha)\psi)} = \frac{1}{n}\log\sum_E e^{\alpha S_n(\phi)}e^{(1-\alpha)S_n(\psi)} \le$$

$$\frac{1}{n}\log\big(\sum_E e^{S_n(\phi)}\big)^\alpha\big(\sum_E e^{S_n(\psi)}\big)^{1-\alpha} \le \alpha\frac{1}{n}\log\big(\sum_E e^{S_n(\phi)}\big) + (1-\alpha)\frac{1}{n}\log\big(\sum_E e^{S_n(\psi)}\big).$$

To conclude the proof use the definition of pressure *via* $E = F_n(\varepsilon)$ that is (n,ε)-separated sets, Theorem 2.2.10.

Proof 2. It is sufficient to prove that the function

$$\hat{P} := \sup_{\mu\in M(X,T)} L_\mu\phi = \mathrm{h}_\mu(T) + \mu\phi$$

is convex, because by variational principle $\hat{P}(\phi) = P(\phi)$.

We need to prove that the set

$$A := \{(\phi, y) \in C(X) \times I\!\!R\} : y \ge \hat{P}(\phi))$$

is convex. Observe however that that by its definition $A = \bigcap_{\mu\in M(X,T)} L_\mu^+$, where by L_μ^+ we denote the upper half space $\{(f,y) : y \ge L_\mu\phi\}$. All the halfspaces L_μ^+ are convex, hence A is convex as their intersection. ♣

Remark 2.5.3. We can write $L_\mu\phi = \mu\phi - (-h_\mu(T))$. The function $\hat{\mathrm{P}}(\phi) = \sup_{\mu\in M(T)} L_\mu\phi$ defined on the space $C(X)$ is called the Legendre-Fenchel transform of the convex function $\mu \mapsto -h_\mu(T)$ on the convex set $M(T)$. We shall abbreviate the name Legendre-Fenchel transform to LF-transform. Observe that this transform generalizes the standard Legendre transformation of a strictly convex function h on a finite dimensional linear space, say $I\!\!R^n$, $y \mapsto \sup_{x\in I\!\!R^n} < x, y > -h(x)$, where $< x, y >$ is the scalar (inner) product of x and y.

Note that $-h_\mu(T)$ is not strictly convex (unless $M(X,T)$ is a one element space) because it is affine, see Th.1.4.7.

Proof 2 just repeats the standard proof that Legendre transform is convex.

In the sequel we will need so called geometric form of the Hahn-Banach theorem (see [Bourbaki, Th.1, Ch.2.5] or Ch. 1.7 of [Edwards, 1995].

Theorem 2.5.4 (Hahn-Banach). Let A be an open convex non-empty subset of a real topological vector space V and let M be a non-empty affine subset of V (linear subspace moved by a vector) which does not meet A. Then there exists a codimension 1 closed affine subset H which contains M and does not meet A.

Suppose now that $P : V \to I\!\!R$ is an arbitrary convex continuous function on a real topological vector space V. We call a continuous linear functional $F : V \to I\!\!R$ *tangent* to P at $x \in V$ if

$$F(y) \leq P(x+y) - P(x) \tag{2.5.2}$$

for every $y \in V$. We denote the set of all such functionals by $V^*_{x,P}$. Sometimes the term *supporting functional* is being used in the literature.

Applying Theorem 2.5.4 we easily prove that for every x the set $V^*_{x,P}$ is non-empty. Indeed, we can consider the open convex set $A = \{(\phi, y) \in V \times I\!\!R\} : y > P(x)\}$ in the vector space $V \times I\!\!R$ with the product topology and the one-point set $M = \{x, P(x)\}$, and define a supporting functional we look for, as having the graph $H - \{x, P(x)\}$ in $V \times I\!\!R$.

We would also like to bring reader's attention to the following another general fact from functional analysis.

Theorem 2.5.5. Let V be a Banach space and $P : V \to I\!\!R$ be a convex continuous function. Then for every $x \in V$ the function P is differentiable at x in every direction (Gateaux differentiable), or in a dense in the weak topology set of directions, if and only if $V^*_{x,P}$ is a singleton.

Proof. Suppose first that P is not differentiable at some point x and direction y. Choose an arbitrary $F \in V^*_{x,P}$. Non-differentiability in the direction $y \in V$ implies that there exist $\varepsilon > 0$ and a sequence $\{t_n\}_{n\geq 1}$ converging to 0 such that

$$P(x + t_n y) - P(x) \geq t_n F(y) + \varepsilon |t_n|. \tag{2.5.3}$$

In fact we can assume that all t_n, $n \geq 1$, are positive by passing to a subsequence and replacing y by $-y$ if necessary. We shall prove that (2.5.3) implies the existence of $\hat{F} \in V^*_{x,P} \setminus \{F\}$. Indeed, take $F_n \in V^*_{x+t_n y,P}$. Then, by (2.5.1), we have

$$P(x) - P(x + t_n y) \geq F_n(-t_n y) \tag{2.5.4}$$

The inequalities (2.5.3) and (2.5.4) give

$$t_n F(y) + \varepsilon t_n \leq t_n F_n(y).$$

Hence

$$(F_n - F)(y) \geq \varepsilon. \tag{2.5.5}$$

In the case when P is Lipschitz continuous, and this is the case of topological pressure see (Theorem 2.5.1) which we are mostly interested in, all F_n's, $n \geq 1$, are uniformly bounded. Indeed, let L be a Lipschitz constant of P. Then for every $z \in V$ and every $n \geq 1$

$$F_n(z) \leq P(x + t_n y + z) - P(x + t_n y) \leq L||z||$$

So, $||F_n|| \leq L$ for every $n \geq 1$. Thus, there exists $\hat{F} = \lim_{n\to\infty} F_n$, a weak*-limit of a sequence $\{F_n\}_{n\geq 1}$ (subsequence of the previous sequence). By (2.5.5) $(\hat{F} - F)(y) \geq \varepsilon$. Hence $\hat{F} \neq F$. Since

$$P(x + t_n y + v) - P(x + t_n y) \geq F_n(v) \quad \text{for all } n \text{ and } v \in V$$

passing with n to ∞ and using continuity of P, we conclude that $\hat{F} \in V^*_{x,P}$.

If we do not assume that P is Lipschitz continuous, we restrict F_n to the 1-dimensional space spanned by y i.e. we consider $F_n|_{I\!Ry}$. In view of (2.5.5) for every $n \geq 1$ there exists $0 \leq s_n \leq 1$ such that $F_n(s_n y) - F(s_n y) = \varepsilon$. Passing to a subsequence, we may assume that $\lim_{n\to\infty} s_n = s$ for some $s \in [0, 1]$. Define

$$f_n = s_n F_n|_{I\!Ry} + (1 - s_n)F|_{I\!Ry}$$

Then $f_n(y) = F(y) = \varepsilon$ hence $||f_n|| = ||F|| = \frac{\varepsilon}{||y||}$ for every $n \geq 1$. Thus the sequence $\{f_n\}_{n\geq 1}$ is uniformly bounded and, consequently, it has a weak-* limit $\hat{f} : I\!Ry \to I\!R$. Now we use Theorem 2.5.4 (Hahn-Banach) for the affine set M being the graph of $\hat{f}$ translated by $(x, P(x))$ in $V \times I\!R$. for every $\alpha \in I\!R$ and every $n \geq 1$. We extend M to H and find the linear functional $\hat{F} \in V^*_{x,P}$ whose graph is H. Since $\hat{F}(y) - F(y) = \hat{f}(y) - F(y) = \varepsilon$, $\hat{F} \neq F$.

Suppose now that Proposition 2.5.4, $V^*_{x,P}$ contains at least two distinct linear functionals, say F and $\hat{F}$. So, $F(y) - \hat{F}(y) > 0$ for some $y \in V$. Suppose on the contrary that P is differentiable in every direction at the point x. In particular P is differentiable in the direction y. Hence

$$\lim_{t\to 0} \frac{P(x + ty) - P(x)}{t} = \lim_{t\to 0} \frac{P(x - ty) - P(x)}{-t}$$

and consequently

$$\lim_{t\to 0} \frac{P(x + ty) + P(x - ty) - 2P(x)}{t} = 0.$$

On the other hand, for every $t > 0$, we have $P(x + ty) - P(x) \geq F(t) = tF(y)$ and $P(x - ty) - P(x) \geq \hat{F}(-ty) = -t\hat{F}(y)$, hence

$$\liminf_{t\to 0} \frac{P(x + ty) + P(x - ty) - 2P(x)}{t} \geq F(y) - \hat{F}(y) > 0.$$

, a contradiction.

In fact $F(y) - \hat{F}(y) = \varepsilon > 0$ implies $F(y') - \hat{F}(y') \geq \varepsilon/2 > 0$ for all y' in the neighbourhood of y in the weak topology defined just by $\{y' : (F - \hat{F})(y - y') < \varepsilon/2\}$. Hence P is not differentiable in a weak*-open set of directions. ♣

Let us go back now to our special situation:

Proposition 2.5.6. If $\mu \in M(T)$ is an equilibrium state for $\phi \in C(X)$, then the linear functional represented by μ is tangent to P at ϕ.

Proof. We have

$$\mu(\phi) + h_\mu = P(\phi)$$

and for every $\psi \in C(X)$

$$\mu(\phi + \psi) + h_\mu \le P(\phi + \psi).$$

Subtracting the sides of the equality from the respective sides of the latter inequality we obtain $\mu(\psi) \le P(\phi + \psi) - P(\phi)$ which is just the inequality defining tangent functionals. ♣

As an immediate consequence of Proposition 2.5.6. and Theorem 2.5.5 we get the following.

Corollary 2.5.7. If the pressure function P is differentiable at ϕ in every direction, or at least in a dense in the weak topology set of directions, then there is at most one equilibrium state for ϕ.

Due to this Corollary, in future (see Chapter 4) to prove uniqueness it will be sufficient to prove differentiability of the pressure function in a weak*-dense set of directions.

The next part of this section will be devoted to kind of reversing Proposition 2.5.6 and Corollary 2.5.7. and better understanding of the mutual Legendre-Fenchel transforms $-h$ and P. This is a beautiful topic but will not have applications in the rest of this book. Let us start with a characterization of T invariant measures in the space of all signed measures $C(X)^*$ provided by the pressure function P.

Theorem 2.5.8. For every $F \in C(X)^*$ the following three conditions are equivalent:

(i) For every $\phi \in C(X)$ it holds $F(\phi) \le \mathrm{P}(\phi)$.

(ii) There exists $C \in I\!R$ such that for every $\phi \in C(X)$ it holds $F(\phi) \le \mathrm{P}(\phi) + C$.

(iii) F is represented by a probability invariant measure $\mu \in M(X,T)$.

Proof. (iii) $\Rightarrow$ (i) follows immediately from the variational principle:

$$F(\phi) \le F(\phi) + h_\mu(T) \le \mathrm{P}(\phi) \quad \text{for every} \quad \phi \in C(X).$$

(i) $\Rightarrow$ (ii) is obvious. Let us prove that (ii) $\Rightarrow$ (iii). Take an arbitrary positive $\phi \in C(X)$, i.e. such that for every $x \in X$, $\phi(x) \ge 0$. For every real $t < 0$ we have

$$F(t\phi) \le \mathrm{P}(t\phi) + C$$

Since $t\phi \le 0$ it immediately follows from (2.5.1) that $\mathrm{P}(t\phi) \le \mathrm{P}(0)$. Hence $F(t\phi) \le P(0) + C$. So

$$|t|F(\phi) \ge -(C + \mathrm{P}(0)) \quad \text{hence} \quad F(\phi) \ge \frac{-(C + \mathrm{P}(0))}{|t|}.$$

Letting $t \to -\infty$ we obtain $F(\phi) \ge 0$. We estimate the value of F on constant functions t. For every $t > 0$ we have $F(t) \le \mathrm{P}(t) + C \le \mathrm{P}(0) + t + C$. Hence $F(1) \le 1 + \frac{\mathrm{P}(0)+C}{t}$.

Similarly $F(-t) \le \mathrm{P}(-t) + C = \mathrm{P}(0) - t + C$ and therefore $F(1) \ge 1 - \frac{\mathrm{P}(0)+C}{t}$. Letting $t \to \infty$ we thus obtain $F(1) = 1$. Therefore by Theorem 2.1.1 (Riesz Representation Theorem) the functional F is represented by a probability measure $\mu \in M(X)$. Let us finally prove that μ is T-invariant. For every $\phi \in C(X)$ and every $t \in I\!R$ we have by (i)

$$F(t(\phi \circ T - \phi)) \le \mathrm{P}(t(\phi \circ T - \phi)) + C$$

It immediately follows from Theorem 2.3.1 (Variational principle) that $\mathrm{P}(t(\phi \circ T - \phi)) = \mathrm{P}(0)$. Hence

$$|F(\phi \circ T) - F(\phi)| \le \left|\frac{\mathrm{P}(0) + C}{t}\right|.$$

Thus, letting $|t| \to \infty$, we obtain $F(\phi \circ T) = F(\phi)$, i.e T-invariance of μ. ♣

We shall prove the following.
Corollary 2.5.9. Every functional F tangent to P at $\phi \in C(X)$, i.e. $F \in C(X)^*_{\phi,P}$, is represented by a probability T-invariant measure $\mu \in M(X,T)$.

Proof. Using Theorem 2.5.1, we get for every $\psi \in C(X)$ that

$$F(\psi) \le P(\phi + \psi) - P(\phi) \le P(\psi) + |P(\phi + \psi) - P(\psi)| - P(\phi) \le P(\psi) + ||\phi||_\infty - P(\phi).$$

So condition (ii) of Theorem 2.5.8 holds hence (iii) holds, F is represented by $\mu \in M(X,T)$. ♣

We can now almost reverse Proposition 2.5.6. Namely being a functional tangent to P at ϕ implies being an "almost" equilibrium state for ϕ.

Theorem 2.5.10. $F \in C(X)^*_{\phi,P}$ if and only if F, in other words the measure $\mu = \mu_F \in M(X,T)$ representing F, is a weak*-limit of measures $\mu_n \in M(X,T)$ such that

$$\mu_n\phi + \mathrm{h}_{\mu_n}(T) \to \mathrm{P}(\phi).$$

Proof. In one way the proof is simple. Assume that $\mu = \lim_{n\to\infty} \mu_n$ in the weak* topology and $\mu_n\phi + \mathrm{h}_{\mu_n}(T) \to P(\phi)$. We proceed as in Proof of Theorem 2.5.6. In view of Theorem 2.3.1 (Variational principle) $\mu_n(\psi + \phi) + \mathrm{h}_{\mu_n}(T) \le \mathrm{P}(\phi + \psi)$ which means that $\mu_n(\psi) \le \mathrm{P}(\phi+\psi) - (\mu_n\phi + \mathrm{h}_{\mu_n}(T))$. Thus, letting $n \to \infty$, we get $\mu(\psi) \le \mathrm{P}(\phi+\psi) - \mathrm{P}(\psi)$. This means that $\mu \in C(X)^*_{\phi,\mathrm{P}}$.

Now, let us prove our Theorem in the other direction. Recall again that the function $\mu \mapsto \mathrm{h}_\mu(T)$ on $M(T)$ is affine, hence concave. Denote $\overline{\mathrm{h}}_\mu = \limsup_{\nu\to\mu} \mathrm{h}_\nu(T)$, with $\nu \to \mu$ in weak*-topology. It is also concave and upper semicontinuous on $M(T) := M(X,T)$. In the sequel we shall prefer to consider the function $\mu \mapsto -\overline{\mathrm{h}}_\mu(T)$ which is lower semicontinuous and convex.

Step 1. For every $\vartheta \in C(X)$ we have

$$\mu\vartheta - \sup_{\nu\in M(T)} (\nu\vartheta - -\mathrm{h}_\nu(T)) \le \mu\vartheta - (\mu\vartheta - -\overline{\mathrm{h}}_\mu(T)) = -\overline{\mathrm{h}}_\mu(T)).$$

We obtained here $-\overline{\mathrm{h}}_\nu(T))$ rather than merely $-\mathrm{h}_\nu(T))$ by taking every sequence $\mu_n \to \mu$ writing the right hand side of the above inequality : $\mu\vartheta - (\mu_n\vartheta - -\mathrm{h}_{\mu_n}\vartheta(T))$ and letting $n \to \infty$. So

$$\sup_{\vartheta\in C(X)} \Big(\mu\vartheta - \sup_{\nu\in M(T)} (\nu\vartheta - -\mathrm{h}_\nu(T))\Big) \le -\overline{\mathrm{h}}_\mu(T).$$

This says that the LF-transform of the LF-transform of $-\mathrm{h}_\mu(T)$ is less or equal to $-\overline{\mathrm{h}}_\mu(T)$.

Let us prove now the opposite inequality. We refer to the following corollary of the geometric form of Hahn-Banach Theorem [Bourbaki, Ch.II.§5. Prop.5]:

Let M be a closed convex set in a locally convex vector space V. Then every lower semi-continuous convex function f defined in M is supremum of a family of functions bounded from above by f, which are restrictions to M of continuous affine functions in V.

We shall apply this theorem to $C^*(X)$ endowed with weak*-topology and use the fact that every linear functional continuous in this topology is represented by an element belonging to $C(X)$. (This is a general fact concerning pairs of vector spaces in duality, [Bourbaki, Ch.II.§6. Prop.3.].). Thus, for every $\varepsilon > 0$ there exists $\psi \in C(X)$ such that for every $\nu \in M(T)$

$$(\nu - \mu)(\psi) \le -\mathrm{h}_\nu(T) - -\overline{\mathrm{h}}_\mu + \varepsilon \tag{2.5.6}$$

So

$$\mu\psi - \sup_{\nu\in M(T)} (\nu\psi - -\mathrm{h}_\nu(T)) \ge -\overline{\mathrm{h}}_\mu(T) - \varepsilon.$$

Letting $\varepsilon \to 0$ we obtain

$$\sup_{\vartheta\in C(X)} \Big(\mu\vartheta - \sup_{\nu\in M(T)} (\nu\vartheta - -\mathrm{h}_\nu(T))\Big) \ge -\overline{\mathrm{h}}_\mu(T).$$

Thus we proved the standard fact that the LF-transform of the LF-transform of $-h_\mu(T)$ is back $-\overline{\mathrm{h}}_\mu(T)$. Remind now that by variational principle the LF-transform of $-\mathrm{h}_\nu(T)$, i.e. the supremum $\sup_{\nu\in M(T)}(\nu\vartheta - -\mathrm{h}_\nu(T))$ is pressure $\mathrm{P}(\vartheta)$. We conclude that

$$\overline{\mathrm{h}}_\mu(T) = \inf_{\vartheta\in C(X)} \{\mathrm{P}(\vartheta) - \mu\vartheta\} \tag{2.5.7}$$

Step 2. Fix $\mu \in C(X)^*_{\phi,\mathrm{P}}$. From $\mu\psi \le P(\phi+\psi) - P(\phi)$ we obtain

$$P(\phi+\psi) - \mu(\phi+\psi) \ge \mathrm{P}(\phi) - \mu\phi \quad \text{for all} \quad \psi \in C(X)$$

or

$$\inf_{\psi\in C(X)} \{\mathrm{P}(\psi) - \mu\psi\} \ge \mathrm{P}(\phi) - \mu\phi \tag{2.5.8}$$

(This expresses the fact that the supremum (– infimum above) in the definition of the LF-transform of P at F is attained at ϕ at which F is tangent to P.) By (2.5.7) and (2.5.8) we obtain

$$\overline{\mathrm{h}}_\mu \geq \mathrm{P}(\phi) - \mu\phi \tag{2.5.9}$$

so by the definition of $\overline{\mathrm{h}}_\mu$ there exists a sequence of measures $\mu_n \in M(T)$ such that $\lim_{n\to\infty} \mu_n = \mu$ and $\lim_{n\to\infty} h_{\mu_n} \geq P(\phi) - \mu\phi$. The proof is finished. ♣

Remark. In Step 1 of the above proof it did not matter whether we considered μ tangent to P or an arbitrary $\mu \in M(T)$. In Step 2, where we started with all $\mu \in M(T)$, considering $\varepsilon > 0$ in (2.5.6) is necessary; without $\varepsilon > 0$ this formula may happen to be false, see Example 2.5.13. For $\mu \in C(X)^*_{\phi,\mathrm{P}}$ we obtain from (2.5.9) and inequality $\mathrm{h}_\nu(T) \leq \mathrm{P}(\phi) - \nu\phi$ for every $\nu \in M(T)$ that $h_\nu(T) - \overline{\mathrm{h}}_\mu(T) \leq (\mu - \nu)\phi$ which is just (2.5.6) with $\varepsilon = 0$. So *a posteriori* we know that ε in (2.5.6) can be omitted.

The meaning of this, is that if μ is tangent to P at ϕ then ϕ is tangent to $-\overline{\mathrm{h}}$, the LF-transform of P, at μ.

Conversely, if ψ satisfies (2.5.6) with $\varepsilon = 0$ i.e. ψ is tangent to $-\overline{\mathrm{h}}$ at $\mu \in M(T)$ then in the same way as in Step 2. we can prove the inequality analogous to (2.5.8), namely that

$$\sup_{\nu \in M(T)} \nu\psi - -\overline{\mathrm{h}}_\mu(T) = \mathrm{P}(\psi) \leq \mu\psi - -\overline{\mathrm{h}}_\mu(T).$$

Hence μ is tangent to P at ψ (by the "if" part of Theorem 2.5.10).

Assume now the upper semicontinuity of the entropy $\mathrm{h}_\mu(T)$ as a function of μ. Then we obtain.

Corollary 2.5.11. If the entropy is upper semicontinuous, then a functional $F \in C(X)^*$ is tangent to P at $\phi \in C(X)$ if and only if it is represented by a measure which is an equilibrium state for ϕ.

Proof. This is just the previous Theorem with the observation that $\lim_{n\to\infty} \mathrm{h}_{\mu_n}(T) \leq \mathrm{h}_{\mu_T}(T)$. (Remark that this uses only the upper semicontinuity of the entropy at the measure μ.) ♣

Recall that already the upper semicontinuity above implies the existence of at least one equilibrium state (Lemma 2.4.3)

Now we can complete Corollary 2.5.7.

Corollary 2.5.12. If the entropy is upper semicontinuous then the pressure function P is differentiable at $\phi \in C(X)$ in every direction, or in a set of directions dense in the weak topology, if and only if there is at most one equilibrium state for ϕ.

Proof. This Corollary follows directly from Corollary 2.5.11 and Theorem 2.5.5. ♣

After discussing functionals tangent to P and proving that they coincide with the set of equilibrium states for maps for which the entropy is upper semicontinuous as the function on $M(T)$ the question arises of whether all measures in $M(T)$ are equilibrium states of some continuous functions. The answer given below is no.

Example 2.5.13. We shall construct a measure $m \in M(T)$ which is not an equilibrium state for any $\phi \in C(X)$. Here X is the one sided shift space Σ^2 with the left side shift map σ. Since this map is obviously expansive, it follows from Theorem 2.4.6 that the entropy function is upper semicontinues. Let $m_n \in M(\sigma)$ be the measure equidistributed on the set Per_n of points of period n, i.e.

$$m_n = \sum_{x \in \mathrm{Per}_n} \frac{1}{\mathrm{Card}\ \mathrm{Per}_n} \delta_x$$

where δ_x is the Dirac measure supported by x. m_n converge weak* to $\mu_{\max}$, the measure of maximal entropy: $\log 2$. (Check that this follows for example from the proof of variational principle, part II.) Let $t_n, n = 0, 1, 2, ...$ be a sequence of positive real numbers such that $\sum_{n=0}^{\infty} t_n = 1$. Finally define

$$m = \sum_{n=0}^{\infty} t_n m_n$$

Let us prove that there is no $\phi \in C(X)$ tangent to h at m. Let $\mu_n = R_n \mu_{\max} + \sum_{j=0}^{n-1} t_j m_j$, where $R_n = \sum_{j=n}^{\infty} t_j$. We have of course $\mathrm{h}_{m_n}(\sigma) = 0,\ n = 1, 2, ...$ Also $\mathrm{h}_m(\sigma) = 0$. This follows from the fact that h is affine on $M(\sigma)$, the function h is bounded by the topological entropy $\mathrm{h}_{\mathrm{top}}(\sigma) = \log 2$ and

$$m = \sum_{j=0}^{n-1} t_j m_j + R_n \sum_{j=n}^{\infty} \frac{t_j}{R_n} m_j.$$

Thus

$$\mathrm{h}_{\mu_n}(\sigma) - \mathrm{h}_m(\sigma) = R_n \mathrm{h}_{\mu_{\max}}(\sigma) = R_n \log 2$$

and for an arbitrary $\phi \in C(\Sigma^2)$

$$(\mu_n - m)\phi = (R_n \mu_{\max} - \sum_{j=n}^{\infty} t_j m_j)\phi \le R_n \varepsilon$$

where $\varepsilon \to 0$ as $n \to \infty$ because $m_j \to \mu_{\max}$. The inequality $\mathrm{h}_{\mu_n}(\sigma) - \mathrm{h}_m(\sigma) \ge R_n \log 2$ and the latter inequality prove that ϕ is not tangent to h at m. Indeed $\mathrm{h}_{\mu_n}(\sigma) - \mathrm{h}_m(\sigma) > (\mu_n - m)\phi$ for n large, contrary to (2.5.5) with $\varepsilon = 0$.

By Remark after Theorem 2.5.10 we know that m is not tangent to any ϕ for the pressure function P. In fact it is easy to see it directly: For an arbitrary $\phi \in C(\Sigma^2)$ we have $\mu_{\max}\phi < \mathrm{P}(\phi)$ because $\mathrm{h}_{\mu_{\max}}(\sigma) > 0$, so $m_n\phi < P(\phi)$ for all n large enough as

$m_n \to \mu_{\max}$. Also $m_n\phi \le P(\phi)$ for all n's. So for the average of m_n's namely m we have $m\phi < \mathrm{P}(\phi)$. So ϕ is not an equilibrium state. ♣

The measure m in this example is very non-ergodic, this is necessary as will follow from Exercise 5.

EXERCISES

Exercise 1. Let $T : X \to X$ and $S : Y \to Y$ be two continuous maps of compact metric spaces respectively. Show that $\mathrm{h_{top}}(T \times S) = \mathrm{h_{top}}(T) + \mathrm{h_{top}}(S)$.

Exercise 2. Prove that $T : X \to X$ is an isometry of a compact metric space X, then $\mathrm{h_{top}}(T) = 0$

Exercise 3. Show that if $T : X \to X$ is a local homeomorphism of a compact metric space, then the number $d = \#T^{-1}(x)$ is finite and independent of $x \in X$.

Exercise 4. With the assumtions and notation of Exercise 3, demonstrate that $\mathrm{h_{top}}(T) \ge \log d$

Exercise 5. Prove that if $f : M \to M$ is a C^1 endomorphism of a compact differentiable manifold M, then $\mathrm{h_{top}}(T) \ge \log \deg(f)$. (Hint: see [MP]).

Exercise 6. Let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle and let $f_n : S^1 \to S^1$ be the map defined by the formula $f_n(z) = z^n$. Show that $\mathrm{h_{top}}(f_n) = \log n$.

Exercise 7. Let $\sigma_A : \Sigma_A \to \Sigma_A$ be the shift map generated by the incidence matrix A. Prove that $\mathrm{h_{top}}(\sigma_A)$ is equal to the logarithm of the spectral radius of A.

Exercise 8. Show that for every continuous potential ϕ, $\mathrm{P}(\phi) \le \mathrm{h_{top}}(T) + \sup(\phi)$.

Exercise 9. Provide an example of a transitive diffeomorphism without measures of maximal entropy.

Exercise 10 Provide an example of a transitive diffeomorphism with at least measures of maximal entropy.

Exercise 11. Find a sequnce of continuous maps $T_n : X_n \to X_n$ such that $\mathrm{h_{top}}(T_{n+1}) > \mathrm{h_{top}}(T_n)$ and $\lim_{n\to\infty} \mathrm{h_{top}}(T_n) < \infty$.

Exercise 12. Prove that for an arbitrary convex continuous function $P : V \to \mathbb{R}$ on a real Banach space V the set of tangent functionals: $\bigcup_{x\in V} V^*_{x,P}$ is dense in the norm topology in

$$\{F \in V^* :\ \text{there exists } C \in \mathbb{R} \text{ such that for every } x \in V,\ F(x) \le P(x) + C\}$$

(such functionals are called P-bounded)

Remark. The conclusion is that for P the pressure function on $C(X)$ tangent measures are dense in $M(X,T)$, see Theorem 2.4.6. Hint: This follows from Bishop – Phelps Theorem, see [BP] or Israel's book [I, pp.112-115], which can be stated as follows: For every P-bounded F_0 $x_0 \in V$ and $\varepsilon > 0$ there exists $x \in V$ and $F \in V^*$ tangent to P at X such that

$$\|F - F_0\| \le \varepsilon \text{ and } \|x - x_0\| \le \frac{1}{\varepsilon}\Big(P(x_0) - F_0(x_0) + s(x_0)$$

where $s(F_0) := \sup_{x' \in V} F_0 x' - P(x')$ (the LF-transform of P. The idea of the proof of this theorem is as follows: If we replace P by $Q(x0 := P(x) - F_0(x) + s(x_0)$ the theorem reduces to the case $F_0 \equiv 0, s(F_0) = 0$. For each $x \in V$ consider the cone

$$\mathcal{C}(x) = \{(x', y) : y - Q(x) < -\varepsilon \|x' - x\|\}.$$

There is $x \in V$ such that $\mathcal{C}(x) \cap \mathrm{graph} Q = \{x\}$ Now F can be defined as a functional which graph translated by a constant separates $\mathcal{C}(x)$ from $\{y \geq Q(x)$.

Exercise 13. Prove that in the situation from Exercise 1 for every $x \in V$ $V^*_{x,P}$ is convex and weak*-compact.

Exercise 14. Let E_ϕ denote the set of all equilibrium states for $\phi \in C(X)$.

(i) Prove that E_ϕ is convex.

(ii) Find an example that E_ϕ is not weak*-compact.

(iii) Prove that extremal points of E_ϕ are extremal points of $M(X,T)$.

(iv) Prove that almost all measures in the ergodic decomposition of an arbitrary $\mu \in E_\phi$ belong also to E_ϕ. (One says that every equilibrium state has a unique decomposition into pure, i.e. ergodic, equilibrium states .)

Hints: In (ii) consider a sequence of Smale horseshoes of topological entropies $\log 2$ converging to a point fixed for T. To prove (iii) and (iv) use the fact that entropy is an affine function of measure.

Exercise 15. Find an example showing that the point (iii) of Exercise 3 is false if we consider $C(X)^*_{\phi,P}$ rather than E_ϕ.

Hint: An idea is to have two fixed points p, q and two trajectories $(x_n), (y_n)$ such that $x_n \to p, y_n \to q$ for $n \to \infty$ and $x_n \to q, y_n \to p$ for $n \to -\infty$. Now take a sequence of periodic orbits γ_k approaching $\{p, q\} \cup \{x_n\} \cup \{y_n\}$ with periods tending to ∞. Take their Cartesian products with corresponding invariant subsets A_k's of small horseshoes of topological entropies less than $\log 2$ but tending to $\log 2$, diameters of the horseshoes shrinking to 0 as $k \to \infty$. Then for $\phi \equiv 0$ $C(X)^*_{\phi,P}$ consists only of measure $\frac{1}{2}(\delta_p + \delta_q)$. One cannot repeat the proof in Exercise 3(iii) with the function $\overline{\mathrm{h}}_\mu$ instead of the entropy function h_μ, because $\overline{\mathrm{h}}_\mu$ is no more affine !

This is Peter Walters' example, for details see the preprint [W2].

Exercise 16. Suppose that the entropy function h_μ is upper semicontinuous (then for each $\phi \in C(x)$ $C(X)^*_{\phi,\mathrm{P}} = E_\phi$, see Corollary 2.5.11). Prove that

(i) every $\mu \in M(T)$ which is a finite combination of ergodic masures $\mu = \sum t_j m_j$, $m_j \in M(T)$, is tangent to P more precisely there exists $\phi \in C(X)$ such that $\mu, m_j \in C(X)^*_{\phi,\mathrm{P}}$ and moreover they are equilibrium states for ϕ.

(ii) if $\mu = \int_{M_e(T)} m d\alpha(m)$ where $M_e(X,T)$ consists of ergodic measures in $M(X,T)$ and α is a probability non-atomic measure on $M_e(X,T)$, then there exists $\phi \in C(X)$ which has uncountably many ergodic equilibria in the support of α.

(iii) the set of elements of $C(X)$ with uncountably many ergodic equilibria is dense in $C(X)$.

Hint: By Bishop – Phelps Theorem (Exercise 12) there exists $\nu \in E_\phi$ arbitrarily close to μ. Then in its ergodic decomposition there are all the measures μ_j because all ergodic measures are far apart from each other (in the norm in $C(X)^*$). These measures by Exercise 14 belong to the same E_ϕ what proves (i). For more details and proofs of (ii) and (iii) see [Israel, Theorem V.2.2] or [Ruelle, 1978, 3.17, 6.15].

Remark. In statistical physics the occurence of more then one equilibrium for $\phi \in C(X)$ is called "phase transition". (iii) says that the set of functions with "very rich" phase transition is dense. For the further discussion see also [Israel, V.2].

Exercise 17. Prove the following. Let $P : V \to I\!R$ be a continuous convex function on a real Banach space V with norm $\|\cdot\|_V$. Suppose P is differentiable at $x \in V$ in every direction. Let $W \subset V$ be an arbitrary linear subspace with norm $\|\cdot\|_W$ such that the embedding $W \subset V$ is continuous and the unit ball in $(W, \|\cdot\|_W)$ is compact in $(V, \|\cdot\|_V)$. Then $P|_W$ is differentiable in the sense that there exists a functional $F \in V^*$ such that for $y \in W$ it holds

$$|P(x+y) - P(x) - F(y)| = o(\|y\|_W).$$

Remark. In Chapter 3 we shall discuss W being the space of Hölder continuous functions with an arbitrary exponent $\alpha < 1$ and the entropy function will be upper semicontinuous. So the conclusion will be that uniqueness of the equilibrium state at an arbitrary $\phi \in C(X)$ is equivalent to the differentiability in the direction of this spaece of Hö .lder functions.

Exercise 18. (Walters) Prove that the pressure function P is Frechet differentiable at $\phi \in C(X)$ if and only if P affine in a neighbourhood of ϕ. Prove also the conclusion: P is Frechet differentiable at every $\phi \in C(X)$ if and only if T is uniquely ergodic, namely if $M(X,T)$ consists of one element.

Exercise 19. Prove S. Mazur's Theorem: If $P : V \to I\!R$ is a continuous convex function on a real separable Banach space V then the set of points at which there exists a unique functional tangent to P is dense G_δ.

Remark. In the case of the pressure function on $C(X)$ this says that for a dense G_δ set of functions there exists at most one equilibrium state. Mazur's Theorem contrasts with the theorem from Exercise 16 (iii).

BIBLIOGRPHICAL NOTES

The concept of topological pressure in dynamical context was introduced by D. Ruelle in [Ruelle, 1973] and since then have been studied in many papers and books. Let us mention only [Bowen, 1975], [Wallters, 1976], [Wallters, 1982] and [Ruelle, 1978]. The topological entropy was introduced earlier in [AKM, 1965]. The variational principle (Theorem 2.3.1) has been proved for some maps in [Ruelle, 1973]. The first proofs of this principle in its full generality can be found in [Walters, 1976] and [Bowen, 1975]. The simplest proof presented in this chapter is taken from [Mi, 1976]. In the case of topological entropy (potential $\phi = 0$) the corresponding results have been obtained earlier: Goodwyn in [Goodwyn, 1969] proved the first part of the variational principle, Dinaburg in [Dinaburg, 1971] proved its full version assuming that the space X has finite covering topological dimension and finally

Goodman proved in [Goodman, 1971] the variational principle for topological entropy without any additional assumptions. The concept of equilibrium states and expansive maps in mathematical setting was introduced in [Ruelle, 1973] where the first existence and uniqueness type results have appeared. Since then these concepts have been explored by many authors, in particular in [Bowen, 1975] and [Ruelle, 1978]. The material of Section 2.5 is mostly taken from [Ruelle, 1978], [Israel, 1979] and [Ellis, 1985].

References

N.Bourbaki: Espaces Vectoriels Topologiques. Masson,Paris, 1981.

M.Denker, Ch.Grillenberger, K.Sigmund: Ergodic Theory on Compact Spaces. Lecture Notes in Mathematics, 527, Springer-Verlag, Berlin 1976.