STUDIA MATHEMATICA, T. LXXII. (1982)

The variational principle

by

JEAN MOULIN OLLAGNIER (Villetaneuse) and
DIDIER PINCHON (Paris)

Abstract. This paper propose a proof (in a whole generality) of the variational principle for the pressure of a continuous function on a compact space under the action of an amenable group of homeomorphisms of this compact space. There are essentially two novelties in our proof: on the one hand, the method for approximating a borelian partition by some open cover, and on the other hand the use of the extension of capacities which allows in the second part of the proof to keep the ideas of M. Misiurewicz without using technics of pavability which were absolutely necessary for him.

In their paper of 1965, R. L. Adler, A. G. Konheim and M. H. McAndrew [1] introduced the notion of topological entropy for a continuous map defined from a compact space to itself. This definition was proposed to be an equivalent in the topological frame of the entropy of measure preserving transformations on measured space which was introduced by soviet probabilists [17] and [19]. The first conjecture proposed by these three makers called upon a comparison of the two kinds of entropy: the topological entropy bounds the metric entropy.

The first definition of topological entropy used open covers of the compact space; by introducing finite separated sets, R. Bowen [2] gave an equivalent definition.

L. W. Goodwyn [10] in 1969 showed that the topological entropy bounds the metric entropy.

E. I. Dinaburg [6] and T. N. T. Goodman [9] proved in 1971 that the topological entropy is exactly the upper bound of the metric entropies for all invariant probabilities. This kind of result is called a *variational principle.*

Goodwyn [11] gave a proof of the variational principle for a locally non-recurrent transformation. We must note the contributions of M. Denker in the study of these topics [4], [5].

Some problems of statistical mechanics lead D. Ruelle [18] to generalize in 1973 this result in two ways: he introduced the concept of pressure for a continuous function and that for the action of the group Z^n. The topological entropy is the pressure of the null function. Hypotheses of specification and expansiveness allow him to prove the variational principle for the pressure.

P. Walters studied in [20] the properties of the pressure for the action of one transformation and proved the variational principle by the method of Goodwyn that consists in an approximation of the dynamical system by symbolic systems. The proof of Walters was extended by S. A. Elsanousi [8] to the action of Z^2.

For the action of the semi-group $Z^{+\bar{n}}$, a new and very short proof of the variational principle for the pressure was obtained by M. Misiurewicz [13]. In fact, Misiurewicz's method works for the action of a pavable group [14].

The kind of action studied by Goodwyn (locally non-recurrent dynamical systems) allows to get an approximation theorem by means of symbolic systems. For such systems, one obtain not only the variational principle but also a new definition of the pressure, similar to the definition of the entropy; we do that in [15] by substituting to subshifts of finite type some models of statistical mechanics on a lattice.

This paper proposes a proof (in a whole generality) of the variational principle for the pressure of a continuous function on a compact space under the action of an amenable group of homeomorphisms of this compact space. There are essentially two novelties in our proof: on the one hand, the method for approximating a borelian partition by some open cover, and on the other hand the use of the extension of capacities which allows in the second part of the proof to keep the ideas of Misiurewicz without using technics of pavability which were absolutely necessary for him.

The references are restricted to papers dealing with the variational principle; works about equilibrium measures, for instance, are not quoted. For such expositions, we refer the reader to the excellent bibliography of M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes 527, Springer-Verlag, 1976.

The pressure of a continuous function. Let X be a compact space and G a group of homeomorphisms of X. W denotes the set of all symmetric neighbourhoods of the diagonal Δ of $X\times X$ (it is a basis of the uniform filter). If A is a finite part of G and δ an element of W, we denote by δ_A the element of W

$$\delta_A = \bigcap_{g\in A} (g^{-1}\times g^{-1})(\delta).$$

A subset E of X is said to be δ-separated if any two distinct elements of E are not neighbours of order δ. A δ-separated set is necessarily finite because X is compact.

Let f be a real valued continuous function on X. For a finite part A of G, we denote by f_A the continuous function $\sum_{g\in A} f\circ g$.

If E is a finite subset of X, we set

$$Z(f, E) = \sum_{x\in E} \exp f(x).$$

For an element δ of W and a continuous function f, we denote by $P_1(f, \delta)$ the upper bound of the numbers $Z(f, E)$ on the set of all δ-separated sets E. We then define

$$p_1(f, \delta) = \limsup_{\mathscr{M}} |A|^{-1} \operatorname{Log} P_1(f_A, \delta_A)$$

where $\mathscr{M}$ denotes the ameaning filter on the set of all finite parts of G and $|\cdot|$ the cardinal number of a finite part. We must remark that if δ and δ' are two elements of W such that $\delta\subset\delta'$, then

$$p_1(f, \delta) \geqslant p_1(f, \delta')$$

because for every finite part A of G,

$$P_1(f_A, \delta_A) \geqslant P_1(f_A, \delta'_A).$$

Thus we can put the definition

$$p_1(f) = \lim_{\delta\to\Delta} p_1(f, \delta) = \sup_{\delta\in W} p_1(f, \delta).$$

For a finite open cover α of X and a continuous function f, let us define

$$Z(f, \alpha) = \sum_{O\in\alpha} \sup_{x\in O} \exp f(x).$$

For an element δ of W, we denote by $P_2(f, \delta)$ the lower bound of the numbers $Z(f, \alpha)$ where α runs over the set of all finite open covers of order δ. Recall that a cover α is of order δ if every element O of α satisfies $O\times O\subset\delta$. We then define

$$p_2(f, \delta) = \limsup_{\mathscr{M}} |A|^{-1} \operatorname{Log} P_2(f_A, \delta_A).$$

Here again, if $\delta'\subset\delta$, $p_2(f, \delta')\geqslant p_2(f, \delta)$, that allows to define $p_2(f)$

$$p_2(f) = \lim_{\delta\to\Delta} p_2(f, \delta) = \sup_{\delta\in W} p_2(f, \delta).$$

THEOREM. $p_2(f) = p_1(f)$.

Proof. (1) $p_1(f)\leqslant p_2(f)$. Let δ be an element of W. We have $p_1(f, \delta)\leqslant p_2(f, \delta)$. It is sufficient to show that, for every δ-separated set E and every open cover of order δ, $Z(f, E)\leqslant Z(f, \alpha)$. To each element x of E we associate an open set of α that contains it and this map is one-to-one.

We deduce that for every finite part A of G and every δ of W, $P_1(f_A, \delta_A) \leqslant P_2(f_A, \delta_A)$ and the inequality $p_1(f) \leqslant p_2(f)$.

(2) $p_2(f) \leqslant p_1(f)$. The proof of this inequality results essentially from the following lemma.

LEMMA. *Let φ be a continuous function, ζ a positive real number and γ an element of W such that*

$$\forall (x, y) \in \gamma, \ |\varphi(x) - \varphi(y)| < \zeta$$

and ω an open element of W such that $\omega \circ \omega \subset \gamma$. Then

$$P_2(\varphi, \gamma) \leqslant P_1(\varphi, \omega) \cdot \exp \zeta.$$

Proof of the lemma. To calculate $P_1(\varphi, \omega)$, it is sufficient to take the upper bound of the $Z(\varphi, E)$ on the set of maximal ω-separated sets E, which are necessarily ω-generating sets. This means that the neighbourhoods of order ω of the elements of E are an open cover α of order δ. For such a set E, we have $Z(\varphi, \alpha) \leqslant Z(\varphi, E) \cdot \exp \zeta$ and the conclusion of the lemma follows.

Proof of the inequality. Let η be a positive real number, δ an element of W such that

$$\forall (x, y) \in \delta, \ |f(x) - f(y)| < \eta$$

and ε an element of W such that $\varepsilon \circ \varepsilon \subset \delta$. For every finite part A of G, $\gamma = \delta_A$, $\omega = \varepsilon_A$, $\zeta = |A| \cdot \eta$, $\varphi = f_A$ satisfy the hypothesis of the lemma. Therefore, after taking the lim sup,

$$p_2(f, \varepsilon) \leqslant p_1(f, \delta) + \eta.$$

The inequality $p_2(f) \leqslant p_1(f)$ follows.

The numbers $p_1(f)$ and $p_2(f)$, which are equal when the function f is continuous, can be defined for every bounded function on X. But the continuity of f was of an essential use to prove their equality. We denote by $p(f)$ their common value and we call this number the *pressure* of the continuous function f.

The equivalence of the two preceding definitions of the pressure of a continuous function makes no use of the ameaning property of the filter $\mathscr{M}$. On the other hand, this property is decisive when watching at the pressure as a function on $C(X)$.

THEOREM. *Let f and φ be two continuous functions on X. We denote by s the sub-additive, positively homogeneous functional on $C(X)$ defined by*

$$s(\varphi) = \limsup_{\mathscr{M}} |A|^{-1} \sup_{x \in X} \sum_{g \in A} \varphi \circ g(x).$$

Then

$$p(f+\varphi) - p(f) \leqslant s(\varphi).$$

Proof. It is sufficient to look at the finite rank inequality

$$P_1((f+\varphi)_A, \delta_A) \leqslant P_1(f_A, \delta_A) \cdot \sup_{x \in X} \exp \sum_{g \in A} \varphi \circ g(x).$$

COROLLARIES. *The following properties of the pressure are immediately deduced from the properties of the functional s:*

p is non-decreasing: $p(f) \leqslant p(f+\varphi)$ if φ is a positive continuous function.

p is Lipschitz: $|p(f_1) - p(f_2)| \leqslant \|f_1 - f_2\|$; $p(f+\varphi\circ g - \varphi) = p(f)$ for every continuous φ and every g in G.

The variational principle. For any Borel probability measure μ on X, invariant under the action of G, one can consider the dynamical system $(X, \mathscr{A}, \mu, G)$. $h(\mu)$ denotes the metric entropy of this dynamical system, i.e. $h(\mu) = \sup_\alpha h(\mu, \alpha)$ where the upper bound is taken over the set of all finite Borel partitions of X. $h(\mu, \alpha)$ is the limit along the ameaning filter and the lower bound of the numbers $|A|^{-1} H(\mu, \alpha_A)$ (see [16]).

The main result of this paper is the following theorem.

VARIATIONAL PRINCIPLE. *Let f be a continuous function on X. $p(f)$ is the upper bound of the numbers $h(\mu) + \mu(f)$ when μ runs over the set of all G-invariant Radon probabilities on X.*

Proof of the variational principle. *First part.* The inequality $h(\mu) + \mu(f) \leqslant p(f)$ for a G-invariant Radon probability μ on X immediately results from the following approximation theorem.

THEOREM. *Let μ be a G-invariant Radon probability on X and α a finite Borel partition of X. For every positive real number ξ, there exists a finite open cover δ of X such that*

$$h(\mu, \alpha) + \mu(f) \leqslant p_2(f, \delta) + \xi.$$

The proof of this theorem is essentially based on the introduction of a real number $R(\mu, \delta)$ which measures the covering ratio of a finite open cover δ relatively to a G-invariant probability μ and on the sub-additivity of the function R for δ.

DEFINITIONS. A finite partition α of X is said to be *δ-adapted* if there exists a one-to-one map j from α to δ such that $K \subset j(K)$ for every K in α.

We call *covering ratio of δ for the probability μ* the number $R(\mu, \delta)$ defined by

$$R(\mu, \delta) = \sup_{\alpha, \beta} H(\mu, \alpha/\beta)$$

where α and β run over the set of all δ-adapted partitions.

PROPOSITION. *The function R is sub-additive: if δ_1 and δ_2 are two* *finite open covers of X, then*

$$R(\mu, \delta_1 \vee \delta_2) \leqslant R(\mu, \delta_1) + R(\mu, \delta_2).$$

Proof. A $\delta_1 \vee \delta_2$-adapted partition α has the form $\alpha_1 \vee \alpha_2$ where α_1 and α_2 are δ_1-adapted and δ_2-adapted partitions, respectively. The conditional entropy $H(\mu, \alpha/\beta)$ is sub-additive in α:

$$H(\mu, \alpha_1 \vee \alpha_2/\beta_1 \vee \beta_2) \leqslant H(\mu, \alpha_1/\beta_1 \vee \beta_2) + H(\mu, \alpha_2/\beta_1 \vee \beta_2)$$

and a non-increasing function of the conditional partition β, thus

$$H(\mu, \alpha_1 \vee \alpha_2/\beta_1 \vee \beta_2) \leqslant H(\mu, \alpha_1/\beta_1) + H(\mu, \alpha_2/\beta_2)$$

and the result of the proposition follows.

Proof of the theorem. Let α be a finite Borel partition of X. We construct an open cover δ such that α is δ-adapted by choosing an open neighbourhood of each atom of α.

Let A be a finite subset of G. The partition α_A is δ_A-adapted.

Let δ' be an open cover finer than δ_A and γ a δ'-adapted partition. As δ' is finer than δ_A, we may construct a map θ from δ' to δ_A such that, for every O in δ', $O \subset \theta(O)$. Let j be the one-to-one map from γ to δ'. The partition β, obtained by gluing together the atoms of γ with the same range by $\theta \circ j$ is δ_A-adapted and γ is finer than β. Therefore

$$H(\mu, \alpha_A) \leqslant H(\mu, \beta) + H(\mu, \alpha_A/\beta)$$

and

$$\begin{aligned} H(\mu, \alpha_A) + \mu(f_A) &\leqslant H(\mu, \beta) + R(\mu, \delta_A) + \mu(f_A) \\ &\leqslant R(\mu, \delta_A) + H(\mu, \gamma) + \mu(f_A). \end{aligned}$$

But, by Jensen's inequality

$$H(\mu, \gamma) + \mu(f_A) \leqslant \operatorname{Log} Z(f_A, \gamma)$$

and

$$Z(f_A, \gamma) \leqslant Z(f_A, \delta').$$

By taking the lower bound over δ', it comes

$$H(\mu, \alpha_A) + \mu(f_A) \leqslant R(\mu, \delta_A) + \operatorname{Log} P_2(f_A, \delta_A).$$

The sub-additivity of $R(\mu, \delta_A)$ gives

$$|A|^{-1} H(\mu, \delta_A) + \mu(f) \leqslant R(\mu, \delta) + |A|^{-1} \operatorname{Log} P_2(f_A, \delta_A).$$

Then, taking the limsup along the ameaning filter,

$$h(\mu, \alpha) + \mu(f) \leqslant R(\mu, \delta) + p_2(f, \delta).$$

To achieve the proof of the theorem it remains to show that δ can be chosen such that $R(\mu, \delta)$ is arbitrarily small. As μ is a Radon measure, for every positive ε and every atom K of α, there exists an open neighbourhood O of K such that $\mu(O - K) < \varepsilon$. That $R(\mu, \delta)$ tends to zero with ε is easy to verify.

Proof of the variational principle. *Second part.* The inequality $\sup_{\mathcal{M}(X,G)} h(\mu) + \mu(f) \geqslant p(f)$ arises from the following theorem.

THEOREM. *For every open δ in W there exists a G-invariant probability μ such that*

$$h(\mu) + \mu(f) \geqslant p_1(f, \delta).$$

Proof. Let A be a finite subset of G and E a δ_A-separated set such that

$$Z(f_A, E) \geqslant P_1(f_A, \delta_A) \cdot \exp(-a)$$

where a is an arbitrary fixed positive real number. Let us consider the probability σ_A, with support E, defined by

$$\sigma_A(\{y\}) = Z(f_A, E)^{-1} \exp f_A(y) \quad \text{for} \quad y \text{ in } E.$$

μ_A is the probability $\sum_{g \in A} g(\sigma_A)$.

Let $\mathcal{N}$ be a filter finer than the ameaning filter $\mathcal{M}$ such that

$$\limsup_{\mathcal{M}} |A|^{-1} \operatorname{Log} P_1(f_A, \delta_A) = \lim_{\mathcal{N}} |A|^{-1} \operatorname{Log} P_1(f_A, \delta_A)$$

and $\lim_{\mathcal{N}} \mu_A = \mu$. μ is a G-invariant probability. Let us show that μ is convenient for the announced inequality. Let α be a finite Borel partition of order δ; moreover, we may suppose that the boundaries of atoms of α have μ-measure zero.

Since α is of order δ, every atom of α contains at most one element of E. Thus

$$H(\sigma_A, \alpha_A) + \sigma_A(f_A) = \operatorname{Log} Z(f_A, E) \geqslant \operatorname{Log} P_1(f_A, \delta_A) - a.$$

Let B be a finite subset of G containing the unit element e. For every element b of B, we may write

$$1_A = B^{-1} \sum_{a \in A_b} 1_{Bb^{-1}a} + A_b''$$

where $A_b = \{a, a \in A \text{ and } Bb^{-1}a \subset A\}$ and A_b'' is a function between 0 and 1 whose support verifies

$$|\operatorname{supp} A_b''| \leqslant m_{BB^{-1}}(A).$$

The function m_D, for a finite D, is defined by

$$m_D(A) = |\{g \in A, \exists d \in D, dg \notin A\}|.$$

These functions m_D are strongly sub-additive and have zero mean if G is an amenable group (see the second part of [16]). The map which associates to a finite subset C of G the number $\varphi(\sigma_A, C) = H(\sigma_A, \alpha_C)$ is strongly sub-additive; therefore it has an extension to functions on G with finite support (see again the second part of [16]. So we may write, for every element b of B,

$$H(\sigma_A, \alpha_A) \leqslant \sum_{a\in A_b} |B|^{-1} H(\sigma_A, \alpha_{Bb^{-1}a}) + \varphi(\sigma_A, A_b'') \tag{1}$$

$\varphi(\sigma_A, A_b'')$ is bounded above by $|\mathrm{supp}\, A_b''|\cdot \mathrm{Log}\,|\alpha|$ where $|\alpha|$ denotes the number of atoms in α.

Averaging the inequalities (1) for all the elements b of B leads to

$$H(\sigma_A, \alpha_A) \leqslant |B|^{-1} \sum_{b\in B} \sum_{a\in A_b} |B|^{-1} H(\sigma_A, \alpha_{Bb^{-1}a}) + m_{BB^{-1}}(A)\cdot \mathrm{Log}\,|\alpha|.$$

As $H(\sigma_A, \alpha_{Bb^{-1}a}) = H(b^{-1}a(\sigma_A), \alpha_B)$, adding some positive terms, we get

$$H(\sigma_A, \alpha_A) \leqslant |B|^{-1} \sum_{b\in B} \sum_{b^{-1}a\in A} |B|^{-1} H(b^{-1}a(\sigma_A), \alpha_B) + m_{BB^{-1}}(A)\cdot \mathrm{Log}\,|\alpha|.$$

Then

$$|A|^{-1} H(\sigma_A, \alpha_A) \leqslant |A|^{-1} \sum_{a\in A} |B|^{-1} H(a(\sigma_A), \alpha_B) + |A|^{-1} m_{BB^{-1}}(A)\cdot \mathrm{Log}\,|\alpha|.$$

As entropy is a concave function of the measure

$$|A|^{-1} \sum_{a\in A} H(a(\sigma_A), \alpha_B) \leqslant H(\mu_A, \alpha_B).$$

On the other hand, $\sigma_A(f_A) = |A|\cdot \mu_A(f)$. Therefore

$$|A|^{-1} \mathrm{Log}\, P_1(f_A, \delta_A) \leqslant |B|^{-1} H(\mu_A, \alpha_B) + \mu_A(f) + |A|^{-1} m_{BB^{-1}}(A)\cdot \mathrm{Log}\,|\alpha| + |A|^{-1}\cdot a.$$

Taking the lim sup in A along the filter $\mathcal{N}$, we obtain

$$p_1(f, \delta) \leqslant |B|^{-1} H(\mu, \alpha_B) + \mu(f).$$

As $\inf_B H(\mu, \alpha_B) = h(\mu, \alpha)$,

$$p_1(f, \delta) \leqslant h(\mu, \alpha) + \mu(f). \blacksquare$$

References

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

[2] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, ibid. 153 (1971), 401–414.

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

[4] M. Denker, *Une démonstration nouvelle du théorème de Goodwyn*, C. R. Acad. Sci. Paris 275 (1972), 735–738.

[5] — *Remarques sur la pression pour les transformations continues*, ibid. 279 (1974), 967–970.

[6] E. I. Dinaburg, *The relation between topological entropy and metric entropy*, Soviet Math. Dokl. 11 (1970), 13–16.

[7] E. Eberlein, *On topological entropy of semigroups of commuting transformations*, Astérisque 40 (1977).

[8] S. A. Elsanousi, *A variational principle for the pressure of a continuous Z^2-action*, Amer. J. Math. 99 (1977), 77–100.

[9] T. N. T. Goodman, *Relating topological entropy with measure theoretic entropy*, Bull. London Math. Soc. 3 (1971), 176–180.

[10] L. W. Goodwyn, *Topological entropy bounds measure-theoretic entropy*, Proc. Amer. Math. Soc. 23 (1969), 679–688.

[11] — *Comparing topological entropy with measure-theoretic entropy*, Amer. J. Math. 94 (1972), 366–388.

[12] J. C. Kieffer, *A generalized Shannon–McMillan theorem for the action of an amenable group on a probability space*, Ann. Probability 3 (1975), 1031–1037.

[13] M. Misiurewicz, *A short proof of the variational principe for a Z^{+n}-action on a compact space*, Astérisque 40 (1977), 147–157.

[14] J. Moulin Ollagnier et D. Pinchon, *Groupes pavables et principe variationnel*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 48 (1979), 71–79.

[15] — *Equilibrium measures in statistical mechanics and the variational principle in Riemannian dynamical system*, ibidem, to appear.

[16] — *Une nouvelle démonstration du théorème de E. Folmer*, C.R. Acad. Sci. Paris 287 (1978), 557–560.

[17] V. A. Rochlin, *Lectures on the entropy theory of measure-preserving transformations*, Russian Math. Surveys 22 (1967), 1–52.

[18] D. Ruelle, *Statistical mechanics on a compact set with Z^n-action*, Trans. Amer. Math. Soc. 185 (1973), 237–252.

[19] Ja. G. Sinai, *On the concept of entropy of a dynamical system*, Dokl. Akad. Nauk. SSSR 124 (1959), 768–771.

[20] P. Walters, *A variational principle for the pressure of continouus transformations*, Amer. J. Math. 97 (1976), 937–971.

[21] K. Zieman, *On topological and measure entropy of semi-groups*, Astérisque 51 (1978), 457–472.

DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ PARIS-NORD
and
LABORATOIRE DE PROBABILITÉ
UNIVERSITÉ PIERRE ET MARIE CURIE

Received October 10, 1979 (1577)