Michał Misiurewicz

A Short Proof of the Variational Principle for a $\mathbb{Z}_+^N$ Action on a Compact Space

Publications des séminaires de mathématiques et informatique de Rennes, 1975, fascicule S4
« International Conference on Dynamical Systems in Mathematical Physics », , p. 1-10

<http://www.numdam.org/item?id=PSMIR_1975___S4_A18_0>
A SHORT PROOF OF THE VARIATIONAL PRINCIPLE

FOR A $\mathbb{Z}^N_+$ ACTION ON A COMPACT SPACE

by

Michał Misiurewicz

0. Introduction. Ruelle in [12] introduced the notion of pressure for an action of the group $\mathbb{Z}^N_+$ on a compact metric space. It is a generalization of the notion of topological entropy. The variational principle (proved in [12] under some strong conditions) is a generalization of the Dinaburg's theorem ([5,8,7]) on a connection between the topological and measure entropies. A general proof of the variational principle was given by Walters in [13] (see also Denker [3]) for an action of $\mathbb{Z}_+$ and by Elsanousi in [6] for an action of $\mathbb{Z}^N_+$.

The first part of the proof given below (an inequality $h_{\mu}(T)+\mu f \leq \mathcal{P}(T,f)$) is a natural generalization of the proof from [10]. The second part ($\sup_{\mu} (h_{\mu}(T)+\mu f) \geq \mathcal{P}(T,f)$) is quite new (although the idea is close to the Ruelle's one).

1. Notations.

$\mathbb{Z}^N_+$ denotes the set of all non-negative integers. Let us fix a positive integer $N$.

$G = \mathbb{Z}^N_+$ is a commutative semigroup with respect to addition. For $n \in G$ we denote by $n_i$ the $i$-th coordinate
of all \( \mathbb{Z}^n \), \( n \in G \).

\( \mathcal{U} \) is the set of all neighborhoods of the diagonal in \( X \times X \), directed by the inclusion. It is a uniform structure (uniformity) for \( X \) (see [9]).

2. Definitions of pressure and entropy.

We may in the natural way extend Ruelle’s definition of pressure of \( \mathbb{Z}^n \)-action ([12]).

Let \( n \in G \), \( \sigma \in \mathcal{U} \), \( f \in C(X) \). We define successively:

\[
\Lambda(n) = \{ m \in G : m_i < n_i \text{ for } i = 1, \ldots, N \},
\]

\[
\lambda(n) = \text{Card} \Lambda(n) = n_1 \cdot \ldots \cdot n_N,
\]

\[
\sigma_n = \bigcap_{k \in \Lambda(n)} (T^k \times T^k)^{-1} \sigma,
\]

\[
f_n = \sum_{k \in \Lambda(n)} T^{\sigma_k} f
\]

Of course \( \sigma_n \in \mathcal{U} \), \( f_n \in C(X) \).

A finite subset \( e \) of \( X \) is called:

\( (n, \sigma) \)-separated, if for any \( x, y \in e \), \( x \neq y \), we have \( (x, y) \notin \sigma_n \),

\( (n, \sigma) \)-spanning, if for any \( x \in X \) there exists \( y \in e \) such that \( (x, y) \in \sigma_n \).

We denote \( p(f, e) = \log \sum_{x \in e} \exp f(x) \). We define further

(1) \( P_{n, \sigma}(T, f) = \sup \{ p(f_n, e) : e \text{ is } (n, \sigma) \text{-separated} \} \)

(2) \( P_\sigma(T, f) = \lim_{n \to \infty} \frac{1}{\lambda(n)} P_{n, \sigma}(T, f) \)

Of course, \( P_\sigma(T, f) \geq P_\varepsilon(T, f) \) for \( \sigma \subset \varepsilon \). Therefore it is possible to define the pressure

(3) \( P(T, f) = \lim_{\varepsilon \to 0} P_\varepsilon(T, f) = \sup_{\sigma \in \mathcal{U}} P_\sigma(T, f) \)

In the sequel we shall use the measure entropy function
of n (i = 1, ..., N). For n, m ∈ G let

\[ nm = (n_1 m_1, ..., n_N m_N) \] . The relation \( \preceq \) (n \( \preceq \) m iff \( n_i \preceq m_i \) for i = 1, ..., N) directs G.

\[ X \] is a non-empty compact Hausdorff space.

\[ C(X) \] is the space of all continuous real functions on \( X \) with the norm \( \|f\| = \sup_{x \in X} |f(x)| \).

\[ \mathcal{M}(X) \] is the space of all positive Borel regular normed measures on \( X \). It can be identified with the \( L^\infty \) space of all positive linear functionals on \( C(X) \) having norm 1 (therefore we shall write \( \mu f \) instead of \( \int_X f \, d\mu \) for \( f \in C(X) \)). We consider the weak-* topology on \( \mathcal{M}(X) \) (then \( \mathcal{M}(X) \) is compact).

\( T \) (n \( \mapsto \) \( T^n \)) is an action of \( G \) on \( X \), i.e. a homomorphism of \( G \) into the semigroup of all continuous transformations of \( X \) into itself (i.e. \( T^n : X \rightarrow X \), \( T^{n+m} = T^n T^m \)).

\( T^n : C(X) \rightarrow C(X) \) (for \( n \in G \)) is the operator induced by \( T^n \) (i.e. \( T^n f = f \circ T^n \)).

\( T^n \mathcal{M} : \mathcal{M}(X) \rightarrow \mathcal{M}(X) \) (for \( n \in G \)) is a restriction of the operator induced by \( T^n \) (i.e. \( T^n \mu = \mu \circ T^n \)) to \( \mathcal{M}(X) \). It is easy to check that indeed \( T^n \mathcal{M}(\mathcal{M}(X)) \subset \mathcal{M}(X) \) and that \( T^n \mathcal{M} \) is continuous.

Of course \( T^n \) and \( T^n \mathcal{M} \) are actions of \( G \) on \( C(X) \) and \( \mathcal{M}(X) \), respectively.

\( \mathcal{M}(X, T) \) is a space of all \( T \)-invariant measures (i.e. these elements of \( \mathcal{M}(X) \) which are fixed points of \( T^n \)).
For $\mu \in \mathcal{M}(X,T)$ the entropy $h_\mu(T)$ may be defined in the same way as in [2] for the action of $\mathbb{Z}^N$. For a Borel finite partition $A$ of the space $X$ we define

$$A^n = \bigvee_{k \in \Lambda(n)} (2^k)^{-1}A$$

for $n \in \mathbb{N}$.

$$h_\mu(T,A) = \lim_{n \to \infty} \frac{1}{\lambda(n)} H_\mu(A^n).$$

(it is easy to show that the limit exists, cf. [2]). Finally,

$$h_\mu(T) = \sup \{ h_\mu(T,A) : A \text{ - Borel finite partition} \}.$$

3. The variational principle.

We shall prove the following variational principle:

$$P(T,f) = \sup_{\mu \in \mathcal{M}(X,T)} (h_\mu(T) + \mu(f)).$$

The proof consists of two parts.

**Part I.** $h_\mu(T) + \mu(f) \leq P(T,f)$ for $\mu \in \mathcal{M}(X,T)$.

**Proof.** Let $\mu \in \mathcal{M}(X,T)$. Let us fix $\frac{\lambda}{2} > 0$ and a Borel finite partition $A$. Take $m \in \mathbb{N}$ such that

$$\log 2 \leq \frac{\lambda}{2} \cdot \lambda(m).$$

Let $A_m$ consist of the sets $a_1, \ldots, a_s$. For any of them there exists a compact set $b_i \subset a_i$ such that

$$\mu(a_i \setminus b_i) \leq \frac{\lambda}{s} \log s.$$

Let $b_0 = X \setminus \bigcup_{i=1}^s b_i$. For the partition $B = \{ b_0, b_1, \ldots, b_s \}$ we have

$$H_\mu(A_m \mid B) \leq \mu(b_0) \cdot \log s \leq \frac{\lambda}{2}.$$

We take $\varepsilon = (X \times X) \setminus \bigcup_{i,j=1}^s (b_i \times b_j) \in \mathcal{F}$ and next $\delta \in \mathcal{F}$ such that $\delta \circ \delta \subset \varepsilon$ (i.e. if $(x,y),(y,z) \in \delta$, then $(x,z) \notin \varepsilon$) and

$$|f(x) - f(y)| \leq \frac{\lambda}{2}$$

if $(x,y) \in \delta$.
Let us fix $n \in \mathbb{N}$. There exists a maximal $(\mu, \sigma)$-separated (i.e., being also $(\mu, \sigma)$-spanning) set $e \subseteq X$.

Denote by $C = \bigvee_{k \in \Lambda(n)} \{ (2^k)^{-1} \}_{B}$. Further, denote

$$\alpha(b) = \sup_{x \in b} f_{\mu}(x) \text{ for } b \in C,$$
$$\beta = \sum_{b \in C} \exp \alpha(b).$$

We have $\int_{b} f_{\mu} \, d\mu \leq \alpha(b) \cdot \mu(b)$, therefore

$$H(C) + \int_{b} \frac{\mu(b)}{\beta} \leq \sum_{b \in C} \frac{\mu(b)}{\beta} (\alpha(b) - \log \mu(b)) =$$
$$= \beta \sum_{b \in C} \exp \alpha(b) \cdot \eta \left( \frac{\mu(b)}{\exp \alpha(b)} \right),$$
where $\eta(x) = -x \log x$.

The function $\eta$ is concave, therefore

$$H(C) + \int_{b} \frac{\mu(b)}{\beta} \leq \log \beta$$

$e$ is a $(\mu, \sigma)$-spanning set, hence for every $b \in C$ there exists a point $z(b) \in e$ such that

$$\alpha(b) = \sup \{ f_{\mu}(x) : x \in b, (x, z(b)) \in \sigma_{nm} \}.$$

But if $(x, z(b)) \in \sigma_{nm}$, then for $k \in \Lambda(n)$ $(T^k x, T^k z(b)) \in \sigma$,

thus, in view of the definition of $\sigma$,

$$|(T^{2^k} f)(x) - (T^{2^k} f)(z(b))| \leq \frac{\lambda}{2}.$$  

Hence

$$f_{\mu}(z(b)) \geq \alpha(b) - \frac{\lambda}{2} \cdot \lambda(n).$$

From the definitions of $\sigma$ and $\varepsilon$ we obtain for $y \in e$, $k \in \Lambda(n)$

$$\text{Card} \left\{ a \in B : \exists x \in a \ (T^{2^k} x, T^{2^k} y) \in \sigma \right\} \leq 2,$$

thus for $y \in e$

$$\text{Card} \left\{ b \in C : \exists x \in b \ (x, y) \in \sigma_{nm} \right\} \leq 2 \lambda(n),$$

because $\sigma_{nm} \subseteq \bigcap_{k \in \Lambda(n)} (T^{2^k} x, T^{2^k} y) \in \sigma$. Hence

$$\text{Card} \left\{ b \in C : z(b) = y \right\} \leq 2 \lambda(n).$$

Hence, from (7) and (8) we get:
\[
\lambda(n) \cdot \log 2 + p(f_{nm}, e) \geq \log \beta - \frac{\lambda(n)}{2} \cdot \lambda(n)
\]

But \( \mu f_{nm} = \lambda(nm) \cdot \mu f \), so from this, from (1), (4), (6) and (9) we obtain (notice that \( \lambda(nm) = \lambda(n) \cdot \lambda(m) \)):

\[
\frac{1}{\lambda(nm)} H_{\mu}(C) + \mu f \leq \frac{1}{\lambda(nm)} P_{nm, \delta}(T, f) + 2 \beta
\]

In view of (5), for \( k \in \Lambda(n) \) we have

\[
H_{\mu}((T^{km})^{-1} A_m | (T^{km})^{-1} B) \leq \beta , \text{ therefore}
\]

\[
H_{\mu}(A_{nm} | C) = H_{\mu}(\bigvee_{k \in \Lambda(n)} (T^{km})^{-1} A_m | \bigvee_{k \in \Lambda(n)} (T^{km})^{-1} B) \leq \beta \cdot \lambda(n)
\]

Hence \( H_{\mu}(A_{nm}) \leq H_{\mu}(C) + H_{\mu}(A_{nm} | C) \leq H_{\mu}(C) + \beta \cdot \lambda(n) \),

thus we obtain from (10):

\[
\frac{1}{\lambda(nm)} H_{\mu}(A_{nm}) + \mu f \leq \frac{1}{\lambda(nm)} P_{nm, \delta}(T, f) + 3 \beta . \text{ Taking } \lim \sup \text{ with respect to } n \text{ we get}
\]

\[
h_{\mu}(T, A) + \mu f \leq P_{T, \delta}(T, f) + 3 \beta \leq P(T, f) + 3 \beta .
\]

But \( \beta \) and \( A \) were arbitrary, hence \( h_{\mu}(T) + \mu f \leq P(T, f) \).

\[\Box\]

**Part II.** \[\sup_{\mu \in \mathcal{M}(X, T)} (h_{\mu}(T) + \mu f) \geq P(T, f) \]

**Proof.** Let us fix \( \delta \in \mathcal{H} \). For every \( n \in G \), we choose such an \((n, \delta)\)-separated set \( e_n \) that

\[
p(f_n, e_n) \geq P_{n, \delta}(T, f) - 1
\]

Let us define a measure \( \sigma_n \), concentrated on \( e_n \), by a formula \( \sigma_n(y) = \exp (f_n(y) - p(f_n, e_n)) \) for \( y \in e_n \).

We have \( \sum_{y \in e_n} \sigma_n(y) = 1 \), therefore \( \sigma_n \in \mathcal{M}(X) \).

Let \( \mu_n = \frac{1}{\lambda(n)} \sum_{k \in \Lambda(n)} T^{k+n} \sigma_n \).

For some sequence \( (n_i)_{i=1}^{\infty} \), cofinal with \( G \), we have
We choose some cluster point of the sequence \( (\mu_n)_{n=1}^{\infty} \) and denote it by \( \mu \). Of course, \( \mu \) is also a cluster point of the net \( (\mu_n)_{n \in G} \). For \( g \in C(X) \) and \( k \in G \) fixed, the function \( \overline{\phi} : M(X) \rightarrow \mathbb{R} \), given by the formula \( \overline{\phi}(\nu) = \nu g - \nu(T^k g) \), is continuous, therefore \( \overline{\phi} \mu \) is a cluster point of the net \( (\overline{\phi}\mu_n)_{n \in G} \). We have

\[
|\overline{\phi}\mu_n| \leq \frac{1}{\lambda(n)} \cdot 2 \cdot (\lambda(n) - \lambda(n-k)) \cdot \|g\| \quad \text{for} \quad n > k
\]

because \( \text{Card} \left( \Lambda(n) \setminus (k + \Lambda(n)) \right) = \text{Card} \left( (k + \Lambda(n)) \setminus \Lambda(n) \right) = \lambda(n) - \lambda(n-k) \). But \( \lim_{n \to \infty} \frac{\lambda(n-k)}{\lambda(n)} = 1 \), thus \( \overline{\phi}\mu = 0 \).

Hence \( \nu g = (T^k \nu) \in C(X) \). But \( g \) and \( k \) were arbitrary, therefore \( \mu \in M(X,T) \).

There exists a Borel finite partition \( \Lambda \) of \( X \) such that \( a \times a < \delta \) for \( a \in \Lambda \). Then for \( a \in A^n \), \( a \times a < \delta_n \), therefore \( \text{Card} \left( e_n \cap a \right) < 1 \). Hence

\[
H_{\sigma_n}(\Lambda^n) + \sigma_n f_n = \sum_{y \in e_n} \sigma_n(\{y\}) (f_n(y) - \log \sigma_n(\{y\})) = p(f_n, e_n).
\]

Let us fix \( m, n \in G \), \( n > 2m \). For given \( j \in \Lambda(m) \), let \( s(j) = (E(\frac{n_1-j}{m_1}), \ldots, E(\frac{n_N-j}{m_N})) \) . We have:

\[
\Lambda^n = \bigvee_{r \in \Lambda(s(j))} (T^{r_m+j})^{-1} \Lambda^m \vee \bigvee_{k \in E} (T^k)^{-1} \Lambda^m ,
\]

where

\[
E = \Lambda(n) \setminus (j + \Lambda(ms(j))) . \quad \text{But} \quad \text{Card} \ E = \lambda(n) - \lambda(ms(j)) \leq \lambda(n) - \lambda(n-2m) , \quad \text{thus} \quad p(f_n, e_n) = H_{\sigma_n}(f^n) + \sigma_n f_n \leq \sum_{r \in \Lambda(s(j))} H_{\sigma_n}(T^{r_m+j})^{-1} \Lambda^m + \sigma_n f_n + (\lambda(n) - \lambda(n-2m)) \text{log Card} \ A.
\]
Summing the inequalities obtained for $j \in \Lambda(m)$ we get (notice that for $k \in \Lambda(n)$ there exists a unique $j(m)$ and a unique $r \in \Lambda(s(j))$ such that $k = rm + j$):

\[ \sum_{k \in \Lambda(n)} H \sigma_n^k ((T^k)^{-1} \lambda^m) + \lambda(m) \cdot \sigma_n^f_n \geq \lambda(m) \cdot (f_n, e_n) - (\lambda(n) - \lambda(n-2m)) \cdot \log \text{Card } A \]

We have also

\[ \sigma_n^f_n = \sigma_n \left( \sum_{k \in \Lambda(n)} T^k f \right) = \left( \sum_{k \in \Lambda(n)} T^k \sigma_n^k \right) f_n = \lambda(n) \cdot \mu_n f \]

From the definition of entropy it follows that

\[ H \sigma_n^k ((T^k)^{-1} \lambda^m) = H_{T^k \sigma_n^k} \sigma_n^k (\lambda^m) \]

From the definition of entropy and from the concavity of the function $-x \log x$ it follows that

\[ H \mu_n^m (\lambda^m) \geq \frac{1}{\lambda(n)} \sum_{k \in \Lambda(n)} H_{T^k \sigma_n^k} \sigma_n^k (\lambda^m) \]

The formulas (11) and (13) - (16) give us

\[ \frac{1}{\lambda(m)} H \mu_n^m (\lambda^m) + \mu_n f \geq \frac{1}{\lambda(n)} (p_n, f) - \frac{1}{\lambda(n)} \cdot ((\lambda(n) - \lambda(n-2m)) \log \text{Card } A + 1) \]

The partition $A$ can be chosen in such a way that the boundaries of elements of $A$ have measure $\mu$ zero.

(see [1], Chapt. IV, §5, exerc. 13 d; see also [10]).

Then $A^n$ has the same property. But for a set $a$ with the boundary of measure $\mu$ zero, the function $\mathcal{M}(X) \rightarrow \mathbb{R}$, given by $\nu \mapsto \nu(a)$, is continuous in the point $\mu$.
Hence the function \( \mathcal{W}(X) \to \mathbb{R} \), given by \( y \mapsto H_y(A^m) \)
is also continuous at the point \( \mu \), therefore in view of (12), (17) and the definition of \( \mu \), we have

\[
\frac{1}{\lambda(m)} H_\mu(A^m) + \mu f \geq P(\delta(T, X))
\]

Taking the limit with respect to \( m \) and using the inequality \( h_\mu(T, A) \leq h_\mu(T) \), we obtain \( h_\mu(T) + \mu f \geq P(\delta(T, X)) \)

But \( \delta \) was arbitrary, hence \( \sup_{\mu \in \mathcal{M}(X, T)} (h_\mu(T) + \mu f) \geq P(T, f) \).

4. Remark.

If \( h_\mu(T) + \mu f = P(T, f) \), then \( \mu \) is called an equilibrium state for \( (T, f) \) (measure with maximal entropy in the case of \( f = 0 \)). The above construction shows that if \( P_\delta(T, f) = P(T, f) \) for some \( \delta \in \mathcal{H}^\mu \), then there exists an equilibrium state for \( (T, f) \). In the case of \( N = 1 \), \( f = 0 \), this can be reformulated as follows:

If there exists an open cover \( A \) such that \( h(T, A) = h(T) \), then there exists a measure with maximal entropy.

This is a particular case of the theorem of Denker ([4]), but obtained without assuming \( X \) finite dimensional.

References.


M. Misiurewicz
Mathematical Institute, Univ. of Warsaw
PKiN IX p.
00-901 Warszawa
Poland