ISRAEL JOURNAL OF MATHEMATICS, Vol. 21, Nos. 2–3, 1975

UNILATERAL CODINGS OF BERNOULLI SYSTEMS

BY
DONALD S. ORNSTEIN AND BENJAMIN WEISS

ABSTRACT
A new proof is presented of a theorem of Sinai that says that, if ϕ is an *endomorphism* and π a probability distribution with $H(\pi) \leqq h(\phi)$, then ϕ has an independent partition with distribution π.

If one interprets $\mathbf{Z}$ as a time parameter in $\{\phi^i\}_{i \in \mathbf{Z}}$, where ϕ is an invertible measure preserving transformation, then the usual isomorphism theory deals with non physical codings in the sense that their implementation requires looking infinitely far ahead into the future. For this reason there has been some interest in unilateral isomorphism theory. We propose to give here a new proof of Sinai's theorem [4] in the hope that it will help to answer the questions about unilateral weak isomorphisms between VWB systems. We assume familiarity with standard results as presented in [2].

1. Let ϕ be an ergodic measure preserving transformation of $(X, \mathcal{B}, \mu)$ with positive entropy, and suppose that $\alpha = \{A_1, \cdots, A_a\}$ is a generator with $h(\phi, \alpha) > 0$. Let $\pi = (\pi(1), \pi(2), \cdots, \pi(p))$ be a probability vector with $H(\pi) = -\Sigma_1^p \pi(i) \log \pi(i) \leqq h(\phi, \alpha)$. The theorem we are after may be phrased this way:

THEOREM 1. *If $H(\pi) \leqq h(\phi, \alpha)$, then there is a partition β, measurable with respect to $\bigvee_0^\infty \phi^i \alpha$, such that:*

(1) $$\operatorname{dist} \beta = \pi$$

(2) $$\{\phi^i \beta\}_{i \in \mathbf{Z}} \text{ are independent.}$$

Received November 21, 1974

We use the notation dist β for the probability vector $(\mu(B_1), \mu(B_2), \cdots, \mu(B_p))$. The first lemma that we state is well known and follows from the basic properties of the conditional entropy. Its function here will be to allow us to work at all times with finite partitions and to avoid using the machinery of measurable partitions. The phrase "*$\mathcal{P}$ holds for ε-*a.e. *atom C of a partition γ*" will mean:

$$(3) \qquad \sum_{\{C\in\gamma:\, \mathcal{P} \text{ holds for } C\}} \mu(C) > 1-\varepsilon .$$

LEMMA 2. *Let γ be a finite partition. Then given $\varepsilon > 0$ there is an $n_0 = n_0(\varepsilon, \gamma)$ such that for all $m > n \geqq n_0$ we have that $\gamma | C$ is ε-independent of $(\vee_{n+1}^{m} \phi^i \gamma)|C$ for ε-a.e. atom, C, of $\vee_1^n \phi^i \gamma$.*

Next we shall introduce a certain measure for the deviation from independence that is suited to the kind of iterative construction that will be the main part of the proof. Distances between probability vectors will be measured by $\|\pi - \pi'\| = \Sigma_i |\pi(i) - \pi'(i)|$. For fixed π, $d(\pi, \beta | \eta)$ will measure how close the distribution of β is to π on most atoms of η, and will be defined by:

$$(4) \qquad d(\pi, \beta \mid \eta) = \sum_{E\in\eta} \| \pi - \text{dist}(\beta \mid E) \| \cdot \mu(E) .$$

Clearly $d(\pi, \beta | \eta) = 0$ means that β is independent of η *and* dist $\beta = \pi$. We shall use the notation $\eta' \subset \eta$ if the algebra of η' is contained in the algebra of η, i.e., η is a refinement of η'. The next lemma is readily proved from the definitions.

LEMMA 3. *If $\eta' \subset \eta$ then $d(\pi, \beta | \eta') \leqq d(\pi, \beta | \eta)$.*

Using this we can now define:

$$(5) \qquad \text{Di}(\pi, \beta) = \lim_{n\to\infty} d\left(\pi, \beta \,\middle|\, \bigvee_1^n \phi^i \beta\right) = \sup_{n\geqq 1} d\left(\pi, \beta \,\middle|\, \bigvee_1^n \phi^i \beta\right).$$

The equality of the limit and supremum follows from Lemma 3. Naturally $\text{Di}(\pi, \beta) = 0$ if and only if β is an independent partition with distribution π. We shall also need a result about the continuity of Di if β and β' are ordered partitions of the same space; we will let $d(\beta, \beta')$ denote the measure of the set of points labeled differently by β and β'.

LEMMA 4. *Given β, π and $\varepsilon > 0$, there exists a $\delta > 0$ such that $d(\beta, \beta') < \delta$ implies*

$$\text{Di}(\pi, \beta') \leqq \text{Di}(\pi, \beta) + \varepsilon. \tag{6}$$

PROOF. In lieu of a detailed computation we present the idea of the proof. The n_0 in Lemma 1 depends upon the rate with which $H(\gamma \mid \vee_1^n \phi^i \gamma)$ approaches $h(\phi, \gamma)$. Now if β' is sufficiently close to β, then the n_0 for β' can be taken just as the n_0 for β, and this implies that one can estimate $\text{Di}(\pi, \beta')$ by $d(\pi, \beta \mid \vee_1^{n_0} \phi^i \beta')$, which easily implies the lemma. ■

We shall need a mild sharpening of a theorem that can be found in [1]. The use of it seems to be essential to the argument here.

THEOREM 5. *Let ϕ be an invertible ergodic measure preserving transformation. Let $\beta \subset \vee_{-\infty}^{\infty} \phi^i \alpha$, and $H(\pi) \leqq h(\phi, \alpha) - h(\phi, \beta)$. Then there exists a partition $\gamma \subset \vee_{-\infty}^{\infty} \phi^i \alpha$ such that:*

(i) dist $\pi = \gamma$

(ii) $\{\phi^i \gamma\}_{i \in \mathbb{Z}}$ *are independent*

(iii) $\vee_{-\infty}^{\infty} \phi^i \beta$ *is independent of* $\vee_{-\infty}^{\infty} \phi^i \gamma$.

PROOF. Let $\tilde{\beta}$ be an independent partition such that $\tilde{\beta} \subset \vee_{-\infty}^{\infty} \phi^i \beta$ and $h(\phi, \tilde{\beta}) = h(\phi, \beta)$. Lemma 5′ in [1] implies† that there exists γ satisfying (i), (ii) and (iii) with β replaced by $\tilde{\beta}$. But then (iii) follows by Lemma 2 of [3]. ■

We can now describe in outline the proof of Theorem 1. Denoting by $\mathscr{A} = \vee_0^{\infty} \phi^i \alpha$, we are trying to construct $\beta \subset \mathscr{A}$ such that $\text{Di}(\pi, \beta) = 0$. Suppose then that we have a first approximation $\beta_1 \subset \mathscr{A}$ with $\text{Di}(\pi, \beta_1)$ small. One can obtain such approximations by taking an independent $\tilde{\beta}_1$, approximating $\tilde{\beta}_1$ by a $\beta_1' \subset \vee_{-N}^{\infty} \phi^i \alpha$ and setting $\beta_1 = \phi^N \beta_1'$. Now we must improve β_1, i.e., find a $\beta_2 \subset \mathscr{A}$ such that $\text{Di}(\pi, \beta_2) < \text{Di}(\pi, \beta_1)$, and for the procedure to converge we must do it in such a way that β_2 remains close to β_1. Look at how β_1 divides an atom B of $\vee_1^N \phi^i \beta_1$. Since $\text{Di}(\pi, \beta_1) > 0$ (otherwise we would be done), there are some atoms of $\beta_1 | B$ that are too small compared to π and others too large. We want to change β_1 by moving a little mass from the excessive atoms of $\beta_1 | B$ to the deficient ones.

In doing so we improve the distribution of β_1 on atoms of $\vee_1^N \phi^i \beta_1$—but of course for the new partition β_2, so obtained, we compute $\text{Di}(\pi, \beta_2)$ by looking at β_2 on atoms of $\vee_1^N \phi^i \beta_2$. For this reason we must exercise care in moving the

† The assumption made there that ϕ be mixing is not really needed, as a careful examination of the proof shows.

mass. Now the basic idea is this. Assume that $h(\phi,\beta_1) < h(\phi,\alpha)$, and use Theorem 5 to get an independent $\gamma = \{C_1, X \sim C_1\}$, with $\bigvee_{-\infty}^{\infty}\phi^i\gamma$ independent of $\bigvee_{-\infty}^{\infty}\phi^i\beta_1$. If $\mu(C_1)$ is small, then we modify β_1 to β_2 by moving C_1 intersected with an excessive atom of $\beta_1 \cap B$ ($B \in \bigvee_1^N \phi^i\beta_1$) over to a deficient atom of $\beta_1 \cap B$. The complete independence of γ enables us now to obtain a reduction $\mathrm{Di}(\pi,\beta_1) - \mathrm{Di}(\pi,\beta_2)$ that is roughly proportional to the change $d(\beta_1,\beta_2)$. To be sure, the β_2 obtained in this way is not necessarily contained in $\mathcal{A}$, but for that replacing γ by $\phi^M\gamma$ changes nothing and, for sufficiently large μ, $\phi^M\gamma$ can be approximated arbitrarily well by a partition in $\mathcal{A}$, which by Lemma 4 says that we lose only a small part of what has already been gained.

The main technical complication in carrying this program out is to make sure that, in the process of constructing β_2, we do not suddenly go up in entropy so that $h(\phi,\beta_2) = h(\phi,\alpha)$, which would prevent us from iterating the step just described. To get around this difficulty we take a sequence of probability distributions $\pi_1, \pi_2, \cdots$, with $H(\pi_1) < H(\pi_2) < \cdots H(\pi)$ and $\lim \pi_i = \pi$, and at the ith stage set our sights on π_{i+1} rather than on π. In the last preliminary lemma we write down the properties needed for the sequence $\{\pi_i\}$; the existence of such a sequence follows from the smoothness of $H(\cdot)$.

LEMMA 6. *Given a probability vector* $\pi = (\pi(1), \cdots, \pi(p))$ *with* $H(\pi) > 0$, *there exists a sequence* $\{\pi_i\}$ *with*

(i) $H(\pi_1) < H(\pi_2) < \cdots < H(\pi)$

(ii) $\lim_{i\to\infty} \pi_i = \pi$

(iii) *If* π' *satisfies* $\|\pi' - \pi_{i+1}\| \leqq 2\|\pi_i - \pi_{i+1}\|$, *then* $H(\pi') < H(\pi)$ (*and by compactures* $H(\pi') < H(\pi) - a_i$, $a_i > 0$)

(iv) $\sum_1^\infty \|\pi_i - \pi_{i+1}\| < \infty$.

2. The proof of Theorem 1 will proceed via a number of reductions. Suppose then that $(X, \mathcal{B}, \mu, \phi)$ is fixed once and for all, with α a generator, $\mathcal{A} = \bigvee_0^\infty \phi^i\alpha$ and π a probability vector with $H(\pi) \leqq h(\phi,\alpha)$. Let $\{\pi_i\}$ be a sequence of probability vectors satisfying the conclusions of Lemma 6.

LEMMA 7. *There is an* $\varepsilon > 0$, *such that if* $\mathrm{Di}(\pi_1,\beta_1) < \varepsilon$, $\beta_1 \subset \mathcal{A}$, *then for any* $\delta > 0$ *there is a* $\beta_2 \subset \mathcal{A}$ *satisfying*

(i) $\mathrm{Di}(\pi_2,\beta_2) < \delta$

(ii) $d(\beta_1,\beta) \leqq \|\pi_1 - \pi_2\|$.

We shall first deduce Theorem 1 from Lemma 7 and then continue with the reduction.

PROOF OF THEOREM 1. Applying Lemma 7 successively, after using the argument in Section 1 to get started, we obtain a sequence of partitions $\{\beta_i\} \subset \mathcal{A}$ that satisfy

$$\mathrm{Di}(\pi_i, \beta_i) < \varepsilon_i \tag{1}$$

$$d(\beta_i, \beta_{i+1}) \leqq \|\pi_i - \pi_{i+1}\|, \tag{2}$$

with $\varepsilon_i \to 0$. From (2) and Lemma 6 we have that the β_i converge to a partition $\beta \subset \mathcal{A}$, while from (1) it follows via Lemma 4 that $\mathrm{Di}(\pi, \beta) = 0$, which proves the theorem. ■

The next lemma formulates the idea of having the improvement in β proportional to the change in a way suited to proving Lemma 7. This and a simple iteration will prove Lemma 7.

LEMMA 8. *Let δ_2, $c > 0$ be constants such that $\|\pi' - \pi_2\| \leqq \delta_2$ implies $H(\pi') \leqq h(\phi, \alpha) - c$. Then for any $\delta_1 > 0$ there exists a positive δ which depends only on c, δ_1, π_2 such that, if $\beta_1 \subset \mathcal{A}$ satisfies*

$$\delta_1 \leqq \mathrm{Di}(\pi_2, \beta_1) < \delta_2,$$

there exists a $\bar{\beta}_1 \subset \mathcal{A}$ satisfying

(i) $\mathrm{Di}(\pi_2, \bar{\beta}_1) \leqq \mathrm{Di}(\pi_2, \beta_1) - \delta$

(ii) $d(\beta_1, \bar{\beta}_1) \leqq 4(\mathrm{Di}(\pi_2, \beta_1) - \mathrm{Di}(\pi_2, \bar{\beta}_1))$.

PROOF OF LEMMA 7. We remark first that

$$\mathrm{Di}(\pi_2, \beta_1) \leqq \mathrm{Di}(\pi_1, \beta_1) + \|\pi_2 - \pi_1\|. \tag{4}$$

Thus, if ε is taken as $\|\pi_2 - \pi_1\|$, then by Lemma 6 we can put $\delta_2 = 2\|\pi_2 - \pi_1\|$, and then the hypotheses of Lemma 8 are satisfied for any $\delta_1 < \mathrm{Di}(\pi_2, \beta_1)$. Now given $\varepsilon_2 > 0$ we choose $\delta_1 = \varepsilon_2$. If $\mathrm{Di}(\pi_2, \beta_1) \leqq \delta_1$, then we can take $\beta_2 = \beta_1$. Otherwise, we can successively apply Lemma 8 until $\mathrm{Di}(\pi_2, \bar{\beta}_1)$ drops below δ_1. Using the triangle inequality we obtain finally a $\beta_2 \subset \mathcal{A}$ with $\mathrm{Di}(\pi_2, \beta_2) \leqq \delta_1 = \varepsilon_2$ and

$$d(\beta_1, \beta_2) \leqq 2\mathrm{Di}(\pi_2, \beta_1) \leqq 2\|\pi_2 - \pi_1\| \tag{5}$$

by (4). This proves Lemma 7. ■

We make one further reduction before implementing the main idea of the proof. To simplify the notation we now drop the subscripts on π_i and β_i, even though it is to them that the lemmas will apply.

LEMMA 9. *Let π be a fixed probability vector, $\pi = (\pi(1), \cdots, \pi(k))$, $\beta \subset \mathcal{A}$ a partition with k atoms satisfying $h(\phi,\beta) \leqq h(\phi,\alpha) - c$ for $c > 0$. Let $\delta > 0$ satisfy $-\delta \log \delta - (1-\delta)\log(1-\delta) < c$, and assume further that for all large m, on a set of atoms B of $\bigvee_1^m \phi^i\beta$ of total measure at least δ, we have, for some i, j:*

$$\mu(B_i \mid B) - \pi(i) > \delta, \mu(B_j \mid B) - \pi(j) < -\delta,$$

then there exists a $\tilde{\beta} \subset \mathcal{A}$ satisfying

(i) $\mathrm{Di}(\pi, \tilde{\beta}) \leqq \mathrm{Di}(\pi, \beta) - \delta^2$

(ii) $d(\beta, \tilde{\beta}) \leqq 4 \cdot (\mathrm{Di}(\pi,\beta) - \mathrm{Di}(\pi, \tilde{\beta}))$.

As usual, we first use Lemma 9 to prove Lemma 8 and then finally we shall prove Lemma 9.

PROOF OF LEMMA 8. Comparing the two lemmas, we note first that, since $h(\phi,\beta_1) \leqq H(\mathrm{dist}\ \beta_1)$, it follows from the hypothesis of Lemma 8 that $h(\phi,\beta_1) \leqq h(\phi,\alpha) - c$. We have to show that $\mathrm{Di}(\pi_2, \beta_1) \geqq \delta_1$ implies that the main assumptions of Lemma 9 hold. By the definitions, for all large m

$$d\Big(\pi_2, \beta_1 \,\Big|\, \bigvee_1^m \phi^i \beta_1\Big) \geqq \delta_1/2. \tag{6}$$

Since $\| \pi_2 - \mathrm{dist}(\beta_1 | B)\|$ is at most one, it follows from (6) that for a set of atoms $B \in \bigvee_1^m \phi^i\beta_1$, of total measure at least $\delta_1/8$,

$$\|\pi_2 - \mathrm{dist}(\beta_1|B)\| \geqq \delta_1/8. \tag{7}$$

Now since there are only k atoms in β_1, it follows that for each B on which (7) holds there is an i, j such that

$$\begin{aligned} &\mu(B_i^{(1)}|B) - \pi_2(i) \leqq \frac{1}{2k}\cdot\frac{1}{8}\delta_1 \\ &\mu(B_j^{(1)}|B) - \pi_2(j) \leqq -\frac{1}{2k}\cdot\frac{1}{8}\delta_1. \end{aligned} \tag{8}$$

Now we are in a position to apply Lemma 9 with any $\bar{\delta} < 1/2k \cdot \delta_1/8$ that satisfies $-\bar{\delta}\log\bar{\delta} - (1-\bar{\delta})\log(1-\bar{\delta}) < c$. Then Lemma 8 follows immediately with $\delta = \bar{\delta}^2$. ■

PROOF OF LEMMA 9. Apply Theorem 5 to obtain a partition $\gamma = \{C, X \sim C\}$ with $\mu(C) = \delta$, $\{\phi^i\gamma\}_{-\infty}^{\infty}$ independent and $\bigvee_{-\infty}^{\infty}\phi^i\gamma$ independent of $\bigvee_{-\infty}^{\infty}\phi^i\beta$. Then use Lemma 2 with $\gamma = \beta$ and $\varepsilon = \delta/10$ to find n_0 such that for any $m > n > n_0$ we have for $\delta/10$-a.e. atom B of $\vee_1^n \phi^i\beta$ that β/B is $\delta/10$-independent of $\vee_{n+1}^m \phi^i\beta$.

We first define $\bar\beta$ to improve the independence properties of β, later we shall worry about getting it into $\mathcal{A}$. Let $n > n_0$ be large enough so that: (i) $|\mathrm{Di}(\pi,\beta) - d(\pi,\beta \mid \vee_1^n \phi^i\beta)| < 1/100 \cdot \delta^2$, and (ii) on a set of atoms B of $\vee_1^n \phi^i\beta$ of total measure at least δ we have for some i,j

$$\mu(B_i \mid B) - \pi(i) > \delta, \quad \mu(B_j \mid B) - \pi(j) < -\delta. \tag{9}$$

On each atom B that satisfies (9) we modify β to $\bar\beta$ by defining:

$$\bar B_i \cap B = B_i \cap B \sim B_i \cap B \cap C; \quad \bar B_j \cap B = (B_j \cap B) \cup (B_i \cap B \cap C), \tag{10}$$

while all other atoms on B are left unchanged. Clearly we have that:

$$d\Big(\pi, \bar\beta \,\Big|\, \bigvee_1^n \phi^i\beta\Big) \leqq d\Big(\pi, \beta \,\Big|\, \bigvee_1^n \phi^i\beta\Big) - 2\delta^2, \tag{11}$$

while, since all the change is in the direction of improving the independence, we have

$$d(\beta,\bar\beta) \leqq d(\pi,\beta) \,\Big|\, \bigvee_1^n \phi^i\beta\Big) - d\Big(\pi,\bar\beta \,\Big|\, \bigvee_1^n \phi^i\beta\Big) \tag{12}$$

By the choice of n_0 and Lemma 2 we deduce from (11) that for all $m > n$

$$d\Big(\pi,\bar\beta \,\Big|\, \bigvee_1^m \phi^i\beta\Big) \leqq \mathrm{Di}(\pi,\beta) - 3/2\delta^2. \tag{13}$$

By the independence assumptions on γ we get from (13) immediately that for all $m > n$

$$d\Big(\pi,\bar\beta \,\Big|\, \bigvee_1^m \phi^i\beta \vee \bigvee_1^m \phi^i\gamma\Big) \leqq \mathrm{Di}(\pi,\beta) - 3/2\delta^2. \tag{14}$$

From the definition of $\bar\beta$ we have that $\vee_1^m \phi^i\bar\beta$ is coarser than $\vee_1^{m+n_0+1} \phi^i\beta \vee \vee_1^m \phi^i\gamma$, and thus (14) with Lemma 3 implies

$$\mathrm{Di}(\pi,\bar\beta) \leqq \mathrm{Di}(\pi,\beta) - 3/2\delta^2. \tag{15}$$

Now we push $\bar\beta$ into $\mathcal{A}$ as follows. Nothing changes if we replace γ by $\phi^N\gamma$, and since α was a generator, N can be chosen so that $\phi^N\gamma$ can be approximated arbitrarily well by $\tilde{\gamma} \subset \mathscr{A}$. Applying now Lemma 4 with $\varepsilon = 1/2\delta^2$, and we get a $\tilde{\beta} \subset \mathscr{A}$ that satisfies (i) and (ii), and Lemma (9) is thereby proved. ■

This completes the proof of Theorem 1, the logical order being Lemma 9 ⇒ Lemma 8 ⇒ Lemma 7 ⇒ Theorem 1. The argument has been presented in a different order in the hope of greater clarity. The proof given here can be pushed a bit further to prove a similar result for Markov chains, all of whose transition probabilities are positive.

References

1. D. Ornstein, *Two Bernoulli shifts with infinite entropy are isomorphic*, Advances in Math. **5** (1970), 339–348.

2. D. Ornstein, *Ergodic Theory, Randomness and Dynamical Systems*, Yale Univ. Press, 1974.

3. D. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.

4. Ya. G. Sinai, *Weak isomorphism of transformations with invariant measure*, Mat. Sb. **63** (1964), 23–42; A.M.S. Translations **57** (1966), 123–143.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA, U.S.A.

THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL