A STRICTLY ERGODIC, POSITIVE ENTROPY SUBSHIFT
UNIFORMLY UNCORRELATED TO THE MÔBIUS FUNCTION

TOMASZ DOWNAROWICZ AND JACEK SERAFIN

Abstract. A recent result [DS] shows that there exist positive entropy sub-
shifts satisfying the assertion of Sarnak’s conjecture [S]. More precisely, it is
proved that if \( y = (y_n)_{n \geq 1} \) is a bounded sequence with zero average along ev-
ery infinite arithmetic progression (the Möbius function is an example of such
a sequence) then for every \( N \geq 2 \) there exists a subshift \( \Sigma \) over \( N \) symbols,
with entropy arbitrarily close to \( \log N \), uncorrelated to \( y \).

In the present note, we improve the result of [DS]. First of all, we observe
that the uncorrelation obtained in [DS] is uniform, i.e., for any continuous
function \( f : \Sigma \to \mathbb{R} \) and every \( \epsilon > 0 \) there exists \( n_0 \) such that for any \( n \geq n_0 \)
and any \( x \in \Sigma \) we have

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(T^i x) y_i \right| < \epsilon.
\]

More importantly, by a fine-tuned modification of the construction from [DS]
we create a strictly ergodic subshift, with all the desired properties of the
example in [DS] (uniformly uncorrelated to \( y \) and with entropy arbitrarily
close to \( \log N \)).

The question about these two additional properties (uniformity of uncor-
relation and strict ergodicity) has been posed by Mariusz Lemańczyk in the
context of the so-called strong MOMO (Möbius Orthogonality on Moving Or-
bits) property. Our result shows, among other things, that strong MOMO
is essentially stronger than uniform uncorrelation, even for strictly ergodic
systems.

1. Preliminaries

Let \( y \) be a bounded, real-valued sequence with zero average along every infinite
arithmetic progression, i.e., satisfying, for every \( t \geq 1 \) and \( l \geq 0 \), the condition

\[
\lim_{n} \frac{1}{n} \sum_{i=1}^{n} y_{it+l} = 0.
\]
A sequence as above we call aperiodic. Clearly we may (and will) assume that $|y_n| \leq 1$ for all $n$. An important example of an aperiodic sequence is:

\[
\mu_n = \begin{cases} 
1 & \text{for } n = 1, \\
(-1)^r & \text{if } n \text{ is a product of } r \text{ distinct primes}, \\
0 & \text{otherwise (i.e., if } n \text{ has a repeated prime factor),}
\end{cases}
\]
called the Möbius function $y = \mu$ (see e.g. [S]).

By a topological dynamical system we will mean a pair $\left( X, T \right)$ where $X$ is a compact metric space and $T : X \to X$ is a continuous transformation. Uncorrelation between a system and a sequence will be understood as follows:

**Definition 1.1.** We say that $\left( X, T \right)$ is uncorrelated to $y$ if for each continuous function $f : X \to \mathbb{R}$ and every $x \in X$, we have

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i x) y_i = 0.
\]

If, for each fixed function $f$, the above convergence to 0 is uniform on $X$ then we will say that $\left( X, T \right)$ is uniformly uncorrelated to $y$.

Let $\Lambda$ be a fixed finite alphabet. By a subshift $\Sigma$ we will mean any closed and shift-invariant subset of $\Lambda^\mathbb{N}$. A subshift becomes a dynamical system when regarded together with the action of the shift map. We denote by $\mathcal{M}(\Sigma)$ the collection of all shift-invariant Borel probability measures supported by $\Sigma$; this set is non-empty and compact in the weak-star topology with the distance between measures given by:

\[
d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \sum_{D \in \Lambda^n} |\mu(D) - \nu(D)|,
\]

where $\Lambda^n$ denotes the family of blocks of length $n$ over $\Lambda$, identified with the corresponding cylinder sets. Since, for any $n' < n$, any cylinder corresponding to a block of length $n'$ is a disjoint union of cylinders corresponding to blocks of length $n$, for proving that two measures $\mu$ and $\nu$ are close it suffices to find just one large enough integer $n$ and small enough $\theta > 0$ such that $|\mu(D) - \nu(D)| < \theta$ for all blocks $D$ of length $n$.

If $D = (d_1, \ldots, d_n)$ and $C = (c_1, \ldots, c_n)$ are finite sequences (blocks) of real numbers of the same length, we define their correlation as the average

\[
\overline{DC} = \frac{1}{n} \sum_{i=1}^{n} d_i c_i.
\]

If $C = (c_1, \ldots, c_n)$ and $D$ is a block of length $n'$ where $n' \leq n$ then the frequency of the occurrence of $D$ in $C$ is

\[
\text{freq}_C(D) := \frac{1}{n} \# \{ i = 1, \ldots, n - n' + 1 : c_i \ldots c_{i+n'-1} = D \}.
\]

2. **The main result**

**Theorem 2.1.** Let $y$ be an aperiodic sequence and let $N \geq 2$ be a fixed integer. There exists a subshift $\Sigma$ over $N$ symbols of topological entropy arbitrarily close to $\log N$, uncorrelated to $y$. 
A closer examination reveals that in fact we have obtained a stronger result: the above subshift $\Sigma$ is uniformly uncorrelated to $y$. We will justify this observation soon (see Fact 3.1 below).

The main goal of this short note is to strengthen the hypothesis and prove:

**Theorem 2.2.** Let $y$ be an aperiodic sequence and let $N \geq 2$ be a fixed integer. There exists a strictly ergodic subshift $\Sigma'$ over $N$ symbols of topological entropy arbitrarily close to $\log N$, uniformly uncorrelated to $y$.

The motivation for the above refinement of the former result from [DS] comes partly from the so-called strong MOMO property, introduced by Lemańczyk and coauthors in a recent paper [AKLR]. For a topological dynamical system $(X, T)$ the strong MOMO property is a form of disjointness between the system and the Möbius function. If the Möbius function is replaced by a bounded sequence $y$, we are dealing with an analog of the strong MOMO property, which we will call “strong $y$-MOMO” (although puristically it should be “strong $y$OMO”). It is important that the strong $y$-MOMO property implies uniform uncorrelation between $(X, T)$ and $y$.

The authors of [AKLR] prove that Sarnak’s conjecture is equivalent to its version in which Möbius disjointness is replaced by the strong MOMO property. It is also proved that if $y$ is generic for a Bernoulli measure then no system of positive entropy has the strong $y$-MOMO property. On the other hand, using a disjointness argument, one can show that every zero entropy system has the strong $y$-MOMO property. In this manner, the strong $y$-MOMO property remarkably allows to “distinguish zero from positive entropy” using just one “test sequence”. Our current result implies that uniform uncorrelation to just one “test sequence” $y$ does not allow to fully distinguish between zero and positive entropy, hence uniform uncorrelation is (at least when $y$ is generic for a Bernoulli measure) essentially weaker than the strong $y$-MOMO property.

Existence of a strictly ergodic example is important in the following context: in [CDS] we have shown that it is very unlikely that every strictly ergodic system is uncorrelated to the Möbius function. For that, the measure generated by the Möbius function would have to be uncorrelated to any ergodic measure. This is a very specific and rare property which fails, for instance, if the Chowla conjecture holds. So, it is believed that the Möbius function correlates with many strictly ergodic systems, and under Sarnak’s conjecture, all such systems have positive entropy. In this setup, there are a priori three possibilities:

(1) All strictly ergodic systems with positive entropy correlate with the Möbius function;
(2) Every ergodic system with positive entropy has a strictly ergodic model which correlates with the Möbius function;
(3) There exists an ergodic system with positive entropy whose all ergodic models do not correlate with the Möbius function.

Our result eliminates the option (1) above. Whether (2) or (3) is true, remains at the moment an open question. The option (2) leaves another possibility open: perhaps every ergodic system has a strictly ergodic model which does not correlate with the Möbius function (our guess is that this is not true).
3. The proofs

We need to start by briefly recalling some details of the construction of the subshift $\Sigma$ in Theorem 2.1. All involved subshifts are based on a fixed alphabet $\Lambda$ of cardinality $N \geq 2$. The desired $\Sigma$ is the intersection of a nested sequence of certain subshifts $\Sigma_k$. By definition, $\Sigma_k$ consists of all infinite concatenations (and their shifts) of blocks belonging to a family $G_k \subset \Lambda^{N_k}$ (each element of $G_k$ is a block of length $N_k$). Since $N_k$ tends to infinity, the topological entropy of $\Sigma$ can be computed as the following (nonincreasing) limit:

$$h(\Sigma) = \lim_k h(\Sigma_k) = \lim_k \frac{1}{N_k} \log(\#G_k).$$

The construction begins by setting $N_0 = 1$ and letting $G_0 = \Lambda$ so that $\Sigma_0$ is the full shift on $N$ symbols. In the inductive step $k \geq 1$ we assume that the family $G_{k-1}$ (and thus the subshift $\Sigma_{k-1}$) is already defined and we pass to constructing the family $G_k$. In order to do so, we must fix several parameters, one of which is a positive integer $m_k$ (called the multiplier) equal to the ratio $\frac{N_k}{m_k-1}$ (the family $G_k$ consists of some carefully selected concatenations of $m_k$ blocks from $G_{k-1}$). The initial value of the sequence $\{m_k\}$ is $m_1 = M \geq 81$, subsequently the multipliers tend very slowly and nondecreasingly to infinity (the sequence $\{m_k\}$ has long intervals of constancy and infinitely many jumps up, each only by a unit). The first index $k$ for which $m_k$ assumes a given value $m \geq M$ is denoted by $K_m$ and called the $m$th jump index. The jump indices form a rapidly growing sequence whose speed of growth is regulated by specific conditions. Since the rules give only lower bounds, within the same construction scheme we are free to impose any faster growth.

We also make use of two sequences of positive numbers, $\{\epsilon_k\}$ and $\{\delta_k\}$, both tending very slowly to zero (in this paper we will be dealing with only one sequence $\{\epsilon_k + \delta_k\}$). We note that in order to verify the uncorrelation between $\Sigma$ and $y$ it suffices to consider only some specific continuous functions $f: \Sigma \to \mathbb{R}$, namely the functions with values in $\{-1, 1\}$, depending on finitely many nonnegative coordinates (we call such functions codes). A convenient property of each code $f$ is that it can be applied not only to infinite sequences $x \in \Lambda^\mathbb{N}$ but also to any sufficiently long block $B$ over the alphabet $\Lambda$, producing as an output a slightly shorter block $f(B)$ over $\{-1, 1\}$ (assuming that $B$ is very long we will ignore the difference in lengths without further consequence). We represent the countable family of all codes as an increasing union of finite families $F_k$ and in each inductive step $k$ we consider only the codes from $F_k$.

By definition, the family $G_k$ consists of all concatenations $B$ of $m_k$ blocks from $G_{k-1}$ which pass the following correlation test:

(R) for every $1 \leq j \leq (m_k^2 - 1)N_k$ and every $f \in F_k$, letting $C = y_j^{j+N_k-1}$ we have $|f(B)C| < 2(\epsilon_k + \delta_k)$.

Informally, we demand all images of $B$ under the codes from the finite family $F_k$ to have small correlations with every block of $y$ of length $N_k$, ending before the position $m_k^2N_k$.

We denote by $\gamma_k$ the “probability of passing the correlation test”, i.e., the probability that a block $B \in (G_{k-1})^{m_k}$ satisfies (R). Then the cardinality of the family $G_k$ is given by:

$$\#G_k = (\#G_{k-1})^{m_k}\gamma_k.$$
A composition of the above, applied to $i$ ranging from 1 to $k$ (recall that $m_i = \frac{N_i}{N_{i-1}}$), yields:

$$\#G_k = N^N_k \cdot \frac{N_k}{N_1} \cdot \frac{N_{k-1}}{N_2} \cdots \frac{N_1}{N_{k-1}} \cdot \frac{N_k}{N},$$

which allows us to evaluate the topological entropy of $\Sigma$ as the limit

$$h(\Sigma) = \lim_{k} \frac{1}{N_k} \log(\#G_k) = \log N + \sum_{k=1}^{\infty} \frac{\log(\gamma_k)}{N_k}.$$ 

As long as all probabilities $\gamma_k$ are greater than or equal to $\frac{1}{2}$ (which we later show to be true), we have the following lower estimate of the topological entropy of $\Sigma$:

$$h(\Sigma) \geq \log N - \log 2 \cdot \sum_{k=1}^{\infty} \frac{1}{M_k} = \log N - \frac{\log 2}{M-1}.$$ 

By choosing the initial multiplier $M$ large, the entropy $h(\Sigma)$ can be made as close to $\log N$ as we desire. In this work, we will make use of another immediate consequence of the above entropy formula—an upper bound on the difference between the topological entropies of $\Sigma_p$ and $\Sigma_k$ for any $p < k$:

$$h(\Sigma_p) - h(\Sigma_k) < \log 2 \cdot \sum_{i=p+1}^{\infty} \frac{1}{N_i} = \frac{\log 2}{N_p} \sum_{i=1}^{\infty} \frac{1}{(m_p)^i} = \frac{\log 2}{N_p(m_p - 1)}.$$ 

We will now argue why the uncorrelation between $y$ and $\Sigma$ is uniform.

**Fact 3.1.** The subshift $\Sigma$ is uniformly uncorrelated to $y$.

*Proof.* Of course, it suffices to test uniformity of the uncorrelation only on codes. To this end, we copy the proof of uncorrelation from [DS] and we indicate the place where uniformity is implicit:

Let $f$ be any $\{-1,1\}$-valued function depending on finitely many nonnegative coordinates. Fix some point $x \in \Sigma$ and pick $n \in \mathbb{N}$. Let $k$ be the smallest integer such that $n < m^2 N_k$ (by convention $m$ abbreviates $m_k$). If $f$ is not in $\mathcal{F}_k$ then we simply must pick a larger $n$. So, we can assume that $f \in \mathcal{F}_k$. Now, $x \in \Sigma_k$, which means that $x^n_k$ is a concatenation of the blocks from $G_k$, except that the first and last component blocks may be incomplete. The contribution of these parts in the length is at most $\frac{2N_k}{n}$, and since $n \geq m^2 N_{k-1} \geq (m-1)^2 N_{k-1} > (m-2)N_k$, this contribution is less than $\frac{2}{m-2}$, and such is also the maximal contribution of these parts in the evaluation of the correlation between $x^n_k$ and $y^n_1$. The rest of the correlation is the average of the correlatons of the complete component blocks from $G_k$ with their respective subblocks of length $N_k$ of $y$. Since all these subblocks end before the position $m^2 N_k$, by (R), each of these correlations is less than $2(\epsilon_k + \delta_k)$ in absolute value. Jointly, the absolute value of the correlation of $x^n_k$ with $y^n_1$ does not exceed

$$\frac{2}{m-2} \cdot 1 + \frac{m-4}{m-2} \cdot 2(\epsilon_k + \delta_k).$$

Just notice that the above argument is completely independent of $x \in \Sigma$. The final estimate of the correlation depends exclusively on $f$ and $n$ ($f$ determines the lower bound for $n$, and $n$ determines $k$ and hence both $m$, $\epsilon_k$ and $\delta_k$). $\Box$

We pass to proving the main result of this paper.
Proof of Theorem 2.2. We shall indicate a modification of the construction of $\Sigma$, leading to a new subshift $\Sigma'$ which carries only one invariant measure (and maintains the other properties). Once this is done, obtaining a strictly ergodic (i.e., uniquely ergodic and minimal) subshift uncorrelated to $y$ is trivial: Every minimal subsystem of $\Sigma'$ carries an invariant measure and since such measure is unique, the minimal subsystem is also unique (let us denote it by $\Sigma''$) and supports the same measure. Thus, $\Sigma''$ is the desired strictly ergodic subshift. From now on we will focus on constructing $\Sigma'$. The modification consists of two steps. The first one is in fact no modification at all, as we simply impose a faster growth of the sequence of jump indices $K_m$. The resulting subshift fits in the original scheme and will be still denoted by $\Sigma$. The second modification is more substantial: in the construction of $\Sigma$ we replace the families $G_k$ by their proper subfamilies $G'_k$. In this manner we create a subsystem $\Sigma'$ of $\Sigma$. Clearly, $\Sigma'$ remains uniformly uncorrelated to $y$. We will only need to verify that the entropy of $\Sigma'$ is close to $\log N$ and that $\Sigma'$ is strictly ergodic.

We continue by recalling more details of the original construction in [DS]. The speed of growth of $K_m$ is ruled by two conditions:

(a) Some technical condition which we will not use or change;
(b) $9 \cdot \alpha(m) \cdot \left(\frac{8}{9}\right)^{K_m-1} < \frac{1}{2^m}$, where $\alpha(m) > 0$ depends only on $m$.

For each step number $k$ we define the reference index as $p_k = m_k - M$ (which is always smaller than $k$). The reference index grows with $k$ as slowly as $m_k$ does (remaining constant throughout many steps and jumping up only by a unit). Each block from $G_k$ (which by definition is a concatenation of the blocks from $G_{k-1}$) is clearly also a concatenation of (much shorter) blocks from $G_{p_k}$. This fact will soon play an important role in our modified construction.

The proof of the main statement of [DS] relies on the validity of the following lemma (extracted from Lemma 3.2 in [DS]):

Lemma 3.2. For every $k$ we have:

(A) $\sum_{s=p_k+1}^{k-1} (1 - \gamma_s) < \frac{1}{2^k}$,
(B) a technical condition which we will not use or change,
(C) $\gamma_k > 1 - \alpha(m_k) \left(\frac{8}{9}\right)^{k-1}$.

Note also that (C) combined with (b) and with the obvious fact that $k \geq K_{m_k}$ guarantees that $\gamma_k > 1 - 2^{-(m_k+2)}$ (so $\gamma_k$ is much larger than $\frac{1}{2}$). The proof of the above lemma starts with showing (A) using the condition (b) and the inductively assumed, for all $s < k$, condition (C). The proof of the condition (B) depends only on (A), and then the proof of (C) (the version for $\gamma_k$) relies on (A) and (B). Since we do not invoke (B), it is essential for us that (C) follows from (A). This implication relies on the particular design of the correlation test (R).

We now introduce the first innovation in the construction—a slightly sharper requirement on how large $K_m$ must be. Namely, instead of (b) we demand that

(b') $9 \cdot \alpha(m) \cdot \left(\frac{8}{9}\right)^{K_m-1} < \frac{1}{2^m}$.

Let us discuss some consequences of this modification. Suppose that for all $s < k$ we replace the families $G_s$ by their subfamilies $G'_s$ in such a way that the corresponding probabilities $\gamma'_s = \frac{\#G'_s}{\#G'_{s-1}}$ instead of (C) (for $s$) satisfy a slightly weaker version:

(C') $\gamma'_s > 1 - \alpha(m_s) \left(\frac{8}{9}\right)^{s-1}$. 

Then, as easily verified, our sharpened condition \( (b') \) allows to prove a version of \( (A) \) with these new probabilities replacing the old ones, as follows:

\[
(A') \quad \sum_{s=p_k+1}^{k-1} (1 - \gamma'_s) < \frac{\delta_k}{2}.
\]

Now, if we denote by \( \bar{\gamma}_k \) the probability that a concatenation from \( (G'_k)_{m_k} \) passes the correlation test \( (R) \), then, by the same proof as that of \( (A) \implies (C) \), \( (A') \) implies \( (C) \) for \( \bar{\gamma}_k \), i.e.,

\[
\bar{\gamma}_k > 1 - \alpha(m_k)(\frac{8}{9})^{k-1}.
\]

(The reason why we do not denote \( \bar{\gamma}_k \) by \( \gamma'_k \) will become clear later.)

We need to impose one more requirement on the growth of the jump in dices \( K_{m_k} \).

First we fix a decreasing to zero sequence of positive numbers \( \{r(j)\}_{j \geq 1} \), and for each \( j \) we find (by referring to the comment following (1.2)) \( n(j) \in \mathbb{N} \) and \( \theta(j) > 0 \) such that if two invariant measures \( \mu \) and \( \nu \) satisfy

\[
|\mu(D) - \nu(D)| < \theta(j)
\]

for all blocks (identified with cylinders) \( D \) of length \( n(j) \) then we have \( d(\mu, \nu) < r(j) \).

Now, for each \( j \) we set

\[
\beta(j) = \left(\frac{\theta(j)^2}{128}\right)
\]

and we find the smallest integer \( p(j) \) such that

\[
(3.2) \quad \frac{\log 2}{m_{p(j)}} - 1 < \beta(j) \quad \text{and} \quad \frac{n(j)}{N_{p(j)}} < \frac{\theta(j)}{4}.
\]

The new requirement (on the largeness of \( K_{m_k} \)) which we are about to force applies only to indices \( m = m(j) = p(j) + M_j \geq 1 \). Since throughout this and the following paragraph \( j \) remains fixed, we will skip “(j)” in the denotation of \( m(j) \), \( n(j) \), \( p(j) \), etc. Note that between steps \( p \) and \( K_{m_k} \), the multiplier is at least \( m_p \) thus the ratio \( \frac{N_{K_{m_k}}}{N_p} \) will be larger than or equal to \( m_p K_{m_k}^{-p} \). Finally, recall that \( N \) stands for the cardinality of the alphabet \( \Lambda \).

Here is the requirement: we demand that \( K_{m_k} \) is so large that

\[
(E) \quad 2N^n \exp(-\beta m_p K_{m_k}^{-p}) < \alpha(m) \left( \left(\frac{8}{9}\right)^{K_{m_k}^{-1}} - \left(\frac{8}{9}\right)^{K_{m_k}^{-1}} \right).
\]

This is clearly satisfied if \( K_{m_k} \) is large enough, because the left hand side decreases (with growing \( K_{m_k} \)) to zero with doubly exponential speed, while the right hand side does it only exponentially fast. We remark, that since the values of \( m_p \) and \( m \) depend on jump steps much smaller than \( K_{m_k} \), this inductive definition of the sequence \( \{K_{m_k}\} \) is correct.

With the above choice of the jump steps, the otherwise unmodified scheme leads to a sequence of families \( G_k \), a sequence of subshifts \( \Sigma_k \) and a subshift \( \Sigma \) which has entropy larger than \( \log N - \frac{\log 2}{2M} \) and is uniformly uncorrelated to \( y \). We will now present the second, more substantial, modification of the inductive construction through which we create subfamilies \( G'_k \) of \( G_k \), subsystems \( \Sigma'_k \) of \( \Sigma_k \) and the desired subsystem \( \Sigma' \) of \( \Sigma \). The modification consists in applying, in addition to the correlation test \( (R) \), a new test, which we will call the Bernstein’s test (described below). The new test will be applied only at steps whose indices \( k \) have
that for some $j$, $K$ will abbreviate and that the corresponding probabilities $\gamma'(s) = \frac{\# G'_s}{(\# G'_{s-1})^N}$ satisfy the inequality (C'). Since $j$ is now fixed we return to our previous notational convention. Moreover, we will abbreviate $K_m$ as $k$ (and so $m = m_k$). The goal is now to define $G'_k \subseteq G_k$.

At this point we create a temporary family $\tilde{G}_k$ by applying to the concatenations from $(G'_{k-1})^m$ just the correlation test (R). By the discussion on the preceding page, the probability $\tilde{\gamma}_k$ of passing the test satisfies the condition (C). For brevity we denote by $q$ the ratio $\frac{N_p}{N}$.

For a fixed block $D \in \Lambda^n$, let us define a random variable $X_D$ on $G'_p$ (note that in step $k$, $G'_p$ is already defined) by setting $X_D(b) = \text{freq}_b(D)$, $b \in G'_p$, and let $\bar{X}_D$ denote the expected value of $X_D$.

Now we create the target family $G'_k$ by discarding from $\tilde{G}_k$ all blocks $B$ which fail the Bernstein’s test, i.e., have the following property (since $G'_k \subseteq (G'_p)^q$ we can formulate the condition for all blocks in $(G'_p)^q$):

(F) A block $B \in (G'_p)^q$, $B = b_1b_2 \ldots b_q$, $b_i \in G'_p$, $i = 1, \ldots q$, fails the Bernstein’s test if for at least one block $D \in \Lambda^n$, we have

$$\left| \frac{1}{q} \sum_{i=1}^{q} X_D(b_i) - \bar{X}_D \right| > \sqrt{8\beta}.$$

This completes the definition of $G'_k$ for $k = K_{m(j)}$. We will show in a moment that the corresponding probability $\gamma'_k = \frac{\# G'_k}{(\# G'_{k-1})^N}$ satisfies (C') (the version for $k$). For indices $k$ strictly between $K_{m(j)}$ and $K_{m(j+1)}$ the families $G'_k$ are created according to the original scheme, i.e., using the correlation test only. By the same argument as used above for $\tilde{\gamma}_k$, the resulting probabilities $\gamma'_k$ satisfy (C) (and hence (C') for $k$), so that the construction can be continued for $k = K_{m(j+1)}$.

Modulo the missing proof of (C') for $k = K_{m(j)}$, the families $G'_k$ and the resulting subshift $\Sigma'$ are now determined. Once (C') is proved, we will also know that all the probabilities $\gamma'_k$ are larger than $\frac{1}{2}$, which will in turn imply that the subshift $\Sigma'$ has topological entropy at least $\log N - \frac{\log 2}{N_p}$. In the end we will also need to verify that $\Sigma'$ is strictly ergodic.

We pass to proving the missing condition (C') for $k = K_{m(j)}$. Since $j$ is now fixed, we apply again our notational convention in which “(j)” is skipped and $K_m = k$. The classical Bernstein’s inequality (see e.g. [B]) implies that in the space $(G'_p)^q$ the probability of failing the Bernstein’s test is smaller than $2N^n \exp(-2q\beta)$. We can express this fact as follows:

$$\mathbb{P}(F) < 2N^n \exp(-2q\beta),$$

where $\mathbb{P}$ denotes the normalized counting measure on $(G'_p)^q$ and $F$ symbolizes the event of failing the Bernstein’s test.

Note that, since all the new probabilities $\gamma'_s$ and $\tilde{\gamma}_k$ are larger than $\frac{1}{2}$, the inequality (3.1) is valid for $\Sigma'$ and $\Sigma_k$:

$$h(\Sigma'_p) - h(\Sigma_k) < \frac{\log 2}{N_p(m_p - 1)}.$$
By the definition of $p = p(j)$, the right hand side above is less than $\frac{q}{N_p}$. We have proved that $$\frac{1}{N_k} \log \#G_k > \frac{1}{N_p} \log \#G'_p - \frac{\beta}{N_p},$$ which implies that $$\#G_k > (#G'_p)^q \exp(-q\beta).$$

In other words, the probability that a randomly chosen concatenation $B \in (G'_p)^q$ belongs to $\tilde{G}_k$ is larger than $\exp(-q\beta)$. Let us note this fact as follows:

(3.4) $$P(\tilde{G}_k) > \exp(-q\beta).$$

Dividing the right hand side of (3.4) by the right hand side of (3.3) we obtain an upper estimate on the conditional probability in $\tilde{G}_k$ of failing the Bernstein’s test:

$$P(F|\tilde{G}_k) < 2N^n \exp(-q\beta).$$

Since $q \geq m_p^k - p$, the right hand side above is dominated by the left hand side of the inequality (E) (in which $K_m$ is written as $k$), hence we obtain:

$$P(F|\tilde{G}_k) < \alpha(m)((\frac{\alpha}{9})k^{-1} - (\frac{q}{9})k^{-1}).$$

So, the new probability $\gamma'_k$ that a block from $(G'_{k-1})^m$ passes both tests (and thus belongs in $G'_k$) equals

$$\gamma'_k = \tilde{\gamma}_k \cdot (1 - P(F|\tilde{G}_k)) > (1 - \alpha(m)((\frac{\alpha}{9})k^{-1} - (\frac{q}{9})k^{-1})) > 1 - \alpha(m)((\frac{\alpha}{9})k^{-1}).$$

We have proved that (C’) holds for $k$, as needed.

It remains to show strict ergodicity of $\Sigma'$. Since the sets $M(\Sigma'_k)$ form a nested sequence of compact subsets of $M(\Lambda^N)$ with intersection equal to $M(\Sigma')$ it suffices to prove that the diameters of $M(\Sigma'_k)$ tend to zero along some subsequence. Once again, we fix some index $j$ and return to our notational convention as above. Recall that if two invariant measures $\mu$ and $\nu$ satisfy

$$|\mu(D) - \nu(D)| < \theta$$

for all blocks $D$ of length $n$, then we have $d(\mu, \nu) < r$. Let us call blocks of length $n$ short, the elements of $G'_p$, medium blocks (these have length $N_p$), and the concatenations belonging to $G'_k$ random long blocks (as before, $q = \frac{N}{N_p}$).

If $D$ is a short block and $B$ is a long block, we can approximate the frequency of occurrence of $D$ in $B$ as the weighted average of its frequencies in the medium blocks $b \in G'_p$, as follows:

$$\text{freq}_B(D) = \left[ \sum_{b \in G'_p} \text{freq}_B(b) \cdot \text{freq}_b(D) \right] \pm \frac{\omega}{N_p}.$$ 

The term $\text{freq}_B(b)$ accounts only “regular” occurrences of $b$ in $B$, i.e., the occurrences of $b$ as components in the concatenation of elements of $G'_p$ which constitute $B$. The symbol $\pm \frac{\omega}{N_p}$ stands for an error term whose absolute value does not exceed $\frac{\omega}{N_p}$ (and thus $\frac{q^2}{4}$). It is needed to account the possible occurrences of $D$ covered by two components of this concatenation. The sum in square brackets equals $\frac{1}{q} \sum_{i=1}^q X_D(b_i)$ (using the notation introduced in the paragraph preceding
Since $B$ belongs to $G'_k$ and hence passes the Bernstein’s test, this average differs from $\bar{X}_D$ by less than $\sqrt{8\beta}$ which equals $\frac{\theta}{2}$. Eventually, we have

$$\text{freq}_B(D) = \bar{X}_D \pm \frac{\theta}{2}.$$ 

This clearly implies that the frequencies of $D$ in any two blocks $B, B' \in G'_k$ differ from each other by at most $\theta$. This property applies also to all elements $x \in \Sigma'_k$: for any $x, x' \in \Sigma'_k$ the lower density of occurrence of $D$ in $x$ and the upper density of occurrence of $D$ in $x'$ differ by at most $\theta$ (the error term $\pm \frac{n}{N_p}$ absorbs also these occurrences of $D$ in $x \in \Sigma'_k$ which are covered by two components of the infinite concatenation of the blocks $B \in G'_k$ which constitute $x$). This easily implies that $|\mu(D) - \nu(D)| < \theta$ for any $\mu, \nu \in \mathcal{M}(\Sigma'_k)$. By the choice of $n$ and $\theta$ we obtain that the diameter of $\mathcal{M}(\Sigma'_k)$ is at most $r = r(j)$. Since the sequence $\{r(j)\}$ tends to zero, we have shown that the diameters of $\mathcal{M}(\Sigma'_k)$ tend to zero along the subsequence indexed by $k = K_{m(j)}$, $j \geq 1$. This ends the proof.

\[\square\]

References

- [AKLR] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, T. de la Rue, *Möbius disjointness for models of an ergodic system and beyond*, Israel Journal of Mathematics 228 (2018), 707–751.
- [B] S.N. Bernstein, *Probability Theory (in Russian)*, Gosizdat, Moscow, 1927.
- [CDS] J.-P. Conze, T. Downarowicz, and J. Serafin, *Correlation of sequences and of measures, generic points for joinings and ergodicity of certain cocycles*, Trans. Amer. Math. Soc. 369 (2017), 3421–3441.
- [DS] T. Downarowicz and J. Serafin, *Almost full entropy subshifts uncorrelated to the Möbius function*, International Mathematics Research Notices (2017), https://doi.org/10.1093/imrn/rnx192
- [S] P. Sarnak, *Three lectures on the Möbius function randomness and dynamics*, http://publications.ias.edu/sites/default/files/MobiusFunctionsLectures(2).pdf.