THE STRUCTURE OF STRONGLY STATIONARY SYSTEMS

NIKOS FRANTZIKINAKIS

ABSTRACT. Motivated by a problem in ergodic Ramsey theory, Furstenberg and Katznelson introduced the notion of strong stationarity, showing that certain recurrence properties hold for arbitrary measure preserving systems if they are valid for strongly stationary ones. We construct some new examples and prove a structure theorem for strongly stationary systems. The building blocks are Bernoulli systems and rotations on nilmanifolds.

1. INTRODUCTION

1.1. **Historical background.** In 1975 Szemerédi proved the following long standing conjecture of Erdös and Turàn:

Theorem 1.1 (Szemerédi). *Let Λ be a subset of the integers with positive upper density. Then Λ contains arbitrarily long arithmetic progressions.*

Szemerédi's proof was combinatorial in nature and intricate. In 1977 Furstenberg ([Fu77]) gave an entirely different proof using ergodic theory. He showed that Szemerédi's theorem is equivalent to a statement about multiple recurrence of measure preserving systems and then proved the ergodic version:

Theorem 1.2 (Furstenberg). *Let $(X, \mathcal{B}, \mu, T)$ be a finite measure preserving system and $A \in \mathcal{B}$ be a set with positive measure. Then for every $k \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that*

$$\mu(A \cap T^{-n}A \cap \cdots \cap T^{-nk}A) > 0.$$

Furstenberg's proof launched the field of ergodic Ramsey theory, where problems in combinatorics are translated to recurrence related statements of measure preserving systems and then proved using ergodic theory. Using this approach Furstenberg and Katznelson, and more recently Bergelson and Leibman (among others) established several ergodic theoretic results whose combinatorial implications are not currently attainable by any other methods. This includes a multidimensional and a polynomial extension of Szemerédi's theorem ([FK79], [BL96]), and the density version of Hales-Jewett theorem ([FK91]), the "master" theorem that contains several major results in the field as special cases.

The notion of strong stationarity (Definition 3.1) was introduced in the paper of Furstenberg and Katznelson ([FK91]) in proving the density version of Hales-Jewett

theorem. An important result established in the same paper is that an arbitrary stationary process "majorizes" a strongly stationary one (Section 3.2). From this it follows that several recurrence properties are valid for arbitrary measure preserving systems if they are valid for strongly stationary ones. In particular, it turns out to be sufficient to verify Theorem 1.2 for the specific case of a strongly stationary system. This motivated the problem of determining the structure of strongly stationary systems.

In [Je97] Jenvey proved that every ergodic strongly stationary system is necessarily Bernoulli. Unfortunately, not every ergodic system majorizes an ergodic strongly stationary one, and the ergodic components of a strongly stationary system are not necessarily strongly stationary. In fact nonergodic strongly stationary systems can have completely different structure than the ergodic ones; there exist several distal examples that one has to identify.

We give a structure theorem for strongly stationary systems (Theorems 6.6 and 6.9):

Main Theorem. *(i) Almost every ergodic component of a strongly stationary system is isomorphic to the direct product of a Bernoulli system and a totally ergodic pro-nilsystem (defined in Section 6.1).*
(ii) An extremal strongly stationary system (Definition 4.1) is isomorphic to the direct product of a Bernoulli system and a strongly stationary system associated to some pro-nilmanifold (defined in Section 6.4).

Moreover, we construct new examples of strongly stationary systems (Section 6.4, examples (iv) and (v)).

1.2. **Format of the paper.** After reviewing some preliminary notions and results in Section 2 we define strong stationarity in Section 3. We prove that an arbitrary stationary process majorizes a strongly stationary one and give the basic examples of strongly stationary systems.

The general strongly stationary system is an integral of extremal ones. In Section 4 we give necessary and sufficient conditions for extremality and prove a homogeneity property for extremal strongly stationary systems.

In Section 5 we prove that almost every ergodic component of a strongly stationary system is isomorphic to the direct product of a distal system and a Bernoulli system. Moreover, we show that the distal factor of a strongly stationary system is strongly stationary and coincides with the characteristic factor of the system (defined in Section 5.4). This reduces our problem to determining the structure of distal strongly stationary systems.

Finally, in Section 6 we obtain a structure theorem for distal strongly stationary systems. Using results from Section 5, in conjunction with a recent result of Host and Kra ([HK03]), we show that their ergodic components are pro-nilsystems. Moreover, we construct new examples of distal strongly stationary systems with nonaffine ergodic components. This new set of examples allows us to give a complete classification.

2. Preliminaries

To facilitate the reading, we establish our notation and review some basic results that are used in the sequel. We refer the reader to [Fu81], [Pe89], and [Wa82] for more details.

2.1. Measure preserving systems.

A *measure preserving system* (or just system) is a measure space $(X, \mathcal{B}, \mu)$ together with a measurable measure preserving transformation T on it. Throughout the discussion we assume that all measure spaces are Lebesgue. When there is no danger of confusion we use the bold symbol $\mathbf{X}$ to denote the system $(X, \mathcal{B}, \mu, T)$. We also use the bold symbol $\mathbf{T}$ to denote the operator $\mathbf{T} \colon L^\infty(\mathbf{X}) \to L^\infty(\mathbf{X})$ defined by $(\mathbf{T}f)(x) = f(T(x))$.

Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an invertible measure preserving system and $\mathcal{B}_0$ be a sub-σ-algebra of $\mathcal{B}$. By $\bigvee_{-n}^{m} T^i \mathcal{B}_0$ we denote the σ-algebra spanned by sets of the form $T^i B$ where $B \in \mathcal{B}_0$, $-n \leq i \leq m$. We say that $\mathcal{B}_0$ is T-*generating* if $\bigvee_{-\infty}^{\infty} T^i \mathcal{B}_0 = \mathcal{B}$ up to null sets. If $\mathcal{F}$ is an algebra of bounded $\mathcal{B}$-measurable complex valued functions, we denote by $\mathcal{B}(\mathcal{F})$ the sub-σ-algebra of $\mathcal{F}$-measurable sets, that is, the σ-algebra generated by sets of the form $f^{-1}(A)$, where $f \in \mathcal{F}$ and $A \subset \mathbb{C}$ is open. We say that $\mathcal{F}$ is T-*generating* if $\mathcal{B}(\mathcal{F})$ is T-generating.

2.2. Furstenberg's structure theorem.

Let $\mathbf{X}$ be a measure preserving system, $\mathbf{Y} = (Y, \mathcal{E}, \nu, R)$ be a factor of $\mathbf{X}$, and $\mu = \int \mu_y \, d\nu(y)$ be the disintegration of μ over $\mathbf{Y}$. A function $f \in L^2(\mathbf{X})$ is *compact* relative to the factor $\mathbf{Y}$ if for every $\varepsilon > 0$ there exist functions $g_1, \dots, g_m \in L^2(\mathbf{X})$ such that $\min_{1 \leq s \leq m} \|\mathbf{T}^i f - g_s\|_{L^2(\mu_y)} < \varepsilon$ for every $i \in \mathbb{N}$ and for ν-a.e. $y \in Y$. An extension $\mathbf{X}$ of $\mathbf{Y}$ is *compact* if the set of compact functions relative to $\mathbf{Y}$ is dense in $L^2(\mathbf{X})$.

Starting with the factor of T-invariant functions $\mathbf{D}_0$, we define inductively $\mathbf{D}_{i+1}$ to be the maximal compact extension of $\mathbf{D}_i$. More precisely, we consider the subalgebra generated by the compact functions relative to $\mathbf{D}_i$ and we define $\mathbf{D}_{i+1}$ to be the factor determined by it. We call the factor $\mathbf{D}_k$ the *k-step distal factor* of the system $\mathbf{X}$. The maximal factor that can be exhausted by a transfinite number of compact extensions is called the *distal factor* and is denoted by $\mathbf{D}$.

We say that $\mathbf{X}$ is a *relatively ergodic* extension of $\mathbf{Y}$, if every T-invariant function on $\mathbf{X}$ is μ-a.e. a function on $\mathbf{Y}$ (that is $\mathcal{E}$-measurable).

Let $\mathbf{X}_1$, $\mathbf{X}_2$ be two extensions of $\mathbf{Y}$. The fiber product space is defined as

$$X_1 \times_Y X_2 = \{(x_1, x_2) \in X_1 \times X_2 \colon \pi_1(x_1) = \pi_2(x_2)\},$$

where $\pi_i \colon X_i \to Y$ the factor map. By $\mathcal{B}'$ we denote the restriction of $\mathcal{B}_1 \times \mathcal{B}_2$ on $X_1 \times_Y X_2$, and by μ' the measure defined by the disintegration (over $\mathbf{Y}$) $\mu'_y = \mu^1_y \times \mu^2_y$, where $\mu_1 = \int \mu^1_y \, d\nu(y)$ and $\mu_2 = \int \mu^2_y \, d\nu(y)$. If $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}$, and $\mathbf{Y}$ is the factor

determined by the action of T on a sub-σ-algebra $\mathcal{E}$ of $\mathcal{B}$, then

$$\int f(x_1)f(x_2)\,d\mu' = \int \mathbb{E}(f_1|\mathcal{E})\mathbb{E}(f_2|\mathcal{E})\,d\mu$$

where $\mathbb{E}(f|\mathcal{E})$ denotes the conditional expectation of f given $\mathcal{E}$. We can check that $\mathbf{X}_1 \times_{\mathbf{Y}} \mathbf{X}_2 = (X_1 \times_Y X_2, \mathcal{B}', \mu', T_1 \times T_2)$ is a measure preserving system that extends $\mathbf{Y}$. We say that $\mathbf{X}$ is a *relative weak mixing* extension of $\mathbf{Y}$ if $\mathbf{X} \times_{\mathbf{Y}} \mathbf{X}$ is a relatively ergodic extension of $\mathbf{Y}$.

Furstenberg ([Fu81]) proved the following structure theorem:

Theorem 2.1 (Furstenberg). *Every measure preserving system is a relative weak mixing extension of its distal factor.*

We note that the distal factor is the smallest factor with respect to which the system is relative weak mixing. The following result was proved in [Fu77] for ergodic systems and is needed in the sequel. The argument given there also works for nonergodic systems.

Lemma 2.2 (Furstenberg). *The invariant functions of $\mathbf{X} \times_{\mathbf{Y}} \mathbf{X}$ belong to the closed subspace spanned by functions of the form $f_1(x_1)f_2(x_2)$ where f_1, f_2 are functions compact relative to $\mathbf{Y}$.*

2.3. **Stationary processes.** A Λ-valued *stochastic process* is a sequence of measurable functions (random variables) $\{f_i\}_{i\in\mathbb{Z}}$ defined on a probability space $(X, \mathcal{B}, \mu)$ with values in a compact metric space Λ (the state space).

The *finite dimensional statistics* of a stationary process $\{f_i\}_{i\in\mathbb{Z}}$ is the collection of all measurements $\mu\big(\bigcap_{i=-k}^{k}\{f_i \in A_i\}\big)$, where $k \in \mathbb{N}$ and $A_i \subset \Lambda$ are open.

A stochastic process is *stationary* if its finite dimensional statistics are invariant under translations of the time parameter, that is, $\mu\big(\bigcap_{i=-k}^{k}\{f_{i+r} \in A_i\}\big) = \mu\big(\bigcap_{i=-k}^{k}\{f_i \in A_i\}\big)$ for all $k, r \in \mathbb{N}$ and open sets $A_i \subset \Lambda$.

Two stationary processes $\{f_i\}_{i\in\mathbb{Z}}$ and $\{g_i\}_{i\in\mathbb{Z}}$ (suppose that μ, ν are the corresponding underlying measures) are *equivalent* if they have the same finite dimensional statistics, that is, $\mu\big(\bigcap_{i=-k}^{k}\{f_i \in A_i\}\big) = \nu\big(\bigcap_{i=-k}^{k}\{g_i \in A_i\}\big)$ for every $k \in \mathbb{N}$ and open sets $A_i \subset \Lambda$.

2.4. **Sequence space representations.** Let Λ be a compact metric space. The *sequence space* $\Lambda^{\mathbb{Z}}$ equipped with the product topology is again a compact metric space. We denote by x_i the i-th coordinate of a point $x \in \Lambda^{\mathbb{Z}}$. The Borel σ-algebra $\mathcal{A}^{\mathbb{Z}}$ is generated by the finite dimensional rectangles $\bigcap_{-k}^{k}\{x\colon x_i \in A_i\}$ where each $A_i \subset \Lambda$ is open. A probability measure σ defined on the completion of $\mathcal{A}^{\mathbb{Z}}$ (which we denote again by $\mathcal{A}^{\mathbb{Z}}$) is *stationary* if $\sigma\big(\bigcap_{i=-k}^{k}\{x\colon x_i \in A_i\}\big) = \sigma\big(\bigcap_{i=-k}^{k}\{x\colon x_{i+r} \in A_i\}\big)$ for every $k \in \mathbb{N}$, $r \in \mathbb{Z}$ and open sets $A_i \subset \Lambda$. Having fixed the space Λ we denote by $\mathcal{M}$ the set of all probability measures and by $\mathcal{M}_s$ the space of all stationary measures on the sequence space $\Lambda^{\mathbb{Z}}$. Both $\mathcal{M}$ and $\mathcal{M}_s$ endowed with the weak-star (w^*) topology are compact convex spaces.

Let σ be a stationary measure on the sequence space $\Lambda^{\mathbb{Z}}$. The shift operator S, defined by $(Sx)_k = x_{k+1}$, is continuous and the system $\mathbf{\Lambda}^{\mathbb{Z}} = (\Lambda^{\mathbb{Z}}, \mathcal{A}^{\mathbb{Z}}, \sigma, S)$ is measure preserving. We call it the *sequence space system* determined by the stationary measure σ. Moreover, we call the stationary process $\{x_i\}_{i\in\mathbb{Z}}$ the *sequence space process* determined by σ. To ease notation we denote by x_i both the i-th coordinate of a point x and the function that maps each point to its i-th coordinate.

Let $\{f_i\}_{i\in\mathbb{Z}}$ be a Λ-valued stationary process. If there exists a stationary measure σ on $\Lambda^{\mathbb{Z}}$ that makes the processes $\{f_i\}_{i\in\mathbb{Z}}$ and $\{x_i\}_{i\in\mathbb{Z}}$ equivalent we say that the second process is the *sequence space representation* of the first. The next classical result ([Br92], page 107) is an easy consequence of Kolmogorov's extension theorem.

Proposition 2.3. *Every Λ-valued stationary process $\{f_i\}_{i\in\mathbb{Z}}$ has a $\Lambda^{\mathbb{Z}}$ sequence space representation.*

Let $\mathbf{X}$ be a measure preserving system and $\mathcal{F}$ be a T-generating subalgebra. Suppose that $\mathbf{X}$ is isomorphic to a sequence space system $\mathbf{I}^{\mathbb{Z}} = (I^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \sigma, S)$, where $I = [0,1]$, and the isomorphism $\phi\colon X \to I^{\mathbb{Z}}$ maps sets in $\mathcal{B}(\mathcal{F})$ to x_0-measurable sets. Then we say that the system $\mathbf{I}^{\mathbb{Z}}$ is the *sequence space representation of $\mathbf{X}$ with respect to the subalgebra $\mathcal{F}$.* The next proposition is a variation of a classical result:

Proposition 2.4. *Every invertible measure preserving system $\mathbf{X}$ has a sequence space representation with respect to any T-generating subalgebra $\mathcal{F}$.*

Sketch of the Proof. By a classical result of Rokhlin ([Ro62]) the sub-σ-algebra $\mathcal{B}_0 = \mathcal{B}(\mathcal{F})$ induces a partition $\mathcal{P} = \{P_t\}_{t\in I}$ of X by $\mathcal{B}_0$-measurable sets such that every $\mathcal{B}_0$-measurable set is equal (up to a set of measure zero) to a union of partition elements P_t. For every open set $A \subset I$ we set $A' = \bigcup_{t\in A} P_t$. We define the measure σ on cylinder sets of $I^{\mathbb{Z}}$ by $\sigma\left(\bigcap_{i=-n}^{n}\{x\colon x_i \in A_i\}\right) = \mu\left(\bigcap_{i=-n}^{n}\{x\colon T^i x \in A_i'\}\right)$, and then extended it to the whole sequence space using Kolmogorov's extension theorem. We can check that the advertised sequence space representation of $\mathbf{X}$ is determined by the measure σ.

2.5. **Van der Corput's Lemma.** The following classical lemma will be needed in the sequel ([FW96], page 47):

Lemma 2.5 (Van der Corput). *Let $\{x_n\}_{n\in\mathbb{N}}$ be a bounded sequence of vectors in a Hilbert space. For each m we set*

$$b_m = \overline{\lim}_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N} <x_{n+m}, x_n>\right|.$$

Assume that

$$\overline{\lim}_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M} b_m = 0.$$

Then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n = 0$$

in the L^2 norm.

3. STRONG STATIONARITY

3.1. **Definitions.**

Definition 3.1. (i) A stationary process $\{f_i\}_{i \in \mathbb{Z}}$ is *strongly stationary* if the finite dimensional statistics of $\{f_i\}_{i \in \mathbb{Z}}$ and $\{f_{ni}\}_{i \in \mathbb{Z}}$ are the same for every $n \in \mathbb{N}$.

(ii) An invertible measure preserving system $\mathbf{X}$ is *strongly stationary* if there exists a T-generating function algebra $\mathcal{F}$ such that every process $\{\mathbf{T}^i f\}_{i \in \mathbb{Z}}$ is strongly stationary for $f \in \mathcal{F}$. When we want to also indicate the subalgebra $\mathcal{F}$ we write that $(\mathbf{X}, \mathcal{F})$ is strongly stationary.

Remark. Equivalently, a system $\mathbf{X}$ is strongly stationary if there exists a T-generating function algebra $\mathcal{F}$ such that

$$\int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \int f_0 \, \mathbf{T}^n f_1 \, \cdots \, \mathbf{T}^{kn} f_k \, d\mu$$

for $n \in \mathbb{N}$, $f_i \in \mathcal{F}$.

Let $\mathbf{\Lambda}^{\mathbb{Z}}$ be a sequence space system. The subalgebra $\mathcal{F}_0$ of bounded x_0-measurable functions is T-generating. Strong stationarity with respect to $\mathcal{F}_0$ is equivalent to saying that the maps τ_n defined by $(\tau_n x)_i = x_{ni}$ are measure preserving for every $n \in \mathbb{N}$.

Definition 3.2. A measure σ on the sequence space $\Lambda^{\mathbb{Z}}$ is *strongly stationary* if the system $(\mathbf{\Lambda}^{\mathbb{Z}}, \mathcal{F}_0)$ is strongly stationary.

3.2. **Stationary processes majorize strongly stationary ones.** Let $\{f_i\}_{i \in \mathbb{Z}}$ and $\{g_i\}_{i \in \mathbb{Z}}$ be two Λ-valued stationary processes with underline measures μ, ν correspondingly. We say that $\{f_i\}_{i \in \mathbb{Z}}$ *majorizes* $\{g_i\}_{i \in \mathbb{Z}}$ if

$$\sup_{n \in \mathbb{N}} \mu \Big(\bigcap_{i=-k}^{k} \{f_{in} \in A_i\} \Big) \geq \sup_{n \in \mathbb{N}} \nu \Big(\bigcap_{i=-k}^{k} \{g_{in} \in A_i\} \Big)$$

for every $k \in \mathbb{N}$ and open sets $A_i \subset \Lambda$.

Furstenberg and Katznelson ([FK91]) proved that every stationary process majorizes a strongly stationary one. Actually they established a much more general result using a strong selection theorem. The argument given below was suggested by Y. Peres and gives an easier proof for the case that we are interested.

Theorem 3.3 (Furstenberg and Katznelson). *Every stationary process majorizes a strongly stationary one.*

Proof. By Proposition 2.3 there exists a stationary measure σ on the sequence space $\Lambda^{\mathbb{Z}}$ such that the processes $\{f_i\}_{i\in\mathbb{Z}}$ and $\{x_i\}_{i\in\mathbb{Z}}$ have the same finite dimensional statistics. For $n \in \mathbb{N}$, let τ_n be the map defined on $\Lambda^{\mathbb{Z}}$ by $(\tau_n x)_i = x_{ni}$. It is straightforward to check that the measure $\tau_n\sigma$ defined by $\tau_n\sigma(A) = \sigma(\tau_n^{-1}A)$ is stationary. If we denote by $\mathcal{O}$ the closure in the w^*-topology of the set of all convex combinations of the measures $\tau_n\sigma$, $n \in \mathbb{N}$, then $\mathcal{O}$ is a compact convex subset of $\mathcal{M}_s$. The maps τ_n commute and act continuously and affinely on $\mathcal{O}$ so by the Markov-Kakutani fixed point theorem ([Co85], page 151) they have a common fixed point $\nu \in \mathcal{O}$.

We claim that the stationary process $\{x_i\}_{i\in\mathbb{Z}}$ which is induced by the measure ν on the sequence space $\Lambda^{\mathbb{Z}}$ is strongly stationary and is majorized by the process $\{f_i\}_{i\in\mathbb{Z}}$. Indeed, the invariance over each τ_n proves strong stationarity. Moreover, convex linear combinations of the measures $\tau_n\sigma$ come arbitrarily close to ν in the w^*-topology. Hence, for every $\varepsilon > 0$ and choice of A_i's there exists $n \in \mathbb{N}$ such that

$$\sigma\big(\bigcap_{i=-k}^{k}\{x_{in} \in A_i\}\big) \geq \nu\big(\bigcap_{i=-k}^{k}\{x_i \in A_i\}\big) - \varepsilon.$$

Since for every $n \in \mathbb{N}$ we have $\sigma\big(\bigcap_{i=-k}^{k}\{x_{in} \in A_i\}\big) = \mu\big(\bigcap_{i=-k}^{k}\{f_{in} \in A_i\}\big)$ and $\nu\big(\bigcap_{i=-k}^{k}\{x_i \in A_i\}\big) = \nu\big(\bigcap_{i=-k}^{k}\{x_{in} \in A_i\}\big)$ the result follows. $\square$

We deduce now a similar result for measure preserving systems.

Corollary 3.4. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be any (not necessarily invertible) measure preserving system and A be a $\mathcal{B}$-measurable set. Then there exists an invertible strongly stationary system $\tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$ and $B \in \tilde{\mathcal{B}}$, such that $\mu(A) = \tilde{\mu}(B)$ and*

$$\sup_{n\in\mathbb{N}} \mu\big(\bigcap_{i=0}^{k} T^{-in}A\big) \geq \sup_{n\in\mathbb{N}} \tilde{\mu}\big(\bigcap_{i=0}^{k} \tilde{T}^{-in}B\big)$$

for every $k \in \mathbb{N}$.

Proof. Let $\mathbf{1}_A$ denote the indicator function of the set A. Using a standard argument we extend the one sided stationary process $\{f_i\}_{i\in\mathbb{N}} = \{\mathbf{T}^i\mathbf{1}_A\}_{i\in\mathbb{N}}$ to a two sided one. We denote the two sided extension by $\{f_i\}_{i\in\mathbb{Z}}$. By Theorem 3.3 the process $\{f_i\}_{i\in\mathbb{Z}}$ majorizes a strongly stationary one $\{g_i\}_{i\in\mathbb{Z}}$. Following the proof of Theorem 3.3 we see that $\{g_i\}_{i\in\mathbb{Z}} = \{\mathbf{S}^i x_0\}_{i\in\mathbb{Z}}$, where S is the (invertible) shift transformation on the sequence space $\{0,1\}^{\mathbb{Z}}$ with some appropriately chosen measure σ. We let $B = \{x \in \{0,1\}^{\mathbb{Z}}: x_0 = 1\}$ and set $\tilde{\mathbf{X}} = \big(\{0,1\}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \sigma, S\big)$. If $\mathcal{F}_0$ is the subalgebra of bounded x_0-measurable functions then $(\tilde{\mathbf{X}}, \mathcal{F}_0)$ is strongly stationary. The advertised inequality is valid since $\{f_i\}_{i\in\mathbb{Z}}$ majorizes $\{g_i\}_{i\in\mathbb{Z}}$. Finally, following again the proof of Theorem 3.3 we see that $\mu(A) = \sigma(B)$. $\square$

3.3. **The τ_n's and Jenvey's result.** Let $\mathbf{X}$ be a strongly stationary system. The next proposition was proved in [Je97] and gives useful necessary and sufficient conditions for strong stationarity:

Proposition 3.5 (**Jenvey**). *The measure preserving system* $\mathbf{X}$ *is strongly stationary if and only if there exists a T-generating function algebra $\mathcal{F}$, and a family of measure preserving transformations $\{\tau_n\}_{n \in \mathbb{N}}$, that leave every function in $\mathcal{F}$ invariant and such that the operators $\mathbf{T}$ and $\boldsymbol{\tau_n}$ satisfy the commutation relations*

$$
(1) \qquad\qquad\qquad \boldsymbol{\tau_n}\mathbf{T} = \mathbf{T}^n\boldsymbol{\tau_n}, \quad n \in \mathbb{N}.
$$

Moreover, we can choose the τ_n's to satisfy $\boldsymbol{\tau_{mn}} = \boldsymbol{\tau_m}\boldsymbol{\tau_n}$, for all $m, n \in \mathbb{N}$.

Remark. Equivalently, for the point transformations T and τ_n relation (1) can be written as $(T\tau_n)(x) = (\tau_n T^n)(x)$, for a.e. $x \in X$ and $n \in \mathbb{N}$.

Using the multiple weak-mixing theorem ([Fu81], page 86) it is easy to see that if a strongly stationary system is weak-mixing then it is Bernoulli. In [Je97] Jenvey shows that the same conclusion holds if we just assume ergodicity.

Theorem 3.6 (**Jenvey**). *Every ergodic strongly stationary system is a Bernoulli system.*

We remark that strong stationarity of a system does not imply that of its ergodic components, so we cannot use this theorem to determine the structure of the general strongly stationary system.

Applying the argument in the proof of Theorem 3.6 for the general (not necessarily ergodic) strongly stationary system we can deduce the following:

Proposition 3.7. *If $\mathbf{X}$ is a strongly stationary system and λ is an eigenvalue of T then $\lambda = 1$.*

Since the proof is too long to reproduce we just indicate the strategy. Suppose that χ is a λ-eigenfunction of T, λ is not a root of unity, and that $g \in \bigvee_{-k}^{k} T^i\mathcal{F}$. We have

$$
\int \chi \, g \, d\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \boldsymbol{\tau_n}\chi \, \boldsymbol{\tau_n}g \, d\mu.
$$

To show that the limit on the right is zero we use repeatedly Van der Corput's lemma (Lemma 2.5) and relation (1). As it turns out, it suffices to show that the sequence of functions $u_n = \mathbf{T}^{-n}(\boldsymbol{\tau_{n+m}}\chi \, \overline{\boldsymbol{\tau_n}\chi})$ converges weakly to zero as $n \to \infty$ for every $m \in \mathbb{N}$. From relation (1) it follows that $\boldsymbol{\tau_n}\chi$ is a linear combination of $\lambda^{1/n}$-eigenfunctions. When λ is not a root of unity this easily gives that the sequence u_n consists of "almost" pairwise orthogonal functions and hence converges weakly to zero. It follows that χ is orthogonal to the subalgebra $\bigvee_{-k}^{k} T^i\mathcal{F}$ for every $k \in \mathbb{N}$. Since the algebra $\mathcal{F}$ is T-generating we get that $\chi = 0$. The case where λ is a nontrivial root of unity is trickier but the strategy of the proof is similar.

3.4. **Examples of strongly stationary systems.** We give now the basic examples of strongly stationary systems.

(i) *Exchangeable systems.* A system is exchangeable if there exists a T-generating function algebra $\mathcal{F}$ which has finite dimensional statistics invariant under any permutation of the time parameter. Bernoulli systems are exchangeable. A theorem of de Finetti says that a system is exchangeable if and only if it is a mixture of Bernoulli systems. We can check that exchangeability implies strong stationarity but the converse is not true as the next example shows.

(ii) *Ergodic components circle rotations.* On the 2-torus, with group action addition mod 1 and the Haar measure, define $T(x, y) = (x, y + x)$. To check that the system is strongly stationary we use Proposition 3.5. We let $\mathcal{F}$ be the algebra generated by the exponentials in y and define the maps τ_n by $\tau_n(x, y) = (nx, y)$. Since $e^{iy} \in \mathcal{F}$ and $e^{ix} = T e^{iy} e^{-iy}$, the algebra $\mathcal{F}$ is T-generating. Moreover, τ_n is measure preserving, each τ_n leaves functions in $\mathcal{F}$ invariant, and

$$T\tau_n(x, y) = (nx, y + nx) = \tau_n T^n(x, y).$$

Hence, $\boldsymbol{\tau_n}\mathbf{T} = \mathbf{T}^n\boldsymbol{\tau_n}$ for $n \in \mathbb{N}$.

(iii) *Ergodic components affine transformations on $\mathbb{T}^d$.* On $\mathbb{T}^3$, with group action addition (mod 1) and the Haar measure, define $T(x, y, z) = (x, y + x, z + y)$. To see that the induced measure preserving system is strongly stationary we use again Proposition 3.5. We let $\mathcal{F}$ be the algebra generated by exponentials in z and define the maps τ_n by $\tau_n(x, y, z) = (n^2 x, ny + \binom{n}{2}x, z)$. We can check as before that the algebra $\mathcal{F}$ is T-generating. Each τ_n clearly leave functions in $\mathcal{F}$ invariant and a direct computation shows that it satisfies the right commutation relations.

More generally, on $\mathbb{T}^d$ with group action addition (mod 1) and the Haar measure, define

$$T(x_1, x_2, \ldots, x_d) = (x_1, x_2 + x_1, \ldots, x_d + x_{d-1}).$$

This time $\mathcal{F}$ is the algebra generated by exponentials in x_d and the τ_n's have the form

$$\tau_n(x_1, x_2, \ldots, x_d) = \left(n^{d-1}x_1, n^{d-2}x_2 + \binom{n}{2}x_1, \ldots, nx_{d-1} + \cdots + \binom{n}{d-1}x_1, x_d\right).$$

(iv) *Ergodic components affine maps on more general groups.* In the last two examples $\mathbb{T}$ can be replaced by any connected compact abelian group. The form of the τ_n's and of the algebra $\mathcal{F}$ is similar. The connectedness assumption is needed to guarantee that each τ_n is measure preserving.

Although the previous examples provide an ample supply of strongly stationary systems, the building blocks are always Bernoulli systems and affine transformations on compact abelian groups. In the last section we will see that this is not the case in general (Section 6.4, example (iv)).

4. Extremality

4.1. **Definition of extremality.** Recall that by Proposition 2.4 every strongly stationary system $(\mathbf{X}, \mathcal{F})$ has a sequence space representation $(\mathbf{I}^{\mathbb{Z}}, \mathcal{F}_0)$, where $I = [0, 1]$ and $\mathcal{F}_0$ is the subalgebra of bounded x_0-measurable functions. This representation is completely determined by the measure σ on the sequence space, so characterizing (up to isomorphism) the strongly stationary measure preserving systems is equivalent to characterizing the strongly stationary measures on $I^{\mathbb{Z}}$. Furthermore, we only have to determine the structure of the extremal ones, that is, those that cannot be decomposed nontrivially into a convex combination of strongly stationary measures. We will make this more precise below.

Consider the space $\mathcal{M}_{ss}$ of all strongly stationary measures. Then $\mathcal{M}_{ss}$ is a closed convex subset of the space of stationary measures $\mathcal{M}_s$ which is w^*-compact, metrizable, and locally convex.

Definition 4.1. (i) Let σ be a strongly stationary measure on $I^{\mathbb{Z}}$. Then σ is *extremal* if it cannot be written in the form $\sigma = a\sigma_1 + (1 - a)\sigma_2$ for some $0 < a < 1$ and strongly stationary measures $\sigma_1 \neq \sigma_2$.

(ii) Let $(\mathbf{X}, \mathcal{F})$ be a strongly stationary system and σ be the strongly stationary measure that determines its sequence space representation with respect to $\mathcal{F}$. Then $(\mathbf{X}, \mathcal{F})$ is *extremal* if σ is an extremal strongly stationary measure.

We will use the following integral representation theorem of Choquet ([Ph01]):

Theorem 4.2 (Choquet). *Let X be a metrizable compact convex subset of a locally convex space E and $x_0 \in X$. Then there exists a Borel probability measure μ on X, supported on the extreme points $ext(X)$, that satisfies $x_0 = \int_{ext(X)} x \, d\mu(x)$ (that is, $l(x_0) = \int_{ext(X)} l(x) \, ds(x)$ holds for every l in the dual of X).*

It follows that the general strongly stationary measure or system is an integral of extremal ones. So we can focus our attention on determining the structure of the extremal strongly stationary measures or systems.

4.2. **Necessary and sufficient conditions.** It is well known that the set of extremal points of the space $\mathcal{M}_s$ is the set of ergodic measures (with respect to the shift transformation S), and that different ergodic measures are mutually singular. It is not hard to establish the analogous results for the space of strongly stationary measures $\mathcal{M}_{ss}$. The corresponding action on $\mathcal{M}_{ss}$ is the joint action of S and the τ_n's $((\tau_n x)_i = x_{ni})$.

Proposition 4.3. *Different extremal strongly stationary measures are mutually singular.*

Proof. Let μ_1 and μ_2 be two different extremal strongly stationary measures and let S denote the shift transformation on $I^{\mathbb{Z}}$. Consider the Lebesgue decomposition of μ_1 with

respect to μ_2, that is, write $\mu_1 = a\nu_1 + (1-a)\nu_2$ where ν_1, ν_2 are probability measures such that $\nu_1 \perp \mu_2$ and $\nu_2 \ll \mu_2$. Then

$$\mu_1 = S\mu_1 = a\, S\nu_1 + (1-a)\, S\nu_2,$$
$$\mu_1 = \tau_n\mu_1 = a\, \tau_n\nu_1 + (1-a)\, \tau_n\nu_2.$$

The Lebesgue decomposition is unique, so both S and τ_n preserve ν_1, ν_2. This means that ν_1 and ν_2 are both strongly stationary measures. Since μ_1 is extremal we have either $\mu_1 = \nu_1$ or $\mu_1 = \nu_2$. If $\mu_1 = \nu_1$ then μ_1 and μ_2 are mutually singular. So it remains to show that $\mu_1 \neq \nu_2$. Suppose on the contrary that $\mu_1 = \nu_2$. Then $\mu_1 \ll \mu_2$, so $\mu_1 = \int f d\mu_2$ for some $f \in L^1(\mu_2)$. The choice of f is unique, so we conclude as before that f is S, τ_n-invariant (with respect to μ_2). Since f is nonconstant ($\mu_1 \neq \mu_2$) there exists a S, τ_n-invariant set A such that $0 < \mu_2(A) < 1$. Then μ_2 is a nontrivial convex combination of the induced strongly stationary measures on A and A^c. This contradicts the extremality of μ_2 and completes the proof. □

Proposition 4.4. *A strongly stationary measure σ is extremal if and only if the joint action of S and the τ_n's is ergodic.*

Proof. Suppose that the joint action is not ergodic, that is, there exists a S, τ_n-invariant set A with $0 < \sigma(A) < 1$. Then σ is a nontrivial convex combination of the induced strongly stationary measures on A and A^c. Hence σ is not extremal.

Conversely, suppose that the joint action is ergodic. Let $\sigma = a\sigma_1 + (1-a)\sigma_2$, for some strongly stationary measures σ_1, σ_2. Then σ_1 is absolutely continuous with respect to σ and the corresponding Radon-Nikodym derivative $d\sigma/d\sigma_1$ is a S, τ_n-invariant function. Since the joint action is ergodic, $d\sigma/d\sigma_1$ is constant μ-a.e.. Hence $\sigma = \sigma_1$. This proves that σ is extremal. □

Remarks. (i) It follows that a strongly stationary system $(\mathbf{X}, \mathcal{F})$ is extremal if and only if the joint action of T and the τ_n's (of Proposition 3.5) is ergodic.

(ii) Using this proposition we can easily check that the examples on $\mathbb{T}^d$ given in Section 3.4 are extremal.

4.3. **Homogeneity property.** In this section we will show that the ergodic components of an extremal strongly stationary system enjoy a homogeneity property, in the sense that their structure is similar.

Lemma 4.5. *Let $(\mathbf{X}, \mathcal{F})$ be a strongly stationary system. Then τ_n leaves the sub-σ-algebra of T-invariant sets invariant.*

Proof. Let A be a T-invariant set, that is $T^{-1}A = A$. The commutation relations of Proposition 3.5 give that

$$\tau_n^{-1}A = \tau_n^{-1}T^{-1}A = T^{-n}\tau_n^{-1}A.$$

Hence, $\tau_n^{-1}A$ is left invariant by T^n. By Theorem 3.7 the transformation T does not have nontrivial roots of unity as eigenvalues. It follows that $\tau_n^{-1}A$ is also T-invariant, completing the proof. $\qquad\square$

Lemma 4.6. *Let $(\mathbf{X}, \mathcal{F})$ be an extremal strongly stationary system. If $A \in \mathcal{I}$ has positive measure then $\bigcup_{n \in \mathbb{N}} \tau_n^{-1}(A)$ has full measure.*

Proof. Let $B = \bigcup_{n \in \mathbb{N}} \tau_n^{-1}(A)$. In view of Proposition 4.4 it suffices to show that B is T, τ_n-invariant. Since $\tau_{mn} = \tau_m \tau_n$ we have

$$\tau_m^{-1}(B) = \bigcup_{n \in \mathbb{N}} \tau_m^{-1} \tau_n^{-1}(A) = \bigcup_{n \in \mathbb{N}} \tau_{mn}^{-1}(A) \subset B.$$

So the set B is τ_m-invariant for every $m \in \mathbb{N}$. Moreover, since $T^{-1}B$ is equal to $\bigcup_{n \in \mathbb{N}} T^{-1} \tau_n^{-1}(A)$ and by Lemma 4.5 the set $\tau_n^{-1}(A)$ is T-invariant, B is also T-invariant. $\qquad\square$

Definition 4.7. Let $\mathbf{X}$ be a measure preserving system with ergodic decomposition $\mu = \int \mu_t \, d\lambda(t)$. We say that the sets $A, B \in \mathcal{I}$ with positive λ-measure are *factor power equivalent* (FPE), if for λ-a.e. $b \in B$ there exists $a \in A$ and $n \in \mathbb{N}$ such that $(X, \mathcal{B}, \mu_b, T)$ is a factor of $(X, \mathcal{B}, \mu_a, T^n)$, and vice versa.

We are now ready to prove the advertised homogeneity property.

Theorem 4.8. *Let $(\mathbf{X}, \mathcal{F})$ be an extremal strongly stationary system with ergodic decomposition $\mu = \int \mu_t \, d\lambda(t)$. Then any two sets $A, B \in \mathcal{I}$ with positive measure are FPE.*

Proof. By Lemma 4.5 each τ_n permutes the ergodic fibers. Suppose that τ_n maps the fiber X_a to the fiber X_b ($b = \tau_n a$). The pointwise commutation relations $T\tau_n = \tau_n T^n$ show that $(X, \mathcal{B}, \mu_b, T)$ is a factor of $(X, \mathcal{B}, \mu_a, T^n)$. By Lemma 4.6 if $A \in \mathcal{I}$ has positive measure then $\bigcup_{n \in \mathbb{N}} \tau_n^{-1}(A)$ has full measure. It follows that A is FPE to X. So any two sets $A, B \in \mathcal{I}$ with positive measure are FPE. $\qquad\square$

We call a property "nice" if it is preserved by factors and powers of measure preserving systems. The homogeneity property just established allows us to extend "nice" properties from a nontrivial set of ergodic components to λ-a.e. ergodic component.

As an application, suppose that a nontrivial set (of positive λ measure) of ergodic components of an extremal strongly stationary system $(X, \mathcal{F})$ is weak mixing . We claim that it is a Bernoulli system. Indeed, weak mixing is a "nice" property, so Theorem 4.8 gives that λ-a.e. ergodic component of the system is weak mixing. Strong stationarity gives

$$\int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{T}^n f_1 \, \cdots \, \mathbf{T}^{kn} f_k \, d\mu,$$

for all k, $N \in \mathbb{N}$, $f_i \in \mathcal{F}$. Let $\mu = \int \mu_t \, d\lambda(t)$ be the ergodic decomposition of μ. Letting $N \to \infty$ and using the multiple weak mixing theorem ([Fu81], page 86) we get

$$\int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \int \left(\int f_0 \, d\mu_t \int f_1 \, d\mu_t \, \cdots \int f_k \, d\mu_t \right) \, d\lambda(t).$$

It follows that the system is an integral of Bernoulli systems and being extremal it must be Bernoulli.

5. Reduction to distal systems

5.1. **Characteristic factors.** The notion of a characteristic factor was introduced by Furstenberg in order to facilitate the study of nonconventional ergodic averages. The idea is to find the smallest factor of a system that completely determines the limit behavior of these averages and then work with this simpler system.

Definition 5.1. A factor $\mathbf{Y} = (Y, \mathcal{E}, \nu, T)$ of a system $\mathbf{X}$ is *characteristic for k terms* if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \left(\prod_{i=1}^{k} \mathbf{T}^{in} f_i - \prod_{i=1}^{k} \mathbf{T}^{in} \mathbb{E}(f_i | \mathcal{E}) \right) = 0$$

in $L^2(\mathbf{X})$ for $f_i \in L^\infty(\mathbf{X})$.

Furstenberg ([Fu77]) proved for ergodic systems that the k-step distal factor is characteristic for $k+1$ terms. We want to use this result for general systems (not necessarily ergodic), so for completeness we include a proof that covers the general case.

Theorem 5.2 (Furstenberg). *Let $\mathbf{X}$ be a measure preserving system. Then the factor $\mathbf{D}_{k-1}$ is characteristic for k-terms.*

Proof. We use induction on k. For $k = 1$ this is the context of the L^2-ergodic theorem. Assume that the statement is valid for k, we will establish it for $k+1$. It suffices to show that if one of the f_i's is orthogonal to $\mathcal{D}_k$ then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k+1} \mathbf{T}^{in} f_i = 0$$

in $L^2(\mathbf{X})$. Indeed, add and subtract $\mathbb{E}(f_i | \mathcal{D}_k)$ to every f_i and expand the product. All the terms but the two that we are interested will converge to zero giving us the desired identity.

So suppose that $\mathbb{E}(f_1 | \mathcal{D}_k) = 0$ (the argument is similar if $\mathbb{E}(f_i | \mathcal{D}_k) = 0$ for $i \neq 1$). We apply Van der Corput's lemma (2.5) on the Hilbert space $L^2(\mathbf{X})$ with $a_n = \prod_{i=1}^{k+1} T^{in} f_i$. In order to show that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a_n = 0$$

in $L^2(\mathbf{X})$ it suffices to establish that

$$\overline{\lim}_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} b_m = 0, \tag{2}$$

where

$$b_m = \overline{\lim}_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} < a_{n+m}, a_n > \right| = \overline{\lim}_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=0}^{k} \mathbf{T}^{in} (\mathbf{T}^{(i+1)m} f_i \bar{f}_i) \, d\mu \right|.$$

By the induction hypothesis the last limit is equal to

$$\overline{\lim}_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} \int \prod_{i=0}^{k} \mathbf{T}^{in} \mathbb{E}(\mathbf{T}^{(i+1)m} f_i \bar{f}_i | \mathcal{D}_{k-1}) \, d\mu \right|.$$

Now we make use of the fact that $\mathbb{E}(f_1 | \mathcal{D}_k) = 0$. It follows from Lemma 2.2 that the function $g(x_1, x_2) = f_1(x_1) \bar{f}_1(x_2)$ is orthogonal to the space of invariant functions of $\mathbf{X} \times_{\mathbf{D_{k-1}}} \mathbf{X}$. If $S = T \times T$, applying the L^2-ergodic theorem for the system $\mathbf{X} \times_{\mathbf{D_{k-1}}} \mathbf{X}$ we get

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \int \mathbf{S}^m g \, \bar{g} \, d\mu' = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \int \left| \mathbb{E}(\mathbf{T}^m f_1 \bar{f}_1 | \mathcal{D}_{k-1}) \right|^2 d\mu = 0. \tag{3}$$

Since every f_i is bounded the Cauchy-Schwartz inequality gives

$$\left| \int \prod_{i=0}^{k} \mathbf{T}^{in} \mathbb{E}(\mathbf{T}^{(i+1)m} f_i \bar{f}_i | \mathcal{D}_{k-1}) \, d\mu \right| \le L \int \left| \mathbb{E}(\mathbf{T}^m f_1 \bar{f}_1 | \mathcal{D}_{k-1}) \right|^2 d\mu$$

for some number L that is independent of n. Hence,

$$b_m \le L \int \left| \mathbb{E}(\mathbf{T}^m f_1 \bar{f}_1 | \mathcal{D}_{k-1}) \right|^2 d\mu.$$

From this and (3) it follows that the limit in (2) is 0. This completes the induction. $\square$

Corollary 5.3. *The distal factor of a system is a characteristic factor for k-terms for every $k \in \mathbb{N}$.*

5.2. **Relative Bernoulli extensions.** The notion of a relative Bernoulli extension was introduced in [Th75b].

Definition 5.4. Let $\mathbf{X}$ be an ergodic system and $\mathbf{Y}$ be a factor of $\mathbf{X}$. Then $\mathbf{X}$ is a *relative Bernoulli* extension of $\mathbf{Y}$ if $\mathbf{X}$ is isomorphic to the direct product of a Bernoulli system and $\mathbf{Y}$.

Let $(X, \mathcal{B}, \mu)$ be a measure space. Two *finite* $\mathcal{B}$-measurable partitions $\mathcal{P} = \{P_i\}_{i=1}^{k}$ and $\mathcal{Q} = \{Q_i\}_{i=1}^{l}$ of X are *ε-independent* if

$$\sum_{i,j} |\mu(P_i \cap Q_j) - \mu(P_i)\mu(Q_j)| \le \varepsilon.$$

Let $\mathbf{X}$ be an invertible and ergodic system, $\mathbf{Y}$ be a factor of $\mathbf{X}$, and $\mu = \int \mu_y d\nu(y)$ be the disintegration of μ over $\mathbf{Y}$. A sequence of finite partitions $\{\mathcal{P}_i\}_{i\in\mathbb{Z}}$ is *weak Bernoulli relative to* $\mathbf{Y}$ if for ν-a.e. y the following is true: for given $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for every $m \geq 1$

$$\bigvee_{-m}^{0} \mathcal{P}_i \quad \text{is } \varepsilon\text{-independent of} \quad \bigvee_{N}^{N+m} \mathcal{P}_i.$$

with respect to μ_y.

The following theorem is a consequence of the results of the articles [Th75a] (Propositions 3, 4, and 5) and [Th75b] (Lemma 6). One can also deduce this from Theorem 2 in [Ki84].

Theorem 5.5 (Thouvenot). *Let $\mathbf{X}$ be an invertible, ergodic system, and $\mathbf{Y}$ be a factor of $\mathbf{X}$. Suppose that for some finite T-generating partition $\mathcal{P}$ the sequence of partitions $\{T^i\mathcal{P}\}_{i\in\mathbb{Z}}$ is weak Bernoulli relative to $\mathbf{Y}$. Then $\mathbf{X}$ is a relative Bernoulli extension of $\mathbf{Y}$.*

Note that the relative notion of weak Bernoulli is a stronger property than the relative notion of very weak Bernoulli that was used in [Th75b].

5.3. The relative Bernoulli property.

Lemma 5.6. *Let $(\mathbf{X}, \mathcal{F})$ be a strongly stationary system and $\{\tau_n\}_{n\in\mathbb{N}}$ be the maps defined in Proposition 3.5. Then the spaces $L^2(\mathbf{D}_k)$ and $L^2(\mathbf{D})$ are τ_n-invariant for every $n \in \mathbb{N}$.*

Proof. Let $n \in \mathbb{N}$. By Lemma 4.5 the subspace $L^2(\mathbf{D}_0)$ is τ_n-invariant. From the definition of $\mathbf{D}_k$ and $\mathbf{D}$ it suffices to show that every maximal compact extension of a τ_n-invariant space is also τ_n-invariant. So suppose that $\mathbf{X}$ is a maximal compact extension of $\mathbf{Y}$ and that $L^2(\mathbf{Y})$ is τ_n-invariant. Consider the disintegration $\mu = \int \mu_y d\nu(y)$ of μ over $\mathbf{Y}$. It suffices to show that if f is compact relative to $\mathbf{Y}$ then so is $\tau_n f$. Let $\varepsilon > 0$. There exists a finite set of functions $g_1, \ldots, g_m$ such that

$$\min_{1 \leq s \leq m} \|\mathbf{T}^i f - g_s\|_{L^2(\mu_y)} < \varepsilon$$

for every $i \in \mathbb{N}$, and ν-a.e. $y \in Y$. Write $i = i'n + r$, for some $i' \in \mathbb{N}$ and $0 \leq r \leq n-1$. Using the commutation relations of Proposition 3.5 we get

$$f(\tau_n T^i x) = f(T^{i'} \tau_n T^r x).$$

So

$$\|\mathbf{T}^i(\boldsymbol{\tau_n} f) - \mathbf{T}^r(\boldsymbol{\tau_n} g)\|_{L^2(\mu_y)} = \|f(T^{i'}\tau_n T^r x) - g(\tau_n T^r x)\|_{L^2(\mu_y)} = \|\mathbf{T}^{i'} f - g\|_{L^2(\mu_{y'})},$$

where $y' = \tau_n T^r y$. The last equality is valid since $L^2(\mathbf{Y})$ is invariant under both τ_n and T. It follows that the set of functions $\{\mathbf{T}^r(\boldsymbol{\tau_n} g_s), \ 1 \leq s \leq m, \ 0 \leq r \leq n-1\}$ is fiberwise

a finite ε-net relative to $\mathbf{Y}$ for the orbit $\{\mathbf{T}^i(\boldsymbol{\tau_n}f)\}_{i\in\mathbb{N}}$. This shows that the function $\boldsymbol{\tau_n}f$ is compact relative to $\mathbf{Y}$ and completes the proof. $\square$

Remark. Since we are only going to use the τ_n-invariance of $L^2(\mathbf{D})$ we can avoid the use of Lemma 4.5. Indeed, the distal factor can be exhausted by a sequence (possibly transfinite) of maximal isometric extensions starting from the trivial factor (determined by the algebra of constant functions). We can then use the step by step argument of the previous proof to show that $L^2(\mathbf{D})$ is τ_n-invariant.

Theorem 5.7. *Let $(\mathbf{X}, \mathcal{F})$ be a strongly stationary system. Then its distal factor is strongly stationary and almost every ergodic component of $\mathbf{X}$ is a relative Bernoulli extension of its distal factor.*

Proof. **Step 1.** Let $\mathbf{D} = (D, \mathcal{D}, \nu, T)$ be the distal factor of $\mathbf{X}$. Strong stationarity gives

$$\int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \int f_0 \, \mathbf{T}^n f_1 \, \cdots \, \mathbf{T}^{kn} f_k \, d\mu$$

for all $k, n \in \mathbb{N}$, $f_i \in \mathcal{F}$. Averaging over n and taking the limit as $N \to \infty$ gives

$$\int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{T}^n f_1 \, \cdots \, \mathbf{T}^{kn} f_k \, d\mu.$$

By Theorem 5.2 the last average is equal to

$$(4) \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \mathbb{E}(f_0|\mathcal{D}) \, \mathbf{T}^n \mathbb{E}(f_1|\mathcal{D}) \, \cdots \, \mathbf{T}^{kn} \mathbb{E}(f_k|\mathcal{D}) \, d\nu.$$

The maps $\{\tau_n\}_{n\in\mathbb{N}}$ leave the functions in $\mathcal{F}$ invariant, as well as the space $L^2(\mathbf{D})$ (by Lemma 5.6), hence

$$(5) \qquad \boldsymbol{\tau_n}\mathbb{E}(f_i|\mathcal{D}) = \mathbb{E}(\boldsymbol{\tau_n}f_i|\mathcal{D}) = \mathbb{E}(f_i|\mathcal{D}).$$

Since $\boldsymbol{\tau_n}\mathbf{T} = \mathbf{T}^n\boldsymbol{\tau_n}$ and τ_n is measure preserving for $n \in \mathbb{N}$, we get using (5) that all the integrals in (4) are equal to

$$\int \mathbb{E}(f_0|\mathcal{D}) \, \mathbf{T}\mathbb{E}(f_1|\mathcal{D}) \, \cdots \, \mathbf{T}^k\mathbb{E}(f_k|\mathcal{D}) \, d\nu.$$

This shows that

$$(6) \qquad \int f_0 \, \mathbf{T} f_1 \, \cdots \, \mathbf{T}^k f_k \, d\mu = \int \mathbb{E}(f_0|\mathcal{D}) \, \mathbf{T}\mathbb{E}(f_1|\mathcal{D}) \, \cdots \, \mathbf{T}^k\mathbb{E}(f_k|\mathcal{D}) \, d\nu$$

for $k \in \mathbb{N}$, $f_i \in \mathcal{F}$.

Step 2. We will strengthen (6) to a fiberwise relation and prove the first claim. Call $\mathcal{F}_{\mathcal{D}}$ the algebra generated by functions of the form $\mathbb{E}(f|\mathcal{D})$, where $f \in \mathcal{F}$. Let $\mathcal{D}'$ be the sub-σ-algebra of $\bigvee_{i=0}^{\infty} \mathbf{T}^i \mathcal{F}_{\mathcal{D}}$-measurable sets. Clearly $\mathcal{D}' \subset \mathcal{D}$. Since $\mathcal{D}'$ is T-invariant

it induces a factor $\mathbf{D}' = (D', \mathcal{D}', \nu', T)$ of $\mathbf{X}$. Applying τ_n to the left hand side of the equation below and using the previous averaging technique we get as before that

$$
(7) \qquad \int \prod_{i=0}^{k} \mathbf{T}^i f_i \ \prod_{j=0}^{m} \mathbf{T}^j g_j \ d\mu = \int \prod_{i=0}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D}) \ \prod_{j=0}^{m} \mathbf{T}^j g_j \ d\nu
$$

for $k, m \in \mathbb{N}$, $f_i \in \mathcal{F}$, $g_i \in \mathcal{F}_\mathcal{D}$. Observe that $\mathbb{E}(f|\mathcal{D})$ is $\mathcal{D}'$-measurable for $f \in \mathcal{F}$, so $\mathcal{D}$ can be replaced by $\mathcal{D}'$ in (7). Moreover, all the g_i's are $\mathcal{D}'$-measurable, so (7) takes the following form

$$
(8) \qquad \int \left[\mathbb{E}\Big(\prod_{i=0}^{k} \mathbf{T}^i f_i | \mathcal{D}' \Big) - \prod_{i=0}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D}') \right] \prod_{j=0}^{m} \mathbf{T}^j g_j \ d\nu' = 0.
$$

Since $\bigvee_{i=0}^{\infty} T^i \mathcal{F}_\mathcal{D}$ is dense in $L^2(\mathbf{D}')$ we get

$$
\int \left[\mathbb{E}\Big(\prod_{i=0}^{k} \mathbf{T}^i f_i | \mathcal{D}' \Big) - \prod_{i=0}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D}') \right] g \ d\nu' = 0
$$

for every $g \in L^2(\mathbf{D}')$. This can only happen if

$$
(9) \qquad \mathbb{E}\Big(\prod_{i=0}^{k} \mathbf{T}^i f_i | \mathcal{D}' \Big)(y) = \prod_{i=0}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D}')(y)
$$

for ν'-a.e. $y \in D'$.

Next we claim that $\mathcal{D}' = \mathcal{D}$. Relation (9) easily implies that $\mathbf{X}$ is a relative weak mixing extension of $\mathbf{D}'$. Since $\mathbf{D}$ is the minimal factor with respect to which $\mathbf{X}$ is relative weak mixing, $\mathcal{D}$ must be contained in $\mathcal{D}'$. Thus, $\mathcal{D}' = \mathcal{D}$ and (9) takes the form

$$
(10) \qquad \mathbb{E}\Big(\prod_{i=0}^{k} \mathbf{T}^i f_i | \mathcal{D} \Big)(y) = \prod_{i=0}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D})(y)
$$

for ν-a.e. $y \in D$.

Since every $f \in \mathcal{F}_\mathcal{D}$ is τ_n-invariant and $\mathcal{F}_\mathcal{D}$ is a T-generating algebra for $\mathcal{D}$, the system $\mathbf{D}$ is strongly stationary.

Step 3. We will now prove the second claim. First assume that the sub-σ-algebra $\mathcal{B}(\mathcal{F})$ is determined by a finite partition $\mathcal{P}$. Then $\mathcal{P}$ is T-generating for almost every ergodic component and relation (10) is valid for almost every ergodic component, provided that we replace $\mathcal{D}$ with the distal factor of the corresponding ergodic component. To simplify the notation we assume that $\mathbf{X}$ is ergodic, and we keep in mind that any result we get will be valid for almost every ergodic component.

We claim that the sequence of partitions $\{T^i\mathcal{P}\}_{i\in\mathbb{Z}}$ satisfies the conditions of Theorem 5.5. To see this observe first that if we replace k with $2k$ in (10) and then apply $\mathbf{T}^{-k}$ we

get

$$(11) \qquad \mathbb{E}\Big(\prod_{i=-k}^{k} \mathbf{T}^i f_i | \mathcal{D} \Big)(y) = \prod_{i=-k}^{k} \mathbf{T}^i \mathbb{E}(f_i|\mathcal{D})(y)$$

for every $k \in \mathbb{N}$, ν-a.e. $y \in Y$, and $\mathcal{P}$-measurable functions f_i. Hence,

$$\int fg \, d\mu_y = \int f \, d\mu_y \int g \, d\mu_y$$

for every $k \in \mathbb{N}$, and ν-a.e. y, whenever f is $\bigvee_{-k}^{0} T^i \mathcal{P}$-measurable and g is $\bigvee_{1}^{k} T^i \mathcal{P}$-measurable. So $\{T^i \mathcal{P}\}_{i \in \mathbb{Z}}$ is weak Bernoulli with respect to μ_y for ν-a.e. y. Theorem 5.5 now implies that the system is a relative Bernoulli extension of its distal factor. Hence, almost every ergodic component is isomorphic to the direct product of its distal factor and a Bernoulli system.

In general, since $L^2(\mathbf{X})$ is separable there exists a sequence $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ of finite $\mathcal{F}$-measurable partitions such that $\mathcal{P}_{n+1}$ refines $\mathcal{P}_n$ and $\bigvee_{n=1}^{\infty} \mathcal{P}_n = \mathcal{B}(\mathcal{F})$. Applying the previous argument for the factor that is T-generated by $\mathcal{P}_n$ we get a factor $\mathbf{X}_n$ which is isomorphic to the direct product of a distal system and a Bernoulli system. Since increasing unions of distal systems is distal and of Bernoulli systems Bernoulli ([Or74], page 52), it follows that almost every ergodic component of $\mathbf{X}$ is isomorphic to the direct product of a distal system and a Bernoulli system. This completes the proof. $\square$

5.4. **The distal factor coincides with the $\mathcal{C}$-factor.** Let $\mathbf{X}$ be a measure preserving system. For fixed $k \in \mathbb{N}$ consider the closed subalgebra generated by weak limits (in $L^2(\mathbf{X})$) of the form

$$(12) \qquad \mathrm{wlim}_{l \to \infty} \frac{1}{N_l} \sum_{n=1}^{N_l} \mathbf{T}^{a_1 n} f_1 \, \mathbf{T}^{a_2 n} f_2 \cdots \mathbf{T}^{a_k n} f_k,$$

where we are free to choose any $a_1, \ldots, a_k \in \mathbb{N}$, functions $f_1, \ldots, f_k \in L^{\infty}(\mathbf{X})$, and an increasing sequence $\{N_l\}_{l \in \mathbb{N}}$ of positive integers that guarantees weak convergence. This algebra is conjugation closed and T-invariant so it gives rise to a factor $\mathbf{C}_k$. If in addition we are free to choose any $k \in \mathbb{N}$ we get a factor that extends every $\mathbf{C}_k$. We denote it by $\mathbf{C}$ and call it the $\mathcal{C}$-*factor* or *characteristic factor* of $\mathbf{X}$. Using the notation of Section 5.1 the factor $\mathbf{C}_k$ is characteristic for k terms and the factor $\mathbf{C}$ is characteristic for any number of terms. Moreover, by Theorem 5.2 we have $\mathbf{C}_k \subset \mathbf{D}_{k-1}$ for every $k \in \mathbb{N}$ and so $\mathbf{C} \subset \mathbf{D}$.

First, we prove that the $\mathcal{C}$-factor is τ_n-invariant for every $n \in \mathbb{N}$.

Lemma 5.8. *Let $(\mathbf{X}, \mathcal{F})$ be a strongly stationary system and $\{\tau_m\}_{m \in \mathbb{N}}$ be the maps defined in Proposition 3.5. Then for all $k, m \in \mathbb{N}$ the subspaces $L^2(\mathbf{C}_k)$ and $L^2(\mathbf{C})$ are τ_m-invariant.*

Proof. Let $k, m \in \mathbb{N}$. Using the commutation relations $\boldsymbol{\tau_m} \mathbf{T}^n = \mathbf{T}^{nm} \boldsymbol{\tau_m}$ (see Proposition 3.5) we see that the operator $\boldsymbol{\tau_m}$ maps functions of the form (12) to functions of the form

$$\operatorname*{wlim}_{l \to \infty} \frac{1}{N_l} \sum_{n=1}^{N_l} \mathbf{T}^{a_1 nm}(\boldsymbol{\tau_m} f_1) \ \mathbf{T}^{a_2 nm}(\boldsymbol{\tau_m} f_2) \cdots \mathbf{T}^{a_k nm}(\boldsymbol{\tau_m} f_k),$$

belongs again to $L^2(\mathbf{C}_k)$. Since functions of the form (12) generate $L^2(\mathbf{C}_k)$ the subspace $L^2(\mathbf{C}_k)$ is τ_m-invariant. A similar argument applies for $L^2(\mathbf{C})$. $\square$

Theorem 5.9. *Every zero entropy (and hence every distal) strongly stationary system coincides with its $\mathcal{C}$-factor.*

Proof. Since $\mathcal{C}$ is T-invariant, and $\mathcal{F}$ is T-generating, it suffices to show that every $f \in \mathcal{F}$ is $\mathcal{C}$-measurable. Equivalently, if $f \in \mathcal{F}$ and $g = f - \mathbb{E}(f|\mathcal{C})$, we need to show that $g = 0$. By Lemma 5.8 the subspace $L^2(\mathbf{C})$ is τ_n-invariant. Moreover, f is τ_n-invariant so g is τ_n-invariant for $n \in \mathbb{N}$. Given functions $f_1, \ldots, f_k \in \mathcal{F}$, there exists an increasing sequence $\{N_l\}_{l \in \mathbb{N}}$ of positive integers such that the sequence

$$\frac{1}{N_l} \sum_{n=1}^{N_l} \mathbf{T}^n f_1 \cdots \mathbf{T}^{kn} f_k$$

converges weakly in $L^2(\mathbf{X})$ as $m \to \infty$. Since $\boldsymbol{\tau_n} \mathbf{T} = \mathbf{T}^n \boldsymbol{\tau_n}$, the functions f_i and g are τ_n-invariant, and τ_n is measure preserving for $n \in \mathbb{N}$ we get

$$\int g \, \mathbf{T} f_1 \cdots \mathbf{T}^k f_k \, d\mu = \lim_{l \to \infty} \frac{1}{N_l} \sum_{n=1}^{N_l} \int g \, \mathbf{T}^n f_1 \cdots \mathbf{T}^{kn} f_k \, d\mu$$

for $k \in \mathbb{N}$, $f_i \in \mathcal{F}$. Since $\mathbb{E}(g|\mathcal{C}) = 0$, we have $\mathbb{E}(g|\mathcal{C}_k) = 0$ and by the definition of $\mathcal{C}_k$ the last average is 0. Thus, g is orthogonal to the closed subalgebra spanned by bounded $\bigvee_1^\infty T^i \mathcal{F}$-measurable functions. Since $\mathbf{X}$ has zero entropy we have $\bigvee_1^\infty T^i \mathcal{F} = \bigvee_{-\infty}^\infty T^i \mathcal{F}$. It follows that g is orthogonal to the full algebra of the system. Hence $g = 0$, proving our claim. $\square$

6. Distal strongly stationary systems

6.1. **Nilrotations.** Let G be a locally compact and separable Lie group. The commutator of two elements $g, h \in G$ is the element $[g, h] = g^{-1} h^{-1} g h$. If $A, B \subset G$ we write $[A, B]$ for the subgroup generated by $\{[a, b] : a \in A, b \in B\}$. The lower central series of G is defined as follows, $G^{(0)} = G$, $G^{(i)} = [G, G^{(i-1)}]$. The group G is *nilpotent of order k* if $G^{(k)} = \{e\}$, where e denotes the identity element of G. With G_0 we denote the connected component of the identity element of G. If Γ is a discrete subgroup (not necessarily normal) of an order k nilpotent group such that G/Γ is compact we call G/Γ an *order k nilmanifold.* The group G acts on G/Γ by left translation $T_a(x\Gamma) = (ax)\Gamma$. There exists a unique probability measure on G/Γ that is invariant under left translations, we

denote it by μ and call it the *Haar measure* on $\mathbf{G}/\Gamma$. If G is nilpotent of order k, we call the system $\mathbf{G}/\Gamma = (G/\Gamma, \mathcal{G}/\Gamma, T_a, \mu)$ an *order k nilsystem* and the transformation T_a an *order k nilrotation.* An inverse limit of (order k) nilsystems (nilmanifolds) is called an (order k) *pro-nilsystem* (*pro-nilmanifold*).

The following generalization of a theorem of Parry ([Pa69]) was proved by Leibman [Lei03]. It was originally established by Lesigne [Les91] under an extra hypothesis. Bergelson and Host ([BH03]) showed that this extra hypothesis can be removed, thus providing another independent proof.

Theorem 6.1 (Leibman). *Let $(G/\Gamma, T_a)$ be a nilsystem and set $Z = G/G^{(1)}\Gamma$. If G is spanned by the connected component G_0 and a then*

(i) The nilsystem $(G/\Gamma, T_a)$ is uniquely ergodic if and only if the factor system (Z, T_a) is ergodic.

(ii) If $(G/\Gamma, T_a)$ is ergodic then its Kronecker factor is (Z, T_a).

Remark. As it was noted in [BH03] if the nilsystem $(G/\Gamma, T_a)$ is ergodic then the projection of $< G_0, a >$ on G/Γ being an open invariant set is equal to G/Γ. Hence, $G/\Gamma = < G_0, a > /(\Gamma \cap < G_0, a >)$. So if T_a is ergodic we can assume that $G = < G_0, a >$.

Examples. (i) The prototypical example of a nonabelian order two ergodic nilsystem is defined on the *Heisenberg* nilmanifold. Let

$$G = \left\{ \begin{pmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix}, \ x_i \in \mathbb{R} \right\}, \ \Gamma = \left\{ \begin{pmatrix} 1 & k_1 & k_3 \\ 0 & 1 & k_2 \\ 0 & 0 & 1 \end{pmatrix}, \ k_i \in \mathbb{Z} \right\}, \ a = \begin{pmatrix} 1 & a_1 & a_3 \\ 0 & 1 & a_2 \\ 0 & 0 & 1 \end{pmatrix},$$

where $a_1, a_2 \in \mathbb{R}$ are rationally independent and $a_3 \in \mathbb{R}$. Then G with the standard metric is locally compact and connected. Moreover, if the group action is matrix multiplication then G is nilpotent of order two and G/Γ is compact. So T_a defines a nilsystem on G/Γ. Observe that $G/G^{(1)}\Gamma \simeq \mathbb{T}^2$ and that the rotation on $\mathbb{T}^2$ by (a_1, a_2) is ergodic. It follows from Theorem 6.1 that $(G/\Gamma, T_a)$ is uniquely ergodic and its Kronecker factor is induced by the functions on x_1, x_2.

(ii) Let

$$G = \left\{ \begin{pmatrix} 1 & k & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix}, k \in \mathbb{Z}, x_i \in \mathbb{R} \right\}, \ \Gamma = \left\{ \begin{pmatrix} 1 & k_1 & k_3 \\ 0 & 1 & k_2 \\ 0 & 0 & 1 \end{pmatrix}, k_i \in \mathbb{Z} \right\}, \ a = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix},$$

where b is irrational. Then $(G/\Gamma, T_a)$ is uniquely ergodic and isomorphic to the affine system on $\mathbb{T}^2$ with the Haar measure induced by $T(x, y) = (x + b, y + x)$.

It turns out that every measure preserving system that is induced by some distal affine transformation on compact abelian group with the Haar measure is isomorphic to a nilsystem. But not every nilsystem is isomorphic to an affine system. For example the order two nilsystem of example (i) is not ([Fu90], page 52).

6.2. **Nonconventional ergodic averages.** The following theorem of Host and Kra ([HK03]) is crucial for our study.

Theorem 6.2 (**Host and Kra**). *Let* $\mathbf{X}$ *be an invertible ergodic measure preserving system. Then the averages*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{T}^n f_1 \; \mathbf{T}^{2n} f_2 \; \cdots \; \mathbf{T}^{kn} f_k$$

converge in $L^2(\mathbf{X})$ *for* $f_i \in L^\infty(\mathbf{X})$*. Moreover, the characteristic factor (defined in Section 5.4) for these averages is an order* $k-1$ *pro-nilsystem.*

If T is an ergodic order k nilrotation it is possible to find an explicit formula for the limit. This was done for G connected and $k = 3$ by Lesigne ([Les89]) and for general (not necessarily connected) G and k by Ziegler ([Zi03]). To describe the formula it is convenient to first establish some notation.

Let $\mathbf{G}/\Gamma$ be an order l ergodic nilsystem. It turns out ([Lei02]) that for every $k \in \mathbb{N}$ the set
$$H_k = \left\{ \left(x_1, \; x_1^2 x_2, \; \ldots, \; x_1^{\binom{k}{1}} x_2^{\binom{k}{2}} \cdots x_l^{\binom{k}{l}}\right), \; x_i \in G^{(i-1)} \right\}$$
is a closed subgroup of $G \times \cdots \times G$ (the product has k terms) with group action coordinatewise multiplication. If $\Delta_k = H_k \cap (\Gamma \times \cdots \times \Gamma)$, then the quotient H_k/Δ_k is again a nilmanifold and supports a unique left invariant (under translations by elements in H_k) measure that we denote by ν_{H_k}.

Theorem 6.3 (**Ziegler**). *Let* $(G/\Gamma, T_a)$ *be an order* l *ergodic nilsystem (assume that* $G = < G_0, a >$*). If* $f_1, \ldots, f_k \in L^\infty(G/\Gamma)$ *then for almost every* $x \in G$ *we have*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{T}_a^n f_1(x\Gamma) \; \cdots \; \mathbf{T}_a^{kn} f_k(x\Gamma) =$$

$$\int_{H_k/\Delta_k} f_1(xy_1\Gamma) \; \cdots \; f_k(xy_k\Gamma) \; d\nu_{H_k}(y\Delta_k)$$

where $y = (y_1, \ldots, y_k)$*.*

The next identity will enable us later to give a general method for constructing strongly stationary systems starting from totally ergodic measure preserving systems. It is a consequence of Theorems 6.2 and 6.3.

Theorem 6.4. *Let* $\mathbf{X}$ *be an ergodic measure preserving system such that* T^r *is ergodic for some* $r \in \mathbb{N}$*. Then*

$$(13) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \; \mathbf{T}^n f_1 \cdots \mathbf{T}^{kn} f_k \; d\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \; \mathbf{T}^{rn} f_1 \cdots \mathbf{T}^{krn} f_k \; d\mu$$

for $f_i \in L^\infty(\mathbf{X})$*.*

Proof. Assume first that T_a is an order l nilrotation defined on G/Γ. Since T_a^r is ergodic T_a is also ergodic, so Theorem 6.3 gives that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{G/\Gamma} f_0(x\Gamma) \ \mathbf{T}_a^n f_1(x\Gamma) \ \cdots \ \mathbf{T}_a^{kn} f_k(x\Gamma) \ d\mu(x\Gamma) =$$

$$\int_{G/\Gamma} \int_{H_k/\Delta_k} f_0(x\Gamma) f_1(xy_1\Gamma) \ \cdots \ f_k(xy_k\Gamma) \ d\nu_{H_k}(y\Delta_k) \ d\mu(x\Gamma),$$

where $f_i \in L^\infty(\mathbf{N}/\Gamma)$, and H_k, Δ_k, ν_{H_k} are defined as before. Observe that the integral on the right does not depend on the nilrotation T_a on G/Γ as long as it is ergodic. Since $T_{ra} = T_a^r$ is assumed to be ergodic we get that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{G/\Gamma} f_0 \ \mathbf{T}_a^n f_1(x\Gamma) \ \cdots \ \mathbf{T}_a^{kn} f_k(x\Gamma) \ d\mu(x\Gamma) =$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{G/\Gamma} f_0 \ \mathbf{T}_{ra}^n f_1(x\Gamma) \ \cdots \ \mathbf{T}_{ra}^{kn} f_k(x\Gamma) \ d\mu(x\Gamma) =$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{G/\Gamma} f_0 \ \mathbf{T}_a^{rn} f_1(x\Gamma) \ \cdots \ \mathbf{T}_a^{rkn} f_k(x\Gamma) \ d\mu(x\Gamma).$$

So (13) holds when T is a nilrotation. A standard approximation argument shows that (13) holds when T defines a pro-nilsystem. In the general case we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \ \mathbf{T}^n f_1 \ \cdots \ \mathbf{T}^{kn} f_k \ d\mu =$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \mathbb{E}(f_0|\mathcal{C}) \ \mathbf{T}^n \mathbb{E}(f_1|\mathcal{C}) \ \cdots \ \mathbf{T}^{kn} \mathbb{E}(f_k|\mathcal{C}) \ d\mu,$$

where $\mathcal{C}$ is the characteristic factor of the system. It follows from Theorem 6.2 that the factor $\mathcal{C}$ is a pro-nilsystem, so the limit on the right is equal to

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \mathbb{E}(f_0|\mathcal{C}) \ \mathbf{T}^{rn} \mathbb{E}(f_1|\mathcal{C}) \ \cdots \ \mathbf{T}^{rkn} \mathbb{E}(f_k|\mathcal{C}) \ d\mu.$$

Since the factor $\mathcal{C}$ is characteristic for all terms the last limit is equal to

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \ \mathbf{T}^{rn} f_1 \ \cdots \ \mathbf{T}^{rkn} f_k \ d\mu.$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We remark that identity (13) was proved for $k = 3$ by Host and Kra ([HK02]). It was also shown there how it implies an odd version of the ergodic Szemerédi theorem. For general k we get:

Corollary 6.5. *Let* $\mathbf{X}$ *be a measure preserving system. Assume that* T^r *is ergodic for some* $r \in \mathbb{N}$ *and let* A *be a measurable set with* $\mu(A) > 0$. *Then for every* $0 \le j < r$ *we have*

$$\mu\big(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A\big) > 0$$

for some $n \equiv j \pmod{r}$.

The odd Szemerédi theorem corresponds to the case $r = 2$, $j = 1$.

6.3. **The structure of the ergodic components.** Theorem 6.2 combined with Theorems 5.7 and 5.9 enables us to determine the structure of the ergodic components of the general strongly stationary system.

Theorem 6.6. *Almost every ergodic component of a strongly stationary system is isomorphic to the direct product of a Bernoulli system and a totally ergodic pro-nilsystem.*

Proof. Theorem 5.7 shows that almost every ergodic component of a strongly stationary system is the direct product of a Bernoulli system and a distal strongly stationary system. By Theorem 5.9 a distal strongly stationary system coincides with its $\mathcal{C}$-factor, so Theorem 6.2 shows that almost every ergodic component of the distal factor of the system is a pro-nilsystem. Finally, by Proposition 3.7 the system has no rational eigenvalue different than 1, so a nontrivial set of ergodic components cannot share the same rational eigenvalue provided that it is different than 1. It follows that almost every ergodic component of the system is totally ergodic, which completes the proof. $\square$

We will discuss the structure of the global system in more detail in the next section. We will see that there exist distal strongly systems with nonaffine ergodic components. This new class of examples will allow a complete classification.

6.4. **New examples and structure theorem.** Let σ be a stationary measure on the sequence space $I^{\mathbb{Z}}$. We define a new measure σ_{av} on $I^{\mathbb{Z}}$ by averaging the statistics of σ along arithmetic progressions. More precisely if $\mathcal{F}_0$ is the algebra of bounded x_0-measurable functions and S denotes the shift transformation, we define the measure σ_{av} on cylinder sets by

$$(14) \qquad \int f_0\, \mathbf{S}f_1\ \cdots\ \mathbf{S}^k f_k\, d\sigma_{av} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0\, \mathbf{S}^n f_1\ \cdots\ \mathbf{S}^{kn} f_k\, d\sigma,$$

for $f_i \in \mathcal{F}_0$ (the limit exists by Theorem 6.2). We then extend σ_{av} to the whole sequence space.

Theorem 6.7. *Let* σ *be a totally ergodic stationary measure. Then the measure* σ_{av} *is strongly stationary.*

Proof. A direct computation proves stationarity. Since S is totally ergodic by Theorem 6.4 we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{S}^n f_1 \, \cdots \, \mathbf{S}^{kn} f_k \, d\sigma = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{S}^{rn} f_1 \, \cdots \, \mathbf{S}^{rkn} f_k \, d\sigma$$

for all $r, k \in \mathbb{N}$, $f_i \in \mathcal{F}_0$. Hence

$$\int f_0 \, \mathbf{S} f_1 \, \cdots \, \mathbf{S}^k f_k \, d\sigma_{av} = \int f_0 \, \mathbf{S}^r f_1 \, \cdots \, \mathbf{S}^{rk} f_k \, d\sigma_{av}$$

for all $r, k \in \mathbb{N}$, $f_i \in \mathcal{F}_0$. Since the subalgebra $\mathcal{F}_0$ is S-generating the measure σ_{av} is strongly stationary. $\qquad\square$

We remark that under the hypothesis of the previous theorem the strongly stationary measure σ_{av} can be shown to be extremal. We will not use this fact so we omit its proof.

Proposition 6.8. *Let σ be an extremal strongly stationary measure with ergodic decomposition $\sigma = \int \sigma_t \, d\lambda(t)$. Then for λ almost every t we have $\sigma = (\sigma_t)_{av}$.*

Proof. Strong stationarity gives

$$\int f_0 \, \mathbf{S} f_1 \, \cdots \, \mathbf{S}^k f_k \, d\sigma = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{S}^n f_1 \, \cdots \, \mathbf{S}^{kn} f_k \, d\sigma$$

$$= \int \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0 \, \mathbf{S}^n f_1 \, \cdots \, \mathbf{S}^{kn} f_k \, d\sigma_t \right) d\lambda(t)$$

for all $k \in \mathbb{N}$, $f_i \in \mathcal{F}_0$. It follows that

$$\sigma = \int (\sigma_t)_{av} \, d\lambda(t).$$

By Theorem 3.7 almost every ergodic component of S is totally ergodic so Theorem 6.7 gives that the measures $(\sigma_t)_{av}$ are strongly stationary for λ-a.e. t. Since σ is extremal we have that $\sigma = (\sigma_t)_{av}$ for λ-a.e. t. $\qquad\square$

We will use Theorem 6.7 to construct an ample supply of strongly stationary systems. We briefly describe the strategy. Starting with an arbitrary invertible totally ergodic measure preserving system $\mathbf{X}$ we first consider its sequence space representation with respect to $\mathcal{F} = L^\infty(\mathbf{X})$ (Proposition 2.4). This representation is determined by a stationary measure σ on $I^{\mathbb{Z}}$, so $\mathbf{X}$ is isomorphic to the system $(I^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \sigma, S)$ where S is the shift transformation. Let $\phi\colon X \to I^{\mathbb{Z}}$ be the isomorphism ($\phi(\mathcal{F}) = \mathcal{F}_0 = x_0$-measurable functions). We construct the strongly stationary measure σ_{av} as in (14), that is, we define the measure σ_{av} on cylinder sets by

$$(15) \qquad \int f_0 \, \mathbf{S} f_1 \, \cdots \, \mathbf{S}^k f_k \, d\sigma_{av} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int f_0' \, \mathbf{T}^n f_1' \, \cdots \, \mathbf{T}^{kn} f_k' \, d\mu,$$

where $f_i \in \mathcal{F}_0$ and $f_i' = f_i \circ \phi \in \mathcal{F}$. Finally, we give an explicit description of the statistics of σ_{av}. This way we recover all the examples mentioned in Section 3.4 and we also construct some new ones.

Examples. (i) Suppose that $\mathbf{X}$ is a weak mixing system. Using the multiple weak mixing theorem ([Fu81], page 86) we can check that the resulting strongly stationary measure defined by (15) is a Bernoulli measure.

(ii) Suppose that $\mathbf{X}$ is the system induced by an ergodic rotation on $\mathbb{T}$ with the Haar measure m. If we compute the limit in (15) we find that

$$\int f_0 \, \mathbf{S} f_1 \, \cdots \, \mathbf{S}^k f_k \, d\sigma_{av} = \int_{\mathbb{T}^2} f_0'(y) \, f_1'(y+x) \, \cdots \, f_k'(y+kx) \, dm(y) \, dm(x).$$

We can check that σ_{av} determines the sequence space representation of the strongly stationary system $T(x,y) = (x, y+x)$ on $\mathbb{T}^2$ with the Haar measure m (with respect to the algebra generated by the exponentials in y).

(iii) Suppose that $\mathbf{X}$ is the system induced by an affine transformation on $\mathbb{T}^2$ with the Haar measure defined by $T(x,y) = (x+a, y+x)$, where a is irrational. We can check that the resulting strongly stationary measure defined by (15) determines the sequence space representation of the system $T'(x,y,z) = (x, y+x, z+y)$ on $\mathbb{T}^3$ with the Haar measure (with respect to the algebra generated by the exponentials in z).

(iv) Suppose that $\mathbf{X}$ is an order l totally ergodic nilsystem defined on $X = G/\Gamma$. Using Theorem 6.3 we see that the resulting strongly stationary measure σ_{av} constructed by (15) is defined on cylinder sets as follows

$$(16) \qquad \int f_0 \, \mathbf{S} f_1 \, \cdots \, \mathbf{S}^k f_k \, d\sigma_{av} = \int_{G/\Gamma} \int_{H_k/\Delta_k} f_0'(x\Gamma) f_1'(xy_1\Gamma) \, \cdots \, f_k'(xy_k\Gamma) \, d\nu_{H_k} d\mu,$$

where H_k, Δ_k, ν_{H_k} are defined as in Theorem 6.3. Then the system $(I^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \sigma_{av}, S)$ is strongly stationary with respect to $\mathcal{F}_0$.

(v) Suppose that $\mathbf{X}$ is a totally ergodic pro-nilsystem defined on the inverse limit X of the nilmanifolds X_i (X has to be connected since it supports a totally ergodic pro-nilsystem). Then the resulting strongly stationary measure σ_{av} constructed by (15) is defined on $\phi(X_i)$ by (16). Since $I^{\mathbb{Z}} = \bigcup_i \phi(X_i)$ this uniquely determines the measure σ_{av} on $I^{\mathbb{Z}}$. Observe that the resulting strongly stationary system $(I^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \sigma_{av}, S)$ depends only on the pro-nilmanifold X. We call it *the strongly stationary system associated to the pro-nilmanifold X*.

We remark that in example (iv) if X is the Heisenberg nilmanifold (see Section 6.1 example (i)) then the ergodic components of the strongly stationary system obtained are nilrotations on X, and hence nonaffine ([Fu90], page 52).

This new set of examples enables us to completely determine the structure of the general strongly stationary system. This is the context of our main theorem:

Theorem 6.9. *Every extremal strongly stationary system* $(\mathbf{X}, \mathcal{F})$ *is isomorphic to the direct product of a Bernoulli system and a strongly stationary system associated to some pro-nilmanifold.*

Proof. Suppose that the ergodic components of $\mathbf{X}$ are the systems $\mathbf{X}_t$. Let σ, σ_t be the measures that determine the sequence space representations of $\mathbf{X}$ and $\mathbf{X}_t$ with respect to $L^\infty(\mathbf{X})$ and $L^\infty(\mathbf{X}_t)$. Then $\sigma = \int \sigma_t \, d\lambda$ is the ergodic decomposition of σ. Since σ is an extremal strongly stationary measure by Proposition 6.8 we have $\sigma = (\sigma_t)_{av}$ for λ-a.e. t. Theorem 6.6 implies that such a measure σ_t has the form $\rho \times \tau$ where ρ and τ determine the sequence space representation of a Bernoulli system and a totally ergodic pro-nilsystem $\mathbf{N}$ correspondingly. Then $\sigma = (\rho \times \tau)_{av} = \rho_{av} \times \tau_{av}$. The measure ρ_{av} induces a Bernoulli system $\mathbf{B}$ (example (i)) and the measure τ_{av} induces a strongly stationary system $\mathbf{P}$ associated to the pro-nilmanifold N (example (v)). Hence, $\mathbf{X}$ is isomorphic to the direct product $\mathbf{B} \times \mathbf{P}$ and the result follows. $\qquad\square$

References

[BH03] V. Bergelson, B. Host. *A Khinchine-Roth Theorem.* Preprint (2003).

[BL96] V. Bergelson, A. Leibman. *Polynomial extensions of van der Waerden's and Szemerédi's theorems.* J. Amer. Math. Soc. 9, (1996), no. 3, 725-753.

[Br92] L. Breiman. *Probability.* Society for Industrial & Applied Mathematics, Reprint edition, (1992).

[Co85] J. B. Conway. *A course in functional analysis.* Springer-Verlag, New York, (1985).

[Fu77] H. Furstenberg. *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions.* J. d'Analyse Math., 71, (1977), 204-256.

[Fu81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, (1981).

[Fu90] H. Furstenberg. *Nonconventional ergodic averages.* Proceedings of Symposia in Pure Mathematics, Volume 50 (1990).

[FK79] H. Furstenberg, Y. Katznelson. *An ergodic Szemerédi theorem for commuting transformations.* J. Analyse Math., 34, (1979), 275-291.

[FK91] H. Furstenberg, Y. Katznelson. *A density version of the Hales-Jewett theorem.* J. d'Analyse Math., 57, (1991), 64-119.

[FW96] H. Furstenberg, B. Weiss. *A mean ergodic theorem for* $(1/N) \sum_{n=1}^{N} f(T^n x) \, g(T^{n^2} x)$. Convergence in ergodic theory and probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ. 5, de Gruyter, Berlin, (1996), 193-227.

[HK02] B. Host, B. Kra. *An odd Furstenberg-Szemerédi Theorem and quasi-affine systems.* J. d'Anal. Math., 86, (2002), 183-220.

[HK03] B. Host, B. Kra. *Nonconventional ergodic averages and nilmanifolds.* To appear. Annals of Math. Available at http://www.math.psu.edu/kra/publications.html

[Je97] E. Jenvey. *Strong stationarity and De Finetti's theorem.* J. d'Analyse Math., 73, (1997), 1-18.

[Ki84] J. Kieffer. *A simple development of Thouvenot relative isomorphism theory.* Annals of Probability, 12, (1984), 204-211.

[Lei02] A. Leibman. *Polynomial mappings of groups.* Israel Journal of Matematics, 129, (2002), 29-60.

[Lei03] A. Leibman. *Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold.* Preprint (2002). Available at http://www.math.ohio-state.edu/leibman/preprints.

[Les89] E. Lesigne. *Théorèmes ergodiques pour une translation sur un nilvariété.* Ergodic Theory Dynam. Systems, 9, (1989), no. 1, 115-126.

[Les91] E. Lesigne. *Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodique.* Ergodic Theory Dynam. Systems, 11, (1991), no. 2, 379-391.

[Or74] D. Ornstein. *Ergodic theory, randomness, and dynamical systems.* Yale Mathematical Monographs, No. 5, Yale University Press, New Haven, Conn.-London, (1974).

[Pa69] W. Parry. *Ergodic properties of affine transformations and flows on nilmanifolds.* Amer. J. Math., 91, (1969), 757-771.

[Pe89] K. Petersen. *Ergodic theory.* Cambridge University Press, Cambridge, (1989).

[Ph01] R. Phelps. *Lectures on Choquet's theorem.* Second Edition, Lecture Notes in Mathematics, 1757, Spinger-Verlag, Berlin, (2001).

[Ro62] V. A. Rokhlin. *On the fundamental ideas of measure theory.* Translations of the American Mathematical Society (Series 1), 10, (1962), 1-54.

[Th75a] J. P. Thouvenot. *Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli.* Israel J. Math., 21, no 2-3, (1975), 177-207.

[Th75b] J. P. Thouvenot. *Remarques sur les systèmes dynamiques donnés avec plusieurs facteurs.* Israel J. Math., 21, no 2-3, (1975), 215–232.

[Wa82] P. Walters. *An introduction to ergodic theory.* Graduate Texts in Mathematics, 79, Springer-Verlag, New York-Berlin, (1982).

[Zi03] T. Ziegler. *A nonconventional ergodic theorem for a nilsystem.* To appear in Ergodic Theory Dyn. Systems. Available at http://www.math.ohio-state.edu/∼tamar

Department of Mathematics, McAllister Building, Pennsylvania State University, University Park, PA 16802

E-mail address: nikos@math.psu.edu