Thomas Ward

Two constructions in ergodic theory

March 14, 2017

Supplementary material to
"Ergodic theory with a view towards number theory"

These notes describe two constructions of measure-preserving transformations, one probabilistic and one geometric. The notes are a supplement to the book [5]. The material is all standard, and in particular may largely be found in the monograph of Cornfeld, Fomin and Sinaĭ [4]. It is assembled here to provide a convenient source.

Contents

Chapter 1
Constructions from Probability Theory

Examples of measure-preserving transformations (and, more generally, of group actions that preserve a probability measure) are of central importance in ergodic theory. Any continuous map of a compact metric space gives rise to at least one measure-preserving map (and possibly to many measure-preserving maps). Many of the applications in number theory arise from a specific type of construction of measure-preserving group actions built from lattices in Lie groups. Here we give a brief introduction to some rather different constructions which are also of great importance in ergodic theory.

The close connection between probability and ergodic theory is largely concerned with the theory of entropy; here we indicate the connection between measure-preserving systems and stationary stochastic processes, and go on to describe the important class of *Gaussian* dynamical systems[1]. Some terminology from probability will be used, and in particular the conditional expectation is used in Section 1.3.2.

1.1 Stationary Processes and Measure-Preserving Maps

Let $(\Omega, \mathcal{A}, P)$ be a probability space (that is, P is a measure on the σ-algebra $\mathcal{A}$ of subsets of Ω, and $P(\Omega) = 1$) and let $(f_n)_{n\in\mathbb{Z}}$ be a sequence of real-valued elements of $L^2_P(\Omega)$ with the following properties: For any finite set $F = \{n_1, n_2, \dots, n_r\} \subset \mathbb{Z}$, the probability measure P_F defined by

$$P_F(A) = P\left(\{\omega \in \Omega \mid (f_1(\omega), \dots, f_r(\omega)) \in A\}\right)$$

[1] We follow the treatment of Cornfeld, Fomin and Sinaǐ [4] closely; other convenient sources include the lecture notes of Totoki [35] and the notes of Neveu [29] and Maruyama [26]. The decomposition in Section 1.3.1 and part of the spectral analysis of Gaussian systems is due to Itô [16], [17]; further spectral analysis was carried out by Fomin [9], [10]. Theorem 2 was proved by Girsanov [12] and Theorem 3 by Foiaş and Strătilă [8]. Veršik [38], [37] showed that Gaussian systems with countable Lebesgue spectrum are all measurably isomorphic.

for any Borel set $A \subset \mathbb{R}^r$ only depends on the set F up to translation, in the sense that $P_F = P_{F+n}$ for any $n \in \mathbb{Z}$. In particular, the measure $P_{\{n\}}$ is independent of $n \in \mathbb{Z}$. The functions $f_n \in L^2_P(\Omega)$ are called *random variables*, and any such sequence $(f_n)_{n\in\mathbb{Z}}$ is called a *stationary stochastic process*. For any random variable $f \in L^2_P(\Omega)$, write $\mathbb{E}(f) = \mathbb{E}_P(f) = \int f \,\mathrm{d}P$ for the *expectation* or *mean* of f. A stationary stochastic process $(f_n)_{n\in\mathbb{Z}}$ is called *centered* if the expectation $\mathbb{E}(f_0) = 0$, and the function $c : n \mapsto \mathbb{E}(f_0 f_n)$ is the *covariance function* of the process.

Definition 1 (Kolmogorov representation). Let $(f_n)_{n\in\mathbb{Z}}$ be a stationary stochastic process on the probability space $(\Omega, \mathcal{A}, P)$. The associated measure-preserving system is $(X, \mathcal{B}, \mu, T)$, where

- $X = \mathbb{R}^{\mathbb{Z}}$ and $\mathcal{B}$ is the Borel σ-algebra defined by the product topology on X;
- for any finite set $F \subset \mathbb{Z}$, μ is the probability measure on X uniquely determined[2] by the property that $\mu(\{x \in X \mid \pi_F(x) \in A\}) = P_F(A)$ for any Borel set $A \subset \mathbb{R}^F$, where $\pi_F : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^F$ denotes the projection onto the coordinates in F; and
- the map T is the left shift on X.

Notice that the stochastic process can be recovered from the associated dynamical system, since $f_n = \pi_{\{n\}}$ by construction. More generally, if $(X, \mathcal{B}, \mu, T)$ is any measure-preserving system, and $f \in L^2_\mu$ then the functions defined by $f_n = U_T^n f$ form a stationary stochastic process.

The measure-preserving system constructed in Definition 1 from the stationary stochastic process defined by a function f in L^2_μ on a measure-preserving system $(X, \mathcal{B}, \mu, T)$ is a factor of $(X, \mathcal{B}, \mu, T)$ (see [5,]).

1.2 Gaussian Dynamical Systems

Gaussian probability distributions arise naturally in probability and statistics, and in particular arise via the central limit theorem whenever independent identically distributed random variables are averaged. In this section we show how Gaussian measures may be used to construct a measure-preserving system from a suitable measure on the circle. Just as in the simple case of compact group automorphisms (see [5,]), we are able to understand ergodic properties of Gaussian automorphisms by studying a suitable decomposition of the space of square-integrable functions under the action of the associated isometry.

Recall that a real $r \times r$ matrix C is called *positive-definite* if $\mathbf{x}^{\mathrm{t}} C \mathbf{x} \geqslant 0$ for any $\mathbf{x} \in \mathbb{R}^r$, with equality only for $\mathbf{x} = 0$.

[2] This property determines μ by the Kolmogorov consistency theorem [5,].

Definition 2. Let $(X, \mathcal{B}, \mu)$ be a probability space. A list $(x_1, \dots, x_r)$ of real-valued random variables on $(X, \mathcal{B}, \mu)$ has an r-dimensional *Gaussian distribution* if there is a positive-definite $r \times r$ matrix C (the covariance matrix) with the property that

$$\mu\left(\{x \in X \mid x_1 \in C_1, \dots, x_r \in C_r\}\right)$$
$$= \frac{1}{(2\pi)^{r/2}\sqrt{\det(C)}} \int_{C_1} \cdots \int_{C_r} \mathrm{e}^{-\frac{1}{2}(\mathbf{s}-\overline{\mathbf{x}})^{\mathrm{t}} C^{-1}(\mathbf{s}-\overline{\mathbf{x}})}\, \mathrm{d}s_1 \dots \mathrm{d}s_r,$$

where $\overline{\mathbf{x}} = \int_X \mathbf{x}\, \mathrm{d}\mu = (\overline{x}_1, \dots, \overline{x}_r)$ is the mean of the list of random variables.

Notice that an r-dimensional Gaussian distribution is completely determined by the means $\overline{x}_i$ and the covariances

$$c(x_i, x_j) = \mathbb{E}(x_i x_j) = \int_X x_i x_j \, \mathrm{d}\mu$$

for $1 \leqslant i, j \leqslant r$.

Definition 3. Let $X = \mathbb{R}^{\mathbb{Z}}$ and let $\mathcal{B}$ be the Borel σ-algebra on X. A probability measure μ defined on $\mathcal{B}$ is called *Gaussian* if the joint distribution of $(x_{n_1}, \dots, x_{n_r})$ is an r-dimensional Gaussian distribution for any $r \geqslant 1$ and integers $n_1 < \cdots < n_r$.

The parameters defining the r-dimensional Gaussian distribution for the random variables $(x_{n_1}, \dots, x_{n_r})$ for all r determine μ by the Kolmogorov consistency theorem. The measure μ is called *stationary* (invariant under the shift map) if $\overline{x}_n = m$ is independent of $n \in \mathbb{Z}$ and

$$c(x_{n_1}, x_{n_2}) = c(x_{n_1+n}, x_{n_2+n})$$

for all $n_1, n_2, n \in \mathbb{Z}$. In this case we define the *correlation function* to be

$$c(n) = c(x_0, x_n).$$

Since the map $(x_n)_{n \in \mathbb{Z}} \mapsto (x_n - m)_{n \in \mathbb{Z}}$ transfers a Gaussian measure to a Gaussian measure with zero mean, we assume from now on that the mean m is zero. Notice that the correlation function is symmetric, since

$$\begin{aligned} c(-n) &= c(x_0, x_{-n}) \\ &= c(x_n, x_0) \\ &= \mathbb{E}(x_n x_0) \\ &= c(x_0, x_n) = c(n) \end{aligned}$$

for any $n \in \mathbb{Z}$. Small values of the correlation function correspond to near-independence of the random variables, and for this reason Gaussian processes are useful in the study of various kinds of asymptotic independence for the measurable dynamical system given by the shift map T on X.

Example 1. Two simple extreme examples of Gaussian behavior are independence and identity.

1. Define a measure on $\mathbb{R}$ by setting

$$\mu_0(A) = \frac{1}{\sqrt{2\pi}} \int_A \mathrm{e}^{-\frac{1}{2}s^2}\,\mathrm{d}s$$

and then let $\mu = \prod_{\mathbb{Z}} \mu_0$ be the infinite product measure. This measure is invariant under the left shift map, and defines a Gaussian dynamical system with correlation function

$$c(n) = \begin{cases} 1 & \text{if } n = 0; \\ 0 & \text{if not.} \end{cases}$$

This is an example of a Bernoulli shift on an infinite alphabet.

2. Define μ_0 as above, and now define μ on $\mathbb{R}^{\mathbb{Z}}$ by

$$\mu\left(\{x \in X \mid x_1 \in C_1, \dots, x_r \in C_r\}\right) = \mu_0\left(C_1 \cap \cdots \cap C_r\right).$$

This defines a Gaussian dynamical system with $c(n) = 1$ for all n, and is measurably the same as the identity map.

Definition 4. A function $c : \mathbb{Z} \to \mathbb{C}$ is *positive-definite* if, for any set

$$\{n_1, \dots, n_r\} \subset \mathbb{Z}$$

and complex numbers $a_1, \dots, a_r$ we have

$$\sum_{j=1}^{r} \sum_{i=1}^{r} a_j \overline{a_i} c(n_j - n_i) \geqslant 0.$$

Since the covariance matrix C associated to a Gauss measure is positive-definite, the correlation function c is positive-definite in the sense of Definition 4, and is symmetric. It follows by the Herglotz–Bochner theorem that there is a unique finite measure ρ on $\mathbb{T}$ with the property that

$$c(n) = \int_0^1 \mathrm{e}^{2\pi \mathrm{i} n s}\,\mathrm{d}\rho(s) \tag{1.1}$$

for all $n \in \mathbb{Z}$. The measure ρ is called the *spectral measure* of the Gaussian measure μ, and ρ determines μ (since we have assumed that the mean m is zero). The symmetry property $c(n) = c(-n)$ implies that $\rho(B) = \rho(-B)$ for any Borel set $B \subset \mathbb{T}$.

Example 2. The spectral measure in Example 1(1) is the Lebesgue measure on the circle, and the relation in equation (1.1) is simply the familiar orthogonality of characters. The spectral measure associated with Example 1(2) is δ_0, the point mass at zero.

We will see later that these crude examples, in which an atomic spectral measure is associated to a non-ergodic measure-preserving system while the spectral measure for the infinite Bernoulli shift is non-atomic, reflect more general results (see Theorems 1 and 4).

Lemma 1. *There is a one-to-one correspondence between correlation functions of stationary Gaussian measures μ with zero mean, and symmetric positive-definite functions $c : \mathbb{Z} \to \mathbb{R}$.*

PROOF. We have seen above that the correlation function of such a Gaussian measure is positive-definite. Assume therefore that $c : \mathbb{Z} \to \mathbb{R}$ is positive-definite; we wish to exhibit a unique corresponding Gaussian measure. Since c is symmetric and positive-definite, the $r \times r$ matrix C defined by

$$C_{ij} = c(j - i)$$

is symmetric and positive-definite for any $r \geqslant 1$. It follows that there is an orthogonal matrix O for which

$$OCO^{\mathrm{t}} = \operatorname{diag}(\lambda_1, \dots, \lambda_r)$$

for some $\lambda_i \geqslant 0$, $1 \leqslant i \leqslant r$. Let $R = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ and $A = (a_{kj}) = RO$, so that by construction $A^{\mathrm{t}}A = C$. Let $\mathbf{z} = (z_1, \dots, z_r)$ be a list of independent real random variables, each with a probability density function

$$\tfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}.$$

Now define real random variables $\mathbf{x} = (x_1, \dots, x_r)$ by $\mathbf{x} = \mathbf{z}A$. Then, since $z_1, \dots, z_r$ are independent, we have

$$\begin{aligned}
\mathbb{E}\left(\mathrm{e}^{\mathrm{i}\langle \mathbf{u}, \mathbf{x}\rangle}\right) &= \mathbb{E}\left(\mathrm{e}^{\mathrm{i}\sum_k z_k \left(\sum_j u_j a_{kj}\right)}\right) \\
&= \mathrm{e}^{-\frac{1}{2}\sum_k \left(\sum_j u_j a_{kj}\right)^2} \\
&= \mathrm{e}^{-\frac{1}{2}\mathbf{u}^{\mathrm{t}} A^{\mathrm{t}} A \mathbf{u}} = \mathrm{e}^{-\frac{1}{2}\mathbf{u}^{\mathrm{t}} C \mathbf{u}}.
\end{aligned}$$

Thus $\mathbf{x}$ has a joint Gaussian distribution with covariance matrix C, so we have constructed a measure μ_F on $\mathbb{R}^F$ where $F = \{n_1, \dots, n_r\}$. Now if

$$F = \{n_1, \dots, n_r\} \subset \{n_1, \dots, n_s\} = E,$$

it is clear that the projection $\mathbb{R}^E \to \mathbb{R}^F$ sends μ_E to μ_F, so the Kolmogorov consistency theorem gives a Gaussian measure μ on $\mathbb{R}^{\mathbb{Z}}$ as required. $\square$

Thus the distribution of any finite set of coordinates in a Gaussian process is determined by finitely many parameters, and this simple structure allows many explicit constructions to be made.

Now let $(X,\mathcal{B},\mu,T)$ be a Gaussian system, and let $H_1^{(\mathbb{R})}$ be the closure of the space spanned by $\{x_n \mid n\in\mathbb{Z}\}$ in L^2_μ; clearly $H_1^{(\mathbb{R})}$ is a U_T-invariant closed real subspace of L^2_μ.

Lemma 2. *1. Any function $f\in H_1^{(\mathbb{R})}$ has a Gaussian probability distribution.*

2. There is an isomorphism $\theta^{(\mathbb{R})}$ from the subspace of real-valued functions ϕ in L^2_ρ with $\phi(t)=\phi(-t)$ to $H_1^{(\mathbb{R})}$, with the property that

$$U_T\left(\ \theta^{(\mathbb{R})}\phi(\cdot)\ \right)=\theta^{(\mathbb{R})}\left(\mathrm{e}^{2\pi\mathrm{i}\cdot}\phi(\cdot)\right). \tag{1.2}$$

Proof. If a sequence of random variables $f_1,f_2,\dots$ converges to f in L^2_μ, then the corresponding probability density functions converge weakly, and hence the values on characteristic functions converge uniformly on any bounded interval. Now characteristic functions of Gaussian distributions converge only to characteristic functions of Gauss distributions, proving (1).

Associate the random variable $f=\sum_{k=1}^r a_k x_{n_k}$ to the function

$$\phi(t)=\sum_{k=1}^r a_k \mathrm{e}^{\mathrm{i}n_k t}$$

for any finite set $\{n_1,\dots,n_r\}\subset\mathbb{Z}$. Clearly $\phi(t)=\overline{\phi(-t)}$, and

$$\mathbb{E}(f^2)=\int|\phi(t)|^2\,\mathrm{d}\rho(t).$$

Thus this defines a one-to-one linear isometry $\theta^{(\mathbb{R})}$ on symmetric trigonometric polynomials in $L^2_\rho(\mathbb{T})$ with equation (1.2), and this extends to the symmetric functions in $L^2_\rho(\mathbb{T})$ by continuity. □

Extend $\theta^{(\mathbb{R})}$ to all of L^2_ρ by writing $H_1=H_1^{(\mathbb{R})}+\mathrm{i}H_1^{(\mathbb{R})}$ and associating to each $f=f_1+\mathrm{i}f_2\in H_1$ the function $\phi_1+\mathrm{i}\phi_2$, where $\phi_i=(\theta^{(\mathbb{R})})^{-1}f_i$ for $i=1,2$. This defines a map θ from L^2_ρ to H_1.

Lemma 3. *The map θ is an isomorphism between L^2_ρ and H_1; under this isomorphism $\theta\left(\mathrm{e}^{\mathrm{i}\cdot}\phi(\cdot)\right)=U_T\theta\left(\phi(\cdot)\right)$.*

Proof. A function $\phi\in L^2_\rho(\mathbb{T})$ may be decomposed as $\phi=\phi_1+\mathrm{i}\phi_2$, where

$$\phi_1(t)=\tfrac{1}{2}[\phi(t)+\overline{\phi(-t)}]$$

and

$$\phi_2(t) = \tfrac{1}{2\mathrm{i}}[\phi(t) - \overline{\phi(-t)}],$$

so that $\phi_1(t) = \overline{\phi_1(-t)}$ and $\phi_2(t) = \overline{\phi_2(-t)}$. The lemma follows by Lemma 2. □

This gives our first result relating properties of the spectral measure ρ to properties of the Gaussian dynamical system.

Theorem 1. *If a Gaussian measure-preserving system is ergodic, then its associated spectral measure is non-atomic.*

PROOF. Suppose that there is some $t_0 \in \mathbb{T}$ for which $\rho(\{t_0\}) = \rho(\{-t_0\}) > 0$. Define a function $\phi \in L^2_\rho(\mathbb{T})$ by

$$\phi(t) = \begin{cases} 1 & \text{if } t = t_0; \\ 0 & \text{if } t \neq t_0. \end{cases}$$

Then, by Lemma 2, $f = \theta(\phi)$ is a non-zero complex random variable whose real and imaginary parts satisfy a non-trivial two-dimensional Gaussian distribution. Now

$$\theta^{-1}U_T f = \mathrm{e}^{2\pi\mathrm{i}(\cdot)}\phi(\cdot) = \mathrm{e}^{2\pi\mathrm{i}t_0}\phi(\cdot) = \mathrm{e}^{2\pi\mathrm{i}t_0}\theta^{-1}f,$$

so $U_T|f| = |f|$, and the function $|f|$ is invariant under T. Since f satisfies a non-trivial two-dimensional Gaussian distribution, $|f|$ is not almost everywhere equal to a constant, so T is not ergodic. □

In order to show the converse of Theorem 1, we will need a more sophisticated decomposition of L^2_μ, and this is provided by Itô's theory of multiple stochastic integrals. Part of this important theory constructs a decomposition of L^2_μ into an orthogonal direct sum of subspaces H_m in such a way that each H_m is isomorphic as a Hilbert space to the subspace of $L^2_{\rho^m}(\mathbb{T}^m)$ (where ρ^m denotes the measure $\rho \times \cdots \times \rho$ on $\mathbb{T}^m$) comprising the functions which are even in each variable and symmetric under permutation of the variables. Moreover, under this isomorphism, U_T corresponds to multiplication by the function $\mathbf{t} \mapsto \mathrm{e}^{2\pi\mathrm{i}(t_1+\cdots+t_m)}$ (Proposition 1). Once this decomposition is available, characterizing the basic ergodic properties of Gaussian dynamical systems is relatively straightforward (see Section 1.3.2). While there are many convenient sources for this material, including the monograph of Cornfeld, Fomin and Sinaĭ [4] and the lecture notes of Totoki [35], it is less familiar than the decomposition of $L^2_{m_X}(X)$ functions on a compact abelian group X afforded by harmonic analysis (see [5, Sect. C.3]), and so is included here for completeness.

The construction of the spaces H_m and the isomorphisms is lengthy, and will take up much of Section 1.3.1. The construction will use the geometry of Hilbert space to construct suitable spaces of functions, and averaging over actions of the symmetric group to obtain the required symmetry properties.

1.2.1 Characterizing Gaussian Automorphisms

A more general notion (giving rise to a more abstract approach to Gaussian measure-preserving systems) is the following.

Definition 5. Let $(X, \mathcal{B}, \mu, T)$ be an invertible measure-preserving system. A function $h_0 \in L^2_\mu$ with $\int h_0 \,\mathrm{d}\mu = 0$ is called a *Gaussian element* if for any $r \geqslant 1$ and $n_1 < \cdots < n_r$, the random variables $h_{n_1}, \dots, h_{n_r}$, where $h_n = U_T^n h_0$, satisfy a joint Gaussian distribution, meaning that for sets $C_1, \dots, C_r \in \mathcal{B}_\mathbb{R}$,

$$\mu\left(\{x \in X \mid h_{n_1}(x) \in C_1, \dots, h_{n_r}(x) \in C_r\}\right) = \alpha \int_{C_1} \cdots \int_{C_r} \mathrm{e}^{-\frac{1}{2}\langle A\mathbf{s}, \mathbf{s}\rangle} \,\mathrm{d}s_1 \dots \mathrm{d}s_r,$$

where A^{-1} is the covariance matrix defined by

$$A^{-1} = \left(\langle h_{n_i}, h_{n_j}\rangle\right)_{1 \leqslant i,j \leqslant r},$$

and

$$\alpha^{-1} = \int_\mathbb{R} \cdots \int_\mathbb{R} \mathrm{e}^{-\frac{1}{2}\langle A\mathbf{s}, \mathbf{s}\rangle} \,\mathrm{d}s_1 \dots \mathrm{d}s_r.$$

Definition 6. A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is called a Gaussian automorphism if there is a Gaussian element $h_0 \in L^2_\mu$ with $\int h_0 \,\mathrm{d}\mu = 0$ with the property that the smallest σ-algebra containing all the sets

$$\{x \in X \mid h_n(x) \in C\}$$

for some $n \in \mathbb{Z}$ and $C \in \mathcal{B}_\mathbb{R}$ is all of $\mathcal{B}$.

1.3 Spectral Analysis of Gaussian Processes

From now on, we assume that $(X, \mathcal{B}, \mu, T)$ is a Gaussian measure-preserving system whose spectral measure ρ is non-atomic. In order to describe the mixing properties of T in terms of ρ we need to construct a suitable decomposition of L^2_μ. Once this is done we are able to use the decomposition in much the same way as characters are used to relate ergodic and mixing properties of a group endomorphism to algebraic properties of the dual homomorphism.

1.3.1 Hermite–Itô Polynomials

Let $f : X \to \mathbb{R}$ be a random variable on the probability space $(X, \mathcal{B}, \mu)$ with

$$\mu\left(\{x \in X \mid f(x) \in (\alpha, \beta)\}\right) = \frac{1}{\sqrt{2\pi b}} \int_\alpha^\beta \mathrm{e}^{-t^2/2b} \,\mathrm{d}t,$$

where $b = \int f^2 \, \mathrm{d}\mu$. That is, f has a Gaussian distribution with mean zero and variance b.

Definition 7. Let f be a random variable with Gaussian distribution and zero mean. Then the *mth Hermite–Itô polynomial*[3] is the random variable

$$\mathbf{:}f^m\mathbf{:} = f^m + a_1 f^{m-1} + \cdots + a_{m-1} f + a_m$$

where the coefficients are determined by the $(m-1)$ linear relations given by requiring that

$$\int \mathbf{:}f^m\mathbf{:}\; f^p \, \mathrm{d}\mu = 0$$

for $p = 0, 1, \ldots, m-1$.

Notice that $\mathbf{:}f^m\mathbf{:} \in L^2_\mu$ for any $m \geqslant 1$, so we may think of these polynomials using the geometry of the Hilbert space L^2_μ. In particular, $\mathbf{:}f^m\mathbf{:}$ differs from f^m by the projection onto the space spanned by $\{f^p \mid 0 \leqslant p \leqslant m-1\}$ of the vector f^m, as illustrated in Figure 1.1. It follows that $\langle \mathbf{:}f^m\mathbf{:}, \mathbf{:}f^n\mathbf{:} \rangle = 0$ for $m \neq n$.

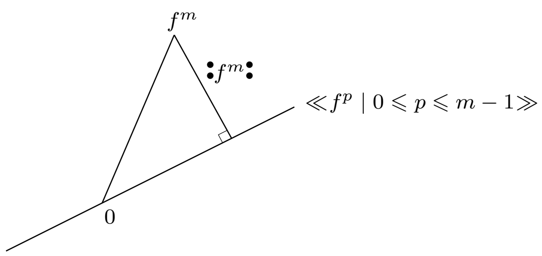

Fig. 1.1: The mth Hermite–Itô polynomial $\mathbf{:}f^m\mathbf{:}$.

The notion of Hermite–Itô polynomials extends to m-tuples of random variables as follows.

Definition 8. Let $f_1, \ldots, f_m$ be Gaussian random variables in $H_1^{(\mathbb{R})}$. The Hermite–Itô polynomial $\mathbf{:}f_1, \ldots, f_m\mathbf{:}$ is the orthogonal projection of the product $f_1 \cdots f_m$ onto the subspace spanned by

$$\{f_1' \cdots f_p' \mid 0 \leqslant p \leqslant m-1, f_i' \in H_1^{(\mathbb{R})}\}.$$

Also define $\overset{\circ}{\circ}\, f_1, \ldots, f_m \,\overset{\circ}{\circ}$ to be the orthogonal projection of the product $f_1 \cdots f_m$ onto the subspace spanned by

[3] The notation $\mathbf{:}f^m\mathbf{:}$ for the mth Hermite–Itô polynomial of f (which only makes sense formally if the parameters f and m are kept separate) comes from physics, and is used in the monograph of Cornfeld, Fomin and Sinaĭ [4].

$$\{f_{i_1}\cdots f_{i_p} \mid 1 \leqslant i_1,\dots,i_p \leqslant m, p<m\}.$$

Thus for any $p<m$ and $f_i' \in H_1^{(\mathbb{R})}$ we have

$$\int {:}f_1,\dots,f_m{:}\; f_1'\cdots f_p' \,\mathrm{d}\mu = 0, \tag{1.3}$$

by definition.

Lemma 4. ${:}f_1,\dots,f_m{:} = {\scriptstyle\circ\atop\circ} f_1,\dots,f_m {\scriptstyle\circ\atop\circ}$.

It follows that the random variables $f_1',\dots,f_p'$ used in Definition 8 may be chosen from the original list of random variables $f_1,\dots,f_m$.

Proof of Lemma 4. Any random variable $f' \in H_1^{(\mathbb{R})}$ has a unique decomposition $f' = \sum_{k=1}^m c_k f_k + g$, where $\int g f_k \,\mathrm{d}\mu = 0$ for $1 \leqslant k \leqslant m$. Writing in this way $f_j' = \sum_{k=1}^m c_{jk} f_k + g_j$ for each j, the product $f_1'\cdots f_p'$ may be written as a sum of expressions $F_\ell(f_1,\dots,f_m)G_{p-\ell}(g_1,\dots,g_m)$, where F_ℓ is a polynomial of degree ℓ and $G_{p-\ell}$ is a polynomial of degree $(p-\ell)$. By the orthogonality of the g_i from the f_i we have

$$\int {\scriptstyle\circ\atop\circ} f_1,\dots,f_m {\scriptstyle\circ\atop\circ}\; F_\ell(f_1,\dots,f_m)G_{p-\ell}(g_1,\dots,g_m)\,\mathrm{d}\mu$$

$$= \int {\scriptstyle\circ\atop\circ} f_1,\dots,f_m {\scriptstyle\circ\atop\circ}\; F_\ell(f_1,\dots,f_m)\,\mathrm{d}\mu \int G_{p-\ell}(g_1,\dots,g_m)\,\mathrm{d}\mu = 0$$

by the construction of ${\scriptstyle\circ\atop\circ} f_1,\dots,f_m {\scriptstyle\circ\atop\circ}$, which proves the lemma. □

It follows in particular that if $f_1,\dots,f_m$ are pairwise orthogonal, then

$${:}f_1,\dots,f_m{:} = f_1\cdots f_m. \tag{1.4}$$

Definition 9. Let $H_m^{(\mathbb{R})}$ (respectively H_m) be the L^2_μ-closure of the real linear span (complex linear span, respectively) of the sets of linear combinations of functions ${:}f_1,\dots,f_m{:}$ with $f_i \in H_1^{(\mathbb{R})}$. Let $Q_m^{(\mathbb{R})} \subset L^2_{\rho^m}(\mathbb{T}^m)$ be the Hilbert space of functions $\phi : \mathbb{T}^m \to \mathbb{C}$ with the following properties:

1. ϕ is symmetric with respect to the m variables;
2. $\phi(-t_1,\dots,-t_m) = \overline{\phi(t_1,\dots,t_m)}$; and
3. $\|\phi\|_2 = \left[\int_{\mathbb{T}^m} |\phi(\mathbf{t})|^2 \,\mathrm{d}\rho^m(\mathbf{t})\right]^{1/2} < \infty$.

It is clear that $\langle H_m, H_n\rangle = 0$ for $m \neq n$, $H_m^{(\mathbb{R})} \subset H_m$, $U_T H^{(\mathbb{R})} = H_m^{(\mathbb{R})}$, and $U_T H_m = H_m$ for any $m \geqslant 1$.

Theorem 2. *For any $m \geqslant 1$, there is an isometry $\theta_m^{(\mathbb{R})} : Q_m^{(\mathbb{R})} \to H_m^{(\mathbb{R})}$ with the following properties:*

1. $\theta_m^{(\mathbb{R})} Q_m^{(\mathbb{R})} = H_m^{(\mathbb{R})}$, and

2. the isomorphism $(\theta_m^{(\mathbb{R})})^{-1} : H_m^{(\mathbb{R})} \to Q_m^{(\mathbb{R})}$ *sends the operator* U_T *to multiplication by the function* $\mathbf{t} \mapsto \mathrm{e}^{2\pi \mathrm{i}(t_1+\cdots+t_m)}$.

Proof. This is a lengthy proof, and in several places we will make use of the shorthand $f(t)$ or $f(\cdot)$ for the function $t \mapsto f(t)$.

Let $\theta_1^{(\mathbb{R})} = \theta^{(\mathbb{R})}$ be the map constructed in Lemma 2, which satisfies the stated properties for $m = 1$. Now fix $m > 1$ and let Σ_m denote the symmetric group on the symbols $\{1, \ldots, m\}$. Given $\phi \in L^2_{\rho^m}(\mathbb{T}^m)$, define the *symmetrization* $\Sigma_m[\phi]$ to be the function defined by

$$\Sigma_m[\phi](t_1, \ldots, t_m) = \sum_{\pi \in \Sigma_m} \phi(t_{\pi(1)}, \ldots, t_{\pi(m)}).$$

The group $\Sigma_m \times \{\pm 1\}$ acts on $\mathbb{T}^m$ by

$$(t_1, \ldots, t_m) \overset{(\pi,\pm 1)}{\longmapsto} (\pm t_{\pi(1)}, \ldots, \pm t_{\pi(m)}).$$

We claim that the set

$$D = \{\mathbf{t} \in \mathbb{T}^m \mid t_1 + \cdots + t_m > m/2, t_1 > t_2 > \cdots > t_m\}$$

is a fundamental domain for the action of $\Sigma_m \times \{\pm 1\}$ in the following sense:

$$\rho_m(\partial D) = 0,$$

and for any point $\mathbf{t} \in \mathbb{T}^m$ whose orbit under the action does not meet ∂D, the orbit meets D in a single point. To see this, notice that if $\mathbf{t}$ is a point with $t_i \neq t_j$ for $i \neq j$, then there is some $\pi \in \Sigma_m$ with $t_{\pi(1)} > \cdots > t_{\pi(m)}$. Clearly either $\sum_i t_{\pi(i)} > m/2$ or $\sum_i -t_{\pi(i)} > m/2$ (where we are writing elements of $\mathbb{T}$ as real numbers in $[0, 1)$ and summing them as real numbers), so under the action of ± 1 we can satisfy the first condition defining D. Finally, the boundary ∂D is a union of finitely many hyperplanes, and the ρ^m-measure of each of these is zero since ρ is non-atomic by assumption.

We now wish to define $\theta_m^{(\mathbb{R})}$ on $Q_m^{(\mathbb{R})}$. Let $D_1, \ldots, D_m$ be Borel subsets of $[0, 1)$ with

$$(D_i \cup (-D_i)) \cap (D_j \cup (-D_j)) = \emptyset \tag{1.5}$$

for $i \neq j$ and with $D_1 \times \cdots \times D_m \subset D$. Associate to such a list of sets the functions

$$\phi_i^+ = \chi_{D_i \cup (-D_i)} \tag{1.6}$$

and

$$\phi_i^- = \mathrm{i}\left(\chi_{D_i} - \chi_{(-D_i)}\right) \tag{1.7}$$

for $1 \leqslant i \leqslant m$. For any $\mathbf{e} = (e_1, \ldots, e_m) \in \{\pm 1\}^m$, define a function $\phi \in Q_m^{(\mathbb{R})}$ by

$$\phi(\mathbf{t}) = \sqrt{\frac{1}{m!}}\Sigma_m \left[\prod_{i=1}^{m} \phi_i^{e_i}(t_i)\right]. \tag{1.8}$$

For any Borel set $B \subset [0,1)$ write

$$\Phi_B = \theta_1^{(\mathbb{R})}(\chi_B(\cdot)),$$

$$\Phi_B^+ = \Phi_{B\cup(-B)},$$

and

$$\Phi_B^- = \mathrm{i}\left(\Phi_B - \Phi_{(-B)}\right).$$

By construction, the functions $\Phi_{D_i}^+$ and $\Phi_{D_i}^-$ lie in $H_1^{(\mathbb{R})}$, and in particular are Gaussian random variables with

$$\left\langle \Phi_{D_i}^+, \Phi_{D_i}^- \right\rangle = \int \phi_i^+ \phi_i^- \,\mathrm{d}\rho(t) = 0,$$

so $\Phi_{D_i}^+$ and $\Phi_{D_i}^-$ are independent. For the function $\phi \in Q_m^{(\mathbb{R})}$ in equation (1.8), define

$$\theta_m^{(\mathbb{R})}(\phi) = \prod_{i=1}^{m} \Phi_{D_i}^{e_i}; \tag{1.9}$$

this then determines $\theta_m^{(\mathbb{R})}$ on all such functions. Notice that the expression in equation (1.9) is well-defined, since the presentation of the function ϕ in the form of equation (1.8) is itself unique.

We claim that $\theta_m^{(\mathbb{R})}(\phi) \in H_m^{(\mathbb{R})}$. From the assumption in equation (1.5), it follows that distinct terms in the right-hand side of equation (1.9) are independent. Thus by equation (1.4) we have

$$:\theta_m^{(\mathbb{R})}(\phi): = \theta_m^{(\mathbb{R})},$$

so that $\theta_m^{(\mathbb{R})}(\phi) \in H_m^{(\mathbb{R})}$. Moreover,

$$\begin{aligned}\|\phi\|_2^2 &= \frac{1}{m!}\left\|\Sigma_m\left[\prod_{i=1}^{m}\phi_i^{e_i}(\cdot)\right]\right\|_2^2 \\ &= \left\|\prod_{i=1}^{m}\phi_i^{e_i}(\cdot)\right\|_2^2 = 2^m\left\|\prod_{i=1}^{m}\phi_i(\cdot)\right\|_2^2 = 2^m\prod_{i=1}^{m}\rho(D_i),\end{aligned}$$

while the independence of the functions $\Phi_{D_i}^{\pm}$ for distinct i shows that

$$\left\|\theta_m^{(\mathbb{R})}(\phi)\right\|_2^2 = \left\|\prod_{i=1}^{m}\Phi_{D_i}^{e_i}\right\|_2^2 = \prod_{i=1}^{m}\|\phi_i^{e_i}(\cdot)\|_2^2 = 2^m\prod_{i=1}^{m}\rho(D_i),$$

so $\theta_m^{(\mathbb{R})}$ acts as an isometry on the set of functions of the form in equation (1.8).

We now extend the domain of $\theta_m^{(\mathbb{R})}$ in successive stages. Given functions

$$\phi_1, \ldots, \phi_k$$

of the form in equation (1.8) with disjoint supports, and constants $a_1, \ldots, a_k \in \mathbb{R}$, write $\phi(\mathbf{t}) = \sum_{i=1}^k a_i \phi_i(\mathbf{t})$, and write A for the set of all functions obtained in this way. The disjoint supports ensure that the representation of a function in A in this form is unique, so the map $\theta_m^{(\mathbb{R})}$ extends uniquely to A by setting

$$\theta_m^{(\mathbb{R})}(\phi) = \sum_{i=1}^k a_i \theta_m^{(\mathbb{R})}(\phi_i).$$

Since the functions $\phi_1, \ldots, \phi_k$ have disjoint supports, the extension of $\theta_m^{(\mathbb{R})}$ to A remains an isometry.

Now suppose that $D^{(1)}, D^{(2)}, \ldots, D^{(k)}$ are disjoint rectangles in $\mathbb{T}^m$, and let

$$c_1, \ldots, c_k \in \mathbb{C}$$

be constants. Denote by A' the set of all functions of the form

$$\phi(\mathbf{t}) = \sum_{i=1}^k \left(c_i \Sigma_m[\chi_{D^{(j)}}(\mathbf{t})] + \overline{c_j} \Sigma_m[\chi_{-D^{(j)}}(\mathbf{t})] \right);$$

we claim that $A' = A$. The inclusion $A \subset A'$ is clear from the definitions. For the reverse inclusion, notice that the functions in A and in A' are invariant under the action of $\Sigma_m \times \{\pm 1\}$ by construction, so it is enough to show that $A_D \supset A'_D$, where A_D and A'_D denote the functions on D obtained by restriction to D. As a real vector space A'_D is spanned by functions of the form $\chi_{\widetilde{D}}$ and $\mathrm{i}\chi_{\widetilde{D}}$, where

$$\widetilde{D} = D_1 \times \cdots \times D_m \subset D.$$

Now for $\mathbf{t} \in D$,

$$\chi_{\widetilde{D}}(\mathbf{t}) = \Sigma_m \left[\prod_{i=1}^m \phi_i^+(t_i) \right]$$

and

$$\mathrm{i}\chi_{\widetilde{D}}(\mathbf{t}) = \Sigma_m \left[\phi_1^- \prod_{i=2}^m \phi_i^+(t_i) \right],$$

so both functions lie in A'.

Since A' is also a real vector space, this shows that $A \supset A'$.

We now claim that the closure of A in $L^2_{\rho^m}(\mathbb{T}^m)$ coincides with $Q_m^{(\mathbb{R})}$. To see this, let ϕ be an element of $Q_m^{(\mathbb{R})}$ and fix $\epsilon > 0$. Choose a function

$$\phi^{(\epsilon)} \in L^2_{\rho^m}(\mathbb{T}^m)$$

that vanishes on an ϵ-neighborhood of $\mathbb{T}^m \setminus D$, for which

$$\int_D |\phi - \phi^{(\epsilon)}| \, \mathrm{d}\rho^m < \epsilon.$$

We can also find a set of disjoint rectangles $D^{(1)}, \ldots, D^{(k)} \subset D$ and constants

$$c_1, \ldots, c_k \in \mathbb{C}$$

for which

$$\left\| \phi^{(\epsilon)} - \sum_{i=1}^{k} c_i \chi_{D^{(i)}} \right\|_2 < \epsilon.$$

Write

$$\phi_\epsilon = \sum_{i=1}^{k} \left(c_i \Sigma_m[\chi_{D^{(i)}}] + \overline{c_i} \Sigma_m[\chi_{D^{(i)}}] \right),$$

so that $\|\phi - \phi_\epsilon\| < (1 + 2m!)\epsilon$. As above, this density for the functions restricted to D shows that the closure of A is $Q_m^{(\mathbb{R})}$. Since the map $\theta_m^{(\mathbb{R})}$ is an isometry on the dense subspace A, it extends to an isometry on $Q_m^{(\mathbb{R})}$. We claim that

$$\theta_m^{(\mathbb{R})} \left(Q_m^{(\mathbb{R})} \right) = H_m^{(\mathbb{R})}; \tag{1.10}$$

it is clear from the construction that $\theta_m^{(\mathbb{R})} \left(Q_m^{(\mathbb{R})} \right) \subset H_m^{(\mathbb{R})}$. The real linear span of the set of Hermite–Itô polynomials of the form

$$:\prod_{i=1}^{m} (c_i \Phi_{D_i} + \overline{c_i} \Phi_{-D_i}):$$

where D_i is an interval of the form

$$D_{p,q} = [\frac{p}{2^q}, \frac{p+1}{2^q})$$

with $1 \leqslant q < \infty$ and $0 \leqslant p < 2^q$, is dense in $H_m^{(\mathbb{R})}$. The identity

$$c\Phi_{D_i} + \overline{c}\Phi_{-D_i} = \Re(c)\Phi^+_{D_i} + \Im(c)\Phi^-_{D_i}$$

shows that the same real vector space is spanned by the Hermite–Itô polynomials of the form

$$:\prod_{i=1}^{m} \Phi^{\pm}_{D_i}: \;. \tag{1.11}$$

Thus it is sufficient to show that each function of the form in equation (1.11) is an element of $\theta_m^{(\mathbb{R})}\left(Q_m^{(\mathbb{R})}\right)$.

Any interval $D_{p,q}$ is a disjoint union of intervals of the form $D_{\ell,n}$ if n is large enough, so a function of the form in equation (1.11) may be represented as a finite sum of Hermite–Itô polynomials of the form

$$:\prod_{i=1}^{m} \Phi^{\pm}(D_{p_i,n}):,$$

corresponding to the decomposition of the rectangle $\prod_{i=1}^{m} D_i$ into disjoint rectangles $\prod_{i=1}^{m} D_{p_i,n}$. Write a sum of the form in equation (1.11) as $\sum_n^1 + \sum_n^2$ where $\sum_n^1$ is taken over all the parameter choices $(p_1, \dots, p_m)$ for which the collection

$$D_{p_1,n}, \dots, D_{p_m,n}$$

is a family of disjoint intervals, and $\sum_n^2$ comprises the remaining terms.

By construction, the Gaussian random variables appearing in $\sum_n^1$ are independent, so

$$:\prod_{i=1}^{m} \Phi^{\pm}(D_{p_i,n}): = \prod_{i=1}^{m} \Phi^{\pm}(D_{p_i,n}) = \sqrt{\frac{1}{m!}}\Sigma_m \left[\prod_{i=1}^{m} \phi^{\pm}_{D_{p_i,n}}(t_i)\right]$$

(where the implicit signs in the product on the left-hand side are carried through to the right-hand side, and the notation $\phi^{\pm}_{D_{p_i,n}}$ is the obvious extension of that given in equation (1.6) and equation (1.7)).

In particular, $\sum_n^1 \in \theta_m^{(\mathbb{R})}\left(Q_m^{(\mathbb{R})}\right)$, so it will be enough to show that

$$\| \textstyle\sum_n^2 \|_2^2 = \mathbb{E}\left(\left(\sum_n^2\right)^2\right) \to 0$$

as $n \to \infty$.

Fix a term in $\sum_n^2$, and enumerate the intervals $D_{p_i,q}$ which appear in it in ascending order of p_i as

$$D_1 \subset \dots \subset D_{m'}$$

for some $m' < m$ (since the intervals $D_{p,q}$ as p varies partition $[0,1)$, and by hypothesis not all the intervals appearing are disjoint, two at least must coincide). Thus this summand may be written in the form

$$\Pi = :\prod_{i=1}^{m'} \left(\Phi^{+}_{D_i}\right)^{r_i^+} \left(\Phi^{-}_{D_i}\right)^{r_i^-}: = \prod_{i=1}^{m'} :\left(\Phi^{+}_{D_i}\right)^{r_i^+} \left(\Phi^{-}_{D_i}\right)^{r_i^-}: \tag{1.12}$$

by Lemma 4, with $r_i^+ + r_i^- > 1$ for at least one value of i. We estimate the L^2-norm of the function $\sum_n^2$ as follows. By orthogonality of the Hermite–Itô polynomials, the inner product of distinct products of the form equation (1.12) vanish, so

$$\begin{aligned}\left\langle \textstyle\sum_n^2, \sum_n^2 \right\rangle &= \sum_{\Pi} \langle \Pi, \Pi \rangle \\ &= \sum \prod_{i=1}^{m'} \mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^+)^{r_i^+} (\Phi_{D_i}^-)^{r_i^-} \boldsymbol{:}^{\,2} \right)\end{aligned}$$

Since the random variables Φ_D^+ and Φ_D^- are independent Gaussians, we have

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^+)^{r_i^+} (\Phi_{D_i}^-)^{r_i^-} \boldsymbol{:}^{\,2} \right) &= \mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^+)^{r_i^+} \boldsymbol{:}^{\,2} \, \boldsymbol{:}(\Phi_{D_i}^-)^{r_i^-} \boldsymbol{:}^{\,2} \right) \\ &= \mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^+)^{r_i^+} \boldsymbol{:}^{\,2} \right) \mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^-)^{r_i^-} \boldsymbol{:}^{\,2} \right) \\ &\leqslant C_{r_i^+} \rho^{r_i^+}(D_i) \times C_{r_i^-} \rho^{r_i^-}(D_i) \\ &\leqslant C_m \rho^{r_i^+ + r_i^-}(D_i);\end{aligned}$$

it follows that

$$\prod_{i=1}^{m'} \mathbb{E}\left(\boldsymbol{:}(\Phi_{D_i}^+)^{r_i^+} (\Phi_{D_i}^-)^{r_i^-} \boldsymbol{:}^{\,2} \right) \leqslant C_m \rho^m (D_{p_1,n} \times \cdots \times D_{p_m,n})$$

where $D_{p_1,n} \times \cdots \times D_{p_m,n}$ is the original rectangle corresponding to the summand from $\sum_n^2$, and C_m depends only on m. Thus $\left\langle \sum_n^2, \sum_n^2 \right\rangle \leqslant C_m \rho^m(B_n)$, where B_n is the union of those rectangles

$$D_{p_1,n} \times \cdots \times D_{p_m,n}$$

which intersect some hyperplane defined by $t_i = t_j$, $t_i = -t_j$ on $\mathbb{T}^m$. Since the measure ρ is assumed to be non-atomic, each of these hyperplanes is a ρ^m-null set, so $\rho^m(B_n) \to 0$ as $n \to \infty$, completing the proof that any function of the form in equation (1.11) is an element of $\theta_m^{(\mathbb{R})}\left(Q_m^{(\mathbb{R})}\right)$ and therefore showing part (1) of Theorem 2.

It remains to show part (2); to do this we first show that

$$\boldsymbol{:}\prod_{i=1}^{m} x_{n_i} \boldsymbol{:}$$

is the function

$$\mathbf{t} \longmapsto \theta_m^{(\mathbb{R})}\left(\sqrt{\frac{1}{m!}}\varSigma_m\left[\mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^m n_i t_i}\right]\right). \tag{1.13}$$

To this end, fix $\epsilon \in (0,1)$, and for each i with $1 \leqslant i \leqslant m$, choose constants $a_{p,n}^{(i)}$ so that the random variable

$$f_i^{(\epsilon)} = \sum a_{p,n}^{(i)} \varPhi_{D_{p,n}}^{\pm} \in H_1^{(\mathbb{R})}$$

has

$$\left\| x_{n_i} - f_i^{(\epsilon)} \right\|_2 = \left\| \mathrm{e}^{2\pi\mathrm{i}n_i\cdot} - \phi_i^{(\epsilon)}(\cdot) \right\|_2 \leqslant \epsilon,$$

where

$$\phi_i^{(\epsilon)}(t) = \sum_p a_{p,n}^{(i)} \phi_{D_{p,n}}^{\pm}(t).$$

Write $\varPhi^{(\epsilon)} = \prod_{i=1}^m f_i^{\epsilon}$; then

$$\left\| \prod_{i=1}^m x_{n_i} - \varPhi^{(\epsilon)} \right\| \leqslant C\epsilon$$

for some constant C, and therefore

$$\left\| :\prod_{i=1}^m x_{n_i}: \; - \; :\varPhi^{(\epsilon)}: \right\| \leqslant C\epsilon.$$

Now by construction we may write

$$\varPhi^{(\epsilon)} = \sum a_{p_1,n}^{(1)} \cdots a_{p_m,n}^{(m)} \prod_{i=1}^m \varPhi_{D_{p_i,n}}^{\pm}.$$

Decompose the last sum into $\sum^1 + \sum^2$ and correspondingly write

$$\varPhi^{(\epsilon)} = \varPhi^{(\epsilon,1)} + \varPhi^{(\epsilon,2)},$$

where the sum $\sum^1$ involves all those collections $p_1, \dots, p_m$ with the property that

$$\prod_{i=1}^m D_{p_1,n} \cap \partial D = \emptyset.$$

As argued above, $\| \sum^2 \|_2 \to 0$ as $n \to \infty$, so for sufficiently large n we have

$$\left\| :\prod_{i=1}^m x_{n_i}: \; - \; :\varPhi^{(\epsilon,1)}: \right\| \leqslant C\epsilon.$$

Write $\phi^{(\epsilon)}(\mathbf{t}) = \prod_{i=1}^m \phi_i^{\epsilon}(t_i)$, so that

$$\left\| \mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^{m} n_i t_i} - \phi^{(\epsilon)}(\mathbf{t}) \right\|_2 \leqslant C\epsilon,$$

and hence

$$\left\| \sqrt{\frac{1}{m!}} \Sigma_m \left[\mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^{m} n_i t_i} \right] - \sqrt{\frac{1}{m!}} \Sigma_m \left[\phi^{(\epsilon)}(\mathbf{t}) \right] \right\|_2 \leqslant C\epsilon.$$

As above, we may write

$$\phi^{(\epsilon)}(\mathbf{t}) = \sum a^{(1)}_{p_1,n} \cdots a^{(m)}_{p_m,n} \prod_{i=1}^{m} \phi^{\pm}_{D_{p_i,n}}(t_i)$$

and split this sum in the same way into $\sum^1 + \sum^2$ with corresponding decomposition $\phi^{(\epsilon)} = \phi^{(\epsilon,1)} + \phi^{(\epsilon,2)}$. As before, $\|\phi^{(\epsilon,2)}\|_2 \leqslant \epsilon$ for large enough n, so

$$\left\| \sqrt{\frac{1}{m!}} \Sigma_m \left[\mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^{m} n_i t_i} \right] - \sqrt{\frac{1}{m!}} \Sigma_m \left[\phi^{(\epsilon,1)}(\mathbf{t}) \right] \right\|_2 \leqslant C\epsilon.$$

Now $\theta^{(\mathbb{R})}_m$ is an isometry, so it follows that

$$\left\| \theta^{(\mathbb{R})}_m \left(\sqrt{\frac{1}{m!}} \Sigma_m \left[\mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^{m} n_i t_i} \right] \right) - \theta^{(\mathbb{R})}_m \left(\sqrt{\frac{1}{m!}} \Sigma_m \left[\phi^{(\epsilon,1)}(\mathbf{t}) \right] \right) \right\|_2 \leqslant C\epsilon.$$

By construction of the map $\theta^{(\mathbb{R})}_m$ we have

$$\theta^{(\mathbb{R})}_m \left(\sqrt{\frac{1}{m!}} \Sigma_m \left[\phi^{(\epsilon,1)}(\mathbf{t}) \right] \right) = {:}\Phi^{(\epsilon,1)}{:}\,,$$

so we deduce that

$$\left\| {:}\prod_{i=1}^{m} x_{n_i}{:}\; - \theta^{(\mathbb{R})}_m \left(\sqrt{\frac{1}{m!}} \Sigma_m \left[\mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^{m} n_i t_i} \right] \right) \right\|_2 \leqslant C\epsilon,$$

which proves equation (1.13) since ϵ was arbitrary.

Now in the space $Q^{(\mathbb{R})}_m$ consider the group of operators $\{V^n \mid n \in \mathbb{Z}\}$, where V^n is defined by

$$(V^n \phi)(\mathbf{t}) = \mathrm{e}^{2\pi \mathrm{i} n \sum_{i=1}^{m} t_i} \phi(\mathbf{t}).$$

We claim that $U^n_T = \theta^{(\mathbb{R})}_m V^n (\theta^{(\mathbb{R})}_m)^{-1}$ (from which it follows that the space $H^{(\mathbb{R})}_m$ is U_T-invariant). Write $f = {:}\prod_{i=1}^{m} x_{n_i}{:}$; since $U^n_T f = {:}\prod_{i=1}^{n} x_{n_i+n}{:}$ it follows that

$$\begin{aligned}
U_T^n f &= \theta_m^{(\mathbb{R})}\left(\sqrt{\frac{1}{m!}}\Sigma_m\left[\mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^m (n_i+n)t_i}\right]\right)\\
&= \theta_m^{(\mathbb{R})}\left(\mathrm{e}^{2\pi\mathrm{i}n\sum_{i=1}^m t_i}\sqrt{\frac{1}{m!}}\Sigma_m\left[\mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^m n_i t_i}\right]\right)\\
&= \theta_m^{(\mathbb{R})}V^n(\theta_m^{(\mathbb{R})})^{-1}f;
\end{aligned}$$

since the linear space of such functions generates the whole space $H_m^{(\mathbb{R})}$ the proof is completed. □

We have constructed spaces of real-valued functions $H_m^{(\mathbb{R})}$ for $m \geqslant 1$; add to these $H_0^{(\mathbb{R})} = \mathbb{R}$, the one-dimensional space of real constant functions.

Theorem 3. *The space $H^{(\mathbb{R})}$ of real-valued functions in $L^2_\mu(X)$ decomposes as the direct sum $\bigoplus_{m=0}^\infty H_m^{(\mathbb{R})}$.*

Proof. For any collection $n_1, \dots, n_p$ of integers define the real subspace

$$H_{n_1,\dots,n_p} \subset H^{(\mathbb{R})}$$

to be the space of functions in $L^2_\mu(X)$ that depend only on the random variables $x_{n_1}, \dots, x_{n_p}$. Since the closure of the sum of all the real subspaces $H_{n_1,\dots,n_p}$ is all of $H^{(\mathbb{R})}$, it is sufficient to check that

$$H_{n_1,\dots,n_p} \subset \bigoplus_{m=0}^{\infty} H_m^{(\mathbb{R})}.$$

Suppose that the random variables x_{n_i} for $1 \leqslant i \leqslant p$ have a joint probability density function of the form

$$\psi(\mathbf{s}) = \frac{\sqrt{\det A}}{(2\pi)^{p/2}}\mathrm{e}^{-\frac{1}{2}\langle A\mathbf{s},\mathbf{s}\rangle},$$

where A is a symmetric positive-definite matrix. Any function $f \in H_{n_1,\dots,n_p}$ can be identified with a function f with the property that

$$\int f^2(\mathbf{s})\psi(\mathbf{s})\,\mathrm{d}s_1 \dots \mathrm{d}s_p < \infty.$$

The matrix A may be diagonalized, so up to a linear change of variables we may assume that

$$\psi(\mathbf{s}) = \frac{1}{(2\pi)^{p/2}}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^p t_i^2}.$$

For any indicator function χ_B and $\epsilon > 0$, there is[4] a polynomial P_ϵ in p variables with

$$\frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} |\chi_B(t) - P_\epsilon(t)|^2 \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^p t_i^2} \,\mathrm{d}\mathbf{t} < \epsilon.$$

The polynomial P_ϵ can be written as a sum of products of Hermite polynomials in one variable, and each such product corresponds to an element of the space

$$\bigoplus_{m=0}^{\infty} H_m^{(\mathbb{R})},$$

proving the theorem. □

All that remains is to extend the decomposition to complex-valued functions. Write $H_m = H_m^{(\mathbb{R})} + \mathrm{i}H_m^{(\mathbb{R})}$, $Q_m = Q_m^{(\mathbb{R})} + \mathrm{i}Q_m^{(\mathbb{R})}$ for all $m \geqslant 1$. By construction H_m is a closed subspace of L^2_μ and $H_m \perp H_n$ for $m \neq n$. Given a function $\phi \in Q_m$, define

$$\phi_1(t) = \frac{1}{2}\left(\phi(t) + \overline{\phi(-t)}\right)$$

and

$$\phi_2(t) = \frac{1}{2\mathrm{i}}\left(\phi(t) - \overline{\phi(-t)}\right),$$

so that $\phi = \phi_1 + \mathrm{i}\phi_2$ and we can extend the function constructed in Theorem 2 to the complex space Q_m by setting $\theta_m(\phi) = \theta_m^{(\mathbb{R})}(\phi_1) + \mathrm{i}\theta_m^{(\mathbb{R})}(\phi_2)$. Similarly, we extend the operator V to Q_m by defining

$$(V^n\phi)(t) = \mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^m t_i}\phi(t).$$

By Theorems 2 and 3 the spaces H_m and Q_m have the following properties.

Proposition 1. *With the notation above:*

1. *$U_T H_m = H_m$ for $m \geqslant 1$;*
2. *$L^2_\mu = \bigoplus_{m=0}^\infty H_m$, where $H_0 = \mathbb{C}$ is the space of complex constant functions; and*
3. *the isometry $\theta_m : Q_m \to H_m$ has $U_T^n = \theta_m V^n (\theta_m)^{-1}$ for all $m \geqslant 0$ and $n \in \mathbb{Z}$.*

---

[4] This is simply the statement that the classical polynomials introduced by Hermite [14] are dense in the Hilbert space associated to the inner-product $\langle f, g\rangle = \int_{-\infty}^{\infty} f(t)g(t)\mathrm{e}^{-t^2/2}\,\mathrm{d}t$; see Hewitt and Stromberg [15, Ex. 16.25] for example.

1.3.2 Mixing Properties of Gaussian Systems

Theorem 1 is the most basic result connecting properties of the spectral measure ρ to dynamical properties of the associated Gaussian dynamical system: if μ is ergodic for T then ρ must be non-atomic. The first result that becomes available using the decomposition of L^2_μ constructed in Section 1.3 is a strong converse to this result.

Theorem 4. *Let $(X,\mathcal{B},\mu,T)$ be a Gaussian measure-preserving system with associated spectral measure ρ and correlation function $c:\mathbb{Z}\to\mathbb{R}$.*

1. *If ρ is non-atomic, then $(X,\mathcal{B},\mu,T)$ is ergodic.*
2. *The system $(X,\mathcal{B},\mu,T)$ is mixing if and only if $c(n)\to 0$ as $|n|\to\infty$.*

Proof. (1) Assume that ρ is non-atomic, let $f\in L^2_\mu$ be an eigenfunction of U_T, and write $U_Tf=\kappa f$ for some $\kappa\in\mathbb{S}^1$. Expand f using the orthogonal decomposition from Section 1.3 into $f=\sum_{m=0}^{\infty}f_m$ with $f_m\in H_m$ (and hence $\langle f_m,f_n\rangle=0$ for $m\neq n$). The spaces H_m are invariant under U_T and f is an eigenfunction, so all the components f_m are eigenfunctions of U_T with the same eigenvalue also. Let $\phi_m=(\theta_m)^{-1}f_m$ for $m\geqslant 0$, so that by Proposition 1(3) above,

$$\begin{aligned}(V^n\phi_m)(\mathbf{t})&=\mathrm{e}^{2\pi\mathrm{i}n\sum_{i=1}^m t_i}\phi_m(\mathbf{t})\\&=\big(V_n(\theta_m)^{-1})f_m\big)(\mathbf{t})\\&=(\theta_m)^{-1}U_T^nf_m\\&=(\theta_m)^{-1}(\kappa f_m)(\mathbf{t})=\kappa\left((\theta_m)^{-1}f_m\right)(\mathbf{t})=\kappa\phi_m(\mathbf{t}),\end{aligned}$$

so that multiplication by the function

$$\mathbf{t}\longmapsto \mathrm{e}^{2\pi\mathrm{i}n\sum_{i=1}^m t_i}$$

sends ϕ_m to $\kappa\phi_m$ as elements of Q_m. Thus, for $m\geqslant 1$,

$$\begin{aligned}&\int\left|\mathrm{e}^{2\pi\mathrm{i}n\sum_{i=1}^m t_i}\phi_m(\mathbf{t})-\kappa\phi_m(\mathbf{t})\right|^2\,\mathrm{d}\rho^m(\mathbf{t})\\&=\int\left|\mathrm{e}^{2\pi\mathrm{i}n\sum_{i=1}^m t_i}-\kappa\right|^2|\phi_m(\mathbf{t})|^2\,\mathrm{d}\rho^m(\mathbf{t})=0,\end{aligned}$$

so $\phi_m(\mathbf{t})=0$ almost everywhere with respect to ρ^m on the complement of the set

$$\{\mathbf{t}\mid \mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^m t_i}=\kappa\},$$

which is a null set with respect to ρ^m since, by assumption, ρ is non-atomic. Hence for $m\geqslant 1$ we have $\phi_m=0$ almost surely, and hence $f=f_0$ is a constant and T is weak-mixing.

Turning to the proof of (2), recall that if we think of the coordinate x_0 as the random variable $x_0:X\to\mathbb{R}$ we have

$$c(n) = \langle U_T^n x_0, x_0 \rangle$$

for all $n \in \mathbb{Z}$. Thus, if T is mixing, we must have $c(n) \to 0$ as $|n| \to 0$ (see [5, Ex. 2.7.5]). Conversely, assume that $c(n) \to 0$ as $|n| \to 0$, let $f_m \in H_m$ for some $m \geqslant 1$, and write $\phi_m = (\theta_m)^{-1} f_m$. By Proposition 1,

$$\langle U_T^n f_m, f_m \rangle = \int \mathrm{e}^{2\pi \mathrm{i} \sum_{i=1}^m t_i} |\phi_m(\mathbf{t})|^2 \,\mathrm{d}\rho^m(\mathbf{t}). \tag{1.14}$$

Write ρ^{*m} for the m-fold convolution $\rho * \cdots * \rho$, and let $\psi(t)$ be the conditional expectation of the function $|\phi_m(\mathbf{t})|^2$ under the condition that $\sum_{i=1}^m t_i = t$. Then

$$\langle U_T^n f_m, f_m \rangle = \int \mathrm{e}^{2\pi \mathrm{i} t} \psi(t) \,\mathrm{d}\rho^{*m}$$

(indeed, this identity for all functions characterizes the conditional expectation). By construction, $\phi_m \in Q_m$, so $\int \psi(t) \,\mathrm{d}\rho^{*m}(t) < \infty$. By assumption,

$$\int \mathrm{e}^{2\pi \mathrm{i} n t} \,\mathrm{d}\rho^{*m}(t) = c^m(n) \to 0$$

as $|n| \to \infty$, so the Riemann–Lebesgue lemma [5, Lem. C.16] shows that

$$\langle U_T^n f_m, f_m \rangle \to 0$$

as $n \to \infty$. Clearly

$$\left\langle U_T^n \sum_{m=1}^N f_m, \sum_{n=1}^N f_m \right\rangle = \sum_{n=1}^N \langle U_T^n f_m, f_m \rangle,$$

so we also have $\langle U_T^n f, f \rangle \to 0$ as $n \to \infty$ for any function f that can be written in the form $f = \sum_{m=1}^N f_m$ with $f_m \in H_m$. The set of such functions is dense in the set of functions in L^2_μ with zero integral, so we have proved that T is mixing by the characterization of mixing due to Rényi [33]: a measure-preserving transformation T is mixing if and only if

$$\mu(A \cap T^{-n} A) \to \mu(A)^2$$

for all $A \in \mathcal{B}$. □

As an application of Theorem 4, notice that by the Riemann–Lebesgue lemma ([5, Lemma C.16]), if ρ is absolutely continuous with respect to Lebesgue measure, then $c(n) \to 0$ as $|n| \to \infty$, so the corresponding Gaussian measure-preserving system is mixing. On the other hand, there are non-atomic measures on $\mathbb{S}^1$ that are singular with respect to Lebesgue measure yet have the property that $c(n) \to 0$ as $|n| \to \infty$. The corresponding Gaussian measure-preserving systems are therefore weak- but not strong-mixing.

1.4 A Rigid $\mathbb{Z}^d$-Action With Mixing Shapes

In this section we describe an example due to Ferenczi and Kamiński [6] that illuminates what is possible in general for measure-preserving $\mathbb{Z}^d$-actions with $d > 1$; the treatment[5] is taken from [40]. This example is included to illustrate some of the power and flexibility of Gaussian constructions, and it should be contrasted with the algebraic examples of mixing behavior from [5, Sect. 8.2]. A finite set $\{\mathbf{n}_1, \dots, \mathbf{n}_r\} \subset \mathbb{Z}^d$ is a *mixing shape* for a measure-preserving $\mathbb{Z}^d$-action T on $(X, \mathcal{B}, \mu)$ if

$$\mu\left(\bigcap_{j=1}^{r} T^{-k\mathbf{n}_j} A_j\right) \to \prod_{j=1}^{r}$$

as $k \to \infty$. The example is constructed as a Gaussian process, and the construction from Section 1.2 extends easily to define $\mathbb{Z}^d$-actions as follows. Let $(\Omega, \mathcal{F}_0) = \prod_{\mathbf{n}\in\mathbb{Z}^d}(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ–algebra on $\mathbb{R}$. Let $\xi_{\mathbf{n}}(\omega)$ be the $\mathbf{n}$th coordinate of $\omega \in \Omega$, and let μ be a probability measure on $(\Omega, \mathcal{F}_0)$ with the property that for any k–tuple of integer vectors $\mathbf{n}_1, \dots, \mathbf{n}_k$ of the k–dimensional random variable $(\xi_{\mathbf{n}_1}, \dots, \xi_{\mathbf{n}_k})$ is a k–dimensional Gaussian law, and the joint distribution is stationary in the sense that $\mu^{(\mathbf{n}_1+\mathbf{m}, \dots, \mathbf{n}_k+\mathbf{m})} = \mu^{(\mathbf{n}_1, \dots, \mathbf{n}_k)}$ for any $\mathbf{m} \in \mathbb{Z}^d$. Let $\mathcal{F}$ denote the completion of $\mathcal{F}_0$ under μ. Then $(\Omega, \mathcal{F}, \mu, \{\xi_{\mathbf{n}}\}_{\mathbf{n}\in\mathbb{Z}^d})$ is a d–dimensional Gaussian stationary sequence. Assume that $\mathbb{E}\{\xi_{\mathbf{n}}\} = 0$ for each $\mathbf{n} \in \mathbb{Z}^d$. The covariance function $R : \mathbb{Z}^d \to \mathbb{C}$ may be expressed in terms of a (symmetric) spectral measure ρ on $\mathbb{T}^d$ as

$$R(\mathbf{n}) = \mathbb{E}\{\xi_{\mathbf{n}+\mathbf{m}}\xi_{\mathbf{m}}\} = \int_0^1 \cdots \int_0^1 \mathrm{e}^{-2\pi\mathrm{i}(n_1 s_1 + \cdots + n_d s_d)} \rho(\,\mathrm{d}s_1 \dots \mathrm{d}s_d).$$

Conversely, if ρ is a symmetric finite measure on $\mathbb{T}^d$, then there is a unique d–dimensional Gaussian stationary sequence whose spectral measure is ρ. Associated to any Gaussian stationary sequence of the above form there is a measure–preserving $\mathbb{Z}^d$–action α, defined by the shift on Ω. Standard approximation arguments give the following. Let C denote the class of functions $f : \Omega \to \mathbb{C}$ with the property that $f(\omega) = F(\xi_{\mathbf{m}_1}(\omega), \dots, \xi_{\mathbf{m}_t}(\omega))$ for some $\mathbf{m}_1, \dots, \mathbf{m}_t$ and some bounded continuous function $F : \mathbb{R}^t \to \mathbb{C}$. Let α be a Gaussian $\mathbb{Z}^d$–action. Then, in order to check any mixing property, it is sufficient to check it for functions in the class C.

For each $\mathbf{n} \in \mathbb{Z}^d$, the $\mathbb{Z}$–action generated by the transformation $\alpha_{\mathbf{n}}$ is again Gaussian, on $(\Omega, \mathcal{F}_{\mathbf{n}})$, where $\mathcal{F}_{\mathbf{n}}$ is the sub-σ-algebra of $\mathcal{F}$ generated by the projections $\{\xi_{k\mathbf{n}}\}_{k\in\mathbb{Z}}$. The spectral measure of $\alpha_{\mathbf{n}}$ is $\rho_{\mathbf{n}} = \rho\psi_{\mathbf{n}}^{-1}$, where the map $\psi_{\mathbf{n}} : \mathbb{T}^d \to \mathbb{T}$ is given by $\psi_{\mathbf{n}}(s_1, \dots, s_d) = n_1 s_1 + \cdots + n_d s_d$ modulo 1.

[5] Consult Totoki [35] or Cornfeld, Fomin and Sinaĭ [4] for the details of the approximation arguments needed.

Now choose $1, \beta_1, \dots, \beta_d$ linearly independent over $\mathbb{Q}$, and let

$$f(t) = (\beta_1 t, \dots, \beta_d t)$$

for $t \in \mathbb{T}$. Let $\imath : \mathbb{T}^d \to \mathbb{T}^d$ be the involution

$$\imath(t_1, \dots, t_d) = (1 - t_1, \dots, 1 - t_d).$$

Define a symmetric, singular, continuous measure ρ on $\mathbb{T}^d$ by

$$\rho = \frac{1}{2}\left(m_{\mathbb{T}^d} f^{-1} + m_{\mathbb{T}^d}(\imath \circ f)^{-1}\right),$$

and let α be the Gaussian $\mathbb{Z}^d$-action with spectral measure ρ. Then

$$R(\mathbf{n}) = \frac{\sin(2\pi(n_1\beta_1 + \cdots + n_d\beta_d))}{2\pi(n_1\beta_1 + \cdots + n_d\beta_d)}. \tag{1.15}$$

Choose a sequence $\mathbf{n}_j = (n_1^{(j)}, \dots, n_d^{(j)}) \to \infty$ with $n_1^{(j)}\beta_1 + \cdots + n_d^{(j)}\beta_d \to 0$ as $j \to \infty$. Then $R(\mathbf{n}_j) \to 1$ as $j \to \infty$, so the $2t$–dimensional random Gaussian vector

$$\Phi_j(\omega) = \left(\xi_{\mathbf{m}_1}(\omega), \dots, \xi_{\mathbf{m}_t}(\omega), \xi_{\mathbf{m}_1 - \mathbf{n}_j}(\omega), \dots, \xi_{\mathbf{m}_t - \mathbf{n}_j}(\omega)\right)$$

has covariance matrix

$$\begin{pmatrix} V_{00}^{(j)} & V_{10}^{(j)} \\ V_{01}^{(j)} & V_{11}^{(j)} \end{pmatrix},$$

where $V_{00}^{(j)} = V_{11}^{(j)}$ is the covariance matrix V of

$$(\xi_{\mathbf{m}_1}(\omega), \dots, \xi_{\mathbf{m}_t}(\omega)),$$

and the (p,q)th entry of $V_{01}^{(j)}$ is

$$\mathbb{E}\{\xi_{\mathbf{m}_p}\xi_{\mathbf{m}_q - \mathbf{n}_j}\} = R(\mathbf{m}_p - \mathbf{m}_q + \mathbf{n}_j) \to R(\mathbf{m}_p - \mathbf{m}_q)$$

as $j \to \infty$ by our choice of $\mathbf{n}_j$. Thus $V_{01}^{(j)} \to V$ and similarly $V_{10}^{(j)} \to V$ as $j \to \infty$. It follows that

$$\mu\left(\alpha_{\mathbf{n}_j}(A) \cap A)\right) \to \mu(A)$$

as $j \to \infty$ for any $A \in \mathcal{F}$, so α is rigid.

Now let $S = \{\mathbf{n}_1, \dots, \mathbf{n}_r\} \subset \mathbb{Z}^d$ be a fixed shape, and define a random vector of dimension rt by $\Psi_k(\omega) = \left(\xi_{\mathbf{m}_i - k\mathbf{n}_j}(\omega) \mid i = 1, \dots, t; j = 1, \dots, r\right)$. This vector is Gaussian with zero mean and covariance matrix

$$V_k = \begin{pmatrix} V_k^{11} & V_k^{12} & \dots & V_k^{1r} \\ \vdots & & & \vdots \\ V_k^{r1} & V_k^{r2} & \dots & V_k^{rr} \end{pmatrix},$$

where $V_k^{j\ell}$ is the $t \times t$ matrix whose (p,q)th element is

$$v_{(p,q)}^{(j,\ell)}(k) = \mathbb{E}\left(\xi_{\mathbf{m}_p - k\mathbf{n}_j} \xi_{\mathbf{m}_q - k\mathbf{n}_\ell}\right) = \begin{cases} R(\mathbf{m}_p - \mathbf{m}_q) & \text{if } j = \ell; \\ R(\mathbf{m}_p - \mathbf{m}_q + k\mathbf{n}_\ell - k\mathbf{n}_j) & \text{if } j \neq \ell. \end{cases}$$

Notice that $V_0 = V_k^{jj}$ is the covariance matrix of $(\xi_{\mathbf{n}_1}, \dots, \xi_{\mathbf{n}_t})$. For $j \neq \ell$, it is clear from equation (1.15) that

$$\lim_{k \to \infty} v_{(p,q)}^{(j,\ell)}(k) = 0,$$

so that

$$\lim_{k \to \infty} V_k = \begin{pmatrix} V_0 & 0 & \dots & 0 \\ 0 & V_0 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & V_0 \end{pmatrix}.$$

It follows that α is mixing for all shapes.

Chapter 2
Geometric and Combinatorial Constructions

We have seen in [5, Chap. 3] the Gauss map, which is of interest because of the direct connection to continued fractions and Diophantine approximation. Noticing that the Gauss measure is preserved by this map gives rise to an important example of a measure-preserving transformation, but it is not clear *a priori* what would make one write down that measure. In [5, Ex. 2.9] we used the Kolmogorov consistency theorem [5, Th. A.11] to write down convenient measures on finite product spaces $A^{\{1,2,\dots,n\}}$ and used this to construct measure-preserving systems in which the map is the left shift on $A^{\mathbb{Z}}$.

In this appendix we construct some different examples, which may initially be thought of as generalizing the circle rotation in [5, Ex. 2.2]. In the circle rotation we start with Lebesgue measure, which enjoys certain geometrical properties (in particular, it is invariant under translations), and use that to write down a map that respects enough of the geometric structure to preserve the measure. As in Chapter 1, we can do no more than introduce these constructions here, and point at some surveys of the more sophisticated results[1].

2.1 Interval Exchange Transformations

Interval exchanges are a natural class of measure-preserving dynamical systems (the order-preserving piecewise isometries of an interval); they are defined by a small amount of combinatorial data yet have a very rich structure[2].

[1] Several surveys are mentioned in the notes; a particularly significant *lacuna* in both this appendix and in Chapter 1 concerns spectral properties; there are surveys on this aspect by Katok and Thouvenot [22] and Goodson [13], and a monograph by Nadkarni [28].

[2] Interval exchange transformations were introduced by Oseledec [31] and Katok and Stepin [21]. Keane [23] showed that there are only finitely many invariant ergodic measures and characterized a dense orbit property (called minimality by analogy with the case of continuous maps); Keynes and Newton [25] and Keane [24] showed that there may be

Definition 10. Let $X = [0,1)$, $r \geqslant 2$, let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_r)$ be a probability vector with $\alpha_i > 0$ for $1 \leqslant i \leqslant r$, and let $\pi \in \Sigma_R$ be a permutation of $\{1, \dots, r\}$. Associate to this data vectors $\boldsymbol{\beta} = (\beta_0 = 0, \beta_1, \dots, \beta_r)$ where $\beta_i = \sum_{j=1}^{i} \alpha_j$ and $\boldsymbol{\beta}^\pi = (\beta_0^\pi = 0, \beta_1^\pi, \dots, \beta_r^\pi)$ where $\beta_i^\pi = \sum_{j=1}^{i} \alpha_{\pi^{-1}(j)}$. The *interval exchange transformation* associated to the data $(\boldsymbol{\alpha}, \pi)$ is the map $T : X \to X$ defined by

$$T(x) = x - \beta_{i-1} + \beta^\pi_{\pi(i)-1}$$

for $x \in [\beta_{i-1}, \beta_i)$.

Some properties follow at once from Definition 10.

- Writing $A_i = [\beta_{i-1}, \beta_i)$ for $1 \leqslant i \leqslant r$ we see that $\{A_1, \dots, A_r\}$ is a partition of $[0,1)$ into half-open intervals, with the length of A_i being α_i.
- The interval exchange defined by $(\boldsymbol{\alpha}^\pi, \pi^{-1})$, where

$$\boldsymbol{\alpha}^\pi = (\alpha_{\pi^{-1}(1)}, \dots, \alpha_{\pi^{-1}(r)}),$$

  is the inverse of T.
- T acts as a piecewise rotation, so preserves Lebesgue measure.
- $T(\beta_i) = \beta^\pi_{\pi(i+1)-1}$ for $0 \leqslant i \leqslant r-1$.

We will also speak of T as being an interval exchange on the ordered partition $(A_1, \dots, A_r)$.

Example 3. For $r = 2$ we must have

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2 = 1 - \alpha_1)$$

and

$$\boldsymbol{\beta} = (0, \alpha_1, 1).$$

Write $\Sigma_2 = \{\imath, (12)\}$, where $\imath$ is the identity and (12) swaps 1 and 2.

1. If $\pi = \imath$ then $\alpha^\pi = \alpha$ so $\boldsymbol{\beta}^\pi = \boldsymbol{\beta}$, and $T(x) = x$ is the identity map on $[0,1)$.

several ergodic measures even when Lebesgue measure is ergodic and the lengths of the intervals satisfy strong irrationality conditions; Rauzy [32] introduced a powerful inducing construction for interval exchanges; Masur [27] and Veech [36] showed that almost every minimal interval exchange is uniquely ergodic; Nogueira and Rudolph [30] showed that almost every interval exchange under a mild condition on the permutation has no continuous eigenfunction; Avila and Forni [2] showed that almost every interval exchange is either weak-mixing or a circle rotation. A detailed survey of interval exchange transformations, with particular emphasis on the geometry surrounding them and with many examples and more complete references, is given by Viana [39]. Some combinatorial and arithmetic aspects of interval exchanges are discussed in the collection edited by Berthé, Ferenczi, Mauduit and Siegel [7]. Our short and elementary introduction follows that of Keane [23] and Cornfeld, Fomin, and Sinaĭ [4] closely.

2. If $\pi = (12)$ then $\boldsymbol{\alpha}^\pi = (1-\alpha_1, \alpha_1)$ and $\boldsymbol{\beta}^\pi = (0, 1-\alpha_1, 1)$, so

$$T(x) = x + (1-\alpha_1) \pmod 1.$$

Thus an interval exchange on two intervals is simply a circle rotation, and therefore many properties are clear. In the non-trivial case $\pi = (12)$, almost every choice of α_1 gives a map with the property that every point is dense (a theorem of Kronecker), the orbit of every point is equidistributed (a theorem of Weyl), and Lebesgue measure is the only invariant measure (see [5, Ex. 4.11]). Much of the more advanced work on interval exchanges is concerned with understanding the extent to which those properties survive for $r > 2$.

Example 4. Let $r = 3$, $\pi = (13)$ and assume that $\alpha_3 \leqslant \alpha_1$. Then

$$\boldsymbol{\beta} = (0, \alpha_1, \alpha_1 + \alpha_2, 1)$$

and

$$\boldsymbol{\beta}^\pi = (0, \alpha_3, \alpha_2 + \alpha_3, 1),$$

and the resulting interval exchange

$$T(x) = \begin{cases} x + \alpha_2 + \alpha_3 & \text{if } x \in A_1; \\ x - \alpha_1 + \alpha_3 & \text{if } x \in A_2; \\ x - \alpha_1 - \alpha_2 & \text{if } x \in A_3 \end{cases}$$

is illustrated in Figure 2.1.

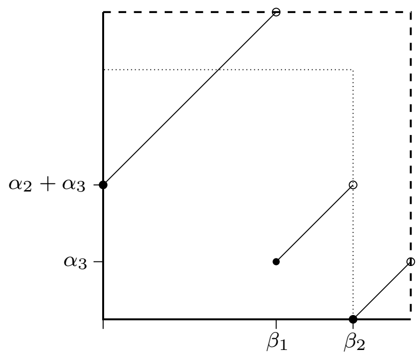

Fig. 2.1: An interval exchange on three intervals.

The interval $Y = A_1 \cup A_2$ is indicated in Figure 2.1, and the map

$$S : Y \to Y$$

induced by T is shown in Figure 2.2; we have

$$S(x) = \begin{cases} T(x) & \text{if } x \in A_2 \cup \left(A_1 \cap T^{-1}Y\right); \\ T^2(x) & \text{if not,} \end{cases}$$

and thus $S(x) = x + (\alpha_1 - \alpha_3)$ modulo $\alpha_1 + \alpha_2$. It follows that the induced map S is a circle rotation by $\frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2}$, and we may use the rather complete understanding of the dynamical properties of S to study T.

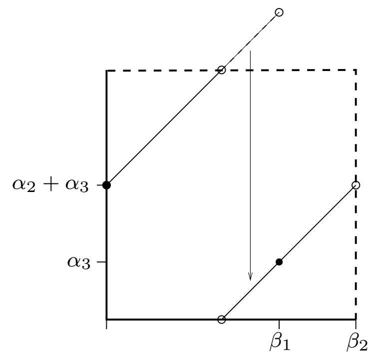

Fig. 2.2: The induced map on $A_1 \cup A_2$ is a rotation.

Notice that for $n \geqslant 2$ the map T^n is an interval exchange on intervals of the form

$$A_{i_0} \cap T^{-1}A_{i_1} \cap \cdots \cap T_{i_{n-1}}^{-(n-1)};$$

each such interval is either empty or is a half-open interval. Write $\mathcal{L}(T)$ for the set of left end points of the intervals defining an interval exchange T, so that, for $n \geqslant 2$,

$$\mathcal{L}(T^n) = \bigcup_{i=0}^{n-1} T^{-i}\left(\mathcal{L}(T)\right)$$

and

$$\mathcal{L}(T^{-n}) = \bigcup_{i=1}^{n} T^{i}\left(\mathcal{L}(T)\right).$$

By construction, any interval exchange is continuous from the right on all of $[0, 1)$ (and is an isometry at all but finitely many points). It follows that if T is an interval exchange and $T(x) = x$ then the entire exchanged interval containing x (which contains an interval of the form $[x, x+\delta)$ for some $\delta > 0$) must be fixed pointwise by T. Since T^n is also an interval exchange, the same property holds for points of any period. Thus there is an extreme dichotomy, in which an interval exchange again resembles a circle rotation: either T has no periodic points at all, or there is some $n \geqslant 1$ that pointwise fixes a non-trivial interval. In the latter case, it is clear that Lebesgue measure cannot be ergodic for T.

Theorem 5. *The following properties of the interval exchange T defined on the r intervals $(A_1, \dots, A_r)$ are equivalent.*

1. *T has no periodic points;*
2. *for any choice of $i_0, \dots, i_n$,*

$$\operatorname{diam}(A_{i_0} \cap T(A_{i_1}) \cap \cdots \cap T^n(A_{i_n})) \to 0 \tag{2.1}$$

as $n \to \infty$;

3. *the set of orbits of left end-points, $\mathcal{L}^*(T) = \bigcup_{k \geqslant 0} T^k \mathcal{L}(T)$, is dense in $[0, 1)$.*

Proof. (2) $\iff$ (3): For each $n \geqslant 1$ the intervals of the form in equation (2.1) form a partition ξ_n of $[0, 1)$ into half-open intervals; by construction each of the intervals in ξ_n is a union of intervals in ξ_{n+1}. The union of all the left end-points of intervals appearing in some ξ_n is exactly $\mathcal{L}^*(T)$, so conditions (2) and (3) are equivalent.

(2) $\implies$ (1): If (1) is not satisfied, then by the discussion above there is some $n \geqslant 1$ for which T^n acts as the identity on some interval of positive diameter, showing that (2) and (3) cannot hold.

(1) $\implies$ (3): Assume that (3) does not hold, so that the non-empty open set

$$O = [0, 1) \setminus \overline{\mathcal{L}^*(T)},$$

which is a union of open intervals, is invariant under T. It follows that T isometrically permutes these intervals in some way; since $[0, 1)$ can only contain a finite number of non-empty open intervals of given length, it follows that all the points of O are periodic points for T, so (1) does not hold. □

2.1.1 Mixing for Interval Exchanges

The simplest interval exchange is a circle rotation, which is known to not be mixing (indeed, it is not weak mixing). In this section we give an argument due to Katok to show that an interval exchange is never mixing.

Lemma 5. *Suppose that $([0, 1), \mathcal{B}, \mu, T)$ is an invertible measure-preserving transformation, and that we can find sequences (ξ_n), $(\mathbf{k}_n)$ and (s_n) with the following properties:*

1. *there is an S for which $s_n \leqslant S$ for all $n \geqslant 1$;*

2. *$\xi_n = \{A_1^{(n)}, \dots, A_{s_n}^{(n)}\}$ is a measurable partition of $[0,1)$ in the sense that $\mu\left(A_i^{(n)} \cap A_j^{(n)}\right) = 0$ for $i \neq j$ and $\mu\left(A_1^{(n)} \cup \dots \cup A_{s_n}^{(n)}\right) = 1$ for all $n \geqslant 1$;*
3. *$\mathbf{k}_n = \left(k_1^{(n)}, \dots, k_{s_n}^{(n)}\right)$ with $\min_{1 \leqslant j \leqslant s_n} k_j^{(n)} \to \infty$ as $n \to \infty$; and*
4. *for any $A \in \mathcal{B}$,*
$$\mu\left(A \triangle \left(\bigcup_{j=1}^{s_n} T^{k_j^{(n)}}\left(A_j^{(n)} \cap A\right)\right)\right) \to 0$$
as $n \to \infty$.

Then $([0,1), \mathcal{B}, \mu, T)$ is not mixing.

Proof. Let $A \in \mathcal{B}$. By the assumption (4),
$$\mu\left(A \cap \left(\bigcup_{j=1}^{s_n} T^{k_j^{(n)}}\left(A_j^{(n)} \cap A\right)\right)\right) \to \mu(A) \tag{2.2}$$
as $n \to \infty$; on the other hand
$$\begin{aligned}\mu\left(A \cap \left(\bigcup_{j=1}^{s_n} T^{k_j^{(n)}}\left(A_j^{(n)} \cap A\right)\right)\right) &= \mu\left(\bigcup_{j=1}^{s_n}\left(A \cap T^{k_j^{(n)}}\left(A_j^{(n)} \cap A\right)\right)\right) \\ &\leqslant \sum_{j=1}^{s_n} \mu\left(A \cap T^{k_j^{(n)}}\left(A_j^{(n)} \cap A\right)\right) \\ &\leqslant \sum_{j=1}^{s_n} \mu\left(A \cap T^{k_j^{(n)}}(A)\right).\end{aligned}$$

If $([0,1), \mathcal{B}, \mu, T)$ is mixing, then properties (1) and (3) imply that
$$\limsup_{n\to\infty} \mu\left(A \cap T^{k_j^{(n)}}(A)\right) \leqslant s\left(\mu(A)\right)^2,$$
which contradicts equation (2.2) whenever $\mu(A) < \frac{1}{s}$ (we may assume that μ is non-atomic, since an atomic invariant measure is never mixing). It follows that T cannot be mixing. □

Theorem 6 (Katok[3]). *If $T : [0,1) \to [0,1)$ is an interval exchange transformation, and μ is any Borel measure on $[0,1)$ preserved by T, then the measure-preserving system $([0,1), \mathcal{B}, \mu, T)$ is not mixing.*

[3] This result is due to Katok, and may be found in the notes by Katok, Sinaĭ and Stepin [20].

Proof. We may assume that the measure μ is non-atomic, and by taking the ergodic decomposition of μ [5, Th. 6.2] we may further assume without loss of generality that μ is ergodic for T. We will show that T is not mixing with respect to μ by exhibiting the combinatorial structure in the hypothesis of Lemma 5.

Write $T = T_0$, and assume that T_0 is an interval exchange without periodic points defined on the intervals $A_1^{(0)}, \dots, A_{k^{(0)}}^{(0)}$ partitioning $I^{(0)} = [0,1)$. We will construct a nested sequence

$$I^{(0)} \supset I^{(1)} \supset \cdots I^{(n)} \supset \cdots$$

of half-open intervals, and denote by T_n the transformation induced by T_0 on the interval $I^{(n)}$, with the first return time r_{A_n} defined as in [5, Eqn. 2.36] denoted by $r^{(n)}$ (see [5, Sect. 2.9]). The next lemma states that any such induced transformation is an interval exchange of constrained complexity.

Lemma 6. *If T is an interval exchange on k intervals and $[a,b) \subset [0,1)$, then the induced transformation $T_{[a,b)}$ is also an interval exchange on no more than $k+2$ intervals.*

Proof. In the notation of Definition 10, the left end-points of the intervals defining T are $0, \beta_1, \dots, \beta_{k-1}$. For each point $x \in D = \{a, b, \beta_1, \dots, \beta_{k-1}\}$, let

$$s(x) = \min\{s \geqslant 0 \mid T^{-s}(x) \in [a,b)\}$$

if any such s exist. There are no more than $k+1$ elements of

$$\{T^{-s(x)}(x) \mid x \in D\},$$

so they divide the interval $[a,b)$ into $k_1 \leqslant k+2$ half-open intervals denoted

$$A'_1, \dots, A'_{k_1}.$$

For each half-open interval A'_i let

$$r_i = \min\{r \geqslant 1 \mid T^r A'_i \cap [a,b) \neq \emptyset\}.$$

By construction, for $1 \leqslant j \leqslant r_i$ the transformation T^j is continuous on A'_i.

We claim that $T^{r_i} A'_i \subset [a,b)$, since if not there is some j with $1 \leqslant j < r_i$ for which the half-open interval would contain a point $x \in D$, which implies that $s(x) = j$, so the point $T^{-j}(x) = T^{-s(x)}(x)$ would lie in A'_i, contradicting the definition of A'_i. It follows that $r_{[a,b)}(x) = r_i$ for $x \in A'_i$; the transformation $T_{[a,b)}$ (which coincides with T^{r_i} on A'_i) is an interval exchange on the intervals $A'_1, \dots, A'_{k_1}$. □

Returning to the proof of Theorem 6, the transformation $T_i : I^{(i)} \to I^{(i)}$ is an interval exchange of a set of half-open intervals $I_j^{(i)}$ for $1 \leqslant j \leqslant k^{(i)}$, and we denote by $r_j^{(i)}$ the value of $r^{(i)}$ on $I_j^{(i)}$. Since the original transformation T is assumed to have no periodic points, we must have $k^{(i)} \geqslant 2$, so the function $r^{(i)}$ is not everywhere constant. Write $r_{j_0(i)}^{(i)} = \max\{r_j^{(i)} \mid 1 \leqslant j \leqslant k^{(i)}\}$, and define the interval $I^{(i)}$ to be $I_{j_0(i)}^{(i)}$. By definition of the return functions $r^{(i+1)}$, we have

$$r^{(i+1)}(x) \geqslant r_{j_0(i)}^{(i)} \tag{2.3}$$

for $i \geqslant 0$ and $x \in I^{(i+1)}$. Since $r^{(i)}$ is not everywhere constant,

$$r_{j_0(i+1)}^{(i+1)} > r_{j_0(i)}^{(i)}. \tag{2.4}$$

As a result of equations (2.3) and (2.4),

$$m^{(i)} = \min\{r_j^{(i)} \mid 1 \leqslant j \leqslant r^{(i)}\} \to \infty \tag{2.5}$$

as $i \to \infty$. By Kac's theorem [5, Th. 2.44], we have

$$m^{(i)} \operatorname{diam}(I^{(i)}) \leqslant \operatorname{diam}(I^{(0)}) = 1,$$

so equation (2.5) implies that

$$\operatorname{diam}(I^{(i)}) \to 0 \tag{2.6}$$

as $i \to \infty$.

For any $i \geqslant 0$ and j, $1 \leqslant j \leqslant k^{(i)}$, let $T_{i,j}$ denote the transformation of the half-open interval $I_j^{(i)}$ induced by T_0. By construction, each $T_{i,j}$ is an interval exchange defined on half-open intervals which we may denote by $I_{j,\ell}^{(i)} \subset I_j^{(i)}$ with $1 \leqslant \ell \leqslant k_j^{(i)}$.

By Lemma 6 and the construction of T_i and $T_{i,j}$, we have

$$k^{(i)} \leqslant k^{(0)} + 2 \tag{2.7}$$

and

$$k_j^{(i)} \leqslant k^{(0)} + 2. \tag{2.8}$$

For any $x \in I_j^{(i)}$, the return function $r_j^{(i)}$ corresponding to the induced transformation $T_{i,j}$ is bounded below by $r^{(i)}$, so that if $r_j^{(i)}(x) = r_{j,\ell}^{(i)}$ for $x \in I_{j,\ell}^{(i)}$, then

$$r_{j,\ell}^{(i)} \geqslant m^{(i)} \tag{2.9}$$

for all ℓ, $1 \leqslant \ell \leqslant k_j^{(i)}$.

Now for $i \geqslant 0$, $1 \leqslant j \leqslant k^{(i)}$, and $1 \leqslant \ell \leqslant k_j^{(i)}$, define

$$A_{j,\ell}^{(i)} = \bigcup_{p=0}^{k_j^{(i)}} T^p I_{j,\ell}^{(i)}$$

and

$$k_{j,\ell}^{(i)} = r_{j,\ell}^{(i)}.$$

Since the set $M_1 = \bigcup_{j,\ell \geqslant 1} A_{j,\ell}^{(i)}$ is invariant under T, the ergodicity of μ implies that $\mu(M_1) = 1$ or $\mu(M_1) = 0$. If $\mu(M_1) = 1$ then

$$\xi_i = \{A_{j,\ell}^{(i)} \mid 1 \leqslant j \leqslant k^{(i)}, 1 \leqslant \ell \leqslant k_j^{(i)}\}$$

is a partition of $([0,1), \mathcal{B}, \mu)$. If $\mu(M_1) = 0$, then the set $[0,1) \setminus M_1$ is a T-invariant union of a finite collection of half-open intervals, and we may carry out the construction above with $[0,1) \setminus M_1$ in the role of $I^{(0)}$. Since $T = T_0$ is an exchange of a finite collection of intervals we can, by repeating the construction several times if need be, eventually obtain a partition ξ_i as above so that the associated transformations T_i and $T_{i,j}$ possess the properties in equations (2.3)–(2.8).

We claim that the construction of the partitions ξ_i and the numbers $k_{j,\ell}^{(i)}$ satisfy the hypotheses of Lemma 5, completing the proof of Theorem 6. The inequalities (2.7) and (2.8) show that the number of elements $A_{j,\ell}^{(i)}$ in the partition ξ_i does not exceed $s = (r^{(0)} + 2)^2$. By the inequality (2.9) and equation (2.5), we must have

$$\min_{j,\ell} r_{j,\ell}^{(i)} \to \infty$$

as $i \to \infty$. Now suppose that $C \in \mathcal{B}$ and fix $\delta > 0$. Letting j and p range over the intervals $1 \leqslant j \leqslant k^{(i)}$ and $1 \leqslant p \leqslant r_j^{(i)}$, define sets

$$C_{\delta,i}^{+} = \{x \mid x \in T^p I_j^{(i)} \text{ and } \mu(C \cap T^p I_j^{(i)}) \geqslant (1-\delta)\mu(I_j^{(i)})\},$$

$$C_{\delta,i}^{-} = \{x \mid x \in T^p I_j^{(i)} \text{ and } \mu(C \cap T^p I_j^{(i)}) \leqslant \delta\mu(I_j^{(i)})\},$$

and

$$C_{\delta,i} = \{x \mid x \in T^p I_j^{(i)} \text{ and } \delta\mu(I_j^{(i)}) \leqslant \mu(C \cap T^p I_j^{(i)}) \leqslant (1-\delta)\mu(I_j^{(i)})\}.$$

By the Lebesgue density theorem [5, Th. A.24] and the fact that μ is not atomic, equation (2.6) implies that

$$\mu(C_{\delta,i}) \to 0, \tag{2.10}$$

$$\mu(C \triangle C_{\delta,i}^{+}) \to 0,$$

and

$$\mu\left(([0,1)\setminus C)\triangle C^{-}_{\delta,i}\right)\to 0 \tag{2.11}$$

as $i\to\infty$. We have

$$T^{k^{(i)}_{j,\ell}}I^{(i)}_{j,\ell}\subset\bigcup_{p=0}^{k^{(i)}_j}T^pI^{(i)}_j,$$

and the unions are pairwise disjoint for different values of j, so

$$\begin{aligned}\mu\left(C\triangle\left(\bigcup_{j,\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\right)\right)&\\ &=\sum_{j,p}\mu\left(T^pI^{(i)}_j\right)\mu\left(C\triangle\bigcup_{\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\big|T^pI^{(i)}_j\right)\\ &=\textstyle\sum^{+}_{\delta,i}+\sum^{-}_{\delta,i}+\sum_{\delta,i},\end{aligned}$$

where the three sums are taken over those j and p for which $T^pI^{(i)}_j$ is contained in the sets $C^{+}_{\delta,i}$, $C^{-}_{\delta,i}$ and $C_{\delta,i}$ respectively. Since $\sum_{\delta,i}\leqslant\mu(C_{\delta,i})$, so equation (2.10) implies that

$$\textstyle\sum_{\delta,i}\to 0 \tag{2.12}$$

as $i\to\infty$. Now

$$T^{k^{(i)}_{j,\ell}}I^{(i)}_{j,\ell}\cap T^pI^{(i)}_j=T^{k^{(i)}_{j,\ell}}\left(T^pI^{(i)}_{j,\ell}\right),$$

so that $T^pI^{(i)}_j\subset C^{+}_{\delta,i}$ implies that

$$\begin{aligned}\mu\left(C\triangle\bigcup_{\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\big|T^pI^{(i)}_j\right)&\leqslant\mu\left(T^pI^{(i)}_j\setminus\bigcup_{\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\big|T^pI^{(i)}_j\right)\\ &\qquad+\delta\\ &=\mu\left(T^pI^{(i)}_j\setminus\bigcup_{\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\right.\\ &\qquad\left.\cap T^pI^{(i)}_j\big|T^pI^{(i)}_j\right)+\delta\\ &=\mu\left(T^pI^{(i)}_j\setminus\bigcup_{\ell}T^{k^{(i)}_{j,\ell}}\left(I^{(i)}_{j,\ell}\cap C\right)\big|T^pI^{(i)}_j\right)\\ &\qquad+\delta. \qquad (2.13)\end{aligned}$$

Since $\bigcup_{\ell}T^{k^{(i)}_{j,\ell}+p}I^{(i)}_{j,\ell}=T^pI^{(i)}_j$, and the sets in the union are pairwise disjoint for different values of ℓ, and the set $C\cap T^pI^{(i)}_j$ differs from $T^pI^{(i)}_j$ by a set of measure no more than δ, we have

$$\bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(T^{p} I_{j,\ell}^{(i)} \cap C\right) \subset T^{p} I_{j}^{(i)}$$

and

$$\begin{aligned}
\mu\left(\bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(T^{p} I_{j,\ell}^{(i)} \cap C\right)\right) &= \sum_{\ell} \mu\left(T^{k_{j,\ell}^{(i)}}\left(T^{p} I_{j,\ell}^{(i)} \cap C\right)\right) \\
&= \sum_{\ell} \mu\left(T^{p} I_{j,\ell}^{(i)} \cap C\right) \\
&= \mu\left(T^{p} I_{j}^{(i)} \cap C\right) \geqslant (1-\delta) \mu\left(T^{p} I_{j}^{(i)}\right).
\end{aligned}$$

It follows that

$$\mu\left(T^{p} I_{j}^{(i)} \setminus \bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(I_{j,\ell}^{(i)} \cap C\right) \middle| T^{p} I_{j}^{(i)}\right) \leqslant \delta,$$

which with equation (2.13) gives

$$\mu\left(C \triangle \bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(I_{j,\ell}^{(i)} \cap C\right) \middle| T^{p} I_{j}^{(i)}\right) < 2\delta,$$

and therefore

$$\textstyle\sum_{\delta,i}^{+} < 2\delta. \tag{2.14}$$

Finally, we wish to estimate the size of $\sum_{\delta,i}^{-}$. We have

$$\begin{aligned}
\textstyle\sum_{\delta,i}^{-} &= \mu\left(\left(C \triangle\left(\bigcup_{j,\ell} T^{k_{j,\ell}^{(i)}}\left(I_{j,\ell}^{(i)} \cap C\right)\right)\right) \cap C_{\delta,i}^{-}\right) \\
&= \mu\left(C \cap C_{\delta,i}^{-}\right) + \sum_{j,p} \mu\left(T^{p} I_{j}^{(i)}\right) \mu\left(\bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(I_{j,\ell}^{(i)} \cap C\right) \middle| T^{p} I_{j}^{(i)}\right).
\end{aligned}$$

Since $C \cap C_{\delta,i}^{-} \subset ([0,1) \setminus C) \triangle C_{\delta,i}^{-}$, equation (2.11) implies that

$$\mu\left(C \cap C_{\delta,i}^{-}\right) \to 0 \tag{2.15}$$

as $i \to \infty$. Arguing as for the case $\sum_{\delta,i}^{+}$, we have

$$\mu\left(\bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(I_{j,\ell}^{(i)} \cap C\right) \middle| T^{p} I_{j}^{(i)}\right) = \mu\left(\bigcup_{\ell} T^{k_{j,\ell}^{(i)}}\left(T^{p} I_{j,\ell}^{(i)} \cap C\right) \middle| T^{p} I_{j}^{(i)}\right),$$

$$\bigcup_{\ell} T^{k_{j,\ell}^{(i)}} \left(T^p I_{j,\ell}^{(i)} \cap C \right) \subset T^p I_j^{(i)},$$

and

$$\mu \left(\bigcup_{\ell} T^{k_{j,\ell}^{(i)}} \left(T^p I_{j,\ell}^{(i)} \cap C \right) \right) = \mu \left(T^p I_j^{(i)} \cap C \right) \leqslant \delta\mu \left(T^p I_j^{(i)} \right),$$

so the second sum in $\sum_{\delta,i}^{-}$ is bounded above by δ.

Thus the bounds from (2.12), (2.14) and (2.15) together give

$$\mu \left(C \triangle \bigcup_{j,\ell} T^{k_{j,\ell}^{(i)}} \left(I_{j,\ell}^{(i)} \cap C \right) \right) \to 0$$

as $i \to \infty$, which proves the theorem by Lemma 5. □

2.1.2 Invariant Measures for Interval Exchanges

Clearly an interval exchange transformation may have many invariant measures if it has periodic points (in the extreme case, the identity map is an interval exchange). The next example shows that Lebesgue measure may not be the only invariant measure for an interval exchange transformation, even if it is assumed to not have any periodic points.

Example 5. Let $\alpha \in (0, \frac{1}{2})$ be irrational, and let T be the interval exchange defined by the data $\boldsymbol{\alpha} = (\alpha, \frac{1}{2} - \alpha, \alpha, \frac{1}{2} - \alpha)$ and $\pi = (12)(34) \in \Sigma_4$. Then

$$\boldsymbol{\beta} = (0, \alpha, \tfrac{1}{2}, \tfrac{1}{2} + \alpha, 1)$$

and

$$\boldsymbol{\beta}^{\pi} = (0, \tfrac{1}{2} - \alpha, \tfrac{1}{2}, 1 - \alpha, 1),$$

and the resulting map is show in Figure 2.3; it is clear that this map has no periodic points since α is irrational.

Thus the interval $[0, \frac{1}{2})$ is invariant under T, so Lebesgue measure is not ergodic for T. It follows that there must be at least two ergodic invariant measures for T. Indeed, there are two ergodic measures given by normalized Lebesgue measure restricted to $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$.

Theorem 7 (Keane[4]). *Suppose that T is an interval exchange without periodic points defined on r intervals, and let μ be a T-invariant Borel probability*

[4] Corollary 1 is due to Keane [23]; our treatment is taken from the monograph of Cornfeld, Fomin, and Sinaĭ [4], who mention that this was also shown by Zemlyakov.

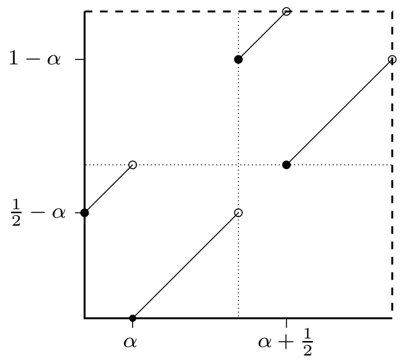

Fig. 2.3: An interval exchange with more than one ergodic invariant measure.

measure on $[0,1)$. *Then* $[0,1)$ *is a union of no more than* r *subsets of positive measure invariant under* T.

Proof. The measure μ is non-atomic since there are no periodic points. Write

$$\mathcal{H} = \{f \in L^2_\mu \mid U_T d = f\}$$

for the subspace of invariant functions. Assume that $[0,1)$ is partitioned into k sets $A_1,\dots,A_k$, invariant under T, of positive μ-measure, and define

$$\mathcal{H}' = \{f \in L^2_\mu \mid f \text{ is a.e. constant on each } A_i\}.$$

Clearly $\mathcal{H}' \subset \mathcal{H}$, so the dimension of $\mathcal{H}$ as a complex vector space is no smaller than k. Thus the theorem will follow if we can establish that

$$\dim(\mathcal{H}) \leqslant r. \tag{2.16}$$

For any function $h \in L^2_\mu$, let

$$\mathcal{H}(h) = \overline{\ll U_T^n h \mid n \in \mathbb{Z} \gg}$$

be the closure of the linear span of the orbit of h under U_T and write P_T for the orthogonal projection onto the space $\mathcal{H}$ of invariant functions. Write

$$h^\perp = h - P_T h,$$

and notice that for any $h \in L^2_\mu$ the space $\mathcal{H}(P_T h)$ is one-dimensional (if h has a non-trivial projection onto $\mathcal{H}$) or zero-dimensional. By the mean ergodic theorem [5, Th. 2.21],

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n h \to P_T h$$

as $N \to \infty$, so $P_T h \in \mathcal{H}(h)$. It follows that $h^\perp = h - P_T h \in \mathcal{H}(h)$, and thus $\mathcal{H}(h^\perp) \subset \mathcal{H}(h)$. Now U_T is unitary and by construction $h^\perp \perp \mathcal{H}$, so $\mathcal{H}(h^\perp) \perp \mathcal{H}$ and $\mathcal{H}(h^\perp) \perp \mathcal{H}(P_T h)$. Thus

$$\mathcal{H}(h) = \mathcal{H}(h^\perp) \oplus \mathcal{H}(P_T h).$$

Assume now that we have found functions $h_1, \dots, h_p \in L^2_\mu$ with the property that

$$L^2_\mu = \mathcal{H}(h_1) + \cdots + \mathcal{H}(h_p). \tag{2.17}$$

Then

$$\mathcal{H} = P_T h_1 + \cdots + P_T h_p,$$

so $\dim(\mathcal{H}) \leqslant p$. Thus equation (2.16) will follow if we can find r functions satisfying equation (2.17).

Let $h_1 = \chi_{A_1}, \dots, h_r = \chi_{A_r}$, where the A_i are the intervals exchanged by T. By Theorem 5(2) and the fact that μ is non-atomic, the linear span of the indicator functions of the sets of the form

$$A_{i_0,\dots,i_m} = A_{i_0} \cap T A_{i_1} \cap \cdots \cap T^m A_{i_m}$$

(for $m \geqslant 1$ and $i_k \in \{1, \dots, r\}$ for each k) is dense in L^2_μ. Thus it is enough to show that any such function can be written in the form

$$\sum_{i=1}^{r} \sum_{k=0}^{m} c_{ik} U_T^{-k} \chi_{A_i} \in \mathcal{H}(\chi_{A_1}) \oplus \mathcal{H}(\chi_{A_2}) \oplus \cdots \oplus \mathcal{H}(\chi_{A_r}). \tag{2.18}$$

We prove equation (2.18) by induction on the parameter m. For $m = 0$ this is clear, so assume that equation (2.18) holds for m. Then

$$\chi_{A_{i_0,\dots,i_{m+1}}} = \chi_{A_{i_0}} \chi_{T A_{i_1} \cap \cdots \cap T^{m+1} A_{i_{m+1}}} = \chi_{A_{i_0}} U_T^{-1} \chi_{A_{i_1,\dots,i_{m+1}}}.$$

By the inductive hypothesis, the function $\chi_{A_{i_1,\dots,i_{m+1}}}$ can be put in the form of equation (2.18). Consider the intervals A_{i_0} as i_0 varies in their order in the interval $[0,1)$ (that is, in the order $i_0 = 1, i_1 = 2, \dots$). Now

$$A_{i_1} \cap T A_{i_2} \cap \cdots \cap T^m A_{i_{m+1}} \subset A_{i_1},$$

so the sets of the form $T A_{i_1} \cap T^2 A_{i_2} \cap \cdots \cap T^{m+1} A_{i_{m+1}}$ are half-open intervals, which we shall also consider in the natural order from left to right in $[0,1)$.

Suppose first that $i_0 = 1$. If

$$T A_{i_1} \cap T^2 A_{i_2} \cap \cdots \cap T^{m+1} A_{i_{m+1}} \subset A_1$$

then

$$\chi_{A_{1,i_1,\dots,i_{m+1}}} = U_T^{-1} \chi_{A_{i_1,\dots,i_{m+1}}},$$

which is in the desired form. If

$$TA_{i_1} \cap T^2 A_{i_2} \cap \cdots \cap T^{m+1} A_{i_{m+1}} \not\subset A_1 \tag{2.19}$$

then

$$\chi_{A_{1,i_1,\dots,i_{m+1}}} = \chi_{A_1} - \sum \chi_{A_{1,i'_1,\dots,i'_{m+1}}},$$

where the sum is taken over all choices of tuples $(i'_1, \dots, i'_{m+1})$ for which

$$TA_{i'_1,\dots,i'_{m+1}} \subset A_1.$$

Using the case equation (2.19), each of these summands can be put in the form desired, proving the result by induction in the case $i_0 = 1$.

Now assume that $i_0 = 2$ (the case $i_0 > 2$ follows by a simple modification of the argument). There are again two possibilities. If

$$TA_{i_1} \cap T^2 A_{i_2} \cap \cdots \cap T^{m+1} A_{i_{m+1}} \subset A_1 \cup A_2,$$

then

$$\chi_{A_{2,i_1,\dots,i_{m+1}}} = U_T^{-1} \chi_{A_{i_1,\dots,i_{m+1}}} - \chi_{A_{1,i_1,\dots,i_{m+1}}},$$

which may be put into the desired form using the case $i_0 = 1$. If

$$TA_{i_1} \cap T^2 A_{i_2} \cap \cdots \cap T^{m+1} A_{i_{m+1}} \not\subset A_1 \cup A_2,$$

then there are two possibilities. If $TA_{i_1,\dots,i_{m+1}} \cap (A_1 \cup A_2) = \emptyset$ then

$$\chi_{A_{2,i_1,\dots,i_m}} = 0.$$

If $TA_{i_1,\dots,i_{m+1}} \cap (A_1 \cup A_2) \neq \emptyset$ then $\chi_{A_{2,i_1,\dots,i_{m+1}}}$ can be written in the form

$$U_T^{-1} \chi_{A_{i_1,\dots,i_{m+1}}} - \sum \chi_{A_{1,i'_1,\dots,i'_{m+1}}} - \sum \chi_{A_{2,i''_1,\dots,i''_{m+1}}},$$

where the summands each correspond to a case already considered. This proves equation (2.18) and hence the theorem. □

Corollary 1. *If T is an interval exchange on r intervals, then T has no more than r distinct ergodic invariant measures.*

Proof. If $\mu_1, \dots, \mu_p$ are ergodic T-invariant measures then, by the ergodic theorem, we may find disjoint T-invariant sets $B_1, \dots, B_p$ such that

$$\mu_i(B_j) = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

If $\mu = \frac{1}{p} \sum_{i=1}^{p} \mu_i$ of the measures, then by construction μ is T-invariant and $\mu(A_i) = \frac{1}{p} > 0$, so the statement follows from Theorem 7. □

2.2 Cutting and Stacking

Relaxing some of the conditions imposed in the definition of interval exchange transformations (Definition 10), and in particular permitting there to be infinitely many intervals, gives a larger class of transformations[5]. The extent of this increased flexibility is shown in Theorem 8, which should be contrasted with Theorem 6. Cutting and stacking is motivated by the Kakutani–Rokhlin lemma [5, Lem. 2.45], which says that up to an arbitrarily small part of the space X, a measure-preserving systems $(X, \mathcal{B}, \mu, T)$ with μ non-atomic has a geometrical presentation as a Rokhlin tower. Notice that a geometrical presentation of a transformation in this way is implicitly using the material from [5, Sect. A.6], in that it presents measurable subsets of a non-atomic Borel probability space as intervals of the real line. This identification will become explicit in Theorem 8.

The construction of a measure-preserving transformation by cutting and stacking resembles Definition 10 but organizes the combinatorial information differently[6].

2.2.1 Cutting and Stacking Transformations

Let $(X, \mathcal{B}, \mu)$ be a Borel probability space, and assume that μ is non-atomic. A *stack* $\mathsf{S} = \mathsf{stack}(h, w)$ of *height* h and *width* w is a collection of h intervals each of length w. Such a stack implicitly defines a transformation on all levels of the stack apart from the top level by requiring that the stack behave like a Rokhlin tower for the transformation (see Figure 2.4).

The support of the stack S, denoted $|\mathsf{S}|$, is the union of all the intervals in S. A *cutting* of a stack $\mathsf{stack}(h, w)$ is defined whenever we are given positive widths $w_1, \dots, w_b$ with $w_1 + \cdots + w_b = w$; it comprises b disjoint stacks S_i, obtained by dividing each level of S into b intervals of widths $w_1, \dots, w_b$ (read from left to right) and then defining S_i to be the stack formed by the ith interval at each level, as illustrated in Figure 2.5, so that $|\mathsf{S}| = \bigcup_{i=1}^{b} |\mathsf{S}_i|$.

Given stacks $\mathsf{S}_1 = \mathsf{stack}(h_1, w)$ and $\mathsf{S}_2 = \mathsf{stack}(h_2, w)$ of the same width, we may concatenate them to form $\mathsf{S}_1\mathsf{S}_2 = \mathsf{stack}(h_1 + h_2, w)$ by placing the h_2 levels of S_2 above the h_1 levels of S_1 and treating them as a single stack of height $h_1 + h_2$.

A general cutting and stacking construction starts with a finite collection of stacks. The final transformation is built up in infinitely many stages, each of which involves cutting the stacks into smaller stacks and then adding some

[5] A related but different family of constructions, in particular better adapted to building smooth models with controlled measurable properties, may be found in the notes of Katok [19].

[6] We follow the notation and presentation of Arnoux, Ornstein and Weiss [1] in this section.

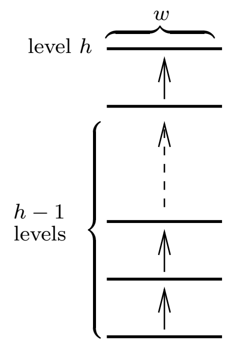

Fig. 2.4: Part of a transformation defined by a stack $\mathsf{S} = \mathsf{stack}(h, w)$.

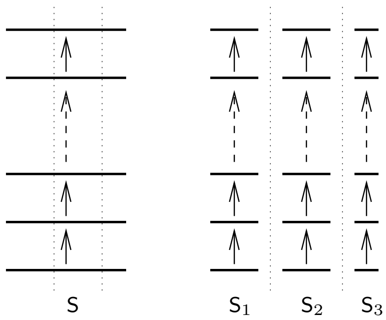

Fig. 2.5: Cutting a stack S into three stacks.

new levels to the resulting stacks (not necessarily adding the same number of levels to each). Two conditions are needed to make this infinite process define a measure-preserving transformation on a Borel probability space: the sum of all the widths of all the intervals used must be finite (and we can then always arrange for it to be 1 by normalizing), and the maximum width of any of the stacks at the nth stage must converge to zero. Once both of these conditions are satisfied, the end product is automatically a Lebesgue measure-preserving ergodic transformation[7] of an interval (which may be normalized to be $[0, 1)$). Notice that the final transformation is not defined as a limit of other transformations: there is one single transformation, and the infinite sequence of stages defines it on more and more of the space. Once the transformation is defined on a part of the space, that definition is not subsequently changed at a later stage. The residual set on which it is not defined shrinks in measure to zero as $n \to \infty$.

[7] We only need the measure-preserving part of this statement. Ergodicity and many other features and examples of cutting and stacking may be found in the paper of Shields [34]; Friedman [11] gives an accessible introduction.

Definition 11. A *cutting and stacking* construction is the transformation of an interval defined by starting with a finite collection of stacks

$$\mathsf{S}_i = \mathsf{stack}(h_i, w_i)$$

for $1 \leqslant i \leqslant a$, and then iterating the following procedure:

1. cut each stack S_i into b_i stacks of width w_{ij}, where $\sum_{j=1}^{b_i} w_{ij} = w_i$, resulting in new stacks $\mathsf{S}_{ij} = \mathsf{stack}(h_i, w_{ij})$;
2. add new levels to the stacks S_{ij} to form $\mathsf{S}_{ij}^{-}\mathsf{S}_{ij}\mathsf{S}_{ij}^{+}$, where
$$\mathsf{S}_{ij}^{-} = \mathsf{stack}(f_{ij}, w_{ij})$$
and
$$\mathsf{S}_{ij}^{+} = \mathsf{stack}(g_{ij}, w_{ij})$$
for various $f_{ij}, g_{ij} \geqslant 0$; and
3. partition $\{(i,j) \mid 1 \leqslant i \leqslant a, 1 \leqslant j \leqslant b_i\}$ into sets $I_1, \dots, I_{a'}$ with the property[8] that w_{ij} takes a single value w'_k on I_k, and concatenate in order the stacks S_{ij} with $(i,j) \in I_k$ into a new stack S'_k, before relabeling the resulting stacks to repeat.

Notice that the height of the stack S'_k is $\sum_{(i,j)\in I_k}(f_{ij} + h_i + g_{ij})$. The choices that determine the transition from one stage to the next are the set of cutting widths w_{ij}, the heights f_{ij} of the stacks S_{ij}^{-} added below, the heights g_{ij} of the stacks S_{ij}^{+} added above, and the choice of the aggregated sets I_k.

Example 6. Construct a cutting and stacking transformation as follows. Begin with the intervals $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$ labeled 0 and 1; at this stage the transformation T (which will eventually be defined on all of $[0,1)$) is not defined anywhere. Pass to the second stage by cutting each stack into four equal intervals, and stacking them in pairs in such a way that all four pairs of binary digits appear vertically. Choice is involved in doing this, and one possible choice is illustrated in Figure 2.6(a). At this stage the transformation is defined on half the space, and the graph of the part of the transformation defined is shown in Figure 2.6(b).

At the next stage, each of the second stage columns of width $\frac{1}{8}$ is cut into 8 stacks of width $\frac{1}{64}$, which are then stacked again in such a way that all possible binary blocks of length 4 appear in the stacks when viewed vertically. At this point the transformation is defined on all but $\frac{1}{4}$ of the space.

After n stages there are $2^{2^{n-1}}$ stacks, each of which has height 2^{n-1} and width $1/2^{n-1}2^{2^{n-1}}$, with a bijection between the labels on the columns and the binary sequences of length 2^{n-1}, with the transformation defined on all

[8] The sets I_k need not be maximal with respect to this property, and the ordering of $I_1, \dots$ is arbitrary.

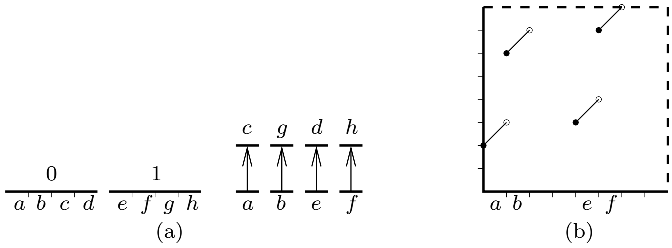

Fig. 2.6: The first stage, and the corresponding graph.

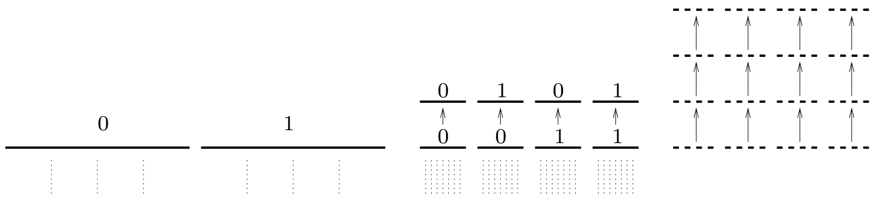

Fig. 2.7: The first three stages of Example 6.

but the top layer, which has measure $1/2^{n-1}$. At each stage of the construction every point $x \in [0,1)$ belongs to some stack and, since the widths of the stack shrink to zero, almost every $x \in [0,1)$ eventually belongs to a level other than the top level of some stack, and therefore the transformation T is defined almost everywhere on $[0,1)$.

We may define a map $\theta : [0,1) \to \{0,1\}^{\mathbb{Z}}$, defined almost everywhere, by $T^n x \in A_{\theta(x)_n}$ where $A_0 = [0, \frac{1}{2})$ and $A_1 = [\frac{1}{2}, 1)$. This map is an isomorphism between the map constructed and the Bernoulli shift on two symbols with $(\frac{1}{2}, \frac{1}{2})$-measure.

Thus (in particular) it is possible for a transformation defined by cutting and stacking to be mixing. The next result shows that much more is true. Recall from [5, Exercise 2.9.2] that an invertible measure-preserving system $(X, \mathcal{B}, \mu, T)$ is called *aperiodic* if $\mu\left(\{x \in X \mid T^k(x) = x\}\right) = 0$ for all $k \in \mathbb{Z} \setminus \{0\}$.

Theorem 8 (Arnoux, Ornstein, Weiss). *Let $(X, \mathcal{B}, \mu, T)$ be an invertible aperiodic measure-preserving system on a Borel probability space. Then there is a transformation $([0,1), \mathcal{B}_{[0,1)}, S, m)$, obtained by cutting and stacking, that is measurably isomorphic to $(X, \mathcal{B}, \mu, T)$.*

Here $\mathcal{B}_{[0,1)}$ denotes the Borel σ-algebra on $[0,1)$ and m denotes Lebesgue measure. This result means that, up to measurable isomorphism, every measure-preserving system may be obtained by cutting and stacking[9]. If

[9] Like the Jewett–Krieger theorem discussed in [5, p. 119], this result in some ways needs to be interpreted in reverse: it does not assert that the class of aperiodic measure-preserving

we are happy to ignore a small part of the space, the Kakutani–Rokhlin lemma [5, Lem. 2.45] makes Theorem 8 trivial. However, part of the meaning of the Kakutani–Rokhlin lemma is that any aperiodic measure-preserving system can be presented in a way that places all of the interesting dynamics onto a set of arbitrarily small measure. Thus Theorem 8 requires an infinite sequence of Rokhlin towers, and a check that they can be constructed in such a way that the combinatorics of the transition from one tower to the next is no more complicated than that allowed by the cutting and stacking construction. In order to do this a slight refinement of the Kakutani–Rokhlin lemma is needed.

For a measure-preserving system, write $\mathsf{R} = \mathsf{tower}(B, n, \epsilon)$ for a Rokhlin tower with base $B \in \mathcal{B}$ and residual set of measure exactly ϵ. Thus

$$B, TB, \dots, T^{n-1}B$$

are disjoint sets, and $\mu\left(\bigcup_{i=0}^{n-1} T^i B\right) = 1 - \epsilon$. As with a stack, we write $|\mathsf{R}|$ for the support $\bigcup_{i=0}^{n-1} T^i B$ of the Rokhlin tower.

Lemma 7. *Let $\mathsf{X} = (X, \mathcal{B}, \mu, T)$ be an aperiodic measure-preserving system. Then we may find a sequence of Rokhlin towers $\mathsf{R}_i = \mathsf{tower}(B_i, n_i, \epsilon_i)$ for X with the property that for any $i \geqslant 1$,*

$$|\mathsf{R}_i| \subset |\mathsf{R}_{i+1}| \setminus \left(B_{i+1} \cup T^{n_{i+1}-1} B_{i+1}\right) \tag{2.20}$$

and for which $\mu\left(|\mathsf{R}_i|\right) \to 1$ as $i \to \infty$.

Proof. Choose a sequence $\epsilon_i \searrow 0$ as $i \to \infty$, and numbers

$$n_1 < n_2 < \cdots$$

with $\sum_{i=\ell}^{\infty} \frac{n_i}{n_{i+1}} < \epsilon_\ell / 4$ for all $\ell \geqslant 1$.

Start with Rokhlin towers

$$\mathsf{R}_1^1 = \mathsf{tower}(B_1^1, n_1, \tfrac{1}{2}\epsilon_1)$$

and

$$\mathsf{R}_2^1 = \mathsf{tower}(B_2^1, n_2, \tfrac{1}{2}\epsilon_2).$$

We wish to modify R_1^1 so as to ensure equation (2.20) without changing it too much. If $|\mathsf{R}_1^1| \cap \left(B_2^1 \cup T^{n_2-1} B_2^1\right) \neq \emptyset$ (one of the ways in which this may happen is illustrated in Figure 2.8), then we delete from the tower R_1^1 a whole column of height n_1 (that is, a set of the form $\bigcup_{i=0}^{n_1-1} T^i A$ for some $A \subset B_1^1$) so remove that intersection. This results in a new Rokhlin tower $\mathsf{R}_1^2 = \mathsf{tower}(B_1^2, n_1, \epsilon_1^2)$ with

---

systems has a simple structure up to isomorphism; instead it means that the class of transformations constructed by cutting and stacking has a very complex structure.

$$\left(B_2^1 \cup T^{n_2-1}B_2^1\right) \cap |\mathsf{R}_1^2| = \emptyset.$$

Since we have deleted a column of height n_1, in each level deleting a set of measure no larger than $\frac{1}{n_2}$, we also have

$$\mu\left(|\mathsf{R}_1^2|\right) \geqslant \mu\left(|\mathsf{R}_1^1|\right) - 2\frac{n_1}{n_2}.$$

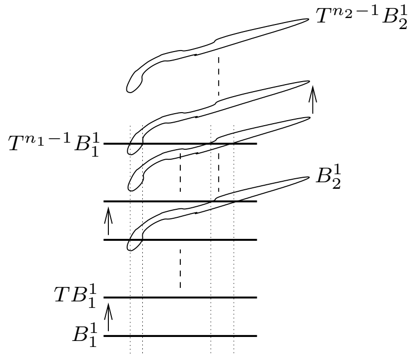

Fig. 2.8: A Rokhlin tower R_2^1 that does not contain R_1^1 in its interior.

We repeat: for a Rokhlin tower $\mathsf{R}_3^1 = \mathsf{tower}(B_3^1, n_3, \frac{1}{2}\epsilon_3)$ and modify R_2^1 to some tower R_2^2 by deleting whole columns of height n_2; this at most removes a set of measure $2\frac{n_2}{n_3}$ from R_1^2 to give a new tower R_1^3. After going through an entire sequence of towers R_i^1 we end up with a sequence of towers $\mathsf{R}_1^\infty, \mathsf{R}_2^\infty, \dots$ with the desired properties: each $\mathsf{R}_i^\infty = \mathsf{tower}(B_i, n_i, \epsilon_i')$ with $\epsilon_i' \leqslant \epsilon_i$ and with equation (2.20). $\square$

PROOF OF THEOREM 8. Let $\xi_1 \subset \xi_2 \subset \cdots$ be a sequence of finite measurable partitions of X with

$$\bigvee_{i=1}^{\infty} \xi_i \underset{\mu}{=} \mathcal{B}.$$

Using Lemma 7, choose a sequence of towers $\mathsf{R}_1 = \mathsf{tower}(B_1, n_1, \epsilon_1), \mathsf{R}_2, \dots$ with equation (2.20), $n_i \nearrow \infty$ and with $\epsilon_i \searrow 0$ as $i \to \infty$. We use these towers to construct a cutting and stacking transformation isomorphic to $(X, \mathcal{B}, \mu, T)$.

For the initial collection of stacks, take the tower R_1 and subdivide it into columns that are *pure* with respect to the partition ξ_1: that is, write B_1 as the disjoint union of sets $B_{1,1} \sqcup \cdots \sqcup B_{1,a_1}$ where each $B_{1,i}$ is maximal with the property that the sets $B_{1,i}, TB_{1,i}, \dots, T^{n_1-1}B_{1,i}$ lie in a single set in ξ_1. This defines stacks $\mathsf{S}_{1,i} = \mathsf{stack}(n_1, \mu(B_{1,i}))$ for $1 \leqslant i \leqslant a_1$. Each of these stacks corresponds to a Rokhlin tower $\mathsf{tower}(B_{1,i}, n_1)$, and each level in $\mathsf{S}_{1,i}$ can be labeled with the label of the element of ξ_1 to which the corresponding level of $\mathsf{tower}(B_{1,i}, n_1, \epsilon_{1,i})$ belongs.

Now divide the tower R_2 into maximal columns pure with respect to ξ_2 by writing $B_2 = B_{2,1} \sqcup \cdots \sqcup B_{2,a_2}$. The requirement that the stacks in the second stage of the cutting and stacking construction correspond to the towers $\mathsf{tower}(B_{2,i}, n_i, \epsilon_{2,i})$ determines the cutting widths, the index sets, and the number of new labels (some choice is involved in assigning new levels between successive R_1-columns in the towers based on $B_{2,i}$ to the top of the lower or the bottom of the upper R_1-column). This is possible since by construction $|\mathsf{R}_1|$ lies inside the interior of R_2. As before, the stacks can be labeled using the labels inherited from ξ_2 seen in R_2.

Repeating this procedure infinitely often gives a cutting and stacking construction that defines a transformation $S : [0,1) \to [0,1)$ almost everywhere, with partitions $\xi_1' \subset \xi_2' \subset \cdots$ in such a way that the process defined by ξ_i under T is measurably isomorphic to the process defined by ξ_i' under S, and so that $\bigvee_{i=0}^{\infty} \xi_i'$ generates all of $\mathcal{B}_{[0,1)}$. This gives an isomorphism between T and S as required. □

2.2.2 Chacon's Transformation

In this section we describe a simple example of a measure-preserving system that is weak-mixing but not mixing[10].

Let $a_0 = 0$, let $a_n = \sum_{i=1}^{n} \frac{2}{3^{i+1}}$, and let $A_n = [\frac{2}{3} + a_{n-1}, \frac{2}{3} + a_n)$ for $n \geqslant 1$. The sets A_n are disjoint, and $\bigcup_{n=1}^{\infty} A_n = [\frac{2}{3}, 1)$. We define a transformation T on $[0,1)$ by cutting and stacking as follows.

Start with the interval $[0, \frac{2}{3})$, and cut it into three intervals of equal length,

$$[0, \tfrac{2}{3}) = [0, \tfrac{2}{9}) \sqcup [\tfrac{2}{9}, \tfrac{4}{9}) \sqcup [\tfrac{4}{9}, \tfrac{2}{3}).$$

Add the set $A_1 = [\frac{2}{3}, \frac{8}{9})$ above the middle interval, and then stack the four intervals up as shown in Figure 2.9(a); the part of the transformation defined at this stage is shown in Figure 2.9(b).

At the next stage we start with the stack in the center of Figure 2.9, of height 4 and width $\frac{2}{9}$, cut it into three stacks each of width $\frac{2}{27}$, add the set $A_2 = [\frac{24}{27}, \frac{26}{27})$ above the middle stack, and then place the middle stack (of height 5) above the left-most stack, and finally place the right-most stack onto the top of the middle stack. The result is a stack with base $[0, \frac{2}{27})$ and height 13, which defines a map on all but $\frac{1}{27}$th of the space, as illustrated in Figure 2.10.

[10] As mentioned in [5, p. 67], the earliest construction of such a system seems to be Gaussian; this cutting and stacking construction is related to one found by von Neumann and Kakutani in 1940 but only published much later by Kakutani [18] and modified by Chacon [3]; we follow Friedman [11] for this section.

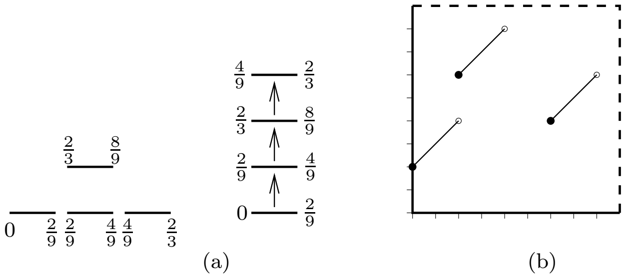

Fig. 2.9: The first stage in constructing the Chacon map.

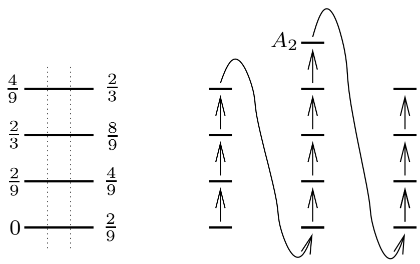

Fig. 2.10: The second stage in constructing the Chacon map.

Notice that the map defined by Figure 2.10 extends that defined in Figure 2.9: it does not change the values where they are already defined, but extends the map by defining it on the interval $[\frac{14}{27}, \frac{16}{27})$.

At the nth stage we have a stack of some height h_n and with base $[0, \frac{2}{3^n})$. We cut this into three stacks of equal width, and place the set A_n above the middle stack. Stacking in the same way (middle onto the left, and then right-most onto the middle) results in a stack of height $h_{n+1} = 3h_n + 1$ with base $[0, \frac{2}{3^{n+1}})$.

If $x \in [0, \frac{2}{3})$ then x is not on the top level of the nth stack for a sufficiently large n, so $T(x)$ is defined. If $x \in [\frac{2}{3}, 1)$, then $x \in A_n$ for some unique n, so $T(x)$ is also defined. Thus we have defined a Lebesgue measure-preserving map $T : [0, 1) \to [0, 1)$, which we call the Chacon map.

Theorem 9. *The Chacon map T is weak-mixing, but not mixing, with respect to Lebesgue measure.*

Proof. Let $A = [0, \frac{2}{9})$, so that A is a union of disjoint levels in the stack constructed at level n for each $n \geqslant 2$. If B is one of those levels, then (see Figure 2.10 for the $n = 2$ case of this) $T^{h_n} B_1 = B_2$, so

$$m\left(T^{h_n} B \cap B\right) \geqslant \tfrac{1}{3} m(B).$$

Since A is a disjoint union of these levels, we must also have

$$m\left(T^{h_n}A\cap A\right)\geqslant \tfrac{1}{3}m(A)=\tfrac{2}{27},$$

which contradicts mixing since $\mu(A)^2=\frac{4}{81}<\frac{2}{27}$ and $h_n\to\infty$ as $n\to\infty$.

We will use the characterization of weak-mixing from [5, Th. 2.36(5)], but before doing so need to confirm that T is ergodic with respect to Lebesgue measure. To this end, let A be a measurable subset of $[0,1)$ with $m(A)>0$. By the Lebesgue density theorem we may a find a point $a\in A$ with the property that for any $\epsilon>0$ there is some $\delta>0$ such that if $I\ni a$ is any interval with $m(I)<\delta$, then $m(A\cap I)\geqslant(1-\epsilon)m(I)$. Fix $\epsilon>0$ and choose n large enough to ensure that T has been defined on a subset of $[0,1)$ with measure at least $1-\delta$ and that x lies on the rth level J in the nth stack. Since T acts by translation on the levels,

$$m\left(T^i(A\cap J)\right)=m(A\cap J)>(1-\epsilon)m(J)$$

whenever T^iJ is still a level in that stack. It follows that

$$\begin{aligned}
m\left(\bigcup_{j\in\mathbb{Z}}T^jA\right)&\geqslant m\left(\bigcup_{j\in\mathbb{Z}}T^j(A\cap J)\right)\\
&\geqslant\sum_{j=1-r}^{h_n-r}m\left(T^j(A\cap J)\right)\\
&>\sum_{j=1-r}^{h_n-r}(1-\epsilon)m(J)=(1-\epsilon)h_nm(J)=(1-\epsilon).
\end{aligned}$$

Since $\epsilon>0$ was arbitrary, we conclude that the orbit $\bigcup_{j\in\mathbb{Z}}T^jA$ has full measure, so T is ergodic.

Now let f be a measurable function on $[0,1)$ with $U_Tf=\mathrm{e}^{\mathrm{i}a}f$ for some a (that is, an eigenvalue for T). Then $|f|$ is invariant under T so, by ergodicity, $|f|$ is a constant almost everywhere. By normalizing we may assume that $|f|=1$ almost everywhere, so that

$$f(x)=\mathrm{e}^{\mathrm{i}\theta(x)}$$

for some measurable map $\theta:[0,1)\to[0,2\pi)$. By Lusin's theorem (see [5, Th. A.20]), for any $\epsilon>0$ we may find a closed set F with $m(F)>1-\epsilon$ with the property that θ is uniformly continuous on F: given $\epsilon'>0$ there is a $\delta'>0$ such that $x_1,x_2\in F$ and $|x_1-x_2|<\delta'$ implies that $|\theta(x_1)-\theta(x_2)|<\epsilon'$. By Lebesgue's density theorem there is a point $x\in F$ so that F has Lebesgue density 1 at x. Let $\epsilon''>0$ be given; choose n large enough to ensure that $\frac{2}{3^n}<\delta'$ and with the property that there is a level J in the nth stack containing x with $m(J \cap F) > (1-\epsilon'')m(F)$. If ϵ'' is small enough, then we can find points $x_1, x_2, x_3 \in J$ arranged as in Figure 2.11.

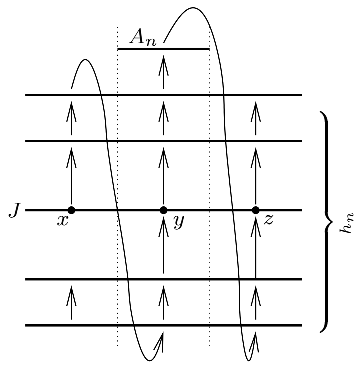

Fig. 2.11: The nth stack in Chacon's map.

By the construction of the map T we must then have

$$\mathrm{e}^{\mathrm{i}\theta(y)} = f(y) = \mathrm{e}^{\mathrm{i}h_n a}\mathrm{e}^{\mathrm{i}\theta(x)}$$

and

$$\mathrm{e}^{\mathrm{i}\theta(z)} = f(z) = \mathrm{e}^{\mathrm{i}(h_n+1)a}\mathrm{e}^{\mathrm{i}\theta(x)},$$

so that

$$\theta(y) = h_n a + \theta(x) \tag{2.21}$$

and

$$\theta(z) = (h_n + 1)a + \theta(y), \tag{2.22}$$

where both equalities are meant modulo 1. By our choice of n, $|x - y| < \delta'$ and $|z - y| < \delta'$, so that taking equation (2.21) from equation (2.22) gives

$$|a + \theta(y) - \theta(x)| = |\theta(z) - \theta(y)| < \epsilon'$$

and hence

$$|a| \leqslant \epsilon' + |\theta(y) - \theta(x)| < 2\epsilon'.$$

Since $\epsilon' > 0$ was arbitrary, it follows that $a = 0$, and therefore T is weakly mixing. □

Exercise: Use the proof of Theorem 8 to show that any aperiodic measure-preserving system on a Borel probability space is measurably isomorphic to an interval exchange transformation $T : [0, 1) \to [0, 1)$ on infinitely many intervals $[t_{j-1}, t_j)$ with the properties that

1. $\lim_{j\to\infty} t_j = 1$;

2. T is a translation $x \mapsto x + a_j$ on each $[t_{j-1}, t_j)$;
3. the only accumulation point of the set $\{t_{j-1} + a_j\} \cup \{t_j + a_j\}$ is 1; and
4. T is 1-to-1.

References

1. P. Arnoux, D. S. Ornstein, and B. Weiss, 'Cutting and stacking, interval exchanges and geometric models', *Israel J. Math.* **50** (1985), no. 1-2, 160–168.
2. A. Avila and G. Forni, 'Weak mixing for interval exchange transformations and translation flows', *Ann. of Math. (2)* **165** (2007), no. 2, 637–664.
3. R. V. Chacon, 'Transformations having continuous spectrum', *J. Math. Mech.* **16** (1966), 399–415.
4. I. P. Cornfeld, S. V. Fomin, and Y. G. Sinaĭ, *Ergodic theory* (Springer-Verlag, New York, 1982).
5. M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, in *Graduate Texts in Mathematics* **259** (Springer-Verlag, New York, 2011).
6. S. Ferenczi and B. Kamiński, 'Zero entropy and directional Bernoullicity of a Gaussian $\mathbb{Z}^2$-action', *Proc. Amer. Math. Soc.* **123** (1995), no. 10, 3079–3083.
7. N. P. Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, in *Lecture Notes in Mathematics* **1794** (Springer-Verlag, Berlin, 2002). Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
8. C. Foiaş and Ş. Strătilă, 'Ensembles de Kronecker dans la théorie ergodique', *C. R. Acad. Sci. Paris Sér. A-B* **267** (1968), A166–A168.
9. S. Fomin, 'On the theory of dynamical systems with continuous spectrum', *Doklady Akad. Nauk SSSR (N.S.)* **67** (1949), 435–437.
10. S. Fomin, 'On dynamical systems in a space of functions', *Ukrain. Mat. Žurnal* **2** (1950), no. 2, 25–47.
11. N. A. Friedman, 'Replication and stacking in ergodic theory', *Amer. Math. Monthly* **99** (1992), no. 1, 31–41.
12. I. V. Girsanov, 'Spectra of dynamical systems generated by stationary Gaussian processes', *Dokl. Akad. Nauk SSSR (N.S.)* **119** (1958), 851–853.
13. G. R. Goodson, 'A survey of recent results in the spectral theory of ergodic dynamical systems', *J. Dynam. Control Systems* **5** (1999), no. 2, 173–226.
14. C. Hermite, 'Sur un nouveau développement en série de fonctions', *C. R. Acad. Sci. Paris* **58** (1864), 93–100, 266–273.
15. E. Hewitt and K. Stromberg, *Real and abstract analysis. A modern treatment of the theory of functions of a real variable* (Springer-Verlag, New York, 1965).
16. K. Itô, 'Multiple Wiener integral', *J. Math. Soc. Japan* **3** (1951), 157–169.
17. K. Itô, 'Complex multiple Wiener integral', *Jap. J. Math.* **22** (1952), 63–86.
18. S. Kakutani, 'Examples of ergodic measure preserving transformations which are weakly mising but not strongly mixing', in *Recent advances in topological dynamics (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund)*, pp. 143–149. Lecture Notes in Math., Vol. 318 (Springer, Berlin, 1973).

19. A. Katok, *Combinatorial constructions in ergodic theory and dynamics*, in *University Lecture Series* **30** (American Mathematical Society, Providence, RI, 2003).
20. A. Katok, J. G. Sinaĭ, and A. M. Stepin, 'The theory of dynamical systems and general transformation groups with invariant measure', in *Mathematical analysis, Vol. 13 (Russian)*, pp. 129–262 (Akad. Nauk SSSR Vsesojuz. Inst. Naučn. i Tehn. Informacii, Moscow, 1975).
21. A. Katok and A. M. Stepin, 'Approximations in ergodic theory', *Uspehi Mat. Nauk* **22** (1967), no. 5 (137), 81–106.
22. A. Katok and J.-P. Thouvenot, 'Spectral properties and combinatorial constructions in ergodic theory', in *Handbook of dynamical systems. Vol. 1B*, pp. 649–743 (Elsevier B. V., Amsterdam, 2006).
23. M. Keane, 'Interval exchange transformations', *Math. Z.* **141** (1975), 25–31.
24. M. Keane, 'Non-ergodic interval exchange transformations', *Israel J. Math.* **26** (1977), no. 2, 188–196.
25. H. B. Keynes and D. Newton, 'A "minimal", non-uniquely ergodic interval exchange transformation', *Math. Z.* **148** (1976), no. 2, 101–105.
26. G. Maruyama, 'The harmonic analysis of stationary stochastic processes', *Mem. Fac. Sci. Kyūsyū Univ. A.* **4** (1949), 45–106.
27. H. Masur, 'Interval exchange transformations and measured foliations', *Ann. of Math. (2)* **115** (1982), no. 1, 169–200.
28. M. G. Nadkarni, *Spectral theory of dynamical systems*, in *Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]* (Birkhäuser Verlag, Basel, 1998).
29. J. Neveu, 'Sur les processus aléatoires gaussiens', *C. R. Acad. Sci. Paris Sér. A-B* **265** (1967), A747–A750.
30. A. Nogueira and D. Rudolph, 'Topological weak-mixing of interval exchange maps', *Ergodic Theory Dynam. Systems* **17** (1997), no. 5, 1183–1209.
31. V. I. Oseledec, 'The spectrum of ergodic automorphisms', *Dokl. Akad. Nauk SSSR* **168** (1966), 1009–1011.
32. G. Rauzy, 'Échanges d'intervalles et transformations induites', *Acta Arith.* **34** (1979), no. 4, 315–328.
33. A. Rényi, 'On mixing sequences of sets', *Acta Math. Acad. Sci. Hungar.* **9** (1958), 215–228.
34. P. C. Shields, 'Cutting and stacking: a method for constructing stationary processes', *IEEE Trans. Inform. Theory* **37** (1991), no. 6, 1605–1617.
35. H. Totoki, *Ergodic theory*, in *Lecture Notes Series, No. 14* (Matematisk Institut, Aarhus Universitet, Aarhus, 1969).
36. W. A. Veech, 'Gauss measures for transformations on the space of interval exchange maps', *Ann. of Math. (2)* **115** (1982), no. 1, 201–242.
37. A. M. Veršik, 'On the theory of normal dynamical systems', *Dokl. Akad. Nauk SSSR* **144** (1962), 9–12.
38. A. M. Veršik, 'A spectral and metrical isomorphism of certain normal dynamical systems', *Dokl. Akad. Nauk SSSR* **144** (1962), 255–257.
39. M. Viana, 'Ergodic theory of interval exchange maps', *Rev. Mat. Complut.* **19** (2006), no. 1, 7–100.
40. T. Ward, 'Three results on mixing shapes', *New York J. Math.* **3A** (1997/98), no. Proceedings of the New York Journal of Mathematics Conference, June 9–13, 1997, 1–10 (electronic).

Springer

http://www.springer.com/978-0-85729-020-5

Ergodic Theory
with a view towards Number Theory
Einsiedler, M.; Ward, Th.
2011, XVII, 481 p., Hardcover
ISBN: 978-0-85729-020-5