MEASURE-PRESERVING SYSTEMS

KARL PETERSEN

WITH NOTES BY MATTHEW BONZEK, JOSHUA CLEMONS, JEANETTE OLLI, NATHAN PENNINGTON, AND INDRANI RAO; AND SOME SECTIONS ON JOININGS BASED ON LECTURES BY ANDRES DEL JUNCO

Abstract. These notes are from a graduate course given in Spring 2007 at the University of North Carolina at Chapel Hill. A major portion will be published as part of the Springer Online Encyclopedia on Complexity. They provide an introduction to the subject of measure-preserving dynamical systems, discussing the dynamical viewpoint; how and from where measure-preserving systems arise; the construction of measures and invariant measures; some basic constructions within the class of measure-preserving systems; and some mathematical background on conditional expectations, Lebesgue spaces, and disintegrations of measures.

Contents

Received by the editors November 11, 2008.
1991 *Mathematics Subject Classification.* 37A30, 37A05.

1. The dynamical viewpoint

Sometimes introducing a dynamical viewpoint into an apparently static situation can help to make progress on apparently difficult problems. For example, equations can be solved and functions optimized by reformulating a given situation as a fixed point problem, which is then addressed by iterating an appropriate mapping. Besides practical applications, this strategy also appears in theoretical settings, for example modern proofs of the Implicit Function Theorem. Moreover, the introduction of the ideas of change and motion leads to new concepts, new methods, and even new kinds of questions. One looks at actions and orbits and instead of always seeking exact solutions begins perhaps to ask questions of a qualitative or probabilistic nature: what is the general behavior of the system, what happens for most initial conditions, what properties of systems are typical within a given class of systems, and so on. Much of the credit for introducing this viewpoint should go to Henri Poincaré [41].

1.1. **Two examples.** Consider two particular examples, one simple and the other not so simple. Decimal or base 2 expansions of numbers in the unit interval raise many natural questions about frequencies of digits and blocks. Instead of regarding the base 2 expansion $x = .x_0x_1\dots$ of a fixed $x \in [0,1]$ as being given, we can regard it as arising from a dynamical process. Define $T : [0,1] \to [0,1]$ by $Tx = 2x \mod 1$ (the fractional part of $2x$) and let $\mathcal{P} = \{P_0 = [0,1/2), P_1 = [1/2,1]\}$ be a partition of $[0,1]$ into two subintervals. We code the orbit of any point $x \in [0,1]$ by 0's and 1's by letting $x_k = i$ if $T^k x \in P_i, k = 0,1,2,\dots$. Then reading the expansion of x amounts to applying to the coding the shift transformation and projection onto the first coordinate This is equivalent to following the orbit of x under T and noting which element of the partition $\mathcal{P}$ is entered at each time. Reappearances of blocks amount to recurrence to cylinder sets as x is moved by T, frequencies of blocks correspond to ergodic averages, and Borel's theorem on normal numbers is seen as a special case of the Ergodic Theorem.

Another example concerns Szemerédi's Theorem [49], which states that every subset $A \subset \mathbb{N}$ of the natural numbers which has positive upper density contains arbitrarily long arithmetic progressions: given $L \in \mathbb{N}$ there are $s, m \in \mathbb{N}$ such that $s, s+m, \dots, s+(L-1)m \in A$. Szemerédi's proof was ingenious, direct, and long. Furstenberg [17] saw how to obtain this result as a corollary of a strengthening of Poincaré's Recurrence Theorem in ergodic theory, which he then proved. Again we have an apparently static situation: a set $A \subset \mathbb{N}$ of positive density in which we seek arbitrarily long regularly spaced subsets. Furstenberg proposed to consider the characteristic function $\mathbf{1}_A$ of A as a point in the space $\{0,1\}^{\mathbb{N}}$ of 0's and 1's and to form the orbit closure X of this point under the shift transformation σ. Because A has positive density, it is possible to find a shift-invariant measure μ on X which gives positive measure to the cylinder set $B = [1] = \{x \in X : x_1 = 1\}$. Furstenberg's Multiple Recurrence Theorem says that given $L \in \mathbb{N}$ there is $m \in \mathbb{N}$ such that $\mu(B \cap T^{-m}B \cap \dots \cap T^{-(L-1)m}B) > 0$. If y is a point in this intersection, then y contains a block of L 1's, each at distance m from the next. And since y

is in the orbit closure of $\mathbf{1}_A$, this block also appears in the sequence $\mathbf{1}_A \in \{0,1\}^{\mathbb{N}}$, yielding the result.

Aspects of the dynamical argument remain in new combinatorial and harmonic-analytic proofs of the Szemerédi Theorem by T. Gowers [20, 21] and T. Tao [50], as well as the extension to the (density zero) set of prime numbers by B. Green and T. Tao [22, 51].

1.2. **A range of actions.** Here is a sample of dynamical systems of various kinds:

(1) A semigroup or group G acts on a set X. There is given a map $G \times X \to X$, $(g, x) \to gx$, and it is assumed that

$$g_1(g_2x) = (g_1g_2)x \quad \text{for all } g_1, g_2 \in G, x \in X \tag{1.1}$$

$$ex = x \quad \text{for all } x \in X, \text{ if } G \text{ has an identity element } e. \tag{1.2}$$

(2) A continuous linear operator T acts on a Banach or Hilbert space V.

(3) $\mathcal{B}$ is a Boolean σ-algebra (a set together with a zero element 0 and operations $\vee, \wedge, '$ which satisfy the same rules as $\emptyset, \cup, \cap, ^c$ (complementation) do for σ-algebras of sets); $\mathcal{N}$ is a σ-ideal in $\mathcal{B}$ ($N \in \mathcal{N}, B \in \mathcal{B}, B \wedge N = B$ implies $B \in \mathcal{N}$; and $N_1, N_2, \dots \in \mathcal{N}$ implies $\vee_{n=1}^{\infty} N_n \in \mathcal{N}$); and $S : \mathcal{B} \to \mathcal{B}$ preserves the Boolean σ-algebra operations and $S\mathcal{N} \subset \mathcal{N}$.

(4) $\mathcal{B}$ is a Boolean σ-algebra, μ is a countably additive positive (nonzero except on the zero element of $\mathcal{B}$) function on $\mathcal{B}$, and $S : \mathcal{B} \to \mathcal{B}$ is as above. Then $(\mathcal{B}, \mu)$ is a *measure algebra* and S is a *measure algebra endomorphism*.

(5) $(X, \mathcal{B}, \mu)$ is a measure space (X is a set, $\mathcal{B}$ is a σ-algebra of subsets of X, and $\mu : \mathcal{B} \to [0, \infty]$ is countably additive: If $B_1, B_2, \dots \in \mathcal{B}$ are pairwise disjoint, then $\mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n)$); $T : X \to X$ is measurable ($T^{-1}\mathcal{B} \subset \mathcal{B}$) and nonsingular ($\mu(B) = 0$ implies $\mu(T^{-1}B) = 0$—or, more stringently, μ and μT^{-1} are equivalent in the sense of absolute continuity).

(6) $(X, \mathcal{B}, \mu)$ is a measure space, $T : X \to X$ is a one-to-one onto map such that T and T^{-1} are both measurable (so that $T^{-1}\mathcal{B} = \mathcal{B} = T\mathcal{B}$), and $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. (In practice often T is not one-to-one, or onto, or even well-defined on all of X, but only after a set of measure zero is deleted.) This is the case of most interest for us, and then we call $(X, \mathcal{B}, \mu, T)$ a *measure-preserving system.* We also allow for the possibility that T is not invertible, or that some other group (such as $\mathbb{R}$ or $\mathbb{Z}^d$) or semigroup acts on X, but the case of $\mathbb{Z}$ actions will be the main focus of this article.

(7) X is a compact metric space and $T : X \to X$ is a homeomorphism. Then (X, T) is a *topological dynamical system.*

(8) M is a compact manifold ($\mathcal{C}^k$ for some $k \in [1, \infty]$) and $T : M \to M$ is a diffeomorphism (one-to-one and onto, with T and T^{-1} both $\mathcal{C}^k$). Then (M, T) is a *smooth dynamical system.* Such examples can arise from solutions of an autonomous differential equation given by a vector field on

M. Recall that in $\mathbb{R}^n$, an ordinary differential equation initial-value problem $x' = f(x), x(0) = x_0$ has a unique solution $x(t)$ as long as f satisfies appropriate smoothness conditions. The existence and uniqueness theorem for differential equations then produces a flow according to $T_t x_0 = x(t)$, satisfying $T_{s+t} x_0 = T_s(T_t x_0)$. Restricting to a compact invariant set (if there is one) and taking $T = T_1$ (the time 1 map) gives us a smooth system (M, f).

Naturally there are relations and inclusions among these examples of actions. Often problems can be clarified by forgetting about some of the structure that is present or by adding desirable structure (such as topology) if it is not. There remain open problems about representation and realization; for example, taking into account necessary restrictions, which measure-preserving systems can be realized as smooth systems preserving a smooth measure? Sometimes interesting aspects of the dynamics of a smooth system can be due to the presence of a highly non-smooth subsystem, for example a compact lower-dimensional invariant set. Thus one should be ready to deal with many kinds of dynamical systems.

2. Where do measure-preserving systems come from?

2.1. **Systems in equilibrium.** Besides physical systems, abstract dynamical systems can also represent aspects of biological, economic, or other real world systems. Equilibrium does not mean stasis, but rather that the changes in the system are governed by laws which are not themselves changing. The presence of an invariant measure means that the probabilities of observable events do not change with time. (But of course what happens at time 2 can still depend on what happens at time 1, or, for that matter, at time 3.)

We consider first the example of the wide and important class of Hamiltonian systems. Many systems that model physical situations, for example a large number of ideal charged particles in a container, can be studied by means of Hamilton's equations. The state of the entire system at any time is supposed to be specified by a vector $(q, p) \in \mathbb{R}^{2n}$, the *phase space*, with q listing the coordinates of the positions of all of the particles, and p listing the coordinates of their momenta. We assume that there is a time-independent *Hamiltonian function* $H(q, p)$ such that the time development of the system satisfies *Hamilton's equations*:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, \dots, n. \tag{2.1}$$

Often the Hamiltonian function is the sum of kinetic and potential energy:

$$H(q, p) = K(p) + U(q). \tag{2.2}$$

The potential energy $U(q)$ may depend on interactions among the particles or with an external field, while the kinetic energy $K(p)$ depends on the velocities and masses of the particles.

As discussed above, solving these equations with initial state (q, p) for the system produces a flow $(q, p) \to T_t(q, p)$ in phase space. According to *Liouville's Theorem*, this flow preserves Lebesgue measure on $\mathbb{R}^{2n}$. Calculating dH/dt by means of the Chain Rule and using Hamilton's equations shows that H is constant on orbits of the flow, and thus each set of constant energy $X(H_0) = \{(q,p) : H(q,p) = H_0\}$ is an invariant set. Thus one should consider the flow restricted to the appropriate invariant set. It turns out that there are also natural invariant measures on the sets $X(H_0)$, namely the ones given by rescaling the volume element dS on $X(H_0)$ by the factor $1/||\nabla H||$. For details, see [33].

Systems in equilibrium can also be hiding inside systems not in equilibrium, for example if there is an attractor supporting an SRB measure (for Sinai, Ruelle, and Bowen). Suppose that $T : M \to M$ is a diffeomorphism on a compact manifold as above, and that m is a version of Lebesgue measure on M, say given by a smooth volume form. We consider m to be a "physical measure", corresponding to laboratory measurements of observable quantities, whose values can be determined to lie in certain intervals in $\mathbb{R}$. Quite possibly m is not itself invariant under T, and an experimenter might observe strange or chaotic behavior whenever the state of the system gets close to some compact invariant set X. The dynamics of T restricted to X can in fact be quite complicated—maybe a full shift, which represents completely undeterministic behavior (for example if there is a horseshoe present), or a shift of finite type, or some other complicated topological dynamical system. Possibly $m(X) = 0$, so that X is effectively invisible to the observer except through its effects. It can happen that there is a T-invariant measure μ supported on X such that

$$\frac{1}{n}\sum_{k=0}^{n-1} mT^{-k} \to \mu \quad \text{weak }^*, \tag{2.3}$$

and then the long-term equilibrium dynamics of the system is described by (X, T, μ). For a recent survey on SRB measures, see [57].

2.2. **Stationary stochastic processes.** A *stationary process* is a family $\{f_t : t \in T\}$ of random variables (measurable functions) on a probability space $(\Omega, \mathcal{F}, P)$. Usually T is $\mathbb{Z}, \mathbb{N}$, or $\mathbb{R}$. For the remainder of this section let us fix $T = \mathbb{Z}$ (although the following definition could make sense for T any semigroup). We say that the process $\{f_n : n \in \mathbb{Z}\}$ is *stationary* if its finite-dimensional distributions are translation invariant, in the sense that for each $r = 1, 2, \dots$, each $n_1, \dots, n_r \in \mathbb{Z}$, each choice of Borel sets $B_1, \dots, B_r \subset \mathbb{R}$, and each $s \in \mathbb{Z}$, we have

$$P\{\omega : f_{n_1}(\omega) \in B_1, \dots, f_{n_r}(\omega) \in B_r\} = P\{\omega : f_{n_1+s}(\omega) \in B_1, \dots, f_{n_r+s}(\omega) \in B_r\}. \tag{2.4}$$

The f_n represent measurements made at times n of some random phenomenon, and the probability that a particular finite set of measurements yield values in certain ranges is supposed to be independent of time.

Each stationary process $\{f_n : n \in \mathbb{Z}\}$ on $(\Omega, \mathcal{F}, P)$ corresponds to a shift-invariant probability measure μ on the set $\mathbb{R}^{\mathbb{Z}}$ (with its Borel σ-algebra) and a single observable, namely the projection π_0 onto the 0'th coordinate, as follows. Define

$$\phi : \Omega \to \mathbb{R}^{\mathbb{Z}} \quad \text{by } \phi(\omega) = (f_n(\omega))_{-\infty}^{\infty}, \tag{2.5}$$

and for each Borel set $E \subset \mathbb{R}^{\mathbb{Z}}$, define $\mu(E) = P(\phi^{-1}E)$. Then examining the values of μ on cylinder sets— for Borel $B_1, \dots, B_r \subset \mathbb{R}$,

$$\mu\{x \in \mathbb{R}^{\mathbb{Z}} : x_{n_i} \in B_i, i = 1, \dots, r\} = P\{\omega \in \Omega : f_{n_i}(\omega) \in B_i, i = 1, \dots, r \tag{2.6}$$

—and using stationarity of (f_n) shows that μ is invariant under σ. Moreover, the processes (f_n) on Ω and $\pi_0 \circ \sigma^n$ on $\mathbb{R}^{\mathbb{Z}}$ have the same finite-dimensional distributions, so they are equivalent for the purposes of probability theory.

3. Construction of measures

We review briefly (following [45]) the construction of measures largely due to C. Carathéodory [8], with input from M. Fréchet [14], H. Hahn [23], A. N. Kolmogorov [35], and others, then discuss the application to construction of measures on shift spaces and of stochastic processes in general.

3.1. **The Carathéodory construction.** A *semialgebra* is a family $\mathcal{S}$ of subsets of a set X which is closed under finite intersections and such that the complement of any member of $\mathcal{S}$ is a finite disjoint union of members of $\mathcal{S}$. Key examples are

(1) the family $\mathcal{H}$ of half-open subintervals $[a, b)$ of $[0, 1)$;

(2) in the space $X = A^{\mathbb{Z}}$ of doubly infinite sequences on a finite alphabet A, the family $\mathcal{C}$ of *cylinder sets* (determined by fixing finitely many entries)

$$\{x \in A^{\mathbb{Z}} : x_{n_1} = a_1, \dots, x_{n_r} = a_r\}; \tag{3.1}$$

(3) the family $\mathcal{C}_1$ of *anchored cylinder sets*

$$\{x \in A^{\mathbb{N}} : x_1 = a_1, \dots, x_r = a_r\} \tag{3.2}$$

in the space $X = A^{\mathbb{N}}$ of one-sided infinite sequences on a finite alphabet A.

An *algebra* is a family of subsets of a set X which is closed under finite unions, finite intersections, and complements. A σ*-algebra* is a family of subsets of a set X which is closed under countable unions, countable intersections, and complements. If $\mathcal{S}$ is a semialgebra of subsets of X, the *algebra* $\mathcal{A}(\mathcal{S})$ *generated* by $\mathcal{S}$ is the smallest algebra of subsets of X which contains $\mathcal{S}$. $\mathcal{A}(\mathcal{S})$ is the intersection of all the subalgebras of the set 2^X of all subsets of X and consists exactly of all finite disjoint unions of elements of $\mathcal{S}$. Given an algebra $\mathcal{A}$, the σ*-algebra* $\mathcal{B}(\mathcal{A})$ *generated by* $\mathcal{A}$ is the smallest σ-algebra of subsets of X which contains $\mathcal{A}$.

A *nonnegative set function* on $\mathcal{S}$ is a function $\mu : \mathcal{S} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ if $\emptyset \in \mathcal{S}$. We say that such a μ is

- *finitely additive* if whenever $S_1, \dots, S_n \in \mathcal{S}$ are pairwise disjoint and $S = \cup_{i=1}^n S_i \in \mathcal{S}$, we have $\mu(S) = \sum_{i=1}^n \mu(S_i)$;
- *countably additive* if whenever $S_1, S_2 \dots \in \mathcal{S}$ are pairwise disjoint and $S = \cup_{i=1}^\infty S_i \in \mathcal{S}$, we have $\mu(S) = \sum_{i=1}^\infty \mu(S_i)$; and

- *countably subadditive* if whenever $S_1, S_2 \cdots \in \mathcal{S}$ and $S = \cup_{i=1}^{\infty} S_i \in \mathcal{S}$, we have $\mu(S) \leq \sum_{i=1}^{\infty} \mu(S_i)$.

Notice that in order to check countable subadditivity of a nonnegative set function on a semialgebra, it is enough to check countable subadditivity on each countable family of *pairwise disjoint* sets. (Disjointify and apply the semialgebra properties.)

A *measure* is a countably additive nonnegative set function defined on a σ-algebra.

Proposition 3.1. *Let $\mathcal{S}$ be a semialgebra and μ a nonnegative set function on $\mathcal{S}$. In order that μ have an extension to a finitely additive set function on the algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$, it is necessary and sufficient that μ be finitely additive on $\mathcal{S}$.*

Proof. The stated condition is obviously necessary. Conversely, given μ which is finitely additive on $\mathcal{S}$, it is natural to define

$$\mu(\bigcup_{i=1}^{n} S_i) = \sum_{i=1}^{n} \mu(S_i) \tag{3.3}$$

whenever $A = \cup_{i=1}^{n} S_i$ (with the S_i pairwise disjoint) is in the algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$. It is necessary to verify that μ is then well defined on $\mathcal{A}(\mathcal{S})$, since each element of $\mathcal{A}(\mathcal{S})$ may have more than one representation as a finite disjoint union of members of $\mathcal{S}$. But, given two such representations of a single set A, forming the common refinement and applying finite additivity on $\mathcal{S}$ shows that μ so defined assigns the same value to A both times: if

$$A = \bigcup_{i=1}^{n} S_i = \bigcup_{j=1}^{m} R_j \quad \text{with all } S_i, R_j \in \mathcal{S}, \tag{3.4}$$

then

$$\begin{aligned} \sum_i \mu(S_i) &= \sum_{i,j} \mu(S_i \cap R_j) = \sum_i \sum_j \mu(S_i \cap R_j) \\ &= \sum_j \sum_i \mu(R_j \cap S_i) = \sum_j \mu(R_j). \end{aligned} \tag{3.5}$$

Then finite additivity on $\mathcal{A}(\mathcal{S})$ of the extended μ is clear. □

Proposition 3.2. *Let $\mathcal{S}$ be a semialgebra and μ a nonnegative set function on $\mathcal{S}$. In order that μ have an extension to a countably additive set function on the algebra $\mathcal{A}(\mathcal{S})$ generated by $\mathcal{S}$, it is necessary and sufficient that μ be (i) finitely additive and (ii) countably subadditive on $\mathcal{S}$.*

Proof. Conditions (i) and (ii) are clearly necessary. If μ is finitely additive on $\mathcal{S}$, then by Proposition 3.1 μ has an extension to a finitely additive nonnegative set function, which we will still denote by μ, on $\mathcal{A}(\mathcal{S})$.

Let us see that this extension μ is countably subadditive on $\mathcal{A}(\mathcal{S})$. Suppose that $A_1, A_2, \cdots \in \mathcal{A}(\mathcal{S})$ are pairwise disjoint and their union $A \in \mathcal{A}(\mathcal{S})$. Then A is a finite disjoint union of sets in $\mathcal{S}$, as is each A_i:

$$(3.6)\qquad \begin{aligned} A &= \bigcup_{i=1}^{\infty} A_i, \quad \text{each } A_i = \bigcup_{k=1}^{n_i} S_{ik}, \\ A &= \bigcup_{j=1}^{m} R_j, \quad \text{each } A_i \in \mathcal{A}(\mathcal{S}),\ \text{each } S_{ik}, R_j \in \mathcal{S}. \end{aligned}$$

Since each $R_j \in \mathcal{S}$, by countable subadditivity of μ on $\mathcal{S}$, and using $R_j = R_j \cap A$,

$$(3.7)\qquad \begin{aligned} \mu(R_j) &= \mu(\bigcup_{i=1}^{\infty} \bigcup_{k=1}^{n_i} S_{ik} \cap R_j) \\ &\leq \sum_{i=1}^{\infty} \sum_{k=1}^{n_i} \mu(S_{ik} \cap R_j), \end{aligned}$$

and hence, by finite additivity of μ on $\mathcal{A}(\mathcal{S})$,

$$(3.8)\qquad \begin{aligned} \mu(A) &= \sum_{j=1}^{m} \mu(R_j) \\ &\leq \sum_{i=1}^{\infty} \sum_{k=1}^{n_i} \sum_{j=1}^{m} \mu(S_{ik} \cap R_j) \\ &= \sum_{i=1}^{\infty} \sum_{k=1}^{n_i} \mu(S_{ik}) = \sum_{i=1}^{\infty} \mu(A_i). \end{aligned}$$

Now finite additivity of μ on an algebra $\mathcal{A}$ implies that μ is *monotonic* on the algebra: if $A, B \in \mathcal{A}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$. Thus if $A_1, A_2, \dots \in \mathcal{A}(\mathcal{S})$ are pairwise disjoint and their union $A \in \mathcal{A}(\mathcal{S})$, then for each n we have $\sum_{i=1}^{n} \mu(A_i) = \mu(\cup_{i=1}^{n} A_i) \leq \mu(A)$, and hence $\sum_{i=1}^{\infty} \mu(A_i) \leq \mu(A)$. □

Theorem 3.3. *In order that a nonnegative set function μ on an algebra $\mathcal{A}$ of subsets of a set X have an extension to a (countably additive) measure on the σ-algebra $\mathcal{B}(\mathcal{A})$ generated by $\mathcal{A}$, it is necessary and sufficient that μ be countably additive on $\mathcal{A}$.*

Here is a sketch of how the extension can be constructed. Given a countably additive nonnegative set function μ on an algebra $\mathcal{A}$ of subsets of a set X, one defines the *outer measure* μ^* that it determines on the family 2^X of all subsets of X by

$$(3.9)\qquad \mu^*(E) = \inf\{\sum_{i=1}^{\infty} \mu(A_i) : A_i \in \mathcal{A}, E \subset \cup_{i=1}^{\infty} A_i\}.$$

Then μ^* is a nonnegative, countably subadditive, monotonic set function on 2^X.

Define a set E to be μ^*-*measurable* if for all $T \subset X$,

$$(3.10)\qquad \mu^*(T) = \mu^*(T \cap E) + \mu^*(T \cap E^c).$$

This ingenious definition can be partly motivated by noting that if μ^* is to be finitely additive on the family $\mathcal{M}$ of μ^*-measurable sets, which should contain X, then at least this condition must hold when $T = X$. It is amazing that then this definition readily, with just a little set theory and a few ϵ's, yields the following theorem.

Theorem 3.4. *Let μ be a countably additive nonnegative set function on an algebra $\mathcal{A}$ of subsets of a set X, and let μ^* be the outer measure that it determines on the family 2^X of all subsets of X as above. Then the family $\mathcal{M}$ of μ^*-measurable subsets of X is a σ-algebra containing $\mathcal{A}$ (and hence $\mathcal{B}(\mathcal{A})$) and all subsets of X which have μ^* measure 0. The restriction $\mu^*|\mathcal{M}$ is a (countably additive) measure which agrees on $\mathcal{A}$ with μ. If μ is σ-finite on $\mathcal{A}$ (so that there are $X_1, X_2, \dots \in \mathcal{A}$ with $\mu(X_i) < \infty$ for all i and $X = \cup_{i=1}^{\infty} X_i$), then μ on $\mathcal{B}(\mathcal{A})$ is the* only *extension of μ on $\mathcal{A}$ to $\mathcal{B}(\mathcal{A})$.*

In this way, beginning with the semialgebra $\mathcal{H}$ of half-open subintervals of $[0, 1)$ and $\mu[a, b) = b-a$, one arrives at Lebesgue measure on the σ-algebra $\mathcal{M}$ of Lebesgue measurable sets and on $\mathcal{B}(\mathcal{H})$ of Borel sets.

3.2. **Measures on shift spaces.** The measures that determine stochastic processes are also frequently constructed by specifying data on a semialgebra of cylinder sets. Given a finite alphabet A, denote by $\Omega(A) = A^{\mathbb{Z}}$ and $\Omega^+(A) = A^{\mathbb{N}}$ the sets of two and one-sided sequences, respectively, with entries from A. These are compact metric spaces, with $d(x, y) = 2^{-n}$ when $n = \inf\{|k| : x_k \neq y_k\}$. In both cases, the *shift transformation* σ defined by $(\sigma x)_n = x_{n+1}$ for all n is a homeomorphism.

Suppose (cf. [3]) that for every $k = 1, 2, \dots$ we are given a function $g_k : A^k \to [0, 1]$, and that these functions satisfy, for all k,

(1) $g_k(B) \geq 0$ for all $B \in A^k$;

(2) $\sum_{i \in A} g_{k+1}(Bi) = g_k(B)$ for all $B \in A^k$;

(3) $\sum_{i \in A} g_1(i) = 1$.

Then Theorems 3.3 and 3.4 imply that there is a unique measure μ on the Borel subsets of $\Omega^+(A)$ such that for all $k = 1, 2, \dots$ and $B \in A^k$

$$\mu\{x \in \Omega^+(A) : x_1 \dots x_k = B\} = g_k(B). \tag{3.11}$$

If in addition the g_k also satisfy

(4) $\sum_{i \in A} g_{k+1}(iB) = g_k(B)$ for all $k = 1, 2, \dots$ and all $B \in A^k$,

then there is a unique *shift-invariant* measure μ on the Borel subsets of $\Omega^+(A)$ (also $\Omega(A)$) such that for all n, all $k = 1, 2, \dots$ and $B \in A^k$

$$\mu\{x \in \Omega^+(A) : x_n \dots x_{n+k-1} = B\} = g_k(B). \tag{3.12}$$

This follows from the Carathéodory theorem by beginning with the semialgebra $\mathcal{C}_1$ of anchored cylinder sets or the semialgebra $\mathcal{C}$ of cylinder sets determined by finitely many consecutive coordinates, respectively.

There are two particularly important examples of this construction. First, let our finite alphabet be $A = \{0, \dots, d-1\}$, and let $p = (p_0, \dots, p_{d-1})$ be a probability vector: all $p_i \geq 0$ and $\sum_{i=0}^{d-1} p_i = 1$. For any block $B = b_1 \dots b_k \in A^k$, define

$$g_k(B) = p_{b_1} \dots p_{b_k}. \tag{3.13}$$

The resulting measure μ_p is the product measure on $\Omega(A) = A^{\mathbb{Z}}$ of infinitely many copies of the probability measure determined by p on the finite sample space A. The measure-preserving system $(\Omega, \mathcal{B}, \mu, \sigma)$ (with $\mathcal{B}$ the σ-algebra of Borel subsets of $\Omega(A)$, or its completion), is denoted by $\mathcal{B}(p)$ and is called the *Bernoulli system* determined by p. This system models an infinite number of independent repetitions of an experiment with finitely many outcomes, the i'th of which has probability p_i on each trial.

This construction can be generalized to model stochastic processes which have some memory. Again let $A = \{0, \dots, d-1\}$, and let $p = (p_0, \dots, p_{d-1})$ be a probability vector. Let P be a $d \times d$ *stochastic matrix* with rows and columns indexed by A. This means that all entries of P are nonnegative, and the sum of the entries in each row is 1. We regard P as giving the transition probabilities between pairs of elements of A. Now we define for any block $B = b_1 \dots b_k \in A^k$

$$g_k(B) = p_{b_1} P_{b_1 b_2} P_{b_2 b_3} \dots P_{b_{k-1} b_k}. \tag{3.14}$$

Using the g_k to define a nonnegative set function $\mu_{p,P}$ on the semialgebra $\mathcal{C}_1$ of anchored cylinder subsets of $\Omega^+(A)$, one can verify that $\mu_{p,P}$ is (vacuously) finitely additive and countably subadditive on $\mathcal{C}_1$ and therefore extends to a measure on the Borel σ-algebra of $\Omega^+(A)$, and its completion. The resulting stochastic process is a (one-step, finite-state) *Markov process*. If p and P also satisfy

$$pP = p, \tag{3.15}$$

then condition (4) above is satisfied, and the Markov process is stationary. In this case we call the (one or two-sided) measure-preserving system the *Markov shift* determined by p and P. Points in the space are conveniently pictured as infinite paths in a directed graph with vertices A and edges corresponding to the nonzero entries of P. A process with a longer memory, say of length m, can be produced by repeating the foregoing construction after recoding with a *sliding block code* to the new alphabet A^m: for each $\omega \in \Omega(A)$, let $(\phi(\omega))_n = \omega_n \omega_{n+1} \dots \omega_{n+m-1} \in A^m$.

3.3. **The Kolmogorov Consistency Theorem.**

There is a generalization of this method to the construction of stochastic processes indexed by any set T. (Most frequently $T = \mathbb{Z}, \mathbb{N}, \mathbb{R}, \mathbb{Z}^d$, or $\mathbb{R}^d$). We give a brief description, following [4].

Let T be an arbitrary index set. We aim to produce a $\mathbb{R}$-valued stochastic process indexed by T, that is to say, a Borel probability measure P on $\Omega = \mathbb{R}^T$, which has prespecified finite-dimensional distributions. Suppose that for every ordered k-tuple $t_1, \dots, t_k$ of *distinct* elements of T we are given a Borel probability measure $\mu_{t_1 \dots t_k}$ on $\mathbb{R}^k$. Denoting $f \in \mathbb{R}^T$ also by $(f_t : t \in T)$, we want it to be the case that, for each k, each choice of distinct $t_1, \dots t_k \in T$, and each Borel set $B \subset \mathbb{R}^k$,

$$P\{(f_t : t \in T) : (f_{t_1}, \dots, f_{t_k}) \in B\} = \mu_{t_1 \dots t_k}(B). \tag{3.16}$$

For consistency, we will need, for example, that

$$\mu_{t_1t_2}(B_1 \times B_2) = \mu_{t_2t_1}(B_2 \times B_1), \quad \text{since} \tag{3.17}$$

$$P\{(f_{t_1}, f_{t_2}) \in A_1 \times A_2\} = P\{(f_{t_2}, f_{t_1}) \in A_2 \times A_1\}. \tag{3.18}$$

Thus we assume:

(1) For any $k = 1, 2, \dots$ and permutation π of $1, \dots, k$, if $\phi_\pi : \mathbb{R}^k \to \mathbb{R}^k$ is defined by

$$\phi_\pi(x_{\pi 1}, \dots, x_{\pi k}) = (x_1, \dots, x_k), \tag{3.19}$$

then for all k and all Borel $B \subset \mathbb{R}^k$

$$\mu_{t_1 \dots t_k}(B) = \mu_{t_{\pi 1} \dots t_{\pi k}}(\phi_\pi^{-1} B). \tag{3.20}$$

Further, since leaving the value of one of the f_{t_j} free does not change the probability in (3.16), we also should have

(2) For any $k = 1, 2, \dots$, distinct $t_1, \dots, t_k, t_{k+1} \in T$, and Borel set $B \subset \mathbb{R}^k$,

$$\mu_{t_1 \dots t_k}(B) = \mu_{t_1 \dots t_k t_{k+1}}(B \times \mathbb{R}). \tag{3.21}$$

Theorem 3.5 (Kolmogorov Consistency Theorem [35]). *Given a system of probability measures $\mu_{t_1 \dots t_k}$ as above indexed by finite ordered subsets of a set T, in order that there exist a probability measure P on $\mathbb{R}^T$ satisfying (3.16) it is necessary and sufficient that the system satisfy (1) and (2) above.*

When $T = \mathbb{Z}, \mathbb{R}$, or $\mathbb{N}$, as in the example with the g_k above, the problem of consistency with regard to permutations of indices does not arise, since we tacitly use the order in T in specifying the finite-dimensional distributions.

In case T is a semigroup, by adding conditions on the given data $\mu_{t_1 \dots t_k}$ it is possible to extend this construction also to produce *stationary* processes indexed by T, in parallel with the above constructions for $T = \mathbb{Z}$ or $\mathbb{N}$.

4. Invariant measures on topological dynamical systems

4.1. **Existence of invariant measures.** Let X be a compact metric space and $T : X \to X$ a homeomorphism (although usually it is enough just that T be a continuous map). Denote by $\mathcal{C}(X)$ the Banach space of continuous real-valued functions on X with the supremum norm and by $\mathcal{M}(X)$ the set of Borel probability measures on X. Given the weak* topology, according to which

$$\mu_n \to \mu \quad \text{if and only if} \quad \int_X f_n \, d\mu \to \int_X f \, d\mu \quad \text{for all } f \in \mathcal{C}(X), \tag{4.1}$$

$\mathcal{M}(X)$ is a convex subset of the dual space $\mathcal{C}(X)^*$ of all continuous linear functionals from $\mathcal{C}(X)$ to $\mathbb{R}$. With the weak* topology it is metrizable and (by Alaoglu's Theorem) compact.

Denote by $\mathcal{M}_T(X)$ the set of T-invariant Borel probability measures on X. A Borel probability measure μ on X is in $\mathcal{M}(X)$ if and only if

$$\mu(T^{-1}B) = \mu(B) \quad \text{for all Borel sets } B \subset X, \tag{4.2}$$

equivalently,

$$\mu(fT) = \int_X f \circ T \, d\mu = \int_X f \, d\mu \quad \text{for all } f \in \mathcal{C}(X). \tag{4.3}$$

Proposition 4.1. *For every compact topological dynamical system (X, T) (with X not empty) there is always at least one T-invariant Borel probability measure on X.*

Proof. Let m be any Borel probability measure on X. For example, we could pick a point $x_0 \in X$ and let m be the point mass δ_{x_0} at x_0 defined by

$$\delta_{x_0}(f) = f(x_0) \quad \text{for all } f \in \mathcal{C}(X). \tag{4.4}$$

Form the averages

$$A_n m = \frac{1}{n} \sum_{i=0}^{n-1} m T^{-i}, \tag{4.5}$$

which are also in $\mathcal{M}(X)$. By compactness, $\{A_n m\}$ has a weak* cluster point μ, so that there is a subsequence

$$A_{n_k} m \to \mu \quad \text{weak}^*. \tag{4.6}$$

Then $\mu \in \mathcal{M}(X)$; and μ is T-invariant, because for each $f \in \mathcal{C}(X)$

$$|\mu(fT) - \mu(f)| = \lim_{k \to \infty} \frac{1}{n_k} |\mu(fT^{n_k}) - \mu(f)| = 0, \tag{4.7}$$

both terms inside the absolute value signs being bounded. □

4.2. **Ergodicity and unique ergodicity.** Among the T-invariant measures on X are the *ergodic* ones, those for which $(X, \mathcal{B}, \mu, T)$ (with $\mathcal{B}$ the σ-algebra of Borel subsets of X) forms an ergodic measure-preserving system. This means that there are no proper T-invariant measurable sets:

$$B \in \mathcal{B}, \mu(T^{-1}B \triangle B) = 0 \quad \text{implies } \mu(B) = 0 \text{ or } 1. \tag{4.8}$$

Equivalently (using the Ergodic Theorem), $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if for each $f \in L^1(X, \mathcal{B}, \mu)$

$$\frac{1}{n} \sum_{k=1}^{n-1} f(T^k x) \to \int_X f \, d\mu \quad \text{almost everywhere .} \tag{4.9}$$

It can be shown that the ergodic measures on (X, T) are exactly the *extreme points* of the compact convex set $\mathcal{M}_T(X)$, namely those $\mu \in \mathcal{M}_T(X)$ for which there do not exist $\mu_1, \mu_2 \in \mathcal{M}_T(x)$, $\mu_1 \neq \mu_2$, and $s \in (0, 1)$ such that

$$\mu = s\mu_1 + (1 - s)\mu_2. \tag{4.10}$$

The Krein-Milman Theorem states that in a locally convex topological vector space such as $\mathcal{C}(X)^*$ every compact convex set is the closed convex hull of its extreme points. Thus every nonempty such set has extreme points, and so there always exist ergodic measures for (X, T). A topological dynamical system (X, T) is called *uniquely ergodic* if there is only *one* T-invariant Borel probability measure on X, in which case, by the foregoing discussion, that measure must be ergodic.

Later we will see many examples of topological dynamical systems which are uniquely ergodic and of others which are not. For now, we just remark that translation by a generator on a compact monothetic group is always uniquely ergodic, while group endomorphisms and automorphisms tend to be not uniquely ergodic. Bernoulli and (nonatomic) Markov shifts are not uniquely ergodic, because they have many periodic orbits, each of which supports an ergodic measure.

5. Finding finite invariant measures equivalent to a quasi-invariant measure

Let $(X, \mathcal{B}, m)$ be a σ-finite measure space, and suppose that $T : X \to X$ is an invertible *nonsingular* transformation. Thus we assume that T is one-to-one and onto (maybe after a set of measure 0 has been deleted), that T and T^{-1} are both measurable, so that

$$T\mathcal{B} = \mathcal{B} = T^{-1}\mathcal{B}, \tag{5.1}$$

and that T and T^{-1} preserve the σ-ideal of sets of measure 0:

$$m(B) = 0 \quad \text{if and only if } m(T^{-1}B) = 0 \quad \text{if and only if } m(TB) = 0. \tag{5.2}$$

In this situation we say that m is *quasi-invariant* for T.

A nonsingular system $(X, \mathcal{B}, m, T)$ as above may model a nonequilibrium situation in which events that are impossible (measure 0) at any time are also impossible at any other time. When dealing with such a system, it can be useful to know whether there is a T-invariant measure μ that is equivalent to m (in the sense of absolute continuity–they have the same sets of measure 0–in which case we write $\mu \sim m$), for then one would have available machinery of the measure-preserving situation, such as the Ergodic Theorem and entropy in their simplest forms. Also, it is most useful if the measures are σ-finite, so that tools such as the Radon-Nikodym and Tonelli-Fubini theorems will be available.

We may assume that $m(X) = 1$. For if $X = \cup_{i=1}^{\infty} X_i$ with each $X_i \in \mathcal{B}$ and $m(X_i) < \infty$, disjointifying (replace X_i by $X_i \setminus X_{i-1}$ for $i \geq 2$) and deleting any X_i that have measure 0, we may replace m by

$$\sum_{i=1}^{\infty} \frac{m|_{X_i}}{2^i m(X_i)}. \tag{5.3}$$

Definition 5.1. Let $(X, \mathcal{B}, m)$ be a probability space and $T : X \to X$ a nonsingular transformation. We say that $A, B \in \mathcal{B}$ are T-*equivalent*, and write $A \sim_T B$, if there are two sequences of pairwise disjoint sets, $A_1, A_2, \dots$ and $B_1, B_2, \dots$ and integers $n_1, n_2, \dots$ such that

$$A = \bigcup_{i=1}^{\infty} A_i, \quad B = \bigcup_{i=1}^{\infty} B_i, \quad \text{and } T^{n_i} A_i = B_i \quad \text{for all } i. \tag{5.4}$$

Definition 5.2. Let $(X, \mathcal{B}, m, T)$ be as above. A measurable set $A \subset X$ is called *T-nonshrinkable* if A is not T-equivalent to any proper subset: whenever $B \subset A$ and $B \sim_T A$ we have $m(A \setminus B) = 0$.

Theorem 5.3 (Hopf [29])**.** *Let $(X, \mathcal{B}, m)$ be a probability space and $T : X \to X$ a nonsingular transformation. There exists a finite invariant measure $\mu \sim m$ if and only if X is T-nonshrinkable.*

Proof. We present just the easy half. If $\mu \sim m$ is T-invariant and $X \sim_T B$, with corresponding decompositions $X = \cup_{i=1}^{\infty} X_i, B = \cup_{i=1}^{\infty} B_i$, then

$$\begin{aligned} \mu(B) &= \sum_{i=1}^{\infty} \mu(B_i) = \sum_{i=1}^{\infty} \mu(T^{n_i} X_i) \\ &= \sum_{i=1}^{\infty} \mu(X_i) = \mu(X), \end{aligned} \tag{5.5}$$

so that $\mu(X \setminus B) = 0$ and hence $m(X \setminus B) = 0$.

For the converse, one tries to show that if X is T-nonshrinkable, then for each $A \in \mathcal{B}$ the following limit exists:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} m(T^k A). \tag{5.6}$$

□

The condition of T-nonshrinkability not being easy to check, subsequent authors gave various necessary and sufficient conditions for the existence of a finite equivalent invariant measure:

5.4 Dowker [12]**.** *Whenever $A \in \mathcal{B}$ and $m(A) > 0$, $\liminf_{n\to\infty} m(T^n A) > 0$.*

5.5 Calderón [6]**.** *Whenever $A \in \mathcal{B}$ and $m(A) > 0$, $\liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{\infty} m(T^k A) > 0$.*

5.6 Dowker [13]**.** *Whenever $A \in \mathcal{B}$ and $m(A) > 0$, $\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{\infty} m(T^k A) > 0$.*

Hajian and Kakutani [24] showed that the condition

$$m(A) > 0 \quad \text{implies} \limsup_{n\to\infty} m(T^n A) > 0 \tag{5.7}$$

is *not* sufficient for existence of a finite equivalent invariant measure. They also gave another necessary and sufficient condition.

Definition 5.7. A measurable set $W \subset X$ is called *wandering* if the sets $T^i W, i \in \mathbb{Z}$, are pairwise disjoint. W is called *weakly wandering* if there are infinitely many integers n_i such that $T^{n_i} W$ and $T^{n_j} W$ are disjoint whenever $n_i \neq n_j$.

Theorem 5.8 (Hajian-Kakutani [24])**.** *Let $(X, \mathcal{B}, m)$ be a probability space and $T : X \to X$ a nonsingular transformation. There exists a finite invariant measure $\mu \sim m$ if and only if there are no weakly wandering sets of positive measure.*

6. Finding σ-finite invariant measures equivalent to a quasi-invariant measure

6.1. **First necessary and sufficient conditions.** While being able to replace a quasi-invariant measure by an equivalent finite invariant measure would be great, it may be impossible, and then finding a σ-finite equivalent measure would still be pretty good. Hopf's nonshrinkability condition was extended to the σ-finite case by Halmos:

Theorem 6.1 (Halmos [25]). *Let $(X, \mathcal{B}, m)$ be a probability space and $T : X \to X$ a nonsingular transformation. There exists a σ-finite invariant measure $\mu \sim m$ if and only if X is a countable union of T-nonshrinkable sets.*

Another necessary and sufficient condition is given easily in terms of solvability of a *cohomological functional equation* involving the Radon-Nikodym derivative w of mT with respect to m, defined by

$$m(TB) = \int_B w\,dm \quad \text{for all } B \in \mathcal{B}. \tag{6.1}$$

Proposition 6.2 [25]. *Let $(X, \mathcal{B}, m)$ be a probability space and $T : X \to X$ a nonsingular transformation. There exists a σ-finite invariant measure $\mu \sim m$ if and only if there is a measurable function $f : X \to (0, \infty)$ such that*

$$f(Tx) = w(x) f(x) \quad \text{a.e.}. \tag{6.2}$$

Proof. If $\mu \sim m$ is σ-finite and T-invariant, let $f = dm/d\mu$ be the Radon-Nikodym derivative of m with respect to μ, so that

$$m(B) = \int_B f\,d\mu \quad \text{for all } B \in \mathcal{B}. \tag{6.3}$$

Then for all $B \in \mathcal{B}$, since $\mu T = \mu$,

$$\begin{aligned} m(TB) &= \int_{TB} f\,d\mu = \int_B fT\,d\mu, \quad \text{while also} \\ m(TB) &= \int_B w\,dm = \int_B wf\,dm, \end{aligned} \tag{6.4}$$

so that $fT = wf$ a.e..

Conversely, given such an f, let

$$\mu(B) = \int_B \frac{1}{f}\,dm \quad \text{for all } B \in \mathcal{B}. \tag{6.5}$$

Then for all $B \in \mathcal{B}$

$$\begin{aligned} \mu(TB) &= \int_{TB} \frac{1}{f}\,dm = \int_B \frac{1}{fT}\,dmT \\ &= \int_B \frac{1}{fT} w\,dm = \int_B \frac{1}{f}\,dm = \mu(B). \end{aligned} \tag{6.6}$$

□

6.2. **Conservativity and recurrence.**

Definition 6.3. A nonsingular system $(X, \mathcal{B}, m, T)$ (with $m(X) = 1$) is called *conservative* if there are no wandering sets of positive measure. It is called *completely dissipative* if there is a wandering set W such that

$$m(\bigcup_{i=-\infty}^{\infty} T^i W) = m(X). \tag{6.7}$$

Note that if $(X, \mathcal{B}, m, T)$ is completely dissipative, it is easy to construct a σ-finite equivalent invariant measure. With W as above, define $\mu = m$ on W and push μ along the orbit of W, letting $\mu = mT^{-n}$ on each $T^n W$. We want to claim that this allows us to restrict attention to the conservative case, which follows once we know that the system splits into a conservative and a completely dissipative part.

Theorem 6.4 (Hopf Decomposition [30])**.** *Given a nonsingular map T on a probability space $(X, \mathcal{B}, m)$, there are disjoint measurable sets C and D such that*

(1) *$X = C \cup D$;*
(2) *C and D are invariant: $TC = C = T^{-1}C, \quad TD = D = T^{-1}D$;*
(3) *$T|_C$ is conservative;*
(4) *If $D \neq \emptyset$, then $T|_D$ is completely dissipative.*

Proof. Assume that the family $\mathcal{W}$ of wandering sets with positive measure is nonempty, since otherwise we can take $C = X$ and $D = \emptyset$. Partially order $\mathcal{W}$ by

$$W_1 \leq W_2 \quad \text{if} \quad m(W_1 \setminus W_2) = 0. \tag{6.8}$$

We want to apply Zorn's Lemma to find a maximal element in $\mathcal{W}$. Let $\{W_\lambda : \lambda \in \Lambda\}$ be a chain (linearly ordered subset) in $\mathcal{W}$. Just forming $\cup_{\lambda \in \Lambda} W_\lambda$ may result in a nonmeasurable set, so we have to use the measure to form a measure-theoretic essential supremum of the chain. So let

$$s = \sup\{m(W_\lambda) : \lambda \in \Lambda\}, \tag{6.9}$$

so that $s \in (0, 1]$. If there is a λ such that $m(W_\lambda) = s$, let W be that W_λ. Otherwise, for each k choose $\lambda_k \in \Lambda$ so that

$$s_k = m(W_{\lambda_k}) \uparrow s, \tag{6.10}$$

and let

$$W = \bigcup_{k=1}^{\infty} W_{\lambda_k}. \tag{6.11}$$

We claim that in either case W is an upper bound for the chain $\{W_\lambda : \lambda \in \Lambda\}$. In both cases we have $m(W) = s$.

Note that if $\lambda, \tau \in \Lambda$ are such that $m(W_\lambda) \leq m(W_\tau)$, then $W_\lambda \leq W_\tau$. For if $W_\tau \leq W_\lambda$, then $m(W_\tau \setminus W_\lambda) = 0$, and thus

$$\begin{aligned} m(W_\tau) &= m(W_\tau \cap W_\lambda) + m(W_\tau \setminus W_\lambda) = m(W_\tau \cap W_\lambda) \\ &\leq m(W_\tau \cap W_\lambda) + m(W_\lambda \setminus W_\tau) \\ &= m(W_\lambda) \leq m(W_\tau), \end{aligned} \tag{6.12}$$

so that $m(W_\lambda \setminus W_\tau) = 0, W_\lambda \leq W_\tau$, and hence $W_\lambda = W_\tau$.

Thus in the first case $W \in \mathcal{W}$ is an upper bound for the chain. In the second case, by discarding the measure 0 set

$$Z = \bigcup_{k=1}^{\infty} (W_{\lambda_k} \setminus W_{\lambda_{k+1}}), \tag{6.13}$$

we may assume that W is the *increasing* union of the W_{λ_k}. Then $W \geq W_{\lambda_k}$ for all k, and W is wandering: if some $T^n W \cap W \neq \emptyset$, then there must be a k such that $T^n W_{\lambda_k} \cap W_{\lambda_k} \neq \emptyset$.

Moreover, $W_\lambda \leq W$ for all $\lambda \in \Lambda$. For let $\lambda \in \Lambda$ be given. Choose k with $s_k = m(W_{\lambda_k}) > m(W_\lambda)$. By the above, we have $W_{\lambda_k} \geq W_\lambda$. Since W is the increasing union of the W_{λ_k}, we have $W \geq W_{\lambda_k}$ for all k. Therefore $W \geq W_\lambda$, and W is an upper bound in $\mathcal{W}$ for the given chain.

By Zorn's Lemma, there is a maximal element W^* in $\mathcal{W}$. Then $D = \cup_{i=-\infty}^{\infty} T^i W^*$ is T-invariant, $T|_D$ is completely dissipative, and $C = X \setminus D$ cannot contain any wandering set of positive measure, by maximality of W^*, so $T|_C$ is conservative. □

Because of this decomposition, when looking for a σ-finite equivalent invariant measure we may assume that the nonsingular system $(X, \mathcal{B}, m, T)$ is conservative, for if not we can always construct one on the dissipative part.

Remark 6.5. If $(X, \mathcal{B}, m)$ is nonatomic and $T : X \to X$ is nonsingular, invertible, and *ergodic*, in the sense that if $A \in \mathcal{B}$ satisfies $T^{-1}A = A = TA$ then either $m(A) = 0$ or $m(A^c) = 0$, then T is conservative. For if W is a wandering set of positive measure, taking any $A \subset W$ with $0 < m(A) < m(W)$ and forming $\cup_{i=-\infty}^{\infty} T^i A$ will produce an invariant set of positive measure whose complement also has positive measure.

We want to reduce the problem of existence of a σ-finite equivalent invariant measure to that of a finite one by using first-return maps to sets of finite measure. For this purpose it will be necessary to know that every conservative nonsingular system is *recurrent*: almost every point of each set of positive measure returns at some future time to that set. This is easy to see, because for each $B \in \mathcal{B}$, the set

$$B^* = \bigcup_{i=1}^{\infty} T^{-i} B \tag{6.14}$$

is wandering. In fact much more is true.

Theorem 6.6 [25]**.** *For any nonsingular system $(X, \mathcal{B}, m, T)$ the following properties are equivalent:*

(1) The system is incompressible*: for each $B \in \mathcal{B}$ such that $T^{-1}B \subset B$, we have $m(B \setminus T^{-1}B) = 0$.*

(2) The system is recurrent*; for each $B \in \mathcal{B}$, with B^* defined as above, $m(B \setminus B^*) = 0$.*

(3) The system is conservative*: there are no wandering sets of positive measure.*

(4) The system is infinitely recurrent*: for each $B \in \mathcal{B}$, almost every point of B returns to B infinitely many times, equivalently,*

$$m(B \setminus \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}B) = m(B \setminus \bigcap_{n=0}^{\infty} T^{-n}B^*) = 0. \tag{6.15}$$

There is a very slick proof by F. B. Wright [56] of this result in the even more general situation of a Boolean σ-algebra homomorphism (reproduced in [40]).

6.3. **Using first-return maps, and counterexamples to existence.** Now given a nonsingular conservative system $(X, \mathcal{B}, m, T)$ and a set $B \in \mathcal{B}$, for each $x \in B$ there is a smallest $n_B(x) \geq 1$ such that

$$T^{n_B}(x) \in B. \tag{6.16}$$

We define the *first-return map* $T_B : B \to B$ by

$$T_B(x) = T^{n_B(x)}(x) \quad \text{for all } x \in B. \tag{6.17}$$

Using derivative maps, it is easy to reduce the problem of existence of a σ-finite equivalent invariant measure to that of existence of finite equivalent invariant measures, in a way.

Theorem 6.7 (see [15])**.** *Let T be a conservative nonsingular transformation on a probability space $(X, \mathcal{B}, m)$. Then there is a σ-finite T-invariant measure $\mu \sim m$ if and only if there is an increasing sequence of sets $B_n \in \mathcal{B}$ with $\cup_{n=1}^{\infty} B_n = X$ such that for each n the first-return map T_{B_n} has a* finite *invariant measure equivalent to m restricted to B_n.*

Proof. Given a σ-finite equivalent invariant measure μ, let the B_n be sets of finite μ-measure that increase to X. Conversely, given such a sequence B_n with finite invariant measures μ_n for the first-return maps T_{B_n}, extend μ_1 in the obvious way to an (at least σ-finite) invariant measure on the full orbit $A_1 = \cup_{i=-\infty}^{\infty} T^i B_1$. Then replace B_2 by $B_2 \setminus A_1$, and continue. □

There are many more checkable conditions for existence of a σ-finite equivalent invariant measure in the literature. There are also examples of invertible ergodic nonsingular systems for which *there does not exist* any σ-finite equivalent invariant measure due to Ornstein [37] and subsequently Chacon [9], Brunel [5], L. Arnold [2], and others.

6.4. **Invariant measures for maps of the interval or circle.** Finally we mention sample theorems from a huge array of such results about existence of finite invariant measures for maps of an interval or of the circle.

Theorem 6.8 ("Folklore Theorem" [1]). *Let $X = (0,1)$ and denote by m Lebesgue measure on X. Let $T : X \to X$ be a map for which there is a finite or countable partition $\alpha = \{A_i\}$ of X into half-open intervals $[a_i, b_i)$ satisfying the following conditions. Denote by A_i^0 the interior of each interval A_i. Suppose that*

(1) for each i, $T : A_i^0 \to X$ is one-to-one and onto;

(2) T is $\mathcal{C}^2$ on each A_i^0;

(3) there is an n such that

$$\inf_i \inf_{x \in A_i^0} |(T^n)'(x)| > 1; \tag{6.18}$$

(4) for each i,

$$\sup_{x,y,z \in A_i^0} \left|\frac{T''(x)}{T'(y)T'(z)}\right| < \infty. \tag{6.19}$$

Then for each measurable set B, $\lim_{n\to\infty} m(T^{-n}B) = \mu(B)$ exists and defines the unique T-invariant ergodic probability measure on X that is equivalent to m.

Moreover, the partition α is weakly Bernoulli *for T, so that the natural extension of T is isomorphic to a Bernoulli system.*

A key example to which the theorem applies is that of the *Gauss map* $Tx = 1/x$ mod 1 with the partition for which $A_i = [1/(i+1), 1/i)$ for each $i = 1, 2, \ldots$. Coding orbits to $\mathbb{N}^{\mathbb{N}}$ by letting $a(x) = a_i$ if $x \in A_i$ carries T to the shift on the continued fraction expansion $[a_1, a_2, \ldots]$ of x. It was essentially known already to Gauss that T preserves the measure whose density with respect to Lebesgue measure is $1/((1+x)\log 2)$.

Theorem 6.9 (see [28]). *Let $X = S^1$, the unit circle, and let $T : X \to X$ be a (noninvertible) $\mathcal{C}^2$ map which is* expanding, *in the sense that $|T'(x)| > 1$ everywhere. Then there is a unique finite invariant measure μ equivalent to Lebesgue measure m, and in fact μ is ergodic and the Radon-Nikodym derivative $d\mu/dm$ has a continuous version.*

7. Constructions

We give examples of some of the ways that one can make new systems from old ones. Construction of measure-preserving systems from scratch, as by cutting and stacking, will also be discussed in the section on examples. Unless stated otherwise, in the following we will be discussing measure-preserving transformations on Lebesgue spaces (see section 9.1).

7.1. **Factors.** We say that a measure-preserving system $(Y, \mathcal{C}, \nu, S)$ is a *factor* of a measure-preserving system $(X, \mathcal{B}, \mu, T)$ if (possibly after deleting a set of measure 0 from X) there is a measurable map $\phi X \to Y$ which carries T to S and μ to ν, that is,

$$\begin{aligned} \phi^{-1}\mathcal{C} &\subset \mathcal{B}, \\ \phi T &= S\phi, \quad \text{and} \\ \mu T^{-1} &= \nu. \end{aligned} \tag{7.1}$$

For Lebesgue spaces, factors of $(X, \mathcal{B}, \mu, T)$ correspond perfectly with T-invariant complete sub-σ-algebras of $\mathcal{B}$. According to Rokhlin's theory of Lebesgue spaces [43] (see Section 9.1), factors also correspond perfectly to certain kinds of partitions of X.

7.2. **Isomorphisms.** We say that measure=preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \mu, S)$ are *isomorphic* if (possibly after deleting sets of measure 0 from X and Y) there is a one-to-one onto factor map $\phi : X \to Y$ whose inverse is a factor map $Y \to X$. A factor map $\phi : X \to Y$ between Lebesgue spaces is an isomrphism if and only if $\phi^{-1}\mathcal{C} = \mathcal{B}$ up to sets of measure 0.

7.3. **Products.** Given measure-preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$, their *product* consists of their completed product measure space with the transformation $T \times S : X \times Y \to X \times Y$ defined by $(T \times S)(x, y) = (Tx, Sy)$ for all $(x, y) \in X \times Y$. Given any family of measure-preserving transformations on probability spaces, their direct product is defined similarly.

7.4. **Skew products.** If $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system, $(Y, \mathcal{C}, \nu)$ is a probability space, and $\{S_x : x \in X\}$ is a family of measure-preserving maps $Y \to Y$ such that the map that takes (x, y) to $S_x y$ is jointly measurable in the two variables x and y, then there is a *skew product system* consisting of the product probability space of X and Y together with the map $T \ltimes S : X \times Y \to X \times Y$ defined by

$$(T \ltimes S)(x, y) = (Tx, S_x y). \tag{7.2}$$

Often Y is a group, ν is a measure on Y invariant under translations, and there is given a measurable function, or *cocycle*, $f : X \to Y$ which is used to define S_x by $S_x y = f(x)y$.

7.5. **Induced transformations.** Since by the Poincaré Recurrence Theorem measure-preserving transformations on probability spaces are recurrent, given any set B of positive measure it is possible to define, for almost all $x \in B$,

$$n_B(x) = \inf\{n \geq 1 : T^n x \in B\} \tag{7.3}$$

and

$$T_B x = T^{n_B(x)} x. \tag{7.4}$$

Then (after perhaps discarding as usual a set of measure 0) $T_B : B \to B$ is a measurable transformation which preserves the probability measure $\mu_B = \mu/\mu(B)$. The system $(B, \mathcal{B} \cap B, \mu_B, T_B)$ is called an *induced* or *derivative* transformation [32].

The construction of the derivative transformation T_B presents the forward orbit of B as a *tower* or *skyscraper* over B. For each $n = 1, 2, \ldots$, let

$$B_n = \{x \in B : n_B(x) = n\}. \tag{7.5}$$

Then $B_1, B_2, \ldots$ form a partition of B, which we think of as the bottom floor. The next floor is made up of $TB_2, TB_3, \ldots$, which form a partition of $TB \setminus B$, and so on. All these sets are disjoint. A *column* is a part of the tower of the form $B_n \cup TB_n \cup \cdots \cup T^{n-1}B_n$ for some $n = 1, 2, \ldots$. The action of T on the entire tower is pictured as mapping each x not at the top of its column straight up to the point Tx above it on the next level, and mapping each point on the top level of its column somewhere back in B.

The tower picture also suggests reversing the derivative construction to make another kind of induced transformation, sometimes called a *primitive* transformation. Given a measure-preserving system $(X, \mathcal{B}, \mu, T)$ and a finite or countable partition $X_1, X_2, \ldots$ of X into measurable sets, for each $n = 1, 2, \ldots$ we let $Y_n^0 = X_n$ and put $n-1$ copies $Y_n^1, \ldots, Y_n^{n-1}$ of X_n above X_n, then define $\hat{T}$ on the resulting naturally defined (but possibly infinite-measure) measure space $\hat{X}$ as in the preceding paragraph: if $\hat{x} \in \hat{Y}_n^k$ for some $k < n-1$, we define $\hat{T}\hat{x}$ to be the corresponding point in Y_n^{k+1}, while if $\hat{x} \in Y_n^{n-1}$, we map $\hat{x}$ to wherever T would map the $x \in X$ that is below $\hat{x}$: formulaically, $\hat{T}\hat{x} = T\hat{T}^{-n+1}\hat{x}$.

7.6. **Flow under a function.** The construction of primitive transformations has a sort of limiting version in which a $\mathbb{Z}$ action leads to an $\mathbb{R}$ action. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $f : X \to (0, \infty)$ a measurable "ceiling" function on X. The set

$$X^f = \{x, t) : 0 \le f(x) < t\}, \tag{7.6}$$

with measure given locally by the product of μ on X with Lebesgue measure m on $\mathbb{R}$, is a measure space in a natural way, with finite measure if f is integrable. We define an action of $\mathbb{R}$ on X^f by letting points flow at unit speed up the vertical lines $\{(x, t) : 0 \le t < f(x)\}$ under the graph of f until it hits the ceiling, then jump to Tx, and so on: defining $f_n(x) = f(x) + \cdots + f(T^n x)$,

(7.7)

$$T_s(x,t) = \begin{cases} (x, s+t) & \text{if } 0 \le s+t < f(x), \\ (Tx, s+t-f(x)) & \text{if } f(x) \le s+t < f(x)+f(Tx) \\ \cdots & \\ (T^n x, s+t-[f(x)+\cdots+f(T^{n-1}x)] & \text{if } f_{n-1}(x) \le s+t < f_n(x). \end{cases}$$

7.7. **Rokhlin's Lemma (with notes by J. Olli and N. Pennington).** The following result is the fundamental starting point for many constructions in ergodic theory, from representing arbitrary systems in terms of cutting and stacking or adic systems, to constructing useful partitions and symbolic codings of abstract systems, to connecting convergence theorems in abstract ergodic theory with those in harmonic analysis. It allows us to picture arbitrarily long stretches of the action of a measure-preserving transformation as translation within the set of integers. In the ergodic nonatomic case the statement follows readily from the construction of derivative transformations.

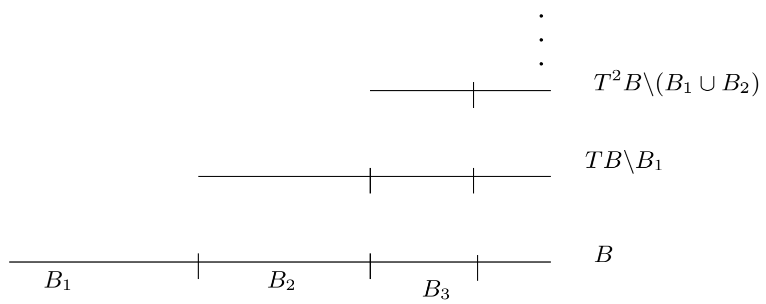
Lemma 7.1 (Rokhlin's Lemma)**.** *Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Suppose that $(X, \mathcal{B}, \mu)$ is nonatomic and $T : X \to X$ is ergodic, or, more generally, $(X, \mathcal{B}, \mu, T)$ is* aperiodic*: the set $\{x \in X : \text{ there is } n \in \mathbb{N} \text{ such that } T^n x = x\}$ of periodic points has measure 0. Then given $n \in \mathbb{N}$ and $\epsilon > 0$, there is a measurable set $B \subset X$ such that the sets $B, TB, \dots, T^{n-1}B$ are pairwise disjoint and $\mu(\cup_{k=0}^{n-1} T^k B) > 1 - \epsilon$.*

Definition 7.2. A measure space $(X, \mathcal{B}, \mu)$ is called *non-atomic* if it has no point masses. We recall that a measure space has no point masses if, for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists a $B \subset A$ with $B \in \mathcal{B}$ such that $0 < \mu(B) < \mu(A)$.

If a measure space has no point masses then, given any $A \in \mathcal{B}$ with $\mu(A) > 0$ and any $t \in [0, \mu(A)]$, there exists an $A_t \in \mathcal{B}$ such that $A_t \subset A$ and $\mu(A_t) = t$.

We mention that the unit interval with Lebesgue measure has this property. For more elaboration, see 7.1 of the typed class notes.

Now let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $T : X \to X$ be ergodic. Also let $B \in \mathcal{B}$, with $\mu(B) > 0$. Then we can define the *derivative transformation* (also called the *first return map*) $T_B : B \to B$ in the following manner. Defining $n_B(x) = \inf\{n : T^n x \in B\}$, we have $T_B(x) = T^{n_B(x)} x$ for all $x \in B$. X decomposes into a *tower* over B, so that $\mathcal{X}$ is the primitive transformation (see 7.5) over $(B, \mathcal{B} \cap B, \mu_B = \mu/\mu(B), T_B)$.



Here $B_n = \{x \in B : T^n x \in B, T^j x \notin B, 1 < j < n\}$.

We next observe that $T(B \cup TB \cup T^2 B \cup ...) \subset B \cup TB \cup T^2 B \cup ...$, and since both sides are sets of positive measure, this implies that $B \cup TB \cup ... = X$. We are now ready for Rokhlin's Lemma.

Lemma 7.3 (Rokhlin)**.** *Let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be a non-atomic measure-preserving ergodic system. Then given any $N = 1, 2, ...$ and any $\epsilon > 0$ there is a set $F \in \mathcal{B}$ such that $F, TF, ..., T^{N-1}F$ are pairwise disjoint and $\mu(\cup_{k=0}^{N-1} T^k F) > 1 - \epsilon$.*

Before proving the lemma, we note that this lemma leads to proofs in the areas of coding, Ornstein theory, combinatorial ergodic theory and ergodic theorems by transference. We also note that the lemma holds in the non-ergodic case if T is *aperiodic*, in the sense that for all n, $\mu\{x : T^n x = x\} = 0$.

Proof. Let $N \geq 1$ and let $\epsilon > 0$. Choose a $B \in \mathcal{B}$ such that $0 < \mu(B) << 1$. Decompose X as a tower over B, as above.

Let $n > N$ be arbitrary, and define $k = [\frac{n-N}{N}]$, where $[\cdot]$ is the least integer function. Then we define F_n as

$$F_n = B_n \cup T^N B_n \cup ... \cup T^{kN} B_n. \tag{7.8}$$

Then define $F = \bigcup_{n>N} F_n$, so that $F, TF, ..., T^{N-1}F$ are pairwise disjoint. Setting $W = \cup_{k=0}^{N-1} T^k F$, we have that

$$\mu(W^c) \leq \sum_{n=1}^{\infty} N\mu(B_n) = N\mu(B). \tag{7.9}$$

By choosing $\mu(B) < \epsilon/n$, it follows that $\mu(F \cup TF \cup ... \cup T^{N-1}) > 1 - \epsilon$, which finishes the proof. □

7.8. **Inverse limit.** Suppose that for each $i = 1, 2, \dots$ we have a Lebesgue probability space $(X_i, \mathcal{B}_i, \mu_i)$ and a measure-preserving transformation $T_i : X_i \to X_i$. Suppose also that for each $i < j$ there is a factor map $\phi_{ji} : (X_j, \mathcal{B}_j, \mu_j, T_j) \to (X_i, \mathcal{B}_i, \mu_i, T_i)$, such that each ϕ_{jj} is the identity on X_j and $\phi_{ji}\phi_{kj} = \phi_{ki}$ whenever $k \geq j \geq i$. Let

$$X = \{x \in \Pi_{i=1}^{\infty} X_i : \phi_{ji} x_j = x_i \text{ for all } j \geq i\}. \tag{7.10}$$

For each j, let $\pi_j : X \to X_j$ be the projection defined by $\pi_j x = x_j$.

Let $\mathcal{B}$ be the smallest σ-algebra of subsets of X which contains all the $\pi_j^{-1}\mathcal{B}_j$. Define μ on each $\pi_j^{-1}\mathcal{B}_j$ by

$$\mu(\pi_j^{-1} B) = \mu_j(B) \quad \text{for all } B \in \mathcal{B}_j. \tag{7.11}$$

Because $\phi_{ji}\pi_j = \pi_i$ for all $j \geq i$, the $\pi_j^{-1}\mathcal{B}_j$ are increasing, and so their union is an algebra. The set function μ can, with some difficulty, be shown to be countably additive on this algebra: since we are dealing with Lebesgue spaces, by means of measure-theoretic isomorphisms it is possible to replace the entire situation by compact metric spaces and continuous maps, then use regularity of the measures involved–see [39, p. 137 ff.]. Thus by Carathéodory's Theorem (3.3) μ extends to all of $\mathcal{B}$. We generally take the completion as well.

Define $T : X \to X$ by $T(x_j) = (Tx_j)$. Then $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system which has all the $(X_j, \mathcal{B}_j, \mu_j, T_j)$ as factors, and any system that factors onto all the $(X_j, \mathcal{B}_j, \mu_j T_j)$ also factors onto $(X, \mathcal{B}, \mu, T)$.

7.9. **Natural extension.** The construction of the preceding section can be used to produce a natural invertible version of a noninvertible system. Let $(X, \mathcal{B}, \mu)$ be a Lebesgue probability space and $T : X \to X$ a map such that $T^{-1}\mathcal{B} \subset \mathcal{B}$ and $\mu T^{-1} = \mu$. For each $i = 1, 2, \dots$ let $(X_i, \mathcal{B}_i, \mu_i, T_i) = (X, \mathcal{B}, \mu, T)$, and $\phi_{ji} = T^{j-i}$ for each $j > i$. Then the inverse limit $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu}, \hat{T})$ of this system is an *invertible* measure-preserving system called the *natural extension* of $(X, \mathcal{B}, \mu, T)$. We have

$$\hat{T}^{-1}(x_1, x_2, \dots) = (x_2, x_3, \dots). \tag{7.12}$$

The original system $(X, \mathcal{B}, \mu, T)$ is a factor of $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu}, \hat{T})$ (using any π_i as the factor map), and any factor mapping from an invertible system onto $(X, \mathcal{B}, \mu, T)$ consists of a factor mapping onto $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu}, \hat{T})$ followed by projection onto the first coordinate.

7.10. **Joinings.** Given measure-preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$, a *joining* of the two systems is a $T \times S$-invariant probability measure on their product measurable space that projects to μ and ν, respectively, under the projections of $X \times Y$ to X and Y, respectively. This concept is the ergodic-theoretic version of the notion in probability theory of a *coupling*. The product measure $\mu \times \nu$ is always a joining of the two systems. If product measure is the *only* joining of the two systems, then we say that they are *disjoint* and write $X \perp Y$ [16]. If $\mathcal{D}$ is any family of systems, we write $\mathcal{D}^{\perp}$ for the family of all measure-preserving systems which are disjoint from every system in $\mathcal{D}$. Extensive recent accounts of the use of joinings in ergodic theory are in [18, 47, 52]; here we mention just a few selected highlights.

Denote by $\mathcal{E}$ the family of all ergodic systems, by $\mathcal{W}$ the family of all weakly mixing systems, by $\mathcal{K}$ the family of all K-systems, by $\mathcal{I}$ the family consisting of the identity transformation on $[0, 1]$, by $\mathcal{R}$ the family consisting of all ergodic translations on compact abelian groups, and by $\mathcal{D}$ the family of all ergodic entropy-zero systems. Then

$$\begin{aligned} \mathcal{E} &= \mathcal{I}^{\perp}, \\ \mathcal{W} &= \mathcal{R}^{\perp} \cap \mathcal{E}, \\ \mathcal{K} &= \mathcal{D}^{\perp} \cap \mathcal{E}. \end{aligned} \tag{7.13}$$

A joining of $(X, \mathcal{B}, \mu, T)$ with itself is, naturally, called a *self-joining*. Besides the product self-joining, there are also the *diagonal self-joining* $\lambda_{\triangle}$ defined by setting, for all bounded measurable functions f on $X \times X$,

$$\int_{X \times X} f(x_1, x_2)\, d\lambda_{\triangle} = \int_X f(x, x)\, d\mu, \tag{7.14}$$

and the *off-diagonal joinings* λ_j defined for $j \in \mathbb{Z}, j \neq 0$, by

$$\int_{X \times X} f(x_1, x_2)\, d\lambda_j = \int_X f(x, T^j x)\, d\mu \tag{7.15}$$

for all bounded measurable f on $X \times X$. Equivalently, for all measurable $A, B \subset X$,

$$\lambda(A \times B) = \mu(A \cap T^{-j} B). \tag{7.16}$$

If these are the only *ergodic* self-joinings of $(X, \mathcal{B}, \mu, T)$, then we say that $(X, \mathcal{B}, \mu, T)$ has *2-fold minimal self-joinings* [46].

Theorem 7.4 [46]**.** *If $(X, \mathcal{B}, \mu, T)$ has 2-fold minimal self-joinings and is* totally ergodic *(every iterate T^k is an ergodic transformation), then the only measure-preserving transformations that commute with T are the powers of T, and $(X, \mathcal{B}, \mu, T)$ has no proper factors (every factor is either trivial or isomorphic to $(X, \mathcal{B}, \mu, T)$).*

Higher Cartesian powers of $(X, \mathcal{B}, \mu, T)$ admit a variety of natural joinings. First, for every choice of $j_1, \ldots, j_n \in \mathbb{Z}$ there is the off-diagonal (or diagonal if all the $j_i = 0$) joining $\lambda_{j_1 \ldots j_n}$ on X^n defined by

$$\int_{X^n} f(x_1, \ldots, x_n)\, d\lambda_{j_1 \ldots j_n} = \int_X f(T^{j_1} x, \ldots, T^{j_n} x)\, d\mu \tag{7.17}$$

for all bounded measurable f on X. We can also form products of joinings of this kind, each of which is defined on a subset of the set of coordinates $\{1, \ldots, n\}$. For each i let $(X_i, \mathcal{B}_i, \mu_i, T_i) = (X, \mathcal{B}, \mu, T)$. For $S = \{s_1 < s_2 < \cdots < s_r\} \subset \{1, \ldots, n\}$ and $j_i \in \mathbb{Z}, i = 1, \ldots, r$, define λ_S on $\prod_{i=1}^r X_{s_i}$ by

$$\lambda_S(A_1 \times \cdots \times A_r) = \mu(T^{j_1} A_1 \cap \cdots \cap T^{j_r} A_r). \tag{7.18}$$

Then for a partition $\{S_1, \ldots, S_m\}$ of $\{1, \ldots, n\}$, we can define λ on X^n by

$$\lambda = \lambda_{S_1} \times \cdots \times \lambda_{S_m}. \tag{7.19}$$

We say that $(X, \mathcal{B}, \mu, T)$ has *n-fold minimal self-joinings* if the only ergodic self-joinings of $(X, \mathcal{B}, \mu, T)$ are such products of (diagonal and) off-diagonal joinings. Rudolph [46] showed, extending the argument of Ornstein [38], that there exist systems with minimal self-joinings of all orders, and that they have astounding dynamical properties which lead to many counterexamples to natural conjectures in ergodic theory, for example a transformation without roots, a transformation with two nonisomorphic square roots, and two systems lacking any common factor that are nevertheless not disjoint.

That the explicit Chacon example has minimal self-joinings of all orders was proved by del Junco, Rahe, and Swanson [11]. King [34] showed that if a system has minimal self-joinings of order 4, then it has minimal self-joinings of all orders. Glasner, Host, and Rudolph [19] strengthened this, showing that minimal self-joinings of order 3 implies minimal self-joinings of all orders.

7.10.1. *Introduction to joinings (based on lecture notes by A. Del Junco, notes by J. Olli and N. Pennington).* Our goal in this introductory section will be to define joinings and list a few of their basic properties. We begin by establishing the convention that $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ are probability spaces with X and Y compact metric spaces.

Definition 7.5. A *joining of μ and ν* is a measure λ on the space $X \times Y$ with *marginals* μ and ν. We call μ and ν *marginals of* λ if $\pi_X \lambda = \mu$ and $\pi_Y \lambda = \nu$, where π is the usual projection map.

Let $\mathcal{H} = \pi_X^{-1}\mathcal{B}$ and $\mathcal{V} = \pi_Y^{-1}\mathcal{C}$. Then $(X, \mathcal{B}, \mu) \approx (X \times Y, \mathcal{H}, \lambda)$ and $(Y, \mathcal{C}, \nu) \approx (X \times Y, \mathcal{V}, \lambda)$. We also have that for $A \subset X$, $\pi_X \lambda(A) = \mu(A)$.

As an immediate observation, we have that $L^2(\mu)$ embeds into $L^2(\lambda)$ by $f \mapsto f \otimes 1$, where $(f \otimes g)(x, y) = f(x)g(y)$. We also have that $\int_X f \, d\mu = \mu(f) = \lambda(f \otimes 1)$ for all $f \in L^2(\mu)$.

We define an operator $P_\lambda : L^2(\mu) \to L^2(\nu)$ by $f \mapsto P_{L^2(\nu)}(f \otimes 1)$. The projection $P_{L^2(\nu)}$ is the conditional expectation operator $E(\cdot|\mathcal{V})$.

If $(X, \mu) = (Y, \nu)$, then one can define a diagonal measure μ_Δ by $\int_{X \times X} f(x_1, x_2) d\lambda = \int_X f(x, x) d\mu_\Delta$. Then $P_{\mu_\Delta} = Id_{L^2(\mu)}$.

We next state a few definitions. Let $\mathcal{X} = (X, \mu, T)$ and $\mathcal{Y} = (Y, \nu, S)$ be measure-preserving systems.

Definition 7.6. A *joining of $\mathcal{X}$ and $\mathcal{Y}$* is a joining λ of μ and ν which is $T \times S$-invariant.

Note that $(X \times Y, T \times S, \lambda)$ is a measure-preserving system and that π_X and π_Y are factor maps onto $\mathcal{X}$ and $\mathcal{Y}$.

Definition 7.7. We say that $\mathcal{X}$ and $\mathcal{Y}$ are *disjoint* if and only if $\lambda = \mu \times \nu$ is the only joining of $\mathcal{X}$ and $\mathcal{Y}$. We will denote $\mathcal{X}$ being disjoint from $\mathcal{Y}$ by $\mathcal{X} \perp \mathcal{Y}$.

Remark 7.8. In the context of these notes, T will also be the notation used for the unitary operator associated with the transformation T. That is, $Tf = f \circ T$.

Our next observation is that $L^2(\mu)$ and $L^2(\nu)$ are $T \times S$-invariant subspaces of $L^2(\lambda)$, so $P_{L^2(\nu)}$ commutes with the unitary operator $T \times S$ on $L^2(\lambda)$ and hence $P_\lambda T = S P_\lambda$, where both are operators from $L^2(\mu)$ to $L^2(\nu)$. Conversely, given an operator $P : L^2(\mu) \to L^2(\nu)$ such that $\|P\|_2 \leq 1, f \geq 0$ implies that $Pf \geq 0$, and $P1 = 1$, we can construct a joining λ for which $P = P_\lambda$.

Relating the measure λ to the inner product on L^2, we have that $\lambda(f \otimes g) = < P_\lambda f, g >_{L^2(\nu)}$. This can be shown by the following:

$$
\begin{aligned}
\lambda(f \otimes g) &= < f \otimes 1, 1 \otimes g >_{L^2(\lambda)} \\
&= < P_{L^2(\nu)}(f \otimes 1), 1 \otimes g > \qquad (7.20) \\
&= < P_\lambda f, g >_{L^2(\nu)} .
\end{aligned}
$$

Let $J(\mathcal{X}, \mathcal{Y})$ ={joinings of $\mathcal{X}$ and $\mathcal{Y}$} $\subset \mathcal{C}(X \times Y)^*$. Then we have the weak-* topology on $J(\mathcal{X}, \mathcal{Y})$, and the space is compact under this topology.

Proposition 7.9. *$\lambda_n \to \lambda$ in $J(\mathcal{X}, \mathcal{Y})$ if and only if $\lambda_n(A \times B) \to \lambda(A \times B)$ for all measurable sets $A \subset X, B \subset Y$.*

The main details of the proof will be omitted. However, the proof of the forward implication involves the following two observations:

(1) $\mathcal{C}(X)$ is dense in $L^2(\mu)$ and $\mathcal{C}(Y)$ is dense in $L^2(\nu)$.
(2) The Cauchy-Schwarz inequality gives that if $f \in L^2(\mu)$ and $g \in L^2(\nu)$ then

$$|\lambda(f \otimes g)| = |< f \otimes 1, 1 \otimes g >_{L^2(\lambda)}| \leq \|f\|_{L^2(\mu)} \cdot \|g\|_{L^2(\nu)}. \tag{7.21}$$

This shows that $f \otimes g$ is in $L^1(\lambda)$. Now (1) and (2) together imply that the convergence can be extended from continuous functions to L^2 functions. That is, $\lambda_n(f \otimes g) \to \lambda(f \otimes g)$ for all $f \in L^2(\mu), g \in L^2(\nu)$. We omit the details of the proof of the reverse implication.

7.10.2. *Some applications of joinings (based on lecture notes by A. del Junco, notes by J. Olli and N. Pennington).*

7.10.2.1. The Weak Ergodic Theorem. Our first application is the "Weak Ergodic Theorem." Choose measure-preserving systems $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ and $\mathcal{Y} = (Y, \mathcal{C}, \nu, S)$.

Theorem 7.10 (Weak Ergodic Theorem)**.** *If T is ergodic, then*

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu(T^n A \cap B) \to \mu(A)\mu(B) \tag{7.22}$$

for all $A, B \in \mathcal{B}$.

To prove this theorem, we require the following result:

Theorem 7.11. *$id_X \perp S$ if S is ergodic.*

Proof. Let λ be a joining of μ and ν that is invariant under $id \times S$, so $(id \times S)\lambda = \lambda$.

For $f \in L^2(\mu)$, $f \otimes 1$ is $id \times S$-invariant. If follows that $P_\lambda(f)$ is S-invariant. Because S is ergodic, we get that $P_\lambda(f)$ is constant a.e. Therefore, $P_\lambda(f) = \mu(f)1_Y$.

For f as above and $g \in L^2(\nu)$,

$$\lambda(f \otimes g) =< P_\lambda(f), g >_{L^2(\nu)}= \mu(f) < 1_Y, f >_{L^2(\nu)}= \mu(f)\nu(g). \tag{7.23}$$

So the joining λ must be product measure. Therefore, $id \perp S$. □

We will now prove the Weak Ergodic Theorem.

Proof. Consider the measure μ_Δ on $X \times X$. What needs to be shown is equivalent to showing that

$$\mu_N = \frac{1}{N}\sum_{n=0}^{N-1}(T^n \times id_X)\mu_\Delta \to \mu \times \mu. \tag{7.24}$$

That is, we want to know the limit of $\mu_N(A \times B)$ for $A, B \in \mathcal{B}$. Note that

$$\mu_\Delta(C \times D) = \mu(C \cap D). \tag{7.25}$$

Let λ be a limit point of $\{\mu_N\}$. If suffices to show that the only possibility for λ is $\mu\times\mu$. We have that μ_N is approximately invariant for large N, which implies that λ is invariant under $T \times id_X$. Also, λ projects to μ on both coordinates, since each μ_N does. Thus, λ is a joining of (X, μ, T) and (X, μ, id_X). (Each μ_N is a joining of (X, μ, T) and (X, μ, id_X) and the set of joinings is closed.) T is ergodic by hypothesis, so we can apply the previous theorem and conclude that $\lambda = \mu \times \mu$. □

Building on this concept, we have a theorem of Furstenberg, but first we require a definition.

Definition 7.12. T is *weak mixing* if $T \times T$ is ergodic. Equivalently, T is *weak mixing* if there exists a sequence $\{n_i\}$ such that $\mu(T^{n_i}A \cap B) \to \mu(A)\mu(B)$ for all measurable sets A, B.

Note that if this condition is satisfied and $f \in L^2(\mu)$ is a nonconstant eigenfunction with eigenvalue ξ then $\xi^{n_i} < f, f >=< T^{n_i}f, f >\to< f, 1 >< 1, f >= 0$, so that $f = 0$ in $L^2(\mu)$.

Theorem 7.13 (Furstenberg)**.** *If T is weak mixing, then*

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu(A \cap T^nB \cap T^{2n}C) \to \mu(A)\mu(B)\mu(C) \tag{7.26}$$

for all measurable sets A, B, and C.

Proof (Ryzhikov [48]*).* Let μ_Δ be the diagonal measure on $X \times X \times X$. Then the statement to be proved is equivalent to

$$A_N = \frac{1}{N}\sum_{n=0}^{N-1}(id_X \times T \times T^2)^n\mu_\Delta \to \mu \times \mu \times \mu. \tag{7.27}$$

Take λ to be a limit point of $\{A_N\}$. If λ is projected onto the second and third coordinates,

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu(T^nB \cap T^{2n}C) = \frac{1}{N}\sum_{n=0}^{N-1}\mu(B \cap T^nC) \to \mu(B)\mu(C) \tag{7.28}$$

because T is weak mixing. We have that λ is a joining of (id_X, μ) and $(T\times T^2, \mu\times\mu)$. Because T is weak mixing, so are T^2 and $T \times T^2$. Therefore $T \times T^2$ is also ergodic. From the previous theorem, we have that $\lambda = \mu \times \mu \times \mu$. □

Remark 7.14. This result extends to the intersection of an arbitrary number of sets. That is, weak mixing implies "weak mixing of all orders."

Remark 7.15. J.-P. Thouvenot (private communication) points out that following the scheme of this argument while working over the Kronecker factor, one can prove an analogous statement, and then Roth's theorem concerning arithmetic progressions of length three in positive-density subsets of $\mathbb{N}$ is an easy consequence.

7.10.2.2. Discrete Spectrum Theorem. Our second application is the Discrete Spectrum Theorem of P. Halmos and J. von Neumann. Let $\mathcal{X}$ and $\mathcal{Y}$ be measure-preserving systems as before.

Definition 7.16. T has *discrete spectrum* if the eigenfunctions of T span $L^2(\mu)$.

Definition 7.17. Suppose $\phi : (X, \mu) \to (Y, \nu)$ is a one-to-one onto measurable map with a measurable inverse such that $\phi\mu = \nu$ and $\phi T \phi^{-1} = S$. Then ϕ is called an *isomorphism* or *conjugacy* of T and S.

Given an isomorphism $\phi : (X, \mu) \to (Y, \nu)$, define a measure μ_ϕ on $X \times Y$ by $\mu_\phi(A \times B) = \mu(A \cap \phi^{-1}B)$. Then μ_ϕ is a joining of X and Y with the property that $\mathcal{B} \times Y = X \times \mathcal{C}$ modulo μ_ϕ-null sets. Any joining λ with the property that $\mathcal{B} \times Y = X \times \mathcal{C}$ modulo λ-null sets is called *graphic*.

Theorem 7.18 (von Neumann)**.** *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Lebesgue spaces and λ is a graphic joining of $\mathcal{X}$ and $\mathcal{Y}$. Then there is an isomorphism $\phi : (X, \mu) \to (Y, \nu)$ such that $\lambda = \mu_\phi$.*

Let $J(T, S)$ denote the set of joinings of T and S and let $J_e(T, S)$ denote the ergodic joinings of T and S.

Theorem 7.19. *If T and S are ergodic, then $J_e(T, S) \neq \emptyset$.*

Proof. $J(T, S)$ is a compact, convex set, so the Krein-Milman Theorem implies that $J(T, S)$ has at least one extreme point, say λ. We claim that λ is ergodic. Assume that there is a set $A \subset X \times Y$ which is $T \times S$-invariant and $0 < \mu(A) < 1$. Define $\lambda_A(E) = \lambda(E|A) = \lambda(E \cap A)/\lambda(A)$ for all measurable $E \subset X \times Y$, and similarly define λ_{A^c}. We know that $\lambda(A)\lambda_A + (1 - \lambda(A))\lambda_{A^C} = \lambda$. Projecting these measures onto $\mathcal{X}$ gives

$$\lambda(A) \cdot \pi_X \lambda_A + (1 - \lambda(A))\pi_X \lambda_{A^C} = \mu. \tag{7.29}$$

Since μ is ergodic, this implies that $\pi_X \lambda_A$ and $\pi_X \lambda_{A^C}$ are both equal to μ. Similarly, $\pi_Y \lambda_A = \pi_Y \lambda_{A^c} = \nu$. This would imply that λ can be expressed as a nontrivial sum of two other joinings, λ_A and λ_{A^c}, which is impossible because λ is an extreme point. Hence λ must be ergodic. □

Now we are ready to prove the Discrete Spectrum Theorem. Denote by $G(T)$ the set of eigenvalues of T on $L^2(\mu)$.

Theorem 7.20 Discrete Spectrum Theorem (Halmos-von Neumann)**.** *Let T, S be ergodic transformations on Lebesgue spaces (X, μ) and (Y, ν), respectively. If T and S have discrete spectrum and $G(T) = G(S)$, then T and S are conjugate (isomorphic).*

Proof (Lemanczyk). Take $\lambda \in J_e(T, S)$ and $\alpha \in G(T)$. Then there exists an $f \in L^2(\mu)$ such that $Tf = \alpha f$ and a $g \in L^2(\nu)$ such that $Sg = \alpha g$. Now let $f' = f \otimes 1$, $g' = 1 \otimes g$, and $R = T \times S$. Then $Rf' = \alpha f'$ and $Rg' = \alpha g'$. The modulus of an eigenfunction is constant and R is ergodic with respect to λ, so there is a constant $c \neq 0$ such that $f' = cg'$ λ-a.e. Since the eigenfunctions span $L^2(\mu)$ and $L^2(\nu)$, it follows that $L^2(\mu) \otimes 1 = 1 \otimes L^2(\nu) \subset L^2(\lambda)$, so $\mathcal{B} \times Y = X \times \mathcal{C}$. □

7.10.2.3. Examples of disjointness. Recall that T and S being disjoint means that $J(T, S) = \{\mu \times \nu\}$, which is true if and only if $J_e(T, S) = \{\mu \times \nu\}$. We begin with some definitions of some types of maps.

Definition 7.21. A map T is called *rigid* if there exists a sequence $\{n_i\} \to \infty$ such that $T^{n_i} f \to f$ in L^2 for all $f \in L^2$. Equivalently, $\mu(T^{n_i} A \Delta A) \to 0$ for all $A \in \mathcal{B}$.

Definition 7.22. A map T is called *mild mixing* if it has no proper rigid factors. Two equivalent conditions are that there are no nonconstant rigid functions and no proper rigid sets.

We now give some examples of pairs of maps that are disjoint.

(0) $id \perp$ every ergodic S
(1) discrete spectrum maps $\perp$ weakly mixing maps
(2) rigid maps $\perp$ mild mixing maps
(3) 0 entropy maps $\perp$ K-automorphisms

For the proof of (0) see Theorem 7.11. We will also prove (1), omitting the proofs of the other items.

Proof. (Sketch) Suppose T has discrete spectrum and S is weak mixing. Take $f \in L_0^2(\mu) = \{f \in L^2(\mu) : \int_X f d\mu = 0\}$ such that $Tf = \alpha f$, where $\alpha \in G(T)$, and take $g \in L_0^2(\nu)$. Now let $f' = f \otimes 1$, $g' = 1 \otimes g$, and $R = T \times S$. Then $Rf' = \alpha f'$, and there exists a sequence $\{n_i\}$ such that $< S^{n_i} g, h > \to 0$, for all $h \in L_0^2(\nu)$. Without loss of generality, assume that $\alpha^{n_i} \to \beta \in \mathbb{T}$. Then $< R^{n_i} f', R^{n_i} g' > =$ $< f', g' >$ since R is unitary. Thus

$$< f', g' > = < R^{n_i} f', R^{n_i} g' > = \alpha^{n_i} < f', R^{n_i} g' > \to \beta \cdot 0 = 0. \tag{7.30}$$

So $< f', g' > = 0$, and $(L_0^2(\mu) \otimes 1) \perp (1 \otimes L_0^2(\nu))$ in $L^2(\lambda)$. Consequently, the σ-algebras $\pi^{-1}\mathcal{B} = \mathcal{B} \times Y$ and $\pi^{-1}\mathcal{C} = X \times \mathcal{C}$ are independent with respect to λ, which means that λ is product measure on $X \times Y$. □

7.10.2.4. 2-Fold mixing. Let μ_Δ be the diagonal measure on $X^3 = X \times X \times X$. By 2-fold mixing we mean that $(id \times T \times T^2)^n \mu_\Delta \to \mu \times \mu \times \mu$. To check that this holds, it suffices to show that any limit point λ must be the product measure.

Rohlin's Problem asks, does mixing imply 2-fold mixing? That is, does mixing imply that $\mu(A \cap T^n B \cap T^{2n} C) \to \mu(A)\mu(B)\mu(C)$ for all $A, B, C \in \mathcal{B}$? Note that if T is weak mixing, then each 2-dimensional projection of any limit point λ is $\mu \times \mu$.

Definition 7.23. T is *2-fold independently determined* (2-ID) if whenever $\lambda \in J(T, T, T)$ has each 2-dimensional projection equal to $\mu \times \mu$, then $\lambda = \mu \times \mu \times \mu$.

A result of Host states that if the spectral type of T is singular and continuous, then T is 2-ID and hence if it is mixing it is 2-fold mixing. A result of Ryzhikov states that if T is finite-rank mixing then T is 2-ID. One question that remains unanswered is: Are there any 0-entropy weak mixing non 2-ID automorphisms?

7.10.3. *Disjointess (notes by J. Olli and N. Pennington).* We now extend Theorem 7.11 above.

Theorem 7.24. *The set of ergodic measure-preserving systems consists of all those systems which are disjoint from the system* $([0,1], Id)$ *with Lebesgue measure. (Recall that systems are disjoint if the only joining between them is the product measure).*

Proof. Suppose that $(X, \mathcal{B}, \mu, T)$ is ergodic and that λ on $X \times I$ is a joining, where I is the unit interval $[0, 1]$.

Let $f \in L^2(X, \mathcal{B}, \mu)$, and let $g \in L^2(I)$. Recall that $(f \otimes g)(x, y) = f(x)g(y)$ and that these span $L^2(\lambda)$. When $f = 1_A$ and $g = 1_B$, we have that $\lambda(f \otimes g) = \mu(A)m(B)$.

With this understanding of $(f \otimes g)$, we observe that for each n and $h = 0, 1, .., n-1$

$$\begin{aligned} \int_{X \times Y} (f \otimes g) d\lambda &= \int (T \times I)^k (f \otimes g) d\lambda \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \int (f \otimes g)(T^k x, y) d\lambda(x, y) \\ &= \int \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) g(y) d\lambda(x, y). \end{aligned} \tag{7.31}$$

Ergodicity of T implies that $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ converges to $\bar{f}(x) = \int_X f d\mu$ almost everywhere $d\mu$. Here, $\bar{f}$ is the average value of f, not the complex conjugate.

Letting $n \to \infty$, we get

$$\begin{aligned} \int_{X\times Y} (f \otimes g) d\lambda &= \int_{X\times Y} \bar{f}(x) g(y) d\lambda(x,y) \\ &= \int_{X\times Y} \left(\int_X f d\mu \right) g(y) d\lambda(x,y) \\ &= \int_X f d\mu \int_{X\times Y} g(y) d\lambda(x,y) \\ &= \int_X f d\mu \int_Y g dm. \end{aligned} \tag{7.32}$$

Now we prove the other direction. We begin by assuming that $T : X \to X$ is not ergodic, and our goal is to construct a joining that is not product measure.

Since T is not ergodic, there exists a T-invariant set $A \in \mathcal{B}$ with $0 < \mu(A) < 1$. We will use A to make a joining λ of T and id_I which is not $\mu \times m$.

Begin by taking an interval $J \subset I$ with $m(J) = \mu(A)$. Let $P_1 = (A \times J) \subset (X \times I)$ and $P_2 = (A^c \times J^c) \subset (X \times I)$. Then these are both $T \times id$-invariant sets, and we put on $X \times Y$ the measure

$$\lambda = \mu(A)(\mu \times m)_{P_1} + \mu(A^c)(\mu \times m)_{P_2}, \tag{7.33}$$

where

$$(\mu \times m)_E(\cdot) = \frac{(\mu \times m)(\cdot \cap E)}{(\mu \times m)(E)}. \tag{7.34}$$

Since P_1 and P_2 are $T \times id_I$ invariant, λ is a convex combination of invariant probability measures, so it is itself an invariant probability measure.

We also note that $\lambda \equiv 0$ on $A^c \times J$ and on $A \times J^c$, so $\lambda \neq \mu \times m$. To finish the proof, we just need to verify that λ is a joining.

To show this, we let $E \in \mathcal{B}$. Then, abbreviating $\mu \times m = \nu$, we have

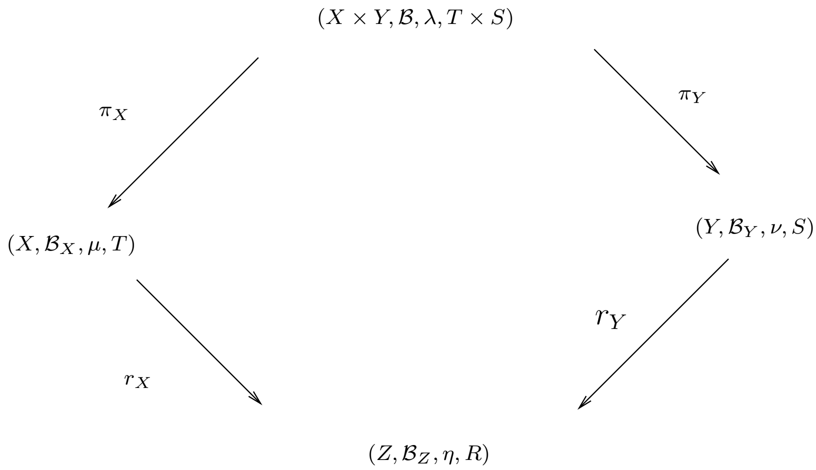
$$\begin{aligned} \lambda(\pi_X^{-1} E) &= \lambda(E \times I) \\ &= \mu(A)\nu_{P_1}(E \times I) + \mu(A^c)\nu_{P_2}(E \times I) \\ &= \mu(A)\frac{\mu(A \cap E)m(J)}{\mu(A)m(J)} + \mu(A^c)\frac{\mu(A^c \cap E)m(J^c)}{\mu(A^c)m(J^c)}. \end{aligned} \tag{7.35}$$

But by our construction, $m(J) = \mu(A)$ and $m(J^c) = \mu(A^c)$, so we get

$$\lambda(\pi_X^{-1} E) = \mu(A \cap E) + \mu(A^c \cap E) = \mu(E). \tag{7.36}$$

Similarly, $\lambda(\pi_I^{-1}) = m$. Since we have constructed a joining that is not product measure, the theorem is done. This λ is an example of a *relatively independent joining*, as will be shown in the next section. $\square$

7.10.4. *Relatively independent joinings over a common factor (notes by J. Olli and N. Pennington).* Suppose that $(X, \mathcal{B}_X, \mu, T)$ and $(Y, \mathcal{B}_Y, \nu, S)$ have a common factor $(Z, \mathcal{B}_Z, \eta, R)$:



The σ-algebra $\mathcal{B}_Z$ is viewed as a sub-σ-algebra of each of $\mathcal{B}_X, \mathcal{B}_Y$ by pulling back: we identify B_Z with $r_X^{-1}\mathcal{B}_Z \subset \mathcal{B}_X$, and similarly we identify B_Z with $r_Y^{-1}\mathcal{B}_Z \subset \mathcal{B}_Y$. Pulling back farther with π_X^{-1} and π_Y^{-1}, each of $\mathcal{B}_Z$, $\mathcal{B}_X$, and $\mathcal{B}_Y$ is identified with an invariant sub-σ-algebra of $\mathcal{B}_X \otimes \mathcal{B}_Y$.

Using this identification, $\mathcal{B}_X = \{W \times Y : W \in \mathcal{B}_X\} = \pi_X^{-1}(\mathcal{B}_X) = \mathcal{H}$, and similarly we have $\mathcal{B}_Y = \{X \times V : V \in \mathcal{B}_Y\} = \pi_Y^{-1}(\mathcal{B}_Y) = \mathcal{V}$.

More on the identification of factors with sub-σ-algebras can be found in the typed notes in the section on Rokhlin theory (see 9.1.1).

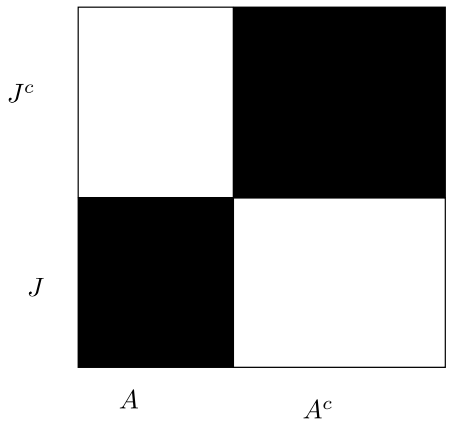
We will define a joining λ which makes $\mathcal{B}_X$ and $\mathcal{B}_Y$ "relatively independent over $\mathcal{B}_Z$". Given $V \in \mathcal{B}_X$ and $W \in \mathcal{B}_Y$, define

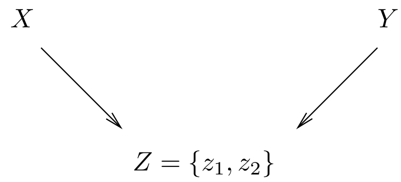
$$\lambda(V \times W) = \int_Z E(1_V | r_X^{-1}\mathcal{B}_Z)(z) E(1_W | r_Y^{-1}\mathcal{B}_Z)(z) d\eta(z). \tag{7.37}$$

Here $E(1_V | r_X^{-1}\mathcal{B}_Z)$ is viewed as a function on X (or $X \times Y$) over any fixed z and is constant on almost every fiber $r_X^{-1}\{z\}$. Analogous statements hold for $E(1_W | r_Y^{-1}\mathcal{B}_Z)$.

In Theorem 7.24 above, we dealt with $\mathcal{X} = (X, \mathcal{B}_X, \mu, T)$ not ergodic and $\mathcal{Y} = (I, m, id)$; they had a common 2-point factor :

$$\begin{aligned} \eta(z_1) &= \mu(A) = m(J) \\ \eta(z_2) &= m(A^c) = m(J^c) \end{aligned} \tag{7.38}$$





For $V \in \mathcal{B}$, we want to know what $E(1_V | r_X^{-1}\mathcal{B}_Z)(\cdot)$ is. We know that this is a function on X which is $r_X^{-1}\mathcal{B}_Z$-measurable and is constant on $r_X^{-1}\{z_1\}$ and $r_X^{-1}\{z_2\}$. We denote these constants by C_1 and C_2, respectively. We also know that $TA = A$, so $r_X^{-1}\{z_1\} = A$ and $r_X^{-1}\{z_2\} = A^c$. Also, $E(1_V | r_X^{-1}\mathcal{B}_Z)$ should have the same integral over each of A and A^c as 1_V:

$$C_1\mu(A) = \int_A 1_V \, d\mu = \mu(A \cap V) \tag{7.39}$$

and

$$C_2\mu(A^c) = \int_{A^c} 1_V \, d\mu = \mu(A^c \cap V). \tag{7.40}$$

So

$$E(1_V | r_X^{-1}\mathcal{B}_Z) = C_1 = \frac{\mu(A \cap V)}{\mu(A)} = \mu(V|A) \text{ on } r_X^{-1}\{z_1\} \tag{7.41}$$

and

$$E(1_V | r_X^{-1}\mathcal{B}_Z) = C_2 = \mu(V|A^c) \text{ on } r_X^{-1}\{z_2\}. \tag{7.42}$$

For $V \in \mathcal{B}_X$, $W \in \mathcal{B}_Y$, according to the above definition of the relatively independent joining λ,

$$\begin{aligned} \lambda(V \times W) &= \int_{\{z_1\}} E(1_V | r_X^{-1}\mathcal{B}_Z))(z) E(1_W | r_Y^{-1}\mathcal{B}_Z))(z) d\eta(z) \\ &\quad + \int_{\{z_2\}} E(1_V | r_X^{-1}\mathcal{B}_Z))(z) E(1_W | r_Y^{-1}\mathcal{B}_Z))(z) d\eta(z) \\ &= \mu(V|A) m(W|J) \eta\{z_1\} + \mu(V|A^c) m(W|J^c) \eta\{z_2\}. \end{aligned} \tag{7.43}$$

Since $\mu(A) = \eta\{z_1\}$ and $\mu(A^c) = \eta\{z_2\}$, we get

$$\lambda(V \times W) = \frac{\mu(V \cap A) m(W \cap J)}{\mu(A)} + \frac{\mu(V \cap A^c) m(W \cap J^c)}{\mu(A^c)}. \tag{7.44}$$

This was how we defined the joining λ in the proof of Theorem 7.24.

8. Cutting and stacking, adic systems (notes by M. Bonzek, J. Clemons, and I. Rao)

8.1. **Cutting and stacking.** Vershik's idea of representing systems defined by cutting and stacking is a fine way to construct examples of measure-preserving systems.

We will begin with the example known as the binary odometer or as the von Neumann-Kakutani adding machine.

Example 8.1. Let us consider the closed interval $X = [0, 1]$ with Lebesgue measure m on the Lebesgue σ-algebra $\mathfrak{M}$. Our aim is to construct a Lebesgue-measure-preserving system. We construct $T: [0, 1) \to [0, 1)$ as follows:

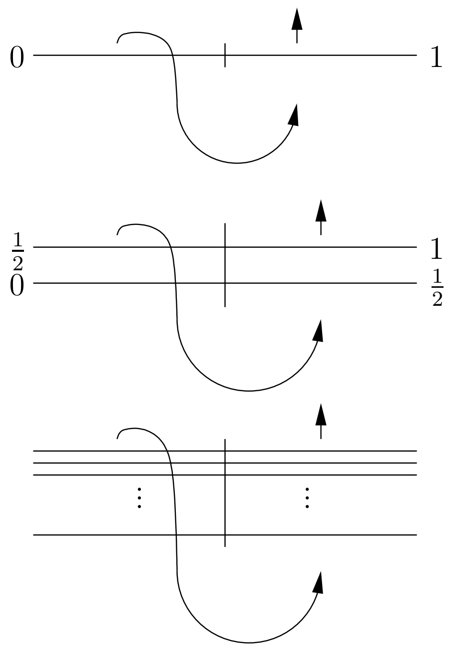


Figure 1. Odometer

$$\text{For } n = 0, 1, 2, ..., \text{ let } I_n = [1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}}). \tag{8.1}$$

Notice that $\bigcup_{n=0}^{\infty} I_n = [0, 1)$.

$$\text{We define } T(x) = x - (1 - \frac{1}{2^n} - \frac{1}{2^{n+1}}), \text{ for all } x \in I_n. \tag{8.2}$$

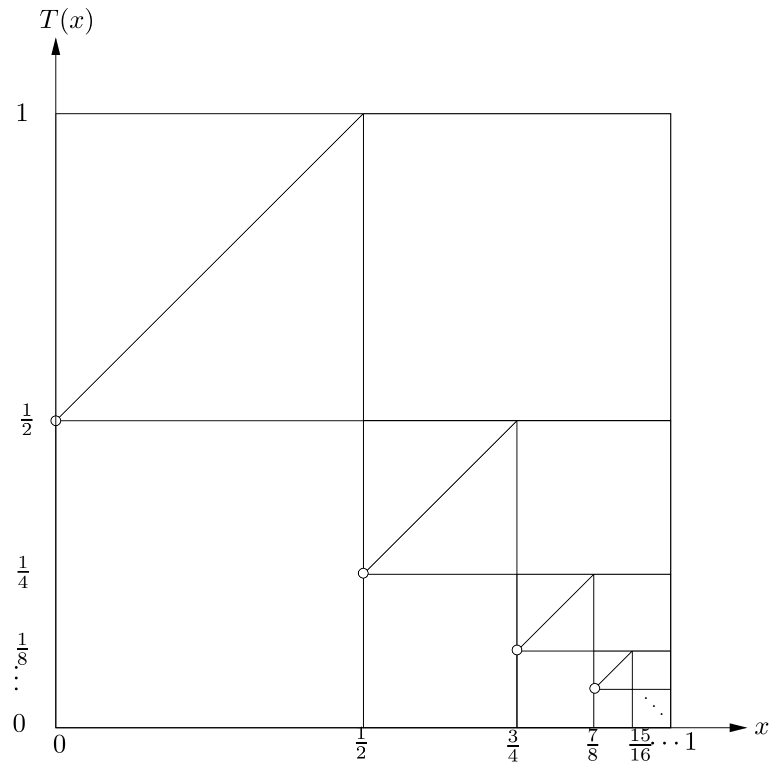


FIGURE 2. Graph of the odometer

This is an m-preserving map, since on each I_n, T is an (affine) linear map of slope 1. It can be proved that T is ergodic with respect to m.

In order to better understand T, consider the map $\pi : \{0,1\}^{\mathbb{N}} \to [0,1]$ which sends every element of $\{0,1\}^{\mathbb{N}}$ to the corresponding dyadic representation of an element of [0,1]; i.e. if $\omega = (\omega_k)_{k\in\mathbb{N}}$, where each ω_k is either 0 or 1, then

$$\pi(\omega) = \sum_{k=1}^{\infty} \frac{\omega_k}{2^k}. \tag{8.3}$$

This map is one-to-one everywhere except at every dyadic rational in [0,1], where it is two-to-one. Thus π is bijective on a set of full measure.

Let $\Omega^+ = \{0,1\}^{\mathbb{N}}$. Define $S : \Omega^+ \to \Omega^+$ by $S(\omega) = \omega + \mathbf{1}$, where $\mathbf{1} = (1,0,0,0,0,........)$ and the addition is performed modulo two in each coordinate, but by carrying to the right. One can think of this as adic addition in $\prod_{i=0}^{\infty} \mathbb{Z}_2$, which is a compact abelian group. We obtain the following commutative diagram:

$$\begin{array}{ccc} \Omega^+ & \xrightarrow{S} & \Omega^+ \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ [0,1] & \xrightarrow{T} & [0,1] \end{array} \tag{8.4}$$

Observe that we can get a simple rule for calculating $S(\omega)$: Look for the first 0 in the dyadic representation of ω and change it to 1, then change all the preceding places to 0, keeping the remaining places as they are.

We define a metric d on Ω^+ by $d(\omega, \eta) = 1/2^k$, where k is the first place where ω and η disagree. Let $\mathfrak{T}$ be the topology induced by d on Ω^+. This topology is that of a Cantor set. Let $\mathcal{B}$ be the Borel σ-algebra on [0,1) and $\mathcal{B}(\mathfrak{T})$ the Borel σ-algebra on Ω^+.

Lemma 8.2. *$([0,1), \mathcal{B}, T, m)$ and $(\Omega^+, \mathcal{B}(\mathfrak{T}), S, \pi^* m)$ are isomorphic as measure-preserving dynamical systems.*

Proof. We will first prove that π is measurable. It suffices to show that π is continuous. We know, the sets

$$B_n^i = \left(\frac{i}{2^n}, \frac{i+1}{2^n}\right) \text{ where } n \in \mathbb{N}, 0 \leq i \leq 2^n \tag{8.5}$$

form a basis for the topology on [0,1). We also observe that

$$\pi^{-1}(B_n^i) = \text{the cylinder set corresponding to the binary expansion of } \frac{i}{2^n}. \tag{8.6}$$

Thus we see that π is continuous and hence measurable. By the definition of $\pi^* m$, we easily see that $\pi^* m(\pi^{-1}(B)) = m(B)$ for all $B \in \mathcal{B}$.

Since the set of dyadic rationals is of measure 0, and π is one-to-one everywhere except on rationals and onto, we can conclude that π is an isomorphism of measure-preserving dynamical systems. □

Then $(\Omega^+, \mathcal{B}(\mathfrak{T}), S, \pi * m)$ is an entropy-0 dynamical system with a unique invariant Borel probability measure and discrete spectrum. Since π is an isomorphism, so is $([0,1), \mathcal{B}, T, m)$.

We observe that $\Omega^+ = \overline{\{k \cdot \mathbf{1} : k \in \mathbb{Z}, k \geq 0\}}$. If we are given a cylinder set $[x_0...x_n]$, then $(x_0 + 2x_1 + \cdots + 2^n x_n) \cdot \mathbf{1} \in [x_0...x_n]$.

Theorem 8.3. *If G is a compact, totally disconnected, monothetic topological group, i.e. there is $g \in G$, such that $G = \overline{\{g^k : k \in \mathbb{Z}\}}$, then $T : G \to G$ given by $T(x) = gx$ is minimal.*

Proof. We need to show that for any given $x \in G$ the set $\{T^k x : k \in \mathbb{Z}\}$ is dense in G.

Let $x \in G$ be arbitrary, but fixed. Let $y \in G$ be arbitrary and U be any neighborhood of y. Since G is a topological group, Ux^{-1} is a neighborhood of yx^{-1}. By the hypothesis, there exists $k \in \mathbb{Z}$ such that $g^k \in Ux^{-1}$. Then $g^k x \in U$, i.e. $T^k x \in U$. This proves that $\{T^k x : k \in \mathbb{Z}\}$ is dense in G. Since $x \in G$ was arbitrary, we can conclude that $T : G \to G$ is minimal. □

Thus $([0, 1), \mathcal{B}, T, m)$ is minimal.

8.2. **The adic representation of the binary odometer.** Sequences $\omega \in \Omega^+$ can be thought of as labeling infinite paths in an infinite directed graded graph, with the edge on the left labeled 0 and the one on the right labeled 1. Edges are ordered according to the order of their labels.

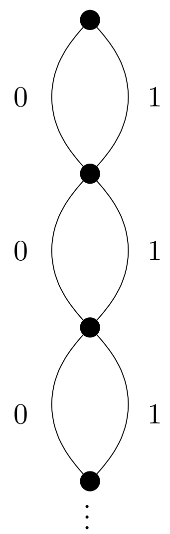


Figure 3. The odometer as an adic

Here is an algoritm for computing $S(\omega)$: Look for the first non-maximal edge (labeled 0). Increase it to the next largest one, minimizing all the preceding ones. For the maximal path, define $S(1, 1, 1, ...) = (0, 0, 0, ...)$.

8.3. **Adic (Bratelli-Vershik) systems.** We will describe a transformation on the space of infinite paths on a directed, graded, infinite graph. The set of vertices, $\mathcal{V}$, is the disjoint union of countably many finite, non-empty sets, $\mathcal{V}_n, n \geq 0$. We assume that $\mathcal{V}_0$ consists of a single vertex, v_0. The set of edges, $\mathcal{E}$, is the disjoint union of countably many finite, non-empty sets, $\mathcal{E}_n$.

Moreover, assume that for all $n \geq 0$ each vertex in $\mathcal{V}_n$ has at least one downward outgoing edge and for all $n \geq 1$ each vertex in $\mathcal{V}_n$ has at least one incoming edge from the level above.

Specify an order on the set of incoming edges down to each vertex by labeling the minimal edge as 0, the next one as 1, and so on. Often it is possible to draw the diagram so that the edges entering a vertex increase in order from left to right.

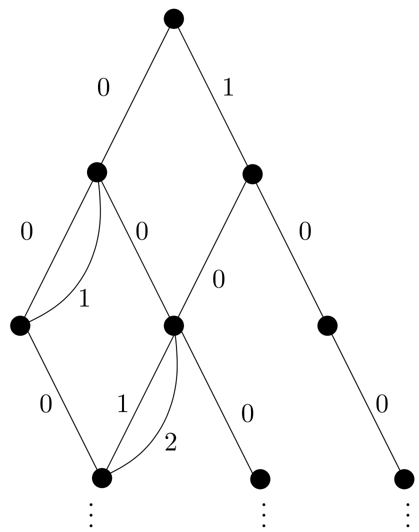


Figure 4. A typical adic

Let $X = \{x_0x_1x_2.... : x_i \in \mathcal{E}_i, \text{ for all } i\}$ denote the set of infinite downward paths from v_0.

Let $\mathcal{E}_{max}$ denote the set of maximal edges and $\mathcal{E}_{min}$ denote the set of minimal edges.

Then let X_{max} denote the set of paths in X all of whose edges are in $\mathcal{E}_{max}$, and let X_{min} denote the set of paths in X all of whose edges are in $\mathcal{E}_{min}$.

We give X the following metric

$$d(x, y) = \frac{1}{2^n}, \tag{8.7}$$

where $n = \min\{j | x_j \neq y_j\}$. By Tychonoff's theorem, X is a compact metric space.

Proposition 8.4. *X is totally disconnected.*

Proof. It suffices to show that every cylinder set is clopen. (The cylinder sets form a base for the topology on X from the definition of the metric.) Every cylinder set is open. We will show that the complement is also open. Let $y \in X \backslash [x_0x_1x_2...x_n]$. Thus, $y = y_0y_1y_2...$ with $y_i \neq x_i$ for some $i = 0, ..., n$. So $y \in [y_0y_1y_2...y_n]$, which is disjoint from $[x_0x_1x_2...x_n]$. We conclude that $X \backslash [x_0x_1x_2...x_n]$ is open. □

We say two elements $x, y \in X$ are *comparable* if there is an N such that $x_n = y_n$ for all n $\geq N$. Let N_0 be the smallest such. Then we say $x < y$ if and only if $x_{N_0-1} < y_{N_0-1}$ in our given ordering of $\mathcal{E}_{N_0-1}$, i.e. compare x_n and y_n the last time they differ.

Define $T(x)$ = smallest $y > x$, if there is one.

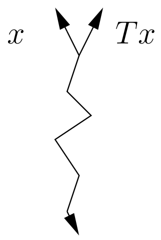


FIGURE 5. Action of the adic

We have the following algorithm to compute $T(x)$:

Find the first n for which x_n is not maximal. Increase that edge to the next largest one, say x_n'. Do not change any edges beyond (below) x_n. Continue the path upward from x_n' by following minimal edges up to v_0. Thus $T(x)$ is the successor of x in the same sense as in the binary odometer.

Similarly, T^{-1} is defined everywhere on X except on X_{min}.

Starting at the minimal path from v_0 to v, which corresponds to a cylinder set C_1, T maps to $T(C_1), ..., T^n(C_1)$ with $T^n(C_1)$ corresponding to the maximal path from v_0 to v. This implies that T is Borel measurable with respect to the smallest σ-algebra containing the cylinder sets).

There is a natural procedure for converting between adic and cutting and stacking representations of a system. The vertices at level n correspond to stacks; the paths correspond to levels of a stack. There is an edge from $v \in \mathcal{V}_n$ to $v' \in \mathcal{V}_{n+1}$ if and only if some part of the stack corresponding to v appears in the stack corresponding to v'. The ordering of the edges corresponds to the order of levels going up the v' stack.

Proposition 8.5. *The map $T : X \backslash X_{max} \to X \backslash X_{min}$ is a homeomorphism.*

Proof. It is clear that T is one-to-one since T^{-1} is defined as

$$T^{-1}y = \max\{x : y > x\}.$$

To show that T is a homeomorphism it suffices to show that

$$d(x, y) \leq d(Tx, Ty) \tag{8.8}$$

and

$$d(u, v) \leq d(T^{-1}u, T^{-1}v) \tag{8.9}$$

for all $x, y \in X\backslash X_{max}$ and all $u, v \in X\backslash X_{min}$. To prove (8.8) and (8.9) notice that if x and y first disagree at the n'th place, then Tx and Ty will still disagree at the n'th place or sooner. Thus $d(x, y) \leq d(Tx, Ty)$. The same argument will hold for T^{-1}. Thus T is a homeomorphism. $\square$

8.4. **Invariant measures.** Suppose that non-negative weights $w(x_i)$ are assigned to edges x_i in such a way that the weights of the edges leaving each vertex downward sum to 1 and weights on edges connecting the same two vertices are equal:

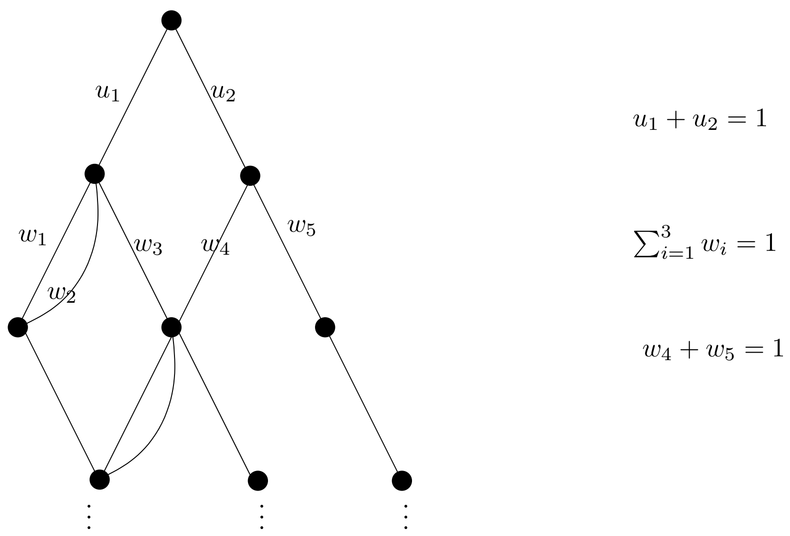


Figure 6. Weights on edges of an adic

We would like to use these weights to define a T-invariant Borel measure on X. Note that the anchored cylinder sets which form a basis for the topology on X form a semialgebra, $\mathcal{S}$ (see Section 3.2). If we can define a measure on $\mathcal{S}$ that is finitely additive and subadditive on $\mathcal{S}$ then the Carathéodory construction (Proposition 3.2, Theorem 3.3, and Theorem 3.4) will give us a measure.

We can define the following "Markovian" measure on anchored cylinder sets:

$$\mu[x_0 x_1 ... x_n] = w(x_0)w(x_1) \cdots w(x_n), \tag{8.10}$$

where $w(x_i)$ is the weight of the edge x_i.

Proposition 8.6. *Suppose $S_1, S_2, ... \in \mathcal{S}$ and $\bigcup_{i=1}^{\infty} S_i \in \mathcal{S}$. Then*

$$\mu(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mu(S_i). \tag{8.11}$$

Proof. The proposition follows directly from the following observation. If

$$\bigcup_{i=1}^{\infty} [x_{i0} x_{i1} ... x_{ik_i}] = [x_0 x_1 ... x_n], \tag{8.12}$$

then by the compactness of $[x_0x_1...x_n]$ the union can be expressed as a finite union

$$\bigcup_{j=1}^{m}[x_{i_j0}x_{i_j1}...x_{i_jk_{i_j}}] = [x_0x_1...x_n], \tag{8.13}$$

and by finite additivity we are done. □

We now determine when such a measure is T-invariant, by which we mean that $\mu(C) = \mu(TC)$ for each anchored cylinder set C on which T is defined. (Since $T : X\backslash X_{max} \to X\backslash X_{min}$, we only have to consider μ on $X\backslash X_{max}$ in the domain and on $X\backslash X_{min}$ in the range. Often X_{max} and X_{min} are countable, so if μ is nonatomic, we will have $\mu(X_{max}) = 0 = \mu(X_{min})$ and this problem will not arise.

Proposition 8.7. *A Borel probability measure μ on X is adic-invariant if and only if all anchored cylinder sets from v_0 to each fixed vertex v have the same measure.*

Proof. Suppose that all anchored cylinder sets from v_0 to each fixed vertex v have the same measure. Consider a cylinder set $[x_0...x_n]$ on which T is defined. The points in the cylinder set consist of paths from v_0 to some v_{n+1} in $\mathcal{V}_{n+1}$. Notice that $T[x_0...x_n]$ also consists of paths from v_0 to the same v_{n+1}. Thus $\mu[x_0...x_n] = \mu T[x_0...x_n]$.

Suppose that a Borel probability measure μ is adic-invariant. If C is a cylinder set of paths determined by a non-maximal path from v_0 to v, then TC is another such cylinder set of paths from v_0 to v. It is necessary that C and TC have the same measure. □

Proposition 8.8. *If μ is an adic-invariant Borel measure on X, then it is Markovian in the above sense.*

Proof. We delete from the path space X the union of all cylinders of measure 0. What remains is still an adic system given by a Bratteli diagram. Since all cylinders determined by paths from the root to a fixed vertex have the same measure, on each edge x_n we can then put the weight

$$w(x_n) = \frac{\mu[x_0x_1...x_{n-1}x_n]}{\mu[x_0x_1...x_n]} \tag{8.14}$$

and then we have

$$\mu[x_0x_1...x_n] = \frac{\mu[x_0x_1...x_n]}{\mu[x_0x_1...x_{n-1}]}\frac{\mu[x_0x_1...x_{n-1}]}{\mu[x_0x_1...x_{n-2}]}\cdots = \prod_{i=1}^{n} w(x_i). \tag{8.15}$$

□

8.5. **More examples.**

Example 8.9. The (q_n) odometers

Let (q_n) be a sequence of positive integers. A (q_n) odometer is the following system.

$$T : \prod_{i=0}^{\infty} \mathbb{Z}_{q_i} \to \prod_{i=0}^{\infty} \mathbb{Z}_{q_i} \tag{8.16}$$

is defined by $Tx = x + (1, 0, 0, ...)$ with carry to the right. The associated Bratteli diagram is:

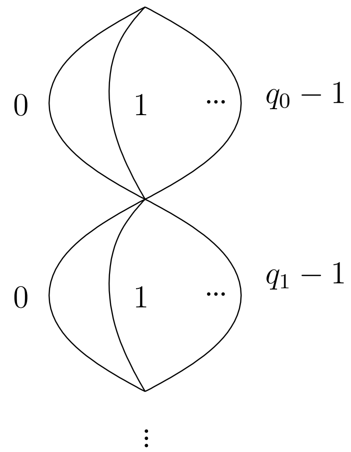


FIGURE 7. The q-odometer

For example, if $X = \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4 \times \cdots$ and $x = 102...$, then $Tx = 012...$, $T^2x = 112...$, $T^3x = 022...$, $T^4x = 122...$, and $T^5x = 003....$

Example 8.10. The Chacon example

The Chacon example is the 3-odometer with spacers. This system is weakly mixing but not strongly mixing. It is also a prime system (no proper factors) and has minimal self-joinings. the following picture shows the first few steps of the cutting and stacking construction.

The adic representation is seen in Figure 8.10.

Example 8.11. The Fibonacci substitution subshift and rotation by $1/\varphi$

The Fibonacci substitution subshift, the golden-mean adic system, and translation by $1/\varphi$ mod 1 are all representations of the same system.

First consider the substitution $\zeta : 0 \to 01, 1 \to 0$.

Beginning with 0,	0
0 becomes 01	01
0 becomes 01 and 1 becomes 0	010

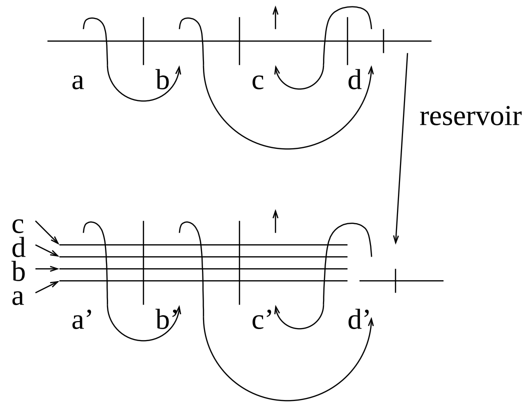


FIGURE 8. The Chacon system by cuting and stacking

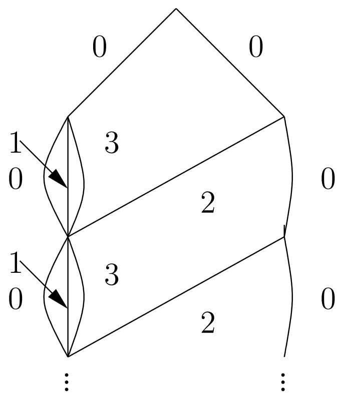


FIGURE 9. The Chacon system as an adic

01001

01001010

and so on.

Notice that the number of digits in each row is a Fibonacci number. This is no coincidence; for each digit in row n, there is a 0 in row $n+1$. For each 0 in row $n+1$, there is an additional digit in row $n+2$. So the number of digits in row $n+2$ is the sum of the numbers of digits in rows $n+1$ and n.

Let $\omega^+ \in \{0,1\}^{\mathbb{Z}_+}$ be the fixed point of the substitution. Then let $\omega = \dots 000.\omega^+ \in \{0,1\}^{\mathbb{Z}}$. Next pick $z \in \{\sigma^n \omega : n \geq 0\}^- \in \{0,1\}^{\mathbb{Z}}$. (Here σ signifies the shift.) Alternatively, we may let z be the fixed point of ζ^2 applied to 1.0, so that $z = \dots 01001.01001010 \dots$.

Finally we define $X = \overline{\mathcal{O}(z)}$, the orbit closure of z. Then we have a system (X, σ). This system is minimal (every orbit is dense) and uniquely ergodic.

The adic representation is as in Figure 8.11.

Now we define a new system. The space is $X = [0, 1)$ and the map is $x \to x + \alpha \bmod 1$, where α is the positive root of $\alpha + \alpha^2 = 1$. So $1/\alpha$ is the golden mean, and this system is translation by α mod 1 on the unit interval. The cutting and stacking system representation looks like the following:

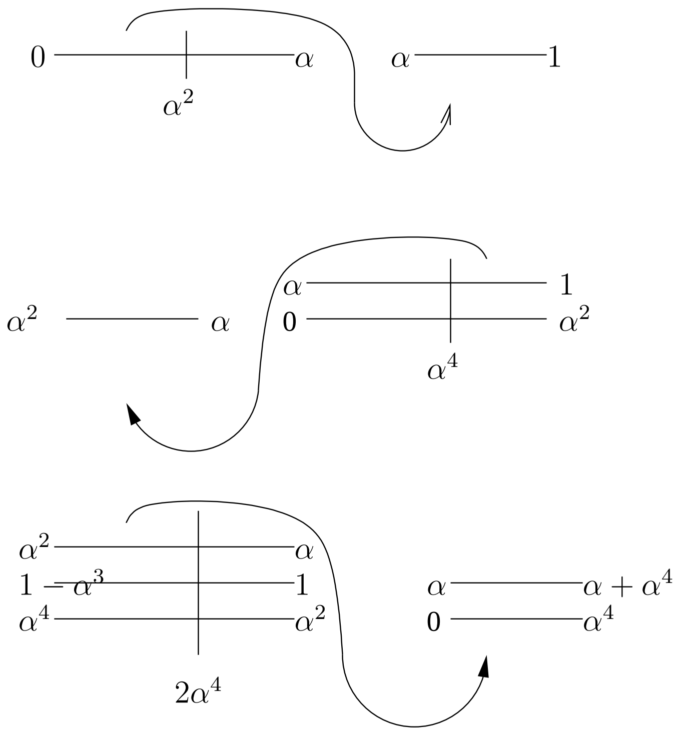


Figure 10. Translation by α as cutting and stacking

We can obtain an adic representation of this system. Before we apply the translation, $x \in [0, 1)$ is in one of two pieces of the interval, one (A) with length α and the other (B) of length α^2. On the first iteration, part A can be split into two pieces: one piece, of length α^2, is sent to part B. The other piece, with length α^3, is sent back into A. Rescaling to 1, this gives us a split of part A into pieces of lengths α and α^2. Part B, on the other hand, is sent into part A. So the adic representation, with weights, is as in Figure 8.11.

This relates to the Fibonacci substitution and rotation by the golden mean as follows. Label each left vertex by 0 and each right vertex by 1. Looking downward, each 0 can be followed by either 0 or 1, but 1 can only be followed by 0. Thus the infinite path space is in correspondence with the golden mean shift of finite type.

The adic transformation can be given a different coding by keeping track of the first vertex of $T^n x$ for all $n \in \mathbb{Z}$. The codings of the minimal paths to each vertex can be obtained by concatenating blocks, as shown in Figure 12.

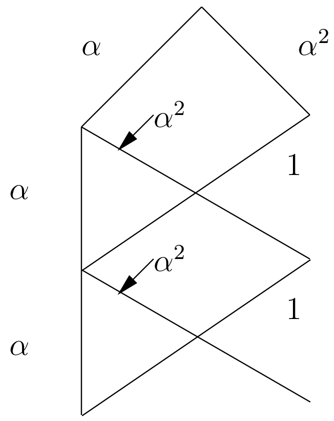


FIGURE 11. Translation by α as an adic system

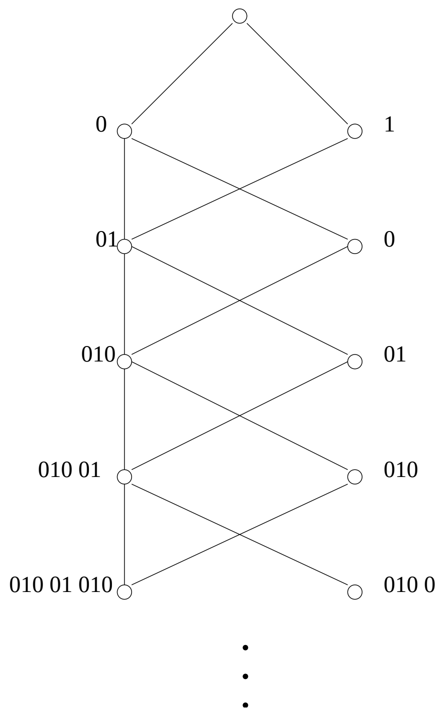


FIGURE 12. Coding the Fibonacci adic by the first vertex

The set of all these codings coincides with the minimal Sturmian subshift consisting of the closure of the set of all the sequences $(1 - \mathbf{1}_{[1-\alpha,1)}(x + k\alpha)), k \in \mathbb{Z}$, $x \in [0, 1]$.

Example 8.12. The Pascal adic system

First define this system by the adic diagram with weights $r, 1 - r \in (0, 1)$:

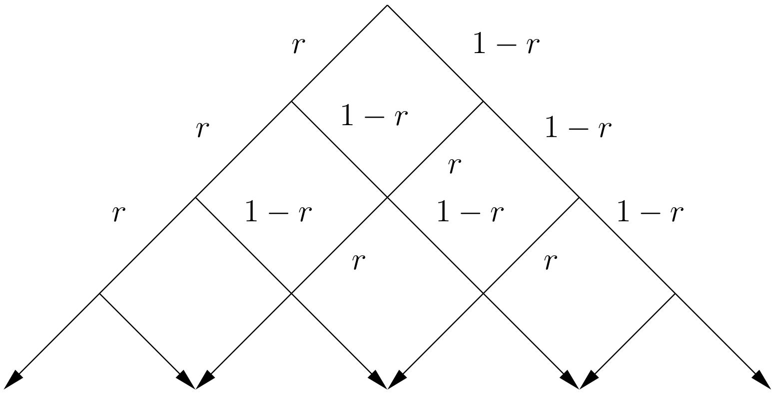


Figure 13. The Pascal adic

Now we give the cutting and stacking representation. For each vertex there is a stack. The height of the stack corresponding to vertex v is the number of paths from v_0 to v, which is also known for historical reasons as the *dimension* of v. In this case, $\dim(n, k) = C(n, k)$, the binomial coefficient n choose k.

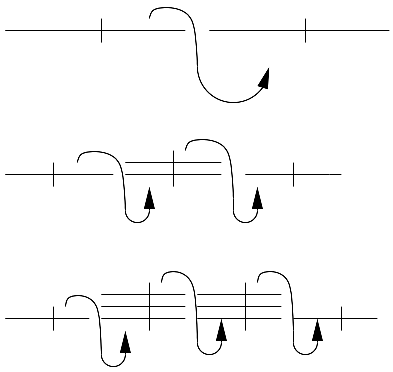


Figure 14. The Pascal system by cutting and stacking

Example 8.13. The Euler adic system

Figure 15 shows the related system of Euler; the numbers along the edges indicate multiple edges.

In this case, $\dim(n, k) = A(n, k)$, the Eulerian numbers. These numbers are important in combinatorics.

Example 8.14. The staircase

This system is defined by cutting and stacking: just put j spacers above the j'th piece, then restack.

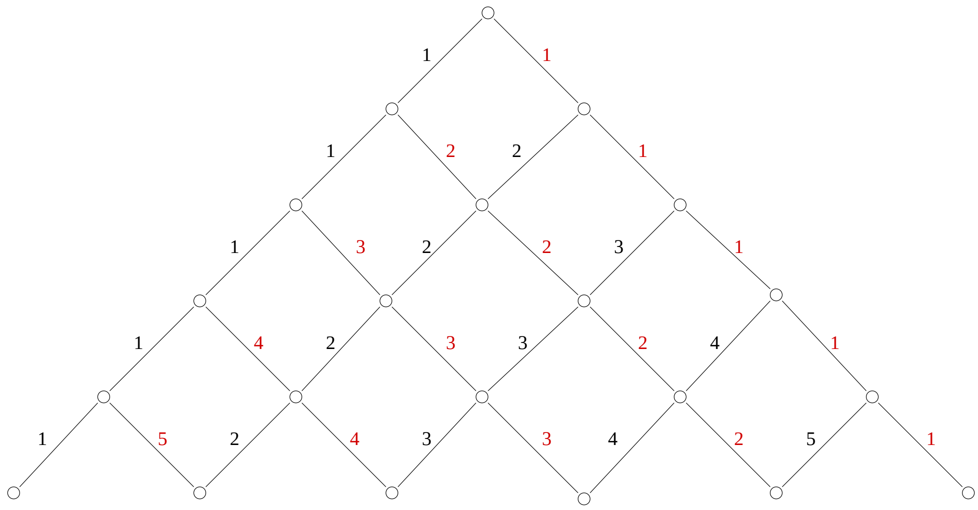


Figure 15. The Euler graph (down to level 5)

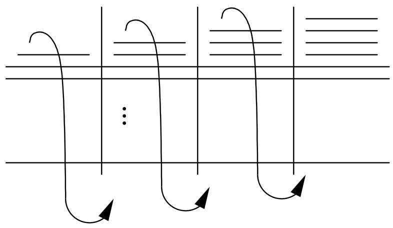


Figure 16. The staircase system

This is a Lebesgue-measure-preserving system. M. Smorodinsky conjectured, and T. Adams proved, that this system is strongly mixing. This was the first explicit example of a rank-1 strongly mixing system.

Definition 8.15. A measure-preserving system has *rank 1* if the levels of Rokhlin towers generate the full σ-algebra. Equivalently, it is constructible by cutting and stacking with one stack and spacers. (A rank k system is constructible with k stacks).

Figure 17 shows the adic representation of a general rank-1 system. Edge labels indicate the numbers of edges connecting pairs of vertices. At the first level below the root, we have one column consisting of a single level and a reservoir (for spacers), also consisting of a single level. Then the column is cut into c_1 pieces and the reservoir into $s_1 + 1$ pieces, represented by the edges from level 1 to level 2. These edges are ordered according to where the spacers are inserted between the pieces of the column. The cutting and stacking continues as described by the rest of the diagram, with the numbers of edges c_k and $s_k + 1$ and specified orderings.

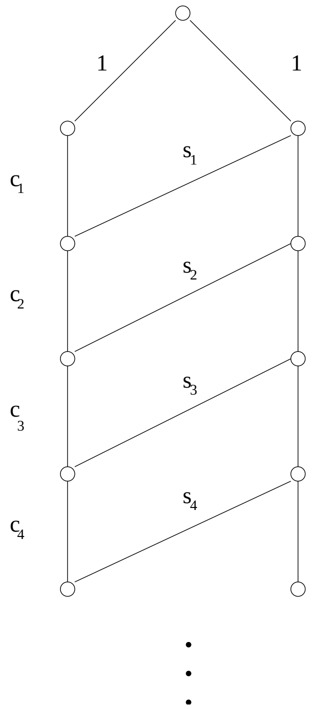


FIGURE 17. A general rank-1 system as an adic

Remark 8.16. The Morse system has rank 2 (this was proved by del Junco). Rank is involved in some interesting unsolved questions, e.g. does the Pascal system (see Example 8.14) have finite rank?

Remark 8.17. A system of finite rank has entropy 0.

8.6. Vershik's representation theorem.

Theorem 8.18 (Vershik's version of the Jewett-Krieger Theorem [53, 54])**.** *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system on a nonatomic Lebesgue space. Then it is isomorphic to an adic system with an invariant probability measure (given by weights on the edges). Moreover, we can arrange that*

(1) *The adic system has a unique maximal path and a unique minimal path. (This property is called* essentially simple.*)*
(2) *There are $n_1 < n_2 < \cdots$ so that for each i there is a path from every vertex at level n_i to every vertex at level n_{i+1}. This property is called* simple *or* primitive.
(3) *The adic system is uniquely ergodic.*

Remark 8.19. (1) implies that by defining $T(x_{max}) = x_{min}$ we can obtain a homeomorphism $X \to X$ of the path space to itself.

Remark 8.20. *Simple* is equivalent, possibly after some modification of the diagram (telescoping), to *minimal* (which, we recall, means that every orbit is dense).

Essentially simple is equivalent to *essentially minimal*, which means that there exists a unique minimal set, $\overline{\mathcal{O}(x_{max})}$. Equivalently, for every neighborhood U of x_{max}, $\bigcup_{n=-\infty}^{\infty} T^n U = X$.

Proof (of Theorem 8.18). We show first that we can find measurable sets $A \subset X$ with bounded return times. Take any measurable set $E \subset X$ with $\mu(E) > 0$ and form the tower decomposition of X with respect to E (see figure 8.6).

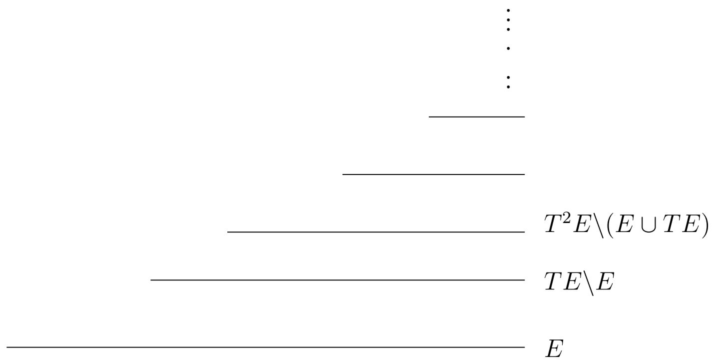


FIGURE 18. A general tower decomposition

Now fix N. Let

$$F = \bigcup_{n=N+1}^{\infty} [T^n E \backslash \left(E \cup TE \cup \cdots \cup T^{n-1}E\right)]. \tag{8.17}$$

Let $A = E \cup F$. Then for all $x \in A$, there is $n \in [0, N]$ such that $T^n x \in A$.

To make the illustrations for the first few steps of the construction less complicated, we show that in fact we can choose A so that the return times to A are always either 2 or 3. Using Rokhlin's Lemma (7.1)start with E such that $\mu(E) > 0$ and $E \cap TE = \emptyset$.

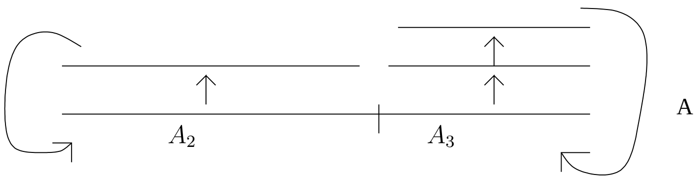


FIGURE 19. A special tower decomposition

This gives us the first level in a cutting and stacking, equivalently adic, representation of $T : X \to X$.

Now look at the system $(A, \mathcal{B} \cap A, \mu_A, T_A)$, where $\mu_A = \mu/\mu(A)$ and T_A is the first-return map.

Find $B = B_2 \cup B_3 \subset A$ with return times (to B) of 2 on B_2 and 3 on B_3, with respect to T_A.

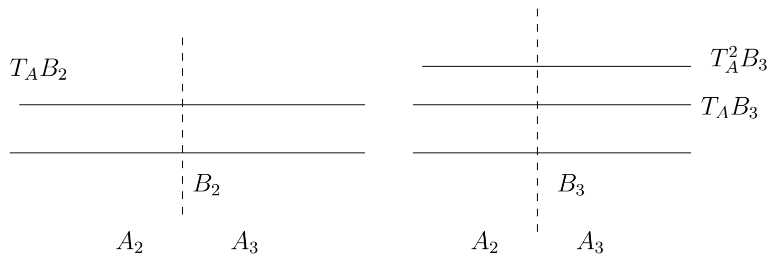


Figure 20. Further cutting and stacking

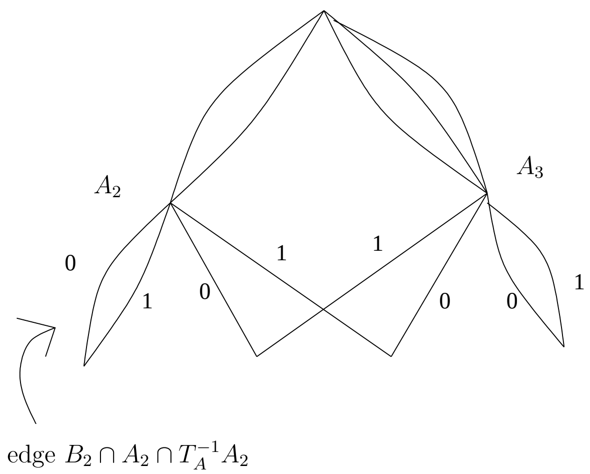


Figure 21. A further stage of the adic construction

Partition A by $\{A_2, A_3\} \vee \{B_2, T_A B_2\} \vee T_A^{-1} \{A_2, A_3\}$.

Partition A also by $\{A_2, A_3\} \vee \{B_3, T_B B_3, T_B^2 B_3\} \vee T_A^{-1} \{A_2, A_3\} \vee T_A^{-2} \{A_2, A_3\}$.

These partitions tell us how further to cut and stack the diagram in Figure 8.6, equivalently, how to add another row of vertices, with connecting edges, in Figure 8.6. The latter is the union of the following two diagrams:

Let $(Y, \mathcal{C}, \nu, S)$ be the system determined by the eventual infinite Bratteli diagram constructed in this way, with Y the space of infinite paths, $\mathcal{C}$ the σ-algebra generated by the cylinder sets, ν the measure on Y corresponding to μ on X, and S the adic transformation. Recall that each vertex v corresponds to a column in the cutting and stacking process, and each path from the root to v corresponds to a level in that column.

To get the isomorphism, we have to make sure that the subsets of X that correspond to cylinder sets in Y, which are the levels of the towers in the cutting

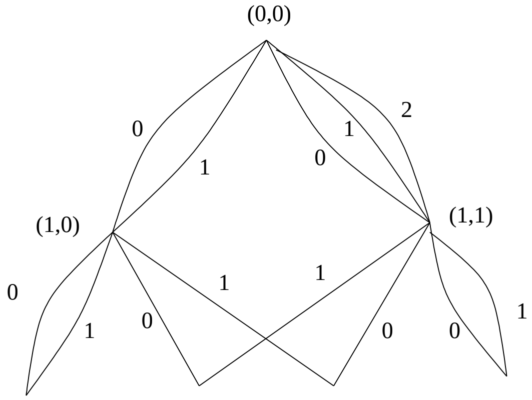


FIGURE 22. The left half of the next stage

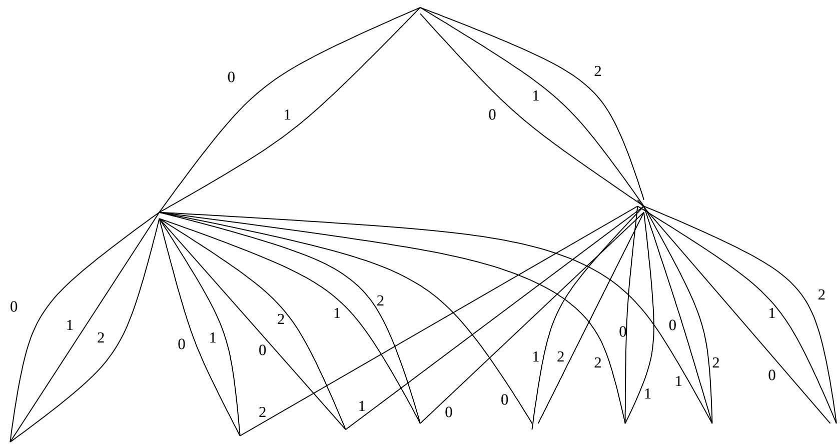


FIGURE 23. The right half of the next stage

and stacking representation, generate the σ-algebra $\mathcal{B}$ of X. Take a sequence of measurable sets $E_1, E_2, \dots$ which generate the σ-algebra $\mathcal{B}$. At stage n, refine the partitions involved by $\mathcal{B}(E_1, \dots, E_n)$. This produces a sequence of more complicated diagrams, the limit of which has this desired property.

The further desirable properties (1), (2), (3) of the Bratteli-Vershik representation follow from combining the Jewett-Krieger theorem [26, 31, 36] on the representation of ergodic systems by uniquely ergodic homeomorphisms of the Cantor set with the Herman-Putnam-Skau theorem [27] on the representation of Cantor minimal systems by adics.

□

9. Some mathematical background

9.1. **Lebesgue spaces.**

Definition 9.1. Two measure spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ are *isomorphic* (sometimes also called *isomorphic mod 0*) if there are subsets $X_0 \subset X$ and $Y_0 \subset Y$ such that $\mu(X_0) = 0 = \nu(Y_0)$ and a one-to-one onto map $\phi : X \setminus X_0 \to Y \setminus Y_0$ such that ϕ and ϕ^{-1} are measurable and $\mu(\phi^{-1}C) = \nu(C)$ for all measurable $C \subset Y \setminus Y_0$.

Definition 9.2. A *Lebesgue space* is a finite measure space that is isomorphic to a measure space consisting of a (possibly empty) finite subinterval of $\mathbb{R}$ with the σ-algebra of Lebesgue measurable sets and Lebesgue measure, possibly together with countably many atoms (point masses).

The *measure algebra* of a measure space $(X, \mathcal{B}, \mu)$ consists of the pair $(\hat{\mathcal{B}}, \hat{\mu})$, with $\hat{\mathcal{B}}$ the Boolean σ-algebra (see Section 1.2, 3) of $\mathcal{B}$ modulo the σ-ideal of sets of measure 0, together with the operations induced by set operations in $\mathcal{B}$, and $\hat{\mu}$ is induced on $\hat{\mathcal{B}}$ by μ on $\mathcal{B}$. Every measure algebra $(\hat{\mathcal{B}}, \hat{\mu})$ is a metric space with the metric $d(A, B) = \hat{\mu}(A \triangle B)$ for all $A, B \in \hat{\mathcal{B}}$. It is *nonatomic* if whenever $A, B \in \hat{\mathcal{B}}$ and $A < B$ (which means $A \wedge B = A$), either $A = 0$ or $A = B$. A *homomorphism of measure algebras* $\psi : (\hat{\mathcal{C}}, \hat{\nu}) \to (\hat{\mathcal{B}}, \hat{\mu})$ is a Boolean σ-algebra homomorphism such that $\hat{\mu}(\hat{C}) = \hat{\nu}(C)$ for all $\hat{C} \in \hat{\mathcal{C}}$. The inverse of any factor map $\phi : X \to Y$ from a measure space $(X, \mathcal{B}, \mu)$ to a measure space $(Y, \mathcal{C}, \nu)$ induces a homomorphism of measure algebras $(\hat{\mathcal{C}}, \hat{\nu}) \to (\hat{\mathcal{B}}, \hat{\mu})$. We say that a measure algebra is *normalized* if the measure of the maximal element is 1: $\hat{\mu}(0') = 1$.

We work within the class of Lebesgue spaces because (1) they are the ones commonly encountered in the wide range of naturally arising examples; (2) they allow us to assume if we wish that we are dealing with a familiar space such as $[0, 1]$ or $\{0, 1\}^{\mathbb{N}}$; and (3) they have the following useful properties.

- (Carathéodory [7]) Every normalized and nonatomic measure algebra whose associated metric space is separable (has a countable dense set) is measure-algebra isomorphic with the measure algebra of the unit interval with Lebesgue measure.
- (von Neumann [55])Every complete separable metric space with a Borel probability measure on the completion of the Borel sets is a Lebesgue space.
- (von Neumann [55]) Every homomorphism $\psi : (\hat{\mathcal{C}}, \hat{\nu}) \to (\hat{\mathcal{B}}, \hat{\mu})$ of the measure algebras of two Lebesgue spaces $(Y, \mathcal{C}, \nu)$ and $(X, \mathcal{B}, \mu)$ comes from a factor map: there are a set $X_0 \subset X$ with $\mu(X_0) = 0$ and a measurable map $\phi : X \setminus X_0 \to Y$ such that ψ coincides with the map induced by ϕ^{-1} from $\hat{\mathcal{C}}$ to $\hat{\mathcal{B}}$.

9.1.1. *Rokhlin theory.* V. A. Rokhlin [43] provided an axiomatic, intrinsic characterization of Lebesgue spaces. The key ideas are the concept of a basis and the correspondence of factors with complete sub-σ-algebras and (not necessarily finite or countable) measurable partitions of a special kind.

Definition 9.3. A *basis* for a complete measure space $(X, \mathcal{B}, \mu)$ is a countable family $\mathcal{C} = \{C_1, C_2, \dots\}$ of measurable sets which

generates $\mathcal{B}$: For each $B \in \mathcal{B}$ there is $C \in \mathcal{B}(\mathcal{C})$ (the smallest σ-algebra of subsets of X that contains $\mathcal{C}$) such that $B \subset C$ and $\mu(C \setminus B) = 0$; and

separates the points of X: For each $x, y \in X$ with $x \neq y$, there is $C_i \in \mathcal{C}$ such that either $x \in C_i, y \notin C_i$ or else $y \in C_i, x \notin C_i$.

Coarse sub-σ-algebras of $\mathcal{B}$ may not separate points of X and thus may lead to equivalence relations, partitions, and factor maps. Partitions of the following kind deserve careful attention.

Definition 9.4. Let $(X, \mathcal{B}, \mu)$ be a complete measure space and ξ a partition of X, meaning that up to a set of measure 0, X is the union of the elements of ξ, which are pairwise disjoint up to sets of measure 0. We call ξ an *R-partition* if there is a countable family $D = \{D_1, D_2, \dots\}$ of ξ-saturated sets (that is, each D_i is a union of elements of ξ) such that

$$\begin{aligned} &\text{for all distinct } E, F \in \xi, \text{ there is } D_i \text{ such that either} \\ &\qquad E \subseteq D_i, F \nsubseteq D_i \quad (\text{so } F \subset D_i^c) \\ &\qquad \text{or } F \subseteq D_i, E \nsubseteq D_i \quad (\text{so } E \subset D_i^c). \end{aligned} \tag{9.1}$$

Any such family D is called a *basis* for ξ.

Note that each element of an R-partition is necessarily measurable: if $C \in \xi$ with basis $\{D_i\}$, then

$$C = \bigcap \{D_i : C \subseteq D_i\}. \tag{9.2}$$

Every countable or finite measurable partition of a complete measure space is an R-partition. The orbit partition of a measure-preserving transformation is often *not* an R-partition. (For example, if the transformation is ergodic, the corresponding factor space will be trivial, consisting of just one cell, rather than corresponding to the partition into orbits as required.)

For any set $B \subset X$, let $B^0 = B$ and $B^1 = B^c = X \setminus B$.

Definition 9.5. A basis $\mathcal{C} = \{C_1, C_2, \dots\}$ for a complete measure space $(X, \mathcal{B}, \mu)$ is called *complete*, and the space is called *complete with respect to the basis*, if for every 0,1-sequence $e \in \{0, 1\}^{\mathbb{N}}$,

$$\bigcap_{i=1}^{\infty} C_i^{e_i} \neq \emptyset. \tag{9.3}$$

$\mathcal{C}$ is called *complete mod 0* (and $(X, \mathcal{B}, \mu)$ is called *complete mod 0 with respect to* $\mathcal{C}$, if there is a complete measure space $(X', \mathcal{B}', \mu')$ with a complete basis $\mathcal{C}'$ such that X is a full-measure subset of X', and $C_i = C_i' \cap X$ for all $i = 1, 2, \dots$.

From the definition of basis, each intersection in (9.3) contains at most one point. The space $\{0,1\}^{\mathbb{N}}$ with Bernoulli $1/2, 1/2$ measure on the completion of the Borel sets has the complete basis $C_i = \{\omega : \omega_i = 0\}$.

Proposition 9.6. *If a measure space is complete mod 0 with respect to one basis, then it is complete mod 0 with respect to every basis.*

Theorem 9.7. *(*[43]*) A measure space is a Lebesgue space (that is, isomorphic mod 0 with the usual Lebesgue measure space of a possibly empty subinterval of* $\mathbb{R}$ *possibly together with countably many atoms) if and only if it has a complete basis.*

In a Lebesgue space $(X, \mathcal{B}, \mu)$ there is a one-to-one onto correspondence between complete sub-σ-algebras of $\mathcal{B}$ (that is, those for which the restriction of the measure yields a complete measure space) and R-partitions of X:

Given an R-partition ξ, let $\mathcal{B}(\xi)$ denote the σ-algebra generated by ξ, which consists of all sets in $\mathcal{B}$ that are ξ*-saturated*—unions of members of ξ–and let $\overline{\mathcal{B}}(\xi)$ denote the completion of $\mathcal{B}(\xi)$ with respect to μ.

Conversely, given a complete sub-σ-algebra $\mathcal{C} \subset \mathcal{B}$, define an equivalence relation on X by $x \sim y$ if for all $A \in \mathcal{C}$, either $x, y \in A$ or else $x, y \in A^c$. The measure algebra $(\hat{\mathcal{C}}, \hat{\mu})$ has a countable dense set $\hat{\mathcal{C}}_0$ (take a countable dense set $\{\hat{B}_i\}$ for $(\hat{\mathcal{B}}, \hat{\mu})$ and, for each i, j for which it is possible, choose $\hat{C}_{ij}$ within distance $1/2^j$ of $\hat{B}_i$). Then representatives $C_i \in \mathcal{C}$ of the $\hat{\mathcal{C}}_{ij}$ will be a basis for the partition ξ corresponding to the equivalence relation $\sim$.

Given any family $\{\mathcal{B}_\lambda\}$, of complete sub-σ-algebras of $\mathcal{B}$, their *join* is the intersection of all the sub-σ-algebras that contain their union:

$$\bigvee_\lambda \mathcal{B}_\lambda = \mathcal{B}(\bigcup_\lambda \mathcal{B}_\lambda), \tag{9.4}$$

and their *infimum* is just their intersection:

$$\bigwedge_\lambda \mathcal{B}_\lambda = \mathcal{B}(\bigcap_\lambda \mathcal{B}_\lambda). \tag{9.5}$$

These σ-algebra operations correspond to the supremum and infimum of the corresponding families of R-partitions. We say that a partition ξ_1 is *finer* than a partition ξ_2, and write $\xi_1 \geq \xi_2$, if every element of ξ_2 is a union of elements of ξ_1. Given any family $\{\xi_\lambda\}$ of R-partitions, there is a coarsest R-partition $\bigvee_\lambda \xi_\lambda$ which refines all of them, and a finest R-partition $\bigwedge_\lambda \xi_\lambda$ which is coarser than all of them. We have

$$\bigvee_\lambda \mathcal{B}(\xi_\lambda) = \mathcal{B}(\bigvee_\lambda \xi_\lambda), \quad \bigwedge_\lambda \mathcal{B}(\xi_\lambda) = \mathcal{B}(\bigwedge_\lambda \xi_\lambda). \tag{9.6}$$

Now we discuss the relationship among factor maps $\phi : X \to Y$ from a Lebesgue space $(X, \mathcal{B}, \mu)$ to a complete measure space $(Y, \mathcal{C}, \nu)$, complete sub-σ-algebras of $\mathcal{B}$, and R-partitions of X. Given such a factor map ϕ, $\mathcal{B}_Y = \phi^{-1}\mathcal{C}$ is a complete sub-σ-algebra of $\mathcal{B}$, and the equivalence relation $x_1 \sim x_2$ if $\phi(x_1) = \phi(x_2)$ determines an R-partition ξ_Y. (A basis for ξ can be formed from a countable dense set in $\hat{\mathcal{B}}_Y$ as above.)

Conversely, given a complete sub-σ-algebra $\mathcal{C} \subset \mathcal{B}$, the identity map $(X, \mathcal{B}, \mu) \to (X, \mathcal{C}, \mu)$ is a factor map. Alternatively, given an R-partition of X, we can form a measure space $(X/\xi, \mathcal{B}(\xi), \mu_\xi)$ and a factor map $\phi_\xi : X \to X/\xi$ as follows. The space X/ξ is just ξ itself; that is, the points of X/ξ are the members (cells, or atoms) of the partition ξ. $\mathcal{B}(\xi)$ consists of the ξ-saturated sets in $\mathcal{B}$ considered as subsets of ξ, and μ_ξ is the restriction of μ to $\mathcal{B}(\xi)$. Completeness of $(X, \mathcal{B}, \mu)$ forces completeness of $(X/\xi, \mathcal{B}(\xi), \mu_\xi)$. The map $\phi_\xi : X \to X/\xi$ is defined by letting $\phi(x) = \xi(x) =$ the element of ξ to which x belongs. Thus for a Lebesgue space $(X, \mathcal{B}, \mu)$, there is a perfect correspondence among images under factor maps, complete sub-σ-algebras of $\mathcal{B}$, and R-partitions of X.

Theorem 9.8. *If $(X, \mathcal{B}, \mu)$ is a Lebesgue space and $(Y, \mathcal{C}, \nu)$ is a complete measure space that is the image of $(X, \mathcal{B}, \mu)$ under a factor map, then $(Y, \mathcal{C}, \nu)$ is also a Lebesgue space.*

Theorem 9.9. *Let $(X, \mathcal{B}, \mu)$ be a Lebesgue space, $(Y, \mathcal{C}, \nu)$ a separable measure space (that is, one with a countable basis as above, equivalently one with a countable dense set in its measure algebra), and $\phi : X \to Y$ a measurable map ($\phi^{-1}\mathcal{C} \subset \mathcal{B}$). Then ϕ is also* forward measurable*: if $A \subset X$ is measurable, then $\phi(A) \subset Y$ is measurable.*

Theorem 9.10. *Let $(X, \mathcal{B}, \mu)$ be a Lebesgue space.*

(1) Every measurable subset of X, with the restriction of $\mathcal{B}$ and μ, is a Lebesgue space. Conversely, if a subset A of X with the restrictions of $\mathcal{B}$ and μ is a Lebesgue space, then A is measurable ($A \in \mathcal{B}$).

(2)The product of countably many Lebesgue spaces is a Lebesgue space.

(3) Every measure algebra isomorphism of $(\hat{\mathcal{B}}, \hat{\mu})$ (defined as above) is induced by a point isomorphism mod 0.

9.1.2. *Disintegration of measures.* Every R-partition ξ of a Lebesgue space $(X, \mathcal{B}, \mu)$ has associated with it a *canonical system of conditional measures*: Using the notation of the preceding section, for μ_ξ-almost every $C \in X/\xi$, there are a σ-algebra $\mathcal{B}_C$ of subsets of C and a measure m_C on $\mathcal{B}_C$ such that:

(1) $(C, \mathcal{B}_C, m_C)$ is a Lebesgue space;

(2) for every $A \in \mathcal{B}$, $A \cap C \in \mathcal{B}_C$ for μ_ξ-almost every $C \in \xi$;

(3) for every $A \in \mathcal{B}$, the map $C \to m_C(A \cap C)$ is $\mathcal{B}(\xi)$-measurable on X/ξ;

(4) for every $A \in \mathcal{B}$,

$$\mu(A) = \int_{X/\xi} m_C(A \cap C)\, d\mu_\xi(C). \tag{9.7}$$

It follows that for $f \in L^1(X)$, (a version of) its conditional expectation (see the next section) with respect to the factor algebra corresponding to ξ is given by

$$\mathbb{E}(f|\mathcal{B}(\xi)) = \int_C f\, dm_C \quad \text{on } \mu_\xi\text{-a.e. } C \in \xi, \tag{9.8}$$

since the right-hand side is $\mathcal{B}(\xi)$-measurable and for each $A \in \mathcal{B}(\xi)$, its integral over any $B \in \mathcal{B}(\xi)$ is, as required, $\mu(A \cap B)$ (use the formula on $B/(\xi|B)$).

It can be shown that a canonical system of conditional measures for an R-partition of a Lebesgue space is essentially unique, in the sense that any two measures m_C and m'_C will be equal for μ_ξ-almost all $C \in \xi$. Also, any partition of a Lebesgue space that has a canonical system of conditional measures must be an R-partition.

These conditional systems of measures can be used to prove the ergodic decomposition theorem and to show that every factor situation is essentially projection of a skew product onto the base (see [44]).

Theorem 9.11. *Let $(X, \mathcal{B}, \mu)$ be a Lebesgue space. If ξ is an R-partition of X, $\{(C, \mathcal{B}_C, m_C)\}$ is a canonical system of conditional measures for ξ, and $A \in \mathcal{B}$, define $\mu(A|C) = m_C(A \cap C)$. Then:*

(1) *for every $A \in \mathcal{B}$, $\mu(A|\xi(x))$ is a measurable function of $x \in X$;*
(2) *if (ξ_n) is an increasing sequence of R-partitions of X, then for each $A \in \mathcal{B}$*

$$\mu(A|\xi_n(x)) \to \mu(A|\bigvee_n \xi_n(X)) \quad \text{a.e. } d\mu; \tag{9.9}$$

(3) *if (ξ_n) is a decreasing sequence of R-partitions of X, then for each $A \in \mathcal{B}$*

$$\mu(A|\xi_n(x)) \to \mu(A|\bigwedge_n \xi_n(X)) \quad \text{a.e. } d\mu. \tag{9.10}$$

This is a consequence of the Martingale and Reverse Martingale Convergence Theorems. The statements hold just as well for $f \in L^1(X)$ as for $f = \mathbf{1}_A$ for some $A \in \mathcal{B}$.

9.2. **Conditional expectation.** Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space, $f \in L^1(X)$, and $\mathcal{F} \subset \mathcal{B}$ a sub-σ-algebra of $\mathcal{B}$. Then

$$\nu(F) = \int_F f\, d\mu \tag{9.11}$$

defines a finite signed measure on $\mathcal{F}$ which is absolutely continuous with respect to μ restricted to $\mathcal{F}$. So by the Radon-Nikodym Theorem there is a function $g \in L^1(X, \mathcal{F}, \mu)$ such that

$$\nu(F) = \int_F g\, d\mu \quad \text{for all } F \in \mathcal{F}. \tag{9.12}$$

Any such function g, which is unique as an element of $L^1(X, \mathcal{F}, \mu)$ (and determined only up to sets of μ-measure 0) is called a version of the *conditional expectation* of

f with respect to $\mathcal{F}$, and denoted by

$$g = \mathbb{E}(f|\mathcal{F}). \tag{9.13}$$

As an element of $L^1(X, \mathcal{B}, \mu)$, $\mathbb{E}(f|\mathcal{F})$ is characterized by the following two properties:

$$\mathbb{E}(f|\mathcal{F}) \quad \text{is } \mathcal{F}\text{-measurable}; \tag{9.14}$$

$$\int_F \mathbb{E}(f|\mathcal{F})\, d\mu = \int_F f\, d\mu \quad \text{for all } F \in \mathcal{F}. \tag{9.15}$$

We think of $\mathbb{E}(f|\mathcal{F})(x)$ as our expected value for f if we are *given the information in* $\mathcal{F}$, in the sense that for each $F \in \mathcal{F}$ we know whether or not $x \in F$. When $\mathcal{F}$ is the σ-algebra generated by a finite measurable partition α of X and f is the characteristic function of a set $A \in \mathcal{B}$, the conditional expectation gives the conditional probabilities of A with respect to all the sets in α:

$$\mathbb{E}(\mathbf{1}_A|\mathcal{F})(x) = \mu(A|\alpha(x)) = \mu(A \cap F)/\mu(F) \quad \text{if } x \in F \in \alpha. \tag{9.16}$$

We write $\mathbb{E}(f) = \mathbb{E}(f|\{\emptyset, X\}) = \int_X f\, d\mu$ for the expectation of any integrable function f. A measurable function f on X is *independent* of a sub-σ-algebra $\mathcal{F} \subset \mathcal{B}$ if for each $(a, b) \subset \mathbb{R}$ and $F \in \mathcal{F}$ we have

$$\mu(f^{-1}(a, b)) \cap F) = \mu(f^{-1}(a, b))\, \mu(F). \tag{9.17}$$

A function $\tau : \mathbb{R} \to \mathbb{R}$ is *convex* if whenever $t_1, \dots, t_n \geq 0$ and $\sum_{i=1}^n t_i = 1$,

$$\phi(\sum_{i=1}^n t_i x_i) \leq \sum_{i=1}^n t_i\, \phi(x_i) \quad \text{for all } x_1, \dots, x_n \in \mathbb{R}. \tag{9.18}$$

Theorem 9.12. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathcal{F} \subset \mathcal{B}$ a sub-σ-algebra.*

(1) $\mathbb{E}(\cdot|\mathcal{F})$ *is a positive contraction on $L^p(X)$ for each $p \geq 1$.*
(2) *If $f \in L^1(X)$ is $\mathcal{F}$-measurable, then $\mathbb{E}(f|\mathcal{F}) = f$ a.e.. If $f \in L^\infty(X)$ is $\mathcal{F}$-measurable, then $\mathbb{E}(fg|\mathcal{F}) = f\mathbb{E}(g|\mathcal{F})$ for all $g \in L^1(X)$.*
(3) *If $\mathcal{F}_1 \subset \mathcal{F}_2$ are sub-σ-algebras of $\mathcal{B}$, then $\mathbb{E}(\mathbb{E}(f|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(f|\mathcal{F}_1)$ a.e. for each $f \in L^1(X)$.*
(4) *If $f \in L^1(X)$ is independent of the sub-σ-algebra $\mathcal{F} \subset \mathcal{B}$, then $\mathbb{E}(f|\mathcal{F}) = \mathbb{E}(f)$ a.e..*
(5) *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, f and $\phi \circ f \in L^1(X)$, and $\mathcal{F} \subset \mathcal{B}$ is a sub-σ-algebra, then $\phi(\mathbb{E}(f|\mathcal{F})) \leq \mathbb{E}(\phi \circ f|\mathcal{F})$ a.e..*

9.3. **The Spectral Theorem.** A *separable Hilbert space* is one with a countable dense set, equivalently a countable orthonormal basis. A *normal operator* is a continuous linear operator S on a Hilbert space $\mathcal{H}$ such that $SS^* = S^*S$, S^* being the adjoint operator defined by $(Sf, g) = (f, S^*g)$ for all $f, g \in \mathcal{H}$. A continuous linear operator S is *unitary* if it is invertible and $S^* = S^{-1}$. Two operators S_1 and S_2 on Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, are called *unitarily equivalent* if there is a unitary operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ which carries S_1 to S_2: $S_2U = US_1$. The following brief account follows [10, 42].

Theorem 9.13. *Let $S : \mathcal{H} \to \mathcal{H}$ be a normal operator on a separable Hilbert space $\mathcal{H}$. Then there are mutually singular Borel probability measures $\mu_\infty, \mu_1, \mu_2, \dots$ such that S is unitarily equivalent to the operator M on the direct sum Hilbert space*

$$L^2(\mathbb{C}, \mu_\infty) \oplus L^2(\mathbb{C}, \mu_1) \oplus \left(\bigoplus_{k=1}^{2} L^2(\mathbb{C}, \mu_2) \right) \oplus \cdots \oplus \left(\bigoplus_{k=1}^{m} L^2(\mathbb{C}, \mu_m) \right) \oplus \dots \tag{9.19}$$

defined by

$$\begin{aligned} M((f_{\infty,1}(z_{\infty,1}), f_{\infty,2}(z_{\infty,2}), \dots), (f_{1,1}(z_{1,1})), (f_{2,1}(z_{2,1}), f_{2,2}(z_{2,2})), \dots) = \\ (z_{\infty,1} f_{\infty,1}(z_{\infty,1}), z_{\infty,2} f_{\infty,2}(z_{\infty,2}), \dots), (z_{1,1} f_{1,1}(z_{1,1})), \\ (z_{2,1} f_{2,1}(z_{2,1}), z_{2,2} f_{2,2}(z_{2,2})), \dots). \end{aligned} \tag{9.20}$$

The measures μ_i are supported on the *spectrum* $\sigma(S)$ of S, the (compact) set of all $\lambda \in \mathbb{C}$ such that $S - \lambda I$ does not have a continuous inverse. Some of the μ_i may be 0. They are uniquely determined up to absolute continuity equivalence. The smallest absolute continuity class with respect to which all the μ_i are absolutely continuous is called the *maximum spectral type of* S. A measure representing this type is $\sum_i \mu_i / 2^i$. We have in mind the example for which $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and $Sf = f \circ T$ (the "Koopman operator") for a measure-preserving system $(X, \mathcal{B}, \mu, T)$ on a Lebesgue space $(X, \mathcal{B}, \mu)$, which is *unitary*: it is linear, continuous, invertible, preserves scalar products, and has spectrum equal to the unit circle.

The proof of Theorem 9.13 can be accomplished by first decomposing $\mathcal{H}$ (in a careful way) into the direct sum of pairwise orthogonal *cyclic subspaces* $\mathcal{H}_n$: each $\mathcal{H}_n$ is the closed linear span of $\{S^i (S^*)^j f_n : i, j \geq 0\}$ for some $f_n \in \mathcal{H}$. This means that for each n the set $\{p(S, S^*) f_n : p \text{ is a polynomial in two variables}\}$ is dense in $\mathcal{H}_n$. Similarly, by the Stone-Weierstrass Theorem the set $\mathcal{P}_n$ of all polynomials $p(z, \overline{z})$ is dense in the set $\mathcal{C}(\sigma(S|\mathcal{H}_n))$ of continuous complex-valued functions on $\sigma(S|\mathcal{H}_n)$. We define a bounded linear functional ϕ_n on $\mathcal{P}_n$ by

$$\phi(p) = (p(S, S^*) f_n, f_n) \tag{9.21}$$

and extend it by continuity to a bounded linear functional on $\mathcal{C}(\sigma(S|\mathcal{H}_n))$. It can be proved that this functional is positive, and therefore, by the Riesz Representation Theorem, it corresponds to a positive Borel measure on $\sigma(S|\mathcal{H}_n)$.

The various L^2 spaces and multiplication operators involved in the above theorem can be amalgamated into a coherent whole, resulting in the following convenient form of the Spectral Theorem for normal operators

Theorem 9.14. *Let $S : \mathcal{H} \to \mathcal{H}$ be a normal operator on a separable Hilbert space $\mathcal{H}$. There are a finite measure space $(X, \mathcal{B}, \mu)$ and a bounded measurable function $h : X \to \mathbb{C}$ such that S is unitarily equivalent to the operator of multiplication by h on $L^2(X, \mathcal{B}, \mu)$.*

The form of the Spectral Theorem given in Theorem 9.13 is useful for discussing absolute continuity and multiplicity properties of the spectrum of a normal operator. Another form, involving spectral measures, has useful consequences such as the functional calculus.

Theorem 9.15. *Let $S : \mathcal{H} \to \mathcal{H}$ be a normal operator on a separable Hilbert space $\mathcal{H}$. There is a unique projection-valued measure E defined on the Borel subsets of the spectrum $\sigma(S)$ of S such that $E(\sigma(S)) = I$ (= the identity on $\mathcal{H}$);*

$$(E(\bigcup_{i=1}^{\infty} A_i))f = \sum_{i=1}^{\infty}(E(A_i))f \tag{9.22}$$

whenever $A_1, A_2, \ldots$ are pairwise disjoint Borel subsets of $\sigma(S)$ and $f \in \mathcal{H}$, with the series converging in norm; and

$$S = \int_{\sigma(S)} \lambda \, dE(\lambda). \tag{9.23}$$

Spectral integrals such as the one in (9.23) can be defined by reducing to complex measures $\mu_{f,g}(A) = (E(A)f, g)$, for $f, g \in \mathcal{H}$ and $A \subset \sigma(S)$ a Borel set. Given a bounded Borel measurable function ϕ on $\sigma(S)$, the operator

$$V = \phi(S) = \int_{\sigma(S)} \phi(\lambda) \, dE(\lambda) \tag{9.24}$$

is determined by specifying that

$$(Vf, g) = \int_{\sigma(S)} \phi(\lambda) \, d\mu_{f,g} \quad \text{for all } f, g \in \mathcal{H}. \tag{9.25}$$

Then

$$S^k = \int_{\sigma(S)} \lambda^k \, dE(\lambda) \quad \text{for all } k = 0, 1, \ldots. \tag{9.26}$$

These spectral integrals sometimes behave a bit strangely:

$$\begin{gathered} \text{If } V_1 = \int_{\sigma(S)} \phi_1(\lambda) \, dE(\lambda) \quad \text{and } V_2 = \int_{\sigma(S)} \phi_2(\lambda) \, dE(\lambda), \text{ then} \\ V_1 V_2 = \int_{\sigma(S)} \phi_1(\lambda) \, \phi_2(\lambda) \, dE(\lambda). \end{gathered} \tag{9.27}$$

Finally, if $f \in \mathcal{H}$ and ν is a finite positive Borel measure that is absolutely continuous with respect to $\mu_{f,f}$, then there is g in the closed linear span of $\{S^i(S^*)^j f : i, j \geq 0\}$ such that $\nu = \mu_{g,g}$.

Theorem 9.15 can be proved by applying Theorem 9.14, which allows us to assume that $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and S is multiplication by $h \in L^\infty(X, \mathcal{B}, \mu)$. For any Borel set $A \subset \sigma(S)$, let $E(A)$ be the projection operator given by multiplication by the $0, 1$-valued function $\mathbf{1}_A \circ h$.

References

[1] Roy L. Adler, *F-expansions revisited*, Recent advances in topological dynamics (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Nedlund), Springer, Berlin, 1973, pp. 1–5. Lecture Notes in Math., Vol. 318. MR0389825 (52 #10655)

[2] Leslie K. Arnold, *On σ-finite invariant measures*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **9** (1968), 85–97. MR0227362 (37 #2947)

[3] Patrick Billingsley, *Ergodic Theory and Information*, Robert E. Krieger Publishing Co., Huntington, N.Y., 1978. Reprint of the 1965 original. MR524567 (80b:28017)

[4] ———, *Probability and Measure*, 3rd ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995. A Wiley-Interscience Publication. MR1324786 (95k:60001)

[5] Antoine Brunel, *Sur les mesures invariantes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** (1966), 300–303 (French). MR0207957 (34 #7770)

[6] Alberto P. Calderón, *Sur les mesures invariantes*, C. R. Acad. Sci. Paris **240** (1955), 1960–1962 (French). MR0069257 (16,1008b)

[7] Constantin Carathéodory, *Die Homomorphieen von Somen und die Multiplikation von Inhaltsfunktionen*, Annali della R. Scuola Normale Superiore di Pisa (Ser. 2) **8** (1939), 105-130.

[8] ———, *Vorlesungen über Reelle Funktionen*, Third (corrected) edition, Chelsea Publishing Co., New York, 1968 (German). MR0225940 (37 #1530)

[9] R. V. Chacon, *A class of linear transformations*, Proc. Amer. Math. Soc. **15** (1964), 560–564. MR0165071 (29 #2362)

[10] John B. Conway, *A Course in Functional Analysis*, 2nd ed., Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, 1990. MR1070713 (91e:46001)

[11] A. del Junco, M. Rahe, and L. Swanson, *Chacon's automorphism has minimal self-joinings*, J. Analyse Math. **37** (1980), 276–284. MR583640 (81j:28027)

[12] Yael Naim Dowker, *On measurable transformations in finite measure spaces*, Ann. of Math. (2) **62** (1955), 504–516. MR0072929 (17,353a)

[13] ———, *Sur les applications mesurables*, C. R. Acad. Sci. Paris **242** (1956), 329–331 (French). MR0074486 (17,593i)

[14] Maurice Frechet, *Des familles et fonctions additives d'ensembles abstraits*, Fund. Math. **5** (191924), 206–251 (French).

[15] Nathaniel A. Friedman, *Introduction to Ergodic Theory*, Van Nostrand Reinhold Co., New York, 1970. Van Nostrand Reinhold Mathematical Studies, No. 29. MR0435350 (55 #8310)

[16] Harry Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49. MR0213508 (35 #4369)

[17] ———, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256. MR0498471 (58 #16583)

[18] Eli Glasner, *Ergodic Theory via Joinings*, Mathematical Surveys and Monographs, vol. 101, American Mathematical Society, Providence, RI, 2003. MR1958753 (2004c:37011)

[19] E. Glasner, B. Host, and D. Rudolph, *Simple systems and their higher order self-joinings*, Israel J. Math. **78** (1992), no. 1, 131–142. MR1194963 (94e:28007)

[20] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588. MR1844079 (2002k:11014)

[21] *Erratum: "A new proof of Szemerédi's theorem" [Geom. Funct. Anal.* **11** *(2001), no. 3, 465–588; 1844079] by W. T. Gowers*, Geom. Funct. Anal. **11** (2001), no. 4, 869. MR1866805

[22] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, available at `arXiv:math.NT/0404188`.

[23] Hans Hahn, *Über die Multiplikation total-additiver Mengenfunktionen*, Annali Scuola Norm. Sup. Pisa **2** (1933), 429–452 (German).

[24] Arshag B. Hajian and Shizuo Kakutani, *Weakly wandering sets and invariant measures*, Trans. Amer. Math. Soc. **110** (1964), 136–151. MR0154961 (27 #4904)

[25] Paul R. Halmos, *Invariant measures*, Ann. of Math. (2) **48** (1947), 735–754. MR0021963 (9,137d)

[26] G. Hansel and J. P. Raoult, *Ergodicity, uniformity and unique ergodicity*, Indiana Univ. Math. J. **23** (1973/74), 221–237. MR0318449 (47 #6996)

[27] Richard H. Herman, Ian F. Putnam, and Christian F. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Internat. J. Math. **3** (1992), no. 6, 827–864. MR1194074 (94f:46096)

[28] Anatole Katok and Boris Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza. MR1326374 (96c:58055)

[29] Eberhard Hopf, *Theory of measure and invariant integrals*, Trans. Amer. Math. Soc. **34** (1932), no. 2, 373–393. MR1501643

[30] ———, *Ergodentheorie*, 1st ed., Ergebnisse der Mathematik und ihrer Grenzgebiete; 5. Bd., 2, J. Springer, Berlin, 1937.

[31] Robert I. Jewett, *The prevalence of uniquely ergodic systems*, J. Math. Mech. **19** (1969/1970), 717–729. MR0252604 (40 #5824)

[32] Shizuo Kakutani, *Induced measure preserving transformations*, Proc. Imp. Acad. Tokyo **19** (1943), 635–641. MR0014222 (7,255f)

[33] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications Inc., New York, N. Y., 1949. Translated by G. Gamow. MR0029808 (10,666c)

[34] Jonathan L. King, *Ergodic properties where order* 4 *implies infinite order*, Israel J. Math. **80** (1992), no. 1-2, 65–86. MR1248927 (94i:28014)

[35] A. N. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea Publishing Co., New York, 1956. Translation edited by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid. MR0079843 (18,155e)

[36] Wolfgang Krieger, *On unique ergodicity*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Univ. California Press, Berkeley, Calif., 1972, pp. 327–346. MR0393402 (52 #14212)

[37] Donald S. Ornstein, *On invariant measures*, Bull. Amer. Math. Soc. **66** (1960), 297–300. MR0146350 (26 #3872)

[38] ———, *On the root problem in ergodic theory*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Univ. California Press, Berkeley, Calif., 1972, pp. 347–356. MR0399415 (53 #3259)

[39] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original. MR2169627 (2006d:60004)

[40] Karl Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989. Corrected reprint of the 1983 original. MR1073173 (92c:28010)

[41] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste. Tomes I, II,III*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Librairie Scientifique et Technique Albert Blanchard, Paris, 1987 (French). Reprint of the 1892 original; With a foreword by J. Kovalevsky; Bibliothèque Scientifique Albert Blanchard. [Albert Blanchard Scientific Library]. MR926906 (89e:01054a)

[42] Heydar Radjavi and Peter Rosenthal, *Invariant subspaces*, 2nd ed., Dover Publications Inc., Mineola, NY, 2003. MR2003221 (2004e:47010)

[43] V. A. Rohlin, *On the fundamental ideas of measure theory*, Amer. Math. Soc. Translation **1952** (1952), no. 71, 55. MR0047744 (13,924e)

[44] ———, *New progress in the theory of transformations with invariant measure.*, Russian Math. Surveys **15** (1960), no. 4, 1–22. MR0132155 (24 #A2002)

[45] H. L. Royden, *Real Analysis*, 3rd ed., Macmillan Publishing Company, New York, 1988. MR1013117 (90g:00004)

[46] Daniel J. Rudolph, *An example of a measure preserving map with minimal self-joinings, and applications*, J. Analyse Math. **35** (1979), 97–122. MR555301 (81e:28011)

[47] ———, *Fundamentals of Measurable Dynamics*, Oxford Science Publications, The Clarendon Press Oxford University Press, New York, 1990. Ergodic theory on Lebesgue spaces. MR1086631 (92e:28006)

[48] V. V. Ryzhikov, *Joinings, wreath products, factors and mixing properties of dynamical systems*, Izv. Ross. Akad. Nauk Ser. Mat. **57** (1993), no. 1, 102–128 (Russian, with Russian summary); English transl., Russian Acad. Sci. Izv. Math. **42** (1994), no. 1, 91–114. MR1220583 (94h:28016)

[49] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245. Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR0369312 (51 #5547)

[50] Terence Tao, *Szemerédi's regularity lemma revisited*, Contrib. Discrete Math. **1** (2006), no. 1, 8–28 (electronic). MR2212136

[51] ———, *Arithmetic progressions and the primes*, Collect. Math. **Vol. Extra** (2006), 37–88. MR2264205

[52] J.-P. Thouvenot, *Some properties and applications of joinings in ergodic theory*, Ergodic theory and its connections with harmonic analysis (Alexandria, 1993), London Math. Soc. Lecture Note Ser., vol. 205, Cambridge Univ. Press, Cambridge, 1995, pp. 207–235. MR1325699 (96d:28017)

[53] A. M. Vershik, *A theorem on Markov periodic approximation in ergodic theory*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **115** (1982), 72–82, 306 (Russian). Boundary value problems of mathematical physics and related questions in the theory of functions, 14. MR660072 (84m:28030)

[54] ———, *The adic realizations of the ergodic actions with the homeomorphisms of the Markov compact and the ordered Bratteli diagrams*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **223** (1995), no. Teor. Predstav. Din. Sistemy, Kombin. i Algoritm. Metody. I, 120–126, 338 (English, with English and Russian summaries); English transl., J. Math. Sci. (New York) **87** (1997), no. 6, 4054–4058. MR1374314 (97b:28024)

[55] J. von Neumann, *Einige Sätze über messbare Abbildungen*, Ann. of Math. (2) **33** (1932), no. 3, 574–586 (German). MR1503077

[56] Fred B. Wright, *The recurrence theorem*, Amer. Math. Monthly **68** (1961), 247–248. MR0133427 (24 #A3261)

[57] Lai-Sang Young, *What are SRB measures, and which dynamical systems have them?*, J. Statist. Phys. **108** (2002), no. 5-6, 733–754. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. MR1933431 (2003g:37042)

Department of Mathematics, CB 3250, Phillips Hall, University of North Carolina, Chapel Hill, NC 27599 USA

E-mail address: petersen@math.unc.edu