Chapter 3
Spectral Theory

3.1 The spectral approach to ergodic theory

A basic problem in ergodic theory is to determine whether two ppt are measure theoretically isomorphic. This is done by studying *invariants*: properties, quantities, or objects which are equal for any two isomorphic systems. The idea is that if two ppt have different invariants, then they cannot be isomorphic. Ergodicity and mixing are examples of invariants for measure theoretic isomorphism.

An effective method for inventing invariants is to look for a *weaker* equivalence relation, which is better understood. Any invariant for the weaker equivalence relation is automatically an invariant for measure theoretic isomorphism. The spectral point of view is based on this approach.

The idea is to associate to the ppt $(X,\mathscr{B},\mu,T)$ the operator $U_T : L^2(X,\mathscr{B},\mu) \to L^2(X,\mathscr{B},\mu)$, $U_t f = f \circ T$. This is an isometry of L^2 (i.e. $\|U_T f\|_2 = \|f\|_2$ and $\langle U_T f, U_T g\rangle = \langle f,g\rangle$). It is useful here to think of L^2 as a Hilbert space over $\mathbb{C}$.

Definition 3.1. Two ppt $(X,\mathscr{B},\mu,T)$, $(Y,\mathscr{C},\nu,S)$ are called *spectrally isomorphic*, if their associated L^2–isometries U_T and U_S are *unitarily equivalent*, namely if there exists a linear operator $W : L^2(X,\mathscr{B},\mu) \to L^2(Y,\mathscr{C},\nu)$ s.t.

1. W is invertible;
2. $\langle Wf, Wg\rangle = \langle f,g\rangle$ for all $f,g \in L^2(X,\mathscr{B},\mu)$;
3. $WU_T = U_S W$.

It is easy to see that any two measure theoretically isomorphic ppt are spectrally isomorphic, but we will see later that there are Bernoulli schemes which are spectrally isomorphic but not measure theoretically isomorphic.

Definition 3.2. A property of ppt is called a *spectral invariant*, if whenever it holds for $(X,\mathscr{B},\mu,T)$, it holds for all ppt which are spectrally isomorphic to $(X,\mathscr{B},\mu,T)$.

Proposition 3.1. *Ergodicity and mixing are spectral invariants.*

Proof. Suppose $(X,\mathscr{B},\mu,T)$ is a ppt, and let U_T be as above. The trick is to phrase ergodicity and mixing in terms of U_T.

Ergodicity is equivalent to the statement "all invariant L^2–functions are constant", which is the same as saying that $\dim\{f : U_T f = f\} = 1$. Obviously, this is a spectral invariant.

Mixing is equivalent to the following statement: $\dim\{f : U_T f = f\} = 1$, and

$$\langle f, U_T^n g\rangle \xrightarrow[n\to\infty]{} \langle f,1\rangle\overline{\langle g,1\rangle} \text{ for all } f,g\in L^2.$$

To see that this property is preserved by spectral isomorphisms, note that if $\dim\{f : U_T f = f\} = 1$, then any unitary equivalence W satisfies $W1 = c$ with $|c| = 1$. □

The spectral point of view immediately suggests the following invariant.

Definition 3.3. Suppose $(X,\mathscr{B},\mu,T)$ is a ppt. If $f : X \to \mathbb{C}$, $f \in L^2$ satisfies $f\circ T = \lambda f$, then we say that f is an *eigenfunction* and that λ is an *eigenvalue*. The *point spectrum* T is the set $H(T) := \{\lambda \in \mathbb{C} : f\circ T = \lambda f\}$.

$H(T)$ is a countable subgroup of the unit circle (problem 3.1). Evidently $H(T)$ is a spectral invariant of T.

It is easy to see using Fourier expansions that for the irrational rotation R_α, $H(R_\alpha) = \{\alpha^k : k \in \mathbb{Z}\}$ (problem 3.2), thus irrational rotations by different angles are non-isomorphic.

Here are other related invariants:

Definition 3.4. Given a ppt $(X,\mathscr{B},\mu,T)$, let $V_d := \overline{\text{span}}\{\text{eigenfunctions}\}$. We say that $(X,\mathscr{B},\mu,T)$ has

1. *discrete spectrum* (sometime called *pure point spectrum*), if $V_d = L^2$,
2. *continuous spectrum*, if $V_d = \{\text{constants}\}$ (i.e. is smallest possible),
3. *mixed spectrum*, if $V_d \neq L^2, \{\text{constants}\}$.

Any irrational rotation has discrete spectrum (problem 3.2). Any mixing transformation has continuous spectrum, because a non-constant eigenfunction $f\circ T = \lambda$ satisfies

$$\langle f, f\circ T^{n_k}\rangle \xrightarrow[n\to\infty]{} \|f\|_2^2 \neq \left(\int f\right)^2$$

along any $n_k \to \infty$ s.t. $\lambda^{n_k} \to 1$. (To see that $\|f\|_2 \neq (\int f d\mu)^2$ for all non-constant functions, apply Cauchy-Schwarz to $f - \int f$, or note that non-constant L^2 functions have positive variance.)

The invariant $H(T)$ is tremendously successful for transformations with discrete spectrum:

Theorem 3.1 (Discrete Spectrum Theorem). *Two ppt with discrete spectrum are measure theoretically isomorphic iff they have the same group of eigenvalues.*

But this invariant cannot distinguish transformations with continuous spectrum. In particular - it is unsuitable for the study of mixing transformations.

3.2 Weak mixing

3.2.1 Definition and characterization

We saw that if a transformation is mixing, then it does not have non-constant eigenfunctions. But the absence of non-constant eigenfunctions is not equivalent to mixing (see problems 3.8–3.10 for an example). Here we study the dynamical significance of this property. First we give it a name.

Definition 3.5. A ppt is called *weak mixing*, if every $f \in L^2$ s.t. $f \circ T = \lambda f$ a.e. is constant almost everywhere.

Theorem 3.2. *The following are equivalent for a ppt $(X, \mathscr{B}, \mu, T)$ on a Lebesgue space:*

1. *weak mixing;*
2. *for all $E, F \in \mathscr{B}$, $\frac{1}{N}\sum_{k=0}^{N-1} |\mu(E \cap T^{-n}F) - \mu(E)\mu(F)| \xrightarrow[N\to\infty]{} 0$;*
3. *for every $E, F \in \mathscr{B}$, $\exists \mathscr{N} \subset \mathbb{N}$ of density zero (i.e. $|\mathscr{N} \cap [1,N]|/N \xrightarrow[N\to\infty]{} 0$) s.t. $\mu(E \cap T^{-n}F) \xrightarrow[\mathscr{N}\not\ni n\to\infty]{} \mu(E)\mu(F)$;*
4. *$T \times T$ is ergodic.*

Proof. We prove $(2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1)$. The remaining implication $(1) \Rightarrow (2)$ requires additional preparation, and will be shown later.

The implication $(2) \Rightarrow (3)$ is a general fact from calculus (Koopman–von Neumann Lemma): If a_n is a bounded sequence of non-negative numbers, then $\frac{1}{N}\sum_{n=1}^{N} a_n \to 0$ iff there is a set of zero density $\mathscr{N} \subset \mathbb{N}$ s.t. $a_n \xrightarrow[\mathscr{N}\not\ni n\to\infty]{} 0$ (Problem 3.3).

We show that $(3) \Rightarrow (4)$. Let $\mathscr{S}$ be the semi-algebra $\{E \times F : E, F \in \mathscr{B}\}$ which generates $\mathscr{B} \otimes \mathscr{B}$, and fix $E_i \times F_i \in \mathscr{S}$. By (3), $\exists \mathscr{N}_i \subset \mathbb{N}$ of density zero s.t.

$$\mu(E_i \cap T^{-n}F_i) \xrightarrow[\mathscr{N}_i\not\ni n\to\infty]{} \mu(E_i)\mu(F_i) \quad (i = 1, 2).$$

The set $\mathscr{N} = \mathscr{N}_1 \cup \mathscr{N}_2$ also has zero density, and

$$\mu(E_i \cap T^{-n}F_i) \xrightarrow[\mathscr{N}\not\ni n\to\infty]{} \mu(E_i)\mu(F_i) \quad (i = 1, 2).$$

Writing $m = \mu \times \mu$ and $S = T \times T$, we see that this implies that

$$m[(E_1 \times E_2) \cap S^{-n}(F_1 \times F_2)] \xrightarrow[\mathscr{N}\not\ni n\to\infty]{} m(E_1 \times F_1)m(E_2 \times F_2),$$

whence $\frac{1}{N}\sum_{k=0}^{N-1} m[(E_1 \times F_1) \cap S^{-n}(E_2 \times F_2)] \xrightarrow[N\to\infty]{} m(E_1 \times F_1)m(E_2 \times F_2)$. In summary, $\frac{1}{N}\sum_{k=0}^{N-1} m[A \cap S^{-n}B] \xrightarrow[N\to\infty]{} m(A)m(B)$ for all $A, B \in \mathscr{S}$.

Since $\mathscr{S}$ generates $\mathscr{B} \otimes \mathscr{B}$ the above holds for all $A, B \in \mathscr{B} \otimes \mathscr{B}$, and this implies that $T \times T$ is ergodic.

Proof that (4)$\Rightarrow$ (1): Suppose T were not weak mixing, then T has an non-constant eigenfunction f with eigenvalue λ. The eigenvalue λ has absolute value equal to one, because $|\lambda|\|f\|_2 = \||f|\circ T\|_2 = \|f\|_2$. Thus

$$F(x,y) = f(x)\overline{f(y)}$$

is $T\times T$–invariant. Since f is non-constant, F is non-constant, and we get a contradiction to the ergodicity of $T\times T$.

The proof that (1)$\Rightarrow$ (2) is presented in the next section. $\square$

3.2.2 Spectral measures and weak mixing

It is convenient to introduce the following notation $U_T^n := (U_T^*)^{|n|}$ where $n<0$, where U_T^* is the unique operator s.t. $\langle U_T^* f, g\rangle = \langle f, U_T g\rangle$ for all $g \in L^2$. This makes sense even if U_T is not invertible. The reader can check that when U_T is invertible, $U_T^{-1} = (U_T)^{-1}$, so that there is no risk of confusion.

We are interested in the behavior of $U_T^n f$ as $n\to\pm\infty$. To study it, it is enough to study $U_T : H_f \to H_f$, where $H_f := \overline{\operatorname{span}}\{U_T^n f : n\in\mathbb{Z}\}$.

It turns out that $U_T : H_f \to H_f$ is unitarily equivalent to the operator $M : g(z)\mapsto zg(z)$ on $L^2(S^1, \mathscr{B}(S^1), \nu_f)$ where ν_f is some finite measure on S^1, called the *spectral measure* of f, which contains all the information on $U_T : H_f \to H_f$.

To construct it, we need the following important tool from harmonic analysis. Recall that The n-th *Fourier coefficient* of μ is the number $\widehat{\mu}(n) = \int_{S^1} z^n d\mu$.

Theorem 3.3 (Herglotz). *A sequence $\{r_n\}_{n\in\mathbb{Z}}$ is the sequence of Fourier coefficients of a positive Borel measure on S^1 iff $r_{-n} = \overline{r_n}$ and $\{r_n\}$ is* positive definite*: $\sum_{n,m=-N}^{N} r_{n-m} a_m \overline{a_n} \geq 0$ for all sequences $\{a_n\}$ and N. This measure is unique.*

It is easy to check that $r_n = \langle U_T^n f, f\rangle$ is positive definite (to see this expand $\langle \Sigma_{n=-N}^N a_n U_T^n f, \Sigma_{m=-N}^N a_m U_T^m f\rangle$ noting that $\langle U_T^n f, U_T^m f\rangle = \langle U_T^{n-m} f, f\rangle$).

Definition 3.6. Suppose $(X, \mathscr{B}, \mu, T)$ is a ppt, and $f\in L^2\setminus\{0\}$. The *spectral measure* of f is the unique measure ν_f on S^1 s.t. $\langle f\circ T^n, f\rangle = \int_{S^1} z^n d\nu_f$ for $n\in\mathbb{Z}$.

Proposition 3.2. *Let $H_f := \overline{\operatorname{span}}\{U_T^n f : n\in\mathbb{Z}\}$, then $U_T : H_f \to H_f$ is unitarily equivalent to the operator $g(z)\mapsto zg(z)$ on $L^2(S^1, \mathscr{B}(S^1), \nu_f)$.*

Proof. By the definition of the spectral measure,

$$\left\|\sum_{n=-N}^{N} a_n z^n\right\|^2_{L^2(\nu_f)} = \left\langle \sum_{n=-N}^{N} a_n z^n, \sum_{m=-N}^{N} a_m z^m\right\rangle = \sum_{n,m=-N}^{N} a_n\overline{a}_m \int_{S^1} z^{n-m} d\nu_f(z)$$

$$= \sum_{n,m=-N}^{N} a_n\overline{a}_m \langle U_T^{n-m} f, f\rangle = \sum_{n,m=-N}^{N} a_n\overline{a}_m \langle U_T^n f, U_T^m f\rangle = \left\|\sum_{n=-N}^{N} a_n U_T^n f\right\|^2_{L^2(\mu)}$$

In particular, if $\Sigma_{n=-N}^{N} a_n U_T^n f = 0$ in $L^2(\mu)$, then $\Sigma_{n=-N}^{N} a_n z^n = 0$ in $L^2(\nu_f)$. It follows that $W : U_T^n f \mapsto z^n$ extends to a linear map from $\mathrm{span}\{U_T^n f : n \in \mathbb{Z}\}$ to $L^2(\nu_f)$.

This map is an isometry, and it is bounded. It follows that W extends to an linear isometry $W : H_f \to L^2(\nu_f)$. The image of W contains all the trigonometric polynomials, therefore $W(H_f)$ is dense in $L^2(\nu_f)$. Since W is an isometry, its image is closed (exercise). It follows that W is an isometric bijection from H_f onto $L^2(\nu_f)$.

Since $(WU_t)[g(z)] = z[Wg(z)]$ on $\mathrm{span}\{U_T^n f : n \in \mathbb{Z}\}$, $WU_T g(z) = zg(z)$ on H_f, and so W is the required unitary equivalence. □

Proposition 3.3. *If T is weak mixing ppt on a Lebesgue space, then all the spectral measures of $f \in L^2$ s.t. $\int f = 0$ are non-atomic (this explains the terminology "continuous spectrum").*

Proof. Suppose $f \in L^2$ has integral zero and that ν_f has an atom $\lambda \in S^1$. We construct an eigenfunction (with eigenvalue λ). Consider the sequence $\frac{1}{N}\sum_{n=0}^{N-1} \lambda^{-n} U_T^n f$. This sequence is bounded in norm, therefore has a weakly convergent subsequence (here we use the fact that L^2 is separable — a consequence of the fact that $(X, \mathscr{B}, \mu)$ is a Lebesgue space):

$$\frac{1}{N_k}\sum_{n=0}^{N-1} \lambda^{-k} U_T^k \xrightarrow[N\to\infty]{w} g.$$

The limit g must satisfy $\langle U_T g, h\rangle = \langle \lambda g, h\rangle$ (check!), therefore it must be an eigenfunction with eigenvalue λ.

But it could be that $g = 0$. We rule this out using the assumption that $\nu_f\{\lambda\} \neq 0$:

$$\begin{aligned}
\langle g, f\rangle &= \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=0}^{N_k-1} \lambda^{-n}\langle U_T^n f, f\rangle = \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=0}^{N_k-1} \int \lambda^{-n} z^n d\nu_f(z) \\
&= \nu_f\{\lambda\} + \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=0}^{N_k-1} \int_{S^1\setminus\{\lambda\}} \lambda^{-n} z^n d\nu_f(z) \\
&= \nu_f\{\lambda\} + \lim_{k\to\infty} \int_{S^1\setminus\{\lambda\}} \frac{1}{N_k} \frac{1-\lambda^{-N_k} z^{N_k}}{1-\lambda^{-1} z} d\nu_f(z).
\end{aligned}$$

The limit is equal to zero, because the integrand tends to zero and is uniformly bounded (by one). Thus $\langle g, f\rangle = \nu_f\{\lambda\} \neq 0$, whence $g \neq 0$. □

Lemma 3.1. *Suppose T is a ppt on a Lebesgue space. If T is weak mixing, then for every $f \in L^2$, $\frac{1}{N}\sum\limits_{k=0}^{N-1} |\int f \cdot f \circ T^n d\mu - (\int f d\mu)^2| \xrightarrow[N\to\infty]{} 0.$*

Proof. It is enough to treat the case when $\int f d\mu = 0$. Let ν_f denote the spectral measure of f, then

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\left|\int f\cdot f\circ T^n d\mu\right|^2 &= \frac{1}{N}\sum_{k=0}^{N-1}|\langle U_T^n f,f\rangle|^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left|\int_{S^1} z^n d\nu_f(z)\right|^2\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\left(\int_{S^1} z^n d\nu_f(z)\right)\overline{\left(\int_{S^1} z^n d\nu_f(z)\right)}\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\int_{S^1}\int_{S^1} z^n\overline{w}^n d\nu_f(z)d\nu_f(w)\\ &= \int_{S^1}\int_{S^1}\frac{1}{N}\left(\sum_{k=0}^{N-1} z^n\overline{w}^n\right)d\nu_f(z)d\nu_f(w)\end{aligned}$$

The integrand tends to zero and is bounded outside $\Delta := \{(z,w) : z = w\}$. If we can show that $(\nu_f \times \nu_f)(\Delta) = 0$, then it will follow that $\frac{1}{N}\sum\limits_{k=0}^{N-1}|\int f\cdot f\circ T^n d\mu|^2 \xrightarrow[N\to\infty]{} 0$. This is indeed the case: T is weak mixing, so by the previous proposition ν_f is non-atomic, whence $(\nu_f\times\nu_f)(\Delta) = \int_{S^1}\nu_f\{w\}d\nu_f(w) = 0$ by Fubini-Tonelli.

It remains to note that by the Koopman - von Neumann theorem, for every bounded non-negative sequence a_n, $\frac{1}{N}\sum_{k=1}^N a_n^2 \to 0$ iff $\frac{1}{N}\sum_{k=1}^N a_n \to 0$, because both conditions are equivalent to saying that a_n converges to zero outside a set of indices of density zero. $\square$

We can now complete the proof of the theorem in the previous section:

Proposition 3.4. *If T is weak mixing, then for all $f,g\in L^2$,*

$$\frac{1}{N}\sum_{k=0}^{N-1}\left|\int g\cdot f\circ T^n d\mu - \left(\int f d\mu\right)\left(\int g d\mu\right)\right| \xrightarrow[N\to\infty]{} 0. \tag{3.1}$$

Proof. Assume first T is invertible, then U_T is invertible, with a bounded inverse (equal to $U_{T^{-1}}$). Fix $f\in L^2$, and set

$$S(f) := \overline{\text{span}}\{U_T^k f : k\in\mathbb{Z}\}.$$

Write $L^2 = S(f) + \{\text{constants}\} + [S(f) + \text{constants}]^\perp$.

1. Every $g\in S(f)$ satisfies (3.1), because $S(f)$ is generated by functions of the form $g := U_T^k f$, and these functions satisfy (3.1) by Lemma 3.1.
2. Every constant g satisfies (3.1) trivially.
3. Every $g\perp S(f)\oplus\{\text{constants}\}$ satisfies (3.1) because $\langle g, f\circ T^n\rangle$ is eventually zero.

It follows that every $g\in L^2$ satisfies (3.1).

Now consider the case of a non-invertible ppt. Let $(\widetilde{X},\widetilde{\mathscr{B}},\widetilde{\mu},\widetilde{T})$ be the natural extension. A close look at the definition of $\widetilde{\mathscr{B}}$ shows that if $\widetilde{f}:\widetilde{X}\to\mathbb{R}$ is $\widetilde{\mathscr{B}}$–measurable, then the value of $\widetilde{f}(\ldots,x_{-1},x_0,x_1,\ldots)$ is completely determined by x_0. Moreover, $\widetilde{f}:\widetilde{X}\to\mathbb{C}$ is of the form $f\circ\widetilde{\pi}$ where f is $\mathscr{B}$–measurable. Thus every eigenfunction for $\widetilde{T}$ is a lift of an eigenfunction for T. It follows that if T is weak mixing, then $\widetilde{T}$ is weak mixing.

By the first part of the proof, $\widetilde{T}$ satisfies (3.1). Since T is a factor of T, it also satisfies (3.1). $\square$

3.3 The Koopman operator of a Bernoulli scheme

In this section we analyze the Koopman operator of an invertible Bernoulli scheme. The idea is to produce an orthonormal basis for L^2 which makes the action of U_T transparent.

We cannot expect to diagonalize U_T: Bernoulli schemes are mixing, so they have no non-constant eigenfunctions. But we shall we see that we can get the following nice structure:

Definition 3.7. An invertible ppt is said to have *countable Lebesgue spectrum* if L^2 has an orthonormal basis of the form $\{1\} \cup \{f_{\lambda,j} : \lambda \in \Lambda, j \in \mathbb{Z}\}$ where Λ is countable, and $U_T f_{\lambda,j} = f_{\lambda,j+1}$ for all i, j.

The reason for the terminology is that the spectral measure of each $f_{\lambda,j}$ is proportional to the Lebesgue measure on S^1 (problem 3.6).

Example. The invertible Bernoulli scheme with probability vector $(\frac{1}{2}, \frac{1}{2})$ has countable Lebesgue spectrum.

Proof. The phase space is $X = \{0,1\}^{\mathbb{Z}}$. Define for every finite non-empty $A \subset \mathbb{Z}$ the function $\varphi_A(\underline{x}) := \prod_{j \in A}(-1)^{x_j}$. Define $\varphi_\varnothing := 1$. Then,

1. if $A \neq B$, then $\varphi_A \perp \varphi_B$;
2. $\mathrm{span}\{\varphi_A : A \subset \mathbb{Z} \text{ finite}\}$ is algebra of functions which separates points, and contains the constants.

By the Stone-Weierstrass theorem, $\overline{\mathrm{span}}\{\varphi_A : A \subset \mathbb{Z} \text{ finite}\} = L^2$, so $\{\varphi_A\}$ is an orthonormal basis of L^2. This is called the *Fourier–Walsh* system.

Note that $U_T \varphi_A = \varphi_{A+1}$, where $A+1 := \{a+1 : a \in A\}$. Take Λ the set of equivalence classes of the relation $A \sim B \Leftrightarrow \exists c$ s.t. $A = c+B$. Let A_λ be a representative of $\lambda \in \Lambda$. The basis is $\{1\} \cup \{\varphi_{A_\lambda + n} : \lambda \in \Lambda, n \in \mathbb{Z}\} = \{\text{Fourier Walsh functions}\}$. $\square$

It is not easy to produce such bases for other Bernoulli schemes. But they exist. To prove this we introduce the following sufficient condition for countable Lebesgue spectrum, which turns out to be satisfied by many smooth dynamical systems:

Definition 3.8. An invertible ppt $(X, \mathscr{B}, \mu, T)$ is called a *K automorphism* if there is a σ–algebra $\mathscr{A} \subset \mathscr{B}$ s.t.

1. $T^{-1}\mathscr{A} \subset \mathscr{A}$;
2. $\mathscr{A}$ generates $\mathscr{B}$: $\sigma(\bigcup_{n \in \mathbb{Z}} T^{-n}\mathscr{A}) = \mathscr{B} \mod \mu$;[1]
3. the tail of $\mathscr{A}$ is trivial: $\bigcap_{n=0}^{\infty} T^{-n}\mathscr{A} = \{\varnothing, X\} \mod \mu$.

[1] $\mathscr{F}_1 \subset \mathscr{F}_2 \mod \mu$ is for all $F_1 \in \mathscr{F}_2$ there is a set $F_2 \in \mathscr{F}_2$ s.t. $\mu(F_1 \triangle F_2) = 0$, and $\mathscr{F}_1 = \mathscr{F}_2 \mod \mu$ iff $\mathscr{F}_1 \subset \mathscr{F}_2 \mod \mu$ and $\mathscr{F}_2 \subset \mathscr{F}_1 \mod \mu$.

Proposition 3.5. *Every invertible Bernoulli scheme has the K property.*

Proof. Let $(S^{\mathbb{Z}}, \mathscr{B}(S^{\mathbb{Z}}), \mu, T)$ be a Bernoulli scheme, i.e. $\mathscr{B}(S^{\mathbb{Z}})$ is the sigma algebra generated by *cylinders* ${}_{-k}[a_{-k}, \dots, a_\ell] := \{x \in S^{\mathbb{Z}} : x_i = a_i \ \ (-k \le i \le \ell)\}$, T is the left shift map, and $\mu({}_k[a_{-k}, \dots, a_\ell]) = p_{a_{-k}} \cdots p_{a_\ell}$.

Call a cylinder *non-negative*, if it is of the form ${}_0[a_0, \dots, a_n]$. Let $\mathscr{A}$ be the sigma algebra generated by all non-negative cylinders. It is clear that $T^{-1}\mathscr{A} \subset \mathscr{A}$ and that $\bigcup_{n\in\mathbb{Z}} T^{-n}\mathscr{A}$ generates $\mathscr{B}(S^{\mathbb{Z}})$. We show that the measure of every element of $\bigcap_{n=0}^{\infty} T^{-n}\mathscr{A}$ is either zero or one. Probabilists call the elements of this intersection *tail events*. The fact that every tail event for a sequence of independent identically distributed random variables has probability zero or one is called "Kolmogorov's zero–one law".

Two measurable sets A, B are called *independent*, if $\mu(A \cap B) = \mu(A)\mu(B)$. For Bernoulli schemes, any two cylinders with non-overlapping set of indices is independent (check). Thus for every cylinder B of length $|B|$,

$$B \text{ is independent of } T^{-|B|}A \text{ for all non-negative cylinders } A.$$

It follows that B is independent of every element of $T^{-|B|}\mathscr{A}$ (a monotone class theorem argument). Thus every cylinder B is independent of every element of $\bigcap_{n\ge1} T^{-n}\mathscr{A}$. Thus every element of $\mathscr{B}$ is independent of every element of $\bigcap_{n\ge1} T^{-n}\mathscr{A}$ (another monotone class theorem argument).

This means that every $E \in \bigcap_{n\ge1} T^{-n}\mathscr{A}$ is independent of itself. Thus $\mu(E) = \mu(E \cap E) = \mu(E)^2$, whence $\mu(E) = 0$ or 1. $\square$

Proposition 3.6. *Every K automorphism on a non-atomic standard probability space has countable Lebesgue spectrum.*

Proof. Let $(X, \mathscr{B}, \mu, T)$ be a K automorphism of a non-atomic standard probability space. Since $(X, \mathscr{B}, \mu)$ is a non-atomic standard space, $L^2(X, \mathscr{B}, \mu)$ is (i) infinite dimensional, and (ii) separable.

Let $\mathscr{A}$ be a sigma algebra in the definition of the K property. Set $V := L^2(X, \mathscr{A}, \mu)$. This is a closed subspace of $L^2(X, \mathscr{B}, \mu)$, and

1. $U_T(V) \subseteq V$, because $T^{-1}\mathscr{A} \subset \mathscr{A}$;
2. $\bigcup_{n\in\mathbb{Z}} U_T^n(V)$ is dense in $L^2(X, \mathscr{B}, \mu)$, because $\bigcup_{n\in\mathbb{Z}} T^{-n}\mathscr{A}$ generates $\mathscr{B}$, so every $B \in \mathscr{B}$ can be approximated by a finite disjoint union of elements of $\bigcup_{n\in\mathbb{Z}} T^{-n}\mathscr{A}$;
3. $\bigcap_{n=1}^{\infty} U_T^n(V) = \{\text{constant functions}\}$, because $\bigcap_{n\ge1} T^{-n}\mathscr{A} = \{\varnothing, X\} \mod \mu$.

Now let $W := V \ominus U_T(V)$ (the orthogonal complement of $U_T(V)$ in V). For all $n > 0$, $U_T^n(W) \subset U_T^n(V) \subset U_T(V) \perp W$. Thus $W \perp U_T^n(W)$ for all $n > 0$. Since U_T^{-1} is an isometry, $W \perp U_T^n(W)$ for all $n < 0$. It follows that

$$L^2(X, \mathscr{B}, \mu) = \{\text{constants}\} \oplus \bigoplus_{n\in\mathbb{Z}} U_T^n(W) \quad \text{(orthogonal sum)}.$$

If $\{f_\lambda : \lambda \in \Lambda\}$ is an orthonormal basis for W, then the above implies that

$$\{1\} \cup \{U_T^n f_\lambda : \lambda \in \Lambda\}$$

is an orthonormal basis of $L^2(X, \mathscr{B}, \mu)$ (check!).

This is almost the full countable Lebesgue spectrum property. It remains to show that $|\Lambda| = \aleph_0$. $|\Lambda| \leq \aleph_0$ because $L^2(X, \mathscr{B}, \mu)$ is separable. We show that Λ is infinite by proving $\dim(W) = \infty$. We use the following fact (to be proved later):

$$\forall N \; \exists A_1, \ldots, A_N \in \mathscr{A} \text{ pairwise disjoint sets, with positive measure.} \tag{3.2}$$

Suppose we know this. Pick $f \in W \setminus \{0\}$ ($W \neq \{0\}$, otherwise $L^2 = \{\text{constants}\}$ and $(X, \mathscr{B}, \mu)$ is atomic). Set $w_i := f 1_{A_i} \circ T$ with $A_1, \ldots, A_N$ as above, then (i) w_i are linearly independent (because they have disjoint supports); (ii) $w_i \in V$ (because $T^{-1}A_i \in T^{-1}\mathscr{A} \subset \mathscr{A}$, so w_i is $\mathscr{A}$–measurable); and (iii) $w_i \perp U_T(V)$ (check, using $f \in W$). It follows that $\dim(W) \geq N$. Since N was arbitrary, $\dim(W) = \infty$.

Here is the proof of (3.2). Since $(X, \mathscr{B}, \mu)$ is non-atomic, $\exists B_1, \ldots, B_N \in \mathscr{B}$ pairwise disjoint with positive measure. By assumption, $\bigcup_{n \in \mathbb{Z}} T^n \mathscr{A}$ generates $\mathscr{B}$, thus we can approximate B_i arbitrarily well by elements of $\bigcup_{n \in \mathbb{Z}} T^n \mathscr{A}$. By assumption, $\mathscr{A} \subseteq T\mathscr{A}$. This means that we can approximate B_i arbitrarily well by sets from $T^n \mathscr{A}$ by choosing n sufficiently large. It follows that $L^2(X, T^n \mathscr{A}, \mu)$ has dimension at least N. This forces $T^n \mathscr{A}$ to contain at least N pairwise disjoint sets of positive measure. It follows that $\mathscr{A}$ contains at least N pairwise disjoint sets of positive measure. ☐

Corollary 3.1. *All systems with countable Lebesgue spectrum, whence all invertible Bernoulli schemes, are spectrally isomorphic.*

Proof. Problem 3.7. ☐

But it is not true that all Bernoulli schemes are measure theoretically isomorphic. To prove this one needs new (non-spectral) invariants. Enter the *measure theoretic entropy*, which we discuss in the next chapter.

Problems

3.1. Suppose $(X, \mathscr{B}, \mu, T)$ is an ergodic ppt on a Lebesgue space, and let $H(T)$ be its group of eigenvalues.

1. show that if f is an eigenfunction, then $|f| = \text{const.}$ a.e., and that if $\lambda, \mu \in H(T)$, then so do $1, \lambda\mu, \lambda/\mu$.
2. Show that eigenfunctions of different eigenvalue are orthogonal. Deduce that $H(T)$ is a countable subgroup of the unit circle.

3.2. Prove that the irrational rotation R_α has discrete spectrum, and calculate $H(R_\alpha)$.

3.3. Koopman - von Neumann Lemma
Suppose a_n is a bounded sequence of non-negative numbers. Prove that $\frac{1}{N}\sum_{n=1}^{N} a_n \to$

0 iff there is a set of zero density $\mathscr{N} \subset \mathbb{N}$ s.t. $a_n \xrightarrow[\mathscr{N} \not\ni n \to \infty]{} 0$. Guidance: Fill in the details in the following argument.

1. Suppose $\mathscr{N} \subset \mathbb{N}$ has density zero and $a_n \xrightarrow[\mathscr{N} \not\ni n \to \infty]{} 0$, then $\frac{1}{N}\sum_{n=1}^{N} a_n \to 0$.
2. Now assume that $\frac{1}{N}\sum_{n=1}^{N} a_n \to 0$.
   a. Show that $\mathscr{N}_m := \{k : a_k > 1/m\}$ form an increasing sequence of sets of density zero.
   b. Fix $\varepsilon_i \downarrow 0$, and choose $k_i \uparrow \infty$ such that if $n > k_i$, then $(1/n)|\mathscr{N}_i \cap [1,n]| < \varepsilon_i$. Show that $\mathscr{N} := \bigcup_i \mathscr{N}_i \cap (k_i, k_{i+1}]$ has density zero.
   c. Show that $a_n \xrightarrow[\mathscr{N} \not\ni n \to \infty]{} 0$.

3.4. Here is a sketch of an alternative proof of proposition 3.4, which avoids natural extensions (B. Parry). Fill in the details.

1. Set $H := L^2$, $V := \bigcap_{n\geq 0} U_T^n(H)$, and $W := H \ominus U_T H := \{g \in H, g \perp U_T H\}$.
   a. $H = V \oplus [(U_T H)^\perp + (U_T^2)^\perp + \cdots]$
   b. $\{U_T^k H\}$ is decreasing, $\{(U_T^k H)^\perp\}$ us increasing.
   c. $H = V \oplus \bigoplus_{k=1}^{\infty} U_T^k W$ (orthogonal space decomposition).
2. $U_T : V \to V$ has a bounded inverse (hint: use the fact from Banach space theory that any bounded linear operator between mapping one Banach space *onto* another Banach space which is one-to-one, has a bounded inverse).
3. (3.1) holds for any $f, g \in V$.
4. if $g \in U_T^k W$ for some k, then (3.1) holds for all $f \in L^2$.
5. if $g \in V$, but $f \in U_T^k W$ for some k, then (3.1) holds for f, g.
6. (3.1) holds for all $f, g \in L^2$.

3.5. Show that every invertible ppt with countable Lebesgue spectrum is mixing, whence ergodic.

3.6. Suppose $(X, \mathscr{B}, \mu, T)$ has countable Lebesgue spectrum. Show that $\{f \in L^2 : \int f = 0\}$ is spanned by functions f whose spectral measures ν_f are equal to the Lebesgue measure on S^1.

3.7. Show that any two ppt with countable Lebesgue spectrum are spectrally isomorphic.

3.8. Cutting and Stacking and Chacon's Example
This is an example of a ppt which is weak mixing but not mixing. The example is a certain map of the unit interval, which preserves Lebesgue's measure. It is constructed using the method of "cutting and stacking" which we now explain.

Let $A_0 = [1, \frac{2}{3})$ and $R_0 := [\frac{2}{3}, 1]$ (thought of as reservoir).

Step 1: Divide A_0 into three equal subintervals of length $\frac{2}{9}$. Cut a subinterval B_0 of length $\frac{2}{9}$ from the left end of the reservoir.

- Stack the three thirds of A_0 one on top of the other, starting from the left and moving to the right.
- Stick B_0 between the second and third interval.
- Define a partial map f_1 by moving points vertically in the stack. The map is defined everywhere except on $R \setminus B_0$ and the top floor of the stack. It can be viewed as a partially defined map of the unit interval.

Update the reservoir: $R_1 := R \setminus B_0$. Let A_1 be the base of the new stack (equal to the rightmost third of A_0).

Step 2: Cut the stack vertically into three equal stacks. The base of each of these thirds has length $\frac{1}{3} \times \frac{2}{9}$. Cut an interval B_1 of length $\frac{1}{3} \times \frac{2}{9}$ from the left side of the reservoir R_1.

- Stack the three stacks one on top of the other, starting from the left and moving to the right.
- Stick B_1 between the second stack and the third stack.
- Define a partial map f_2 by moving points vertically in the stack. This map is defined everywhere except the union of the top floor floor and $R_1 \setminus B_1$.

Update the reservoir: $R_2 := R_1 \setminus B_1$. Let A_2 be the base of the new stack (equal to the rightmost third of A_1).

Step 3: Cut the stack vertically into three equal stacks. The base of each of these thirds has length $\frac{1}{3^2} \times \frac{2}{9}$. Cut an interval B_2 of length $\frac{1}{3^2} \times \frac{2}{9}$ from the left side of the reservoir R_2.

- Stack the three stacks one on top of the other, starting from the left and moving to the right.
- Stick B_2 between the second stack and the third stack.
- Define a partial map f_3 by moving points vertically in the stack. This map is defined everywhere except the union of the top floor floor and $R_2 \setminus B_2$.

Update the reservoir: $R_3 := R_2 \setminus B_2$. Let A_3 be the base of the new stack (equal to the rightmost third of A_2)

Continue in this manner, to obtain a sequence of partially defined maps f_n. There is a canonical way of viewing the intervals composing the stacks as of subintervals of the unit interval. Using this identification, we may view f_n as partially defined maps of the unit interval.

1. Show that f_n is measure preserving where it is defined (the measure is Lebesgue's measure). Calculate the Lebesgue measure of the domain of f_n.
2. Show that f_{n+1} extends f_n (i.e. the maps agree on the intersection of their domains). Deduce that the common extension of f_n defines an invertible probability preserving map of the open unit interval. This is Chacon's example. Denote it by $(I, \mathscr{B}, m, T)$.
3. Let ℓ_n denote the height of the stack at step n. Show that the sets $\{T^i(A_n) : i = 0, \ldots, \ell_n,\ n \geq 1\}$ generate the Borel σ–algebra of the unit interval.

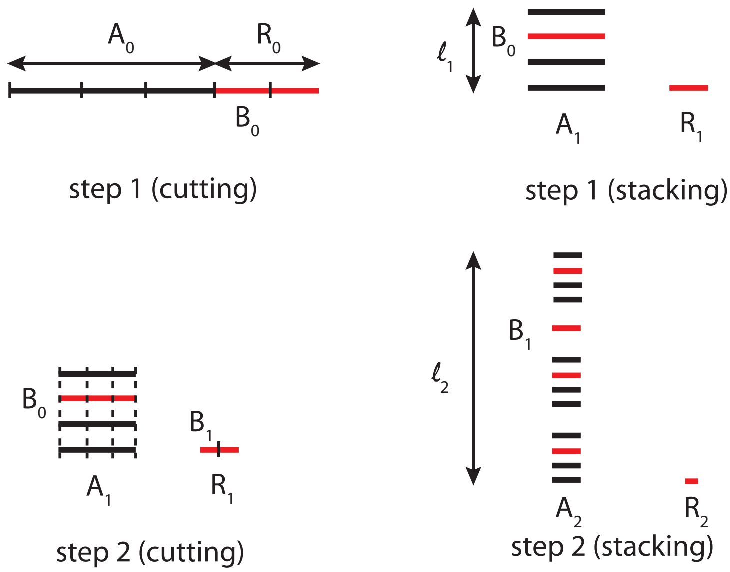

Fig. 3.1 The construction of Chacon's example

3.9. (Continuation) Prove that Chacon's example is weak mixing using the following steps. Suppose f is an eigenfunction with eigenvalue λ.

1. We first show that if f is constant on A_n for some n, then f is constant everywhere. (A_n is the base of the stack at step n.)
   a. Let ℓ_k denote the height of the stack at step k. Show that $A_{n+1} \subset A_n$, and $T^{\ell_n}(A_{n+1}) \subset A_n$. Deduce that $\lambda^{\ell_n} = 1$.
   b. Prove that $\lambda^{\ell_{n+1}} = 1$. Find a recursive formula for ℓ_n. Deduce that $\lambda = 1$.
   c. The previous steps show that f is an invariant function. Show that any invariant function which constant on A_n is constant almost everywhere.
2. We now consider the case of a general L^2– eigenfunction.
   a. Show, using Lusin's theorem, that there exists an n such that f is nearly constant on most of A_n. (Hint: part 3 of the previous question).
   b. Modify the argument done above to show that any L^2–eigenfunction is constant almost everywhere.

3.10. (Continuation) Prove that Chacon's example is not mixing, using the following steps.

1. Inspect the image of the top floor of the stack at step n, and show that for every n and $0 \le k \le \ell_{n-1}$, $m(T^k A_n \cap T^{k+\ell_n} A_n) \ge \frac{1}{3} m(T^k A_n)$.

2. Use problem 3.8 part 3 and an approximation argument to show that for every Borel set E and $\varepsilon > 0$, $m(E \cap T^{\ell_n} E) \geq \frac{1}{3} m(E) - \varepsilon$ for all n. Deduce that T cannot be mixing.

Notes to chapter 3

The spectral approach to ergodic theory is due to von Neumann. For a thorough modern introduction to the theory, see Nadkarni's book [1]. Our exposition follows in parts the books by Parry [2] and Petersen [1]. A proof of the discrete spectrum theorem mentioned in the text can be found in Walters' book [3]. A proof of Herglotz's theorem is given in [2].

References

1. Nadkarni, M. G.: *Spectral theory of dynamical systems.* Birkhäuser Advanced Texts: Birkhäuser Verlag, Basel, 1998. x+182 pp.
2. Parry, W.: *Topics in ergodic theory.* Cambridge Tracts in Mathematics, 75. Cambridge University Press, Cambridge-New York, 1981. x+110 pp.
3. Petersen, K.: *Ergodic theory.* Corrected reprint of the 1983 original. Cambridge Studies in Advanced Mathematics **2** Cambridge University Press, Cambridge, 1989. xii+329 pp.
4. Walters, P.: *An introduction to ergodic theory.* Graduate Texts in Mathematics, **79** Springer-Verlag, New York-Berlin, 1982. ix+250 pp.