June 29, 2020

## 1. Poincaré Recurrence Theorem

Let $(X, \mathcal{B}, \mu)$ be a measure space. A measurable map $T : X \mapsto X$ is called *measure preserving* if $\mu(T^{-1}(A)) = \mu(A)\ \forall A \in \mathcal{B}$.

**Theorem 1.1.** *(Poincaré Recurrence Theorem, 1890) Let $T$ be a measure preserving map on a finite measure space $(X, \mu)$ and $E \subseteq X$ measurable with $\mu(E) > 0$, then for $\mu$-a.a. $x \in E$ $T^n x \in E$ for infinitely many $n \in \mathbb{N}$.*

*Proof.* For $N \geq 0$ set $E_N := \cup_{n=N}^{\infty} T^{-n}(E)$. Then $\cap_{N=0}^{\infty} E_N$ is the set of points in $X$, for which the sequence $x, Tx, T^2x, \ldots$ has infinitely many elements in $E$. We want to show $\mu(F) = \mu(E)$ for $F := E \bigcap \cap_{N=0}^{\infty} E_N$.

We have $T^{-1}(E_N) = E_{N+1}$, so $\mu(E_N) = \mu(E_{N+1})$ and $\mu(E_N) = \mu(E_0)\,\forall N \in \mathbb{N}$. Since $E_0 \supset E_1 \supset E_2 \supset \ldots$ we get $\mu(\cap_{N=0}^{\infty} E_N) = \mu(E_0)$ and as $E \subset E_0$ we see that $\mu(F) = \mu(E \cap E_0) = \mu(E)$.

□

## 2. Ergodic Theorems

{utiso}

**Lemma 2.1.** *If $(X, \mathcal{B}, \mu)$, $1 \leq p$ is a measure space and $T : X \to X$ a measure preserving map, then the operator $U_T : L^p(X, \mu) \mapsto L^p(X, \mu)$, $U_T f = f \circ T$ is an isometry. If $T$ is invertible, then $U_T$ is unitary.*

*Proof.* For a simple function $f_n$ we have $\int_X f_n\, d\mu = \int_X U_T f_n\, d\mu$ since $T$ is measure preserving. Assume $f$ is positive and integrable and $(f_n)$ a monotone sequence of simple functions converging to $f$ from below, then by Lebesgue convergence theorem

$$\int_X U_T f\, d\mu = \lim_{n\to\infty} \int_X U_T f_n\, d\mu = \lim_{n\to\infty} \int_X f_n\, d\mu = \int_X f\, d\mu.$$

□

**Theorem 2.2.** *(Mean Ergodic Theorem, v. Neumann 1932) Let $T$ a measure preserving map on a measure space $(X, \mathcal{B}, \mu)$ and $f \in L^2(X)$. Then there is $\bar{f} \in L^2(X)$ with*

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k - \bar{f} \right\|_2 = 0.$$

*Proof.* Let $\mathcal{M} := \{f : U_T f = f\}$ be the fixed point set of the isometry $U_T$. Then $\mathcal{M}$ is a closed linear subspace of $L^2(X, \mu)$.

We claim that $\mathcal{M} = \mathcal{N}^{\perp}$ for $\mathcal{N} := \{g - U_T g : g \in L^2(X, \mu)\}$.

We have

$$\begin{aligned} h \in \mathcal{N}^{\perp} \quad &\Leftrightarrow \quad (g - U_T g, h) = 0 \quad \forall g \\ &\Leftrightarrow \quad (g, h - U_T^* h) = 0 \quad \forall g \Leftrightarrow h = U_T^* h. \end{aligned}$$

Since $\|U_T\| = 1$ we have $\|U_T^*\| = 1$, so $h = U_T^* h$ gives

$$\|h\|^2 = (h, U_T^* h) = (U_T h, h),$$

which by virtue of $\|U_T h\| \leq \|h\|$ gives $U_T h = h$, so $\mathcal{M} \supset \mathcal{N}^{\perp}$.

As above $f - U_T f = 0$ gives $\|f\|^2 = (f, U_T^* f) = (U_T f, f)$ and $U_T^* f = f$, so

$$f \in \mathcal{M} \Rightarrow (f - U_T^* f, g) = 0 \quad \forall g \Rightarrow (f, g - U_T g) = 0 \quad \forall g \Rightarrow f \in \mathcal{N}^{\perp}$$

and we have shown $\mathcal{M} = \mathcal{N}^{\perp}$.

For $f_1 \in \mathcal{N}$ we have

$$\frac{1}{n} \sum_{k=0}^{n-1} U_T^k f_1 = \frac{1}{n} \sum_{k=0}^{n-1} (U_T^k g - U_T^{k+1} g) = \frac{1}{n}(g - U_T^n g) \to 0.$$

Since $\frac{1}{n} \sum_{k=0}^{n-1} U_T^k$ is a contraction it follows that this limit vanishes for $f_1$ in the closure $\overline{\mathcal{N}}$.

For $f_2 \in \mathcal{M}$ we have $\frac{1}{n} \sum_{k=0}^{n-1} U_T^k f_2 = f_2$.

For $f \in L^2(X, \mu)$ let $f = f_1 + f_2$ with $f_1 \in \overline{\mathcal{N}}$ and $f_2 \in \mathcal{M}$. Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} U_T^k f = f_2,$$

so the claim follows with $\bar{f} = f_2 = Pf$, where $P$ is the orthogonal projector on $\mathcal{M}$.

□

{maxerg} **Theorem 2.3.** *(Maximal Ergodic Theorem, Wiener, Yoshida & Kakutani 1939) Let $(X, \mu)$ be a finite measure space and $T : X \mapsto X$ a measure preserving map. For $f \in L^1(X, \mu)$ we have*

$$\int_{\{x \colon f^*(x) > 0\}} f \, d\mu \geq 0 \text{ with } f^*(x) := \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x).$$

*Proof.* (Garsia) Set $f_0 = 0$, $f_n := f + U_T f + U_T^2 f + \ldots + U_T^{n-1} f$ and $F_N := \max\{f_n : 0 \leq n \leq N\}$. Since $F_N \geq f_0$ we have $F_N \geq 0$. Furthermore $F_N \in L^1(X, \mu)$, $F_N \geq f_n$, so by positivity of $U_T$ we get $U_T F_N \geq U_T f_n$ and $U_T F_N + f \geq f_{n+1}$ for $0 \leq n \leq N$.

It follows that

$$U_T F_N(x) + f(x) \geq \max_{1 \leq n \leq N} f_n(x)$$

and for $x \in A_N := \{x : F_N(x) > 0\}$:

$$U_T F_N(x) + f(x) \geq \max_{0 \leq n \leq N} f_n(x) = F_N(x),$$

so $f \geq F_N - U_T F_N$ auf $A_N$. We get

$$\begin{aligned}
\int_{A_N} f\, d\mu &\geq \int_{A_N} F_N\, d\mu - \int_{A_N} U_T F_N\, d\mu \\
&= \int_X F_N\, d\mu - \int_{A_N} U_T F_N\, d\mu \qquad (F_N = 0 \text{ on } X \setminus A_N) \\
&\geq \int_X F_N\, d\mu - \int_X U_T F_N\, d\mu \qquad (F_N \geq 0) \\
&= 0 \qquad \text{since } U_T \text{ is an isometry.}
\end{aligned}$$

As $A_N$ is increasing with $\cup_N A_N = \{x : f^*(x) > 0\}$ the claim follows. □

{cmet}

**Corollary 2.4.** *Let $T$ be a measure preserving map on a measure space $(X, \mu)$ and $g \in L^1(X)$. For*

$$B_\alpha := \left\{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i(x)) > \alpha \right\}$$

*and $A \subset X$ with $T^{-1}A = A$ and $\mu(A) < \infty$*

$$\int_{B_\alpha \cap A} g\, d\mu \geq \alpha \mu(B_\alpha \cap A)$$

*obtains.*

*Proof.* For $\mu(X) < \infty$, $A = X$ and $f := g - \alpha$, $B_\alpha := \cup_{N=0}^\infty \{x : F_N(x) > 0\}$, so by the Maximal Ergodic Theorem 2.3 $\int_{B_\alpha} f\, d\mu > 0$. It follows that $\int_{B_\alpha} g\, d\mu \geq \alpha \mu(B_\alpha)$.

The general case $A \subset X$ follows by considering the restriction of $T$ to $A$, since by assumption $T$ maps $A$ to itself. □

{inderg}

**Theorem 2.5.** *(Individual Ergodic Theorem, Birkhoff 1931) Let $T$ be a measure preserving map on a measure space $(X, \mu)$ and $f$ a $\mu$-integrable function on $X$. Then $\frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$ converges for almost all $x$ to a function $\bar{f} \in L^1(X, \mu)$ and $\bar{f} \circ T = \bar{f}$ almost everywhere.*

*If $\mu$ is finite $\int \bar{f}\, d\mu = \int f\, d\mu$ obtains.*

*Proof.* Assume $f$ real. Let

$$\bar{f} := \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \quad \text{and } \underline{f} := \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)),$$

then $\bar{f} \circ T = \bar{f}$ and $\underline{f} \circ T = \underline{f}$. It remains to show that $\bar{f} = \underline{f} \in L^1(X, \mu)$.

For $\alpha, \beta \in \mathbb{R}$ set

$$E_{\alpha,\beta} := \{x \in X : \alpha < \bar{f}(x),\ \underline{f}(x) < \beta\}$$

then $E_{\alpha,\beta}$ is a $T$-invariant subset of $X$ and

$$\{x \in X : \underline{f}(x) < \bar{f}(x)\} = \bigcup_{\beta < \alpha \in \mathbb{Q}} E_{\alpha,\beta}.$$

We have for

$$B_\alpha := \left\{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) > \alpha \right\}$$

$E_{\alpha,\beta} \cap B_\alpha = E_{\alpha,\beta}$, so by Corollary 2.4

$$\int_{E_{\alpha,\beta}} f \, d\mu = \int_{E_{\alpha,\beta} \cap B_\alpha} f \, d\mu \geq \alpha \mu(E_{\alpha,\beta} \cap B_\alpha) = \alpha \mu(E_{\alpha,\beta})$$

Replacing $f, \alpha, \beta$ by $-f, -\beta, -\alpha$ in the above computation gives $\overline{-f} = -\underline{f}$; $\underline{-f} = -\bar{f}$ and

$$\int_{E_{\alpha,\beta}} f \, d\mu \leq \beta \mu(E_{\alpha,\beta}).$$

We now have $\alpha \mu(E_{\alpha,\beta}) \leq \beta \mu(E_{\alpha,\beta})$, so for $\beta < \alpha$ we have $\mu(E_{\alpha,\beta}) = 0$ and $\bar{f} = \underline{f}$ $\mu$-almost everywhere.

We next show that $\bar{f} \in L^1(X, \mu)$: Let

$$g_n := \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \right| \leq \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))|,$$

then $\int g_n \, d\mu \leq \int |f| \, d\mu$ and by Fatou's Lemma and Lemma 2.1

$$\int |\bar{f}| \, d\mu = \int \lim g_n \, d\mu \leq \liminf \int g_n \, d\mu \leq \int |f| \, d\mu.$$

Finally we show $\int f \, d\mu = \int \bar{f} \, d\mu$ for $\mu(X) < \infty$:

Set $D_k^n := \left\{ x \in X : \frac{k}{n} \leq \bar{f}(x) < \frac{k+1}{n} \right\}$ for $k \in \mathbb{Z}$, $n \geq 1$. Applying Corollary 2.4 and observing that for $\epsilon > 0$ we have $D_k^n \cap B_{\frac{k}{n} - \epsilon} = D_k^n$ we obtain $\int_{D_k^n} f \, d\mu \geq \left(\frac{k}{n} - \epsilon\right) \mu(D_k^n)$, so $\int_{D_k^n} f \, d\mu \geq \frac{k}{n} \mu(D_k^n)$ and

$$\int_{D_k^n} \bar{f} \, d\mu \leq \frac{k+1}{n} \mu(D_k^n) \leq \frac{1}{n} \mu(D_k^n) + \int_{D_k^n} f \, d\mu.$$

It follows that $\int_X \bar{f} \, d\mu \leq \frac{1}{n} \mu(X) + \int_X f \, d\mu$ and therefore $\int_X \bar{f} \, d\mu \leq \int_X f \, d\mu$.

Taking $-f$ instead of $f$ in the above computation we get $\int_X \overline{-f}\, d\mu \leq -\int_X f\, d\mu$, so $\int_X \underline{f}\, d\mu \geq \int_X f\, d\mu$. Since $\bar{f} = \underline{f}$ almost everywhere we get

$$\int_X \bar{f}\, d\mu = \int_X f\, d\mu.$$

□

**Theorem 2.6.** *($L^p$-Ergodic Theorem, V. Neumann) For $1 \leq p < \infty$ and $T$ a measure preserving map on a probability space $(X, \mu)$ and $f \in L^p(X)$ $\frac{1}{n}\sum_{k=0}^{n-1} U_T^k f$ converges in $L^p$-norm to a function $\bar{f} \in L^p$ as $n \to \infty$ with $U_T \bar{f} = \bar{f}$.*

*Proof.* For $f$ bounded and measurable we have by the individual ergodic theorem 2.5 $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \bar{f}(x)$ almost everywhere with $\bar{f} \in L^\infty(X)$. By Lebesgue dominated convergence theorem we get

$$\left\| \frac{1}{n}\sum_{k=0}^{n-1} U_T^k f - \bar{f} \right\|_p \to 0, \quad \bar{f} \in L^\infty(X, \mu).$$

It follows from the individual ergodic theorem 2.5that $\bar{f} \circ T = \bar{f}$.

If $f$ is unbounded and $\epsilon > 0$ we find a bounded function $f_\epsilon$ with $\|f - f_\epsilon\|_p < \epsilon$. Since the operators $\frac{1}{n}\sum_{k=0}^{n-1} U_T^k$ are contractions $\bar{f}_\epsilon$ converges as $\epsilon \to 0$ to some $\bar{f}$ in $L^p$-norm. It follows that

$$\left\| \frac{1}{n}\sum_{k=0}^{n-1} U_T^k f - \bar{f} \right\|_p \leq \left\| \frac{1}{n}\sum_{k=0}^{n-1} U_T^k f_\epsilon - \bar{f}_\epsilon \right\|_p < 2\epsilon$$

for sufficiently large $n$, so $f$ converges to $\bar{f}$ in $L^p$-norm. □

## 3. Ergodic Maps

On a probability space $(X, \mathcal{B}, \mu)$ a measure preserving map $T$ which satisfies $\mu(B) \in \{0,1\}$ for all $B \in \mathcal{B}$ with $\mathrm{T}^{-1}(B) = B$ is *ergodic*. So ergodic maps allow no measure theoretical non trivial decomposition in $T$-invariant subsets.

{er1}

**Theorem 3.1.** *The following are equivalent:*

i) *$T$ is ergodic;*
ii) *$B \in \mathcal{B}$ with $\mu(\mathrm{T}^{-1}(B) \triangle B) = 0$ implies $\mu(B) \in \{0,1\}$;*
iii) *$A \in \mathcal{B}$ with $\mu(A) > 0$ implies $\mu\left(\cup_{n=1}^{\infty} T^{-n}A\right) = 1$;*
iv) *For $A, B \in \mathcal{B}$ with $\mu(A) > 0, \mu(B) > 0$ there is $n \in \mathbb{N}$ with $\mu(T^{-n}A \cap B) > 0$;*
v) *For $A, B \in \mathcal{B}$ $\frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) \to \mu(A)\mu(B)$.*

*Proof.* i)$\Rightarrow$ ii): Assume $B \in \mathcal{B}$ with $\mu(\mathrm{T}^{-1}(B) \triangle B) = 0$. For $n \geq 0$

$$T^{-n}B \triangle B \subset \cup_{i=0}^{n-1} T^{-i-1}B \triangle T^{-i}B = \cup_{i=0}^{n-1} T^{-i}(\mathrm{T}^{-1}(B) \triangle B),$$

so $\mu(T^{-n}B \triangle B) \leq n\mu(\mathrm{T}^{-1}(B) \triangle B) = 0$. This yields

$$\mu(\cup_{i=n}^{\infty} T^{-i}B \triangle B) \leq \sum_{i=n}^{\infty} \mu(T^{-i}(B) \triangle B) = 0$$

and since $\cup_{i=n}^{\infty} T^{-i}B$ is decreasing in $n$ with $\mu(\cup_{i=n}^{\infty} T^{-i}B) = \mu(B)$ we obtain for $B_\infty := \cap_{n=0}^{\infty} \cup_{i=n}^{\infty} T^{-i}B$: $\mu(B_\infty \triangle B) = 0$ and therefore $\mu(B_\infty) = \mu(B)$. We see that

$$\mathrm{T}^{-1}(B_\infty) = \cap_{n=0}^{\infty} \cup_{i=n}^{\infty} T^{-i-1}B = \cap_{n=0}^{\infty} \cup_{i=n+1}^{\infty} T^{-i}B = B_\infty.$$

Since $T$ is ergodic and $\mu(B_\infty) = \mu(B)$ we obtain $\mu(B) = \mu(B_\infty) \in \{0,1\}$.

ii)$\Rightarrow$ iii): Assume $\mu(A) > 0$ and set $A_1 := \cup_{i=1}^{\infty} T^{-i}A$. We have $\mathrm{T}^{-1}(A_1) \subset A_1$ and since $T$ is measure preserving $\mu(\mathrm{T}^{-1}(A_1) \triangle A_1) = 0$. By assumption $\mu(A_1) \in \{0,1\}$ but $\mathrm{T}^{-1}(A) \subset A_1$ and $\mu(\mathrm{T}^{-1}(A)) = \mu(A) > 0$, so $\mu(A_1) = 1$.

iii)$\Rightarrow$ iv): By assumption

$$0 < \mu(B) = \mu(B \cap \cup_{i=1}^{\infty} T^{-i}(A)) = \mu(\cup_{i=1}^{\infty} B \cap T^{-i}(A)),$$

so $\mu(B \cap T^{-n}(A)) > 0$ for some $n$.

iv)$\Rightarrow$ i): For $B \in \mathcal{B}$ with $\mathrm{T}^{-1}(B) = B$ and $0 < \mu(B) < 1$ we get $0 = \mu(B \cap (X \setminus B)) = \mu(T^{-n}(B) \cap X \setminus B)$ for all $n$ contradicting iv).

ii)$\Rightarrow$ v): For $f := \chi_A$ the individual ergodic theorem 2.5 implies

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x) \to \bar{f} \text{ with } \bar{f} = \bar{f} \circ T \text{ a.e.}$$

Applying ii) to the sets $B := \{x : \bar{f}(x) \in [a,b]\}$ we see that $\bar{f}$ is constant a.e., so $\bar{f}(x) = \int \bar{f}\, d\mu = \int f\, d\mu = \mu(A)$ a.e., so

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i x)\chi_B \to \mu(A)\chi_B \text{ for } \mu \text{ a.a. } x$$

and

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

by Lebesgue convergence theorem.

[v)⇒ i)]: For $E \in \mathcal{B}$ with $T^{-1} E = E$ we apply v) to $A = B = E$, then

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu(E) \to \mu(E)^2$$

so $\mu(E) = \mu(E)^2$ and $\mu(E) \in \{0,1\}$.

□

{er2}

**Theorem 3.2.** *Let $(X, \mathcal{B}, \mu)$ be a probability space with a measure preserving map $T$. Then the following are equivalent:*

i) *$T$ is ergodic;*
ii) *A measurable function $f$ on $X$ with $f \circ T(x) = f(x)\, \forall x \in X$ is constant almost everywhere;*
iii) *A measurable function $f$ on $X$ with $f \circ T(x) = f(x)$ for almost all $x \in X$ is constant almost everywhere;*
iv) *A measurable square integrable function $f$ on $X$ with $f \circ T(x) = f(x)\, \forall x \in X$ is constant almost everywhere;*
v) *A measurable square integrable function $f$ on $X$ with $f \circ T(x) = f(x)$ for almost all $x \in X$ is constant almost everywhere;*

*Proof.* Clearly iii)⇒ii)⇒iv) and iii)⇒v)⇒iv). We show i)⇒iii) and iv)⇒i).

i)⇒iii): Assume $T$ is ergodic and $f$ measurable with $f = f \circ T$ almost everywhere. For

$$A_{k,n} := \left\{x : \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\right\}$$

we have $T^{-1} A_{k,n} \triangle A_{k,n} \subset \{x : f \circ T(x) \neq f(x)\}$. It follows that $\mu(T^{-1} A_{k,n} \triangle A_{k,n}) = 0$, so by Theorem 3.1 $\mu(A_{k,n}) \in \{0,1\}$. We see that for all $n$ there is $k(n)$ with $\mu(A_{k(n),n}) = 1$. So $\mu(\cap_n A_{k(n),n}) = 1$ and $f$ is constant on $\cap_n A_{k(n),n}$, so $f$ is constant almost everywhere.

iv)$\Rightarrow$i): Let $\mathrm{T}^{-1} E = E$ for some $E \in \mathcal{B}$. Then $\chi_E$ is square integrable with $\chi_E \circ T(x) = \chi_E(x)$ for all $x$, so by assumption $\chi_E = 0$ a.e. or $\chi_E = 1$ a.e. and it follows that $\mu(E) \in \{0, 1\}$.

□

**Corollary 3.3.** *For $f \in L^1(X)$ and $T$ ergodic we have*

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \to \int_X f \, d\mu \ a.e..$$

*Proof.* By the individual ergodic theorem 2.5 the sums converge to $\bar{f} = \bar{f} \circ T$ a.e. but $\int_X f \, d\mu = \int_X \bar{f} \, d\mu = \bar{f}$ a.e.. □

**Example 3.4.** *Consider the map $T : [0, 1) \mapsto [0, 1), x \mapsto 2x \bmod 1$ with Lebesgue measure $\lambda$ on $[0, 1)$. $T$ is measure preserving since $\lambda(\mathrm{T}^{-1}((a, b))) = \lambda((a, b))$ for all intervals $(a, b)$.*

*For $f \in L^2([0, 1))$ with Fourier expansion $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{i2\pi nx}$ we have $f \circ T(x) = \sum_{n \in \mathbb{Z}} b_n e^{i2\pi nx}$ with $b_{2n} = a_n$ and $b_{2n+1} = 0$ $\forall n \in \mathbb{N}$. Therefore $f = f \circ T$ a.e. implies $a_n = 0$ for $n \neq 0$, so $f$ is constant a.e. and by Theorem 3.2 $T$ is ergodic.*

As an application we get:

**Theorem 3.5.** *(Borel Theorem on Normal Numbers, Borel 1909) Almost all (w.r.t. Lebesgue measure) numbers in $[0, 1)$ are normal in basis 2, that is in the binary representation of almost all numbers the digit 1 appears with asymptotic density 1/2.*

*Proof.* Let $T$ be the ergodic map $x \mapsto 2x$ as in the example above. Allmost all numbers in $[0, 1)$ have a unique binary representation. If the binary representation of some $x \in [0, 1)$ is $0, a_1 a_2 \ldots$, with $a_i \in \{0, 1\}$, then $Tx$ has binary representation $Tx = 0, a_2 a_3 \ldots$. We have $\chi_{[1/2,1)}(x) = a_1$, so the asymptotic density of the digit 1 is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_{[1/2,1)}(T^i x),$$

which since $T$ is ergodic is just $\int_{[0,1)} \chi_{[1/2,1)}(x) \, d\lambda(x) = 1/2$. □

**Example 3.6.** *Irrational rotations of the torus. Consider the map $Tx = x + \alpha \bmod 1$ for some $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. We show that $T$ is ergodic:*

*For $f \in L^2(\mathbb{T})$ with Fourier expansion $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi inx}$, we have*

$$U_T f(x) = f(x + \alpha) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi in\alpha} e^{2\pi inx}.$$

*Now $U_T f(x) = f(x)$ gives $a_n = a_n e^{2\pi in\alpha}$, therefore $a_n = 0$ for $n \neq 0$ and $f(x) = a_0$, so by Theorem 3.2 $T$ is ergodic.*

## 4. Mixing maps and spectral properties

By Theorem 3.1 we have for ergodic maps $T$:

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

$\forall A,B\in\mathcal{B}$, which motivates the following definitions: A measure preserving map $T$ is *weak mixing* if $\frac{1}{n}\sum_{k=0}^{n-1}|\mu(T^{-k}A\cap B)-\mu(A)\mu(B)|\to 0$ $\forall A,B\in\mathcal{B}$. A measure preserving map $T$ is *strong mixing* if $|\mu(T^{-k}A\cap B)-\mu(A)\mu(B)|\to 0$ $\forall A,B\in\mathcal{B}$.

Clearly strong mixing implies weak mixing, which implies ergodicity.

{smix}

**Theorem 4.1.** *$T$ is strong mixing iff for all $f,g\in L^2(X)$ $(U_T^n f,g)\to(f,1)(1,g)$ as $n\to\infty$.*

Therefore $T$ is strong mixing iff $U_T^n f\to\int_X f\,d\mu$ in the weak topology of $L^2(X)$.

*Proof.* If $T$ is strong mixing the asserted convergence holds for characteristic functions, therefore for simple functions, and since simple functions are dense in $L^2(X)$ for all functions.

The converse is obvious by considering characteristic functions. □

{ersiev}

**Theorem 4.2.** *$T$ is ergodic iff 1 is a simple eigenvalue of $U_T\in L(L^2(X))$.*

*Proof.* Constant functions are $U_T$-invariant, so 1 is always an eigenvalue of $U_T$.

If 1 is a simple eigenvalue, then for $B=\mathrm{T}^{-1}(B)$ we have $U_T\chi_B=\chi_B$, so $\chi_B$ is constant $\mu$-a.e., so $T$ is ergodic.

The converse implication is an immediate consequence of Theorem 3.2 v). □

**Theorem 4.3.** *If $T$ is ergodic then the eigenvalues of $U_T$ are simple and the set of eigenvalues is a subgroup of $\{z:|z|=1\}$.*

*Proof.* Since $U_T$ is an isometry by Lemma 2.1 all eigenvalues $\lambda$ of $U_T$ satisfy $|\lambda|=1$. So if $f_\lambda$ is eigenfunction with eigenvalue $\lambda$, then

$$U_T|f_\lambda|=|U_Tf_\lambda|=|\lambda f_\lambda|=|f_\lambda|$$

and $|f_\lambda|$ is constant a.e. by Theorem 4.2.

If $f_\nu$ is an eigenfunction with eigenvalue $\nu$ we have, since $f_\lambda\neq 0$ a.e.,

$$U_T\left(\frac{f_\nu}{f_\lambda}\right)=\frac{\nu}{\lambda}\frac{f_\nu}{f_\lambda}$$

a.e. and $f_\nu/f_\lambda$ is an eigenfunction with eigenvalue $\nu/\lambda$. With Theorem 4.2 it follows that the eigenvalues define a subgroup.

Taking $\nu = \lambda$ and applying Theorem 4.2 we see that $f_\nu = cf_\lambda$ a.e., so all eigenvalues are simple. □

Recall that a subset $S$ of $\mathbb{N}$ has *density 0* if

$$\lim_{N\to\infty} \frac{1}{N}(\#\{n \in S : n \le N\}) = 0,$$

where $\#A$ denotes the cardinality of the set $A$.

{wmix} **Theorem 4.4.** *For a measure preserving map $T$ the following are equivalent:*

i) *$T$ is weak mixing;*
ii) $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |(U_T^k f, g) - (f,1)(1,g)| = 0\, \forall f, g \in L^2(X)$;
iii) *For $A, B \in \mathcal{S}$, $\mathcal{S}$ being a semiring generating $\mathcal{B}$, there is $E \subset \mathbb{N}$ with density 0 and*

$$\lim_{\binom{n\to\infty}{n\notin E}} \mu(T^{-n}A \cap B) = \mu(A)\mu(B);$$

iv) *$T \times T$ is weak mixing on $X \times X$;*
v) *$T \times S$ is ergodic on $X \times Y$ for all ergodic transformations $S$ on $Y$;*
vi) *$T \times T$ is ergodic;*

We first need:

{kvn} **Lemma 4.5.** *(Koopman, v. Neumann, 1932) Let $f : \mathbb{N} \to \mathbb{R}_+$ be bounded. Then $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(k) = 0$ iff there is $E \subset \mathbb{N}$ with density 0 such that $\lim_{\binom{n\to\infty}{n\notin E}} f(n) = 0$.*

*Proof.* Since any sequence converging to 0 has arithmetic means converging to 0

$$\lim_{\binom{n\to\infty}{n\notin E}} f(n) = 0 \quad \text{gives} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{\binom{0\le k<n}{n\notin E}} f(k) = 0.$$

Clearly any bounded function $f$ satisfies $\lim_{n\to\infty} \frac{1}{n}\sum_{\binom{0\le k<n}{n\in E}} f(k) = 0$ if $E$ has density 0, so $\lim_{n\to\infty} \frac{1}{n}\sum_{0\le k<n} f(k) = 0$ if $\lim_{\binom{n\to\infty}{n\notin E}} f(n) = 0$.

Conversely assume $\lim_{n\to\infty} \frac{1}{n}\sum_{0\le k<n} f(k) = 0$.

We define a density fuction

$$d(f,n) := \frac{1}{n}\sum_{k=0}^{n-1} f(k), \quad d(A,n) := d(\chi_A, n)$$

and set

$$E_0 := \{n \in \mathbb{N} : 1 \leq f(n)\}, \quad E_i := \{n \in \mathbb{N} : 2^{-i} \leq f(n) < 2^{-i+1}\}.$$

Then for all $i$ $E_i$ has density 0 and therefore $\cup_{i=0}^{L} E_i$ has density 0 for all $L$.

Set $n_L := \max(\{n : d(\cup_{i=0}^{L} E_i, n) > \frac{1}{L}\} \cup \{1\})$. It is easily seen that $n_{L+1} \geq n_L$ and $n_L \to \infty$ as $L \to \infty$ unless $f(n) = 0$ for a.a $n$.

For

$$E := \bigcup_{L \in \mathbb{N}} E_L \cap (n_L, \infty).$$

we have $\frac{1}{L} \geq d(\cup_{i=0}^{L} E_i, n)$ for $n_L < n$ and $d(\cup_{i=0}^{L} E_i, n) \geq d(E, n)$ for $n \leq n_{L+1}$, so for $n_L < n \leq n_{L+1}$ we get $d(E, n) \leq \frac{1}{L}$. Since $n_L$ goes to infinity this shows that $E$ has density 0. $\square$

*Proof.* (Theorem 4.4)

i) $\Leftrightarrow$ ii)As in the proof of Theorem 4.1 we first consider characteristic functions, then simple functions and finally extend by continuity to all of $L^2(X)$.

i) $\Leftrightarrow$ iii) This is a consequence of applying the Koopman - V. Neumann Lemma 4.5 to $f(n) = |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)|$ and observing that if i) holds for $A, B \in \mathcal{S}$ it holds for $A, B$ in the ring $\mathcal{R}$ generated by $\mathcal{S}$ since elements in $\mathcal{R}$ are finite disjoint unions of sets in $\mathcal{S}$ and finally approximating sets $A, B \in \mathcal{B}$ by sets $\tilde{A}, \tilde{B} \in \mathcal{R}$ with $\mu(A \triangle \tilde{A}) < \epsilon$, $\mu(\triangle \tilde{A}) < \epsilon$.

iii) $\Rightarrow$ iv): By the equivalence of i) and iii) it follows that iii) holds for all $A, B \in \mathcal{B}$ if it holds for all $A, B$ in some semiring generating $\mathcal{B}$. So by assumption for $A, B, C, D \in \mathcal{B}$ there are subsets $E_1, E_2$ of $\mathbb{N}$ of density 0 such that

$$\lim_{\binom{n\to\infty}{n\notin E_1}} \mu(T^{-n}A \cap C) = \mu(A)\mu(C), \quad \lim_{\binom{n\to\infty}{n\notin E_2}} \mu(T^{-n}B \cap D) = \mu(B)\mu(D).$$

$E := E_1 \cup E_2$ is a set with density 0 for which we have

$$\begin{aligned}
& \lim_{\binom{n\to\infty}{n\notin E}} |\mu \times \mu((T \times T)^{-n}(A \times B) \cap (C \times D)) \\
& \qquad -\mu \times \mu(A \times B)\mu \times \mu(C \times D)| \\
= \quad & \lim_{\binom{n\to\infty}{n\notin E}} |\mu(T^{-n}A \cap C)\mu(T^{-n}B \cap D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\
\leq \quad & \lim_{\binom{n\to\infty}{n\notin E}} \mu(T^{-n}A \cap C)|\mu(T^{-n}B \cap D) - \mu(B \cap D)| \\
& \qquad +\mu(B)\mu(D)|\mu(T^{-n}A \cap C) - \mu(A)\mu(C)| = 0.
\end{aligned}$$

By the equivalence of i) and iii) this is just ergodicity of $T \times T$ acting on $X \times X$.

iv) $\Rightarrow$ v): If $T \times T$ is weak mixing on $X \times X$ it follows by considering the sets $A \times X$ that $T$ is weak mixing.

As above we see that if

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu \times \nu((T \times S)^{-k}(A \times C) \cap (B \times D)) \to \mu(A)\mu(B)\nu(C)\nu(D)$$

holds for $A, B$ measurable subsets of $X$ and $C, D$ measurable subsets of $Y$.

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu \times \nu((T \times S)^{-k}(\tilde{A}) \cap (\tilde{B})) \to \mu \times \nu(\tilde{A})\mu \times \nu(\tilde{B})$$

holds for all measurable $\tilde{A}, \tilde{B}$ in $X \times Y$. So by Theorem 3.1 we have to show the above limit for rectangles $A \times C$ and $B \times D$ only:

We have

$$\begin{aligned}
&\frac{1}{n}\sum_{k=0}^{n-1} \mu \times \nu[(T \times S)^{-k}(A \times C) \cap (B \times D)] = \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B)\nu(S^{-k}C \cap D) = \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \mu(A)\mu(B)\nu(S^{-k}C \cap D) \\
&\qquad\qquad + \frac{1}{n}\sum_{k=0}^{n-1} \left[\mu(T^{-k}A \cap B) - \mu(A)\mu(B)\right] \nu(S^{-k}C \cap D).
\end{aligned}$$

Since $T$ is weak mixing

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=0}^{n-1} \left[\mu(T^{-k}A \cap B) - \mu(A)\mu(B)\right] \nu(S^{-k}C \cap D)\right| \\
\leq \quad & \frac{1}{n}\sum_{k=0}^{n-1} \left|\mu(T^{-k}A \cap B) - \mu(A)\mu(B)\right| \to 0
\end{aligned}$$

and as $S$ is ergodic

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu(A)\mu(B)\nu(S^{-k}C \cap D) \to \mu(A)\mu(B)\nu(C)\nu(D),$$

so the claim follows.

v) $\Rightarrow$ vi): Applying v) to $(X \times \{1\}, T \times \mathrm{Id})$ we see that $T$ is ergodic and by v) $T \times T$ is ergodic.

vi) $\Rightarrow$ iii) For $A, B \in \mathcal{B}$ we have

$$\begin{aligned}
& \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) - \mu(A)\mu(B))^2 \\
= \quad & \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}A \cap B)^2 - 2\mu(A)\mu(B)\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-k}A \cap B) \\
& \quad +(\mu(A)\mu(B))^2 \\
= \quad & \frac{1}{n}\sum_{k=0}^{n-1} \mu \times \mu((T \times T)^{-k}(A \times A) \cap (B \times B)) \\
& \quad -2\mu(A)\mu(B)\frac{1}{n}\sum_{k=0}^{n-1} \mu \times \mu((T \times T)^{-k}(A \times X) \cap (B \times X)) \\
& \quad +(\mu(A)\mu(B))^2.
\end{aligned}$$

If $T \times T$ is ergodic this converges as $n \to \infty$ to

$$\begin{aligned}
& \mu \times \mu(A \times A)\mu \times \mu(B \times B) \\
& \quad -2\mu(A)\mu(B)\mu \times \mu(A \times X)\mu \times \mu(B \times X) + (\mu(A)\mu(B))^2 \\
= \quad & (\mu(A)\mu(B))^2 - 2\mu(A)^2\mu(B)^2 + \mu(A)^2\mu(B)^2 = 0.
\end{aligned}$$

By Lemma 4.5 $\frac{1}{n}\sum_{i=1}^{n-1} a_i^2$ converges to 0 iff $\frac{1}{n}\sum_{i=1}^{n-1} |a_i|$ converges to 0, so the claim follows. $\square$

**Theorem 4.6.** *If $T$ is weak mixing all square-integrable eigenfunctions of $U_T$ are constant a.e. If $T$ is measure preserving and invertible with the constant functions being the only square-integrable eigenfunctions of $U_T$, then $T$ is weak mixing.*

*Proof.* Assume $f \in L^2(X)$, $U_T f = \lambda f$, $|\lambda| = 1$, then for $g(x,y) := f(x)\overline{f(y)}$ we have

$$T \times Tg(x,y) = g(Tx,Ty) = f(Tx)\overline{f(Ty)} = \lambda\bar{\lambda}g(x,y) = g(x,y),$$

so $g$ is $T \times T$-invariant. As $T \times T$ is ergodic by Theorem 4.4 we have $g$ is constant $\mu \times \mu$-a.e., so $f$ is constant $\mu$-a.e..

Let $V$ be the closed linear hull of the eigenfunctions of $U_T$. For $f \in V^\perp$, $g \in L^2(X)$ the function $\lambda \mapsto (E_\lambda f, g)$ is continuous in $\lambda$ ($(E_\lambda)_{\lambda\in\sigma}$ denoting the partition of unity w.r.t. the operator $U_T$ as given by the spectral theorem for unitary operators and $\sigma$ the spectrum of $U_T$). So $\int_D d(E_\lambda f, g)\, d(E_\lambda f, g) = 0$ if $D$ is the diagonal $\{(\lambda, \lambda) :\ \lambda \in \sigma\}$ in $\sigma \times \sigma$.

We have if $\mu$ denotes the measure corresponding to the function $\lambda \mapsto E_\lambda f, g)$ of bounded variation

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}|(U_T^k f,g)|^2 =& \frac{1}{n}\sum_{k=0}^{n-1}\left|\int_{\sigma(U_T)} \lambda^k \, d(E_\lambda f, g)\right|^2 \\ =& \frac{1}{n}\sum_{k=0}^{n-1}\left|\int_{\sigma(U_T)} \lambda^k \, d\mu(\lambda)\right|^2 \\ =& \iint_{\sigma\times\sigma} \frac{1}{n}\sum_{k=0}^{n-1} \lambda^k \overline{\zeta^k} \, d\mu(\lambda) d\bar{\mu}(\zeta) \\ =& \iint_{\sigma\times\sigma\setminus D} \frac{1}{n}\frac{1-(\lambda\bar{\zeta})^n}{1-\lambda\bar{\zeta}} \, d\mu(\lambda) d\bar{\mu}(\zeta)\end{aligned}$$

and by Lebesgue convergence theorem ($\left|\frac{1-(\lambda\bar{\zeta})^n}{1-\lambda\bar{\zeta}}\right| = \left|\sum_{i=0}^{n-1}(\lambda\bar{\zeta})^i\right| \le n$):

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|(U_T^k f,g)|^2 = 0 \qquad \forall f \in V^\perp,\ g \in L^2.$$

The Koopman v. Neumann lemma 4.5 yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|(U_T^k f,g)| = 0 \qquad \forall f \in V^\perp,\ g \in L^2.$$

Since constant functions are the only eigenfunctions we see $f-(f,1)1 \in V^\perp$, so

$$\begin{aligned}0 =& \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|(U_T^k f - (f,1)1, g)| \\ =& \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|(U_T^k f, g) - (f,1)\overline{(g,1)}|\end{aligned}$$

and the claim follows by Theorem 4.4. $\square$

## 5. Topological Dynamics

$(X, G)$ with $G$ a group of homeomorphisms on a compact metric space $X$ is a *dynamical system.* If $G$ is generated by a single homeomorphism $T$ we write $(X, T)$ for $(X, \{T^n : n \in \mathbb{Z}\})$.

A subset $Y$ of $X$ is *invariant* if $gY = Y$ for all $g \in G$, so for dynamical systems $(X, T)$ $Y$ is invariant iff $Y = T(Y)$.

A dynamical system $(X, T)$ is *minimal* if $X$ contains no nontrivial invariant subset.

For a dynamical system $(X, G)$ the set $\mathcal{O}(x) := \{gx : g \in G\}$ is called the *orbit* of $x$.

{misu}

**Theorem 5.1.** *For a dynamical system $(X, G)$ there is a closed non empty subset $Y$ of $X$ such that $(Y, G|_Y)$ is a minimal dynamical system.*

*Proof.* Set inclusion gives a partial order on the set $\mathcal{D}$ of non empty closed $T$-invariant subsets of $X$. By compactness the intersection of a linearly ordered subset of $\mathcal{D}$ is non empty. It is also $T$-invariant, so it is a lower bound for this linearly ordered subset of $\mathcal{D}$. The Lemma of Zorn now gives the existence of a minimal element in $\mathcal{D}$, i.e. of a minimal dynamical system. □

{mino}

**Theorem 5.2.** *For a dynamical system $(X, T)$ the following are equivalent:*

i) *$(X, T)$ is minimal;*
ii) *$\overline{\{T^n x : n \in \mathbb{N}\}} = X$ for all $x \in X$;*
iii) *$\overline{\mathcal{O}(x)} = X$ for all $x \in X$.*

*Proof.* i) $\Rightarrow$ ii): For $x \in X$ set $A_n := \overline{\{T^k x : k \geq n\}}$. Then $A := \cap_{n\in\mathbb{N}} A_n$ is by compactness non empty and closed. Since $T$ is a homeomorphism $T(A_k) = A_{k+1}$ and

$$T(A) = T(\cap_{n\in\mathbb{N}} A_n) = \cap_{n\in\mathbb{N}} T(A_n) = \cap_{n\in\mathbb{N}} A_{n+1} = A.$$

So $A$ is $T$-invariant and by minimality of the dynamical system $A = X$. Since $\overline{\{T^n x : n \in \mathbb{N}\}} = A_1 \supset A$ the claim follows.

iii) $\Rightarrow$ i): If $Y$ is a closed $T$-invariant subset of $X$, then $\mathcal{O}(y) \subset Y$ for $y \in Y$, so the assumption implies $X = Y$.

ii) $\Rightarrow$ iii) is obvious. □

The dynamical system $([0, 1], T)$ with $Tx = x^2$ shows that without the assumption of minimality for some $x$ $\overline{\{T^n x : n \in \mathbb{N}\}} \neq \overline{\mathcal{O}(x)}$ and $x$ need not be recurrent. We have however:

{brec}

**Theorem 5.3.** *(Birkhoff Recurrence Theorem, 1927) For a dynamical system $(X, T)$ there is $x \in X$ and a sequence $(n_k)$ in $\mathbb{N}$ such that $T^{n_k} x \to x$ as $k \to \infty$.*

*Proof.* By Theorem 5.1 we may assume w.l.o.g. that $X$ is minimal. Then by Theorem 5.2 for $x \in X$ the sequence $(T^n x)$ has a cluster point $z$ in $X$, since $X$ is compact. By Theorem 5.2 $x \in \overline{\mathcal{O}(z)}$, so for a neighbourhood $U$ of $x$ there is $l_1 \in \mathbb{N}$ with $T^{l_1} z \in U$. By continuity of $T^{l_1}$ we have $T^{l_1} s \in U$ for all $s$ in some neighbourhood $V$ of $z$. $z \in \overline{\mathcal{O}(x)}$, so there is $l_2 \in \mathbb{N}$ with $T^{l_2} x \in V$. It follows that $T^{l_1+l_2} x \in U$. □

- A dynamical system $(X, T)$ is *topologically ergodic* if any closed $T$-invariant proper subset of $X$ has empty interior.
- $(X, T)$ is *topologically weakly mixing* if $(X \times X, T \times T)$ is topologically ergodic.
- $(X, T)$ is *topologically strongly mixing* if for all nonempty open subsets $U, V$ of $X$ there is $n_0 \in \mathbb{N}$ such that $T^n(U) \cap V \neq \emptyset$ for all $n \geq n_0$.

**Theorem 5.4.** *In a dynamical system $(X, T)$ the following are equivalent:*

i) $(X, T)$ *is topologically ergodic;*
ii) *The subset of elements of $X$ with non dense orbit is of first category;*
iii) *There is an element in $X$ with dense orbit;*
iv) *For $U, V$ non empty open subsets of $X$ there is $n \in \mathbb{Z}$ with $T^n U \cap V \neq \emptyset$.*

*Proof.* i) ⇒ ii): If $T$ is topologically ergodic then for any nonempty open subset $U$ of $X$ $\cup_{i\in\mathbb{Z}} T^i(U)$ is a dense subset, so for a cover $B_1, \ldots, B_k$ of $X$ with balls of radius $\epsilon/2$ and some $U_l \subset X$ open, there is $x_l \in U_l, y_l \in B_{l+1}, n_l \in \mathbb{Z}$ with $T^{n_l} x_l = y_l$. By continuity of $T^{n_l}$ there is an open subset $U_{l+1}$ of $U_l$ such that $T^{n_l}(U_{l+1}) \subset B_{l+1}$. Starting with an arbitrary open subset $U_0$ of $X$ we therefore find a sequence $U_0 \supset U_1 \cdots \supset U_k$ of open sets and a sequence $(n_i)$ of integers such that $T^{n_i}(U_{i+1}) \subset B_{i+1}$, $i = 0, \ldots, k-1$. It follows that any $x \in U_k$ has an $\epsilon$-dense orbit, that is for $y \in X$ there is $n_{x,y} \in \mathbb{Z}$ with $d(T^{n_{x,y}} x, y) < \epsilon$. So the interior of the set $V_\epsilon$ of elements with $\epsilon$-dense orbit is dense and we see that the set of elements without dense orbit $\cup_{n\in\mathbb{N}} V_{\frac{1}{n}}^{\complement}$ is of first category.

ii) ⇒ iii) is immediate since compact spaces are of second category.

iii) ⇒ iv): If $x$ has dense orbit then there are $n_1, n_2$ with $T^{n_1} x \in V$, $T^{n_2} x \in U$, so $T^{n_1-n_2}(T^{n_2} x) \in V$ and the claim follows with $n = n_1 - n_2$.

iv) ⇒ i): If the maps $T^i$ generate for any non empty open subset of $X$ a dense subset of $X$, then there is no closed invariant proper subset of $X$ with non empty interior, so $(X, T)$ is topologically ergodic.

□

In a dynamical system $(X,T)$ a point $x \in X$ is *uniformly recurrent* (or *almost periodic*) if for $\epsilon > 0$ there is an increasing sequence $(n_i)$ in $\mathbb{N}$ with $d(T^{n_i}x, x) < \epsilon$ and $\sup_i n_{i+1} - n_i < \infty$.

**Theorem 5.5.** *(Gottschalk, 1944) A point $x$ in a dynamical system $(X,T)$ is uniformly recurrent iff $(\overline{\mathcal{O}(x)}, T)$ is minimal.* {gott}

*Proof.* If $\overline{\mathcal{O}(x)}$ is minimal, then by Theorem 5.2 $\overline{\{T^n y : n \in \mathbb{N}\}} = \overline{\mathcal{O}(y)} = \overline{\mathcal{O}(x)}\,\forall y \in \overline{\mathcal{O}(x)}$. So for $\epsilon > 0$ and $y \in \overline{\mathcal{O}(x)}$ there is $n_y \in \mathbb{N}$ with $d(T^{n_y}y, x) < \epsilon$. By continuity of $T^{n_y}$ there is a neighbourhood $U(y)$ of $y$ with $d(T^{n_y}z, x) < \epsilon$ for $z \in U(y)$. By compactness there is a finite cover of $\overline{\mathcal{O}(y)}$ with such neighbourhoods $U(y_1), \ldots, U(y_k)$.

For $n_0 := \max\{n_1, \ldots, n_k\}$ we have for all $y \in \overline{\mathcal{O}(x)}$ some $w \in \{y, Ty, \ldots, T^{n_0}y\}$ which satisfies $d(x, w) < \epsilon$, so $x$ is almost periodic.

If $\overline{\mathcal{O}(x)}$ is not minimal, then for some $y \in \overline{\mathcal{O}(x)}$ we have $x \notin \overline{\mathcal{O}(y)}$, so for some $\alpha > 0$ we have $d(T^n y, x) > \alpha$ for all $n$. By continuity of $T^n$ $d(T^n z, x) > \alpha$ for all $z$ in some neighbourhood $U_n(y)$ of $y$.

For $k \in \mathbb{N}$ there is $n_0$ with $T^{n_0}x \in \cap_{i=1}^{k} U_i(y)$. So $d(T^{n_0+l}x, x) > \alpha\,\forall l \leq k$ and $x$ is not almost periodic. $\square$

Two elements $x_1, x_2$ of a dynamical system $(X,T)$ are *proximal* if there is $z \in X$ and a sequence $(n_k)$in $\mathbb{N}$ such that $T^{n_k}x_1 \to z$; $T^{n_k}x_2 \to z$ as $k \to \infty$.

Let $X^X$ denote the space of functions mapping $X$ to itself. Endowed with the topology of pointwise convergence it is by Tychonoff's Theorem a compact space. The closure of the set of functions $\{T^n : n \in \mathbb{N}\}$ in $X^X$ is the *Ellis semigroup* $E = E(X,T)$.

**Theorem 5.6.** *For the Ellis semigroup $E$ of a dynamical system $(X,T)$ we have:*

i) *Under composition of functions $E$ is a semigroup;*
ii) *Right multiplication $p \mapsto pq$ is continuous in $E$;*
iii) *Left multiplication by a continuous element of $E$ is continuous;*
iv) *Two points $x_1, x_2 \in X$ are proximal iff there is $p \in E$ with $p(x_1) = p(x_2)$;*
v) $\overline{\{T^n x : n \in \mathbb{N}\}} = Ex \quad \forall x \in X$.

*Proof.* i) For $p, q \in E$ we have to show that the composition $pq$ is in $E$. For $\epsilon > 0$ and $x_1, \ldots, x_k \in X$ we find, since $p \in E$, some $m \in \mathbb{N}$ such that $d(T^m(q(x_i)), pq(x_i)) < \epsilon$ for $i \leq k$. By continuity of the map $T^m$ there are neighbourhoods $U_i$ of $q(x_i)$ such that $d(T^m y, pq(x_i)) < \epsilon$ for $y \in U_i$. As $q \in E$ there is $n$ such that $T^n x_i \in U_i$ and we then obtain $d(T^{m+n}x_i, pq(x_i)) < \epsilon$ for $1 \leq i \leq k$.

ii) This is obvious since the set of $r \in E$ for which $d(pq(x_i), rq(x_i)) < \epsilon$ is a neighbourhood of $p$ in the topology of pointwise convergence.

iii) If $f$ is continuous at the points $p(x_i)$, $i = 1, \ldots, s$, then for $\epsilon > 0$ there is a $\delta > 0$ such that $d(f(p(x_i)), f(y_i)) < \epsilon$ for $d(p(x_i), y_i) < \delta$. Since $\{q \in E : d(p(x_i), q(x_i)) < \delta\}$ is a neighbourhood of $p$ in $E$ left multiplication by a continuous function in $E$ is continuous.

iv) Any cluster point $p$ of the sequence $T^{n_k}$ satisfies $p(x_1) = p(x_2) = z$ if $T^{n_k}(x_i) \to z$ for $k \to \infty$.

Conversely if there is $p$ with $p(x_1) = p(x_2)$, then for any $\epsilon > 0$ there is $n_\epsilon \in \mathbb{N}$ with $d(T^{n_\epsilon}(x_i), p(x_i)) < \epsilon/2$ for $i = 1, 2$, so $d(T^{n_\epsilon}(x_1), T^{n_\epsilon}(x_2)) < \epsilon$.

v) Clearly $Ex \supset \overline{\mathcal{O}(x)}$. Conversely any $y = px$ is the limit of $T^{n_k}x$ for some sequence $(n_k)$, so $y \in \overline{\mathcal{O}(x)}$. $\square$

We define a *right topological semigroup* as a semigroup endowed with a Hausdorff topology under which right multiplication $p \mapsto pq$ is continuous. Note that some authors refer to these semigroups as left topological semigroups!

{ellis} **Theorem 5.7.** *(Ellis, 1958) Any compact right topological semigroup $S$ contains an idempotent, i.e. some element $p \in S$ with $p = p^2$.*

*Proof.* If we order closed subsemigroups of $S$ by set inclusion, compactness and the Lemma of Zorn give the existence of a minimal closed subsemigroup $S_0$. Take $p \in S_0$ and set $M_p := \{q \in S_0 : qp = p\}$. Then $M_p$ is closed since $S$ is right topological. $S_0p$, as the continuous image of a compact set, is a closed subsemigroup of $S_0$, so by minimality $S_0p = S_0$ and there is $q$ in $S_0$ with $qp = p$, so $M_p$ is non empty. Since $M_p$ is also a semigroup it has to be $S_0$ by minimality, therefore $p \in M_p$ i.e. $p$ is idempotent.

We have in fact shown that the minimal closed subsemigroups are just the idempotents. $\square$

{proxp} **Theorem 5.8.** *For $x_0$ in the dynamical system $(X, T)$ the set $L(x_0)$ of cluster points of $\{T^n x_0 : n \geq 0\}$ is $T$-invariant. Any minimal dynamical subsystem $(M, T)$ of $(L(x_0), T)$ contains an element $m_0$ such that $m_0$ and $x_0$ are proximal.*

*Proof.* $TL(x_0) = L(x_0)$ since the cluster points of the set $\{T^n x_0 : n \geq 0\}$ are the cluster points of $\{T^n x_0 : n \geq 1\}$. Since $X$ is a metric space there is for $y \in L(x_0)$ a sequence $(n_k)$ such that $T^{n_k} x_0 \mapsto y$. Therefore any cluster point $p$ of $\{T^{n_k} : k \geq 0\}$ in the Ellis semigroup $E$ satisfies $px_0 = y$, and $Ex_0 \supset L(x_0)$.

Conversely if $px_0 = y$, then there is a sequence $(n_k) \in \mathbb{N}$ with $T^{n_k} x_0 \to y$, so $y \in L(x_0)$. Therefore $Ex_0 = L(x_0)$.

Set $F := \{p \in E : px_0 \in M\}$, then $F$ is non empty and closed since the projection $E \to X : p \mapsto px_0$ is continuous. As $M \subset L(x_0) = Ex_0$ we have $Fx_0 = M$ and since $M$ is $T$-invariant any $p \in E$ maps $M$ to itself. It follows that $F^2 \subset EF \subset F$ and we see that $F$ is a closed semigroup.

By Theorem 5.7 there is an idempotent $u$ in $F$. Then $ux_0 \in M$ and $ux_0$ and $x_0$ are proximal since $uux_0 = ux_0$ and for $(n_k)$ with $T^{n_k} \to u$ at the coordinates $x_0$ and $ux_0$, then $T^{n_k}ux_0 \to u^2x_0 = ux_0$ and $T^{n_k}x_0 \to ux_0$.

□

## 6. Applications to Ramsey Theory

Ramsey theory is a collection of theorems in combinatorics, asserting that if some set is "large" enough it contains some subset with a certain structure. The set considered, the meaning of large and the definition of the structure vary.

For example the classical infinite *Ramsey party theorem* (Ramsey, 1930) states that if we consider a party with infinitely many guests, it is always possible to find either an infinite subset of these guests all of which know each other or an infinite subset of guests in which noone knows anyone else. In other words for any two-coloring of the set of unordered pairs of naturals we can find an infinite subset of $\mathbb{N}$ for which the corresponding pairs are monocromatic, i.e. all pairs have the same color assigned.

In fact pick any guest $a_1$ who knows infinitely many other guests, eliminate all guests who do not know $a_1$, then pick from the remaining guests some $a_2$ who knows infinitely many among the remaining and eliminate all guests who do not know $a_2$ and continue this way. We either get an infinite sequence $a_1, a_2, \dots$ of guests all of which know each other, or we end up with an infinite subset of guests in which noone knows infinitely many of this subset. In this case we pick any guest $b_1$ of this subset, eliminate the finite number of guests who know $b_1$, pick some $b_2$ from the remaining guests, eliminate the finite number of guests who know $b_2$ and carry on this way. This construction never ends since we started with an infinite number of guests and eliminated only finitely many in each step, which yields an infinite sequence of guests $b_1, b_2, \dots$ which do not know each other.

By induction it is easy to generalize this statement to $n-$colorings of the set of unordered pairs in $\mathbb{N}$ and to imply the finite Ramsey theorem: For $k, n \in \mathbb{N}$ there is a number $R(k,n) \in \mathbb{N}$ such that for any $n-$coloring of the unordered pairs of a set with at least $R(k,n)$ there is a subset with $k$ elements, for which all pairs have the same color assigned.

Ramseys party theorem gives a short proof of *Schur's theorem (1916)*: In any $n-$coloring of $\mathbb{N}$ there are two numbers $x, y$ s.t. $x, y$ and $x+y$ have the same color. Simply assign to any pair $z_i < z_j$ of naturals the color of the number $z_j - z_i$. Then by Ramsey's party theorem there are three numbers $z_1 < z_2 < z_3$ with $z_2 - z_1$, $z_3 - z_2$ and $z_3 - z_1$ have the same color. For $x = z_2 - z_1$, $y = z_3 - z_2$ this yields Schur's theorem.

We now show how recurrence properties in dynamical systems can be used to prove theorems in infinite Ramsey theory. Our first application is a generalization of Schur's theorem.

Let $C_1, \dots, C_r$ be a partition of $\mathbb{N}$ and consider a finite or infinite sequence $(p_i)$, $i = 1, 2, \dots$ in $\mathbb{N}$. This sequence is called an *IP-sequence* if there is some $r_0 \leq r$ such that $\sum_j a_j p_j \in C_{r_0}$ for any non-zero sequence $(a_j)$ with $a_j = 1$ for finitely many $j$ and 0 otherwise. So $(p_i)$ is an IP-sequence if all finite sums $p_{j_1} + p_{j_2} + \dots + p_{j_k}$ are in $C_{r_0}$ for any $1 \leq j_1 < j_2 < \dots < j_k$, $k \geq 1$. If $(p_j)$ is an IP-sequence, this can be thought of some $C_{r_0}$ containing an infinite parallelepiped.

It was already shown by Hilbert, that any finite partition of $\mathbb{N}$ contains finite IP-sequences of arbitrary length.

**Theorem 6.1.** *(Hindman, 1974) For any finite partition* $\mathbb{N} = C_1 \cup \dots \cup C_r$ *there is* $i \leq r$ *such that* $C_i$ *contains an infinite IP-sequence.*

*Proof.* (Furstenberg & Weiss)
Set $X := \Lambda^{\mathbb{Z}}$ with $\Lambda := \{1, \dots, r\}$ and

$$\omega(n) = \begin{cases} j & n \geq 1 \quad n \in C_j \\ 1 & n \leq 0. \end{cases}$$

$X$ endowed with the product topology is a compact metrizable space on which we denote left translation by $T$: $Tx(n) = x(n+1)$.

We therefore have to show that $\omega(n)$ has identical values for all $n$ in some infinite IP-set.

If $\omega$ were recurrent we could choose $p_1 > 0$ such that $\omega(p_1) = \omega(0)$ and define $p_{k+1} > 0$ inductively by requiring that $T^{p_{k+1}}\omega(n) = \omega(0)$ for all $n \in S_k := \left\{\sum_{i=1}^k a_i p_i : a_i \in \{0,1\}\right\}$. Then $\omega(n) = \omega(0)$ for all $n \in S_{k+1} = S_k \cup (p_{k+1} + S_k)$ and $(p_i)$ defines an infinite IP-sequence with $r_0 = \omega(0)$.

However $\omega$ need not be recurrent, but by Theorem 5.8 there is a $m$ in $X$, which is proximal to $\omega$ in some minimal subsystem. By Theorems 5.2and 5.5 $\omega$ is uniformly recurrent. This allows us to refine the above argument to find an IP-set on which both $\omega$ and $m$ take value $m(p_1)$:

As $m$ is uniformly recurrent there is $l_0$ s.t. for all $s \in \mathbb{N}$ there is $0 \leq l < l_0$ with $m(0) = m(s + l)$. As $m$ and $\omega$ are proximal we can choose $p_1$ such that $\omega(p_1) = m(p_1) = m(0)$. We define $p_k \in \mathbb{N}$ inductively: Set

$$S_k^* := \left\{ \sum_{i=1}^k a_i p_i : a_i \in \{0,1\} \right\} \setminus \{0\}$$

and assume that $p_1, \dots, p_k$ have been chosen such that $\omega(n) = m(n) = \omega(p_1)$ for $n \in S_k^*$. Then $U_k := \{x \in X : x(n) = \omega(p_1) \, \forall n \in S_k^*\}$ is a neighbourhood of both $m$ and $\omega$.

By uniform recurrence of $m$ there is some $l_k \in \mathbb{N}$ such that $\forall s \in \mathbb{N}$ there is $0 \le l < l_k$ with $T^{s+l}m \in U_k$. Since $m$ and $\omega$ are proximal there is $r_k \in \mathbb{N}$ with $T^{r_k}\omega(n) = T^{r_k}m(n)$ for $0 \le n < l_k + \sum_{i=1}^{k} p_i$. We therefore can choose $0 \le l < l_k$ such that for $p_{k+1} := r_k + l$ we have $T^{p_{k+1}}m \in U_k$. It follows that $\omega(n) = m(n) = m(0)$ for $n \in p_{k+1} + S_k$. Since $S_{k+1} = S_k \cup (p_{k+1} + S_k)$ we have $\omega(n) = m(n) = m(0)$ for $n \in S^*_{k+1}$. □

For a dynamical system $(X,T)$ a closed subset $A$ of $X$ is called *homogeneous*, if there is a group $G$ of homeomorphisms of $X$ commuting with $T$, such that $(A,G)$ is a minimal dynamical system. Note that $A$ need not be $T$-invariant!

{homsub1}

**Proposition 6.2.** *Let $A$ be a homogeneous subset of $X$ and assume that for $\epsilon > 0$ there is $x_0, y_0 \in A$, $n_0 \ge 1$ with $d(T^{n_0}x_0, y_0) < \epsilon$. Then for all $\epsilon > 0$ there is $z \in A$, $n \ge 1$ with $d(T^n z, z) < \epsilon$.*

*Proof.* The set $V := \{(x,y) : d(x,y) < \epsilon/2\}$ is an open subset of the compact set $X \times X$ and the homeomorphisms $g$, $g \in G$ induce continuous maps again denoted by $g$: $X\times X \to X\times X$, $(x,y) \mapsto (gx,y)$.

Since $(A,G)$ is minimal we have for $x \in A : \overline{Gx} = A$, $\cup_{g\in G} g^{-1}V \supset A \times A$ and there are $g_1, \dots, g_k \in G$ such that $g_1^{-1}V \cup \dots \cup g_k^{-1}V$ is a finite cover of $A \times A$, so $\min\{d(g_i x, y) : i \le k\} < \epsilon/2$ for all $x, y \in A$.

Let $\delta > 0$ be such that $d(g_i x, g_i x') < \epsilon/2$ for $d(x,x') < \delta$ and $1 \le i \le k$.

By assumption there is $x_0, y_0 \in A$, $n_1 \ge 1$ with $d(T^{n_1}x_0, y_0) < \delta$, so

$$d(g_i T^{n_1} x_0, g_i y_0) = d(T^{n_1} g_i x_0, g_i y_0) < \epsilon/2 \quad \forall 1 \le i \le k.$$

For $i$ with $d(g_i y_0, y) < \epsilon/2$ we get $d(T^{n_1} g_i x_0, y) < \epsilon$.

Therefore the assumption holds for all $y_0 \in A$, i.e. for $z_0 \in A, \epsilon > 0$ there is $0 < \epsilon_1 < \epsilon/2, z_1 \in A, n_1 \ge 1$ with $d(T^{n_1} z_1, z_0) < \epsilon_1$.

By continuity of $T^{n_1}$ there is $\epsilon_2 < \epsilon_1$ such that $d(T^{n_1}z, z_0) < \epsilon_1$ for $d(z, z_1) < \epsilon_2$. We then choose $n_2, z_2$ with $d(T^{n_2} z_2, z_1) < \epsilon_2$, so we get $d(T^{n_2+n_1} z_2, z_0) < \epsilon/2$.

Proceeding inductively we have sequences $(\epsilon_i), (z_i), (n_i)$, with

$$d(T^{n_i} z_i, z_{i-1}) < \epsilon_i < \epsilon/2$$

and $d(z, z_{i-1}) < \epsilon_i$ implies $d(T^{n_{i-1}} z, z_{i-2}) < \epsilon_{i-1}$, which yields

$$\begin{array}{rcl} d(T^{n_j} z_j, z_{j-1}) & < & \epsilon_j \\ d(T^{n_j+n_{j-1}} z_j, z_{j-2}) & < & \epsilon_{j-1} \\ & \vdots & \\ d(T^{n_j+n_{j-1}+\dots+n_{i+1}} z_j, z_i) & < & \epsilon_{i+1} < \epsilon/2. \end{array}$$

By compactness of $A$ there are $i < j$, such that $d(z_i, z_j) < \epsilon/2$ and $d(T^n z_j, z_i) < \epsilon/2$ for $n = n_{i+1} + \ldots + n_j$, which gives $d(T^n z_j, z_j) < \epsilon$. $\square$

{homsub2}

**Proposition 6.3.** *Let $A$ be a homogeneous subset of $X$ and assume that for $\epsilon > 0$ there is $x_0, y_0 \in A$, $n_0 \geq 1$ with $d(T^{n_0} x_0, y_0) < \epsilon$. Then there is $x \in A$ which is recurrent in $(X, T)$.*

*Proof.* We have to show that $\cup_n E_n \neq A$ for

$$E_n := \{x \in A : d(T^k x, x) \geq 1/n \,\forall k \geq 1\}.$$

Since the sets $E_n$ are closed it suffices by Baire's category theorem to show that the interior $E_n^\circ$ of $E_n$ is empty for all $n$.

So assume $V := E_n^\circ \neq \emptyset$ for some $n$. Then by minimality of $(A, G)$ for $a \in A$ there is $g \in G$ with $ga \in V$ or equivalently $A \subset \cup_{g \in G} gV$, so there are $g_1, \ldots, g_r \in G$ with $\cup_{i \leq r} g_i^{-1} V \supset A$ and a $\delta > 0$ such that $d(g_i x, g_i y) < 1/n$ for $d(x, y) < \delta$ and $i \leq r$.

By Proposition 6.2 there is $z \in A$, $k \geq 1$ with $d(T^k z, z) < \delta$. Take $i \leq r$ such that $z = g_i^{-1} v$ for some $v \in V$, then $d(g_i T^k z, g_i z) = d(T^k v, v) < 1/n$ contradicting $v \in E_n$. $\square$

{fwthm}

**Theorem 6.4.** *(Furstenberg, Weiss, 1978) Let $T_1, \ldots, T_k$ be commuting homeomorphisms of a compact metric space $X$. Then there is $x \in X$ such that for all $\epsilon > 0$ there is $n \geq 1$ with $d(T_i^n x, x) < \epsilon \,\forall i = 1, \ldots, k$.*

*Proof.* For $k = 1$ this is just Birkhoff's recurrence theorem 5.3. We prove by induction.

Let $G$ be the group generated by $T_1, \ldots, T_k$. We may assume by Theorem 5.1 that $(X, G)$ is minimal. Set $T = T_1 \times \cdots \times T_k$, then $T$ and $G$ act on $X^k$ by $T(x_1, \ldots, x_k) = (T_1 x_1, \ldots, T_k x_k)$ and $g(x_1, \ldots, x_k) = (gx_1, \ldots, gx_k)$, $g \in G$. We endow $X^k$ with the metric

$$d((x_1, \ldots, x_k), (y_1, \ldots, y_k)) = \max\{d(x_i, y_i) : i \leq k\}.$$

$g \in G$ and $T$ commute and if $\Delta$ denotes the diagonal in $X^k$ then $(\Delta, G)$ being isomorphic to $(X, G)$ is minimal. Therefore $\Delta$ is a homogeneous subset of $(X^k, T)$.

For $i = 1, \ldots, k-1$ there is by induction hypothesis a $z \in X$ and $n_m \to \infty$ with $T_i^{n_m} T_k^{-n_m} z \to z$ for $i = 1, \ldots, k-1$.

Set $x_0^* = (T_k^{-n_m} z, T_k^{-n_m} z, \ldots, T_k^{-n_m} z)$, $y_0^* = (z, z, \ldots, z) \in X^k$, then for $\epsilon > 0$ there is $m$ such that

$$\begin{aligned} d(T^{n_m} x_0^*, y_0^*) &= d(T_1^{n_m} \times T_2^{n_m} \times \ldots \times T_k^{n_m} x_0^*, y_0^*) \\ &= d((T_1^{n_m} T_k^{-n_m} z, \ldots, T_{k-1}^{n_m} T_k^{-n_m} z, z), (z, \ldots, z)) < \epsilon. \end{aligned}$$

By Proposition 6.3 there is $x^* \in \Delta$ which is recurrent under $T = T_1 \times \ldots \times T_k$.

$\square$

**Theorem 6.5.** *(Van der Waerden, 1927) Any finite partition of* $\mathbb{N} = C_1 \cup \ldots \cup C_r$ *contains a set* $C_{r_0}$ *with arithmetic progressions of arbitrary length, i.e. for* $k \in \mathbb{N}$ *there is* $a, b \in \mathbb{N}$ *with* $a + ib \in C_{r_0}$ *for* $0 \leq i \leq k$.

*Proof.* Set $\Omega := \Lambda^{\mathbb{Z}}$, $\Lambda := \{1, \ldots, r\}$ and

$$d(\omega_1, \omega_2) = 2^{-s},\ s = \inf\{|n| : \omega_1(n) \neq \omega_2(n)\}.$$

We denote left translation on $\Omega$ by $\sigma$: $\sigma\omega(n) = \omega(n+1)$. Define $\omega_0 \in \Omega$ by

$$\omega_0(n) = \begin{cases} j & n \geq 1 \quad n \in C_j \\ 1 & n \leq 0. \end{cases}$$

Set $X := E\omega_0$, where $E$ is the Ellis semigroup generated by $\sigma$. Then $X$ is closed and $\sigma$-invariant.

Applying the Furstenberg-Weiss theorem 6.4 to the space $X$ and the maps $T_i := \sigma^i$, $i = 1, \ldots, k$ and $\epsilon = 1$ shows that there is $x \in X$, $b \geq 1$ such that $d(T_i^b x, x) < 1\ \forall i = 1, 2, \ldots, k$. Therefore all elements $\sigma^{bl} x$, $0 \leq l \leq k$ have the same 0-th coordinate, i.e. $x(0) = x(b) = x(2b) = \ldots = x(kb)$.

As $x$ is a cluster point of $\{\sigma^b \omega_0\}$ there is some $a \geq 0$ such that $\sigma^a \omega_0(lb) = x(lb)$ for $0 \leq l \leq k$, so

$$\omega_0(a) = \omega_0(a + b) = \ldots = \omega_0(a + kb).$$

$\square$