

ERGODIC SETS

JOHN C. OXTOBY

Introduction. Ergodic sets were introduced by Kryloff and Bogoliouboff in 1937 in connection with their study of compact dynamical systems [16]. The purpose of this paper is to review some of the work that has since been done on the theory that centers around this notion, and to present a number of supplementary remarks, applications, and simplifications. For simplicity we shall confine attention to systems with a discrete time. Continuous flows present no difficulty, but the development of a corresponding theory for general transformation groups is still in an incomplete stage. An example due to Kolmogoroff (see [5]) shows that such an extension cannot be made without sacrificing either the invariance or the disjointness of ergodic sets.

In §§1 and 2 we give a brief, but self-sufficient, development of the basic theorems of Kryloff and Bogoliouboff. In §3 we collect some auxiliary results for later use. In §4 a simple characterization of transitive points is obtained. In §5 the distinctive properties of some special types of systems and subsystems are discussed, and in §6 these results are used to discover conditions under which the ergodic theorem holds uniformly. In §7 a generalization to noncompact systems is considered, and in §§8 and 9 some known representation theorems are obtained as an application of ergodic sets. In §10 there is given an example of a minimal set that is not strictly ergodic, similar to one constructed by Markoff.

1. Some corollaries of the ergodic theorem. We shall use the following notations: If $f(p)$ is a real-valued function on a set Ω , and if T is a 1:1 transformation of Ω onto itself, then

$$M(f, p, k) = f_k(p) = \frac{1}{k} \sum_{i=1}^{i=k} f(T^i p) \quad (k = 1, 2, \dots)$$

and

$$M(f, p) = f^*(p) = \lim_{k \rightarrow \infty} M(f, p, k)$$

in case this limit exists. The characteristic function of a set E is

An address delivered before the Washington meeting of the Society, October 27, 1951, by invitation of the Committee to Select Hour Speakers for Eastern Sectional Meetings; received by the editors November 12, 1951.

denoted by χ_E . The upper density of a set E of integers is defined by $\delta^*(E) = \limsup_{k \rightarrow \infty} (1/k) \sum_{n=1}^{k-1} \chi_E(n)$.

Let T be a 1:1 measure preserving transformation of a measure space (Ω, μ) onto itself, and assume that $\mu(\Omega) = 1$. The ergodic theorem states:

(1.1) *If $f \in L_1$, then the sequence $\{f_k(p)\}$ converges a.e. to a limit function $f^* \in L_1$ such that $f^*(Tp) = f^*(p)$ and $\int f^* d\mu = \int f d\mu$.*

(1.2) *If $f \in L_1$ and f is non-negative, then for almost all p either $f^*(p) > 0$ or $f(p) = 0$.*

PROOF. Let $\chi(p)$ be the characteristic function of the set on which $f^*(p)$ is either zero or undefined, and apply (1.1) to the function χf .

A second corollary is the mean ergodic theorem:

(1.3) *If $f \in L_q, 1 \leq q < \infty$, then $f^* \in L_q$ and $\|f_k - f^*\|_q \rightarrow 0$ as $k \rightarrow \infty$.*

PROOF. For any bounded function f the conclusion follows immediately from (1.1). For any other $f \in L_q$ it follows from the inequality $\|f_k\|_q \leq \|f\|_q$ ($k = 1, 2, \dots$) and the fact that bounded functions are dense in L_q .

A less familiar corollary is the following:

(1.4) *If $f \in L_2$, then the limit functions $f^*(p)$ and $\phi^k(p) = \lim_{n \rightarrow \infty} (1/n) \sum_{i=1}^{n-1} (f_k(T^i p) - f^*(p))^2$ ($k = 1, 2, \dots$) are defined a.e., and $\int \phi^k d\mu \rightarrow 0$ as $k \rightarrow \infty$.*

PROOF. By (1.3), $f_k - f^* \in L_2$, hence $(f_k - f^*)^2 \in L_1$. Since $\phi^k(p) = M((f_k - f^*)^2, p)$, it follows from (1.1) that $\phi^k(p)$ is defined a.e. and that $\int \phi^k d\mu = \int (f_k - f^*)^2 d\mu$. The last conclusion then follows from (1.3).

A system (Ω, T, μ) (and the measure μ) is called ergodic if Ω cannot be split into two disjoint invariant measurable sets of positive measure. For such systems (1.1) implies that f^* is constant a.e., and (1.4) can be strengthened as follows:

(1.5) *If the system (Ω, T, μ) is ergodic, and if $f \in L_2$, then as $k \rightarrow \infty$ the sequence $\{\phi^k(p)\}$ defined in (1.4) converges to 0 uniformly a.e., and for any $\alpha > 0$ the sequence $\psi^k(p, \alpha) = \delta^* \{n : |f_k(T^n p) - \int f d\mu| > \alpha\}$ likewise converges to 0 uniformly a.e.; both sequences converge uniformly on the set Ω_1 of points where $f^*(p) = \int f d\mu$ and $\phi^k(p) = \int (f_k - f^*)^2 d\mu$ ($k = 1, 2, \dots$).*

PROOF. For an ergodic system, $\mu(\Omega_1) = 1$, and the first conclusion follows from (1.3). The second conclusion then follows from the inequality

$$\alpha^2 \delta^* \{n : |f_k(T^n p) - f^*(p)| > \alpha\} \leq \phi^k(p),$$

which is implied by the definition of ϕ^k .

This result shows that for ergodic systems the convergence of f_k

to its constant limiting value f^* is *nearly uniform* on almost every orbit, in the sense that if $\epsilon > 0$ and k is sufficiently large, then for almost all p the mean square difference between $f_k(T^n p)$ and $f^*(p)$ as n ranges over the positive integers is less than ϵ , and for all n except a set of small upper density the absolute difference is less than ϵ .

2. The theory of Kryloff and Bogoliouboff. Assume now that Ω is a compact metric space and that T is a homeomorphism of Ω onto itself. Denote by $C(\Omega)$ the space of real-valued continuous functions f on Ω , with the norm $\|f\| = \max |f(p)|$. Any finite Borel measure μ in Ω determines a corresponding positive linear functional $L(f)$ on $C(\Omega)$, namely,

$$L(f) = \int f d\mu \quad (f \in C(\Omega)).$$

The Riesz theorem [28; 12] asserts that, conversely, any positive linear functional on $C(\Omega)$ corresponds in this way to one and only one finite Borel measure μ in Ω . Invariant measures correspond to invariant functionals, and normalized measures to functionals for which $L(1) = 1$. It will be seen that the theory of Kryloff and Bogoliouboff is essentially a series of corollaries of this theorem and the ergodic theorem.

(2.1) *Any compact system (Ω, T) admits at least one normalized invariant Borel measure. If K is a compact subset of Ω , then either $\mu(K) > 0$ for some such measure μ or else $M(\chi_K, p) = 0$ for every p in Ω .*

PROOF. Choose any point $p \in \Omega$ and determine (by a diagonal process) an increasing sequence of positive integers k_i such that the limit

$$L(f) = \lim_{i \rightarrow \infty} M(f, p, k_i)$$

exists for a countable dense set of functions f in $C(\Omega)$. Then the limit $L(f)$ exists for every $f \in C(\Omega)$ and defines an invariant positive linear functional such that $L(1) = 1$. Hence $L(f) = \int f d\mu$ for some normalized invariant Borel measure μ . If K is compact and $M(\chi_K, p)$ is not identically zero, the point p and the sequence $\{k_i\}$ can be so chosen that $\alpha = \limsup_{i \rightarrow \infty} M(\chi_K, p, k_i) > 0$. This implies that $L(f) \geq \alpha$ whenever $f \geq \chi_K$, and therefore $\mu(K) \geq \alpha$.

A point p in Ω is called *quasi-regular* ($p \in Q$) if the mean value $M(f, p)$ is defined for every $f \in C(\Omega)$. A Borel subset E of Ω is said to have *invariant measure one* if $\mu(E) = 1$ for every normalized invariant Borel measure μ . It follows from (2.1) that any such set is nonempty.

(2.2) *Q is a Borel set of invariant measure one.*

PROOF. In order that p belong to Q it is sufficient that the limit $M(f, p)$ exist for every f belonging to a countable dense subset of $C(\Omega)$. For any $f \in C(\Omega)$ the set on which $M(f, p)$ is defined is an $F_{\sigma\delta}$ subset of Ω [17, p. 274], and by (1.1) this set has invariant measure one. Because Q is a countable intersection of such sets, it has the same two properties.

To each quasi-regular point p corresponds uniquely a normalized invariant Borel measure μ_p in Ω , such that

$$M(f, p) = \int f d\mu_p$$

for every $f \in C(\Omega)$. A quasi-regular point p is called a *point of density* ($p \in Q_D$) if $\mu_p(U) > 0$ for every open set U containing p . A quasi-regular point p is called *transitive* ($p \in Q_T$) if μ_p is an ergodic measure. A point p is called *regular* if it belongs to the set $R = Q_D \cap Q_T$.

(2.3) Q_D is a Borel set of invariant measure one.

PROOF. Let $\{U_i\}$ be a countable base of open sets in Ω , and for each i let f^i be a continuous function on Ω such that $f^i(p) = 0$ on $\Omega - U_i$ and $f^i(p) > 0$ on U_i . Then the set E_i on which either $M(f^i, p) > 0$ or $f^i(p) = 0$ is an F_σ relative to the set on which $M(f^i, p)$ is defined, therefore an $F_{\sigma\delta}$ in Ω . By (1.2), E_i has invariant measure one. Since $Q_D = Q \cap \bigcap_{i=1}^{\infty} E_i$, the conclusion follows. This result can be looked upon as a refinement of Poincaré's recurrence theorem (cf. [14]).

(2.4) Q_T is a Borel set of invariant measure one.

PROOF. A quasi-regular point p is transitive if and only if for each f belonging to some countable dense subset D of $C(\Omega)$, $M(f, q) = \int f d\mu_p$ for all q except a set of μ_p -measure zero. Consequently, a point p is transitive if and only if $p \in Q$ and

$$\int_Q (f^*(q) - f^*(p))^2 d\mu_p(q) = 0$$

for every $f \in D$. For any $f \in C(\Omega)$, $f_k(q)$ converges boundedly on Q to $f^*(q)$. Hence for any $p \in Q$,

$$\begin{aligned} \int_Q (f^*(q) - f^*(p))^2 d\mu_p(q) &= \lim_k \int_Q (f_k(q) - f^*(p))^2 d\mu_p(q) \\ &= \lim_k \lim_n \frac{1}{n} \sum_{i=1}^{i=n} (f_k(T^i p) - f^*(p))^2. \end{aligned}$$

The subset of Q on which this iterated limit vanishes is an $F_{\sigma\delta}$ set,

and (1.4) shows that it has invariant measure one.

(2.5) R is a Borel set of invariant measure one.

(2.6) For any bounded Borel measurable function f on Ω , $\int f d\mu_p$ is a Borel measurable function of p on Q , and

$$\int f d\mu = \int_Q \left(\int f d\mu_p \right) d\mu(p)$$

for every finite invariant Borel measure μ .

PROOF. For any positive number M , the class of Borel measurable functions f on Ω , bounded by M , for which the assertions of (2.6) are true is closed with respect to pointwise convergence, and includes all continuous functions f with $\|f\| \leq M$.

The next two theorems follow from (2.5) and (2.6).

(2.7) For any Borel set $E \subset \Omega$, $\mu_p(E)$ is Borel measurable on Q , and $\mu(E) = \int_R \mu_p(E) d\mu(p)$ for every finite invariant Borel measure μ .

(2.8) A Borel set $E \subset \Omega$ has invariant measure zero if and only if $\mu(E) = 0$ for every ergodic measure μ .

For any ergodic measure μ , it follows from (1.1) that $\mu_p = \mu$ for all p except a set of μ -measure zero. The set of all such quasi-regular points is called the *quasi-ergodic set* [4] corresponding to μ , and the part of this set contained in R is called the *ergodic set* corresponding to μ . Each of these sets is invariant and $F_{\sigma\delta}$ (see (3.1)). Distinct ergodic measures correspond to disjoint quasi-ergodic sets. The ergodic sets (and the quasi-ergodic sets) stand in 1:1 correspondence with the totality of ergodic measures, and each ergodic measure vanishes outside the corresponding set. The ergodic and quasi-ergodic sets corresponding to a measure do not differ very essentially, since $Q - R$ has invariant measure zero, but both notions are convenient. It should be noted that the points of any ergodic set all have the same orbit closure, whereas the points of a quasi-ergodic set need not. The ergodic sets constitute a partition of R , the quasi-ergodic sets a partition of Q_T .

3. Some additional properties of the set Q .

(3.1) There exists a Borel measurable function $f(p)$ from Q to the unit interval such that $f(p) = f(q)$ if and only if $\mu_p = \mu_q$.

PROOF. Let $\{f^i\}$ be a generating sequence in $C(\Omega)$, that is, a sequence that spans a linear manifold dense in $C(\Omega)$, and assume further that $\|f^i\| \leq 1$ for every i . In order that $\mu_p = \mu_q$ it is necessary and sufficient that $M(f^i, p) = M(f^i, q)$ for every i . For any $p \in Q$ the sequence $\{M(f^i, p)\}$ defines a point $\phi(p)$ of the Cartesian product of \aleph_0 intervals each equal to $[-1, 1]$, and ϕ is of first Baire class on

Q . (It follows that each quasi-ergodic set is G_δ relative to Q , and therefore $F_{\sigma\delta}$ in Ω .) This product space can, in turn, be mapped onto $[0, 1]$ by a 1:1 Borel measurable mapping [17, p. 358]. By composition one obtains a function $f(p)$ with the required properties.

For any $p \in Q$, the equation $M(f, p) = \int f d\mu_p$ holds for every $f \in C(\Omega)$, by definition. We now show that it holds also for some other functions.

(3.2) *If $p \in Q$, the equation $M(f, p) = \int f d\mu_p$ holds for any bounded Borel measurable function f whose upper and lower limit functions are equal almost everywhere (μ_p). In particular, $\mu_p(E) = M(\chi_E, p)$ for any Borel set E whose boundary has μ_p -measure zero.*

PROOF. For any such f , and any $\epsilon > 0$, there exist continuous functions f_1 and f_2 such that $f_1(p) \leq f(p) \leq f_2(p)$ and $\int f_2 d\mu_p < \int f_1 d\mu_p + \epsilon$, and the conclusion follows immediately.

The class of sets E for which $\mu_p(E)$ can be identified with $M(\chi_E, p)$ is somewhat more extensive than is indicated by (3.2), although simple examples show that it does not necessarily include all compact sets. In this connection the following result is useful. We denote by $O^+(p)$ the positive semi-orbit of p , that is, the set $O^+(p) = \bigcup_0^\infty \{T^i p\}$.

(3.3) *For any compact set $K_0 \subset \Omega$ and any quasi-regular point p there exists a compact set K such that $K_0 \subset K \subset K_0 \cup O^+(p)$ and such that $M(\chi_K, p) = \mu_p(K) = \mu_p(K_0)$.*

PROOF. If p is periodic, we may take $K = K_0$. Otherwise, let U_k denote the $1/k$ -neighborhood of K_0 , and for each positive integer k let f^k be a continuous function such that $f^k(q) = 1$ on K_0 , $f^k(q) = 0$ on $\Omega - U_k$, and $0 \leq f^k(q) \leq 1$ on Ω . Let $\{N_k\}$ be an increasing sequence of positive integers such that $M(f^k, p, n) > \mu_p(K_0) - 1/k$ for all $n > N_k$. Put

$$K = K_0 \cup \bigcup_{k=1}^{\infty} \left(\bigcup_{i=1}^{N_{k+1}} \{T^i p\} \right) \cap U_k.$$

Since K has only a finite number of points outside U_k , it follows that K is compact and that $\limsup_n M(\chi_K, p, n) \leq \mu_p(K_0)$. If $0 < i \leq N_{k+1}$, then either $f^k(T^i p) = 0$ or $T^i p \in K$, hence $M(f^k, p, n) \leq M(\chi_K, p, n)$ for $N_k < n \leq N_{k+1}$, $\mu_p(K_0) - 1/k < M(\chi_K, p, n)$ for all $n > N_k$, and therefore $\mu_p(K_0) = M(\chi_K, p)$. The equation $\mu_p(K_0) = \mu_p(K)$ results from the assumption that p is nonperiodic.

(3.4) *If f is bounded and Borel measurable on Ω , if p is quasi-regular, and if for every $\epsilon > 0$ there exists a Borel set E containing $O^+(p)$ such that $\mu_p(E) > 1 - \epsilon$ and such that the contraction of f to E is continuous, then $\int f d\mu_p = M(f, p)$.*

PROOF. Let A be an upper bound of $|f(q)|$ on Ω . Choose a compact set $K_0 \subset E$ with $\mu_p(K_0) > 1 - \epsilon$ (cf. [12, §43(3)]) and construct a compact set K with the properties stated in (3.3). Since $K \subset E$, the contraction of f to K is continuous and there exists a function $g \in C(\Omega)$ such that $g=f$ on K and $\|g\| \leq A$. Since $|f-g| \leq 2A\chi_{\Omega-K}$, we have

$$\begin{aligned} \left| \int f d\mu_p - f_k(p) \right| &\leq \left| \int f d\mu_p - \int g d\mu_p \right| \\ &\quad + \left| \int g d\mu_p - g_k(p) \right| \\ &\quad + |g_k(p) - f_k(p)| \\ &\leq 2A\mu_p(\Omega - K) + \left| \int g d\mu_p - g_k(p) \right| \\ &\quad + 2A \cdot M(\chi_{\Omega-K}, p, k). \end{aligned}$$

The last term is equal to $2A(1 - M(\chi_K, p, k))$ and therefore tends to $2A(1 - \mu_p(K))$ as $k \rightarrow \infty$. Hence the right member tends to a limit less than $4A\epsilon$, and the conclusion follows. For later reference (§7) we record the following corollary:

(3.5) *If f is bounded and Borel measurable on Ω , if p is quasi-regular, and if there exists an invariant Borel set E containing p such that $\mu_p(E) = 1$ and such that the contraction of f to E is continuous, then $\int f d\mu_p = M(f, p)$.*

4. Two characteristic properties of transitive points. For any $p \in Q$, for any $f \in C(\Omega)$, for any positive integer k , and for any $\alpha > 0$, we may define

$$\phi^k(p) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{i=n} (f_k(T^i p) - f^*(p))^2$$

and

$$\psi^k(p, \alpha) = \delta^* \{n: |f_k(T^n p) - f^*(p)| > \alpha\}.$$

(4.1) *A quasi-regular point p is transitive if and only if $\phi^k(p) \rightarrow 0$ as $k \rightarrow \infty$ for every $f \in C(\Omega)$. Moreover, $\phi^k(p)$ tends to zero uniformly on any quasi-ergodic set.*

PROOF. The first assertion was proved in the course of proving (2.4), the second follows from (1.5).

From the definitions of $\phi^k(p)$ and $\psi^k(p, \alpha)$ it is easy to verify the inequalities

$$\alpha^2 \psi^k(p, \alpha) \leq \phi^k(p) \leq \alpha^2 + 2\|f\| \psi^k(p, \alpha).$$

Hence (4.1) implies the following theorem:

(4.2) *A quasi-regular point p is transitive if and only if $\psi^k(p, \alpha) \rightarrow 0$ as $k \rightarrow \infty$, for every $\alpha > 0$ and for every $f \in C(\Omega)$. The convergence is uniform on any quasi-ergodic set.*

The obvious analogy of the conditions stated in (4.1) and (4.2) to mean convergence and convergence in measure suggests the following definitions:

DEFINITION. *A sequence $\{F_k(n)\}$ of functions of an integer n converges in the mean if the limit $F(n) = \lim_k F_k(n)$ exists for every n , and the mean square difference*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{n=N} (F_k(n) - F(n))^2$$

exists for every k and tends to zero as $k \rightarrow \infty$.

DEFINITION. *A sequence $\{F_k(n)\}$ of functions of an integer n converges in density if the limit $F(n) = \lim_k F_k(n)$ exists for every n , and $\delta^* \{n: |F_k(n) - F(n)| > \alpha\} \rightarrow 0$ as $k \rightarrow \infty$, for every $\alpha > 0$.*

(4.3) *In order that a point $p \in \Omega$ be transitive it is necessary that for every $f \in C(\Omega)$ the sequence $f_k(T^n p)$, regarded as a sequence of functions of n , should converge in density and in the mean, and it is sufficient that convergence in either sense should hold for a generating sequence of functions in $C(\Omega)$.*

We remark that in general convergence in the mean and convergence in density are independent conditions, but that they are equivalent for any bounded convergent sequence such that the mean square difference exists for every k .

5. Uniquely ergodic and strictly ergodic systems. We shall call a system (Ω, T) *uniquely ergodic* if it has a unique normalized invariant Borel measure, or equivalently, if it has only one ergodic set. Following Nemyckii and Stepanov [23], a system (Ω, T) is called *strictly ergodic* if Ω consists of a single ergodic set. (Note that this usage differs slightly from that of Kryloff and Bogoliouboff [16].) Obviously any strictly ergodic system is uniquely ergodic, and any uniquely ergodic system is ergodic with respect to its unique measure.

(5.1) *If (Ω, T) is uniquely ergodic with measure μ , then for any $f \in C(\Omega)$ the sequence $\{f_k(p)\}$ converges uniformly on Ω to $\int f d\mu$, as $k \rightarrow \infty$.*

PROOF. Suppose the conclusion false. Then for some $g \in C(\Omega)$ and some number $\alpha \neq \int g d\mu$ there exists a sequence $\{p_i\} \subset \Omega$ and a sequence $\{k_i\}$ of positive integers such that $k_i \rightarrow \infty$ and $g_{k_i}(p_i) \rightarrow \alpha$. We may assume in addition that the $\lim_i f_{k_i}(p_i)$ exists for every f

belonging to some countable dense set in $C(\Omega)$. It follows that the limit $L(f) = \lim_i f_{k_i}(p_i)$ exists for every $f \in C(\Omega)$ and defines a positive linear functional on $C(\Omega)$. By the Riesz theorem, $L(f) = \int f d\nu$ for some normalized Borel measure ν , and ν is invariant since $L(fT) = L(f)$, hence $\nu = \mu$. In particular, taking $f = g$, we find that $L(g) = \int g d\mu$, and therefore $\alpha = \int g d\mu$, contrary to hypothesis.

It follows that in a uniquely ergodic system $Q_T = Q = \Omega$, $Q_D = R$, and Ω consists of a single quasi-ergodic set. In this case R is closed, since $\Omega - Q_D$ is the union of all open sets of measure zero, and because R consists of a single ergodic set it follows that R is a minimal set [16]. (A nonempty compact set is called *minimal* if it is equal to the orbit closure of each of its points.) There can be no other minimal set, since (2.1) and (2.5) imply that R intersects every minimal set, hence:

(5.2) *A uniquely ergodic system has only one minimal set and only one ergodic set and these sets coincide, both being equal to R .*

A number of partial converses of (5.1) can be stated. It is convenient to formulate a chain of equivalent conditions.

(5.3) *In a compact system (Ω, T) the following conditions are pairwise equivalent:*

- (i) (Ω, T) is uniquely ergodic,
- (ii) for each $f \in C(\Omega)$, $f_k(p)$ converges uniformly on Ω to a constant,
- (iii) for each $f \in C(\Omega)$ there is a sequence $\{k_i\}$ of positive integers such that $f_{k_i}(p)$ converges pointwise on Ω to a constant,
- (iv) Ω contains only one minimal set, and for each $f \in C(\Omega)$ the sequence $\{f_k(p)\}$ converges uniformly on Ω ,
- (v) R is a minimal set.

PROOF. That (i) implies (ii), (iv), and (v) follows from (5.1) and (5.2). The implications (ii) \rightarrow (iii) \rightarrow (i) are obvious. Any compact invariant set contains a minimal set, consequently if Ω contains only one minimal set then every invariant continuous function is constant, hence (iv) implies (ii). That (v) implies (i) is shown by the following theorem.

(5.4) *A minimal set is either an ergodic set or else it contains at least one non quasi-regular point.*

PROOF. Suppose that every point of a minimal set E is quasi-regular. For each $f \in C(\Omega)$ the limit function f^* is defined at every point of E and constant on each orbit. Since the orbit of each point of E is dense in E , it follows that f^* is either constant on E or else the contraction $f^*|_E$ is everywhere discontinuous. The latter alternative cannot occur, since $f^*|_E$ is the limit of a convergent sequence of continuous functions on a compact set. Consequently f^* is constant

on E and each point of E has the same corresponding measure μ_p , equal to μ , say. E contains at least one regular point p_0 , and the orbit of p_0 is dense in E , therefore μ is ergodic and positive for every open set that intersects E . Hence every point of E belongs to the ergodic set E_μ corresponding to μ , and since $\Omega - E$ is open and $\mu(\Omega - E) = 0$ it follows that $E = E_\mu$.

(5.5) *In a compact system (Ω, T) the following conditions on a point $p \in \Omega$ are pairwise equivalent:*

(i) *for each $f \in C(\Omega)$ the sequence $f_k(T^n p)$ converges uniformly in n as $k \rightarrow \infty$,*

(ii) *for each $f \in C(\Omega)$ the functions $f_k(q)$ ($k = 1, 2, \dots$) are equiuniformly continuous on $\bar{O}(p)$ (=the orbit closure of p),*

(iii) *the subsystem $(\bar{O}(p), T)$ is uniquely ergodic.*

PROOF. (i) implies that for all sufficiently large k , $f_k(q)$ is nearly constant on $\bar{O}(p)$, hence (i) implies (ii). (ii) implies that some subsequence $\{f_{k_i}(q)\}$ converges uniformly on $\bar{O}(p)$ to a constant [9, p. 304], hence (ii) implies (iii). That (iii) implies (i) follows from (5.1).

We shall call a point p *strictly transitive* if it satisfies condition (i) of (5.5). (4.2) shows that any such point is transitive, but the converse is not generally true (see §10 for a counterexample). Following Hedlund [15] a point p is called *almost periodic* if for every neighborhood U of p the set $\{n: T^n p \in U\}$ is relatively dense. According to a theorem of Gottschalk [8], $\bar{O}(p)$ is a minimal set if and only if p is almost periodic. From this result, together with (5.2) and (5.5), it is easy to deduce the following theorem:

(5.6) *In a compact system (Ω, T) the following conditions on a point $p \in \Omega$ are pairwise equivalent:*

(i) *$\bar{O}(p)$ is an ergodic set,*

(ii) *the subsystem $(\bar{O}(p), T)$ is strictly ergodic,*

(iii) *p is almost periodic and strictly transitive,*

(iv) *p is regular and strictly transitive,*

(v) *p belongs to a closed ergodic set.*

A system (Ω, T) is called *uniformly-L-stable* ("stable in the sense of Liapounov" [6]) if the powers of T are equiuniformly continuous on Ω , that is, if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d(p, q) < \delta$ implies $d(T^n p, T^n q) < \epsilon$ for every integer n .

(5.7) *If a compact system is uniformly-L-stable, then every point is regular and strictly transitive.*

PROOF. From uniform-L-stability it follows that every point p satisfies (5.5) (ii), and that if $q \in \bar{O}(p)$, then $p \in \bar{O}(q)$. Consequently every point is strictly transitive and its orbit closure is a minimal set.

The conclusion then follows from (5.6), and has the following corollary:

(5.8) *If a compact system (Ω, T) is uniformly- L -stable and has at least one dense orbit, then the system is strictly ergodic.*

It is well known [13] that any system that fulfills the hypothesis of (5.8) is homeomorphic to a system (G, T) , where G is a (monothetic) compact topological group and T is a translation of G by an element whose powers are dense in G . Hence a proof of (5.8) can also be based on the uniqueness of Haar measure.

6. Uniform ergodic limits. An important difference between the Birkhoff ergodic theorem (1.1) and the mean ergodic theorem (1.3) is that the latter implies that for any $f \in L_q$ the averages

$$f_{n,m}(p) = \frac{1}{n-m} \sum_{i=m+1}^{i=n} f(T^i p) \quad (n > m)$$

converge (in L_q) as $n-m \rightarrow \infty$, in other words, the sequence $f_k(T^n p)$ converges (in L_q) *uniformly in n* as $k \rightarrow \infty$. On the other hand, (1.1) implies only that this sequence converges (a.e.) as $k \rightarrow \infty$ for *any fixed value of n* , and simple examples show that in general there is no uniformity in n , even for strictly ergodic systems. (For example, let T be an irrational rotation of the circle, let U_n be the union of the first 2^{n+1} images of some interval of length 4^{-n-1} , and let $f(p)$ be the characteristic function of the set $U = \bigcup_1^\infty U_n$. For every point p and positive integer k there is a positive integer n such that $f_k(T^n p) = 1$, but for almost all p the $\lim_k f_k(p) = m(U) \leq 1/2$. Consequently the convergence of $f_k(T^n p)$ is nonuniform in n for almost all p .) Nevertheless, it is reasonable to ask for what compact systems it can be asserted that for any continuous function f the averages $f_{n,m}(p)$ converge a.e. as $n-m \rightarrow \infty$. An answer to this question can be formulated as follows:

(6.1) *In any compact system (Ω, T) with normalized invariant Borel measure μ the following conditions are pairwise equivalent:*

- (i) *for each $f \in C(\Omega)$, $\{f_{n,m}(p)\}$ converges a.e. as $n-m \rightarrow \infty$,*
- (ii) *almost all points are strictly transitive,*
- (iii) *almost all points belong to closed ergodic sets.*

PROOF. (i) implies that there is a set E with $\mu(E) = 1$ such that for each $p \in E$ and for each f belonging to a countable dense subset of $C(\Omega)$ the sequence $f_k(T^n p)$ converges uniformly in n as $k \rightarrow \infty$. It follows that the same conclusion holds for every $f \in C(\Omega)$, hence every point of E is strictly transitive, and (i) implies (ii). (ii) implies (iii) by (2.5) and (5.6). The reverse implications are obvious.

(6.2) In any compact system (Ω, T) the following conditions are pairwise equivalent:

- (i) for any $f \in C(\Omega)$ and for any finite invariant Borel measure μ , $\{f_{n,m}(p)\}$ converges a.e. as $n-m \rightarrow \infty$,
- (ii) every ergodic set is closed,
- (iii) the family of minimal sets and the family of ergodic sets coincide.

The following definition was introduced by Fomin [6].

DEFINITION. A compact system (Ω, T) is mean-L-stable ("stable in the mean in the sense of Liapounov") if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d(p, q) < \delta$ implies $d(T^n p, T^n q) < \epsilon$ for all n except a set of upper density less than ϵ .

(6.3) If a compact system is mean-L-stable, then for each $f \in C(\Omega)$ the sequence $\{f_k(p)\}$ is equiuniformly continuous and uniformly convergent on Ω , every ergodic set is closed, every quasi-ergodic set is closed, and every point is strictly transitive.

PROOF. The equiuniform continuity of the functions $f_k(p)$ ($k = 1, 2, \dots$) is an easy consequence of mean-L-stability. From (5.5)(ii) it follows that every point is strictly transitive, and the uniform convergence of $\{f_k(p)\}$ then follows from equicontinuity and pointwise convergence. Consequently $f^* \in C(\Omega)$ whenever $f \in C(\Omega)$, and therefore every quasi-ergodic set is closed. That every ergodic set is closed follows from (5.6)(iv).

The next two theorems are corollaries of (6.3).

(6.4) If a compact mean-L-stable system has at least one dense orbit, then it is uniquely ergodic.

(6.5) If a minimal set is mean-L-stable, then it is strictly ergodic (cf. [6]).

The last three results should be compared with (5.7) and (5.8). Trivial examples (e.g. a system consisting of two invariant points p and q and an orbit $\{p_n\}$ such that $p_n \rightarrow p$ and $p_{-n} \rightarrow q$) show that mean-L-stability is not a necessary condition for every ergodic set to be closed, and (same example with $p=q$) that the hypothesis of (6.4) does not imply the conclusion of (6.5). Interesting examples of minimal sets that are mean-L-stable but not uniformly-L-stable are provided by the Sturmian minimal sets studied by Hedlund [15].

7. Noncompact systems. If T is a homeomorphism of a complete separable metric space Ω , the system (Ω, T) need not admit a finite invariant Borel measure. However, Fomin [4] has shown that in case such a measure exists, the main theorems of Kryloff and Bogoliouboff remain true provided certain definitions are appropriately generalized. (For a generalization in another direction, see

[29].) In this section we shall show how the study of such systems can be reduced to that of compact systems. At the same time a slightly greater degree of generality is attained, since it turns out to be unnecessary to postulate completeness.

By a *Borel system* we shall mean a system (Ω, T) where Ω is a Borel subset of some complete separable metric space, and T is a homeomorphism of Ω onto itself. Two such systems (Ω, T) and (Ω', T') are called *homeomorphic* if there exists a homeomorphism of Ω onto Ω' under which T corresponds to T' . (Ω', T') is called a *subsystem* of (Ω, T) in case Ω' is a Borel subset of Ω invariant under T .

(7.1) *Any Borel system (Ω, T) is homeomorphic to a subsystem of the compact system (Ω^*, S) where Ω^* is the Cartesian product of \aleph_0 Hilbert cubes, and S is the shift transformation on Ω^* , i.e., $S\{\omega_n\} = \{\omega_{n+1}\}$.*

PROOF. By Urysohn's theorem, we may assume that Ω itself is a subset of the Hilbert cube. Then the correspondence $p \rightarrow \{T^n p\}$ defines a 1:1 bicontinuous mapping of Ω onto a Borel subset Ω' of Ω^* under which T corresponds to S .

It is clear that the invariant Borel measures of the system (Ω, T) correspond to those of the system (Ω^*, S) for which $\mu(\Omega^* - \Omega') = 0$. Consequently, a necessary and sufficient condition that (Ω, T) admit no finite invariant Borel measure is that the set Ω' have invariant measure zero. This will be the case if and only if every compact subset of Ω' has invariant measure zero, and for this it is necessary and sufficient that $M(\chi_{K'}, p') = 0$ for every compact set $K' \subset \Omega'$ and for every $p' \in \Omega'$, by (2.1) and (1.2). Hence we obtain the following generalization of (2.1) (cf. [26]).

(7.2) *A necessary and sufficient condition that a Borel system (Ω, T) admit no finite invariant Borel measure is that $M(\chi_K, p) = 0$ for every compact set $K \subset \Omega$ and for every point $p \in \Omega$.*

Let us now denote by $C(\Omega)$ the space of all bounded real-valued continuous functions on Ω . The definition of a quasi-regular point requires modification, since it is no longer true that $M(f, p)$ can always be represented as an integral. We follow Fomin [4] in adopting the following definition:

DEFINITION. *A point $p \in \Omega$ is quasi-regular with respect to the Borel system (Ω, T) if*

- (1) *the mean value $M(f, p)$ exists for each $f \in C(\Omega)$, and*
- (2) *for every $\epsilon > 0$ there is a compact set $K \subset \Omega$ such that $M(\chi_K, p) > 1 - \epsilon$.*

It is clear that this definition is equivalent to that of Kryloff and Bogoliouboff (cf. §2) in case the system is compact. That it is an

appropriate generalization is shown by the following theorem.

(7.3) *If (Ω', T) is a subsystem of a compact system (Ω^*, T) , then a point $p \in \Omega'$ is quasi-regular with respect to (Ω', T) if and only if p is a quasi-regular point of (Ω^*, T) and $\mu_p(\Omega') = 1$.*

PROOF. The necessity of both conditions is clear; that they are sufficient is shown by (3.3) and (3.5).

It follows that with any quasi-regular point of the Borel system (Ω, T) there is associated uniquely a normalized invariant Borel measure μ_p such that $\int f d\mu_p = M(f, p)$ for every $f \in C(\Omega)$, and that the quasi-regular points are the only points for which $M(f, p)$ is defined on $C(\Omega)$ and representable in this way. We remark that (2) cannot be weakened to the requirement that $M(\chi_K, p) > 0$ for some compact set K , since it is possible for a quasi-regular point p in Ω^* to be such that $0 < \mu_p(\Omega') < 1$.

The set Q' of quasi-regular points of (Ω^*, T) for which $\mu_p(\Omega') = 1$ is a Borel subset of Ω^* , by (2.7), and therefore the set Q of quasi-regular points of (Ω, T) is a Borel subset of Ω . If E is any ergodic set of (Ω^*, T) , with ergodic measure μ , then either $\mu(E \cap \Omega') = 1$ or $\mu(E \cap \Omega') = 0$; hence $\mu(\Omega' - Q') = 0$. It follows from (2.8) that $\Omega' - Q'$ has invariant measure zero, and therefore Q has invariant measure one relative to (Ω, T) . Similar reasoning shows that if points of density, transitive points, and regular points are defined as in §2, then not only (2.2) but also (2.3), (2.4), and (2.5) hold for any Borel system. The remaining theorems of §2 then follow at once from the corresponding theorems for compact systems. It should be emphasized, however, that all of these theorems are vacuous unless the Borel system admits at least one finite invariant measure.

8. Decomposition into a direct sum of ergodic systems. One of the most fruitful conceptions in ergodic theory has been the idea that any system can be regarded as a combination, in some sense, of ergodic systems. The first result of this sort was obtained by von Neumann [24]. For compact systems, the theory of Kryloff and Bogoliouboff gave a still more detailed analysis. Subsequently, Halmos, Ambrose, and Kakutani [10; 2] obtained von Neumann's theorem in a purely measure-theoretic form, with no explicit topological assumptions. It was for this purpose that Halmos introduced the notion of a direct sum of measure spaces. More recently, the work of Dieudonné [3; 11], Nikodým [25], and Maharam [18] on Boolean σ -algebras has led to a more precise version of the decomposition theorem for measure spaces, in which some of the direct summands are combined into a direct product. In this section and the next it will be shown how some of these decomposition theorems for

measure spaces can also be obtained from the ergodic set decomposition.

Let (Ω, T) be a compact system, and let R be the set of regular points. Let f be a Borel measurable mapping of R into the unit interval, such that $f(p) = f(q)$ if and only if $\mu_p = \mu_q$ (cf. (3.1)), and let X denote the range of f . Let g be a 1:1 Borel measurable mapping of R onto the unit interval [17, p. 358]. Then the equations $x = f(p)$, $y = g(p)$ define a 1:1 Borel measurable mapping ϕ of R onto a Borel subset Y of the unit square, such that for each $x \in X$ the ergodic set $f^{-1}(x)$ is mapped onto the x -section Y_x of Y . Let ν_x denote the Borel measure in Y_x corresponding to the ergodic measure in $f^{-1}(x)$. Let \mathcal{X} be the Borel field consisting of all sets $A \subset X$ such that $f^{-1}(A)$ is a Borel subset of R . Corresponding to any invariant Borel measure μ in Ω define $\nu(A) = \mu f^{-1}(A)$, for every $A \in \mathcal{X}$. Then for any Borel set $E \subset Y$ we have, by (2.7),

$$\begin{aligned} \mu(\phi^{-1}E) &= \int_R \mu_p(\phi^{-1}E) d\mu(p) \\ &= \int_R \nu_{f(p)}(E \cdot Y_{f(p)}) d\mu(p) = \int_X \nu_x(E \cdot Y_x) d\nu(x). \end{aligned}$$

Thus ϕ represents the measure space (R, μ) as a direct sum of the measure spaces (Y_x, ν_x) over (X, \mathcal{X}, ν) . T corresponds to a Borel measurable transformation $T' = \phi T \phi^{-1}$ of Y that leaves invariant each x -section Y_x , and on Y_x is ergodic with respect to ν_x . Thus ϕ represents (R, T, μ) as a direct sum of ergodic systems. It should be noted that *the same decomposition serves for every finite invariant Borel measure μ* , the only variable term being the measure function ν . Moreover, every finite measure ν on \mathcal{X} corresponds in this way to one and only one invariant measure μ . For if E is any Borel subset of R , then $\mu_p(E)$ is measurable relative to $f^{-1}(\mathcal{X})$, hence $\nu_x(\phi(E) \cdot Y_x)$ is measurable \mathcal{X} , and for any finite measure ν on \mathcal{X} the equation

$$\mu(E) = \int_X \nu_x(\phi(E) \cdot Y_x) d\nu(x)$$

defines a corresponding invariant Borel measure in R . A 1:1 correspondence is thus established between the invariant Borel measures μ of the system (Ω, T) and the finite measure functions ν on the measurable space (X, \mathcal{X}) .

This correspondence between μ and ν determines at the same time a correspondence between their completions μ^* and ν^* , and ϕ represents (R, μ^*) as a direct sum of the measure spaces (Y_x, ν_x^*) over

(X, ν^*) , in the sense that if E is measurable μ^* then $\phi(E)$ is measurable ν_x^* for all x except a set of ν^* -measure zero, and $\nu_x^*(\phi(E) \cdot Y_x)$ is a ν^* -measurable function of x whose integral is equal to $\mu^*(E)$.

Now let T be any 1:1 measure-preserving transformation on a normal measure space (Ω, m^*) . (A normal measure space is one that is isomorphic to the unit interval, cf. [13]. Two measure spaces are said to be isomorphic if there exists a 1:1 measure preserving transformation from almost all of one to almost all of the other. Two systems (Ω, T, μ) and (Ω', T', μ') are isomorphic if T corresponds to T' under some isomorphism.) We are seeking an isomorphic representation of (Ω, T, m^*) , hence we may assume from the outset that Ω is the unit interval and that m^* is Lebesgue measure. Moreover, up to isomorphism, T can be replaced by an equivalent 1:1 Borel measurable transformation. For each $p \in \Omega$, the sequence of numbers $\{T^n p\}$ may be regarded as a point $\phi_1(p)$ of the \aleph_0 -dimensional cube Ω^∞ , and then ϕ_1 is a 1:1 Borel measurable mapping of Ω onto a Borel subset of Ω^∞ (cf. [17, pp. 290, 397]). Under ϕ_1 , the transformation T corresponds to the shift transformation S , which is a homeomorphism on Ω^∞ . If μ^* denotes the completed Borel measure in Ω^∞ that corresponds to m^* under ϕ_1 , then the system (Ω, T, m^*) is isomorphic to $(\Omega^\infty, S, \mu^*)$. We have already shown how the latter system can be represented as a direct sum of ergodic systems. In this way the following theorem of Ambrose, Halmos, and Kakutani [2] is obtained as a consequence of the ergodic set decomposition.

(8.1) *Any normal system (Ω, T, m^*) is isomorphic to a direct sum of ergodic systems.*

We remark that if T is a 1:1 Borel measurable transformation on a Borel subset Ω of a complete separable metric space, the system (Ω, T) can be mapped by a 1:1 Borel measurable mapping onto a subsystem (Ω', T') of a compact system by the device used in §7. The representation of a compact system described above yields a corresponding representation of (Ω, T) as a direct sum of ergodic systems, *simultaneously for every finite invariant Borel measure*. In other words, the possibility of a universal decomposition into ergodic parts is not limited to compact systems, but extends to any system that is "Borel isomorphic" to a subsystem of a compact system.

It should also be remarked that the possibility of representing a normal system by an isomorphic continuous transformation is not limited to systems with a discrete time; Ambrose and Kakutani [1] have shown that any measurable flow in a normal measure space is isomorphic to a continuous flow.

9. Direct product representation. The following theorem is the principal special case of a theorem of Maharam [18, Theorem 6]. The proof given here was suggested by a somewhat similar proof obtained by Kakutani.

(9.1) *If (Ω, T, m^*) is a normal system such that (1) no subsystem is ergodic relative to m^* , and (2) almost all points are nonperiodic, then (Ω, T, m^*) is isomorphic to a system (Ω', T', m_2^*) , where Ω' is the unit square with plane measure m_2^* , and T' is a transformation that leaves every x -section of Ω' invariant and on it is ergodic with respect to linear measure.*

It should be remarked that both conditions (1) and (2) are obviously necessary, but that neither of them is seriously restrictive. Without them, it is not hard to see that the system can be decomposed into a direct sum of systems all but one of which are comparatively trivial, namely, either periodic or ergodic on an interval. The remaining summand, if present, is covered by the above theorem.

In §8 it was shown that any normal system is isomorphic to a compact system $(\Omega^\infty, S, \mu^*)$. Under the added hypotheses (1) and (2) each ergodic set of the latter system has μ^* -measure zero, and the set of periodic points has μ^* -measure zero. Hence (9.1) is a consequence of the following slightly more precise theorem.

(9.2) *If (Ω, T, μ) is a compact system with normalized invariant Borel measure μ , and if each ergodic set and the set P of periodic points has μ -measure zero, then there exists a 1:1 Borel measurable mapping of almost all of R onto a Borel subset W of the unit square such that (i) μ corresponds to plane Borel measure m_2 in W , (ii) T corresponds to a transformation of W that leaves every x -section of W invariant and on it is ergodic with respect to linear Borel measure m , and (iii) m is the only normalized invariant Borel measure on any x -section of W .*

PROOF. We start with the direct sum representation already obtained in §8, but restrict ϕ to the set $R_0 = R - P$. Under present hypotheses the measures ν and ν_x vanish on every finite subset of X and Y_x respectively. For any $(x, y) \in Y$ define

$$F(x) = \nu(X \cdot [0, x))$$

and

$$G(x, y) = \nu_x(Y_x \cdot [0, 1] \times [0, y)).$$

Then $F(x)$ is a monotone increasing continuous function of x on X , and for each $x \in X$, $G(x, y)$ is a monotone increasing continuous function of y on Y_x . Both functions are Borel measurable on Y , the first because it is continuous, and the second because it corresponds

under ϕ to the Borel measurable function $\mu_p(g^{-1}[0, g(p)])$ on R_0 (cf. [17, p. 285]). Consequently the equations $x' = F(x)$, $y' = G(x, y)$ define a Borel measurable mapping ψ of Y onto a subset Z of the unit square. ψ is not necessarily 1:1, but we shall show that it can be made so by suitably contracting its domain. We first show that it preserves measure.

Consider any rectangle $J = [0, a) \times [0, b)$, where $0 < a < 1$ and $0 < b < 1$. Let $a' = \sup \{x : F(x) < a\}$, and for each $x \in X$ let $b'(x) = \sup \{y : (x, y) \in Y_x, G(x, y) < b\}$. Then $\psi^{-1}J$ consists of all points $(x, y) \in Y$ such that $0 \leq x < a'$ and $0 \leq y < b'(x)$. For any $x \in X$, using the fact that each point has ν_x -measure zero, we have $\nu_x(Y_x \cdot [0, 1] \times [0, b'(x))) = b$. Consequently $\mu(\psi^{-1}J) = \int_{X \cdot [0, a')} (\nu_x \cdot [0, 1] \times [0, b'(x))) d\nu(x) = b \cdot \nu(X \cdot [0, a')) = b \cdot a = m_2(J)$. Let \mathcal{E} be the class of Borel subsets E of the square such that $\mu(\psi^{-1}E) = m_2(E)$. We have just shown that \mathcal{E} contains all rectangles J , and since \mathcal{E} contains proper differences, countable decreasing intersections, and countable disjoint unions of its members, it follows that \mathcal{E} includes all Borel subsets of the square [12, §5(2), §6(3)].

Similarly, if x' is any value for which $x = F^{-1}(x')$ is single-valued, and if \mathcal{E}_x denotes the class of Borel subsets E of $\{x'\} \times [0, 1]$ such that $\nu_x(\psi^{-1}E) = m(E)$, then \mathcal{E}_x includes all intervals $\{x'\} \times [0, b)$, $0 < b < 1$, and has closure properties similar to those of \mathcal{E} . It follows that $\nu_x(\psi^{-1}E) = m(E)$ for every Borel subset E of $Z_{x'}$.

Let Z_0 be the set of all points of Z where ψ^{-1} is multiple-valued. The monotoneity of F and G shows that for all but countably many values of x' , $F^{-1}(x')$ is a single x -section of Y , and that on any such x' -section of Z , Z_0 is at most countable. Since Z_0 is an analytic set [17, p. 402] it follows that $m_2^*(Z_0) = 0$. Enclose Z_0 in a Borel set B_0 with $m_2(B_0) = 0$, and choose B_0 so that every x' -section either contains $Z_{x'}$ or has linear measure zero. (It follows that B_0 includes all x' -sections $Z_{x'}$ for which $f^{-1}(x')$ is multiple-valued.) Put $Y_1 = Y - \psi^{-1}(B_0)$ and $Y_2 = \bigcap_{-\infty}^{\infty} (T')^n Y_1$. Then Y_2 is a Borel subset of Y with $\mu(Y_2) = 1$, and ψ is a 1:1 Borel measurable mapping of Y_2 onto a Borel subset $Z_2 = \psi(Y_2)$ of the unit square. Define $T'' = \psi T' \psi^{-1}$ on Z_2 . Then the systems (Y_2, T', μ) and (Z_2, T'', m_2) correspond under ψ , and likewise $((Y_2)_x, T', \nu_x)$ and $((Z_2)_{x'}, T'', m)$ for every x for which $(Y_2)_x$ is nonempty. Moreover, m is the only normalized invariant Borel measure in $(Z_2)_{x'}$, since ν_x is the only such measure in $(Y_2)_x$. To complete the proof of (9.2) it suffices to follow ψ by a trivial 1:1 Borel measurable transformation of the square that involves only the values of x' on a nullset and maps Z_2 onto a set W with the property that W_x is nonempty for every x in $0 \leq x \leq 1$.

10. **The shift transformation on sequence space.** Let X be the set of all mappings $x(n)$ of the integers into the set $\{0, 1\}$, and let T be the transformation that takes $x(n)$ into $x(n+1)$. If we define $d(x, y) = \max \{1/(|n| + 1) : x_n \neq y_n\}$ for any two points $x \neq y$, then X is a compact metric space and T is a homeomorphism. In this system it is easy to characterize points of the various kinds discussed in §§2 and 5. For instance, a point x is quasi-regular if and only if each block of terms that appears in the sequence $x(n)$ recurs with a definite limiting frequency; x is strictly transitive if and only if for each block B and $\epsilon > 0$ there is a positive integer k such that in any sequence of k successive terms of $x(n)$, B occurs with a frequency that differs by less than ϵ from the frequency with which B occurs in any other sequence of k successive terms of $x(n)$. On the basis of (4.3) it is easy to formulate similar characterizations of transitive points and points of density.

The almost periodic points of this system are the same as *recurrent sequences* in the sense of Birkhoff (cf. [15]). A number of examples and classes of such sequences have been defined and studied by Morse [19], Morse and Hedlund [20; 21; 22], Garcia and Hedlund [7], and Robbins [27]. In most of these examples it is not hard to verify that the sequence defines a point that is not only almost periodic but also strictly transitive. Hence the orbit closure of such a point is a minimal set on which T is strictly ergodic. However, it is also possible to exhibit a minimal set that is not strictly ergodic. The first such example was constructed by Markhoff (cf. [23, p. 533]). We shall give a similar example in the system (X, T) . For this purpose it suffices to define a point x that is almost periodic but not quasi-regular (cf. (5.4)).

Let $\{k_i\}$ ($i \geq 0$) be a sequence of positive integers such that k_i divides k_{i+1} and such that $\sum_{i=1}^{\infty} k_{i-1}/k_i \leq 1/12$ (e.g., the sequence $k_i = 2^{i(i+9)/2}$). Let n and m denote variable integers, and for each positive integer i define

$$E_i = \bigcup_{-\infty < m < \infty} \{n : |n - mk_i| \leq k_{i-1}\}.$$

Since E_i includes all n with $|n| \leq k_{i-1}$, and since $k_i \rightarrow \infty$, it is clear that for each integer n there is a least positive integer $p = p(n)$ such that n belongs to E_p . Define $x(n) = 0$ or 1 according as $p(n)$ is even or odd. We shall show that this point x has the required properties.

In the first place, each of the sets E_1, E_2, \dots, E_p is invariant under translation by k_p . Hence if $p(n) = p$ and $m \equiv n \pmod{k_p}$ then $p(m) = p$. Consequently, if $m \equiv n \pmod{k_i}$ and $|n| \leq k_{i-1}$, then

$p = p(n) \leq i$, k_p divides k_i , $m \equiv n \pmod{k_p}$, $p(m) = p(n)$, and therefore $x(m) = x(n)$. It follows that $d(x, T^n x) \leq 1/(1+k_{i-1})$ for all $n \equiv 0 \pmod{k_i}$. This shows that x is almost periodic (in fact, *regularly almost periodic*, as defined in [7, p. 957]).

Secondly, if $1 \leq j \leq i$ the number of elements of E_j in the interval $0 < n \leq k_i$ is exactly $(k_i/k_j)(2k_{j-1}+1)$. Hence an upper bound to the number of elements of $E_1 \cup E_2 \cup \dots \cup E_i$ in the interval $0 < n \leq k_i$ is

$$\sum_{j=1}^i \frac{3k_i k_{j-1}}{k_j} \leq 3k_i \sum_{j=1}^{\infty} \frac{k_{j-1}}{k_j} < \frac{1}{4} k_i.$$

It follows that $p(n) = i+1$ for at least 3/4 of the numbers n in the interval $0 < n \leq k_i$. Consequently,

$$\left| \frac{1}{k_i} \sum_{n=1}^{k_i} x(n) - \frac{1}{k_{i+1}} \sum_{n=1}^{k_{i+1}} x(n) \right| \geq \frac{1}{2} \quad (i = 1, 2, \dots).$$

But $x(n) = \chi(T^n x)$, where χ denotes the characteristic function of the set of points $y \in X$ for which $y(0) = 1$, and therefore $M(\chi, x)$ is undefined. Since χ is a continuous function on X this shows that x is not quasi-regular.

BIBLIOGRAPHY

1. W. Ambrose and S. Kakutani, *Structure and continuity of measurable flows*, Duke Math. J. vol. 9 (1942) pp. 25-42.
2. W. Ambrose, P. R. Halmos, and S. Kakutani, *The decomposition of measures*, II, Duke Math. J. vol. 9 (1942) pp. 43-47.
3. J. Dieudonné, *Sur le théorème de Lebesgue-Nikodým (III)*, Annales de l'Université de Grenoble vol. 23 (1948) pp. 25-53.
4. S. Fomin, *Finite invariant measures in the flows*, Rec. Math. (Mat. Sbornik) N.S. vol. 12 (1943) pp. 99-108.
5. ———, *On measures invariant under certain groups of transformations*, Izvestiya Akad. Nauk SSSR, Ser. Mat. vol. 14 (1950) pp. 261-274.
6. ———, *On dynamical systems with pure point spectrum*, Doklady Akad. Nauk SSSR. vol. 77 (1951) pp. 29-32.
7. M. Garcia and G. A. Hedlund, *The structure of minimal sets*, Bull. Amer. Math. Soc. vol. 54 (1948) pp. 954-964.
8. W. H. Gottschalk, *Almost periodic points with respect to transformation semi-groups*, Ann. of Math. vol. 47 (1946) pp. 762-766.
9. H. Hahn, *Theorie der reellen Funktionen*, I, Berlin, 1921.
10. P. R. Halmos, *The decomposition of measures*, Duke Math. J. vol. 8 (1941) pp. 386-392.
11. ———, *On a theorem of Dieudonné*, Proc. Nat. Acad. Sci. U.S.A. vol. 35 (1949) pp. 38-42.
12. ———, *Measure theory*, New York, 1950.
13. P. R. Halmos and J. von Neumann, *Operator methods in classical mechanics*, II, Ann. of Math. vol. 43 (1942) pp. 332-350.

14. P. Hartman and A. Wintner, *Asymptotic distributions and the ergodic theorem*, Amer. J. Math. vol. 61 (1939) pp. 977–984.
15. G. A. Hedlund, *Sturmian minimal sets*, Amer. J. Math. vol. 66 (1944) pp. 605–620.
16. N. Kryloff and N. Bogoliouboff, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. of Math. vol. 38 (1937) pp. 65–113.
17. C. Kuratowski, *Topologie I*, 2d ed., Warszawa-Wrocław, 1948.
18. D. Maharam, *Decompositions of measure algebras and spaces*, Trans. Amer. Math. Soc. vol. 69 (1950) pp. 142–160.
19. M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. vol. 22 (1921) pp. 84–100.
20. M. Morse and G. A. Hedlund, *Symbolic dynamics*, Amer. J. Math. vol. 60 (1938) pp. 815–866.
21. ———, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. vol. 62 (1940) pp. 1–42.
22. ———, *Unending chess, symbolic dynamics and a problem in semigroups*, Duke Math. J. vol. 11 (1944) pp. 1–7.
23. V. V. Nemyckii and V. V. Stepanov, *Qualitative theory of differential equations*, 2d ed., Moscow, 1949.
24. J. von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Ann. of Math. vol. 33 (1932) pp. 587–642.
25. O. M. Nikodým, *Tribus de Boole et fonctions mesurables*, C. R. Acad. Sci. Paris vol. 228 (1949) pp. 150–151.
26. J. C. Oxtoby and S. M. Ulam, *On the existence of a measure invariant under a transformation*, Ann. of Math. vol. 40 (1939) pp. 560–566.
27. H. E. Robbins, *On a class of recurrent sequences*, Bull. Amer. Math. Soc. vol. 43 (1937) pp. 413–417.
28. S. Saks, *Integration in abstract metric spaces*, Duke Math. J. vol. 4 (1938) pp. 408–411.
29. K. Yosida, *Simple Markoff process with a locally compact phase space*, Mathematica Japonicae vol. 1 (1948) pp. 99–103.

BRYN MAWR COLLEGE