ON POINCARÉ'S RECURRENCE THEOREM

KONRAD JACOBS
UNIVERSITY OF ERLANGEN

1. Introduction

Let T be a continuous mapping of a Polish space Ω into itself. Assume that some randomness is introduced into Ω by a normalized measure m. Then we may distinguish between macroscopic and microscopic properties of the system; the macroscopic properties concern the behavior of m, whereas the microscopic properties concern the behavior of the individual points of Ω, under the action of T.

Poincaré's classical recurrence theorem (see, for example, Jacobs [6], p. 49, ff.) says, roughly speaking, that macroscopic stationarity implies microscopic recurrence. Here the statement concerning the behavior of the system is weakened in passing from the macroscopic hypothesis to the microscopic conclusion of the theorem. Imagine a sequence of systems, each governing by its microscopic behavior (of one of its points) the macroscopic behavior of the subsequent one. If in the first system T is the identity mapping, then the second one will be macroscopically stationary, hence (according to Poincaré) the third system will be 'macroscopically recurrent.' Poincaré's theorem does not say what the third system will do microscopically. Theorem 3.1 of the present paper asserts that also the third system will be microscopically recurrent, and that recurrence will never get lost throughout the whole dynasty of systems linked in the indicated way: recurrence is a hereditary property.

Of course we first have to make precise the concepts of macroscopic and microscopic recurrence. Section 2 of this paper is devoted to the definition and discussion of recurrence of points and measures under the action of T; for measures, weak topology is adopted, and examples and easy constructions are exhibited.

In section 3 the new recurrence theorem (theorem 3.1) is stated. We prove it in two different ways. I had proved the theorem for mixing measures (section 5) but not for the general case, when I told the problem to V. Strassen (Goettingen). After 24 hours we met again, each having produced a proof for the general case. Strassen's proof is by far the simpler one, employing the same ideas which are used for the classical Poincaré theorem; my proof needs some preparation and uses the ideas of M. Kac [7] (see also Jacobs [6], p. 55, ff.); both proofs are given here.

Section 4 gives the lattice properties of the system of all measures which are recurrent along a given sequence of time points. It is shown that 'extremal' measures are either invariant or periodic or 'wandering' under the action of T. In the latter case, no stronger or equivalent finite T-invariant measure exists. This shows that our recurrence theorem goes strictly beyond the range covered by ergodic theory with an invariant measure, and that it cannot be reduced to the classical Poincaré theorem by passing to a stronger finite invariant measure. Nonperiodic 'extremal' recurrent measures are easily obtained; indeed, it is shown in section 5 that every mixing recurrent measure is even 'universally extremal.'

In section 6, Gaussian recurrent measures are discussed. The idea of the proof of theorem 6.1 was given to me orally by S. M. Vershik in a discussion.

In section 7, recurrent Markov measures with finite state space are investigated by an extension of methods developed in Jacobs [2], [3]. And in section 8, we make an application to the theory of nonstationary noisy channels. A special case of theorem 8.3 has already been announced in Jacobs [5].

I am very much indebted to V. Strassen and A. M. Vershik for their contributions to this paper.

2. Recurrence

1. *Topological preliminaries.* Let $\Omega = \{\omega, \eta, \cdots\}$ be a Hausdorff space, and let $C(\Omega)$ denote the Banach lattice of all bounded real continuous functions on Ω with the usual supremum norm $\|f\| = \sup_{\omega\in\Omega} |f(\omega)|$. Further, let B be the σ-field generated by $C(\Omega)$, that is, the σ-field of Baire sets in Ω. Throughout this we shall assume Ω Polish, that is, separable and metrizable such as to ensure completeness, unless an exception is explicitly stated. Consequently, there will be no distinction between Baire and Borel sets.

Note that $R(B) = \{h, m, \cdots\}$ denotes the L-space of all finite σ-additive real functions on B, that is, of all electric charges in Ω. If Ω is compact, $R(B)$ is simply the dual Banach lattice of $C(\Omega)$. The set $R(B)^+$ of all finite measures on Ω is the positive cone of $R(B)$. Let $V = \{m | m \in R(B)^+, m(\Omega) = 1\}$ be the system of all normalized measures, that is, probability distributions on Ω. Being the cross-section of the convex lattice cone $R(B)^+$ with the hyperplane $\{h | \int dh = 1\}$, it is (algebraically) a simplex whose extremal points are just the normalized one-point masses $\delta_\omega (\omega \in \Omega)$. In addition to the topology given by the total variation norm,

$$\|h\| = |h|(\Omega) = \sup_{\substack{\|f\|\leq 1 \\ f\in C(\Omega)}} |\textstyle\int f\, dh|, \tag{1}$$

we consider in $R(B)$ the *weak topology* induced by $C(\Omega)$:

$$h_k \to h, \qquad \text{(weakly)} \tag{2}$$

means

$$\int f\, dh_k \to \int f\, dh, \qquad (F \in C(\Omega)). \tag{3}$$

Both $R(B)^+$ and V are weakly closed. If Ω is compact, V is weakly compact. The lattice operations, obviously norm continuous, are not weakly continuous in general ($\omega_k \to \omega$ implies $\delta_{\omega_k} \to \delta_\omega$, but $\omega_{\omega_k} \wedge \delta_\omega \to \delta_\omega \wedge \delta_\omega$ is wrong if $\delta_{\omega_k} \neq \delta_\omega$, $(k = 1, 2, \cdots)$), but enjoy some semicontinuity properties. For later use, we mention that the operation $h \to |h|$ in $R(B)$ is weakly lower semicontinuous; that is, (2) implies

$$\liminf_k \int f\, d|h_k| \geq \int f\, d|h|, \qquad (0 \leq f \in C(\Omega)). \tag{4}$$

This follows from the general formula

$$\int f\, d|h| = \sup \sum_k \left|\int f_k\, dh\right|, \tag{5}$$

where the sup goes over all finite decompositions $f = \sum_k f_k$ which fulfill $0 \leq f_k \in C(\Omega)$. Clearly, the expressions under the sup are weakly continuous, which proves the above statement.

Now let T be a continuous mapping of Ω into itself; T is B-measurable. That is, if we define the inverse image UT^{-1} of a set $U \subseteq \Omega$ by $UT^{-1} = \{\omega | \omega T \in U\}$, then

$$UT^{-1} \in B, \qquad (U \in B). \tag{6}$$

The mapping T induces norm continuous mappings—also denoted by T—of $C(\Omega)$ and of $R(B)$, each into itself, which are defined by

$$(fT)(\omega) = f(\omega T), \qquad (\omega \in \Omega, f \in C(\Omega)), \tag{7}$$

$$\int f\, d(hT) = \int fT\, dh, \qquad (f \in C(\Omega),\, h \in R(B)). \tag{8}$$

Clearly T: $R(B) \to R(B)$ is weakly continuous and norm contracting. It even commutes with all finite lattice operations, if it is one-to-one onto and has a measurable inverse. As a rule, we shall not make the latter hypothesis. Both $R(B)^+$ and V are T-invariant.

2. *Recurrent points.*

DEFINITION 2.1. *A point $\omega \in \Omega$ is called* recurrent *(with respect to T), if it returns into each of its neighborhoods; that is, if for every open U containing ω, the set*

$$\mathfrak{t}(\omega, U) = \{t | t > 0,\, \omega T^t \in U\} \tag{9}$$

of its return times *into U is nonempty. Let Ω_{rec} denote the set of all T-recurrent points in Ω.*

It is obvious that for a recurrent ω, the set $\mathfrak{t}(\omega, U)$ is always infinite.

Throughout the paper we shall denote *strictly increasing* sequences $t_0, t_1, \cdots$ of nonnegative integers by gothic letters, such as $\mathfrak{t}, \mathfrak{t}', \cdots$.

DEFINITION 2.2. *Let $\mathfrak{t}$: $t_0, t_1, \cdots$ be given. A point $\omega \in \phi$ is called* recurrent along $\mathfrak{t}$ *or* $\mathfrak{t}$-recurrent *(with respect to T) if*

$$\omega T^{t_k} \underset{k}{\to} \omega. \tag{10}$$

Let $\Omega(\mathfrak{t})$ denote the set of all points of Ω which are recurrent along $\mathfrak{t}$. The points of $\bigcup_{\mathfrak{t}} \Omega(\mathfrak{t})$ are also called sequentially recurrent.

We shall use both definitions also if Ω is an arbitrary topological space. Obviously, sequentially recurrent points are always recurrent. Our standing hypothesis that Ω be Polish implies (i) of the following lemma.

LEMMA 2.1. *Since Ω is Polish,* (i) $\Omega_{\text{rec}} = \bigcup_{\mathfrak{t}} \Omega(\mathfrak{t})$. *Further,* (ii) *for arbitrary $U \subseteq \Omega$, let $U^c = \Omega - U$ be the complement of U, and let*

$$U_{\text{ret}} = \{\omega | \omega \in U,\ \mathfrak{t}(\omega, U) \neq 0\} \tag{11}$$

be the set of all points in U which return to U. Then for every basis $\sum$ of the topology of Ω,

$$\Omega_{\text{rec}} = \bigcap_{U \in \Sigma} (U_{\text{ret}} \cup U^c). \tag{12}$$

The proof is obvious. Note that (i) depends only on the assumption that Ω has a locally countable basis, and that (ii) holds for an arbitrary topological space Ω.

COROLLARY. *If $U \subseteq \Omega$ is measurable, then U_{ret} is measurable. The set Ω_{rec} of all recurrent points is measurable.*

PROOF. Measurability of U_{ret} follows, for example, from

$$U - U_{\text{ret}} = \bigcup_{t \geq 0} UT^{-t} - \bigcup_{t \geq 1} UT^{-t}, \tag{13}$$

if U is measurable. If we choose a countable basis $\sum$ of the topology, measurability of Ω_{rec} follows from this, and from (12). If looking for examples of recurrent points, one should note:

(1) if $\mathfrak{t}$: $t_0, t_1, \cdots$ is such that $t_{k+1} = t_k + 1$ for infinitely many k, then $\Omega(\mathfrak{t})$ consists exactly of all T-fixed points;

(2) if $\mathfrak{t}$: $t_0, t_1, \cdots$ is such that $\liminf_k (t_{k+1} - t_k) = d < \infty$, then there exists a d_0 such that $0 < d_0 \leq d$, and $\Omega(\mathfrak{t})$ consists exactly of all points ω which have period d_0 under T, that is, which fulfill $\omega T^{d_0} = \omega$. If, for example, $t_k = 2k + 1$, then $d = 2$ and $d_0 = 1$;

(3) if $t_k = k!$, $(k = 0, 1, \cdots)$, then $\Omega(\mathfrak{t})$ contains all T-periodic points;

(4) passing to a subsequence increases $\Omega(\mathfrak{t})$;

(5) there may quite well be disjoint sequences $\mathfrak{t}$, $\mathfrak{t}'$ such that $\Omega(\mathfrak{t}) = \Omega(\mathfrak{t}')$. Take, for example, an irrational real $\alpha/2\pi$ and define T: $e^{i\varphi} \rightarrow e^{i(\varphi+\alpha)}$ on $\Omega = \{e^{i\varphi} | \varphi \text{ real}\}$. There is a sequence $\mathfrak{t}$: $t_k \rightarrow \infty$ of even, and also a sequence $\mathfrak{t}'$: $t'_k \rightarrow \infty$ of odd integers such that $e^{i(\varphi+t_k\alpha)} \rightarrow i \leftarrow e^{i(\varphi+t_k'\alpha)}$. We have $\Omega(\mathfrak{t}) = \Omega = \Omega(\mathfrak{t}')$.

Under some additional hypotheses there are general devices assuring the *existence* of recurrent points. We mention the following cases.

(6) Let Ω be compact. A closed nonempty set $M \subseteq \Omega$ which fulfills $MT \subseteq M$ and is minimal with respect to these properties, is called *minimal invariant.* The existence of minimal invariant sets follows from compactness via Zorn's lemma. In our metrizable case, Zorn's lemma may be replaced by some more constructive device. Minimal invariance of M implies that for every $\eta \in M$, the sequence $\eta, \eta T, \eta T^2, \cdots$ is dense in M. Consequently, if U is an open neighborhood of

some $\omega \in M$, then $M \subseteq \bigcup_{t \geq 0} UT^{-t}$, and compactness implies $M \subseteq \bigcup_{0 \leq t \leq L} UT^{-t}$ for some finite $L \geq 0$. In particular, $\mathfrak{t}(\omega, U)$ is nonempty and even dense; two successive elements of $\mathfrak{t}(\omega, U)$ differ by at most $L + 1$. Thus we have obtained points which are recurrent even in a refined sense.

(7) Let m be a finite measure in Ω which is T-invariant, namely, it fulfills $m(GT^{-1}) = m(G)$ $(G \in B)$. Applying this to $G = \bigcup_{t \geq 0} UT^{-1}$, we obtain from equation (13)

$$m(U - U_{\text{ret}}) = 0, \qquad (U \in B). \tag{14}$$

This is Poincaré's classical recurrence theorem. By (12) (for a countable basis $\sum$) we find that m-almost every $\omega \in \Omega$ is recurrent. By means of the individual ergodic theorem it can even be proved that for almost every ω the sets $\mathfrak{t}(\omega, U)$ each have positive frequency in $(0, 1, 2, \cdots)$. In particular, existence of a finite T-invariant measure implies existence of recurrent points. Incidentally, compactness of Ω implies the existence of a finite T-invariant measure. In the next section we shall see how to construct T-recurrent points, if Ω is a product space and T is the shift transformation.

3. *Recurrent sequences.* Let $X = \{x, y, \cdots\}$ be a Polish space. From copies $X_t = X$, $(t = 0, 1, \cdots)$ of X we form the product space

$$\Omega = \prod_{t=0}^{\infty} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X (t \geq 0)\}, \tag{15}$$

of all *sequences* of points in X, with product topology, which is also Polish. The *shift* mapping $T\colon \Omega \to \Omega$ is defined by

$$(x_0, x_1, \cdots)T = (x_1, x_2, \cdots), \qquad (\omega = (x_0, x_1, \cdots) \in \Omega). \tag{16}$$

Since T is continuous, all preceding sections apply.

DEFINITION 2.3. *A sequence $x_0, x_1, \cdots \in X$ is called* recurrent, *if the corresponding point $\omega = (x_0, x_1, \cdots) \in \Omega$ is T-recurrent, and* recurrent along $\mathfrak{t}$ *or $\mathfrak{t}$-recurrent, if ω is recurrent along $\mathfrak{t}$. According to the definition of product topology in Ω, a sequence $x_0, x_1, \cdots \in X$ is recurrent along $\mathfrak{t}$ if and only if*

$$x_{t+t_k} \to x_t, \qquad (t = 0, 1, \cdots). \tag{17}$$

This implies, for example, that two $\mathfrak{t}$-recurrent sequences are equal if they differ only in finitely many components.

Let us see how to get examples.

(1) If $\mathfrak{t}$: $t_0, t_1, \cdots$ is such that $t_{k+1} = t_k + 1$ for infinitely many k, then every $\mathfrak{t}$-recurrent sequence is constant.

(2) Let $t_k = k!$. Then every periodic sequence is recurrent along $\mathfrak{t}$: $t_0, t_1, \cdots$. Since the periodic points are dense in Ω, we find that $\Omega(\mathfrak{t})$ is dense in Ω, but it is a proper subset of Ω, unless X consists of a single point; indeed, it is easy to exhibit sequences ω which are not $\mathfrak{t}$-recurrent, if X has more than one point.

(3) If X is compact, so is Ω, and we obtain recurrent sequences ω as members of minimal T-invariant sets. There are also explicit constructions of such points.

For $X = \langle -1, 1\rangle$ we may, for instance, put $X_k = \cos \alpha k$ for an arbitrarily fixed real α (almost periodicity) (see, furthermore, Morse-Hedlund [9]).

(4) It is easy to construct finite T-invariant measures in Ω; for example, as product or Markov measures. This makes (7) of section 3 applicable to yield another way to recurrent sequences.

(5) Explicit construction of recurrent sequences: recurrence of a sequence $x_0, x_1, \cdots, \in X$ means that every finite section $x_0, \cdots, x_t$ is reproduced in the sequence again and again up to arbitrarily small deviations. This suggests the following: let $\mathfrak{t}$: $t_0, t_1, \cdots$ be such that $t_{k+1} - t_k \underset{k}{\to} \infty$ (after (3) this assumption causes no essential loss of generality). We intend to place points of X onto the integers $0, 1, \cdots$ so as to obtain a recurrent sequence. This placement will now be done in the following way. First, choose $x_0, \cdots, x_{t_1} \in X$ arbitrarily; for all $k > 1$ such that $t_{k+1} - t_k > t_1$, put $x_{t+u} = x_u$ $(0 \leq u \leq t_1)$, thus ensuring that $x_{u+t_k} \to x_u$ $(0 \leq u \leq t_1)$. Second, fill all empty places $u \leq t_2$ with arbitrary $x_u \in X$; for all $k > 2$ such that $t_{k+1} - t_k > t_2$, we may put $x_{t_k+u} = x_u$ $(0 \leq u \leq t_2)$ without destroying the results of our first step, and so on. It is clear that this construction, working as if the topology in X would be the discrete one, is highly flexible. One could also discuss the two-sided infinite product space $\tilde{\Omega} = \prod_{t=-\infty}^{\infty} X_t$, $(X_t = X, t = 0, \pm 1, \cdots)$ with its shift $\tilde{T}$, and ask different questions, for instance, about continuation of T-recurrent sequences in Ω into $\tilde{T}$-recurrent sequences in $\tilde{\Omega}$. We will not discuss these problems here.

4. *Recurrent measures.* We consider the set V of all probability distributions on Ω, endow it with its weak topology, and consider T as a continuous mapping of V into itself. We want to define recurrent probability distributions as recurrent points in V. As V, in general, is not Polish, we have to make a choice between definitions 2.1 and 2.2. We decide for the latter, thus setting up the following definition.

DEFINITION 2.4. *A finite charge distribution $h \in R(B)$ is called* recurrent along $\mathfrak{t}$ *or* $\mathfrak{t}$-recurrent, *if*

$$hT^{t_k} \underset{k}{\to} h, \qquad \text{(weakly)}. \tag{18}$$

Let us denote by

(a) $R(\mathfrak{t})$ the set of all $\mathfrak{t}$-recurrent charge distributions;

(b) $R(\mathfrak{t})^+$ the set of all $\mathfrak{t}$-recurrent finite measures; and

(c) $V(\mathfrak{t})$ the set of all $\mathfrak{t}$-recurrent probability distributions.

The elements of $U_{\mathfrak{t}}R(\mathfrak{t})$, $U_{\mathfrak{t}}R(\mathfrak{t})^+$, $U_{\mathfrak{t}}V(\mathfrak{t})$ are simply called (sequentially) recurrent charges, measures, and probability distributions respectively.

I think there will be no confusion about the symbols $R(B)$, $R(\mathfrak{t})$, and so on.

Well-known criteria for weak convergence (see, for example, Prohorov [10], p. 164) imply the following.

LEMMA 2.2. *Let m be a finite measure in Ω, and $\mathfrak{t}$: $t_0, t_1, \cdots$. Then the following statements are equivalent:*

(i) $m \in R^+(\mathfrak{t})$,

(ii) $\lim_k \inf m(UT^{-t_k}) \geq m(U)$, $\qquad (U \subseteq \Omega \text{ open})$,

(iii) $\lim_k \sup m(FT^{-t_k}) \leq m(F)$, $(F \subseteq \Omega$ closed$)$,

(iv) $\lim_k m(ET^{-t_k}) = m(E)$, $(E \subseteq \Omega$ m-boundaryless$)$.

Here a measurable set E is called m-boundaryless if its boundary ∂E fulfills $m(\partial E) = 0$.

Now we give some easy examples and general points of view so as to make the reader familiar with the concept of a recurrent measure. More complicated examples will be discussed in sections 6 and 7. The structure of $R(\mathfrak{t})$, $R(\mathfrak{t})^+$, and $V(\mathfrak{t})$ will be investigated in sections 4 and 5.

(1) Let X be a compact Polish space and W be the simplex of all probability distributions on X. Endowed with its weak topology, W is a compact Polish space again; hence, we may define the concept of a T-recurrent sequence in W, and construct examples according to section 4. Therefore, let some $\mathfrak{t}$: $t_0, t_1, \cdots$ and a sequence $p_0, p_1, \cdots$ of probability distributions in X be given, such that

$$p_{t+t_k} \underset{k}{\rightarrow} p_t, \qquad \text{(weakly)} \tag{19}$$

for every $t = 0, 1, \cdots$. Let $X_t = X$; $(t = 0, 1, \cdots)$. Let

$$\Omega = \prod_{t=0}^{\infty} X_t = \{\omega = (x_0, x_1, \cdots), x_t \in X, (t = 0, 1, \cdots)\}, \tag{20}$$

and T be the shift in Ω. Let us show that the product measure $m = \prod_{t=0}^{\infty} p_t$ in Ω is $\mathfrak{t}$-recurrent. By the Stone-Weierstrass theorem we have to check

$$\int f\, d(mT^{t_k}) \underset{k}{\rightarrow} \int f\, dm \tag{21}$$

only for functions f which are derived from functions $f_0(x), \cdots, f_n(x) \in C(X)$ by

$$f(\omega) = f_0(x_0) \cdots f_n(x_n), \qquad (\omega = (x_0, x_1, \cdots) \in \Omega). \tag{22}$$

For such a function f we obtain

$$\int f\, d(mT^{t_k}) = \int fT^{t_k}\, dm = \prod_{t=0}^{n} \int f_t\, dp_{t+t_k} \underset{k}{\rightarrow} \prod_{t=0}^{n} \int f_t\, dp_t = \int f\, dm, \tag{23}$$

as desired. If X is finite, m may be called an $\mathfrak{t}$-recurrent Bernoulli distribution, and one may well expect that such distributions appear when a sequence of independent random experiments is subject to recurrent influences from outside.

(2) Let X be finite and $P_t = (P_t(i, k))_{i,k \in X}$, $(t = 1, 2, \cdots)$ be a $\mathfrak{t}$-recurrent sequence in the compact space S of all stochastic $X \times X$-matrices. Passing to a subsequence of $\mathfrak{t}$, we may assume that the partial products converge along $\mathfrak{t}$:

$$P_1 \cdots P_{t_k} \underset{k}{\rightarrow} Q \in S. \tag{24}$$

Let p be any Q-invariant probability distribution. The existence of such p is a well-known fact. It follows that the sequence $(p_t = pP_1 \cdots P_t)$ fulfills

$$p_{t_k} \underset{k}{\rightarrow} p. \tag{25}$$

Now let Ω, T as in (1), and let m be the Markovian measure in Ω with initial distribution p and transition matrices $P_1, P_2, \cdots$. We claim that m is $\mathbf{t}$-recurrent. Indeed, by the Stone-Weierstrass theorem we need only prove

$$(mT^{t_k})(F) = m(F) \tag{26}$$

for special cylinder sets

$$F = \{\omega = (x_0, x_1, \cdots) | x_0 = y_0, \cdots, x_n = y_n\}, \tag{27}$$

(the indicator functions of these sets form a total set in $C(\Omega)$). But for such sets we have

$$\begin{aligned}(mT^{t_k})(F) &= m(FT^{-t_k}) \\ &= p_{t_k}(y_0)P_{t_k+1}(y_0, y_1) \cdots P_{t_k+n}(y_{n-1}, y_n).\end{aligned} \tag{28}$$

This converges to $p(y_0)\, P_1(y_0, y_1) \cdots P_n(y_{n-1}, y_n) = m(F)$.

(3) If $\omega \in \Omega$ is $\mathbf{t}$-recurrent, then δ_ω, the unit mass on the point ω, is a $\mathbf{t}$-recurrent measure.

(4) Minimal invariant sets may be found in V, if V is weakly compact (for example, if Ω is compact); thus we have another way to obtain recurrent measures.

3. The recurrence theorem (hereditary property of recurrence)

According to our introductory announcement, we now prove a theorem which states that recurrence is hereditary from macrophysics to microphysics, thus providing a generalization of the topologized version of Poincaré's classical recurrence theorem.

THEOREM 3.1. *Let T be a continuous mapping of a Polish space Ω into itself, and m a recurrent (with respect to T) finite measure on Ω. Then m-almost every point of Ω is recurrent (with respect to T).*

FIRST PROOF (V. Strassen). Let $U \subseteq \Omega$ be open and $G = \bigcup_{t \geq 0} UT^{-t}$. Then G is open, $GT^{-1} \subseteq G$, and $G - GT^{-1} = U - U_1$, where U_1 is the set of all points in U which return to U. It is sufficient to show $m(GT^{-1}) = m(G)$. Let $\epsilon > 0$ and $f \in C(\Omega)$ be such that $0 \leq f \leq 1_G$ and $\int f\, dm \geq m(G) - \epsilon$. Find $t > 0$ such that $\int fT^t\, dm > \int f\, dm - \epsilon$. Then

$$\begin{aligned}m(G) \geq m(GT^{-1}) \geq m(GT^{-t}) &= \int 1_G T^t\, dm \geq \int fT^t\, dm \\ &> \int f\, dm - \epsilon > m(G) - 2\epsilon.\end{aligned} \tag{29}$$

As $\epsilon > 0$ is arbitrary, the desired statement follows.

SECOND PROOF (Jacobs). (A) For any set $U \subseteq \Omega$, let ∂U denote the boundary of U. The set U is called *m-boundaryless* if $m(\partial U) = 0$ (of lemma 2.2). An open set U is called *strictly m-boundaryless* if every UT^{-t} is m-boundaryless, that is, if $m(\partial(UT^{-t})) = 0$, $(t = 0, 1, \cdots)$. Complements, finite unions, and finite intersections of strictly m-boundaryless sets are again strictly m-boundaryless.

Let us show that there exists a countable basis for the topology, which consists of strictly m-boundaryless sets only. Let the topology be given by the metric $|\cdot, \cdot|$. Then for any given $\omega \in \Omega$ the system of all sets of the form

$$U_\epsilon(\omega) = U_\epsilon = \{\eta | \, |\omega, \eta| < \epsilon\}, \qquad (\epsilon > 0) \tag{30}$$

is a basis of neighborhoods of ω. Let us show that for any $t \geq 0$ and $0 < \epsilon < \epsilon'$,

$$\partial(U_\epsilon T^{-t}) \cap \partial(U_{\epsilon'} T^{-t}) = 0. \tag{31}$$

Indeed, let η belong to the set on the left side of (31), and let $\eta' = \eta T^t$, and W' a neighborhood of η'; then $W = W'T^{-t}$ is a neighborhood of η, and hence contains points of $U_\epsilon T^{-t}$, $U_\epsilon^c T^{-t}$, $U_{\epsilon'} T^{-t}$, and $U_{\epsilon'}^c T^{-t}$ respectively. The images under T^t of such points belong to U_ϵ, U_ϵ^c, $U_{\epsilon'}$, $U_{\epsilon'}^c$, respectively, and to W'. It follows that $\eta' \in \partial U_\epsilon \cap \partial U_{\epsilon'} = \{\zeta | \, |\omega, \zeta| = \epsilon\} \cap \{\zeta | \, |\omega, \zeta| = \epsilon'\}$, which yields a contradiction.

From (31) it follows, that for every $\omega \in \Omega$ the $U_\epsilon(\omega)$ are strictly m-boundaryless if $\epsilon > 0$ avoids a certain countable set. Clearly, there are sequences $\omega_1, \omega_2, \cdots \in \Omega$, $\epsilon_1, \epsilon_2, \cdots > 0$ such that the $U_{\epsilon_i}(\omega_k)$ are strictly boundaryless, and that their finite unions constitute a basis of the topology. These finite unions are again strictly boundaryless, and form a countable system.

(B) It follows that

$$\Omega_r = \bigcap_{k=1}^{\infty} U_{\text{ret}}^k \cap (\Omega' - U^k) \tag{32}$$

for a suitable sequence $U^1, U^2, \cdots$ of strictly m-boundaryless open sets. Hence, it will be sufficient to prove

$$m(U_{\text{ret}}) = m(U) \tag{33}$$

for an arbitrary strictly m-boundaryless open set U.

(C) Let U be open and strictly m-boundaryless. For any $0 < s_1 < s_2 < \cdots < s_\ell$, we have

$$U - U_{\text{ret}} \subseteq U \cap U^c T^{-s_1} \cap U^c T^{-s_2} \cap \cdots \cap U^c T^{-s_\ell}; \tag{34}$$

hence, it is sufficient to show that, for every $\epsilon > 0$, $s_1, \cdots, s_\ell$ may be chosen in such a way that the set on the right-hand side of this relation has m-measure $< \epsilon$.

Exploiting recurrence of m, lemma 2.2, and the fact that arbitrary finite intersections of m-boundaryless sets, such as UT^{-t}, U^cT^{-t}, are m-boundaryless again, we chose $t_1, t_2, \cdots$ successively in the following way. Let $\epsilon_{k\nu} > 0$, $(k = 1, 2, \cdots; \nu \geq k)$ be chosen successively in such a way that $\sum_{\nu \geq k} \epsilon_{k\nu} < \epsilon/2$, $(k = 1, 2, \cdots)$.

Let

(i) $t_1 > 0$ be such that

$$m(UT^{-t_1}) > m(U) - \epsilon_{11}; \tag{35}$$

(ii) $t_2 > 0$ be such that

$$\begin{aligned} m(UT^{-(t_1+t_2)}) = m((UT^{-t_1})T^{-t_2}) &> m(UT^{-t_1}) - \epsilon_{12} \\ &> m(U) - (\epsilon_{11} + \epsilon_{12}), \end{aligned} \tag{36}$$

(37) $$m(U^cT^{-(t_1+t_2)} \cap UT^{-t_2}) = m((U \cap U^cT^{-t_1})T^{-t_2}) > m(U \cap U^cT^{-t_1}) - \epsilon_{22};$$

(iii) t_n be such that

(38) $$m(UT^{-(t_1+\cdots+t_n)}) > m(U) - (\epsilon_{11} + \epsilon_{12} + \cdots + \epsilon_{1n}),$$

(39) $$m(U^cT^{-(t_1+\cdots+t_n)} \cap UT^{-(t_2+\cdots+t_n)}) > m(U \cap U^cT^{-t_1}) - (\epsilon_{22} + \epsilon_{23} + \cdots + \epsilon_{2n}),$$

$$\cdots$$

(40) $$m(U^cT^{-(t_1+\cdots+t_n)} \cap U^cT^{-(t_2+\cdots+t_n)} \cap \cdots \cap U^cT^{-(t_{n-1}+t_n)} \cap UT^{-t_n}) > m(U \cap U^cT^{t_{n-1}} \cap \cdots \cap U^cT^{-(t_1+\cdots+t_{n-1})}) - \epsilon_{nn}.$$

At every step, the sets occurring as the arguments of m in the left sides of these inequalities are mutually disjoint. As m is finite, there is, for n sufficiently large, a $k \leq n$ such that an inequality

(41) $$\frac{\epsilon}{2} > m(U^cT^{-(t_1+\cdots+t_n)} \cap \cdots \cap U^cT^{-(t_k+\cdots+t_n)} \cap UT^{-(t_{k+1}+\cdots+t_n)}) > m(U \cap U^cT^{-t_k} \cap U^cT^{-(t_{k-1}+t_k)} \cap \cdots \cap U^cT^{-(t_1+\cdots+t_k)}) - (\epsilon_{kk} + \cdots + \epsilon_{kn}),$$

holds, from which

(42) $$m(U \cap U^cT^{-s_1} \cap \cdots \cap U^cT^{-s_l}) < \epsilon$$

follows, if we put

(43) $$l = k, \qquad s_1 = t_k, \cdots, s, \qquad s_k = t_1 + \cdots + t_k,$$ q.e.d.

Both of the above proofs leave open the following question: let $\mathfrak{t}$: $t_0, t_1, \cdots$ be such that m is recurrent along $\mathfrak{t}$. Is m-almost every $\omega \in \Omega$ recurrent along some subsequence of $\mathfrak{t}$? In the classical case $\mathfrak{t}$: $0, 1, 2, \cdots$, this is trivially true. We shall prove in section 5 that it is also true for arbitrary $\mathfrak{t}$, if we impose some mixing condition on m. We shall give a third proof for $m(\Omega_{\text{rec}}) = m(\Omega)$ in that special case.

4. Lattice properties of recurrent measures

Let Ω (Polish), T: $\Omega \to \Omega$ (continuous), and $\mathfrak{t}$: $t_0, t_1, \cdots$ be given. Remember the definitions of section 2, subsection 4:

(a) $R(\mathfrak{t}) = \{h | h \in R(B),\ hT^{t_k} \underset{k}{\to} h \text{ (weakly)}\}$,

(b) $R(\mathfrak{t})^+ = R^+(\mathfrak{t}) = \{m | m \in R(\mathfrak{t}),\ m \geq 0\}$,

(c) $V(\mathfrak{t}) = \{m | m \in R^+(\mathfrak{t}),\ m(\Omega) = 1\}$.

THEOREM 4.1. *The following relations hold:*

(1) $R(\mathfrak{t})$ *is a norm-closed linear sublattice of* $R(B)$;

(2) $R^+(\mathfrak{t})$ *is a norm-closed convex cone, and a lattice;*

(3) $V(\mathfrak{t})$ *is a simplex in the algebraic sense;*

(4) *T maps each of $R(\mathbf{t})$, $R^+(\mathbf{t})$, $V(\mathbf{t})$ into itself and commutes with all lattice operations.*

PROOF. We have to prove that $R(\mathbf{t})$ is a sublattice of $R(B)$ and that T commutes with all finite lattice operations. All other statements are either easily derived from that, or admit simple direct proofs. It is sufficient to show that

$$|h| \in R(\mathbf{t}), \qquad (h \in R(\mathbf{t})), \tag{44}$$

$$|h|T = |hT|, \qquad (h \in R(\mathbf{t})), \tag{45}$$

as every lattice operation may be represented by means of $|\cdot|$ and linear operations. We now observe the following:

(A) the operation $h \to |h|$ in $R(B)$ is *weakly lower semicontinuous* (see section 2 (2));

(B) $|hT^t| \leq |h|T^t$ with equality if and only if $|hT^t|(\Omega) = (|h|T^t)(\Omega)$, $(h \in R(B), t \geq 0)$;

(C) assume now $h \in R(\mathbf{t})$, that is, $hT^{t_k} \to h$ (weakly). From (A) (section 2, (2)) and (B) we infer, for arbitrary nonnegative $f \in C(\Omega)$,

$$\liminf_k \int f\, d(|h|T^{t_k}) \geq \liminf_k \int f\, d|hT^{t_k}| \geq \int f\, d|h|. \tag{46}$$

Applying this result to $c - f \in C(\Omega)$ where $c \geq 0$ is a constant such that $c - f \geq 0$, we obtain

$$\limsup_k \int f\, d(|h|T^{t_k}) \leq \int f\, d|h|, \tag{47}$$

and hence,

$$\int f\, d(|h|T^{t_k}) \underset{k}{\rightarrow} \int f\, d|h|, \tag{48}$$

first for $0 \leq f \in C(\Omega)$, and then, obviously, for arbitrary $f \in C(\Omega)$. This shows (44).

(D) It follows from (B) that for $f \equiv 1$,

$$\int f\, d|h| \geq \int f\, d|hT| \geq \int f\, d|hT^2| \geq \cdots, \tag{49}$$

with equality everywhere if and only if $|h|T^t = |hT^t|$, $(t \geq 0)$. But (C) implies that a subsequence of the sequence $\int f\, d|hT^t|$ tends to $\int f\, d|h|$. Hence, equality holds everywhere in (49), and (45) follows.

For $\mathbf{t}$: $0, 1, \cdots$, $V(\mathbf{t})$ is the set of all T-invariant normalized measures in Ω; the extremal points of $V(\mathbf{t})$ are the so-called T-ergodic normalized measures; if Ω is compact, $V(\mathbf{t})$ is weakly compact, and Choquet's barycentric decomposition theorem can be applied to yield ergodic decomposition of an arbitrary $m \in V(\mathbf{t})$. The ergodic decomposition is unique because $V(\mathbf{t})$ is algebraically a simplex (see, for example, Jacobs [6]). We may question to what extent the 'classical situation' described by these statements carries over to arbitrary $\mathbf{t}$: $t_0, t_1, \cdots$.

Theorem 4.1 shows that many statements are still true in the general case. There is, however, one key hypothesis to Choquet's theorem which does not generalize: in general, $V(\mathbf{t})$ is not compact. This is shown by the following.

EXAMPLE. Let $X = \{1, \cdots, a\}$ be finite and

$$\Omega = \prod_{t \geq 0} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X (t \geq 0)\} \tag{50}$$

be the corresponding product space, and let T be the shift operator. If $t_k = k!$, then every T-periodic probability distribution is recurrent along $\mathbf{t}$: $t_0, t_1, \cdots$; that is, $V(\mathbf{t})$ contains all T-periodic normalized measures in Ω. But these are known to be dense in V. On the other hand, $V \neq V(\mathbf{t})$, as is easily shown by constructions as in sections 2.1(6), 2.4(1). Hence $V(\mathbf{t})$ is not weakly closed, and therefore not weakly compact.

Thus we cannot simply apply Choquet's barycentric decomposition theorem to the simplex $V(\mathbf{t})$ in general. It is not hard to show that barycenters of arbitrary mass distributions in $V(\mathbf{t})$ (which need not be concentrated on finitely many points) belong to $V(\mathbf{t})$. In particular, every $m \in V$ which fulfills $m(\Omega(\mathbf{t})) = 1$, belongs to $V(\mathbf{t})$. Theorem 3.1 also yields a barycentric representation of an arbitrary $m \in V(\mathbf{t})$ as a barycenter of a mass distribution M on point masses δ_ω, each of which is recurrent; M is obtained by transport of m by the mapping $\omega \rightarrow \delta_\omega$. But in many cases, M-almost no δ_ω will belong to $V(\mathbf{t})$; for instance, let m be extremal in $V(\mathbf{t})$, but not concentrated at a single point (examples of this situation will be obtained in section 5). This implies that $m(\Omega(\mathbf{t})) = 0$, because if $m(\Omega(\mathbf{t})) > 0$, we could write m as a sum of two nonzero measures from $R^+(\mathbf{t})$ which are not multiples of m (namely, the restriction m_F of m to some subset F of $\Omega(\mathbf{t})$ fulfilling $0 < m(F) < 1$, and $m - m_F$) contradicting the extremality of m in $V(\mathbf{t})$. Hence, M-almost no δ_ω belongs to $V(\mathbf{t})$.

The following theorem generalizes a well-known classical statement.

THEOREM 4.2. *If m, m' are two different extremal points of $V(\mathbf{t})$, then $m \perp m'$.*

PROOF. Clearly $m \wedge m'$ belongs to $R^+(\mathbf{t})$ and has to be a constant multiple of both m and m', because otherwise barycentric representations of m or m' of a type forbidden for extremal points would arise, and $m \wedge m' \neq 0$ would not imply $m = m'$. Hence, the theorem follows.

REMARK. This theorem has special cases in common with the result of Kakutani [8].

DEFINITION 4.1. *A finite measure m is called wandering (with respect to T) if the transported measures $m, mT, mT^2, \cdots$ are pairwise orthogonal:*

$$mT^s \perp mT^t, \qquad (0 \leq s < t) \tag{51}$$

THEOREM 4.3. *Let $m \in V(\mathbf{t})$ be such that $m, mT, mT^2, \cdots$ are extremal points of $V(\mathbf{t})$. Then m is either periodic or wandering.*

PROOF. Assume there exist s, t such that $0 \leq s < t$, $mT^s = mT^t$. Put $t - s = d$. Then $mT^{u+d} = mT^u$ $(u \geq s)$. As m is a weak limit point of the sequence $mT^s, mT^{s+1}, \cdots$, which is periodic of period d, it coincides with some mT^u where $s \leq u < t$. It follows that $mT^d = m$, hence m is periodic. If $mT^s \neq mT^t$, $(0 \leq s < t)$, then m is wandering by theorem 4.2.

We now have to investigate under what circumstances theorem 4.3 will apply. The subsequent theorem 4.4 will show that theorem 4.3 always applies if T is an automorphism and m is an extremal point of $V(\mathfrak{t})$. In section 5 we shall see that theorem 4.3 always applies if m is "mixing along $\mathfrak{t}$."

LEMMA 4.1. *Let S: $\Omega \to \Omega$ be continuous such that $ST = TS$. Then $R(\mathfrak{t})S \subseteq R(\mathfrak{t})$, $R^+(\mathfrak{t})S \subseteq R^+(\mathfrak{t})$, and $V(\mathfrak{t})S \subseteq V(\mathfrak{t})$.*

PROOF. Let, for example, $h \in R(\mathfrak{t})$. Then

$$(hS)R^{t_k} = (hT^{t_k})S \underset{k}{\to} hS, \tag{52}$$

as S clearly induces a weakly continuous mapping of $R(B)$ into itself.

THEOREM 4.4. *Let T be an automorphism of Ω (namely, one-to-one onto, with a continuous inverse T^{-1}). Then T sends extremal points of $V(\mathfrak{t})$ into extremal points of $V(\mathfrak{t})$.*

PROOF. Let m be an extremal point of $V(\mathfrak{t})$. Assume that $mT \in V(\mathfrak{t})$ is not extremal: $mT = \frac{1}{2}(m' + m'')$, $m' \neq m''$, and $m', m'' \in V(\mathfrak{t})$. Let $\overline{m}' = m'T^{-1}$, $\overline{m}'' \in m''T^{-1}$. Lemma 4.1 applied to $S = T^{-1}$ shows that $\overline{m}', \overline{m}'' \in V(\mathfrak{t})$. If $m' \neq m''$, then $\overline{m}' \neq \overline{m}''$. But we have $m = \frac{1}{2}(\overline{m}' + \overline{m}'')$ contradicting extremality.

The following results may be considered as tools for disproving the extremality of a given point $m \in V(\mathfrak{t})$.

THEOREM 4.5. *Let $K \subseteq B$ be a system of measurable sets, and*

(a) $R(\mathfrak{t}, K) = \{h | h \in R(\mathfrak{t}), h(F) = 0\ (F \in K)\}$,

(b) $R^+(\mathfrak{t}, K) = R(\mathfrak{t}, K) \cap R^+(\mathfrak{t})$,

(c) $V(\mathfrak{t}, K) = R(\mathfrak{t}, K) \cap V(\mathfrak{t})$.

Then

(1) *$R(\mathfrak{t}, K)$ is a norm-closed linear sublattice of $V(\mathfrak{t})$,*

(2) *$R^+(\mathfrak{t}, K)$ is a norm-closed convex cone, and a lattice,*

(3) *$V(\mathfrak{t}, K)$ is a simplex.*

For the proof use theorem 4.1.

THEOREM 4.6. *Let $K \subseteq B$ be arbitrarily given. Then for every finite measure m_0 the system*

$$\{m | m \in R^+(\mathfrak{t}, K),\ m \leq m_0\} \tag{53}$$

is a norm-closed lattice and hence contains a unique maximal element $\overline{m}$. If $m_0 \in R^+(\mathfrak{t})$, then

$$\overline{m} \perp (m_0 - \overline{m}), \tag{54}$$

that is, $\overline{m}$ is the restriction of m_0 to a suitable measurable set.

PROOF. All statements but the last are an easy consequence of theorem 4.5. The fact that $R^+(\mathfrak{t}, K)$ is a norm-closed cone implies, together with the maximality of $\overline{m}$ and $m_0 \in R^+(\mathfrak{t})$,

$$\overline{m} = (n\overline{m}) \wedge m_0, \qquad (n = 1, 2, \cdots). \tag{55}$$

This in turn implies (54).

The following example exhibits an application of the above theorems.

EXAMPLE. A set $F \in B$ is called *weakly wandering* if there is a sequence $u_0, u_1, \cdots \to \infty$ such that $FT^{-u_j} \cap FT^{-u_k} = 0$, $(j \neq k)$. Let

$$K = \{F | F \in B,\ F \text{ weakly wandering}\}. \tag{56}$$

Theorem 3.6 implies that for any given $m_0 \in V(\mathfrak{t})$ there is a unique minorant $\overline{m}$ of m_0 which vanishes on every weakly wandering set and is maximal with respect to these properties. Sucheston's [11] extension of Hajian-Kakutani's [1] famous result implies that a finite measure is dominated by a finite T-invariant measure if and only if it vanishes on K. Hence $\overline{m}$ is the maximal $\mathfrak{t}$-recurrent component of m_0 which is dominated by a finite T-invariant measure. Theorem 3.6 says that $\overline{m}$ is the restriction of m_0 to a certain measurable set. For instance, let Ω be the circumference of the unit circle, and T its rotation by an irrational multiple of π. If m_0 consists of at most countably many point masses, then $\overline{m} = 0$. If m_0 is the equidistribution of mass 1 over an arc of positive length, then $\overline{m} = m$. In general, $\overline{m}$ is the absolutely continuous part of m_0 with respect to arc length measure.

5. Mixing

Let Ω (Polish) and $T\colon \Omega \to \Omega$ (continuous) be fixed.

DEFINITION 5.1. *Let* $\mathfrak{t}$: $t_0, t_1, \cdots$ *be given. A normalized measure in* $m \in V$ *is called*

(i) mixing along $\mathfrak{t}$, *if*

$$\lim_{k\to\infty} \left[\int fT^{t_k}g\, dm - \int fT^{t_k}\, dm \int g\, dm\right] = 0, \qquad (f, g \in C(\Omega)); \tag{57}$$

(ii) mixing, *if it is mixing along* $\mathfrak{t}$: $0, 1, \cdots$.

Mixing along a sequence is preserved

(a) if we pass to a subsequence,

(b) if we transport the measure by T.

Clearly every point mass $m = \delta_\omega$ is mixing. If $mT = m$, and m is weakly mixing, then m is in general not mixing (see Jacobs [6]), but there is a sequence $\mathfrak{t}$: $t_0, t_1, \cdots$, even of frequency 1, such that m is mixing along $\mathfrak{t}$ (apply lemma 4.4.1 from Jacobs [4]). Let X be finite and

$$\Omega = \prod_{t=0}^{\infty} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X\ (t \geq 0)\} \tag{58}$$

be the corresponding product space, with shift T. Let $p_0, p_1, \cdots$ be any sequence of normalized measures on X. Then the product measure $m = \prod_{t=0}^{\infty} p_t$ on Ω is mixing. This follows easily (Stone-Weierstrass) from the fact that finite-dimensional cylinder functions are dense in $C(\Omega)$. The same argument holds for the two-sided shift space $\prod_{t=-\infty}^{\infty} X_t$.

Further examples (Markov measures in shift space) will be provided by theorem 7.6. It is easily seen by an approximation procedure that m is mixing along $\mathfrak{t}$ if and only if (i) holds for arbitrary $f \in C(\Omega)$ and $g \in L^1_m$. Passage to $f \in L^1_m$ is, however, in general not allowed, as we shall see later on.

The above definition is formulated in terms of *continuous functions*. We want to pass to an equivalent version in terms of *sets*. For any subset F of Ω, denote by ∂F the boundary of F. Then we have the following theorem.

THEOREM 5.1. *Let $m \in V(\mathbf{t})$. Then the following statements are equivalent:*

(i) *m is mixing along $\mathbf{t}$;*

(ii) *for any two measurable subsets F, G of Ω such that $m(\delta F) = 0$, the relation*

$$\lim_{k\to\infty} [m(FT^{-t_k} \cap G) - m(FT^{-t_k})m(G)] = 0 \tag{59}$$

holds, that is, the sequence FT^{-t_k} is mixing in the sense of Rényi [13].

PROOF. (i) $\Rightarrow$ (ii): Let m be mixing along $\mathbf{t}$. If $m(\partial F) = 0$, then the existence of $f_0, f_1 \in C(\Omega)$ such that $0 \le f_0 \le 1_F \le f_1$ and $\int (f_1 - f_0)\, dm < \epsilon$, where $\epsilon > 0$ is arbitrarily given in advance. If $m \in V(\mathbf{t})$, then for arbitrary $g \in C(\Omega)$ with $0 \le g \le 1$,

$$\begin{aligned} 0 &\le \liminf_k \left[\int 1_F T^{t_k} g\, dm - \int f_0 T^{t_k} g\, dm\right] \\ &\le \limsup_k \left[\int 1_F T^{t_k} g\, dm - \int f_0 T^{t_k} g\, dm\right] \\ &\le \limsup_k \left[\int (f_1 - f_0) T^{t_k} g\, dm\right] \le \lim_k \int (f_1 - f_0) T^{t_k}\, dm \\ &= \int (f_1 - f_0)\, dm < \epsilon. \end{aligned} \tag{60}$$

This implies

$$\begin{aligned} \limsup_k \int 1_F T^{t_k} g\, dm &< \lim_k \int f_0 T^{t_k} g\, dm + \epsilon \\ &= \left[\lim_k \int f_0 T^{t_k}\, dm \cdot \int g\, dm\right] + \epsilon \\ &= \int f_0\, dm \cdot \int g\, dm + \epsilon \\ &\le \int 1_F\, dm \int g\, dm + \epsilon. \end{aligned} \tag{61}$$

A similar argument shows

$$\liminf_k \int 1_F T^{t_k} g\, dm \ge \int 1_F\, dm \int g\, dm - \epsilon. \tag{62}$$

As $\epsilon > 0$ is arbitrary,

$$\lim_k \int 1_F T^{t_k} g\, dm = \int 1_F\, dm \int g\, dm \tag{63}$$

follows, first for $0 \le g \le 1$, hence for arbitrary $f \in C(\Omega)$. Taking $g \equiv 1$, we obtain

$$\lim_k \int 1_F T^{t_k}\, dm = \int 1_F\, dm, \tag{64}$$

and, by combination,

$$\lim_k \left[\int 1_F\, T^{t_k} g\, dm - \int 1_F T^{t_k}\, dm \int g\, dm\right] = 0, \qquad (g \in C(\Omega)). \tag{65}$$

Obviously (65) also holds for all $g \in L_m^1$, hence for all $g \in L_m^1$, especially for $g = 1_G$ with measurable $G \subseteq \Omega$. But this means (59).

(ii) $\Rightarrow$ (i): (59) means (65) for $g = 1_G$. By linear combination and L_m^1-norm approximation, we obtain (65) for $m(\partial F) = 0$ and arbitrary $g \in L_m^1$, in particular, for $g \in C(\Omega)$. Now, every $f \in C(\Omega)$ is uniformly approximable by linear combinations of functions 1_F with $m(\partial F) = 0$. This allows us to prove (65) for arbitrary $f \in C(\Omega)$ in the place of 1_F. Consequently, m is mixing along $\mathfrak{t}$.

Remark that $m \in V(\mathfrak{t})$ was not needed for (ii) $\Rightarrow$ (i).

THEOREM 5.2. *If $m \in V(\mathfrak{t})$ is mixing along $\mathfrak{t}$, then m is an extremal point of $V(\mathfrak{t})$.*

PROOF. An easy approximation procedure shows that (57) holds even for $g \in L_m^1$. Now let $m_0 \in V(\mathfrak{t})$, $m_0 \leq m$, and $dm_0 = g\, dm$, that is, $g \in L_m^1$, $0 \leq g \leq 1$. Recurrence, with (57), now implies for arbitrary $f \in C(\Omega)$,

$$\begin{aligned} \int f\, dm_0 &= \lim_k \int f\, d(m_0 T^{t_k}) = \lim_k \int f T^{t_k}\, dm_0 \\ &= \lim_k \int f T^{t_k} g\, dm = \lim_k \int f T^{t_k}\, dm \int g\, dm \\ &= \lim_k \int f\, d(m T^{t_k}) \int g\, dm = \int f\, dm \int g\, dm, \end{aligned} \tag{66}$$

namely that m_0 is a constant multiple of m. This shows extremality of m in $V(\mathfrak{t})$.

THEOREM 5.3. *If $m \in V(\mathfrak{t})$ is mixing along $\mathfrak{t}$, then m is either periodic or wandering.*

PROOF. As $m, mT, mT^2, \cdots$ are all mixing and all in $V(\mathfrak{t})$ (theorem 4.1, (4)), we need only apply theorem 4.3.

We know how to construct nonperiodic mixing measures in shift space (see section 2, 3(6) and 4(1) and the beginning of section 5). Hence we know quite a few wandering recurrent measures which are, as a rule, not simply point masses, but nonatomic.

Assume that m is a wandering measure which is not concentrated at a single point. Assume for simplicity that T is an automorphism. Then there is a decomposition $\Omega = F + G$ such that

$$F \cap G = 0, \qquad FT^{-1} = F, \qquad GT^{-1} = G, \tag{67}$$

$$m(F)m(G) > 0. \tag{68}$$

Indeed let Ω_0 be a carrier of m such that the sets $\Omega_0 T^{-t}$ (t integer) are mutually disjoint; find a decomposition $\Omega_0 = F_0 + G_0$ such that $F_0 \cap G_0 = \varnothing$, $m(F_0)m(G_0) > 0$, and put $F = \bigcup_{t=-\infty}^{\infty} F_0 T^{-t}$, $G = \Omega - F$. For $f = 1_F$, $g = 1_G$ we obtain $fT^r = f$; hence, $fT^t \cdot g \equiv 0$, $(t = 0, 1, \cdots)$. Consequently, (57) cannot hold for these f, g although m can be constructed as a mixing measure. This shows that it was reasonable to restrict f, g to $C(\Omega)$ in definition 5.1.

THEOREM 5.4. *Let $m \in V(\mathfrak{t})$ be mixing along $\mathfrak{t}$. Then m-almost every point of Ω is recurrent along some subsequence of $\mathfrak{t}$ (which will in general depend on the point).*

PROOF. By the construction given in section (A) of the second proof of theorem 3.1, it is sufficient to show that for every strictly m-boundaryless open set $U \subseteq \Omega$ with $0 < m(U) < 1$,

$$\lim_{n\to\infty} m(U \cap U^cT^{-t_1} \cap \cdots \cap U^cT^{-t_n}) = 0, \tag{69}$$

possibly after passage from $\mathfrak{t}$ to a subsequence. Now every finite intersection of sets of type UT^{-t} or U^cT^{-s} is m-boundaryless. Recurrence, with mixing along $\mathfrak{t}$ (the latter in the set-theoretical form given in theorem 5.1), implies the possibility of writing

$$m(U \cap U^cT^{-t_{\nu_1}} \cap \cdots \cap U^cT^{-t_{\nu_n}}) \leq m(U)m(U^c)^n + \delta, \qquad (0 < \nu_1 < \cdots < \nu_n) \tag{70}$$

for an arbitrary $\delta > 0$ possibly by passage from $\mathfrak{t}$ to a subsequence. By a diagonal procedure (69) is now easily obtained for $m(U^c) < 1$.

6. Recurrent Gaussian distributions

Let $\rho(s, t)$ (s, t integers) be a strictly positive definite function, that is, $\rho(s, t) = \overline{\rho(t, s)}$ and $\sum_{i,k=1}^n \alpha_i\rho(s_i, s_k)\bar{\alpha}_k > 0$ for arbitrary mutually different integers $s_1, \cdots, s_n$ and arbitrary complex $\alpha_1, \cdots, \alpha_n$ such that $\sum_{k=1}^n |\alpha_k| > 0$. Let $X = R^1 = X_t$ (t integer) and $\Omega = \prod_t X_t$, with natural product Borel field B and shift T. Clearly, Ω is a Polish but noncompact space. There exists a unique Gaussian probability distribution m in Ω such that $\rho(s, t)$ is the covariance function of m; that is, $\rho(s, t) = \int \varphi_s\overline{\varphi_t}\, dm$ where φ_t: $\Omega \to R^1$ denotes the t-th component mapping.

LEMMA 6.1. *Let $\rho(s, t)$, $\rho_\nu(s, t)$, ($\nu = 1, 2, \cdots$) be strictly positive definite functions over the integers, and m, m_ν ($\nu = 1, 2, \cdots$) the corresponding Gaussian measures in Ω. Then $m_\nu \to m$* (weakly) *if and only if*

$$\rho_\nu(s, t) \to \rho(s, t), \qquad (s, t \text{ integer}). \tag{71}$$

PROOF. (I). Assume $m_\nu \to m$ (weakly). Let $\varphi_t: \Omega \to R^1$ denote the t-th component function. For fixed integers $s \neq t$, let p, p_ν be the common distributions of φ_s and φ_t under m, m_ν ($\nu = 1, 2, \cdots$); they are nonsingular normal distributions in R^2 fulfilling $p_\nu \to p$ (weakly). From this

$$\int_{R^2} xy\, p_\nu(dx, dy) \to \int_{R^2} xy\, p(dx, dy) \tag{72}$$

is deduced by elementary procedures. Hence $p_\nu(s, t) \to p(s, t)$ follows for $s \neq t$. The proof for $s = t$ is still simpler. I leave the details to the reader.

(II). Assume $\rho_\nu(s, t) \to \rho(s, t)$ (s, t integers). Let p, p_ν be the common distribution of a finite number of component functions of Ω under m, m_ν, ($\nu = 1, 2, \cdots$). Then $p_\nu \to p$ (weakly) is easily deduced. It follows that $m_\nu(F) \to m(F)$ for m-boundaryless finite dimensional cylinder sets. The open sets of this kind form a basis for the topology in Ω. Hence, for every open $G \subseteq \Omega$ and given $\epsilon > 0$, there is an m-boundaryless cylinder set F such that $m(F) > m(G) - \epsilon$. This implies

$$\liminf_{\nu} m_{\nu}(G) \geq \liminf_{\nu} m_{\nu}(F) = m(F) > m(G) - \epsilon. \tag{73}$$

Hence,

$$\liminf_{\nu} m_{\nu}(G) \geq m(G), \qquad (G \subseteq \Omega \text{ open}). \tag{74}$$

Consequently, $m_{\nu} \rightarrow m$ (weakly) after a well-known criterion for weak convergence (see, for example, Prohorov [10]).

As a corollary we obtain the following theorem.

THEOREM 6.1. *Let m be a Gaussian probability distribution in Ω, and $\rho(s, t)$ (s, t integers) its covariance function. Then for every* $\mathbf{t}$: $t_0, t_1, \cdots$,

$$m \in V(\mathbf{t}) \tag{75}$$

is equivalent to

$$\rho(s + t_{\nu}, t + t_{\nu}) \rightarrow \rho(s, t), \qquad (s, t \text{ integers}). \tag{76}$$

PROOF. Clearly, $\rho_{\nu}(s, t) = \rho(s + t_{\nu}, t + t_{\nu})$ is the covariance function of $m_{\nu} = mT^{t_{\nu}}$.

It is easy to construct $\rho(s, t)$ such that (76) will be fulfilled for a certain $\mathbf{t}$: $t_0 < t_{\nu} < \cdots$. Take, for example, a sequence of strictly positive definite 2×2 matrices $Q_u = (Q_u(i, k))$, (u integer, $i = 0, 1$), such that $Q_{u+t_{\nu}}(i, k) \rightarrow Q_u(i, k)$, ($u, i, k$ arbitrary). Define

$$\rho(s, t) = \begin{cases} Q_u(i,k) & \text{if} \quad s = 2u + i, \qquad t = 2u + k;\ i, k = 0, 1 \\ 0 & \text{otherwise.} \end{cases} \tag{77}$$

Then $\rho(s + 2t_{\nu}, t + 2t_{\nu}) = Q_{u + t_{\nu}}(i, k) \rightarrow Q_u(i, k) = \rho(s, t)$, if $s = 2u + i$, $t + 2u + k$. For all other pairs s, t we have $\rho(s + 2t_{\nu}, t + 2t_{\nu}) = \rho(s, t) = 0$. Of course, convex combination does not lead out of the domain of all ρ fulfilling (76) with a fixed sequence $\mathbf{t}$: $t_0, t_1, \cdots$.

THEOREM 6.2. *Let m be a Gaussian measure in Ω and $\rho(s, t)$ (s, t integer) its covariance function. Assume that m is recurrent along* $\mathbf{t}$. *Then m is mixing along* $\mathbf{t}$ *if and only if*

$$\lim_{k \rightarrow \infty} \rho(t, t + t_k) = 0, \qquad (t \text{ integer}). \tag{78}$$

PROOF. (I). Assume (78) holds. Then it is easy to prove the relation

$$\lim_{k} [m(FT^{-t_k} \cap G) - m(FT^{-t_k})m(G)] = 0 \tag{79}$$

for all finite-dimensional cylinder sets $F, G \subseteq \Omega$ which belong to the class A of all cylinder sets of the form $F = F_0 \times F_1$ where F_0 is a conditionally compact set in a finite number of coordinates, and F_1 is the product of copies of R_1 in the other coordinates, and moreover, F_0 is boundaryless with respect to the projection of m to the corresponding finitely many coordinates.

There is a countable system $\overline{A}$ of finite unions of sets of A such that $\overline{A}$ is a basis of the topology of Ω; $\overline{A}$ consists of boundaryless sets. On the other hand, for every fixed $G \in B$, the system of all $F \in B$, such that (79) holds is closed with respect to the formation of proper differences and disjoint unions. It is easy to see that A is closed with respect to the formation of finite intersections; consequently, (79) holds for arbitrary $F \in \bar{A}$ and $G \in A$.

Another simple extension procedure shows that (79) holds even for arbitrary $F \in \bar{A}$ and $G \in B$, and consequently, also for arbitrary F such that $F^c \in A$ and $G \in B$. Now we are able to find for arbitrary m-boundaryless F and arbitrary $\epsilon > 0$ sets F_0, F_1 such that $F_0, F_1^c \in \bar{A}$,

$$F_0 \subseteq F \subseteq F_1, \tag{80}$$

$$m(F_1 - F_0) < \epsilon \tag{81}$$

(exhaust the interior of F and of $\Omega - F$ by sets from the above-mentioned basis). Let $G \subseteq \Omega$ be measurable; then

$$m(F_0T^{t_k} \cap G) \leq m(FT^{-t_k} \cap G) \leq m(F_1T^{-t_k} \cap G). \tag{82}$$

Applying (79) to $F = F_0$ and $F = F_1$, we obtain

$$\lim_k m(F_0T^{-t_k} \cap G) = (\lim_k m(F_0T^{-t_k}))m(G) = m(F_0)M(G) \tag{83}$$

$$\begin{aligned} \lim_k m(F_1T^{-t_k} \cap G) &= (\lim_k m(F_1T^{-t_k}))m(G) \\ &= m(F_1)m(G) \leq m(F_0)m(G) + \epsilon \end{aligned} \tag{84}$$

by recurrence along $\mathfrak{t}$ and a well-known criterion for weak convergence (see Prohorov [10]). Consequently,

$$\limsup_k |m(FT^{-t_k} \cap G) - m(FT^{-t_k})m(G)| \leq \epsilon. \tag{85}$$

As $\epsilon > 0$ is arbitrary, (79) follows for arbitrary m-boundaryless F and measurable G; hence, m is mixing along $\mathfrak{t}$, by theorem 5.1.

(II). Let m be mixing along $\mathfrak{t}$. Equation (78) is then derived by easy approximation procedures.

7. Recurrent finite-state Markov measures

In this section we investigate Markov measures with a *finite state space* $X = \{1, \cdots, a\} \neq \varnothing$ and a recurrent sequence of transition matrices. We shall prove that for appropriate initial distributions the Markov measure is recurrent, and sometimes even mixing.

1. *Distributions in state space.* A *charge distribution* in the state space $X = \{1, \cdots, a\} \neq \varnothing$ is a real vector $h = (h_{(1)}, \cdots, h_{(a)}) = (h_i) \in R^a$; h_i is the charge of state i. The total (absolute) charge

$$\|h\| = \sum_{i=1}^{a} |h_{(i)}| \tag{86}$$

is the *norm* of h. The charge h is called a *mass distribution on* X if

$$h \geq 0, \quad \text{that is,} \quad h_{(i)} \geq 0, \qquad (i = 1, \cdots, a), \tag{87}$$

and a *probability distribution on* X if $h \geq 0$ and $\sum_{i=1}^{a} h_{(i)} = 1$. Let W denote the compact simplex of all probability distributions on X.

For every charge distribution $h \in R^a$, the set $\{i|h_{(i)} \neq 0\}$ is called the *carrier* of h. Two probability distributions p, q on X have disjoint carriers if and only if $\|p - q\| = 2$. A polyhedron $W_0 \subseteq W$, whose vertices (= extremal points) have mutually disjoint carriers, is always a *simplex*. Let γ_r denote the system of all simplices with r vertices such that any two vertices have disjoint carriers $(r = 1, \cdots, a)$. Clearly, $\gamma_1 = \{\{p\}|p \in W\}$, $\gamma_a = \{W\}$. The simplices of γ_1 are the only ones in $\bigcup_{r=1}^{a} \gamma_r$ with norm diameter less than 2.

For closed nonempty subsets W_1, W_2 of R^a we introduce the *set distance*

$$|W', W''| = \sup_{h' \in W'} \inf_{h'' \in W''} \|h' - h''\| + \sup_{h'' \in W''} \inf_{h' \in W'} \|h'' - h'\|. \tag{88}$$

The closed nonempty subsets of W form a compact metric space, and $\gamma_1, \cdots, \gamma_a$ are compact subsets of this space. For later use we extract from Jacobs [2] the following lemmas.

LEMMA 7.1. *If $W_1, W_2 \in \gamma_r$, $W_1 \subseteq W_2$, then $W_1 = W_2$.*

LEMMA 7.2. *There is a monotone function $\eta(\theta) > 0$, defined for $\theta > 0$ sufficiently small, and fulfilling $\eta(\theta) < \frac{1}{4}$ and*

$$\lim_{\theta \to 0+0} \eta(\theta) = 0 \tag{89}$$

such that the following statements hold.

(i) *Let U, $\bar{U} \in \gamma_r$, and let $e^1, \cdots, e^r$ be the vertices of U, and $\bar{e}^1, \cdots, \bar{e}^r$ the vertices of $\bar{U}$. Let θ be such that $\eta(\theta)$ is defined, and assume $|U, \bar{U}| < \theta$. Then there is a unique permutation π of $\{1, \cdots, r\}$ such that*

$$\|\bar{e}^{\rho\pi} - e^\rho\| < \eta(\theta), \qquad (1 \leq \rho \leq r) \tag{90}$$

holds.

(ii) *Let $\hat{U} \in \gamma_r$ with vertices $\hat{e}^1, \cdots, \hat{e}^r$, and let $\bar{\pi}$ be a permutation. Assume that*

$$\|\hat{e}^{\rho\bar{\pi}} - e^\rho\| < \eta(\theta), \qquad (1 \leq \rho \leq r). \tag{91}$$

Then

$$\|\hat{e}^{\rho\pi\bar{\pi}} - e^\rho\| < 2\eta(\theta), \qquad (1 \leq \rho \leq r), \tag{92}$$

that is, when permutations multiply, deviations add.

A linear mapping P of R^a into itself is called *stochastic* if

$$WP \subseteq W. \tag{93}$$

If we identify P with its matrix $(P(i, k))$, $(i, k = 1, \cdots, a)$, then (93) is tantamount to

$$\begin{aligned} &P(i, k) \geq 0, \\ &\sum_{k=1}^{\infty} P(i, k) = 1. \end{aligned} \tag{94}$$

A system γ of nonempty closed subsets of W is called *stochastically connected*, if for any two sets $W_1, W_2 \in \gamma$ there are stochastic mappings P, Q such that $W_1P = W_2$, $W_2Q = W_1$. From

$$\|hP\| \leq \|h\|, \qquad (h \in R^a, P \text{ stochastic}), \tag{95}$$

that is, the *contraction* property of stochastic mappings, the following lemma is easily deduced.

LEMMA 7.3. *Let γ be a stochastically connected system of nonempty closed subsets of W. Assume that there is an r such that $\gamma \cap \gamma_r \neq 0$. Then $\gamma \subseteq \gamma_r$.*

PROOF. Let $W_0 \in \gamma \cap \gamma_r$ and $W_1 \in \gamma$. As W_1 is a stochastic image of W_0, its vertices are at most r in number, and at mutual distances ≤ 2. Since W_0 is a stochastic image of W_1, the vertices of W_1 are at least r, hence exactly r, in number, and at mutual distances ≥ 2. Hence $W_1 \in \gamma_r$.

2. *Recurrent sequences of stochastic mappings.* Let $\mathfrak{M}$ be the system of all stochastic mappings in R^a. The norm

$$\|L\| = \sup_{\|h\| \leq 1} \|hL\| \tag{96}$$

for linear mappings in R^a makes $\mathfrak{M}$ a compact metric space. Assume that a recurrent sequence

$$P_1, P_2, \cdots \tag{97}$$

of stochastic mappings $P_t \in \mathfrak{M}$ is given. Choose a fixed sequence $\mathfrak{t}$: $t_0, t_1, \cdots$ along which the sequence (97) is recurrent:

$$\|P_{t+t_k} - P_t\| \xrightarrow[k]{} 0, \qquad (t = 1, 2, \cdots). \tag{98}$$

(That 1 instead of 0 is the first index for the sequence (97) should not cause any trouble.)

Let

$$\begin{aligned} {}_sP_t &= P_s, \cdots, P_t, && (0 < s \leq t), \\ {}_tP_s &= 1, && (s < t). \end{aligned} \tag{99}$$

Our aim is the investigation of the asymptotic behavior of the sequence ${}_1P_1, {}_1P_2, {}_1P_3, \cdots, {}_1P_t, \cdots$. First of all we look for the asymptotic behavior of the sequence $W\,{}_1P_t$, $(t = 1, 2, \cdots)$ of nonempty polyhedra $\subseteq W$, employing the set distance.

THEOREM 7.1. *There is an integer r with $1 \leq r \leq a$ such that all limit points of the sequence $W\,{}_1P_t$, $(t = 1, 2, \cdots)$ belong to γ_r. If any $W\,{}_1P_t$ has a diameter <2, then $r = 1$. In other terms, the shape of $W\,{}_1P_t$ becomes more and more that of a simplex whose vertices have mutually disjoint carriers; if the diameter of $W\,{}_1P_t$ ever becomes <2, it tends to 0.*

PROOF. Applying the norm continuity of matrix multiplication, we can find a subsequence $\mathfrak{U}$: $u_0, u_1, \cdots$ of the sequence $\mathfrak{t}$: $t_0, t_1, \cdots$, and a sequence $Q_1, Q_2, \cdots \in \mathfrak{M}$ (indeed the Q_k will be of the form ${}_sP_t$) such that

$$u_k > u_0 + \cdots + u_{k-1}, \qquad (k > 0), \tag{100}$$

and

$${}_1P_{u_0+\cdots+u_k} = {}_1P_{u_0+\cdots+u_{k-1}} Q_k \quad {}_{1+u_k}P_{u_0+\cdots+u_{k-1}+u_k}, \qquad (k = 1, 2, \cdots), \tag{101}$$

$$\|{}_uP_v - {}_{u+u_k}P_{v+u_k}\| < 2^{-k}, \qquad (1 \leq u, v \leq u_0 + \cdots + u_{k-1}). \tag{102}$$

Shortening this, let us put

$$P_{(k)} = {}_1P_{u_0+\cdots+u_k}, \tag{103}$$

$$P_{[k]} = {}_{1+u_{k+1}}P_{u_0+\cdots+u_k+u_{k+1}}. \tag{104}$$

Then (101) becomes

$$P_{(k+1)} = P_{(k)}Q_kP_{[k]}, \qquad (k = 0, 1, \cdots), \tag{105}$$

and (102) implies

$$\|P_{[k]} - P_{(k)}\| \underset{k}{\to} 0. \tag{106}$$

It is our aim to pass (using (106)) from (105) to a limit relation of the form $P = PQP$. In order to obtain all the convergences needed, we have to use the compactness of $\mathfrak{M}$, and to pass to subsequences of the sequences $P_{(k)}$, Q_k. The problem arises whether (105) or a similar relation still holds true after passage to subsequences.

For any $k, \ell > 0$, let

$$P_{[k,\ell]} = {}_{1+u_{k+1}+\cdots+u_{k+\ell}}P_{u_0+\cdots+u_k+u_{k+1}+\cdots+u_{k+\ell}}. \tag{107}$$

Then there is a $Q_{k\ell} \in \mathfrak{M}$ such that

$$P_{(k+\ell)} = P_{(k)}Q_{k\ell}P_{[k,\ell]}. \tag{108}$$

We obtain

$$\begin{aligned} \|P_{[k,\ell]} - P_{(k)}\| &= \|{}_{1+u_{k+1}+\cdots+u_{k+\ell}}P_{u_0+\cdots+u_{k+\ell}} - {}_1P_{u_0+\cdots+u_k}\| \\ &\leq \sum_{j=k+1}^{k+\ell-1} \|{}_{1+u_{k+1}+\cdots+u_j+u_{j+1}}P_{u_0+\cdots+u_k+u_{k+1}+\cdots+u_j+u_{j+1}} \\ &\quad - {}_{1+u_{k+1}+\cdots+u_j}P_{u_0+\cdots+u_k+u_{k+1}+\cdots+u_j}\| \\ &\quad + \|{}_{1+u_{k+1}}P_{u_0+\cdots+u_k+u_{k+1}} - {}_1P_{u_0+\cdots+u_k}\| \\ &< \sum_{j=k+1}^{k+\ell-1} 2^{-j} + 2^{-(k+1)} < 2^{-k}, \end{aligned} \tag{109}$$

by (102). If $k_1, k_2, \cdots$ is a strictly increasing sequence of positive integers, and if we put

$$P_{(k_\nu)} = P'_\nu, \qquad Q_{k_\nu,k_{\nu+1}-k_\nu} = Q'_\nu, \qquad P_{[k_\nu,k_{\nu+1}-k_\nu]} = P''_\nu, \tag{110}$$

then (105) results in

$$P'_{\nu+1} = P'_\nu Q'_\nu P''_\nu, \tag{111}$$

and we have

$$\|P'_\nu - P''_\nu\| < 2^{-\nu} \underset{\nu}{\to} 0. \tag{112}$$

Choosing $k_1, k_2, \cdots$ appropriately, we may enforce the convergence of P'_ν towards some $P \in \mathfrak{M}$. By (112) then $P''_\nu \to P$ also follows. For any limit point $Q \in \mathfrak{M}$ of the sequence $Q'_1, Q'_2, \cdots \in \mathfrak{M}$, we now obtain the desired relation

$$P = PQP. \tag{113}$$

From this we deduce

$$WQP = WQPQP \subseteq WP = WPQP \subseteq WQP; \tag{114}$$

hence,

$$WP = WQP = W(QP)^2 = \cdots . \tag{115}$$

Well-known results from the theory of *stationary* finite-state Markov chains now imply that WP is a simplex $\subseteq W$ whose vertices have mutually disjoint carriers: $WP \in \gamma_r$ for some r. As P is the limit of the P'_ν which always have the form ${}_1P_t$, we obtain:

(a) *the system γ of all limit points of the sequence W_1P_t intersects some γ_r.* If we now show that

(b) *the system γ is stochastically connected,* the theorem is proved by lemma 7.3.

Let $W_0, W_1 \in \gamma$. There are sequences $v_\nu, w_\nu \to \infty$ such that $W\,{}_1P_{v_\nu} \to W_0$, $W\,{}_1P_{w_\nu} \to W_1$. We may assume $v_\nu < w_\nu$ and ${}_{v_\nu+1}P_{w_\nu} \to Q \in \mathfrak{M}$; hence, $W_0Q = W_1$ is easily obtained. By symmetry (b) follows.

THEOREM 7.2. *Let $\mathfrak{t}$: $t_0, t_1, \cdots$ be such that* (98) *holds, and let $r > 0$ be as in theorem* 7.1. *Furthermore, let $\mathfrak{t}'$: $t'_0, t'_1, \cdots$ be any subsequence of $\mathfrak{t}$ such that the sequence $W\,{}_1P_{t_k'}$ converges* (*in set distance*) *towards some $W_0 \subseteq W$; by compactness such a $\mathfrak{t}'$ always exists. Put*

$$W_u = W_0\,{}_1P_u, \qquad (u = 0, 1, \cdots). \tag{116}$$

Then

$$W_u \in \gamma_r, \qquad (u = 0, 1, \cdots), \tag{117}$$

$$W_u \subseteq W\,{}_1P_u, \qquad (u = 0, 1, \cdots), \tag{118}$$

$$\lim_{u\to\infty} |W_u, W\,{}_1P_u| = 0. \tag{119}$$

PROOF. From theorem 7.1, $W_0 \in \gamma_r$ follows. Using obvious relations between set distance and transformation norm (see Jacobs [2]), we obtain for any $u \geq 0$

$$\begin{aligned} W_u = W_0\,{}_1P_u &= \lim_k W\,{}_1P_{t_k'}\,{}_1P_u \\ &= \lim_k W\,{}_1P_{t_k'}\,{}_{t_k'+1}P_{t_k'+u} \\ &= \lim_k W\,{}_1P_{t_k'+u}. \end{aligned} \tag{120}$$

By theorem 7.1, the latter limit belongs to γ_r. The same theorem 7.1 also yields the other statements. Of course, one should like to remove the passage from $\mathfrak{t}$ to a subsequence $\mathfrak{t}'$ from the statement of the above theorem. This can be achieved in the following two cases: (a) all P_t are one-dimensional projections, and (b) the sequence P_t $(t = 1, 2, \cdots)$ is almost periodic (see Jacobs [2], [3]). The difficulty in the general case results from the possible nonuniformity of certain limits. In general the passage from $\mathfrak{t}$ to some $\mathfrak{t}'$ is unavoidable, as is shown by the following simple example.

EXAMPLE. Let $a = 2$ and $p_1, p_2, \cdots$ be any nonconvergent sequence of points in W. Let $\mathbf{t}$: $t_0, t_1, \cdots$ be such that $t_0 = 0$, $t_k - t_{k-1} \geq 2$, $(k = 1, 2, \cdots)$, $t_k - t_{k-1} \underset{k}{\rightarrow} \infty$, and construct a recurrent sequence $P_1, P_2, \cdots \in \mathfrak{M}$ such that P_{t_k} is the identity matrix $(k = 0, 1, \cdots)$ and $WP_{t_k-1} = \{p_k\}$, $(k = 1, 2, \cdots)$. This can easily be done by the method exhibited in section 2, subsection 3(6). Clearly, ${}_1P_{t_k} = P_{t_k-1}$, $(k = 1, 2, \cdots)$ is a nonconvergent sequence of one-dimensional stochastic projections, and $W\, {}_1P_{t_k}$ is nonconvergent.

THEOREM 7.3. *Let* $\mathbf{t}$: $t_0, t_1, \cdots$ *be such that* (98) *holds and* $\lim_k W\, {}_1P_{t_k} = W_0$ *exists. Then the sequence* $W_t = W_0\, {}_1P_t$, $(t = 0, 1, \cdots)$ *is recurrent along* $\mathbf{t}$:

$$\lim_k |W_{u+t_k}, W_u| = 0, \qquad (u = 0, 1, \cdots). \tag{121}$$

PROOF. By the proof of theorem 7.2,

$$W_u = \lim_k W\, {}_1P_{t_k+u}. \tag{122}$$

On the other hand, $W_0 \subseteq W$ implies

$$W_{u+t_k} \subseteq W\, {}_1P_{t_k+u}. \tag{123}$$

All limit points, for $k \rightarrow \infty$, of the left side belong to γ_r, as γ_r is compact and $W_{u+t_k} \in \gamma_r$. They are all contained in the limit W_u of the right side, which is also in γ_r. Consequently, W_{u+t_k} has, for $k \rightarrow \infty$, only one limit point, namely W_u (lemma 7.1). This proves the theorem.

Let $r > 0$ be determined according to theorem 7.1. A simplex $W_0 \in \gamma_r$ is called *admissible* with respect to the sequence $P_1, P_2, \cdots$ if $W_u = W_0\, {}_1P_u \in \gamma_r$, $(u = 0, 1, \cdots)$, and if the sequence $W_0, W_1, \cdots \in \gamma_r$ is recurrent along a sequence, along which the sequence $P_1, P_2, \cdots$ is also recurrent.

THEOREM 7.4. *Let* W_0 *be an admissible simplex. Then there is a sequence* $\mathfrak{U}$: $u_0, u_1, \cdots$ *such that for every* $p \in W_0$ *the sequence* $p_t = p\, {}_1P_t$, $(t = 0, 1, \cdots)$ *is recurrent along* $\mathfrak{U}$:

$$p_{v+u_k} \rightarrow p_v, \qquad (v = 0, 1, \cdots). \tag{124}$$

PROOF. Let $e^1, \cdots, e^r$ be an enumeration of the r vertices of the simplex W_0. Put $e^\rho_t = e^\rho\, {}_1P_t$. Then $e^1_t, \cdots, e^r_t$ are the r vertices of the simplex W_t. It is sufficient to find $\mathfrak{U}$: $u_0, u_1, \cdots$ such that

$$e^\rho_{v+u_k} \underset{k}{\rightarrow} e^\rho_v, \qquad (1 \leq \rho \leq r;\ v = 0, 1, \cdots) \tag{125}$$

holds.

If $r = 1$, we can—according to theorem 7.3—take $\mathfrak{U} = \mathbf{t}$, where $\mathbf{t}$ is any sequence such that (98) holds.

If $r > 1$, the idea of proof is the following. The recurrence of the sequence $W_0, W_1, \cdots$ implies that every W_u is again and again nearly congruent with simplices W_{u+t}. By lemma 7.2, such an approximate congruence implies approximate coincidence of vertices. However, the vertex numbers will in general not coincide; as a matter of fact, there will be some permutation of vertex numbers. The main objective of the technique applied below will be to obtain identity permutation of vertex numbers. For this purpose we observe that permutations multiply if subsequent time intervals are glued together. We shall glue together $r!$ time intervals whose attached permutations are equal, thus obtaining the identity permutation.

The details of the proof are a bit involved. Choose $\epsilon > 0$, $w > 0$, and determine $\theta > 0$ such that $\eta(\theta)$ is defined and is $< \epsilon, < \frac{1}{4}$, according to lemma 7.2. Passing, if necessary, to a subsequence of $\mathfrak{t}$, we may write

$$(126) \qquad |W_{v+u}, W_{v+u+t_k}| < \frac{\theta}{r!+1}, \qquad (k \geq 0, 0 \leq v \leq w, 0 \leq u \leq t_1 + \cdots + t_{k-1}),$$

$$(127) \qquad \|{}_{v+1+t_k}P_{v+u_k+t_k} - {}_{v+1}P_{v+u_k}\| < \eta(\theta) \qquad (0 \leq v \leq w;\ 0 \leq u \leq t_1 + \cdots + t_{k-1}).$$

For $u = 0$, we obtain

$$(128) \qquad |W_v, W_{v+t_k}| < \theta, \qquad (0 \leq v \leq w).$$

Lemma 7.2 now implies the existence of permutations π_{vk} of $(1, \cdots, r)$ such that

$$(129) \qquad \|e_{v+t_k}^{\rho\pi_{vk}} - e_v^\rho\| < \eta(\theta), \qquad (1 \leq \rho \leq r;\ 0 \leq v \leq w).$$

Passing once more to a subsequence of $\mathfrak{t}$, we may write

$$(130) \qquad \pi_{v1} = \pi_{v2} = \cdots = \pi_v.$$

Now put

$$(131) \qquad u_0 = 0, \qquad u_k = t_0 + \cdots + t_{k-1}, \qquad (k > 0).$$

From (126) we deduce

$$(132) \qquad \left.\begin{aligned} |W_v, W_{v+u_k}| &< \theta \\ |W_v, W_{v+u_k+t_k}| &< \theta \\ |W_{v+u_k}, W_{v+u_k+t_k}| &< \theta \end{aligned}\right\} \qquad (0 \leq k \leq r, 0 \leq v \leq w).$$

By lemma 7.2, there exist unique permutations

$$(133) \qquad {}_{v+1}\pi_{v+u_k},\ {}_{v+1+u_k}\pi_{v+u_k+t_k},\ {}_{v+1}\pi_{v+t_k} = \pi_v,\ {}_{v+1}\pi_{v+u_k+t_k}$$

such that

$$(134) \qquad {}_{v+1}\pi_{v+u_k+t_k} = {}_{v+1}\pi_{v+u_k}\ {}_{v+1+u_k}\pi_{v+u_k+t_k}$$

holds. Denoting approximations up to $\eta(\theta)$ by $\approx$ and putting ${}_{v+1+u_k}\pi_{v+u_k+t_k} = \pi$ for short, we obtain (using (128) and (127))

$$(135) \qquad \begin{aligned} e_{v+u_k+t_k}^{\rho\pi_v} &= e_{v+t_k}^{\rho\pi_v}\ {}_{v+1+t_k}P_{v+u_k+t_k} \\ &\approx e_v^\rho\ {}_{v+1+t_k}P_{v+u_k+t_k} \approx e_v^\rho\ {}_{v+1}P_{v+u_k} \\ &= e_{v+u_k}^\rho \approx e_{v+u_k+t_k}^{\rho\pi}. \end{aligned}$$

The sum of deviations is $< 3\eta(\theta) < 1$. Hence, the approximation is sufficiently close to yield $\pi_v = \pi$, and consequently,

$$ {}_{v+1+u_k}\pi_{v+u_k+t_k} = {}_{v+1}\pi_{v+t_k} = \pi_v. \tag{136} $$

This shows that

$$ {}_{v+1}\pi_{v+u_k+t_k} = {}_{v+1}\pi_{v+u_k}\pi_v. \tag{137} $$

Thus we obtain inductively

$$ {}_{v+1}\pi_{v+u_k} = \pi_v^k, \tag{138} $$

and hence, for $k = r!$,

$$ {}_{v+1}\pi_{v+u_k} = 1. \tag{139} $$

This implies

$$ \|e^\rho_{v+u_k} - e^\rho_v\| < \eta(\theta) < \epsilon, \qquad (0 \leq v \leq w). \tag{140} $$

Clearly, the construction can be carried out so as to make u_k as large as desired. The result of our construction may thus be described as follows: for every $\epsilon > 0$ and every $w > 0$, there are arbitrarily large u such that

$$ \|e^\rho_{v+u} - e^\rho_v\| < \epsilon, \qquad (0 \leq v \leq w). \tag{141} $$

This implies the statement of the theorem.

THEOREM 7.5. *Let W_0 be an admissible simplex and $e_0^1, \cdots, e_0^r$ be an enumeration of its vertices. Put $W_t = W_0\, {}_1P_t$, $e_t^\rho = e_0^\rho\, {}_1P_t$, $(t \geq 0;\ 1 \leq \rho \leq r)$. Let $u \geq 0$ be fixed and $p \in W$ such that the carrier of p is contained in the carrier of some e_u^ρ. Put $p_t = p\, {}_{u+1}P_t$, $(t \geq u)$. Then the carrier of p_t is contained in the carrier of e_t^ρ, $(t \geq u)$ and*

$$ \|p_t - e_t^\rho\| \underset{t}{\rightarrow} 0. \tag{142} $$

PROOF. If W_0 is admissible with respect to $P_1, P_2, \cdots$, then W_u is admissible with respect to $P_{u+1}, P_{u+2}, \cdots$. Hence, it is sufficient to prove the theorem in the case $u = 0$.

There is an $\alpha > 0$ such that $p_0 \leq \alpha e_0^\rho$. This implies $p_t \leq \alpha e_t^\rho$, and the statement concerning carriers follows. Formula (142) follows now from $p_t \in W\, {}_1P_t$ and (119). Indeed, a vector in W which is close to W_t, and whose carrier is contained in the carrier of e_t^ρ, is close to e_t^ρ (it keeps distance 2 from all other e_0^σ $(\sigma \neq \rho)$). The details of the argument are left to the reader (see also Jacobs [2]).

3. *Mixing and recurrence of Markov measures.* Let us now pass from state space X to shift space Ω (see section 2, subsection 4, example 2)).

THEOREM 7.6. *Let W_0 be an admissible simplex, and p one of its vertices. In $\Omega = \prod_{t \geq 0} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X\}$, let T be the shift, and m the Markov measure with initial distribution p and transition matrices $P_1, P_2, \cdots$. Then m is mixing along* $\mathbf{t}$: $0, 1, \cdots$:

$$ \lim_{t \to \infty} \left| \int fT^t g\, dm - \int fT^t\, dm \int g\, dm \right| = 0, \qquad (f, g \in C(\Omega)). \tag{143} $$

PROOF. By the Stone-Weierstrass theorem, it is sufficient to prove (143) in the special case where $f = 1_F$, $g = 1_G$; F, G are special finite-dimensional cylinder sets, for example,

$$(144) \qquad F = [y_0, \cdots, y_s], \qquad G = [z_0, \cdots, z_n], \qquad (y_0, \cdots, y_s, z_0, \cdots, z_n \in X).$$

For $t > u$ we have

$$\int fT^t \, dm = (p\,{}_1P_t)(y_0)P_{t+1}(y_0, y_1) \cdots P_{t+s}(y_{s-1}, y_s),$$

$$(145) \qquad \int g \, dm = p(z_0)P_1(z_0, z_1) \cdots P_u(z_{u-1}, z_n),$$

$$\int fT^t g \, dm = p(z_0)P_1(z_0, z_1) \cdots P_u(z_{u-1}, z_u)$$
$$\qquad {}_{u+1}P_t(z_u, y_0)P_{t+1}(y_0, y_1) \cdots P_{t+s}(y_{s-1}, y_s).$$

If z_u is not in the carrier of $p\,{}_1P_u$, then $\int fT^t g \, dm = \int g \, dm = 0$ $(t > u)$, and (143) holds trivially. If z_u is in the carrier of $p\,{}_1P_u$, we need only apply theorem 7.5 to every vector q which corresponds to unit mass in z_u, in order to obtain $\|{}_{u+1}P_t(z_u, \cdot) - p_t\| \to 0$. By the above formulae this implies (143).

THEOREM 7.7. *Let W_0 be an admissible simplex and let $p \in W_0$ be arbitrary. In $\Omega = \Pi_{t \geq 0} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X\}$, let T be the shift, and m the Markov measure with initial distribution p and transition matrices $P_1, P_2, \cdots$. Let $\mathfrak{U}$: $u_0, u_1 \cdots$ be determined according to theorem 7.4. Then m is recurrent along $\mathfrak{U}$:*

$$(146) \qquad mT^{u_k} \underset{k}{\to} m, \qquad \text{(weakly).}$$

PROOF. By the Stone-Weierstrass theorem, it is sufficient to prove

$$(147) \qquad \int fT^{u_k} \, dm \underset{k}{\to} \int f \, dm$$

for $f = 1_F$ where $F = [y_0, \cdots, y_s]$, with $y_0, \cdots, y_s \in X$ arbitrarily chosen. Formulae (145), (100), and (124) (the latter for $v = 0$) then show that

$$(148) \qquad \int fT^{u_k} \, dm = (p_1P_{u_k})(y_0)P_{1+u_k}(y_0, y_1) \cdots P_{s+u_k}(y_{s-1}, y_s)$$
$$\to p(y_0)P_1(y_0, y_1) \cdots P_s(y_{s-1}, y_s) = \int f \, dm.$$

8. Applications to channel theory

Let X, X' be nonempty finite sets and

$$(149) \qquad \Omega = \prod_{t=0}^{\infty} X_t = \{\omega = (x_0, x_1, \cdots) | x_t \in X (t = 0, 1, \cdots)\},$$

$$(150) \qquad \Omega' = \prod_{t=0}^{\infty} X'_t = \{\omega' = (x'_0, x'_1, \cdots) | x'_t \in X' \ (t = 0, 1, \cdots)\}$$

be the corresponding one-sided infinite product spaces. If we put $\tilde{X} = X \times X'$, then $\tilde{\Omega} = \Omega \times \Omega'$ may be considered as the one-sided product space corresponding to $\tilde{X}$. Shift will be denoted by T in each of the spaces Ω, Ω', $\tilde{\Omega}$. Note that they may be considered as compact metric spaces, the shift being continuous. Indicator functions of finite dimensional cylinder sets are continuous; their linear combinations are dense in the Banach space of all continuous functions on each of these spaces (the Stone-Weierstrass theorem). Hence, we may express most statements concerning weak topology in terms of cylinder sets as well as in terms of continuous functions.

Let us recall the language of information theory. The sets X and X' are called the input and output alphabets; the points of Ω and Ω' are called the input and output messages respectively. Probability distributions on Ω, Ω', and $\tilde{\Omega}$ are called input and output and compound sources respectively. A stochastic kernel C from Ω to Ω' is called a channel. An input source m, together with a channel C, determines an output source $m' = mC$, and a compound source $\tilde{m} = m \times C$, according to

$$m'(F') = \int_{\Omega} m(d\omega)\, C(\omega, F'), \qquad (F' \subseteq \Omega' \text{ measurable}) \tag{151}$$

$$\tilde{m}(F \times F') = \int_{F} m(d\omega)\, C(\omega, F'), \qquad (F \subseteq \Omega,\ F' \subseteq \Omega' \text{ measurable}), \tag{152}$$

DEFINITION 8.1. *A channel C is said to be*

(a) stationary, *if*

$$C(\omega T, F') = C(\omega, F'T^{-1}), \qquad (\omega \in \Omega,\ F' \subseteq \Omega' \text{ measurable}); \tag{153}$$

(b) periodic with period d, *if*

$$C(\omega T^{d}, F') = C(\omega, F'T^{-d}), \qquad (\omega \in \Omega,\ F' \subseteq \Omega' \text{ measurable}); \tag{154}$$

(c) recurrent *along* $\mathfrak{t}$: $t_0, t_1, \cdots$, *if*

$$C(\omega T^{t_k}, F') - C(\omega, F'T^{-t_k}) \underset{k}{\rightarrow} 0, \qquad (\omega \in \Omega,\ F' \subseteq \Omega' \text{ finite-dimensional cylinder set}); \tag{155}$$

(d) mixing *along* $\mathfrak{t}$, *if each* $C(\omega, \cdot)$ *in* Ω' *is mixing along* $\mathfrak{t}$;

(e) *a channel with* finite past history $h \geq 0$, *if* $C(\omega, F')$ *depends only on components* $x_{s-h}, \cdots, x_s, \cdots, x_t$ *of* $\omega \in \Omega$ *for every cylinder set* F' *depending only on the components* x_s *through* x_t *in* Ω';

(f) continuous, *if* $C(\cdot, F') \in C(\Omega)$ *for every finite-dimensional cylinder set* $F' \subseteq \Omega'$;

(g) memoryless, *if there is a sequence* $P_0, P_1, \cdots$ *of stochastic matrices from* X *to* X' *such that*

$$C(\omega, \cdot) = \prod_{t=0}^{\infty} P_t(x_t, \cdot), \qquad (\omega = (x_0, x_1, \cdots) \in \Omega). \tag{156}$$

The correspondence between C *and* $(P_0, P_1, \cdots)$ *is, of course, one-to-one.*

Channels with finite past history clearly are always continuous. Memoryless channels are always mixing, have finite past history; they are stationary, periodic with period d, recurrent along $\mathfrak{t}$, according to whether the sequence $P_0, P_1, \cdots$ is constant, periodic with period d, recurrent along $\mathfrak{t}$, respectively. If C is stationary and $mT = m$, then $\tilde{m}T = \tilde{m}$, $m'T = m'$. If C and m have period d, so have $\tilde{m}$ and m'. If C is continuous, and m and C are recurrent along $\mathfrak{t}$, then $\tilde{m}$ and m' are recurrent along $\mathfrak{t}$. If C is continuous, and C and m are recurrent and mixing along $\mathfrak{t}$, then $\tilde{m}$ and m' are recurrent and mixing along $\mathfrak{t}$. The proofs are left to the reader.

DEFINITION 8.2. (1) *Let P be a channel, and ω, $\bar{\omega}$ two input messages. Then ω and $\bar{\omega}$ are called* discernible through P, *if the probability distributions $P(\omega, \cdot)$, $P(\bar{\omega}, \cdot)$ are orthogonal. If $M \subseteq \Omega$ is such that $\omega, \bar{\omega} \in M$, $\omega \neq \bar{\omega}$ implies discernibility of ω, $\bar{\omega}$ through P, then M is called discernible through P.*

(2) *Let P, $\overline{P}$ be channels. They are called* discernible, *if there is a source m such that mP and $m\overline{P}$ are orthogonal. This is equivalent to the existence of an input message $\omega \in \Omega$ such that $P(\omega, \cdot)$ and $\overline{P}(\omega, \cdot)$ are orthogonal in Ω'.*

THEOREM 8.1. *Let C be a channel which is recurrent and mixing along $\mathfrak{t}$: $t_0, t_1, \cdots$. Let $\omega, \eta \in \Omega(\mathfrak{t}')$ for some subsequence $\mathfrak{t}'$ of $\mathfrak{t}$ be such that*

$$C(\omega, \cdot) \neq C(\eta, \cdot). \tag{157}$$

Then ω and η are discernible through C, that is,

$$C(\omega, \cdot) \perp C(\eta, \cdot). \tag{158}$$

PROOF. The hypotheses imply $C(\omega, \cdot), C(\eta, \cdot) \in V(\mathfrak{t}')$, hence (157) implies (158) by theorems 5.2 and 4.2.

COROLLARY. *Let C be a stationary channel which is mixing along $\mathfrak{t}$: $0, 1, \cdots$ (for instance, a stationary memoryless channel). Let ω, η be two input messages produced in two independent experiments by the same recurrent input source. Then with probability one, ω and η are discernible through C if $C(\omega, \cdot) \neq C(\eta, \cdot)$.*

PROOF. The correct model for the production of the pair of messages (ω, η) is the space $\Omega \times \Omega$ with the product measure $m \times m$. It is rather trivial that $m \times m$ is recurrent (with respect to shift in $\Omega \times \Omega$) if m is recurrent. Thus, by theorem 5.4, $m \times m$-almost every pair $(\omega, \eta) \in \Omega \times \Omega$ belongs to $(\Omega \times \Omega)(\mathfrak{t}')$ for some subsequence $\mathfrak{t}'$ of $\mathfrak{t}$: it has components ω, η which are both recurrent along a sequence $\mathfrak{t}'$, along which C is also recurrent. The corollary then follows from theorem 8.1.

THEOREM 8.2. *Let C, $\overline{C}$ be two continuous channels, both recurrent and mixing along the same $\mathfrak{t}$: $t_0, t_1, \cdots$. Assume that $mC \neq m\overline{C}$ for some source m which is recurrent and mixing along $\mathfrak{t}$. Then $mC \perp m\overline{C}$; hence, C and $\overline{C}$ are discernible.*

The proof is trivial from the preliminary discussion and theorem 8.1.

THEOREM 8.3. *Let C be a continuous channel which is mixing and recurrent along $\mathfrak{t}$: $t_0, t_1, \cdots$. For every $t \geq 0$, let $C^t(\omega, F') = C(\omega, F'T^{-t})$, $C_t(\omega, F') = C(\omega T^t, F')$. Then C^t and C_t are continuous channels which are mixing and recurrent along $\mathfrak{t}$. If $C^t \neq C_t$, then C^t and C_t are discernible.*

PROOF. All statements are obvious except for the last one. Now, $C^t \neq C_t$ implies the existence of a finite-dimensional cylinder set $F' \subseteq \Omega'$ and a point $\omega = (x_0, x_1, \cdots) \in \Omega$ such that $C^t(\omega, F') \neq C_t(\omega, F')$. By continuity, we may find some $s > 0$ such that $C^t(\eta, F') \neq C_t(\eta \cdot F')$ for all $\eta = (y_0, y_1, \cdots)$ such that $y_0 = x_0, \cdots, y_s = x_s$. Passing if necessary to a subsequence of $\mathfrak{t}$, we may assume $t_{k+1} - t_k \xrightarrow[k]{} \infty$.

The construction given in section 2, subsection 3, (6) now yields a point $\eta = (y_0, y_1, \cdots) \in \Omega(\mathfrak{t})$ such that $y_0, \cdots, y_s$ have the prescribed values $x_0, \cdots, x_s$. Clearly, $m = \delta_\eta$ is recurrent and mixing along $\mathfrak{t}$, and $(mP^t)(F') = P^t(\eta, F') \neq P_t(\eta, F') = mP_t(F')$. The preceding theorem now implies $mP^t \perp mP_t$, as required.

COROLLARY. *Let P be a continuous channel which is mixing and recurrent along some sequence $\mathfrak{t}$. Then there are only the following three possibilities:*

(1) *P is stationary,*

(2) *P is periodic,*

(3) *P is wandering in the sense that P^t and P_t are discernible for every $t > 0$.*

PROOF. If (1) and (2) are not true, then $P^t \neq P_t$ for every $t > 0$. The preceding theorem then shows that (3) is true. This result applies especially to memoryless channels arising from recurrent sequences $\Pi_0, \Pi_1, \cdots$ of stochastic matrices.

With every channel P, a linear positive contraction operator, also denoted by P, from the Banach lattice $R(B)$ of all finite charge distributions in Ω into the Banach lattice $R(B')$ of all finite charge distributions in Ω', is associated. Let

$$\|Q\| = \sup_{R(B)\ni h,\ \|h\|\leq 1} \|hQ\| \tag{159}$$

denote the norm of any linear operator Q: $R(B) \to R(B')$. Discernibility of P and $\overline{P}$ implies $\|P - \overline{P}\| = 2$.

Our above corollary thus shows that a channel with the properties mentioned is never recurrent with respect to operator norm topology, except when it is periodic. In a special case of recurrence (namely almost periodicity), and for memoryless channels, this result has been announced in Jacobs [5].

REFERENCES

[1] A. HAJIAN and S. KAKUTANI, "Weakly wandering sets and invariant measures," *Trans. Amer. Math. Soc.*, Vol. 110 (1964), pp. 136–151.

[2] K. JACOBS, "Fastperiodische diskrete Prozesse von endlicher Dimension," *Abh. Math. Sem. Univ. Hamburg*, Vol. 21 (1957), pp. 194–246.

[3] ———, "Fastperiodische Markoffsche Prozesse," *Math. Ann.*, Vol. 134 (1958), pp. 408–427.

[4] ———, *Neuere Methoden und Ergebnisse der Ergodentheorie*, Berlin, Springer, 1960.

[5] ———, "Almost periodic channels," *Symp. Comb. Math. in Prob.*, Aarhus, Aarhus University Press, 1962, pp. 118–126.

[6] ———, "Lectures on ergodic theory," Aarhus, Aarhus University Press, 1963.

[7] M. KAC, "On the notion of recurrence in discrete stochastic processes," *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 1002–1010.

[8] S. KAKUTANI, "On equivalence of infinite product measures," *Ann. of Math.*, Vol. 49 (1948), pp. 214–224.

[9] M. MORSE and G. HEDLUND, "Symbolic dynamics," *Amer. J. Math.*, Vol. 60 (1938), pp. 815–866.

[10] YU. V. PROHOROV, "Convergence of random processes and limit theorems in probability theory," *Teor. Verojatnost. i Primenen.*, Vol. 1 (1956), pp. 177–238.

[11] L. SUCHESTON, "On existence of finite invariant measures," *Math. Z.*, Vol. 86 (1964), pp. 327–336.

[12] J. WOLFOWITZ, "Remarks on the notion of recurrence," *Bull. Amer. Math. Soc.*, Vol. 55 (1942), pp. 394–395.

[13] A. RÉNYI, "On mixing sequences of sets," *Acta Math. Akad. Sci. Hungar.*, Vol. 9 (1958), pp. 215–228.