The proof holds in any number of dimensions, if the constant in Harnack's inequality is changed properly.

* Presented to the American Mathematical Society, Dec. 28, 1931.

† NATIONAL RESEARCH FELLOW.

[1] O. Perron, *Math. Zeit.*, **18,** 42–54 (1923); we follow the notations in this paper.

[2] R. Remak, *Math. Zeit.*, **20,** 126–130 (1924); T. Radó and F. Riesz, *Ibid.* **22,** 41–44 (1925); R. Remak, *J. I. Math.*, **156,** 227–230 (1926). The proof here given is in essence like that of the last-named paper.

[3] I. Petrowsky, *Rec. Math. Moscou*, **35,** 105–110 (1928); N. Wiener, *J. Math. Phys. Mass. Inst. Techn.*, **4,** 21–32 (1925).

---

# PROOF OF THE QUASI-ERGODIC HYPOTHESIS

BY J. V. NEUMANN

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 10, 1931

1. The purpose of this note is to prove and to generalize the quasi-ergodic hypothesis of classical Hamiltonian dynamics[1] (or "ergodic hypothesis," as we shall say for brevity) with the aid of the reduction, recently discovered by Koopman,[2] of Hamiltonian systems to Hilbert space, and with the use of certain methods of ours closely connected with recent investigations of our own of the algebra of linear transformations in this space.[3] A precise statement of our results appears on page 79.

We shall employ the notation of Koopman's paper, with which we assume the reader to be familiar. The Hamiltonian system of $k$ degrees of freedom corresponding with the Hamiltonian function $H(q_1, \ldots, q_k, p_1, \ldots, p_k)$ defines a steady incompressible flow $P \longrightarrow P_t = S_tP$ in the space $\Phi$ of the variables $(q_1, \ldots, q_k, p_1, \ldots, p_k)$ or "phase-space," and a corresponding steady conservative flow of positive density $\rho$ in any invariant sub-space $\Omega \subset \Phi$ ($\Omega$ being, e.g., the set of points in $\Phi$ of equal energy). The Hilbert space $\mathfrak{H}$ consists of the class of measurable functions $f(P)$ having the finite Lebesgue integral $\int_\Omega |f|^2\rho d\omega$, the "inner product"[4] of any two of them $(f, g)$ and "length" $\|f\|$ being defined by the equations

$$(f, g) = \int_\Omega f\bar{g}\rho d\omega; \quad \|f\| = \sqrt{(f,f)}. \tag{1}$$

The transformation $U_t$ is defined as follows:

$$U_tf(P) = f(S_tP) = f(P_t); \tag{2}$$

obviously it has the group property

$$U_t U_s = U_{t+s},\ U_0 = I; \tag{3}$$

and in virtue of the conservative character of the flow, and the resulting invariance of $(U_t f,\ U_t g)$, it is unitary. The spectral reduction of $U_t$ in terms of its "canonical resolution of the identity" $E(\lambda)$[5] is furnished by a theorem due to Stone,[6] and gives us

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda), \tag{4}$$

this being the symbolic expression for the fact that, for all $f$, $g$ of $\mathfrak{H}$, we have, in terms of Stieltjès integrals,

$$(U_t f,\ g) = \int_{-\infty}^{+\infty} e^{it\lambda}\, d(E(\lambda)f,\ g). \tag{4'}$$

The pith of the idea in Koopman's method resides in the conception of the spectrum $E(\lambda)$ reflecting, in its structure, the properties of the dynamical system—more precisely, those properties of the system which are true "almost everywhere," in the sense of Lebesgue sets.

The possibility of applying Koopman's work to the proof of theorems like the ergodic theorem was suggested to me in a conversation with that author in the spring of 1930. In a conversation with A. Weil in the summer of 1931, a similar application was suggested, and I take this opportunity of thanking both mathematicians for the incentive which they furnished me for undertaking the investigations of this paper.

2. For the sake of brevity, we shall introduce the following notation: We shall replace $\rho d\omega$ by $dv$, writing $\int_\Omega - \rho d\omega = \int_\Omega - dv$. By the "weight $\mu\Theta$ of the Lebesgue-measurable set $\Theta(\subset\Omega)$ with respect to the density $\rho$" will be meant the quantity $\mu\Theta = \int_\Theta \rho d\omega = \int_\Theta dv$. By a "zero set" we shall mean a set of zero weight, and hence, since throughout $\Omega$, $0 < \rho_1 < \rho < \rho_2$, a set of zero Lebesgue measure.

If $\Theta$ is a set of points $P$ of $\Omega$ or $\Phi$, we shall denote its characteristic function by $\chi_\Theta = \chi_\Theta(P)$; i.e., $\chi_\Theta(P) = 1$ or $0$ according as $\Theta$ does or does not contain $P$. If $f(P)$ is any measurable function, the set of points for which $f(P) > \lambda$, etc., will be denoted as usual by $[f(P) > \lambda]$, etc. We have the identity $[\chi_{[f>\lambda]} = 1] = [f > \lambda]$, etc.

By the strong convergence of a sequence $f_1, f_2, \ldots$ in $\mathfrak{H}$ $(f_n \longrightarrow f)$ will be meant that $\|f_n - f\| \longrightarrow 0$ as $n \longrightarrow \infty$. By weak convergence $(f_n \stackrel{\cdot}{\longrightarrow} f)$ we mean, on the other hand, that for an arbitrarily chosen $g$ of $\mathfrak{H}$, $(f_n,\ g) \longrightarrow (f,\ g)$ as $n \longrightarrow \infty$. It is shown that $f_n \longrightarrow f$ implies $f_n \stackrel{\cdot}{\longrightarrow} f$, but not conversely.[7] In general, expressions depending for their precise meaning on the nature of the convergence considered will be suffixed by the corresponding convergence symbol, thus we shall write "separable ($\longrightarrow$)," "everywhere dense ($\dot{\longrightarrow}$)," etc. All these notions subsist if $n$ is replaced by one or more continuously varying parameters.

By $\mu$-convergence of a sequence of point sets $\Theta_1, \Theta_2, \ldots$ ($\subset \Omega$ or $\Phi$) will be meant the strong convergence of the corresponding measure functions: $\Theta_n \longrightarrow \Theta$ if $\chi_{\Theta n} \longrightarrow \chi_\Theta$, or, what is the same thing, $\mu[\Theta_n + \Theta - \Theta_n\Theta] \longrightarrow 0$ as $n \longrightarrow \infty$. Clearly $\Theta$, $\underline{\lim}\ \Theta_n$, and $\overline{\lim}\ \Theta_n$, will all differ by at most zero sets.

The greatest lower bound and least upper bound of a set [ ] will be denoted, as usual, by inf [ ] and sup [ ].

3. The starting point of our investigations is the construction of the operator

$$\sigma_{t,s} = \frac{1}{t-s}\int_s^t U_\tau d\tau \quad (s < t), \tag{5}$$

this being, as before, but the symbolic expression of the fact that for all $f, g$ in $\mathfrak{H}$,

$$(\sigma_{t,s} f, g) = \frac{1}{t-s}\int^t (U_\tau f, g)d\tau; \tag{5'}$$

the existence of $\sigma_{t,s}$ is easily proved.[8] We will show that, for each $f$ of $\mathfrak{H}$, $\sigma_{t,s}f$ is convergent ($\longrightarrow$) as $t - s \longrightarrow \infty$, irrespectively of the mode of variation of $s$, $t$.

We have from (5′):

$$\|\sigma_{t,s}f\|^2 = (\sigma_{t,s}f, \sigma_{t,s}f) = \frac{1}{t-s}\int_s^t (U_\tau f, \sigma_{t,s}f)d\tau$$
$$= \frac{1}{(t-s)^2}\int_s^{\cdot}\int_s^t (U_\tau f, U_\sigma f)d\epsilon d\sigma;$$

since $U_\sigma$ is unitary and $U_\sigma^* = U_\sigma^{-1} = U_{-\sigma}$,[9]

$$\|\sigma_{t,s}f\|^2 = \frac{1}{(t-s)^2}\int_s^t\int_s^t (U_{\tau-\sigma}f, f)d\tau d\sigma,$$

which reduces, on making the change of variables $\tau - \sigma = x$, $\tau + \sigma = y$, to

$$\frac{1}{(t-s)^2}\int_{-(t-s)}^{+(t-s)}\int_{2s+|x|}^{2t-|x|} (U_x f, f)\frac{dx\,dy}{2}$$
$$= \frac{1}{(t-s)^2}\int_{-(t-s)}^{+(t-s)} (t - s - |x|)(U_x f, f)dx.$$

This may be calculated with the aid of (4′), inasmuch as the various changes in the order of integration are permissible, on account of the uniform convergence of the Stieltjès integral in (4′)[10] (for all values of $t$—at present, $x$). Thus:

$$\|\sigma_{t,s}f\|^2 = \frac{1}{(t-s)^2}\int_{-(t-s)}^{+(t-s)} (t - s - |x|)\left[\int_{-\infty}^{+\infty} e^{ix\lambda}\, d(E(\lambda)f, f)\right]dx$$

$$= \frac{1}{(t-s)^2}\int_{-\infty}^{+\infty}\left[\int_{-(t-s)}^{+(t-s)} e^{ix\lambda}\,(t - s - |x|)dx\right]d(E(\lambda)f, f)$$

$$= \frac{2}{(t-s)^2}\int_{-\infty}^{+\infty}\left[\int_{0}^{t-s} \cos (x\lambda).(t - s - x).dx\right]d(E(\lambda)f, f)$$

$$= \frac{2}{(t-s)^2}\int_{-\infty}^{+\infty}\frac{1 - \cos (t-.s)\lambda}{\lambda^2}\,d(E(\lambda)f, f)$$

$$= \int_{-\infty}^{+\infty}\left[\frac{\sin {}^1/_2(t-s)\lambda}{{}^1/_2(t-s)\lambda}\right]^2 d(E(\lambda)f, f).$$

This integral has a non-negative integrand and a non-decreasing expression after the $d$-sign; hence we may obtain an upper bound for it as follows: First, break it up into $\int_{-\epsilon}^{+\epsilon}$ and $\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}$; ($\epsilon > 0$, to be considered later). Then replace the integrand in the first part by 1, and that in the second part by $\left[\frac{1}{{}^1/_2(t-s)\epsilon}\right]^2 = \frac{4}{(t-s)^2\epsilon^2}$. Finally, replace the field of integration in the second by $\int_{-\infty}^{+\infty}$. We shall then have

$$\|\sigma_{t,s}f\|^2 \leq \int_{-\epsilon}^{+\epsilon} d(E(\lambda)f, f) + \frac{4}{(t-s)^2\epsilon^2}\int_{-\infty}^{+\infty} d(E(\lambda)f, f)$$

$$= \left\{(E(\epsilon)f, f) - (E(-\epsilon)f, f)\right\} + \frac{4}{(t-s)^2\epsilon^2}(f, f).$$

Hence, as $t - s \longrightarrow +\infty$, $\overline{\lim}\ \|\sigma_{t,s}f\|^2 \leq (E(\epsilon)f, f) - (E(-\epsilon)f, f)$, and if, as $\epsilon \longrightarrow 0$ $\left\{E(\epsilon) - E(-\epsilon)\right\}f \longrightarrow 0$, we shall have, on letting $\epsilon \longrightarrow 0$, $\|\sigma_{t,s}f\|^2 \longrightarrow 0$, so that $\sigma_{t,s}f \longrightarrow 0$.

We now introduce the projection operator $E_0$ defined as follows: $(E(\epsilon) - E(-\epsilon))f \longrightarrow E_0 f$ as $\epsilon \longrightarrow 0$. The existence and projective nature of $E_0$ is easily deduced from the fact that $E(\epsilon) - E(-\epsilon)$ is a non-increasing "function" of $\epsilon$.[11] Thus, we are able to express the condition that $\sigma_{t,s}f \longrightarrow 0$ as $t - s \longrightarrow 0$ in the form: $E_0 f = 0$.

Suppose that $E_0 f = f$; then, since $E(\epsilon) - E(-\epsilon) \geqslant E_0$, we have $E(\epsilon)f = f$, $E(-\epsilon)f = 0$,[12] i.e., $E(\lambda)f = f$ for $\lambda > 0$, $= 0$ for $\lambda < 0$. Hence, for all $g$,

$$(U_t f, g) = \int_{-\infty}^{+\infty} e^{it\lambda}\, d(E(\lambda)f, g) = (f, g),$$

so that, for all values of $t$, $U_t f = f$.

Let $\mathfrak{M}$ be the linear manifold in $\mathfrak{H}$ corresponding with $E_0$. For every

$f$ of $\mathfrak{M}$, $E_0 f = f$; hence $U_t f = f$, and $\sigma_{t,s} f = f$, so that as $t - s \longrightarrow \infty$ we have

$$\sigma_{t,s} f \longrightarrow f = E_0 f. \tag{6}$$

For all $f$ orthogonal to $\mathfrak{M}$ on the other hand, we have

$$\sigma_{t,s} f \longrightarrow 0 = E_0 f. \tag{6'}$$

Now let $f$ be an arbitrary point of $\mathfrak{h}$; we can write $f = f_1 + f_2$, where $f_1$ is in $\mathfrak{M}$, and $f_2$, orthogonal to $\mathfrak{M}$. Then it will follow from (6), (6′), that we still have, as $t - s \longrightarrow 0$,

$$\sigma_{t,s} f \longrightarrow E_0 f. \tag{6''}$$

Throughout $\mathfrak{M}$, $U_t f = f$. Conversely, if $U_t f = f$, it will follow that $\sigma_{t,s} f = f$, and hence by (6″), $f = E_0 f$, i.e., $f$ belongs to $\mathfrak{M}$. Thus, $\mathfrak{M}$ is the class of all solutions of the equation $U_t f = f$ (i.e., the identity in $t$).

4. Let us examine $\mathfrak{M}$ more closely. Its elements $f$ are characterized by $U_t f = f$, and hence, in virtue of (2), by

$$f(P) \equiv f(P_t), \tag{7}$$

the $\equiv$ sign holding for all $t$ but with the possible exception of a zero set of points $P$ (in general, dependent on $t$). Hence, if $f$ is in $\mathfrak{M}$, $\mathfrak{F}(f)$ will be also, provided $\int |\mathfrak{F}(\lambda)|^2 d\lambda$ is finite.

Now $f$ can be expressed as the limit ( $\longrightarrow$ ) of functions of the form $\mathfrak{F}(f)$ where $\mathfrak{F}$ is susceptible of but a finite number of values, and these, in their turn, are linear combinations of similar functions susceptible only of the values 0 and 1.[13] The latter, being of the form $\mathfrak{F}(f)$, belong to $\mathfrak{M}$. If we denote by $\mathfrak{K}$ the class of all functions belonging to $\mathfrak{M}$ and taking on only the values 0 and 1, we may say that $\mathfrak{K}$ spans ( $\longrightarrow$ ) the closed ( $\longrightarrow$ ) linear manifold $\mathfrak{M}$.

If $f$ belongs to $\mathfrak{K}$, we shall write $[f(P) = 1] = \Lambda$, and $f = f_\Lambda (= \chi_\Lambda)$, (cf. § 2). Since $\mu\Lambda = \|f\|^2$, and $f$ is in $\mathfrak{H}$, $\mu\Lambda$ must be finite; and since $f$ is in $\mathfrak{K} \supset \mathfrak{M}$, it follows from (7) that the transformation $P \longrightarrow P_t$ changes $\Lambda$ by at most a zero set. These two properties, its finite weight and invariance under $P \longrightarrow P_t$, characterize $\Lambda$. We shall call any set having these properties a $\Lambda$-set. Evidently $\Omega$ will be a $\Lambda$-set if and only if $\mu\Omega$ is finite.

The class of all $\Lambda$-sets, being a subclass of the class of all measurable sub-sets of $\Omega$, is separable;[14] that is, there exists a sequence $\Lambda_1$, $\Lambda_2$, ... of $\Lambda$-sets such that any $\Lambda$-set can be expressed as the limit (in the sense of $\mu$-convergence) of a subsequence of $\Lambda_1, \Lambda_2, \ldots$. By methods which we have used in another connection,[15] we are able to replace the sequence $\Lambda_1, \Lambda_2, \ldots$ by another sequence of $\Lambda$-sets, $\Lambda_1', \Lambda_2', \ldots,$

such that, firstly, any member of one sequence may be expressed in terms of elements of the other with the finite repetition of the operations of taking the logical sum ($\dotplus$), the logical product ($\times$) and the logical difference ($-$); secondly, $\Lambda_1'$, $\Lambda_2'$, ... may be set into one to one correspondence with certain rational numbers $\rho_n = \rho(\Lambda_n')$, $\Lambda_n' = \Lambda(\rho_n)$, $0<\rho_n<1$, in such a manner that $\rho_m < \rho_n$ implies $\Lambda(\rho_m)C\Lambda(\rho_n)$; and, thirdly, $\inf \rho_n = 0$ and $\sup \rho_n = 1$.

We now define the function $G(P)$ as follows:

$$G(P) \begin{cases} = \inf[\rho_n \text{ for which } P \text{ is in } \Lambda(\rho_n)]; \\ = 1, \text{ when no such } \rho_n \text{ exists.} \end{cases}$$

By its construction, $G(P)$ is invariant under $P \longrightarrow P_t$ (remaining unchanged apart from zero sets); and $\inf G(P) = 0$, $\sup G(P) = 1$. Since $[G(P) \leq \rho_n] = \Lambda(\rho_n)$, it follows that $f_{\Lambda_n'} = f_{\Lambda(\rho_n)} = F(G)$ (for we may set $F(\lambda) = \Lambda$ for $\lambda \leq \rho_n$, $= 0$ for $\lambda > \rho_n$); and therefore this property remains true for every $f_{\Lambda_n}$,[16] and any $f_\Lambda$ of $\mathfrak{K}$ (cf. definition of $\longrightarrow$ for sets). And since every $f$ of $\mathfrak{M}$ is the limit ( $\longrightarrow$ ) of a sequence of linear combinations of functions of $\mathfrak{K}$, it follows that every such $f$ is a function of $G$.[17] Finally if $\lambda < \Lambda$, $\mu[G(P) \leqslant \lambda]$ is finite. For a $\rho_n > \lambda$ may be found, whereupon $[G(P) \leqslant \lambda] \subset [G(P) \leqslant \rho_n] = \Lambda(\rho_n) = \Lambda_n'$, and $\mu\Lambda_n' = \|f_{\Lambda_n'}\|^2$, which is finite for any $f_{\Lambda_n'}$ of $\mathfrak{K}(\subset\mathfrak{H})$.

Any function $G(P)$ like the above, such that $G(P_t) = G(P)$ for all $t$ except perhaps at zero-sets, such that $\lambda' = \inf G(P)$ and $\lambda'' = \sup G(P)$ exist, and that, if $\lambda < \lambda''$, $\mu[G(P) \leqslant \lambda]$ is finite, and which possesses the property that every $f$ of $\mathfrak{M}$ may be expressed as $\mathfrak{F}(G)$, shall be called a universal integral. We have shown that one universal integral always exists; obviously there are infinitely many.[18]

The class $\mathfrak{M}$ coincides with the totality of expressions $\mathfrak{F}(G)$ (with $\int_\Omega |\mathfrak{F}(G(P))|^2 dv$ finite). This makes it possible to express $E_0 f$ in terms of $G$. We shall give below, instead of our original method of computation, an abbreviated method for which we are grateful to Mr. M. H. Stone.

5. For an arbitrary $f$ of $\mathfrak{H}$, $E_0 f$ belongs to $\mathfrak{M}$, and is, accordingly, of the form $\mathfrak{F}(G)$. Let $\bar{\lambda} < \lambda''$ ($= \sup G$), and define $\zeta(\lambda)$ to be 1 for $\lambda < \bar{\lambda}$ and 0 for $\lambda \leqslant \bar{\lambda}$. Let us set $\Lambda(\lambda) = [G(P) \leq \lambda]$ (we have seen that $\mu\Lambda(\lambda)$ is finite). Then we have, on the one hand,

$$\int_\Omega E_0 f(P)\zeta(G(P))dv = \int_\Omega \mathfrak{F}(G(P))\zeta(G(P))dv^{20}$$

$$= \int_{\lambda'}^{\lambda''} \mathfrak{F}(\lambda)G(\lambda)d\mu\Lambda(\lambda) = \int_{\lambda'}^{\bar{\lambda}} \mathfrak{F}(\lambda)d\mu\Lambda(\lambda),$$

and, on the other hand,

$$\begin{aligned}\int_\Omega E_0 f(P)\zeta(G(P))dv &= (E_0 f,\ \zeta(G)) \\ &= (f,\ E_0\zeta(G))^{21} = (f,\ \zeta(G)) \\ &= \int_\Omega f(P)\zeta(G(P))dv = \int_{\Lambda(\bar\lambda)} f(P)dv.\end{aligned}$$

Thus:

$$\int_{\lambda'}^{\bar\lambda} \mathfrak{F}(\lambda)d\mu\Lambda(\lambda) = \int_{\Lambda(\bar\lambda)} f(P)dv. \tag{8}$$

As $\lambda$ increases from $\lambda'$ to $\lambda''$, $\mu\Lambda(\lambda)$ goes in a non-decreasing fashion from 0 to $\mu\Omega$; and $\mu\Lambda(\lambda + 0) = \mu\Lambda(\lambda)$. In intervals $\lambda_1 \leq \lambda < \lambda_2$ where $\mu\Lambda(\lambda)$ is constant, $\Lambda(\lambda)$ changes at most by points $P$ of a zero set, i.e., $\lambda_1 < G(P) < \lambda_2$ can be true at most on a zero set; thus the behavior of $\mathfrak{F}(\lambda)$ in such intervals does not affect the relation $E_0 f = \mathfrak{F}(G)$—we may take $\mathfrak{F}(\lambda)$ constant upon them. It follows that the familiar theorems on the differentiation of Lebesgue integrals may be applied, the independent variable being here $x = \mu\Lambda(\lambda)$. From such considerations it follows that $\dfrac{d\{\int_{\Lambda(\lambda)} f(P)dv\}}{d\{\mu\Lambda(\lambda)\}}$ exists for all $\lambda$ in $\lambda' \leq \lambda < \lambda''$, except for a set of values of $\lambda$ for which the corresponding set of values $x = \mu\Lambda(\lambda)$ is of zero measure,[22] and this derivative is equal to $\mathfrak{F}(\lambda)$. The correspondence $P \longrightarrow x$, obtained by setting $\lambda = G(P)$, $x = \mu\Lambda(\lambda)$, carries a set $\Theta \subset \Omega$ into a set $\theta$ on the $x$-axis so that $\mu\Theta$ = measure of $\theta$.[23]

Thus we have, except for at most a zero set on $\Omega$,

$$E_0 f(P) = \left[\frac{d\{\int_{\Lambda(\lambda)} f(P)dv\}}{d\{\mu\Lambda(\lambda)\}}\right]_{\lambda = f(P)}. \tag{9}$$

This is naturally true for $\lambda = \lambda''$ only when $x = \mu\Lambda(\lambda'')$ exists, i.e., when $\mu\Omega$ is finite.

In the case $\lambda = G(P) = \lambda''$, we carry out the above process with $\zeta(\lambda) = 1$ for $\lambda = \lambda''$, $= 0$ for $\lambda \neq \lambda''$. On setting $\Lambda = [G(P) = \lambda'']$, the following becomes clear: If $\mu\Lambda = 0$, $\mathfrak{F}(\lambda'')$ does not affect $E_0 f(P)$. If $\mu\Lambda = \infty$, $E_0 f(P)$ must be zero; for it is constant on $\Lambda$, and belongs to $\mathfrak{H}$. If $\mu\Lambda$ is $> 0$ and finite, the considerations which lead to (8) show that

$$\mathfrak{F}(\lambda'')\cdot\mu\Lambda = \int_\Lambda f(P)dv. \tag{8'}$$

Since $\Lambda = \Lambda(\lambda'') - \Lambda(\lambda'' - 0) = \Omega - \Lambda(\lambda'' - 0)$, it follows that

$$E_0 f(P) = \frac{\int_{\Omega - \Lambda(\lambda'' - 0)} f(P)dv}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \text{ (for } G(P) = \lambda''\text{)}. \tag{9'}$$

This formula subsists when $\mu[\Omega - \Lambda(\lambda'' - 0)] = 0$ or $\infty$, provided we agree to replace the right-hand member by 0 in the case where the denominator vanishes (and hence the numerator also) or is infinite.

When $\mu\Omega$ is finite, (9′) leads to (9) (cf. [22]); otherwise, it forms its natural generalization.

6. Let $M$ and $N$ be any measurable sub sets of $\Omega$, with $\mu M$, $\mu N$ finite. Then $\chi_M(P)$ and $\chi_N(P)$ are in $\mathfrak{H}$, and we may apply our results to them. It follows from (5′) that $(\sigma_{t,s}\,\chi_N,\ \chi_M)$, which equals $\int_M \sigma_{t,s}\,\chi_N(P)dv$, is equal to

$$\frac{1}{t-s}\int_s^t \left\{ \int_\Omega U_\tau \chi_N(P)\cdot\bar{\chi}_M(P)dv \right\} d\tau$$

$$= \frac{1}{t-s}\int_s^t \mu(S_\tau N \times M)d\tau \quad (S_t P = P_t,\ \text{etc.})$$

$$= \int_m Z_{s,t}\,(P,\ N)dv,$$

where we have set $Z_{s,t}\,(P,\ N)$ equal to $\frac{1}{t-s}$ times the linear measure of the set of $\tau$-values for which $S_{-\tau}P\,(= P_{-\tau})$ is on $N$.[24] That is, $Z_{s,t}(P,\ N)$ is the mean time of sojourn of $P$ in $N$ between the times $s$ and $t$ (actually, with the sign of $\tau$ changed, but this is immaterial). Since the above is true for all $M$, we have

$$\sigma_{t,s}\chi_N(P) = Z_{s,t}(P,\ N), \tag{10}$$

and in virtue of (6″),

$$Z_{s,t}(P,\ N) \longrightarrow E_0\chi_N(P) = \chi_N^0(P) \text{ as } t - s \longrightarrow \infty, \tag{11}$$

in the sense of strong convergence in $\mathfrak{H}$.

On applying (9), (9′) to $f = \chi_N$, we have

$$\chi_N^0(P) = \left[\frac{d\{\mu[\Lambda(\lambda) \times N]\}}{d\{\mu\Lambda(\lambda)\}}\right]_{\lambda = x_n(P)} \tag{12}$$

when $G(P) < \lambda''$, and

$$X_n^0(P) = \frac{\mu[\Omega - N \times \Lambda(\lambda'' - 0)]}{\mu[\Omega - \Lambda(\lambda'' - 0)]} \tag{12'}$$

when $G(P) = \lambda''$. The right-hand members have a meaning except possibly for $P$ on a zero set: in (12), cf.;[22] in (12′), we take 0 when the denominator is infinite, or when it (and hence, the numerator) vanishes.

Let us express the content of (11) in the following three ways: (A) explicitly as strong convergence in $\mathfrak{H}$; (B) as point convergence of a sub sequence;[17] (C) as weak convergence, or rather as the implication of the latter regarding the inner product of (11) with an arbitrary $X_m$ ($\mu_m$, finite).

A. $\int [Z_{s,t}\,(P,\ N) - X_n^0(P)]^2 dv \longrightarrow 0$ as $t - s \longrightarrow +\infty$.

*B.* For every sequence $t_{1,s1}$; $t_2, s_2$; ... with $t_n - s_n \longrightarrow +\infty$, there is a sub sequence $t_{n_\nu}, s_{n_\nu}$ ($\nu = 1, 2, \ldots$) such that for all $P$ of $\Omega$ with the possible exception of a zero set, $Z_{s_{n_\nu}}, t_{n_{q_\nu}}(P, N) \longrightarrow X^0_n(P)$ as $v \longrightarrow \infty$.

*C.* For each subset $M$ of $\Omega$ of finite $\mu M$, $\int_m Z_{s,t}(P, N)dv \longrightarrow \int_M \chi^0_N(P)dv$ as $t - s \longrightarrow +\infty$.

We observe that although $\chi^0_N$ is expressed in (12), (12′) in terms of the non-uniquely determined universal integral $G$, its dependence upon $G$ is only apparent: each of (11), $A$, $B$ or $C$, determines $\chi^0_N$ uniquely.

7. The existence, for each point $P$, of the limit of the mean sojourn $Z_{s,t}(P, M)$ is a consequence of $A$, $B$ or $C$, and applies to any Hamiltonian system. Our system is *ergodic* if and only if this limit, $\chi^0_N(P)$, is independent of $P$, i.e., when

$$X^0_N(P) = C_n, \text{ a constant.} \tag{13}$$

When this is true, we must have in the case $\mu\Omega = \infty$ that $C_n = 0$; for otherwise, $\|\chi^0_N\| = \infty$, whereas $\chi^0_N$ is in $\mathfrak{H}$. But when $\mu\Omega$ is finite, we have: $\int \chi^0_N(P)dv = (\chi^0_N, 1) = (E_0\chi_N, 1) = (\chi_N, E_0 1) = (\chi_N, 1) = \int \chi_N(P)dv = \mu N$. Hence, by (13),

$$C_n = \frac{\mu N}{\mu\Omega}. \tag{13′}$$

This is obviously true, from what was said earlier, when $\mu\Omega = \infty$.

It is now a simple matter to tell whether the system is ergodic or not, and we do not even need the more complete results of § § 4 and 5.

First, suppose that (13) (and consequently (13′)) is true for arbitrary $N$. Let $f$ be an element of $\mathfrak{M}$. Then, on the one hand, we have

$$\int \chi^0_N(P)\bar{f}(P)dv = \frac{\mu N}{\mu\Omega}\int_\Omega \bar{f}(P)dv = \bar{C}\mu\cdot N,$$

and, on the other hand,

$$\begin{aligned}\int_\Omega \chi^0_N(P)\bar{f}(P)dv &= (\chi^0_N, f) = (E_0\chi_N, f) = (\chi_N, E_0 f) = (\chi_N, f)\\ &= \int \chi_N(P)\bar{f}(P)dv = \int_N \bar{f}(P)dv.\end{aligned}$$

From the equality of the final expressions for all $N$, we conclude that $f(P) = C$. Secondly, suppose that, conversely, every function of $\mathfrak{M}$ is a constant. Then $\chi^0_N$, belonging to $\mathfrak{M}$, is a constant, and (13) is true.

Thus the system will be ergodic if and only if $\mathfrak{M}$ consists exclusively of constants. This will be true if and only if $\mathfrak{K}$ consists exclusively of constants,[13] i.e., that the $\Lambda$-sets all differ from 0 or from $\Omega$ at most by zero sets. Thus we have proved the theorem:

E. *The system is ergodic if and only if every measurable set $\Lambda$ remaining invariant under $S_t$ (except for points of a zero set) reduces to* 0 *or to $\Omega$ (except for points of a zero set).*[25]

Since in $E$ the ergodic condition is the non-existence of any measurable $\Lambda$-sets ($\neq 0, \Omega$), one might be tempted to suppose that the ergodic condition as stated in (13) would have to hold for a correspondingly broad class of sets. This, however, is not the case: If (13) is true for all open sets $N$ of finite $\mu N$, it will be true, by continuity, for all $\mu$-limits of such sets (cf. §2),—i.e., for all measurable sets $N$ of finite $\mu N$.[26] Indeed, it is only necessary to require its truth for sets $N$ which are the sums of a finite number of the neighborhoods of an arbitrary "topologically equivalent system of neighborhoods" in $\Omega$,[26] for instance, for sums of finite numbers of spheres.

8. From a purely mathematical standpoint, the question as to the validity and most appropriate generalization of the ergodic hypothesis has been fully answered: these *special* problems have been reduced to the *general* problem of the integrals of the system—the structure of $G(P)$. Thus, the system is either ergodic, or else there is a non-constant $G(P)$, in which case $\Omega$ is decomposable into subsets like $[G(P) = \lambda_1]$, $[\lambda_1 \leq G(P) < \lambda_2]$, etc., upon which the flow has a sort of ergodic character, as is easily shown by means of (12), (12').

But from the point of view of physics, there remains the difficult question as to the existence and nature of $G(P)$ in each particular case. It might happen that there are integrals of the system in the classical sense, i.e., analytic, or at least continuously differentiable, as would be true, for example, if $G(P)$ were of this character; in which case they could be used for the reduction of the dimensionality of $\Omega$ (cf.,[2] p. 315, last line). Or it might happen that no such integrals exist, in spite of the fact that $G(P)$ is non-constant. Conceivably this last situation is impossible when the Hamiltonian $H$ is analytic (or even continuously differentiable); if so, the proof of this fact would be most useful. But it appears that the proof could not be obtained alone from the general formal considerations in Koopman's method, i.e., from (3) and from

$$U_t F(f_1(P), f_2(P), \ldots) = F(U_t f_1(P), U_t f_2(P), \ldots), \qquad (14)$$

(cf.,[2] p 318). For (4), (14) remain true in the case that the one to one map $P \longrightarrow P_t$ of $\Omega$ upon itself is any one-parameter group of the following properties:

a. $P_t$ is a measurable function of $t$,

b. $P \longrightarrow P_t$ maps every measurable $P$-set in a measurable $P_t$-set with the same measure.

Since $a$, $b$ permit of discontinuities of $P_t$, it is easy to give examples with only discontinuous integrals.

Indeed, even when $P \longrightarrow P_t$ is defined by means of equations of the type

$$\frac{\partial}{\partial t} x_1 = \alpha_1(x_1, \ldots, x_l), \ldots, \frac{\partial}{\partial t} x_l = \alpha_l(x_1, \ldots, x_l), \qquad (16)$$

$(P:(x_1, \ldots, x_l))$, and when $b$ is valid—when $P \longrightarrow P_t$ is indeed an "incompressible continuous flow"—there are examples where all the integrals are discontinuous, and yet are not constants. (In an example, $l = 2$, $\alpha_1/\alpha_2$ is continuous in $P$, but $\alpha_1$, $\alpha_2$ themselves, discontinuous.)

We shall not pursue this question further.

We may observe, in conclusion, how remarkable it is that the concept of Lebesgue measure should play so important a rôle in a so essentially physical a question as the validity of the ergodic hypothesis, or, more generally, in the value of the limit of the mean sojourn, $\lim\limits_{t-s \to +\infty} Z(P, N)$.

Even in the case where $N$ is an open set or, indeed, the sum of a finite number of spheres—which has an immediate physical significance—the function $\chi_N^0(P)$ given by the above limit does only need to be measurable! In the last analysis, one is always brought to the cardinal question: "Does $P$ belong to $\Theta$ or not?" where the set $\Theta$ is merely assumed to be measurable. The opinion is generally prevalent that from the point of view of empiricism such questions are meaningless, for example, when $\Theta$ is everywhere dense—for every measurement is of limited accuracy. The author believes, however, that this attitude must be abandoned, and gives the following reason as an argument:

Suppose that $\Omega$, in which $P$ varies, have a finite measure, $m\Omega$. Since $\Theta$ is measurable, it follows from a familiar theorem of Lebesgue that

$$\lim_{\epsilon \to 0} \frac{m[\Theta \times K(P, \epsilon)]}{mK(P, \epsilon)},$$

(where $K(P, \epsilon)$ is a sphere of center $P$ and radius $\epsilon$), exists at each point of $\Theta$, and $= 1$ with the exception of a zero set.[27] Similarly for all points of $\Omega - \Theta$, where it is zero, with the same exception. The same is true when the spheres are replaced by many other sorts of figures, e.g., cubes.[27] Consider a sequence of partitions of $\Omega$ into systems of disjoint cells, $Z_1^{(n)}$, $\ldots$, $Z_{k_n}^{(n)}$ $(n = 1, 2, \ldots)$, such that the maximum diameter $\epsilon_n$ of $Z_1^{(n)}, \ldots$, $Z_{k_n}^{(n)}$ approaches zero as $n \longrightarrow \infty$. The limited accuracy of measurements finds its expression in the fact that we have to consider different order of accuracy (viz., 1, 2, ...); where, by an experiment of order of accuracy $n$, shall be meant the mere process of distinguishing in which $Z_\nu^{(n)}$ $(\nu = 1, \ldots, k_n)$ $P$ lies.

Suppose that a measurement of order $n$ has established that, for instance, $P$ lies in $Z_\nu^{(n)}$; then the (geometric) probability that $P$ belong to $\Theta$ is $\dfrac{m[Z_\nu^{(n)} \times \Theta]}{mZ_\nu^{(n)}}$. So that if

$$\frac{m[Z_\nu^{(n)} \times \Theta]}{mZ_\nu^{(n)}} < \delta \text{ or } > 1 - \delta \quad (\delta > 0), \tag{15}$$

we know with a probability $> 1 - \delta$ the answer to the question, "Is $P$ in $\Theta$ or not?" The fact that we will be able to answer this question with a probability $>1 - \delta$ of being right has the *a priori* probability (i.e., before the observation is made) of

$$w_n^{(\delta)} = \frac{\sum_\nu^{\prime(n)} m\ Z_\nu^{(n)}}{\sum\limits_{\nu=1}^{K_n} m\ Z_\nu^{(n)}} = \frac{\sum_\nu' m\ Z_\nu^{(n)}}{m\Omega},$$

where $\sum_\nu^{\prime(n)}$ represents the summation over all values of $\nu$ satisfying (15). If we could prove that $w_n^{(\delta)} \longrightarrow 1$ as $n \longrightarrow \infty$, it would become clear that, granted a sufficiently high accuracy of experiment, the above question could be answered with an arbitrarily great degree of certainty—i.e., the question has physical meaning. (This is seen by taking, e.g., $w_n^{(\delta)} > 1 - \delta$).

Suppose that $w_n^{(\delta)} \longrightarrow 1$ as $n \longrightarrow \infty$ is untrue. Then for infinitely many values of $n$, $w_n^{(\delta)} \leq 1 - \eta$ (for a certain $\eta > 0$); so that if $\sum_\nu^{\prime\prime(n)}$ implies summation over all values of $v$ for which (15) is violated, we shall have

$$m \sum_\nu^{\prime\prime(n)} Z_\nu^{(n)} \geqq \eta\ m\Omega.$$

The set $\Xi$ of all points $P$ which belong to infinitely many such sets $\sum_\nu^{\prime\prime(n)} Z_\nu^{(u)}$ will then also have a measure $\geq \eta m\Omega > 0$, in virtue of a theorem of Arzela's.[28] If $P$ is on $\Xi$, it lies on infinitely many sets $\sum_{1\nu}^{\prime\prime(n)} Z_\nu^{(n)}$; suppose it to be, for example, on $Z_{\nu_n}^{(n)}$. Since $\nu_n$ belongs, for infinitely many values of $n$, to $\sum_\nu^{\prime\prime(n)}$, (15) is violated for these values, the ratio $\dfrac{m[\Theta \times Z_{\nu_n}^{(n)}]}{mZ_{\nu_n}^{(n)}}$ determines neither the limit 0 nor 1. But this is in contradiction with the theorem of Lebesgue, in the case where the $Z_\nu^{(n)}$'s are such that its hypothesis applies (e.g., when $Z_\nu^{(n)}$'s are cubes).

[1] For the formulation and critique of this theorem, cf., e.g., *Entykl. d. Math. Wiss.*, **4**, Art. 32 on Statistical Mechanics, by P. and T. Ehrenfest, specially 30–36. The original formulations are to be found in *Wien. Ber.*, **63**, [2] 679 (1871) (Boltzmann), and *Cambr. Phil. Soc. Trans.*, **12**, 547 (1879) (Maxwell).

[2] These PROCEEDINGS, **17**, [5] 315–318 (May, 1931).

[3] Cf., e.g., the discussion in the author's paper, "Allgemein Eigenwerttheorie Hermitescher Funktionaloperatoren" (*Math. Ann.*, **102**, [1] 108–111 (1929)). This paper, as well as the author's paper, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren" (*Math. Ann.*, **102**, 3 (1929)), will be referred to in the present paper under the abbreviations E and A, respectively.

[4] Cf. E, 54–55, 109.

[5] Cf. E, 91–92.

[6] These PROCEEDINGS, **16,** [2] 172–175 (Feb., 1930); also, cf. a paper soon to appear in the *Ann. Math.*

[7] Cf., regarding these concepts, the article of Hellinger and Toeplitz in the *Math. Encyklopädie,* **2,** c. 13, 1435 (1928); further, cf. A, 378–381.

[8] Cf., e.g., the similar proof in E, 112, top.

[9] $R^*$ is the adjoint of $R$ in the terminology of matrices: the conjugate-transposed matrix. Cf., e.g., 112.

[10] The integrand, $e^{i t\lambda}$, is uniformly bounded, the expression after the $d$-sign, $(E(\gamma), f)$, of bounded variation: $\int_{-\infty}^{+\infty} d(E(\gamma)f, f) = (f, f)$.

[11] E, 91, 77–78.

[12] Cf. the theory of projection operators outlined in E, 74–78; similarly for the discussion to follow.

[13] Cf. E, 110.

[14] Cf. Hansdorff, *Mengenlehre,* 127 (1927), line 6; or E, 110.

[15] Cf. E, 110.

[16] Cf. the corresponding construction in the proof of theorem 10; in A, 401–402. There, the permutable projections $E_1, E_2, \ldots$ and $F_1, F_2, \ldots$ took the place of the $\Lambda$-sets $\Lambda_1, \Lambda_2, \ldots$ and $\Lambda_1', \Lambda_2', \ldots$; but this distinction can be abolished by replacing each $\Lambda$-set by the operator, $E_\lambda$: $f(P) \longrightarrow \chi_\Lambda(P)f(P)$.

[17] For clearly, $X_{\Lambda+M} = \chi_\Lambda + \chi_M - \chi_\Lambda \cdot \chi_M$, $\chi_{\Lambda \times M} = \chi_\Lambda \chi \cdot_M$, $\chi_{\Lambda - M} = \chi_\Lambda - \chi_\Lambda \chi_M$.

[18] If the sequency $f_1, f_2, \ldots$ converges ($\longrightarrow$), a subsequency will converge in any point, excepted a 0-set (cf., f.i. E, 111). Therefore a limit of functions of $G$ is a function of $G$.

[19] Thus, e.g., each $T(G)$ is one, if $T(\gamma)$ is a monotonically increasing function.

[20] On the other hand, our construction shows that it suffices to confine $\mathfrak{F}$ to the second Baire class (the convergence is pointwise convergence except for zero sets).

[21] This transformation from Lebesgue to Stieltjès integrals goes back to Lebesgue. Cf. *Ann. de l'École Normale,* **3,** 27 (1910), p. 407. It is sufficient to establish it for a real variable, for it is then easily extended to an arbitrary $\Omega$, which may always be mapped in a measure-preserving manner upon the real axis. Maps of this sort are given by Lebesgue (loc. cit.) for $n$-dimensional space, and may easily be extended to $\Omega$.

[22] Since $\zeta(G)$ belongs to $\mathfrak{M}$, it is left unchanged by $E_0$.

[23] At points of discontinuity of $x = \mu\Lambda(\lambda)$, where $x$ experiences a jump of a whole interval, the differential quotient has a meaning, and is equal to the difference quotient between $\lambda + 0$ and $\lambda - 0$:

$$\frac{\int_{\Lambda(\lambda+0)-\Lambda(\lambda-0)} f(P)dv}{\mu[\Lambda(\lambda + 0) - \Lambda(\lambda - 0)]}.$$

[24] Cf. the author's paper, "Über Funktionen von Funktionaloperatoren," *Ann. Math.,* **32,** [2] 196 (1931), (Satz 3), as well as the reference (20).

[25] This transformation consists in a change in the order of integration; since in the Lebesgue integrals that appear, everything is bounded, it is permissible.

[26] For $\mu\Omega = \infty$, naturally only the former will come into question.

[27] Cf. E, 110.

[28] The generalization to other figures is to be found in its broadest form in Carathéodory, *Vorlesungen über reelle Funktionen* (Leipzig-Berlin, 1918), 492–494, in particular, Theorem 3. The function $f(P)$ appearing there is to be defined as $f(P) = X\Theta(P)$.

[29] Cf., e.g., de la Valleé Poussin, *Cours d'Analyse infinitésimale,* **1,** 2 (Louvain-Paris, 1909), pp. 68–69.