PACIFIC JOURNAL OF MATHEMATICS
Vol. 20, No. 1, 1967

SOME METRICAL THEOREMS IN NUMBER THEORY

WALTER PHILIPP

In this paper some metrical theorems on Diophantine approximation, continued fractions and θ-adic expansions are proved.

In the first part some of the common properties of the following transformations from the unit interval onto itself are investigated. Denote by $\{\alpha\}$ the fractional part of x,

A. $T : \alpha \to \{a\alpha\}$ $\quad a > 1$ integer

which describes the expansion of α in the scale a

B. $T : \alpha \to \left\{\frac{1}{\alpha}\right\}$

which describes the continued fractions

C. $T : \alpha \to \{\theta\alpha\}$ $\quad \theta > 1$ noninteger

which describes the expansion of α as a θ-adic fraction.

The main theorem of the first part (Theorem 2) gives an estimate of the number of solutions of the system of inequalities

$$T^k\alpha \in I_k \qquad\qquad 1 \leqq k \leqq n$$

where n is an integer, T is any of these three transformations and (I_k) is an arbitrary sequence of intervals contained in the unit interval.

It generalises and refines well known theorems on the distribution function of the sequence $(T^k\alpha)$. Theorem 2 follows from a very general theorem—a quantitative Borel-Cantelli Lemma.

It is also shown that T is strongly mixing (Theorem 1). The second part of the paper deals with the metric theory of continued fractions. Theorems of LeVeque and Bernstein are refined.

1. Frequently a real number α is represented in one of the following ways:

A. in the scale a, where $a > 1$ is an integer,

B. as a continued fraction,

C. as a θ-adic fraction, where $\theta > 1$ is a noninteger.

Let us recall some of the properties of these representations:

A. If $a > 1$ denotes an integer then every $\alpha \in [0, 1)$ can be written as

$$\alpha = \sum_{k=1}^{\infty} \frac{c_k}{a^k} = \sum_{k=1}^{n} \frac{c_k}{a^k} + \frac{y_{n+1}}{a^n}$$

where the digits c_k are nonnegative integers less than a and $0 \leqq y_n < 1$. The representation can be made unique.

Define on $[0, 1)$ the transformation T

$$T : \alpha \to \{a\alpha\}$$

Clearly for $n \geqq 0$ we have

$$y_{n+1} = y_{n+1}(\alpha) = T^n\alpha = \{a^n\alpha\} .$$

As is well known T preserves the Lebesgue-measure and T is ergodic. (For definitions and theorems in ergodic theory see Halmos [5] pp. 5–37).

B. Every $\alpha \in (0, 1]$ can be expressed as a simple continued fraction

$$\alpha = [a_1, a_2, \cdots] \tag{1}$$

where the partial quotients $a_i = a_i(\alpha)$ are positive integers. Again, the representation can be made unique. If α is given by (1) then the finite continued fraction

$$[a_1, a_2, \cdots, a_n] = \frac{p_n}{q_n}$$

is called the n-th convergent. It is in its lowest terms.

Define on $(0, 1]$ the transformation T

$$T : \alpha \to \left\{\frac{1}{\alpha}\right\}$$

or else

$$T([a_1(\alpha), a_2(\alpha), \cdots]) = [a_2(\alpha), a_3(\alpha), \cdots] .$$

Then clearly for $n \geqq 0$ we have

$$a_{n+1}(\alpha) = a_1(T^n\alpha)$$

and

$$z_{n+1}(\alpha) = T^n\alpha = [a_{n+1}(\alpha), a_{n+2}(\alpha), \cdots] .$$

Define on $(0, 1]$ a measure μ by setting

$$\mu(E) = \frac{1}{\log 2}\int_E \frac{dx}{1+x}$$

for every Lebesgue-measurable set E. Knopp [13] proved that T is ergodic with respect to the Lebesgue-measure. Ryll-Nardzewski [21] showed that T preserves μ and that μ is equivalent to the Lebesgue-measure (i.e. both are absolutely continuous to each other).

C. For a fixed noninteger $\theta > 1$ define the transformation T on $[0, 1)$

$$T\colon \alpha \to \{\theta\alpha\} .$$

Put

$$x_1 = \alpha,\ x_2 = \{\theta x_1\},\ \cdots,\ x_{n+1} = \{\theta x_n\},\ \cdots$$
$$\lambda_1 = [\theta\alpha],\ \cdots,\ \lambda_n = [\theta x_n],\ \cdots$$

For $n \geqq 0$ we have $x_{n+1} = x_{n+1}(\alpha) = T^n\alpha$. Clearly $0 \leqq \lambda_k < \theta$, $0 \leqq x_n < 1$ and

$$\alpha = \sum_{k=1}^{\infty} \frac{\lambda_k}{\theta^k} = \sum_{k=1}^{n} \frac{\lambda_k}{\theta^k} + \frac{x_{n+1}}{\theta^n} .$$

Rényi [19] proved that there exists a unique measure μ invariant under T and equivalent to the Lebesgue measure and furthermore that T is ergodic with respect to μ.

Write $\lambda_k(1) = \lambda_k$ and $x_k(1) = t_k$. Denote by $\varphi_\gamma(t)$ the characteristic function of $[0, \gamma)$. Put (Gel'fond [4])

$$\sigma(t) = \frac{1}{\tau} \sum_{k=1}^{\infty} \frac{\varphi_{t_k}(t)}{\theta^{k-1}} \qquad \tau = \sum_{k=1}^{\infty} \frac{t_k}{\theta^{k-1}} .$$

Cigler [3] showed that this unique measure μ is defined by setting

$$\mu(E) = \int_E \sigma(\alpha) d\alpha$$

for every measurable set $E \subset [0, 1)$.

The explicit formula for the invariant measure has been also found, independently by Parry [17] who additionally remarked that T is even weakly mixing.

Now it is necessary to say a few words about the notation. In the remainder of the § 1 and in § 4 T always means any one of the three transformations and μ always stands for the invariant measure associated with T as described in Sections 1 ABC. For example Theorem 2 ABC in fact consists of three theorems and should be interpreted to mean that Theorem 2 holds for each of the three transformations and further that in Theorem 2A $\mu(I)$ — where $I = (a, b)$ is an interval—stands for $b - a$, that in Theorem 2B $\mu(I)$ means

$$\frac{1}{\log 2} \int_a^b \frac{dx}{1 + x} = \frac{1}{\log 2} (\log (1 + b) - \log (1 + a))$$

and that likewise in Theorem 2C $\mu(I)$ denotes $\int_a^b \sigma(t) dt$. Throughout the paper "almost all" always means all except a set of Lebesgue-measure 0.

In the first part of this paper some of the common properties of these transformations are investigated. I shall prove:

THEOREM 1 A, B, C. *T is strongly mixing.*

Theorem 1A is well known. It holds even for a compact connected abelian group (Hartman, Ryll-Nardzewski [9], p. 169).

This paper was already typed when Professor Krickeberg in a letter kindly called my attention to a paper of Rohlin [20]. Rohlin showed that a wide class of transformations are—what he calls—exact endomorphisms and consequently that they are mixing of every degree which implies strongly mixing. Since the proof I give is different from Rohlin's proof and since Theorem 1 is a straightforward application of some lemmas used to prove Theorem 2, I did not withdraw Theorem 1. Furthermore its proof can be easily extended to show the mixing property of every degree.

In analogy to some results on Diophantine approximation we get:

THEOREM 2A, B, C. *Let T be any of the three transformation associated with its invariant measure μ as described in* A, B, C. *Let (I_n) be an arbitrary sequence of intervals contained in the unit interval. For any positive integer N and $x \in [0, 1]$ denote by $A(N, x)$ the number of positive integers $n \leqq N$ such that $T^n x \in I_n$. Put*

$$\phi(N) = \sum_{n \leqq N} \mu(I_n) .$$

Then

$$A(N, x) = \phi(N) + O(\phi^{1/2}(N) \log^{3/2+\varepsilon} \phi(N)) \qquad \varepsilon > 0$$

for almost all $x \in [0, 1)$.

LeVeque [15] has proved theorems of the same type as Theorem 2A for arbitrary sequences (a_n) of integers instead of (a^n) under certain assumptions on the intervals I_n. Recently, his results have been extended by Walker [26]. The novelty in Theorem 2A is the arbitrariness of the intervals I_n; in particular that we can dispense the assumption that the sequence $(\mu(I_n))$ is decreasing. This is not a contradiction to a theorem of Cassels ([1], p. 215) since with Cassels' notation every subsequence of (a^n) is again a Σ-sequence and so the method of proof of Cassels' theorem does not apply to our case.

Theorem 2 is a generalization and refinement of some well known results on distribution functions of certain sequences. $\mu(x)$ is called the distribution function or distribution measure of the sequence $(x_n)(0 \leqq x_n \leqq 1)$ if for all $0 \leqq x \leqq 1$

$$\lim_{n \to \infty} \frac{A(n, x)}{n} = \mu(x) .$$

Here $A(\mathrm{n}, x)$ denotes the number of positive integers $k \leqq n$ such that $x_k < x$ (see Cigler und Helmberg [4], § 7). In each of the cases ABC the individual ergodic theorem implies at once that the sequence $(T^n\alpha)$ has the distribution function $\mu(x)$ for almost all α — $\mu(x)$ is the measure of the interval $(0, x)$ — μ invariant under T. These results are well known (H. Weyl [27], Ryll-Nardzewski [21], Gel'fond [7]). The case A follows also from the fact that $(a^n\alpha)$ is uniformly distributed for almost all α. Putting in Theorem 2 $I_n = I = (0, x)$ for $n = 1, 2, \cdots$ we get at once

COROLLARY ABC. *For $0 \leqq x \leqq 1$ denote by $A(n, \alpha, x)$ the number of positive integers $k \leqq n$ such that $T^k\alpha < x$. Then for almost all α*

$$A(n, \alpha, x) = n\mu(x) + O(n^{1/2} \log^{3/2+\varepsilon} n) \qquad \varepsilon > 0$$

where $\mu(x)$ denotes the μ-measure of the interval $(0, x)$ — μ invariant under T.

2. **A quantitative Borel-Cantelli Lemma.** Throughout the rest of the paper $|E|$ denotes the measure of E in the underlying measure-space.

THEOREM 3. *Let $(E_n, n \geqq 1)$ be a sequence of measurable sets in an arbitrary probability space. Denote by $A(N, x)$ the number of integers $n \leqq N$ such that $x \in E_n$. Put*

$$\phi(N) = \sum_{n \leqq N} |E_n| .$$

Suppose that there exists a convergent series $\sum C_k$ with $C_k \geqq 0$ such that for all integers $n > m$ we have

$$(2) \qquad |E_n \cap E_m| \leqq |E_n||E_m| + |E_n| C_{n-m} .$$

Then

$$A(N, x) = \phi(N) + O(\phi^{1/2}(N) \log^{3/2+\varepsilon} \phi(N)) \qquad \varepsilon > 0$$

for almost all x.

Theorems of this kind have been proved by LeVeque [15] and W.M. Schmidt [22] for particular sequences of sets on the real line.

In an earlier draft I obtained the error term $O(\phi^{2/3}(N)\log^{1/2+\varepsilon}\phi(N))$ using a well known device of H. Weyl [27]. However, it was pointed out to me that W.M. Schmidt's [22] modification of Rademacher's method for orthogonal sums gives a better estimate. In the following proof I shall use Schmidt's method.

Proof. In case that $\phi(\infty) < \infty$ the theorem follows from the convergence part of the Borel-Cantelli lemma even without assuming (2). So we may assume that $\phi(N) \to \infty$. Denote by $\psi_n(x)$ the characteristic function of E_n. For $m < n$ put

$$A(m, n, x) = \sum_{i=m+1}^{n} \psi_i(x)$$

and

$$\phi(m, n) = \int A(m, n, x) d\mu(x) .$$

Then clearly $\phi(0, N) = \phi(N)$ and $A(0, N, x) = A(N, x)$. Using (2) we obtain

$$\begin{aligned} &\int (A(m, n, x) - \phi(m, n))^2 d\mu(x) \\ &= 2 \sum_{m<i<j\leqq n} |E_i \cap E_j| + \phi(m, n) - \phi^2(m, n) \\ &\leqq \phi(m, n) - \phi^2(m, n) + 2 \sum_{m<i<j\leqq n} |E_i||E_j| + 2 \sum_{m<i<j\leqq n} |E_j| C_{j-i} \\ &\leqq \phi(m, n) + 2 \sum_{m<j\leqq n} |E_j| \cdot \sum_{m<i<j} C_{j-i} \\ &= O(\phi(m, n)) . \end{aligned} \tag{3}$$

For integer $u \geqq 0$ we define N_u to be the largest integer N such that $\phi(N) < u$. Denote by L_r the set of intervals $(u, v]$ with $u = t\cdot 2^s$, $v = (t+1)2^s \leqq 2^r$ where $s \geqq 0$, $t \geqq 0$ and $r \geqq 1$ are integer. Now

$$\sum \phi(N_u, N_v) \leqq \phi(N_{2^r}) < 2^r$$

where the summation is extended over all $(u, v] \in L_r$ corresponding to a fixed s since these intervals cover $(0, 2^r]$ exactly once and thus the corresponding intervals $(N_u, N_v]$ cover $(0, N_{2^r}]$ exactly once. Clearly $s \leqq r$ and so

$$\sum_{(u,v]\in L_r} \phi(N_u, N_v) < (r+1)2^r . \tag{4}$$

Put

$$Z_r = Z(r, x) = \sum_{(u,v]\in L_r} (A(N_u, N_v, x) - \phi(N_u, N_v))^2 .$$

Then (3) and (4) imply

$$\int Z_r d\mu(x) = O(r2^r)$$

or

$$\int \frac{Z^r}{2^r r^{2+\varepsilon}} d\mu(x) = O(r^{-1-\varepsilon}) .$$

Therefore

$$(5) \qquad Z_r = O(2^r r^{2+\varepsilon})$$

for almost all x. If w is an integer and $2^{r-1} < w \leqq 2^r$ then $(0, w]$ can be represented as the union of at most r intervals of L_r and thus so can $(0, N_w]$. Hence

$$A(N_w, x) - \phi(N_w) = \sum A(N_u, N_v, x) - \phi(N_u, N_v))$$

where the sum is over at most $r + 1$ intervals $(u, v] \in L_r$. This equation together with (5) and Cauchy's inequality yields

$$(A(N_w, x) - \phi(N_w))^2 = O(2^r r^{3+\varepsilon})$$

almost everywhere. Hence the theorem is true for all $N = N_w$. If N is arbitrary find w such that $N_w \leqq N < N_{w+1}$. We obviously have

$$A(N_w, x) \leqq A(N, x) \leqq A(N_{w+1}, x)$$

and

$$\phi(N_w) \leqq \phi(N) \leqq \phi(N_{w+1}) .$$

Since

$$\phi(N_{w+1}) \leqq \phi(N_w) + O(1)$$

the result follows.

3. The overlap estimates.

3A. This section deals with the situation as described in Section 1A.

LEMMA 1. *Let $P > 1$ be an integer and let $E = (x_1, x_2)$ be an interval and F be a measurable set both contained in* $[0, 1]$. *Then*

$$| E \cap S^{-1}F | = | E | \, | F | + 2P^{-1} | F | \eta$$

where $| \eta | \leqq 1$ *and S denotes the transformation* $\alpha \to \{P\alpha\}$.

Proof. If $\rho(t)$ denotes the characteristic function of F with period 1 we have

$$|E \cap S^{-1}F| = \int_{x_1}^{x_2} \rho(Pt)dt = P^{-1}\int_{Px_1}^{Px_2} \rho(t)dt$$
$$= |E||F| + 2P^{-1}|F|\eta \qquad |\eta| \leqq 1 .$$

3B. We use now the notation introduced in Section 1B. We begin with a lemma which is essentially due to Khintchine [10, 11] (see also [12] p. 89).

LEMMA 2. *Let* $E = \{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k\}$ *and let* F *be a measurable set. Then*[1]

$$|E \cap T^{-n-k}F| = |E||F|(1 + 0(q^{\sqrt{n}})) \qquad q < 1 .$$

Proof. For $0 < x \leqq 1$ denote by $\varphi_n(x)$ the (μ-)measure of the set—Khintchine used the Lebesgue-measure—

$$(6) \qquad \{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k, T^{n+k}\alpha < x\} .$$

Then $\varphi_n(x)$ satisfies the functional equation

$$(7) \qquad \varphi_n(x) = \sum_{g=1}^{\infty}\left(\varphi_{n-1}\left(\frac{1}{g}\right) - \varphi_{n-1}\left(\frac{1}{g+x}\right)\right) .$$

In fact, $T^{n+k}\alpha < x$ is equivalent with

$$0 \leqq \frac{1}{T^{n+k-1}\alpha} - \left[\frac{1}{T^{n+k-1}\alpha}\right] < x$$

or else with

$$\frac{1}{g+x} < T^{n+k-1}\alpha \leqq \frac{1}{g}$$

for $g = 1, 2, \cdots$. But this implies (7).

Now we are going to compute $\varphi_0(x)$ and its derivatives; $\varphi_0(x)$ is the measure of the set

$$M = \{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k, T^k\alpha < x\} .$$

Consequently M is just the interval with the endpoints p_k/q_k, $(p_k + p_{k-1}x)/(q_k + q_{k-1}x)$ where $p_k/q_k = [r_1, \cdots, r_k]$. This follows from the fact that $\alpha \in M$ is equivalent with

$$\alpha = [r_1, r_2 \cdots, r_k + T^k\alpha], T^k\alpha < x .$$

Hence, by computation

[1] Throughout the rest of the paper q, ρ and the constant implied by 0 are absolute constants unless otherwise stated.

$$\log 2\cdot\varphi_0(x) = (-1)^k\Big(\log\Big(1+\frac{p_k+p_{k-1}x}{q_k+q_{k-1}x}\Big)-\log\Big(1+\frac{p_k}{q_k}\Big)\Big)$$

$$\log 2\cdot\varphi_0'(x) = ((p_k+q_k)q_k + x((p_{k-1}+q_{k-1})q_k + (p_k+q_k)q_{k-1}) + q_{k-1}(p_{k-1}+q_{k-1})x^2)^{-1}$$

$$\log 2\cdot\varphi_0''(x) = -\frac{(p_{k-1}+q_{k-1})q_k + (p_k+q_k)q_{k-1} + 2q_{k-1}(p_{k-1}+q_{k-1})x}{((p_k+q_k)q_k + x((p_{k-1}+q_{k-1})q_k + (p_k+q_k)q_{k-1} + q_{k-1}(p_{k-1}+q_{k-1})x^2)^2}$$

and

$$\log 2\cdot|E| = (-1)^k \log\Big(1+\frac{(-1)^k}{(p_k+q_k)(q_k+q_{k-1})}\Big).$$

Write

$$\psi_n(x) = |E|^{-1}\varphi_n(x).$$

Then $\psi_n(x)$ also satisfies (7) and by the last set of formulas we obtain

$$0 < |\psi_0'(x)| \leqq 4 \quad\text{and}\quad |\psi_0''(x)| < 32 \qquad x\in(0,1].$$

Hence a theorem of Kuzmin yields (see [10] or [12] p. 78)

$$\Big|\psi_n'(x) - \frac{1}{\log 2}\cdot\frac{1}{1+x}\Big| \leqq c_1 q^{\sqrt{n}}$$

where c_1 and $q<1$ are absolute positive constants. Integrating from a to b we obtain

$$|\psi_n(b)-\psi_n(a) - |F|| \leqq c_1(b-a)q^{\sqrt{n}} \leqq c|F|q^{\sqrt{n}}.$$

This proves the lemma for the case that $F=(a,b)$ is an interval. Now $\psi_n(x)$ defines a normed measure ψ_n on $[0,1)$ in the usual way. We rewrite the last inequality as

$$C_n'\mu(I) \leqq \psi_n(I) \leqq C_n\mu(I)$$

where C_n', C_n are constants. It follows that this inequality holds for arbitrary measurable sets $F\subset[0,1)$.

Note. If $\varphi_n(x)$ is defined to be the Lebesgue-measure of the set (6) a theorem of Szüsz [24] gives a sharper estimate. Strangely enough the hypotheses of Szüsz' theorem are not satisfied in our case.

LEMMA 3. *Let F be a measurable set and let $E=(p_k/q_k,\, p_k'/q_k')$ where p_k/q_k, p_k'/q_k' are kth convergents. Then*

$$|E\cap T^{-(n+k)}F| = |E||F|(1+0(q^{\sqrt{n}})) \qquad q<1.$$

Proof. This is an immediate consequence of Lemma 2 since E is the union of at most countably many disjoint intervals for which the partial quotients $a_j(\alpha)$, $j \leqq k$ are constant.

LEMMA 4. *Let E be an interval and F be a measurable set. Then*

$$|E \cap T^{-n}F| = |E|\,|F| + |F|\,0(q^{\sqrt{n}}) \qquad q < 1\,.$$

Proof. Put $k = [n/2]$ and let $E = (x, y)$. Then there exist convergents p_k/q_k, p'_k/q'_k such that x and y are contained in intervals with endpoints

$$x \in \left(\frac{p_k}{q_{k'}}, \frac{p_k + p_{k-1}}{q_k + q_{k-1}}\right), \qquad y \in \left(\frac{p'_k}{q'_k}, \frac{p'_k + p'_{k-1}}{q'_k + q'_{k-1}}\right)$$

respectively. Call the intersection of E with these two intervals E_1 and E_2. Then $E = E_0 \cup E_1 \cup E_2$ where E_0 is of type described in Lemma 3. Hence by Lemma 3 and using the T preserves μ

$$\big||E \cap T^{-n}F| - |E|\,|F|\big| \leqq C\,|E_0|\,|F|\,q^{\sqrt{k}} + |F|\,(|E_1| + |E_2|)\,.$$

But for $i = 1, 2$

$$|E_i| \leqq \frac{1}{q_k^2} \leqq \frac{3}{2}\left(\frac{2}{3}\right)^k$$

Observing that $q < 1$ implies $q^{2^{-1/2}} < 1$ we get the result.

Later we need

LEMMA 5. *Let $E = \{\alpha \mid a_1(\alpha) \geqq M\}$ and $F = \{\alpha \mid a_1(\alpha) \geqq N\}$ for positive integers M, N. Then*

$$|E \cap T^{-n}F| = |E|\,|F|\,(1 + O)(q^{\sqrt{n}})) \qquad q < 1\,.$$

Proof. This follows at once from Lemma 2 since for example

$$E = \bigcup_{i=M}^{\infty} \{\alpha \mid a_1(\alpha) = i\}\,.$$

3C. In this section we use the notation introduced in section 1C. In the next lines $\lambda(E)$ stands for the Lebesgue-measure of E.

LEMMA 6. (*Gel'fond* [7]) *For $t \leqq 1$ let $E = [0, t)$ be an interval and let F be a measurable set. Then*

$$\lambda(E \cap T^{-n}F) = \lambda(E)\,|F| + |F|\,O(\rho^{-n}) \qquad 1 < \rho < \theta\,.$$

Proof. Apart from the factor $|F|$ before the 0 symbol this is just formula (12) in Gel'fond's paper. But inspection of its proof shows that we may pull out the factor $|F|$ of the 0-symbol and this symbol still has the required properties.

LEMMA 7. *Let E be an interval and let F be a measurable set. Then*

$$|E \cap T^{-n}F| = |E||F| + |F|O(\rho^{-n}) \qquad \rho > 1\,.$$

Proof. It is enough to show the lemma in case that $E = [0, t)$. Let f be the characteristic function of F. Using Lemma 6 we obtain

$$\begin{aligned}
|E \cap T^{-n}F| &= \int_0^1 \varphi_t(\alpha) f(x_{n+1}(\alpha))\sigma(\alpha)d\alpha \\
&= \frac{1}{\tau}\sum_{k=1}^{\infty}\frac{1}{\theta^{k-1}}\int_0^1 \varphi_t(\alpha) f(x_{n+1}(\alpha))\varphi_{t_k}(\alpha)d\alpha \\
&= \frac{1}{\tau}\sum_{k=1}^{\infty}\frac{1}{\theta^{k-1}}\int_0^{\min(t,t_k)} f(x_{n+1}(\alpha))d\alpha \\
&= \frac{1}{\tau}|F|\sum_{k=1}^{\infty}\frac{\min(t,t_k)}{\theta^{k-1}} + |F|O(\rho^{-n}) \\
&= |E||F| + |F|O(\rho^{-n})\,.
\end{aligned}$$

4. Proof of Theorem 1 and 2.

4.1 *Proof of Theorem* 1. We have to show that for every pair of measurable sets $A_1, A_2 \subset [0, 1]$

$$\lim_{n \to \infty} |A_1 \cap T^{-n}A_2| = |A_1||A_2|\,.$$

We approximate A_i $(i = 1, 2)$ in measure by a finite union E_i $(i = 1, 2)$ of disjoint intervals arbitrarily closely: Given $\varepsilon > 0$ we can find E_i such that

$$|A_i \Delta E_i| < \varepsilon \qquad (i = 1, 2)\,.$$

By Lemmas 1, 4, 7 we easily get for all sufficiently large n

$$||E_i \cap T^{-n}E_2| - |E_1||E_2|| < \varepsilon\,.$$

Using that T preserves μ we have

$$\begin{aligned}
||A_1 \cap T^{-n}A_2| - |A_1||A_2|| &\leqq ||A_1 \cap T^{-n}A_2| - |E_1 \cap T^{-n}A_2|| \\
&+ ||E_1 \cap T^{-n}A_2| - |E_1 \cap T^{-n}E_2|| \\
&+ ||E_1 \cap T^{-n}E_2| - |E_1||E_2|| + ||E_1||E_2|| - |A_1||A_2|| < 5\varepsilon
\end{aligned}$$

for all sufficiently large n.

Rohlin's theorem that T is mixing of every degree follows in the same way from:

LEMMA 8ABC. *Let $n_1, \cdots, n_r$ be positive integers and let $I_0, I_1, \cdots, I_r$ be intervals all contained in the unit interval. Then*

$$|I_0 \cap T^{-n_1}I_1 \cap \cdots \cap T^{-(n_1+\cdots+n_r)}I_r| = \prod_{i=0}^{r} |I_i| + |I_r| O(q^{\min \sqrt{n_j}}) \qquad q < 1$$

where the constant implied by O only depends on r.

For a proof we only have to apply Lemmas 1, 4, 7 several times.

4.2. In order to prove Theorem 2 we put $E_n = T^{-n}I_n$. Using Lemmas 1, 4, 7 and the fact that T preserves μ we obtain for $n > m$

$$\begin{aligned} |E_n \cap E_m| &= |I_m \cap T^{-(n-m)}I_n| \\ &\leqq |I_m||I_n| + |I_n| \cdot O(q^{\sqrt{n-m}}) \\ &= |E_m||E_n| + |E_n| \cdot O(q^{\sqrt{n-m}}) \qquad q < 1 . \end{aligned}$$

Observing that all the measures involved are equivalent to the Lebesgue-measure we get Theorem 2 as an application of Theorem 3.

5. Some metrical theorems on continued fractions. In this section some metrical theorems on continued fractions are proved. I use the same notation as in sections 1B and 3B. The main result (Theorem 4) is a refinement of a theorem of Khintchine [11]. LeVeque [14] has outlined a proof of a weaker form of Theorem 4. Several applications of Theorem 4 are given. Finally a well known theorem (e.g. see [12], p. 67) of Bernstein is sharpened.

Again it will be illustrated that it is more natural to use the measure μ invariant under T rather than the Lebesgue-measure. In the theorems though "almost all" always means "all except a set of Lebesgue-measure 0." But μ and the Lebesgue-measure have the same null-sets and hence-considered from this point of view it makes no difference which of them we use.

5.1. The general theorem.

THEOREM 4. *Let $f(u_1, \cdots, u_k)$ be a nonnegative function defined for all k-tuples $(u_1, \cdots, u_k)$ with positive entries and satisfying*

$$\int_0^1 (f(a_1(x), \cdots, a_k(x))^2 dx < \infty . \tag{8}$$

Then for integers $n \geqq 1$ and fixed $k \geqq 1$

$$(9)\qquad \begin{aligned} &\frac{1}{n}\sum_{i=0}^{n-1} f(a_1(T^i x), \cdots, a_k(T^i x)) \\ &= \frac{1}{\log 2}\int_0^1 f(a_1(x), \cdots, a_k(x))\frac{dx}{1+x} + O(n^{-1/2}\log^{3/2+\varepsilon} n) \end{aligned}$$

almost everywhere.

If k is not fixed but if instead n and k are linked by

$$(10)\qquad 2^k \leqq n < 2^{k+1}$$

then (9) *holds with $n^{-1/2}\log^{2+\varepsilon} n$ instead of $n^{-1/2}\log^{3/2+\varepsilon} n$ in the error term.*

Proof. Since T preserves μ we have for integer $i \geqq 0$

$$(11)\qquad \int_0^1 f(a_1(T^i x), \cdots, a_k(T^i x))d\mu(x) = \alpha_i = \alpha < \infty$$

by (8). Further for integers $0 \leqq i < j$ and $m \geqq 0$

$$\begin{aligned} I(i, j) = \int_0^1 & f(a_1(T^{i+m}x), \cdots, a_k(T^{i+m}x))\cdot f(a_1(T^{j+m}x), \cdots, \\ & a_k(T^{j+m}x))d\mu(x) = \sum_1^\infty f(r_1, \cdots, r_k)\cdot f(r_1', \cdots, r_k') \\ & \cdot|E \cap T^{-(j-i)}F| \end{aligned}$$

where the summation is extended over all the r's and

$$E = \{\alpha\colon\ a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k\}$$

and

$$F = \{\alpha\colon\ a_1(\alpha) = r_1', \cdots, a_k(\alpha) = r_k'\}\ .$$

By Lemma 2 we obtain for $j > i + k$

$$\begin{aligned} I(i, j) &= \sum_1^\infty f(r_1, \cdots, r_k)\,|E|\cdot \sum_1^\infty f(r_1', \cdots, r_k')\,|F|\cdot(1 + O(q^{\sqrt{j-i-k}})) \\ &= \alpha^2(1 + O(q^{\sqrt{j-i-k}}))\ . \end{aligned}$$

By (8) we clearly have for $j \leqq i + k$

$$I(i, j) \leqq \alpha^2\ .$$

Hence we obtain by (11)

$$\int_0^1\Big(\sum_{i=0}^{n-1} f(a_1(T^{i+m}x), \cdots, a_k(T^{i+m}x)) - \alpha n\Big)^2 d\mu(x)$$

$$(12) \quad = \sum_{i,j=0}^{n-1}\int_0^1 f(a_1(T^{i+m}x), \cdots, a_k(T^{i+m}x))\cdot f(a_1(T^{j+m}x), \cdots,$$

$$a_k(T^{j+m}x)d\mu(x) - \alpha^2 n^2 \leqq \alpha^2\cdot O\Big(\sum_{\substack{i,j=0\\ j>k+i}}^{n-1} q^{\sqrt{j-i-k}}\Big) + \alpha^2 kn = O(n)$$

if k is fixed. If n and k are linked by (10) then we get in (12) only the estimate $O(n\log n)$. In both cases Theorem 6 (p. 649) of Gál and Koksma [6] gives the result.

De Vroedt [25] has obtained earlier a result of this type also for a narrow class of functions only. His theorem and mine have a non-empty overlap: To see this put in Theorem 4 $k = 1$ and $f(u_1) = f(u) = \log u$. We get

COROLLARY 1 (*de Vroedt*). *Almost everywhere*

$$(a_1(x)\cdots a_n(x))^{1/n} = \prod_{n=1}^{\infty}\Big(1 + \frac{1}{n^2+2n}\Big)^{\log n/\log 2} + O(n^{-1/2}\log^{3/2+\varepsilon} n)\ .$$

This is a refinement of a well known theorem of Khintchine [10]. Applying Theorem 1 for the characteristic function of the set $\{x\colon a_1(x) = p\}$ we get a refinement of a theorem of Lévy [19] on the frequency of the digit p.

COROLLARY 2 (*de Vroedt*). *Denote by $h(n, p, x)$ the number of positive integers $k \leqq n$ such that $a_k(x) = p$. Then for almost all x*

$$h(n, p, x) = \frac{1}{\log 2}\log\Big(1 + \frac{1}{p(p+2)}\Big) + O(n^{-1/2}\log^{3/2+\varepsilon} n)\ .$$

Doeblin [5] using quite different methods sketched a proof of the law of the iterated logarithm in both cases. Recently, Stackelberg [23] has announced a proof different from that of Doeblin.

As another application of Theorem 4 I shall prove:

THEOREM 5. *Denote by $q_n(x)$ the denominator of the nth convergent to x. Then*

$$\sqrt[n]{q_n(x)} = \exp(\pi^2/12\log 2) + O(n^{-1/2}\log^{2+\varepsilon} n)$$

almost everywhere.

This improves slightly LeVeque's [14] refinement of a well known theorem of Khintchine [11] and Lévy [17]. Furthermore the proof will not depend on Lévy's result. For this reason we need some lemmas.

5.2. *Some lemmas.* The first one can be proved by induction or by taking the transposed matrices.

LEMMA 9. *If*

$$\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix}$$

then

$$\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} q_{k-1} & q_k \\ p_{k-1} & p_k \end{pmatrix} .$$

If we consider $a_1, \cdots, a_k$ as partial quotients of a continued fraction then p_k/q_k is just the kth convergent. So the lemma describes the relation between the two fractions $[a_1, \cdots, a_k]$ and $[a_k, \cdots, a_1]$. This is used in

LEMMA 10. *Let*

$$E = \{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k\}$$

and

$$F = \{\alpha \mid a_1(\alpha) = r_k, \cdots, a_k(\alpha) = r_1\}$$

then

$$| E | = | F | .$$

Proof. Put $[r_1, \cdots, r_k] = p_k/q_k$ and $[r_k, \cdots, r_1] = p'_k/q'_k$ then

$$\log 2 \cdot | E | = (-1)^k \log\left(1 + \frac{(-1)^k}{(p_k + q_k)(q_k + q_{k-1})}\right)$$

and similarly for $| F |$ with primes. But Lemma 9 yields $q'_k = q_k$, $p'_k = q_{k-1}$, and $q'_{k-1} = p_k$. Hence the result.

LEMMA 11. *Define for* $k \geqq 1$

$$\begin{aligned} f_k(x) &= a_k(x) + [a_{k-1}(x), \cdots, a_1(x)] \quad \textit{if} \\ x &= [a_1(x), \cdots, a_k(x), \cdots]\textit{-irrational} \\ &= 0 \qquad\qquad\qquad \textit{if } x \textit{ is rational} \end{aligned}$$

then

$$\lim_{k \to \infty} \int_0^1 \log f_k(x) d\mu(x) = \frac{\pi^2}{12 \log 2} .$$

Proof. Using Lemma 10 we obtain

$$\begin{aligned}\lambda_k &= \int_0^1 \log f_k(x)d\mu(x) \\ &= \sum_{r_1,\cdots,r_k \geqq 1} \log(r_k + [r_{k-1}, \cdots, r_1])\mu\{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k\} \\ &= \sum_{r_1,\cdots,r_k \geqq 1} \log(r_1 + [r_2, \cdots, r_k])\mu\{\alpha \mid a_1(\alpha) = r_1, \cdots, a_k(\alpha) = r_k\} \\ &= \int_0^1 \log(a_1(x) + [a_2(x), \cdots, a_k(x)])d\mu(x) .\end{aligned}$$

But $\log(a_1(x) + [a_2(x), \cdots, a_k(x)]) \leqq \log(a_1(x) + 1)$ which is integrable. Hence

$$\begin{aligned}\lim_{k\to\infty} \int_0^1 \log f_k(x)d\mu(x) &= \int_0^1 \lim_{k\to\infty} \log(a_1(x) + [a_2(x), \cdots, a_k(x)])d\mu(x) \\ &= \frac{1}{\log 2}\int_0^1 \log\frac{1}{x}\cdot\frac{dx}{1+x} = \frac{\pi^2}{12\log 2} .\end{aligned}$$

5.3. Now we can finish the proof of Theorem 5. Essentially, we use the argument of Khintchine [11]. Let $k \geqq 1$ be an integer. Define for $n \geqq k$

$$\begin{aligned}f_n^{(k)}(x) &= a_n(x) + [a_{n-1}(x), \cdots, a_{n-k+1}(x)] \text{ if} \\ x &= [a_1(x), \cdots, a_n(x), \cdots] \text{ irrational} \\ &= 0 \qquad\qquad \text{if } x \text{ is rational.}\end{aligned}$$

Then $f_n^{(k)}(Tx) = f_{n+1}^{(k)}(x)$. Hence for $n \geqq k$

$$f_n^{(k)}(x) = f_k^{(k)}(T^{n-k}x) = f_k(T^{n-k}x)$$

in the notation of Lemma 11.

We observe

$$\frac{q_n(x)}{q_{n-1}(x)} = f_n(x) . \tag{13}$$

For $n \geqq k$

$$|\log f_n(x) - \log f_n^{(k)}(x)| \leqq |f_n(x) - f_n^{(k)}(x)| \leqq 2^{-(k-3)} \tag{14}$$

by Lemma 4, p. 7 in Cassel's book [2]. Using again this lemma and the first part of the proof of Lemma 11 we get

$$\begin{aligned}\left|\lambda_k - \frac{\pi^2}{12\log 2}\right| &\leqq \int_0^1 |\log(a_1(x) + [a_2(x), \cdots, a_k(x)] - \log(a_1(x) \\ &\qquad + [a_2(x), \cdots])| d\mu(x) \\ &\leqq \int_0^1 |[a_2(x), \cdots, a_k(x)] - [a_2(x), \cdots]| d\mu(x) \\ &\leqq 2^{-(k-4)} .\end{aligned} \tag{15}$$

From now on n and k are linked by (10). Since

$$\int_0^1\left(n^{1/2}\sum_{i=1}^{k-1}\log f_i(x)\right)^2 dx = O(n\log^2 n)$$

we obtain by the mentioned theorem of Gál and Koksma

$$\frac{1}{n}\sum_{i=1}^{k-1}\log f_i(x) = O(n^{-1}\log^{5/2+\varepsilon}) \tag{16}$$

almost everywhere. Hence by Theorem 4, (10), (14), (15), (16)

$$\begin{aligned}&\left|\frac{1}{n+k-1}\sum_{i=1}^{n+k-1}\log f_i(x) - \frac{\pi^2}{12\log 2}\right| \\ &\quad < \frac{k-1}{n}\frac{\pi^2}{12\log 2} + \left|\frac{1}{n}\sum_{i=1}^{k-1}\log f_i(x)\right| + \frac{1}{n}\sum_{i=k}^{n+k-1}\left|\log f_i(x)\right. \\ &\quad \left.-\log f_i^{(k)}(x)\right| + \left|\frac{1}{n}\sum_{i=k}^{n+k-1}\log f_i^{(k)}(x) - \lambda_k\right| + \left|\lambda_k - \frac{\pi^2}{12\log 2}\right| \\ &\quad = O(n^{-1/2}\log^{2+\varepsilon} n)\end{aligned}$$

almost everywhere. Since obviously every positive integer $N > 2$ can be written as $N = n + k$ (n, k subject to (10)) (13) yields the result.

It might be interesting to remark that Khintchine-Lévy's theorem, namely that $\sqrt[n]{q_n(x)} \to \exp(\pi^2/12\log 2)$ almost everywhere, is another interesting application of the individual ergodic theorem. The proof is apart from a few simplifications the same as that of Theorem 2.

COROLLARY. *If $p_n(x)$ denotes the numerator of the nth convergent to x then almost everywhere*

$$\sqrt[n]{p_n(x)} = \sqrt[n]{x}\cdot\exp(\pi^2/12\log 2) + O(n^{-1/2}\log^{2+\varepsilon} n)$$

This follows from the inequality

$$\left|\sqrt[n]{p_n(x)} - \sqrt[n]{xq_n(x)}\right| \leqq \frac{2}{n}$$

for $0 < x \leqq 1$.

5.4. I shall prove now a refinement of a well known theorem of Bernstein (e.g. see [12] p. 67). However, Doeblin [5] states with some misprints that even the law of the iterated logarithm holds in case that $\varphi(n) \to \infty$.

THEOREM 6. *Let $(\varphi(n))$ be any sequence of positive integers and denote by $A(N, x)$ the number of positive integers $n \leqq N$ such that $a_n(x) \geqq \varphi(n)$. Put*

$$\phi(N) = \frac{1}{\log 2} \sum_{n \leq N} \log\left(1 + \frac{1}{\varphi(n)}\right).$$

Then for almost all x

$$A(N, x) = \phi(N) + O(\phi^{1/2}(N) \log^{3/2+\varepsilon} \phi(N)).$$

Proof. Put

$$E_n = \{\alpha \mid a_n(\alpha) \geqq \varphi(n)\} \quad \text{and} \quad U_n = T^{-(n-1)} E_n .$$

Then

$$U_n = \{\alpha \mid a_1(\alpha) \geqq \varphi(n)\} .$$

Using Lemma 5 and the fact that T preserves μ we get for $n > m$

$$\begin{aligned} | E_n \cap E_m | &= | U_m \cap T^{-(n-m)} U_n | \leqq | U_m | | U_n | (1 + O(q^{\sqrt{n-m}})) \\ &= | E_m | | E_n | (1 + O(q^{\sqrt{n-m}})) . \end{aligned}$$

But

$$| E_n | = | U_n | = \frac{1}{\log 2} \int_0^{1/\varphi(n)} \frac{dx}{1+x} = \frac{1}{\log 2} \log\left(1 + \frac{1}{\varphi(n)}\right).$$

Since μ is equivalent to the Lebesgue-measure Theorem 6 follows from Theorem 3.

References

1. J. W. S. Cassels, *Some metrical theorems in Diophantine approximation I*, Proc. Camb. Phil. Soc. **46** (1950), 209-18.
2. ———, *An introduction to Diophantine approximation*, Cambridge, 1957.
3. Johann Cigler, *Ziffernverteilung in θ-adischen Brüchen*, Math. Zeitschr. **75** (1961), 8-13.
4. J. Cigler und G. Helmberg, *Neuere Entwicklungen der Theorie der Gleichverteilung*, Jahresbericht DMV **64** (1961), 1-50.
5. W. Doeblin, *Remarques sur la theorie metrique des fractions continues*, Compositio Math. **7** (1940), 353-371.
6. I. S. Gál and J. F. Koksma, *Sur l'ordre de grandeur des functions sommables*, Proc. Akad. Amsterdam **53** (1950), 638-653.
7. A. O. Gel'fond, *On a general property of number systems*, Izvest. Akad. Nauk **23** (1959), 809-814. (Russian)
8. Paul R. Halmos, *Lectures on ergodic theory*, Math. Soc. Japan, 1956.
9. S. Hartman and C. Ryll-Nardzewski, *Zur Theorie der lokal-kompakten abelschen Gruppen*, Colloquium Math. **4** (1957), 157-188.
10. A. Khintchine, *Metrische Kettenbruchprobleme*, Compositio Math. **1** (1935), 359-382.
11. ———, *Zur metrischen Kettenbruchtheorie*, Compositio Math. **3** (1936), 276-285.
12. ———, *Kettenbrüche*, Teubner, Leipzig, 1956 (there is also an English translation).

13. K. Knopp, *Mengentheoretische Behandlung einiger Probleme...*, Math. Annalen **95** (1926), 409-426.
14. W. J. LeVeque, *On the frequency of small fractional parts in certain real sequences*, Transactions AMS **87** (1958), 237-260.
15. ———, *On the frequency of small fractional parts in certain real sequences III*, Journal reine angew. Math. **202** (1959), 215-220.
16. P. Lévy, *Sur les lois de probabilité dont dependent quotient*, Bull. Soc. Math. France **57** (1929), 178-194.
17. ———, *Sur le developpement en fraction continue d'un nombre choisi au hasard*, Compositio Math. **3** (1936), 286-303.
18. W. Parry, *On the β-expansion of real numbers*, Acta Math. Acad. Sci. Hung. **11** (1960), 401-416.
19. A. Rényi, *Representations for real numbers and their ergodic properties*, Acta Math. Acad. Sci. Hung. **8** (1957), 477-493.
20. V. A. Rohlin, *Exact endomorphisms of a Lebesgue space*, Izvest. Akad. Nauk **25** (1961), 499-530, AMS Translations ser 2. **39** (1964), 1-36.
21. C. Ryll-Nardzewski, *On the ergodic theorems II*, Studia Math. **12** (1951), 74-79.
22. Wolfgang M. Schmidt, *Metrical theorems on fractional parts of sequences*, Transactions AMS **110** (1964), 493-518.
23. O. P. Stackelberg, *On the law of the iterated logarithm for continued fractions*, Notices AMS **11** (1964), p. 772 No. 64T-508.
24. P. Szüsz, *Über einen Kusminschen Satz*, Acta Math. Acad. Sci. Hung. **12** (1961).
25. C. de Vroedt, *Measure-theoretical investigations concerning continued fractions*, Proc. Akad. Amsterdam, Ser A. **65** (1962), 583-591.
26. Thomas P. Walker, *Two metric theorems on Diophantine approximation* (Ph. D. thesis, Univ. of Michigan 1963).
27. H. Weyl, *Uber die Gleichverteilung von Zahlen mod Eins*, Math. Ann. **77** (1916), 313-352.

Received March 15, 1965.

UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS