REPRESENTATIONS FOR REAL NUMBERS AND THEIR ERGODIC PROPERTIES

By

A. RÉNYI (Budapest), member of the Academy

Introduction

We shall consider representations of a real number x by infinite iteration of a positive function $y=f(x)$ in the form of the "f-expansion"

$$x=\varepsilon_0+f(\varepsilon_1+f(\varepsilon_2+f(\varepsilon_3+\cdots)\ldots) \tag{1}$$

where the "digits" $\varepsilon_n=\varepsilon_n(x)$ $(n=0,1,\ldots)$ and the "remainders"

$$r_n(x)=f(\varepsilon_{n+1}+f(\varepsilon_{n+2}+f(\varepsilon_{n+3}+\cdots)\ldots) \qquad (n=0,1,\ldots) \tag{2}$$

are defined by the following recursive relations:

$$\begin{aligned} \varepsilon_0(x)&=[x], & r_0(x)&=(x), \\ \varepsilon_{n+1}(x)&=[\varphi(r_n(x))], & r_{n+1}(x)&=(\varphi(r_n(x))) \qquad (n=0,1,\ldots) \end{aligned} \tag{3}$$

where $[z]$ denotes the integral part and (z) the fractional part of the real number z and $x=\varphi(y)$ is the inverse function of $y=f(x)$. In § 1 we shall investigate what conditions imposed on the function $f(x)$ are sufficient to ensure that every real number x should have a representation in the form of the f-expansion (1).[1]

The representation (1) reduces for $f(x)=\dfrac{x}{q}$ $(q=2,3,\ldots)$ to the q-adic expansion $x=\sum\limits_{n=0}^{\infty}\dfrac{\varepsilon_n}{q^n}$ and for $f(x)=\dfrac{1}{x}$ to the continued fraction representation of x. The case when $f(x)$ is a general decreasing function has been considered previously by B. H. BISSINGER [1]. Our treatment is still more general than his, since we do not suppose the unnecessary condition that $f(x)$ is positive for any $x\geqq 1$ (i. e. that $\varphi(0)=+\infty$). The case when $f(x)$ is a general increasing function has been considered previously by C. I. EVERETT [2]. He supposed the unnecessary condition that $\varphi(1)$ is an integer. We shall not need this restriction. The principal aim of the present paper, however, is not this generalization of the conditions ensuring the validity of

[1] If for some n we have $r_n(x)=0$, then $r_{n+k}(x)$ and $\varepsilon_{n+k}(x)$ are not defined for $k=1,2,\ldots$, and x has the finite representation $x=\varepsilon_0+f(\varepsilon_1+f(\varepsilon_2+\cdots+f(\varepsilon_n)\ldots)$.

the representation (1), but to prove some theorems on the ergodic properties of the digits $\varepsilon_n(x)$ and the remainders $r_n(x)$ which contain as special cases the well-known theorems on q-adic expansions and on continued fractions, respectively (see [5]—[15]). To obtain such theorems we have to impose some additional restrictions on $f(x)$.

The mentioned ergodic properties of an "f-expansion (1) with independent digits" will be investigated in § 2. In § 3 we consider some examples in which our general theorem is applicable; q-adic expansions, continued fractions and the algorithm of W. BOLYAI (see [2], [3], [4]). In § 4 we consider a class of f-expansions, called β-adic expansions ($\beta>1$ not an integer), to which our theorem can not be applied, but another method leads to the same conclusion.[2]

§ 1. Representation theorems

A) We consider first the case when $f(x)$ is a decreasing function. We suppose

A1) $f(1)=1$.

We suppose further

A2) *$f(t)$ is positive, continuous and strictly decreasing for $1\leqq t\leqq T$ and $f(t)=0$ for $t\geqq T$ where $2<T\leqq+\infty$* (in case $T=+\infty$, this means that $\lim\limits_{t\to+\infty}f(t)=0$).

We distinguish three subcases:

A2$_1$) $T=+\infty$; A2$_2$) $2<T<+\infty$ and T is an integer; A2$_3$) $2<T<+\infty$ and T is not an integer.

Let us mention that B. H. BISSINGER considered only the case A2$_1$).

Following BISSINGER, we suppose further that the following condition is also satisfied:[3]

A3) *$|f(t_2)-f(t_1)|\leqq|t_2-t_1|$ for $1\leqq t_1<t_2$ and there is a constant λ such that $0<\lambda<1$ and*

$$|f(t_2)-f(t_1)|\leqq\lambda|t_2-t_1| \quad \textit{if} \quad 1+f(2)<t_1<t_2.$$

We shall prove that conditions A1), A2) and A3) imply that the representation (1) is valid for any real x. (Clearly, it suffices to prove this for $0<x<1$. In what follows we shall always suppose therefore that $0<x<1$.)

[2] The assertions of Theorem 1 have been proved under somewhat more restrictive suppositions and Theorem 2 has been announced without proof in a previous paper (in Hungarian language) [16] of the author.

[3] This condition could be replaced by a less restrictive one as will be pointed out below.

Before proving this, we introduce some notations. Let us define

$$f_1(z_1)=f(z_1),\tag{1.1}$$
$$f_n(z_1, z_2, \dots, z_n)=f_{n-1}(z_1, z_2, \dots, z_{n-2}, z_{n-1}+f(z_n))$$

for $n=2, 3, \dots$. Let us put further

$$C_n(x)=f_n(\varepsilon_1(x), \varepsilon_2(x), \dots, \varepsilon_n(x))\tag{1.2}$$

where the digits $\varepsilon_1(x), \varepsilon_2(x), \dots$ are defined by the recursion (3). We shall call $C_n(x)$ the n-th convergent of x. The validity of (1) means that either we have $r_n(x)=0$ for some n, in which case $x=f(\varepsilon_1+f(\varepsilon_2+\cdots+f(\varepsilon_n)\dots)$, or

$$\lim_{n\to\infty} C_n(x)=x.\tag{1.3}$$

We have to consider only the latter case when $r_n(x)\neq 0$ $(n=1, 2, \dots)$. We have clearly (for $0<x<1$)

$$x=f_n(\varepsilon_1(x), \varepsilon_2(x), \dots, \varepsilon_{n-1}(x), \varepsilon_n(x)+r_n(x)).\tag{1.4}$$

Thus it follows

$$x-C_n(x)=f_n(\varepsilon_1(x), \dots, \varepsilon_n(x)+r_n(x))-f_n(\varepsilon_1(x), \dots, \varepsilon_n(x)),\tag{1.5}$$

and therefore putting

$$u_k=\varepsilon_{k+1}(x)+f_{n-k-1}(\varepsilon_{k+2}(x), \dots, \varepsilon_n(x)+r_n(x))\tag{1.6}$$

and

$$v_k=\varepsilon_{k+1}(x)+f_{n-k-1}(\varepsilon_{k+2}(x), \dots, \varepsilon_n(x))$$

for $k=0, 1, \dots, n-1$, we have

$$x-C_n(x)=r_n(x)\prod_{k=0}^{n-1}\left(\frac{f(u_k)-f(v_k)}{u_k-v_k}\right).\tag{1.7}$$

Now each factor on the right of (1. 7) has an absolute value not exceeding 1. We shall prove that from any two numbers

$$\left|\frac{f(u_k)-f(v_k)}{u_k-v_k}\right|, \quad \left|\frac{f(u_{k+1})-f(v_{k+1})}{u_{k+1}-v_{k+1}}\right| \qquad (k=0, 1, \dots, n-3)$$

at least one does not exceed λ. As a matter of fact, we have

$$u_k=\varepsilon_{k+1}+f(\varepsilon_{k+2}+f(u_{k+2})),$$
$$u_{k+1}=\qquad\varepsilon_{k+2}+f(u_{k+2}),$$

and similarly

$$v_k=\varepsilon_{k+1}+f(\varepsilon_{k+2}+f(v_{k+2})),$$
$$v_{k+1}=\qquad\varepsilon_{k+2}+f(v_{k+2}).$$

Three cases are possible. If $\varepsilon_{k+1}\geqq 2$, then $u_k\geqq 2>1+f(2)$ and $v_k\geqq 2>1+f(2)$ and thus by condition A3) $\left|\frac{f(u_k)-f(v_k)}{u_k-v_k}\right|\leqq\lambda$. If $\varepsilon_{k+1}=1$

and $\varepsilon_{k+2} \geqq 2$, then similarly we obtain $\left|\frac{f(u_{k+1})-f(v_{k+1})}{u_{k+1}-v_{k+1}}\right| \leqq \lambda$. Finally, if $\varepsilon_{k+1}=\varepsilon_{k+2}=1$, then

$$u_k = 1+f(1+f(u_{k+2})) \geqq 1+f(2)$$

and

$$v_k = 1+f(1+f(v_{k+2})) \geqq 1+f(2).$$

Thus our assertion is proved. It follows from (1. 7) that

$$|x-C_n(x)| \leqq \lambda^{\left[\frac{n}{2}\right]-1} \tag{1.8}$$

and (1. 8) clearly implies (1. 3).

The above proof is essentially that of BISSINGER. By the same method it can be shown that it suffices to suppose that $\left|\frac{f(t_2)-f(t_1)}{t_2-t_1}\right| \leqq \lambda < 1$ holds for $t_2 > t_1 \geqq 1+f_{2r-1}(1, 1, \ldots, 1, 2)$ for some r $(r=1, 2, 3, \ldots)$, because in this case from $2r$ consecutive numbers $\left|\frac{f(u_k)-f(v_k)}{u_k-v_k}\right|$ at least one does not exceed λ.

B) Now we consider the case when $f(x)$ is increasing. We suppose first of all

B1) $f(0)=0$.

We suppose further that the following condition is satisfied:

B2) *$f(t)$ is continuous and strictly increasing for $0 \leqq t \leqq T$ and $f(t)=1$ if $t \geqq T$ where $1 < T \leqq +\infty$.* (In case $T=+\infty$, this means $\lim_{t\to+\infty} f(t)=1$.)

We distinguish again three subcases: $B2_1$), $B2_2$), $B2_3$) accordingly as $T=+\infty$, $T<+\infty$ and T is an integer, $T<+\infty$ and T is not an integer, respectively. EVERETT considered only the case $B2_2$).

We need here also a condition on the slope $\frac{f(t_2)-f(t_1)}{t_2-t_1}$. For example, the following condition considered already by EVERETT is *sufficient:*[4]

B3) $\frac{f(t_2)-f(t_1)}{t_2-t_1} < 1$ *for* $0 \leqq t_1 < t_2$.

If B1), B2) and B3) are satisfied, then the f-expansion (1) is valid for any real x. (We may suppose again $0 < x < 1$.) Following EVERETT, this can be shown as follows:

Clearly, the sequence $C_n(x)$ $(n=1, 2, \ldots)$ defined by (1. 2) is non-decreasing and the sequence $D_n(x)$, where $D_n(x)$ is defined as the least value of $C_n(x')$ which is greater than $C_n(x)$ (or 1 if such an x' does not exist), is

[4] This condition can be replaced by a weaker one, cf. [2].

non-increasing and

$$C_n(x) \leqq x < D_n(x). \tag{1.9}$$

Thus

$$\underline{x} = \lim_{n\to\infty} C_n(x) \tag{1.10}$$

and

$$\bar{x} = \lim_{n\to\infty} D_n(x) \tag{1.11}$$

always exist and $\underline{x} \leqq x \leqq \bar{x}$. We have to prove that $\underline{x} = \bar{x} = x$ for any x $(0 < x < 1)$. If this would not hold for all x in $(0, 1)$, then there would exist a finite or denumerable sequence of non-overlapping „gaps" $(\underline{x}, \bar{x})$ in the unit interval, and thus there would exist an x for which $\bar{x} - \underline{x}$ is maximal. For this value of x we would have by condition B3) putting $r_1(x) = y$

$$\bar{x} - \underline{x} = \left(\frac{f(\varepsilon_1(x) + \bar{y}) - f(\varepsilon_1(x) + \underline{y})}{\bar{y} - \underline{y}}\right)(\bar{y} - \underline{y}) < \bar{y} - \underline{y} \tag{1.12}$$

which contradicts our assumption that $\bar{x} - \underline{x}$ is maximal. Thus we have $\bar{x} = \underline{x} = x$ for all x.[5]

The admissible values for $\varepsilon_n(x)$ $(n = 1, 2, \ldots)$ are $1, 2, \ldots$ in case $A2_1$), $1, 2, \ldots, T-1$ in case $A2_2$) and $1, 2, \ldots, [T]$ in case $A2_3$), similarly $0, 1, \ldots$ in case $B2_1$), further $0, 1, \ldots, T-1$ in case $B2_2$) and $0, 1, \ldots, [T]$ in case $B2_3$). Let us call a finite sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ a *canonical sequence* with respect to a given function $f(x)$, which satisfies either conditions A1), A2) and A3) or conditions B1), B2) and B3), if there exists a number x $(0 \leqq x < 1)$ such that $\varepsilon_k(x) = \varepsilon_k$ $(k = 1, 2, \ldots, n)$. There is an essential difference for decreasing $f(x)$ between the case when T is an integer or $T = +\infty$ (cases $A2_1$) and $A2_2$)) and, on the other hand, the case with a finite non-integral T (case $A2_3$)). This difference consists in that in the case of an integer T or $T = +\infty$ all finite sequences $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ consisting of admissible digits, i. e. all sequences of positive integers $< T$ are canonical, while in the case when T is not an integer this is not true. The same difference

[5] J. Czipszer remarked that the above method of the proof, due to Everett, may be combined with the method of Bissinger in the case when $f(x)$ is decreasing, and in this way it can be shown that condition A3) can be replaced by the following weaker condition:

A3*) $|f(t_2) - f(t_1)| \leqq |t_2 - t_1|$ for $1 \leqq t_1 < t_2$

and

$$|f(t_2) - f(t_1)| < |t_2 - t_1| \quad \text{if} \quad \tau - \varepsilon < t_1 < t_2$$

where τ is the solution of the equation $1 + f(\tau) = \tau$ and $0 < \varepsilon < \tau$ is arbitrary. The only essential difference in the proof consists in that $\underline{x}$ and $\bar{x}$ are defined as $\underline{x} = \lim_{n\to\infty} C_{2n}(x)$ and $\bar{x} = \lim_{n\to\infty} C_{2n+1}(x)$, respectively.

exists for increasing $f(x)$ between the case when T is an integer or $T=+\infty$ (cases $B2_1$) and $B2_2$)) and the case when T is finite but not an integer (case $B2_3$)). While in cases $B2_1$) and $B2_2$) every finite sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ of non-negative integers $<T$ is canonical,[6] this is not true in case $B2_3$). By other words, in both cases A) and B) if T is an integer or $T=+\infty$, the values of the digits ε_n of a canonical sequence can be chosen independently, but if T is finite and not an integer, there exists some dependence between the members of a canonical sequence.

We shall call the f-expansions when one of the conditions $A2_1$), $A2_2$) respectively $B2_1$), $B2_2$) is satisfied *f-expansions with independent digits*, and the f-expansions when $A2_3$) respectively $B2_3$) are satisfied *f-expansions with dependent digits*. It should be noted that independence is not meant here in the sense of probability theory, but only in a weaker sense. As a matter of fact, in some cases, (e. g., in the case of the q-adic expansions) the digits $\varepsilon_n(x)$ considered as random variables (on the interval $(0, 1)$ with the Lebesgue measure) are also statistically independent but for most f-expansions with independent digits this is not true. (For example, the digits of a continued fraction are not statistically independent.)

We shall see that the investigation of ergodic properties of f-expansions is much easier for f-expansions with independent digits than for f-expansions with dependent digits. The first case will be considered in § 2; in § 3 the ergodic theory of some special f-expansions with dependent digits, called the β-expansions, and corresponding to $f(x)=\frac{x}{\beta}$ for $0 \leqq x \leqq \beta$ ($\beta>1$ non-integral) is investigated.

§ 2. Ergodic theory of f-expansions with independent digits

In this § we consider only f-expansions with independent digits. Let $f(x)$ satisfy the corresponding conditions of § 1. Then $f(x)$ is derivable almost everywhere and absolutely continuous. Clearly the same holds for $f_n(\varepsilon_1, \ldots, \varepsilon_n+t)$ as a function of t $(0 \leqq t \leqq 1)$.

Let us put

$$H_n(x, t)=\frac{d}{dt} f_n(\varepsilon_1(x), \ldots, \varepsilon_{n-1}(x), \varepsilon_n(x)+t). \tag{2.1}$$

Then $H_n(x, t)$ is defined for any x, for which $\varepsilon_n(x)$ is defined,[7] and for almost

[6] In these cases clearly $D_n(x)=f_n(\varepsilon_1(x), \ldots, \varepsilon_{n-1}(x), \varepsilon_n(x)+1)$.

[7] I. e., except for those x which have a finite representation in the form (1) of length smaller than n.

all t. We shall suppose that $f(x)$ satisfies also the following condition:

$$\text{C)} \qquad \frac{\sup\limits_{0<t<1} |H_n(x,t)|}{\inf\limits_{0<t<1} |H_n(x,t)|} \leqq C$$

where the constant $C \geqq 1$ does not depend neither on x nor on n.

We prove the following

THEOREM 1. *If $f(x)$ satisfies the conditions* A1), A2$_1$) *or* A2$_2$), A3) *and* C); *or the conditions* B1), B2$_1$) *or* B2$_2$), B3) *and* C), *respectively, then for any function $g(x)$ which is L-integrable in the interval* $(0, 1)$ *we have for almost all x*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(r_k(x)) = M(g), \tag{2.2}$$

where $M(g)$ is a finite constant which can be represented in the form

$$M(g) = \int_0^1 g(x)\, h(x)\, dx \tag{2.3}$$

where $h(x)$ is a measurable function, depending only on $f(x)$ and satisfying the inequality

$$\frac{1}{C} \leqq h(x) \leqq C \tag{2.4}$$

where C is the constant figuring in condition C). *The measure*

$$\nu(E) = \int_E h(x)\, dx \tag{2.5}$$

is invariant with respect to the transformation

$$Tx = (\varphi(x)) \qquad (0 < x < 1) \tag{2.6}$$

where $y = \varphi(x)$ is the inverse function of $x = f(y)$.

PROOF. Let $\mathfrak{S}_n = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ denote a canonical sequence of n terms with respect to $f(x)$. The intervals $(f_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n), f_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n + 1))$ do not overlap and if $\mathfrak{S}_n$ runs over all canonical sequences of n terms, these intervals fill out the interval $(0, 1)$. Therefore we have

$$\sum_{\mathfrak{S}_n} |f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, \varepsilon_n + 1) - f_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, \varepsilon_n)| = 1 \tag{2.7}$$

where the summation is to be extended over all canonical sequences $\mathfrak{S}_n$ of n terms.

Let us consider the mapping $Tx = (\varphi(x))$ of the interval $(0, 1)$ onto itself. For any subset E of $(0, 1)$ we denote by $T^{-1}E$ the set of those real numbers x $(0 < x < 1)$ for which $Tx \in E$. We define further $T^{-n}E$ by the recursion: $T^{-n}E = T^{-1}(T^{-(n-1)}E)$ $(n=2, 3, \ldots)$. Clearly $T^{-n}E$ is measurable if E is any measurable subset of $(0, 1)$. Let $I_{a,b}$ denote the interval (a, b) $(0 < a < b < 1)$ and let $\mu(E)$ denote the Lebesgue measure of the set E. Then we have clearly

$$\mu(T^{-n}I_{a,b}) = \sum_{\mathfrak{E}_n} |f_n(\varepsilon_1, \ldots, \varepsilon_n + b) - f_n(\varepsilon_1, \ldots, \varepsilon_n + a)| \tag{2.8}$$

where the summation is to be extended again over all canonical sequences $\mathfrak{E}_n = (\varepsilon_1, \ldots, \varepsilon_n)$ of n terms. Let us denote by $x(\mathfrak{E}_n)$ a number for which

$$\varepsilon_k(x(\mathfrak{E}_n)) = \varepsilon_k \qquad (k = 1, 2, \ldots, n); \tag{2.9}$$

such a number $x(\mathfrak{E}_n)$ exists for any canonical sequence $\mathfrak{E}_n$ by definition. It follows from (2.7) that

$$\sum_{\mathfrak{E}_n} \inf_{0<t<1} |H_n(x(\mathfrak{E}_n), t)| \leqq 1 \leqq \sum_{\mathfrak{E}_n} \sup_{0<t<1} |H_n(x(\mathfrak{E}_n), t)| \tag{2.10}$$

and from (2.8) that

$$\sum_{\mathfrak{E}_n} \inf_{0<t<1} |H_n(x(\mathfrak{E}_n), t)| \leqq \frac{\mu(T^{-n}I_{a,b})}{(b-a)} \leqq \sum_{\mathfrak{E}_n} \sup_{0<t<1} |H_n(x(\mathfrak{E}_n), t)|. \tag{2.11}$$

Comparing (2.10) and (2.11) we obtain by condition C) that

$$\frac{1}{C}\mu(E) \leqq \mu(T^{-n}E) \leqq C\mu(E), \tag{2.12}$$

provided that E is a subinterval of $(0, 1)$. It follows easily that (2.12) holds for any measurable subset E of the interval $(0, 1)$. Thus we have

$$\frac{1}{C}\mu(E) \leqq \frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}E) \leqq C\mu(E) \qquad (n = 1, 2, \ldots) \tag{2.13}$$

where $C \geqq 1$ does not depend on n. According to the theorem of DUNFORD and MILLER ([17], [18]), it follows from the upper inequality of (2.13) that for any L-integrable function $g(x)$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k x) = g^*(x) \tag{2.14}$$

exists for almost all x. But clearly $T^k x = r_k(x)$ $(k = 0, 1, \ldots)$ and thus we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(r_k(x)) = g^*(x) \tag{2.15}$$

for almost all x.

To prove that $g^*(x)$ is (almost everywhere) equal to a constant depending only on $g(x)$, by a well-known argument it suffices to prove that the transformation T is *ergodic* (indecomposable), or by other words, that if E is a measurable invariant set of positive measure, i. e. $T^{-1}E=E$ and $\mu(E)>0$, then $\mu(E)=1$

According to a theorem of K. KNOPP [19], if $\mu(E)>0$ and there exists a class J of subintervals of $(0,1)$ such that a) every open subinterval of $(0,1)$ is the union of a finite or a denumerably infinite sequence of disjoint intervals belonging to J and b) for any $I\in J$ we have $\mu(EI)\geqq \Delta\mu(I)$ where $\Delta>0$ does not depend on I, then $\mu(E)=1$. We shall show that the class J of all intervals $I_{\mathfrak{E}_n}=[f_n(\varepsilon_1,\ldots,\varepsilon_n), f_n(\varepsilon_1,\ldots,\varepsilon_n+1))=[a_{\mathfrak{E}_n}, b_{\mathfrak{E}_n})$ where $\mathfrak{E}_n=(\varepsilon_1,\ldots,\varepsilon_n)$ is a canonical sequence $(n=1,2,\ldots)$ has the properties required by the mentioned theorem of KNOPP. The class J has according to the representation theorems of § 1 the property a). As regards b), let us put

$$E(x)=\begin{cases}1 & \text{if } x\in E,\\ 0 & \text{if } x\bar{\in} E.\end{cases} \tag{2. 16}$$

Then we have

$$\mu(EI_{\mathfrak{E}_n})=\int_{a_{\mathfrak{E}_n}}^{b_{\mathfrak{E}_n}} E(x)\,dx. \tag{2. 17}$$

Introducing in the integral on the right of (2. 17) the new variable t defined by $x=f_n(\varepsilon_1,\ldots,\varepsilon_n+t)$ (i. e. putting $t=r_n(x)=T^n x$) and taking into account that by virtue of the supposition $T^{-1}E=E$ we have $E(T^{-n}x)=E(x)$, further that $\frac{dx}{dt}=H_n(x(\mathfrak{E}_n),t)$ where $x(\mathfrak{E}_n)$ is a number for which $\varepsilon_k(x(\mathfrak{E}_n))=\varepsilon_k$ $(k=1,2,\ldots,n)$, we obtain

$$\mu(EI_{\mathfrak{E}_n})=\int_0^1 E(t)\,|H_n(x(\mathfrak{E}_n),t)|\,dt. \tag{2. 18}$$

It follows by condition C) that

$$\mu(EI_{\mathfrak{E}_n})\geqq\mu(E)\inf_{0<t<1}|H_n(x(\mathfrak{E}_n),t)|\geqq\frac{\mu(E)}{C}\sup_{0<t<1}|H_n(x(\mathfrak{E}_n),t)|. \tag{2. 19}$$

On the other hand,

$$\sup_{0<t<1}|H_n(x(\mathfrak{E}_n),t)|\geqq\int_0^1|H_n(x(\mathfrak{E}_n),t)|\,dt=\mu(I_{\mathfrak{E}_n}). \tag{2. 20}$$

Thus we obtain from (2. 19) and (2. 20)

$$\mu(EI_{\mathfrak{E}_n})\geqq\frac{\mu(E)}{C}\mu(I_{\mathfrak{E}_n}), \tag{2. 21}$$

i. e. the property b) of KNOPP's theorem holds for the class J. Thus T is ergodic, and therefore $g^*(x) = M(g)$ is constant almost everywhere. It remains to prove the existence of the function $h(x)$ satisfying (2. 3) and (2. 4), and the invariance of the measure $\nu(E) = \int_E h(x)\,dx$ with respect to the transformation T.

Let us put for any measurable subset E of $(0, 1)$

$$E(x) = \begin{cases} 1 & \text{for} \quad x \in E, \\ 0 & \text{for} \quad x \notin E \end{cases}$$

and

$$\nu_n(E) = \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}E) = \int_0^1 \left(\frac{1}{n} \sum_{k=0}^{n-1} E(T^k x) \right) dx. \tag{2. 22}$$

As $0 \leqq E(x) \leqq 1$, it follows from the existence almost everywhere of the limit (2. 2) proved above for $g(x) = E(x)$ and LEBESGUE's theorem, that

$$\lim_{n \to \infty} \nu_n(E) = \nu(E) \tag{2. 23}$$

exists for any measurable E. As by (2. 13)

$$\frac{1}{C} \mu(E) \leqq \nu(E) \leqq C\mu(E), \tag{2. 24}$$

$\nu(E)$ is a measure which is equivalent to the Lebesgue measure $\mu(E)$; the ν-measure of the interval $(0, 1)$ is evidently equal to 1.

It follows by (2. 22)

$$\nu_n(T^{-1}E) = \frac{n+1}{n} \nu_{n+1}(E) - \frac{\mu(E)}{n} \tag{2. 25}$$

and therefore

$$\nu(T^{-1}E) = \nu(E), \tag{2. 26}$$

i. e. ν is invariant with respect to the transformation T.

Let us put

$$h(x) = \frac{dV(x)}{dx} \tag{2. 27}$$

where $V(x) = \nu(I_{0,x})$; here $I_{0,x}$ denotes the interval $(0, x)$ $(0 \leqq x \leqq 1)$.

From the invariance of the measure ν with respect to T it follows, as well known, that

$$M(g) = \int_0^1 g(x)\,h(x)\,dx. \tag{2. 28}$$

Thus (2. 3) is proved. (2. 4) follows evidently from (2. 24). Thus Theorem 1 is completely proved.

Let us define the function $e_k(x)$ as follows: If $f(x)$ is decreasing, put for $1 \leqq k < T$

$$e_k(x) = \begin{cases} 1 & \text{for} \quad f(k+1) < x \leqq f(k), \\ 0 & \text{otherwise.} \end{cases}$$

If $f(x)$ is increasing, put for $0 \leqq k < T$

$$e_k(x) = \begin{cases} 1 & \text{for} \quad f(k) \leqq x < f(k+1), \\ 0 & \text{otherwise.} \end{cases}$$

Applying our theorem to $g(x) = e_k(x)$ it follows that the relative frequency of every admissible digit converges to a positive limit, for almost all x, and these limits depend only on the function $f(x)$ and not on x. The values of these limits can be calculated for a given $f(x)$ if we succeed in constructing explicitly the corresponding (uniquely determined) invariant measure ν.

§ 3. Some examples

EXAMPLE 1. Let us put

$$f(x) = \begin{cases} \dfrac{x}{q} & \text{for} \quad 0 \leqq x \leqq q, \\ 1 & \text{for} \quad x > q \end{cases}$$

where $q \geqq 2$ is an integer. Clearly conditions B1), B2$_2$) and B3) are satisfied, further condition C) is also satisfied (with $C = 1$) because $H_n(x, t)$ is identically equal to $\dfrac{1}{q^n}$. Thus we obtain as a special case of our Theorem 1 the theorem of RAIKOFF [6] and the classical theorem of BOREL [5] on normal decimals, respectively. In this special case $\nu(E) = \mu(E)$, i. e. the Lebesgue measure is invariant with respect to the tranformation $Tx = (qx)$.

EXAMPLE 2. Let us put $f(x) = \dfrac{1}{x}$ for $x \geqq 1$. Clearly conditions A1), A2$_1$) and A3) are satisfied. To show that condition C) is also satisfied, we need the well-known formula according to which if $\dfrac{p_k(x)}{q_k(x)}$ denotes the k-th convergent of the continued fraction of x, we have

$$f_n(\varepsilon_1(x), \ldots, \varepsilon_n(x) + t) = \frac{p_{n-1}(x)\,(\varepsilon_n(x) + t) + p_{n-2}(x)}{q_{n-1}(x)\,(\varepsilon_n(x) + t) + q_{n-2}(x)}.$$

It follows that

$$H_n(x, t) = \frac{(-1)^n}{(q_{n-1}(x)(\varepsilon_n(x) + t) + q_{n-2}(x))^2}$$

and thus

$$\frac{\sup\limits_{0<t<1} |H_n(x,t)|}{\inf\limits_{0<t<1} |H_n(x,t)|} = \left(1+\frac{q_{n-1}(x)}{q_n(x)}\right)^2 \leqq 4.$$

Consequently, condition C) is satisfied with $C=4$ and therefore[8] by (2. 12) $\mu(T^{-n}E) \leqq 4\mu(E)$. Thus we obtain as a special case of Theorem 1 the theorem of RYLL-NARDZEWSKI [12].

EXAMPLE 3. Let us consider the case when $f(x) = \sqrt[m]{1+x}-1$ for $0 \leqq x \leqq 2^m-1$ where $m \geqq 2$ is an integer. Conditions B1), B2$_2$) and B3) are clearly satisfied and thus every real number x can be represented in the form

$$x = \varepsilon_0 - 1 + \sqrt[m]{\varepsilon_1 + \sqrt[m]{\varepsilon_2 + \sqrt[m]{\varepsilon_3 + \cdots}}}$$

where the digits ε_n are generated by the recursion

$$\varepsilon_0 = [x], \qquad r_0 = (x),$$

$$\varepsilon_{n+1} = [(1+r_n)^m - 1], \qquad r_{n+1} = ((1+r_n)^m - 1) \qquad (n = 0, 1, \ldots),$$

and thus the digits ε_n are capable of the values $0, 1, \ldots, 2^m-2$. This algorithm may be called the algorithm of W. BOLYAI who used it to approximate the roots of some equations (in the special case $m=2$) in his book "Tentamen..." [3] published in the year 1832.

Let us verify that condition C) is fulfilled. We have clearly

$$\frac{\sup\limits_{0<t<1} H_n(x,t)}{\inf\limits_{0<t<1} H_n(x,t)} = \prod_{j=1}^{n} \left(\frac{\varepsilon_j + \sqrt[m]{\varepsilon_{j+1} + \sqrt[m]{\varepsilon_{j+2} + \cdots + \sqrt[m]{\varepsilon_n + 2}}}}{\varepsilon_j + \sqrt[m]{\varepsilon_{j+1} + \sqrt[m]{\varepsilon_{j+2} + \cdots + \sqrt[m]{\varepsilon_n + 1}}}} \right)^{1-\frac{1}{m}}.$$

Thus, owing to the inequality $\frac{a+c}{a+b} \leqq \frac{c}{b}$ if $0 < b \leqq c$ and $a \geqq 0$, it follows

$$\frac{\sup\limits_{0<t<1} H_n(x,t)}{\inf\limits_{0<t<0} H_n(x,t)} \leqq \prod_{j=1}^{n} \left(1+\frac{1}{\varepsilon_n+1}\right)^{\left(1-\frac{1}{m}\right)\cdot\frac{1}{m^{n-j}}} \leqq 2,$$

i. e. condition C) is satisfied with $C=2$.

[8] It has been shown by HARTMAN that more is true; we have $\mu(T^{-n}E) \leqq 2\mu(E)$ (see [14] and for another proof [16]).

§ 4. The β-expansion of real numbers

In this § we consider the case

$$f(x)=\begin{cases}\dfrac{x}{\beta} & \text{for} \quad 0\leqq x\leqq \beta,\\ 1 & \text{for} \quad \beta<x\end{cases}$$

where $\beta>1$ is not an integer. As conditions B1), B2$_3$) and B3) are clearly satisfied, it follows that every real number x can be represented in the form

$$x=\varepsilon_0+\frac{\varepsilon_1}{\beta}+\frac{\varepsilon_2}{\beta^2}+\cdots+\frac{\varepsilon_n}{\beta^n}+\cdots \tag{4.1}$$

where the digits ε_n can be obtained by the recursion formulae

$$\begin{aligned}\varepsilon_0&=[x], \quad r_0=(x),\\ \varepsilon_{n+1}&=[\beta r_n], \quad r_{n+1}=(\beta r_n) \qquad (n=0,1,\ldots).\end{aligned} \tag{4.2}$$

The digits ε_n which for $n\geqq 1$ are capable of the values $0,1,\ldots,[\beta]$ can be expressed without introducing the remainders r_n as

$$\begin{aligned}\varepsilon_0&=[x],\\ \varepsilon_1&=[\beta(x)],\\ \varepsilon_2&=[\beta(\beta(x))],\\ \varepsilon_3&=[\beta(\beta(\beta(x)))],\\ &\vdots\end{aligned} \tag{4.3}$$

In this case Tx is the transformation $Tx=(\beta x)$ of the interval $(0,1)$ onto itself.

We shall prove

THEOREM 2. *For any function $g(x)$ which is L-integrable in $(0,1)$ we have for almost all x*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}g(r_k(x))=M(g) \tag{4.4}$$

where the constant $M(g)$ does not depend on x. There exists further a measure ν which is equivalent to the Lebesgue measure μ and invariant with respect to the transformation $Tx=(\beta x)$, and for any measurable subset E of the interval $(0,1)$ we have

$$\nu(E)=\int_E h(x)\,dx \tag{4.5}$$

where $h(x)$ is a measurable function and

$$1-\frac{1}{\beta}\leqq h(x)\leqq\frac{1}{1-\dfrac{1}{\beta}}. \tag{4.6}$$

and we have

$$(4.7) \qquad M(g) = \int_0^1 g(x)\,h(x)\,dx.$$

PROOF. The β-expansion is an expansion with *dependent* digits. As a matter of fact, the admissible values for ε_n are $0, 1, \ldots, [\beta]$. But as

$$\sum_{n=1}^{\infty} \frac{[\beta]}{\beta^n} = \frac{[\beta]}{\beta - 1} > 1,$$

there exists a value N for which

$$\sum_{n=1}^{N} \frac{[\beta]}{\beta^n} > 1.$$

This implies that the first N digits can not all be equal to $[\beta]$.

Thus not every sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ formed from the numbers $0, 1, \ldots, [\beta]$ is canonical. Let $S(n)$ denote the number of canonical sequences of order n for $n \geqq 1$ and put $S(0) = 1$. Then $S(n) - S(n-1)$ is the number of those canonical sequences of order n for which $\varepsilon_n \neq 0$, because if $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1})$ is a canonical sequence of order $n-1$, then clearly $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1}, 0)$ is a canonical sequence of order n, and conversely. In general, if $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1}, \varepsilon_n)$ is a canonical sequence of order n, then $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1})$ is a canonical sequence of order $n-1$. Let us consider all canonical sequences $\mathfrak{E}_{n-1} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1})$ of order $n-1$. If $(\varepsilon_1, \ldots, \varepsilon_{n-1}, k)$ is canonical for $k \leqq k_{\mathfrak{E}_{n-1}}$ but not for $k > k_{\mathfrak{E}_{n-1}}$, then the intervals $\left[\frac{\varepsilon_1}{\beta} + \frac{\varepsilon_2}{\beta^2} + \cdots + \frac{\varepsilon_{n-1}}{\beta^{n-1}},\ \frac{\varepsilon_1}{\beta} + \frac{\varepsilon_2}{\beta^2} + \cdots + \frac{\varepsilon_{n-1}}{\beta^{n-1}} + \frac{k_{\mathfrak{E}_{n-1}}}{\beta^n}\right)$ are clearly disjoint, and thus we have

$$\frac{1}{\beta^n}(S(n) - S(n-1)) = \frac{1}{\beta^n} \sum k_{\mathfrak{E}_{n-1}} \leqq 1,$$

consequently

$$(4.8) \qquad S(n) - S(n-1) \leqq \beta^n \qquad (n = 1, 2, \ldots).$$

As $S(0) = 1$, we obtain

$$(4.9) \qquad S(n) \leqq \frac{\beta^{n+1}}{\beta - 1} \qquad (n = 1, 2, \ldots).$$

Let us arrange the $S(n)$ numbers $\frac{\varepsilon_1}{\beta} + \frac{\varepsilon_2}{\beta^2} + \cdots + \frac{\varepsilon_n}{\beta^n}$, where $\mathfrak{E}_n = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is a canonical sequence, and the number 1 according to their order of magnitude. Clearly the distance between any two consecutive terms does not exceed $\frac{1}{\beta^n}$. Thus we have

$$(4.10) \qquad S(n) \geqq \beta^n.$$

From (4. 10) and (4. 9) we obtain incidentally

$$\lim_{n\to\infty} \sqrt[n]{S(n)} = \beta. \tag{4. 11}$$

Now let E denote any measurable subset of the interval $(0, 1)$. As $T^{-n}E$ consists of $S(n)$ sets, each of which has a measure not exceeding $\frac{1}{\beta^n}\mu(E)$, we have

$$\mu(T^{-n}E) \leqq \frac{S(n)\mu(E)}{\beta^n} \leqq \frac{1}{1-\frac{1}{\beta}}\mu(E). \tag{4. 12}$$

On the other hand, $S(n)-S(n-1)$ of the sets mentioned above have the measure exactly equal to $\frac{\mu(E)}{\beta^n}$ and thus we obtain

$$\mu(T^{-n}E) \geqq \frac{(S(n)-S(n-1))\mu(E)}{\beta^n}. \tag{4. 13}$$

It follows by (4. 10) that

$$\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}E) \geqq \frac{1}{n}\left(1+\sum_{k=1}^{n-1}\frac{(S(k)-S(k-1))}{\beta^k}\right)\mu(E) \geqq \left(1-\frac{1}{\beta}\right)\mu(E).$$

Thus we have

$$\left(1-\frac{1}{\beta}\right)\mu(E) \leqq \frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}E) \leqq \frac{1}{1-\frac{1}{\beta}}\mu(E). \tag{4. 14}$$

Applying again the theorem of DUNFORD and MILLER, Theorem 2 follows exactly in the same way as Theorem 1 in § 2. As regards the ergodicity of the transformation $Tx=(\beta x)$, it can be proved in the same way by using KNOPP's theorem as the ergodicity of the transformations $Tx=(f(x))$ considered in § 2. The only difference consists in that we choose now for J the class of those intervals $\left[\frac{\varepsilon_1}{\beta}+\frac{\varepsilon_2}{\beta^2}+\cdots+\frac{\varepsilon_n}{\beta^n}, \frac{\varepsilon_1}{\beta}+\frac{\varepsilon_2}{\beta^2}+\cdots+\frac{\varepsilon_n+1}{\beta^n}\right)$ for which not only the sequence $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ but also $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n+1)$ is canonical.

Let us consider an example.

EXAMPLE 4. Let us take $\beta=\frac{\sqrt{5}+1}{2}$ and put $\alpha=\frac{1}{\beta}=\frac{\sqrt{5}-1}{2}$. Then we have $\alpha+\alpha^2=1$. This implies that each digit $\varepsilon_n=1$ is followed by a digit

$\varepsilon_{n+1}=0$ and there does not exist any other dependence of the digits on each other.[9] This makes it easy to obtain in this special case a complete insight into the set of canonical sequences. It can be shown that in this case

$$h(x)=\begin{cases}\dfrac{5+3\sqrt{5}}{10} & \text{for} \quad 0\leqq x<\dfrac{\sqrt{5}-1}{2},\\ \dfrac{5+\sqrt{5}}{10} & \text{for} \quad \dfrac{\sqrt{5}-1}{2}<x\leqq 1,\end{cases}$$

and thus the limiting frequencies of the digits 0 and 1 are $\dfrac{5+\sqrt{5}}{10}$ and $\dfrac{5-\sqrt{5}}{10}$, respectively.

We hope to return to the explicit determination for an arbitrary $\beta>1$ of the measure which is invariant with respect to the transformation $Tx=(\beta x)$ and is equivalent to the Lebesgue measure (the proof of the existence of which is contained in Theorem 2) at another occasion.

(Received 15 September 1957)

Bibliography

[1] B. H. Bissinger, A generalization of continued fractions, *Bulletin of the Amer. Math. Soc.*, **50** (1944), pp. 868—876.

[2] C. I. Everett, Representations for real numbers, *Bulletin of the Amer. Math. Soc.*, **52** (1946), pp. 861—869.

[3] W. Bolyai, *Tentamen iuventutem studiosam in elementa matheseos purae elementaris ac sublimioris methodo intuitiva evidentiaque huic propria introducendi*, ed. sec. (Budapest, 1897), Vol. I.

[4] Gy. Farkas, A Bolyai-féle algoritmus, *Értekezések a matematikai tudományok köréből*, **8** (1881), pp. 1—8, further see P. Veres, A Bolyai-féle algoritmus, *Mennyiségtani és Term. Tud. Didaktikai Lapok*, **1** (1943), pp. 57—62.

[5] É. Borel, Les probabilités dénombrables et leurs applications arithmétiques, *Rendiconti del Circ. Mat. di Palermo*, **27** (1909), pp. 247—271.

[6] D. Raikoff, On some arithmetical properties of summable functions, *Mat. Sbornik*, **1** (1936), pp. 377—384.

[9] I have shown in [16] that, in this special case, the sequence $\varepsilon_n(x)$ considered as a sequence of random variables (with respect to the Lebesgue measure) forms a Markov chain; and on the basis of this remark proved the assertions of Theorem 2 for the case $\beta=\dfrac{\sqrt{5}+1}{2}$ by means of probabilistic arguments.

[7] R. O. Kuzmin, Sur un problème de Gauss, *Atti del Congresso Internazionale del Matematici Bologna, (1928)*, Vol. VI, pp. 83—89.
[8] P. Lévy, *Théorie de l'addition des variables aléatoires* (Paris, 1954), Ch. IX, pp. 290.
[9] A. Khintchine, Metrische Kettenbruchprobleme, *Comp. Math.*, **1** (1935), pp. 359—382.
[10] A. Khintchine, Zur metrischen Kettenbruchtheorie, *Comp. Math.*, **3** (1936), pp. 276—285.
[11] A. Khintchine, *Kettenbrüche* (Leipzig, 1956).
[12] C. Ryll-Nardzewski, On the ergodic theorems. II. Ergodic theory of continued fractions, *Studia Math.*, **12** (1951), pp. 74—79.
[13] S. Hartman—E. Marczewski—C. Ryll-Nardzewski, Théorèmes ergodiques et leurs applications, *Coll. Math.*, **2** (1951), pp. 109—123.
[14] S. Hartman, Quelques propriétés ergodiques des fractions continues, *Studia Math.*, **12** (1951), pp. 271—278.
[15] F. Riesz, Sur la théorie ergodique, *Commentarii Math. Helv.*, **1** (1944—45), pp. 221—239.
[16] A. Rényi, Valós számok előállítására szolgáló algoritmusokról, *MTA Mat. és Fiz. Oszt. Közl.*, **7** (1957), pp. 265—293.
[17] N. Dunford and D. S. Miller, On the ergodic theorem, *Trans. Amer. Math. Soc.*, **60** (1946), pp. 538—549.
[18] F. Riesz, On a recent generalization of G. D. Birkhoff's ergodic theorem, *Acta Sci. Math. Szeged*, **11** (1948), pp. 193—200.
[19] K. Knopp, Mengentheoretische Behandlung einiger Probleme der diophantischen Approximationen und der transfiniten Wahrscheinlichkeiten, *Math. Annalen*, **95** (1926), pp. 409—426.