BULL. AUSTRAL. MATH. SOC.
VOL. 8 (1973), 191-203.

Arithmetical properties of the digits of the multiples of an irrational number

Kurt Mahler

Little seems to be known about the digits or sequences of digits in the decimal representation of a given irrational number like $\sqrt{2}$ or π. There is no difficulty in constructing an irrational number such that in its decimal representation certain digits or sequences of digits do *not* occur. On the other hand, well known theorems by Tchebychef, Kronecker, and Weyl imply that *some* integral multiple of the given irrational number always has any given finite sequence of digits occuring at least once in its decimal representation: for the fractional parts of the multiplies of the number lie dense in the interval $(0, 1)$.

In the present note I shall prove the following result.

Let α *be any positive irrational number and* N *any positive integer. Then there exists a positive integer* $P = P(N)$ *independent of* α *with the following property. There is an integer* X *satisfying* $1 \leq X \leq P$ *such that the decimal representation of* $X\alpha$ *contains infinitely often every possible sequence of* N *digits* $0, 1, 2, \ldots, 9$.

The proof is elementary. A very similar result can be shown for the digits in the canonical representation of any irrational p-adic number.

The proof given here is carried out for the more general case of

Received 21 October 1972.

the representation of the irrational number α to an arbitrary basis g where g is an integer at least 2 .

1.

Let $g \geq 2$ be a fixed integer, and let

$$D_g = \{0, 1, 2, \ldots, g-1\}$$

be the set of all digits to the basis g . By the *representation* of a positive number or a positive integer we shall always mean its representation to the basis g .

If x is any real number, $[x]$ denotes as usual its integral part and $(x) = x - [x]$ its fractional part.

Denote by α a fixed real *irrational* number satisfying

$$0 < \alpha < 1 ,$$

and by

$$\alpha = \sum_{h=1}^{\infty} a_h g^{-h} \quad (a_h \in D_h \quad \text{for all} \quad h \geq 1)$$

its representation. More generally, if X is any positive integer, let

$$(X\alpha) = \sum_{h=1}^{\infty} a_{X,h} g^{-h} \quad (a_{X,h} \in D_g \quad \text{for all} \quad h \geq 1)$$

be the representation of $(X\alpha)$. The ordered sequences of digits of α and $(X\alpha)$ will be denoted by

$$A = \{a_1, a_2, \ldots\} \quad \text{and} \quad A_X = \{a_{X,1}, a_{X,2}, \ldots\} ,$$

respectively, and their study forms the subject of this note.

2.

Let n be any positive integer, and let

$$B = \{b_0, b_1, \ldots, b_{n-1}\}$$

be any finite ordered set of digits in D_g . By a classical theorem by

H. Weyl, the numbers

$$(X\alpha) \quad (X = 1, 2, 3, \ldots)$$

are uniformly distributed (mod 1) . From this it follows easily that there are infinitely many X for which B is the sequence of certain n consecutive terms of A_X . The same result can also be deduced from Tchebychef's Theorem on inhomogeneous linear approximations, or from Kronecker's Theorem on such approximations.

In the present note, we shall try to determine an upper bound for X depending on n , but not on α , N_n say, such that there always is an integer X in the interval $1 \le X \le N_n$ such that B occurs infinitely often in the sequence A_X .

3.

Denote by m a second positive integer. The two linear forms in x and y ,

$$g^n\left(g^m\alpha x-y\right) \quad \text{and} \quad g^{-n}x$$

have the determinant 1 . Hence, by Minkowski's Theorem on linear forms, there exist two positive integers x and y not both zero such that

(1) $$|g^m\alpha x-y| < g^{-n} \quad \text{and} \quad |x| \le g^n .$$

In fact,

(2) $$1 \le |x| \le g^n .$$

For if $x = 0$, then $y \ne 0$ and therefore

$$1 \le |y| < g^{-n} ,$$

which is impossible.

With x, y also $-x, -y$ is a solution of (1) and (2). Without loss of generality let then from now on

(3) $$1 \le x \le g^n .$$

Assume further that m is sufficiently large so that

$$g^m\alpha \geq 1 .$$

Then

$$g^m\alpha x \geq 1 > g^{-n} \text{ and hence also } y \geq 1 .$$

Thus the following result holds.

LEMMA 1. *Let m be so large that $g^m\alpha \geq 1$. Then there exist two integers x and y depending on m and n such that*

$$(4) \qquad |g^m\alpha x-y| < g^{-n} , \quad 1 \leq x \leq g^n , \quad y \geq 1 .$$

4.

In the lemma just proved, keep the integer n fixed, but allow m to run successively over all integers satisfying $g^m\alpha \geq 1$. By the lemma, there exists to each such m a solution

$$x = x(m) \ , \quad y = y(m)$$

of (4). Thus $x(m)$ always is one of the finitely many numbers

$$1, 2, 3, \ldots, g^n ,$$

while m is allowed to assume infinitely many different values.

It follows that there exists an infinite sequence $S = \{m_k\}$ of integers $m = m_k$ satisfying

$$g^{m_1}\alpha \geq 1 , \quad m_1 < m_2 < m_3 < \ldots$$

such that

$$x(m_1) = x(m_2) = x(m_3) = \ldots , \ = x_0 \text{ say,}$$

retains a constant value x_0 independent of $m \in S$. Thus Lemma 1 can be strengthened as follows.

LEMMA 2. *There exists an infinite sequence S of positive integers $m = m_k$, an integer x_0 independent of $m \in S$, and an integer $y(m)$ depending on $m \in S$, such that always*

(5) $\left|g^m\alpha x_0-y(m)\right| < g^{-n}$, $1 \le x_0 \le g^n$, $y(m) \ge 1$ *for* $m \in S$.

In this lemma, x_0 may still be divisible by g . Denote by g^u the highest power of g which divides x_0 and put

$$x_0 = x_1 g^u ,$$

so that x_1 is not divisible by g and therefore distinct from g^n . The integer u satisfies $0 \le u \le n$ and does not depend on $m \in S$. Add u to all the elements m of S , call the resulting sums again $m = m_k$, and denote from now on by S the sequence of these new integers $m = m_k$. Then Lemma 2 can be replaced by the following stronger result.

LEMMA 3. *There exists an infinite sequence* S *of positive integers* $m = m_k$ *, a constant integer* x_1 *, and an integer* $y(m)$ *depending on* $m \in S$ *, such that*

(6) $\left|g^m\alpha x_1-y(m)\right| < g^{-n}$, $1 \le x_1 \le g^n - 1$, $g \nmid x_1$, $y(m) \ge 1$ *for* $m \in S$.

5.

The number

$$\alpha_m = g^m\alpha x_1 - y(m) \quad \text{where} \quad m \in S$$

cannot vanish because α is irrational; it is therefore either positive or negative. On replacing, if necessary, S by an infinite subsequence, we can in any case assume that α_m has a fixed sign for all the elements m of S . We write

$$S = S^+ \quad \text{or} \quad S = S^- ,$$

depending on whether α_m is positive or negative for all $m \in S$, respectively.

6.

Consider first the case when $S = S^{+}$, and hence, by (6),

$$0 < g^{m}\alpha x_1 - y(m) < g^{-n} < 1 \quad \text{for} \quad m \in S .$$

This means that $g^{m}\alpha x_1$ has the fractional part

$$\left(g^{m}\alpha x_1\right) = g^{m}\alpha x_1 - y(m)$$

and therefore satisfies the inequality

$$0 < \left(g^{m}\alpha x_1\right) < g^{-n} .$$

Hence the representation of $\left(g^{m}\alpha x_1\right)$ *begins with* n *digits zero.* Now the sequence A_{x_1}, as defined in §1, is obtained from the similar sequence $A_{g^{m}x_1}$ by adding at the beginning certain m digits the values of which are immaterial. Furthermore, this relation holds for all the elements m of $S = S^{+}$. *Hence the sequence* A_{x_1} *contains infinitely many subsequences at least of length* n *and consisting entirely of the digit* 0.

7.

A slightly different result holds when $S = S^{-}$. Now

$$0 > g^{m}\alpha x_1 - y(m) > -g^{-n} > -1 \quad \text{for} \quad m \in S .$$

This implies that

$$1 > g^{m}\alpha x_1 - [y(m)-1] > 1 - g^{-n} \quad \text{for} \quad m \in S$$

and that $y(m) - 1$ is the integral part of $g^{m}\alpha x_1$, hence that

$$1 - g^{-n} < \left(g^{m}\alpha x_1\right) < 1 \quad \text{for} \quad m \in S .$$

This inequality means that *the representation of* $\left(g^{m}\alpha x_1\right)$ *begins with* n *digits* $g - 1$. By a consideration similar to that in **§6** we deduce then that in the present case *the sequence* A_{x_1} *contains infinitely many subsequences at least of length* n *and consisting entirely of the digit* $g - 1$.

This result for $S = S^-$ can be put in a more convenient equivalent form. For this purpose put

$$\alpha^* = 1 - \alpha .$$

Then

$$1 - (x_1\alpha) \quad \text{and} \quad (x_1\alpha^*)$$

are identical because both numbers lie between 0 and 1 , and the difference

$$1 - x_1\alpha - x_1\alpha^* = 1 - x_1$$

is an integer. All but the first digit of $(x_1\alpha^*)$ are therefore obtained by subtracting the corresponding digit of $(x_1\alpha)$ from $g - 1$.

In analogy to A_X denote by A_X^* the ordered sequence of digits of $(X\alpha^*)$. On combining the result just proved with that obtained in §6, we arrive at the following result.

LEMMA 4. *Let* α *be an irrational number in the interval* $0 < \alpha < 1$ *, and put* $\alpha^* = 1 - \alpha$. *To every positive integer* n *there exists a positive integer* x_1 *satisfying*

$$1 \le x_1 \le g^n - 1$$

such that either in A_{x_1} *or in* $A_{x_1}^*$ *there are infinitely many subsequences at least of length* n *and consisting only of zeros.*

From now on denote by α_0 that one of the two numbers α and α^* to which the last lemma applies.

8.

In the representation

$$(x_1\alpha_0) = \sum_{h=1}^{\infty} a^0_{x_1,h} g^{-h} \quad \left(a^0_{x_1,h} \in D_g \quad \text{for all} \quad h \geq 1\right)$$

(the superscript 0 denotes that the representation is that of the fractional part of $x_1\alpha_0$), there are by Lemma 4 infinitely many sequences of at least n consecutive digits equal to zero; but since α_0 is irrational, there are of course also infinitely many digits distinct from zero.

These facts can be applied as follows. Denote by H an arbitrarily large positive integer. There exists then a *smallest* suffix h_0 greater than H such that

$$a^0_{x_1,h} = 0 \quad \text{for} \quad h_0 \leq h \leq h_0 + n - 1 ,$$

and there also exists a *smallest* suffix h_1 for which both

$$a^0_{x_1,h_1} \neq 0 \quad \text{and} \quad h_1 \geq h_0 + n .$$

Hence, in particular,

$$a^0_{x_1,h} = 0 \quad \text{if} \quad h_0 \leq h \leq h_1 - 1 .$$

With h_0 and h_1 so defined, put

$$s = \sum_{h=1}^{h_0-1} a^0_{x_1,h} g^{-h} \quad \text{and} \quad t = \sum_{h=h_1}^{\infty} a^0_{x_1,h} g^{h_1-h-1} ,$$

so that

$$(x_1\alpha_0) = s + g^{-(h_1-1)} t .$$

Evidently,

$$t = \sum_{j=1}^{\infty} a^0_{x_1,h_1+j-1} g^{-j} .$$

Here the digit in the first term is not less than 1 ; there are infinitely many digits not zero; and none of the digits is greater than $g - 1$. Therefore

(7) $$g^{-1} < t < 1 .$$

9.

Denote now by

$$B = \{b_0, b_1, \ldots, b_{n-1}\}$$

an arbitrary ordered sequence of n digits, and further put

$$b = b_0 g^{n-1} + b_1 g^{n-2} + \ldots + b_{n-1} .$$

If $b_0 = b_1 = \ldots = b_{n-1} = 0$, then, by Lemma 4, the sequence B occurs infinitely often in the ordered sequence of digits of $(x_1\alpha_0)$. Let this case therefore be excluded. Thus at least one of the n digits b_i is distinct from zero, all are at most $g - 1$, and it follows that

(8) $$1 \leq b \leq g^n - 1 .$$

The consecutive terms of the arithmetic progression

$$t, 2t, 3t, \ldots$$

are irrational and of distance less than 1. Therefore the *open* interval between any two consecutive integers contains at least one element of the progression.

It follows thus in particular that there is a positive integer x_2 such that

(9) $$b < x_2 t < b + 1 .$$

In other words, b is the integral part

$$b = [x_2 t]$$

of $x_2 t$. From (7), (8), and (9),

$$x_2 < \frac{b+1}{t} < g.g^n ,$$

whence, since x_2 is an integer,

$$(10) \qquad 1 \le x_2 \le g^{n+1} - 1 .$$

By (9) and by the definition of b, the number x_2t has the representation

$$x_2t = b_0g^{n-1} + b_1g^{n-2} + \dots + b_{n-1} + \sum_{j=1}^{\infty} b_j^* g^{-j}$$

where the b_j^* are certain digits the exact values of which play no role in the following considerations. This representation implies that

$$(11) \quad g^{-(h_1-1)}x_2t = b_0g^{-h_1+n} + b_1g^{-h_1+n-1} + \dots + b_{n-1}g^{-(h_1-1)} + \sum_{j=1}^{\infty} b_j^* g^{-h_1-j+1} .$$

On the other hand, since x_2 is an integer, the denominator of x_2s is a divisor of g^{h_0-1}; the highest *negative* power of g that occurs in the development to the basis g of x_2s is then at most $g^{-(h_0-1)}$, and here $h_0 - 1 < h_1 - n$.

Since evidently $x_2(x_1\alpha_0) - (x_1x_2\alpha_0)$ is a non-negative integer, and since

$$(x_2(x_1\alpha_0)) = (x_1x_2\alpha_0) ,$$

we have then found that in the representation

$$(x_1x_2\alpha_0) = \sum_{h=1}^{\infty} a^0_{x_1x_2,h}g^{-h}$$

the sequence of n consecutive digits

$$(12) \qquad a^0_{x_1x_2,h} , \text{ where } h_1 - n \le h \le h_1 - 1 ,$$

is identical with the given sequence B.

10.

In the construction just given, let now H tend to infinity. The values of $h_1 = h_1(H)$ and $x_2 = x_2(H)$ will vary with H, and h_1, being greater than H, will likewise tend to infinity. On the other hand, for all values of H the integer x_2 is restricted to the finite interval (10). It is then possible to select an infinite increasing sequence of integers H for which x_2 remains constant.

With this fixed value of x_2, put

$$X = x_1 x_2 ;$$

then, by Lemma 4 and by (10),

$$1 \leq X \leq \left(g^n-1\right)\left(g^{n+1}-1\right) < g^{2n+1} . \tag{13}$$

With this choice of X is has just been proved that, *for every ordered sequence* B *of* n *digits, the representation of at least one of the two numbers*

$$(X\alpha) \quad \textit{and} \quad (X\alpha^*)$$

contains infinitely many subsequences of n *consecutive digits identical with the corresponding digits of* B .

11.

Next associate with B the new ordered sequence of n digits

$$B^* = \{g-b_0-1,\ g-b_1-1,\ \ldots,\ g-b_{n-1}-1\} .$$

It is obvious that, when B runs over all ordered sequences of n digits, B^* does the same, and vice versa.

Further,

$$\alpha + \alpha^* = 1 , \quad 0 < (X\alpha) < 1 , \quad 0 < (X\alpha^*) < 1 , \quad X\alpha + X\alpha^* = X ,$$

and therefore

$$(X\alpha) + (X\alpha^*) = 1 .$$

Hence, whenever the sequence B occurs at some position

$h_1 - n \le h \le h_1 - 1$ in the representation of $(X\alpha)$, the second sequence B^* occurs at the same position in the representation of $(X\alpha^*)$; and naturally an analogous result holds with the two sequences B and B^* interchanged.

Thus, from what has been proved in §10, we obtain the following result.

THEOREM 1. *Let α be an arbitrary positive irrational number, n a positive integer, and $B = \{b_0, b_1, \ldots, b_{n-1}\}$ any ordered sequence of n digits*

$$0, 1, 2, \ldots, g-1 .$$

Then there exists a positive integer X satisfying

$$1 \le X < g^{2n+1}$$

such that B occurs infinitely often in the sequence of digits of the representation of $(X\alpha)$ and hence also that of $X\alpha$ to the basis g.

12.

One particular case of Theorem 1 has special interest.

It is known from combinatoric that for every positive integer N there exists an ordered sequence $B = \{b_0, b_1, \ldots, b_{n-1}\}$ of

$$n = g^N + N - 1$$

digits $0, 1, 2, \ldots, g-1$ such that the g^N subsequences

$$\{b_j, b_{j+1}, \ldots, b_{j+N-1}\} \quad (j = 0, 1, \ldots, g^N-1)$$

of B are exactly all g^N possible ordered sequences of N digits. On identifying the sequence B of Theorem 1 with this special sequence, the following result is found.

THEOREM 2. *Let α be an arbitrary positive irrational number, and N any positive integer. Then there exists an integer X satisfying*

$$1 \le X < g^{2g^N+2N-1}$$

such that every possible sequence of N digits occurs infinitely often in *the sequence of digits of the development of* $X\alpha$ *to the basis* g .

By way of example, let $N = 1$, and $g = 10$. The theorem shows then that for every positive irrational number α every digit $0, 1, \ldots, 9$ occurs infinitely often in the decimal representation of $X\alpha$ where X is a certain integer satisfying

$$1 \le X < 10^{21} .$$

Except for the upper bound for X , Theorem 2 is essentially best possible. For one can easily construct real numbers α with the following property.

To every positive integer X *there exists at least one sequence* B *which occurs at most finitely often in the representation of* $(X\alpha)$.

With very little change the method of this note can be applied to the canonical representation of irrational p-adic numbers when results completely analogous to Theorems 1 and 2 can be proved.

Department of Mathematics,
Institute of Advanced Studies,
Australian National University,
Canberra, ACT.