ACTA ARITHMETICA
103.2 (2002)

Additive properties of certain sets

by

Gergely Dombi (Budapest)

1. Introduction. Let $A = \{a_1, a_2, \ldots\}$ $(a_1 < a_2 < \ldots)$ be an infinite sequence of positive integers. For $n = 1, 2, \ldots$ let $R_1(A,n) = R_1(n)$, $R_2(A,n) = R_2(n)$, $R_3(A,n) = R_3(n)$ denote the number of solutions of

$$x + y = n, \quad x, y \in A,$$
$$x + y = n, \quad x < y,\ x, y \in A,$$
$$x + y = n, \quad x \le y,\ x, y \in A,$$

respectively. Furthermore let $r(k, A, n)$ denote the number of solutions of

$$x_1 + \ldots + x_k = n, \quad x_1, \ldots, x_k \in A,$$

so that $r(2, A, n) = R_1(A, n)$. Finally we use the notation $A(n)$ for the counting function

$$A(n) = \sum_{a \le n,\, a \in A} 1.$$

In this paper we construct sets of positive integers for which the functions $R_i(A,n)$, $r(k,A,n)$ have special properties. First we study the monotonicity property of $r(k,A,n)$, then we examine how much the function $R_2(A,n) = R_2(n)$ determines the corresponding set A. We start from a binary sequence $P = \{p_0, p_1, \ldots\} \in \{1,-1\}^\infty$ with certain pseudorandom properties; actually we deal with the Rudin–Shapiro and Thue–Morse sequences. Then we consider the corresponding pseudorandom set $Q = \{n \mid n \in \mathbb{N},\ p_{n-1} = 1\}$. Then by using certain properties of the sequence P we get arithmetic properties of the set Q.

2. The Rudin–Shapiro set. In [3] P. Erdős, A. Sárközy and V. T. Sós proved that if $r(2,A,n)$ is increasing for $n \ge n_0$ then A contains all the integers from a certain point on, i.e. there exists an integer n_1 with

$$A \cap \{n_1, n_1+1, n_1+2, \ldots\} = \{n_1, n_1+1, n_1+2, \ldots\}.$$

2000 *Mathematics Subject Classification*: Primary 11B34; Secondary 11B83, 11B85.

(See [1]–[3] for related problems.) Our Theorem 1 states that for $k > 4$ the opposite holds: there exists a set of density $1/2$ with $r(k, A, n)$ eventually increasing. Moreover, a constructive proof will be given, in the sense that the required set A will be given explicitly.

THEOREM 1. *There exists an $A \subset \mathbb{N}$ such that $r(k, A, n) = r(n)$ is increasing for every $k > 4$ and $n > n_0(k)$ and the density of A is equal to $1/2$.*

Proof. We will use the Rudin–Shapiro sequence so we briefly recall its definition and some of its basic properties. First we define the Rudin–Shapiro polynomials P_{2^n} and Q_{2^n} by the following recursion:

$$P_{2^n}(z) = \begin{cases} P_{2^{n-1}}(z) + z^{2^{n-1}} Q_{2^{n-1}}(z) & \text{if } n > 0, \\ 1 & \text{if } n = 0 \end{cases}$$

and

$$Q_{2^n}(z) = \begin{cases} P_{2^{n-1}}(z) - z^{2^{n-1}} Q_{2^{n-1}}(z) & \text{if } n > 0, \\ 1 & \text{if } n = 0. \end{cases}$$

The maximum modulus of these polynomials on the unit circle $|z| = 1$ can be estimated as

$$|P_{2^n}(z)| \le \sqrt{2} \cdot 2^{n/2} \quad \text{and} \quad |Q_{2^n}(z)| \le \sqrt{2} \cdot 2^{n/2} \quad \text{for } n = 0, 1, 2, \ldots,$$

since $|P_{2^n}(z)|^2 + |Q_{2^n}(z)|^2 = 2^{n+1}$.

By the construction of the polynomials

$$P_{2^n}(z) = \sum_{i=0}^{2^n - 1} p_i z^i$$

it is easy to see that their coefficients form a coherent sequence $\{p_0, p_1, \ldots\} \in \{-1, 1\}^\infty$, which satisfies the following recursion:

$$\begin{aligned} p_0 &= 1, \\ p_{2n} &= p_n \qquad \text{for } n = 1, 2, \ldots, \\ p_{2n+1} &= (-1)^n p_n \quad \text{for } n = 1, 2, \ldots \end{aligned}$$

(see [6, p. 73]). Let us define the corresponding power series

$$P(z) = \sum_{i=0}^{\infty} p_i z^i.$$

Thus for all $m \in \mathbb{N}$ we can define the polynomials

$$P_m(z) = \sum_{i=0}^{m-1} p_i z^i,$$

and in this general case a similar (but weaker) estimation holds:

$$|P_m(z)| \le (2+\sqrt{2})\, m^{1/2} \quad \text{for } m \in \mathbb{N},\ |z| = 1$$

(see [5]–[8]). We will extend this inequality to the whole $|z| \le 1$ by the Maximum Principle.

Let $A := \{n \in \mathbb{N} \mid p_{n-1} = 1\}$ and let $g(A,z) = g(z) = \sum_{a\in A} z^a$ denote its generating function. With these notations

$$(g(z))^k = \sum_{n\in\mathbb{N}} r(k,A,n) z^n.$$

Hence $r(n) - r(n-1)$ can be calculated by taking the coefficient of z^n in $(1-z)(g(z))^k$; the theorem is equivalent to the fact that this coefficient is nonnegative for all $n > n_0(k)$. In order to prove this we are going to investigate the integral

$$\int_0^1 (1-z)(g(z))^k z^{-n}\, dt$$

for a fixed n, where

$$z = z(t) = re(t) = r\exp(2\pi i t) \quad \text{with } r = e^{-1/(n-k)}.$$

Notice that

$$g(z) = \frac{z}{2}\left(\frac{1}{1-z} + P(z)\right),$$

so

$$\begin{aligned} r(n) - r(n-1) &= \int_0^1 (1-z)\left(\frac{z}{2}\right)^k \left(\frac{1}{1-z} + P(z)\right)^k z^{-n}\, dt \\ &= 2^{-k}\int_0^1 (1-z)\left(\frac{1}{1-z} + P_n(z)\right)^k z^{-(n-k)}\, dt \\ &= 2^{-k}\int_0^1 z^{-(n-k)}(1-z)\sum_{j=0}^{k}\binom{k}{j}\frac{1}{(1-z)^{k-j}} P_n^j(z)\, dt. \end{aligned}$$

Taking the integral of the first summand we get

$$I_0 = \int_0^1 \frac{1}{(1-z)^{k-1}} z^{-(n-k)}\, dt = \binom{n-2}{k-2}.$$

Now we estimate the remainder. Fix $1 \le j \le k-1$ and ε. With the notation $\|x\| = \min(\{x\}, 1-\{x\})$, in the interval

$$J_1 = \left\{ t \;\middle|\; 0 < t < 1,\ \|t\| \le \frac{1}{\sqrt{n}} \cdot \frac{1}{n^\varepsilon} \right\}$$

we have

$$\begin{aligned}
|I_{1,j}| &= \left| \int\limits_{J_1} \left(\frac{1}{1-z}\right)^{k-j-1} \binom{k}{j} P_n^j(z) z^{-(n-k)}\, dt \right| \\
&\le \binom{k}{j} \int\limits_{J_1} \left|\frac{1}{1-z}\right|^{k-j-1} e(2+\sqrt{2})^j n^{j/2}\, dt \\
&= \binom{k}{j} e(2+\sqrt{2})^j n^{j/2} \int\limits_{J_1} \left|\frac{1}{1-z}\right|^{k-j-1} dt \\
&\le \binom{k}{j} e(2+\sqrt{2})^j n^{j/2} n^{k-j-1} e \frac{1}{\sqrt{n}} \cdot \frac{1}{n^{\varepsilon}} \\
&= \binom{k}{j} e^2 (2+\sqrt{2})^j n^{k-j/2-3/2-\varepsilon}
\end{aligned}$$

where in the first inequality we applied our estimate of the Rudin–Shapiro polynomial on the $|z| = r = e^{-1/(n-k)}$ circle, while in the second one we used

$$1 - e^{-x} = x - \frac{x^2}{2!} + \frac{x^3}{3!} - \ldots = x\left(1 - \frac{x}{2!} + \frac{x^2}{3!} - \ldots\right),$$

thus

$$\frac{1}{|1-z|} \le \frac{1}{\big|1-|z|\big|} = \frac{1}{1-e^{-1/(n-k)}} < (n-k)\left(1 + \frac{1}{2(n-k)-1}\right),$$

and so

$$\left|\frac{1}{1-z}\right|^{k-j-1} \le (n-k)^{k-j-1}\left(1 + \frac{1}{2(n-k)-1}\right)^{k-j-1} < \frac{e}{2} n^{k-j-1},$$

for large enough n. In the interval

$$J_2 = \left\{ t \,\middle|\, 0 < t < 1,\ \|t\| \ge \frac{1}{\sqrt{n}} \cdot \frac{1}{n^{\varepsilon}} \right\},$$

since

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots > x - \frac{x^3}{6} = x\left(1 - \frac{x^2}{6}\right) > \frac{5}{6}x \quad \text{for } 0 < x < 1,$$

we have

$$\begin{aligned}
\left|\frac{1}{1-z}\right| = \left|\frac{1}{1-z(t)}\right| &\le \left|\frac{1}{1-z\left(\frac{1}{n^{1/2+\varepsilon}}\right)}\right| \le \left|\frac{1}{r \cdot \sin\left(\frac{2\pi}{n^{1/2+\varepsilon}}\right)}\right| \\
&\le \left|\frac{1}{e^{-1/(n-k)} \cdot \frac{5}{6} \cdot \frac{1}{n^{1/2+\varepsilon}}}\right|,
\end{aligned}$$

thus

$$\left|\frac{1}{1-z}\right|^{k-j-1} < e\left(\frac{6}{5}\right)^k n^{(1/2+\varepsilon)(k-j-1)}$$

in J_2 for large n. Using the above and the Rudin–Shapiro estimate we get

$$\begin{aligned}
|I_{2,j}| &= \left| \int\limits_{J_2} \left(\frac{1}{1-z}\right)^{k-j-1} \binom{k}{j} P_n^j(z) z^{-(n-k)}\, dt \right| \\
&\le \binom{k}{j} \int\limits_{J_2} \left|\frac{1}{1-z}\right|^{k-j-1} e(2+\sqrt{2})^j n^{j/2}\, dt \\
&< \binom{k}{j} (2+\sqrt{2})^j n^{j/2} e^2 \left(\frac{6}{5}\right)^k n^{(1/2+\varepsilon)(k-j-1)} \\
&= \binom{k}{j} (2+\sqrt{2})^j e^2 \left(\frac{6}{5}\right)^k n^{(k-1)/2+\varepsilon(k-j-1)}.
\end{aligned}$$

If we set

$$\varepsilon = \frac{1}{2} \cdot \frac{k-3}{k-1}$$

we can easily verify that the order of the remainder terms $I_{1,j}$ and $I_{2,j}$ are

$$|I_{1,j}| = O(n^{k-2-j/2+1/(k-1)}) \quad \text{and} \quad |I_{2,j}| = O(n^{k-2-j/2+j/(k-1)}),$$

both smaller than the order of the main term $I_0 = \binom{n-2}{k-2}$.

Finally for $j = k$ we can apply the Rudin–Shapiro estimate on the whole circle $r = e^{-1/(n-k)}$ to get

$$|I_k| = \left| \int\limits_0^1 (1-z) P_n^k(z) z^{-(n-k)}\, dt \right| \le 2e(2+\sqrt{2})^k n^{k/2}.$$

Since the number of the remainder terms is $2k-1$, their sum

$$\sum_{j=1}^{k-1} |I_{1,j}| + \sum_{j=1}^{k-1} |I_{2,j}| + |I_k| = O(n^{k-2-1/2+1/(k-1)})$$

still does not exceed the main term, which means that $r(n) - r(n-1)$ is positive for $n > n_0(k)$. Notice that all of our computations—except the last one—are valid for the case $k = 4$, and turn out to be false when $k = 3$.

Using the fact that the Rudin–Shapiro sequence is well distributed in arithmetic progressions (see [5]), it can be easily shown that A is of density $1/2$, which completes our proof.

Remark 1. Combining our method with a theorem of G. Halász on random polynomials (which extensively uses Fourier technique, see [4]) one can prove the existence of many such sets, but due to the probabilistic

methods applied, it does not yield any further construction. We point out that all these examples are of density 1/2.

3. The Thue–Morse set. A. Sárközy asked whether there exist two sets A and B of positive integers with infinite symmetric difference, i.e.

$$|(A \cup B) \setminus (A \cap B)| = \infty,$$

and having

$$R_i(A,n) = R_i(B,n), \quad n > n_0,$$

for $i = 1, 2, 3$. The answer is different for the cases $i = 1$ and $i = 2$. The case $i = 3$ is much more complicated, but our conjecture is that the answer is no. For $i = 1$, the answer is no. Assume that there exist such sets A and B, with $R_1(A,n) = R_1(B,n)$ for $n > n_0$. Since their symmetric difference is infinite, there exists a for which, say, $a \in A$, $a \notin B$, $a > n_0$. Then $R_1(A, 2a)$ is odd while $R_1(B, 2a)$ is even. For $i = 2$ the answer is yes as the following theorem shows.

THEOREM 2. *The set $\mathbb{N}$ of positive integers can be partitioned into two subsets A and B such that $R_2(A,n) = R_2(B,n)$ for all $n \in \mathbb{N}$.*

Proof. Let us consider the Thue–Morse sequence $M = \{m_0, m_1, \ldots\}$ which is defined in the following way. For $n \in \mathbb{N}$ let $S(n)$ denote the sum of digits of its dyadic representation and let

$$m_n = (-1)^{S(n)}.$$

It will be more convenient to use the shifted sequence $T = \{t_1, t_2, \ldots\}$ with $t_n = m_{n-1}$, which satisfies the following recursion:

$$t_1 = 1, \quad t_{2n-1} = t_n, \quad t_{2n} = -t_n \quad \text{for all } n \in \mathbb{N}.$$

Now we will construct the corresponding pseudorandom sets. Let

$$A = \{n \in \mathbb{N} \mid t_n = 1\} \quad \text{and} \quad B = \{n \in \mathbb{N} \mid t_n = -1\}.$$

We have to show that for these sets $R_2(A,n) = R_2(B,n)$. We use induction on n. The theorem holds for $n \le 5$. Suppose that we have proved it for all positive integers less than $n+1$ and let

$$n + 1 = 2^k + l + 1, \quad \text{where} \quad 0 \le l \le 2^k - 1.$$

First consider the decompositions $n + 1 = i + (n + 1 - i)$ with $1 \le i \le l$. Hence

$$S(2^k + i - 1) = S(i-1) + 1 \quad \text{and} \quad S(n-i) = S(l-i) + 1,$$

that is, $m_{2^k+i-1} \ne m_{i-1}$ and $m_{n-i} \ne m_{l-i}$, so $t_{2^k+i} \ne t_i$ and $t_{n+1-i} \ne t_{l+1-i}$, thus if i and $n + 1 - i$ are both in A then $2^k + i$ and $l + 1 - i$ are both in B and vice versa (we call such a decomposition *monochromatic*). So we can get a bijection between those monochromatic decompositions of

$n+1$ from A and B in which one of the summands is between 1 and l, by matching the monochromatic decomposition containing i with the one containing $l+1-i$. Indeed this is one-to-one correspondence, because if $l+1-i=i$ and $i=(l+1)/2 \in A$ then $n+1-i=2^k+i \in B$, so there is no such monochromatic decomposition.

Now consider those decompositions where the summands are between $l+1$ and 2^k. We will use the following lemma.

LEMMA 1. *Fix $k \in \mathbb{N}$. Then exactly one of the following two alternatives holds for all $0 \le l \le 2^k-1$:*

(i) $$|\{x \mid x+y=n+1,\ x,y \in A,\ l+1 \le x < y \le 2^k\}| = |\{x \mid x+y=2^k+1-l,\ x,y \in A,\ x<y\}|$$

and

$$|\{x \mid x+y=n+1,\ x,y \in B,\ l+1 \le x < y \le 2^k\}| = |\{x \mid x+y=2^k+1-l,\ x,y \in B,\ x<y\}|;$$

(ii) $$|\{x \mid x+y=n+1,\ x,y \in A,\ l+1 \le x < y \le 2^k\}| = |\{x \mid x+y=2^k+1-l,\ x,y \in B,\ x<y\}|$$

and

$$|\{x \mid x+y=n+1,\ x,y \in B,\ l+1 \le x < y \le 2^k\}| = |\{x \mid x+y=2^k+1-l,\ x,y \in A,\ x<y\}|.$$

Proof. Let $0 \le l \le 2^k-1$. Then $S(2^k-1-l)=k-S(l)$, so the modulo 2 congruence relation between $S(2^k-1-l)$ and $S(l)$ does not depend on l. This means that the Thue–Morse sequence $M=\{m_0,m_1,\ldots\}$ has the following property: if k is even then

$$\{m_{2^k-1}, m_{2^k-2}, \ldots, m_0\} = \{m_0, m_1, \ldots, m_{2^k-1}\}$$

and if k is odd then

$$\{m_{2^k-1}, m_{2^k-2}, \ldots, m_0\} = \{-(m_0), -(m_1), \ldots, -(m_{2^k-1})\}.$$

So if k is even, $x,y \in A$, $l+1 \le x < y \le 2^k$ and $m_{x-1}=m_{y-1}=1$ then $m_{2^k-1-x+1}=m_{2^k-1-y+1}=1$, so

$$y' = 2^k - x + 1 > 2^k - y + 1 = x'$$

are both in A. Furthermore, if $x+y=n+1$ then

$$2^k-x+1+2^k-y+1 = 2^{k+1}+2-(x+y) = 2^{k+1}+2-(2^k+l+1) = 2^k-l+1.$$

Therefore if k is even then the decompositions

$$x+y=n+1 \quad \text{with } x,y \in A,\ l+1 \le x < y \le 2^k$$

correspond uniquely to the decompositions

$$x' + y' = 2^k - l + 1 \quad \text{with } x', y' \in A,\ x' < y'.$$

Similarly we can make a one-to-one correspondence between the decompositions

$$x + y = n + 1 \quad \text{with } x, y \in B,\ l + 1 \le x < y \le 2^k$$

and

$$x' + y' = 2^k - l + 1 \quad \text{with } x', y' \in B,\ x' < y'.$$

When k is odd then a similar argument shows that the decompositions

$$x + y = n + 1 \quad \text{with } x, y \in A,\ l + 1 \le x < y \le 2^k,$$

correspond to

$$x' + y' = 2^k - l + 1 \quad \text{with } x', y' \in B,\ x' < y'$$

and vice versa. This proves the lemma.

For $1 \le l \le 2^k - 1$ we can combine the induction hypothesis

$$|\{x \mid x + y = 2^k + 1 - l,\ x, y \in A,\ x < y\}| = |\{x \mid x + y = 2^k + 1 - l,\ x, y \in B,\ x < y\}|$$

with the lemma to obtain

$$|\{x \mid x + y = n + 1,\ x, y \in A,\ l + 1 \le x < y \le 2^k\}| = |\{x \mid x + y = n + 1,\ x, y \in B,\ l + 1 \le x < y \le 2^k\}|.$$

For $l = 0$ we can apply the idea of the proof of the lemma to see that either $2^k + 1$ does not have any monochromatic decomposition or it has exactly 2^{k-2} decompositions from A and B. Finally for $l = 2^k - 1$ the only corresponding sum is $2^k + 2^k = 2^{k+1}$, which is not allowed.

Now we can summarize our results. We have already showed that

$$|\{x \mid x + y = n + 1,\ x, y \in A,\ 1 \le \min(x, y) \le l\}| = |\{x \mid x + y = n + 1,\ x, y \in B,\ 1 \le \min(x, y) \le l\}|$$

and

$$|\{x \mid x + y = n + 1,\ x, y \in A,\ l + 1 \le x < y \le 2^k\}| = |\{x \mid x + y = n + 1,\ x, y \in B,\ l + 1 \le x < y \le 2^k\}|,$$

so $R_2(A, n + 1) = R_2(B, n + 1)$, which was to be proved.

4. Unsolved problems. We present some related open problems here.

Problem 1. Let $k = 3, 4$. Is there any set $A \subset \mathbb{N}$ such that $\mathbb{N} \setminus A$ is infinite and $r(k, A, n)$ eventually increasing? Some computational experiments show that for $k = 4$ our example seems to be strictly increasing from the

beginning, but in the case $k = 3$ it is definitely not. According to these computer based results we conjecture that for $k = 3$ the answer is no (similarly to the case $k = 2$) and for $k = 4$ the answer is yes (as it is for $k > 4$).

PROBLEM 2. Let $k \geq 3$. Is there any set $A \subset \mathbb{N}$ with $r(k, A, n)$ eventually increasing and satisfying

$$\lim_{n\to\infty} \frac{A(n)}{n} = \alpha \quad \text{with } \alpha \neq \frac{1}{2}?$$

PROBLEM 3. Is there any pair of sets of positive integers A and B with

$$|(A \cup B) \setminus (A \cap B)| = \infty$$

and

$$R_3(A, n) = R_3(B, n), \qquad n > n_0?$$

PROBLEM 4. For two sets A and B we can consider the $\mathcal{L}_2$ distance of $R_i(A, n)$ and $R_i(B, n)$ for each $i = 1, 2, 3$. If we choose two sets having large gaps between their elements we can have

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} |R_i(A, j) - R_i(B, j)|^2 = 0,$$

so it is natural to normalize with the counting function of their union and study

$$d_i(A, B) = \limsup_{n\to\infty} \frac{1}{(A \cup B)(n)} \sum_{j=1}^{n} |R_i(A, j) - R_i(B, j)|^2.$$

For the Thue–Morse sets A and B by Theorem 2 we have

$$d_2(A, B) = 0 \quad \text{and} \quad d_1(A, B) = d_3(A, B) = 1/2.$$

Our conjecture is that for all sets A and B and for $i = 1, 3$,

$$\liminf_{n\to\infty} \frac{1}{(A \cup B)(n)} \sum_{j=1}^{n} |R_i(A, j) - R_i(B, j)|^2 \geq \frac{1}{2}.$$

PROBLEM 5. Do there exist two sets A and B with

$$|(A \cup B) \setminus (A \cap B)| = \infty$$

and

$$r(k, A, n) = r(k, B, n), \qquad n > n_0?$$

The similar question can be asked for the other two cases

$$r_2(k, A, n) = \sum_{\substack{x_1+\ldots+x_k=n \\ x_1<\ldots<x_k}} 1$$

and

$$r_3(k,A,n) = \sum_{\substack{x_1+\ldots+x_k=n \\ x_1\le\ldots\le x_k}} 1.$$

Acknowledgments. The author is grateful to András Sárközy and Mihály Szalay for their helpful advice and comments during the preparation of this paper.

References

[1] P. Erdős and A. Sárközy, *Problems and results on additive properties of general sequences I*, Pacific J. Math. 118 (1985), 347–357.

[2] P. Erdős, A. Sárközy and V. T. Sós, *Problems and results on additive properties of general sequences III*, Studia Sci. Math. Hungar. 22 (1987), 53–63.

[3] —, —, —, *Problems and results on additive properties of general sequences IV*, in: Number Theory (Ootacamund), Lecture Notes in Math. 1122, Springer, Berlin, 1984, 85–104.

[4] G. Halász, *On a result of Salem and Zygmund concerning random polynomials*, Studia Sci. Math. Hungar. 8 (1973), 369–377.

[5] C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences II*, J. Number Theory 73 (1998), 256–276.

[6] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, Lecture Notes in Math. 1294, Springer, New York, 1987.

[7] W. Rudin, *Some theorems on Fourier coefficients*, Proc. Amer. Math. Soc. 10 (1959), 855–859.

[8] H. S. Shapiro, *Extremal problems for polynomials and power series*, doctoral thesis, M.I.T., 1951.

Department of Differential Equations
Technical University of Budapest
Pf. 91, H-1521 Budapest, Hungary
E-mail: dombi@math.bme.hu

Received on 5.3.2001 (3990)