ACTA ARITHMETICA
108.1 (2003)

Unique representation bases for the integers

by

MELVYN B. NATHANSON (Bronx, NY)

1. Additive bases for the integers. Let A be a set of integers, and let $r_A(n)$ denote the number of representations of n in the form $n = a + a'$, where $a, a' \in A$ and $a \le a'$. The function $r_A(n)$ is called the *representation function* of the set A. An unsolved problem of Erdős and Turán states that if A is a subset of the semigroup $\mathbb{N}_0$ of nonnegative integers and $r_A(n) \ge 1$ for all sufficiently large integers n, then the representation function $r_A(n)$ is unbounded. On the other hand, it is known that the group of integers $\mathbb{Z}$ contains sets A with the property that $r_A(n) \ge 1$ for all $n \in \mathbb{Z}$ and $r(n)$ is bounded.

A set A of integers is called an *additive basis for the integers* if $r_A(n) \ge 1$ for all $n \in \mathbb{Z}$, and a *unique representation basis* if $r_A(n) = 1$ for all $n \in \mathbb{Z}$. The purpose of this paper is to construct a family of arbitrarily sparse unique representation bases for $\mathbb{Z}$. When a greedy algorithm is used in this construction, we obtain a unique representation basis A whose growth is logarithmic in the sense that the number of elements $a \in A$ with $|a| \le x$ is bounded above and below by constant multiples of $\log x$. In the last section of this paper we state some open problems suggested by the additive bases that we have constructed.

2. Bases with arbitrarily slow growth. For sets A and B of integers and for any integer c, we define the *sumset*

$$A + B = \{a + b : a \in A,\ b \in B\}$$

and the *translation*

$$A + c = \{a + c : a \in A\}.$$

2000 *Mathematics Subject Classification*: 11B13, 11B34, 11B05.

Key words and phrases: additive bases, representation functions, density, Erdős–Turán conjecture.

This work was supported in part by grants from the NSA Mathematical Sciences Program and the PSC-CUNY Research Award Program.

For the sumset

$$2A = A + A = \{a + a' : a, a' \in A\},$$

we have the *representation function*

$$r_A(n) = \text{card}\{(a, a') \in A \times A : a \leq a' \text{ and } a + a' = n\}.$$

The *counting function* for the set A is

$$A(y, x) = \text{card}\{a \in A : y \leq a \leq x\}.$$

In particular, $A(-x, x)$ counts the number of integers $a \in A$ such that $|a| \leq x$.

THEOREM 1. *Let $f(x)$ be a function such that $\lim_{x\to\infty} f(x) = \infty$. There exists an additive basis A for the group $\mathbb{Z}$ of integers such that*

$$r_A(n) = 1 \quad \textit{for all } n \in \mathbb{Z},$$

and

$$A(-x, x) \leq f(x) \quad \textit{for all sufficiently large } x.$$

Proof. We shall construct an ascending sequence $A_1 \subseteq A_2 \subseteq \ldots$ of finite sets such that

$$\begin{aligned} |A_k| &= 2k && \text{for all } k \geq 1, \\ r_{A_k}(n) &\leq 1 && \text{for all } n \in \mathbb{Z}, \\ r_{A_k}(n) &= 1 && \text{for all } n \text{ such that } |n| \leq k. \end{aligned}$$

It follows that the infinite set

$$A = \bigcup_{k=1}^{\infty} A_k$$

is a unique representation basis for the integers.

We construct the sets A_k by induction. Let $A_1 = \{0, 1\}$. We assume that for some $k \geq 1$ we have constructed sets $A_1 \subseteq \ldots \subseteq A_k$ such that $|A_k| = 2k$ and

$$r_{A_k}(n) \leq 1 \quad \text{for all } n \in \mathbb{Z}.$$

We define the integer

$$d_k = \max\{|a| : a \in A_k\}.$$

Then

$$A_k \subseteq [-d_k, d_k], \quad 2A_k \subseteq [-2d_k, 2d_k].$$

If both numbers d_k and $-d_k$ belong to A_k, then, since $0 \in A_1 \subseteq A_k$ and $d_k \geq 1$, we would have the following two representations of 0 in the sumset $2A_k$:

$$0 = 0 + 0 = (-d_k) + d_k.$$

This is impossible, since $r_{A_k}(0) \leq 1$, hence only one of the two integers d_k and $-d_k$ belongs to A_k. It follows that if $d_k \notin A_k$, then

$$\{2d_k, 2d_k - 1\} \cap 2A_k = \emptyset,$$

and if $-d_k \notin A_k$, then

$$\{-2d_k, -(2d_k - 1)\} \cap 2A_k = \emptyset.$$

Define the integer b_k by

$$b_k = \min\{|b| : b \notin 2A_k\}.$$

Then $1 \leq b_k \leq 2d_k - 1$.

To construct the set A_{k+1}, we choose an integer c_k such that $c_k \geq d_k$. If $b_k \notin 2A_k$, let

$$A_{k+1} = A_k \cup \{b_k + 3c_k, -3c_k\}.$$

We have

$$b_k = (b_k + 3c_k) + (-3c_k) \in 2A_{k+1}.$$

If $b_k \in 2A_k$, then $-b_k \notin 2A_k$ and we let

$$A_{k+1} = A_k \cup \{-(b_k + 3c_k), 3c_k\}.$$

Again we have

$$-b_k = -(b_k + 3c_k) + 3c_k \in 2A_{k+1}.$$

Since

$$d_k < 3c_k < 3c_k + b_k,$$

it follows that $|A_{k+1}| = |A_k| + 2 = 2(k+1)$. Moreover,

$$d_{k+1} = \max\{|a| : a \in A_{k+1}\} = b_k + 3c_k.$$

For example, since $A_1 = \{0, 1\}$ and $2A_1 = \{0, 1, 2\}$, it follows that $d_1 = b_1 = 1$. For $c_1 \geq 1$ we have

$$A_2 = \{-(1 + 3c_1), 0, 1, 3c_1\}.$$

Then

$$2A_2 = \{-(2 + 6c_1), -(1 + 3c_1), -3c_1, -1, 0, 1, 2, 3c_1, 1 + 3c_1, 6c_1\}$$

and $d_2 = 1 + 3c_1$ and $b_2 = 2$.

We can assume that $b_k \notin 2A_k$, hence $A_{k+1} = A_k \cup \{b_k + 3c_k, -3c_k\}$. (The argument in the case $b_k \in 2A_k$ and $-b_k \notin 2A_k$ is similar.) We shall show that the sumset $2A_{k+1}$ is the disjoint union of the following four sets:

$$2A_{k+1} = 2A_k \cup (A_k + b_k + 3c_k) \cup (A_k - 3c_k) \cup \{b_k, 2b_k + 6c_k, -6c_k\}.$$

If $u \in 2A_k$, then

$$-2c_k \leq -2d_k \leq u \leq 2d_k \leq 2c_k.$$

Suppose that $v = a + b_k + 3c_k \in A_k + b_k + 3c_k$, where $a \in A_k$. The inequalities

$$-c_k \leq -d_k \leq a \leq d_k \leq c_k \quad \text{and} \quad 1 \leq b_k \leq 2d_k - 1 \leq 2c_k - 1$$

imply that

$$2c_k + 1 \leq v \leq 6c_k - 1 < 2b_k + 6c_k.$$

Similarly, if $w = a - 3c_k \in A_k - 3c_k$, then

$$-6c_k < -4c_k \leq w \leq -2c_k.$$

These inequalities imply that the sets $2A_k$, $A_k + b_k + 3c_k$, $A_k - 3c_k$, and $2\{b_k + 3c_k, -3c_k\}$ are pairwise disjoint, unless $c_k = d_k$ and $-2d_k \in 2A_k \cap (A_k - 3d_k)$. If $-2d_k \in 2A_k$, then $-d_k \in A_k$. If $-2d_k \in A_k - 3d_k$, then $d_k \in A_k$. This is impossible, however, because A_k does not contain both integers d_k and $-d_k$.

Since the sets $A_k + b_k + 3c_k$ and $A_k - 3c_k$ are translations, it follows that

$$r_{A_{k+1}}(n) \leq 1 \quad \text{for all integers } n.$$

Let $A = \bigcup_{k=1}^{\infty} A_k$. For all $k \geq 1$ we have $2 = b_2 \leq b_3 \leq \ldots$ and $b_k < b_{k+2}$, hence $b_{2k} \geq k + 1$. Since b_{2k} is the minimum of the absolute values of the integers that do not belong to $2A_{2k}$, it follows that

$$\{-k, -k+1, \ldots, -1, 0, 1, \ldots, k-1, k\} \subseteq 2A_{2k} \subseteq 2A$$

for all $k \geq 1$, and so A is an additive basis. If $r_A(n) \geq 2$ for some n, then $r_{A_k}(n) \geq 2$ for some k, which is impossible. Therefore, A is a unique representation basis for the integers.

We observe that if $x \geq d_1$ and k is the unique integer such that $d_k \leq x < d_{k+1}$, then

$$A(-x, x) = A_{k+1}(-x, x) = \begin{cases} 2k & \text{for } d_k \leq x < 3c_k, \\ 2k+1 & \text{for } 3c_k \leq x < b_k + 3c_k = d_{k+1}. \end{cases}$$

In the construction of the set A_{k+1}, the only constraint on the choice of the number c_k was that $c_k \geq d_k$. Given a function $f(x)$ that tends to infinity, we shall use induction to construct a sequence of integers $\{c_k\}_{k=1}^{\infty}$ such that $A(-x, x) \leq f(x)$ for all $x \geq c_1$. We begin by choosing the integer $c_1 \geq d_1$ so that

$$f(x) \geq 4 \quad \text{for } x \geq c_1.$$

Then

$$A(-x, x) \leq 4 \leq f(x) \quad \text{for } c_1 \leq x \leq d_2.$$

Let $k \geq 2$, and suppose we have selected an integer $c_{k-1} \geq d_{k-1}$ such that

$$f(x) \geq 2k \quad \text{for } x \geq c_{k-1}, \qquad A(-x, x) \leq f(x) \quad \text{for } c_1 \leq x \leq d_k.$$

There exists an integer $c_k \geq d_k$ such that

$$f(x) \geq 2k + 2 \quad \text{for } x \geq c_k.$$

Then

$$\begin{aligned} A(-x, x) &= 2k \leq f(x) && \text{for } d_k \leq x < 3c_k, \\ A(-x, x) &\leq 2k + 2 \leq f(x) && \text{for } 3c_k \leq x \leq d_{k+1}, \end{aligned}$$

hence $A(-x,x) \leq f(x)$ for $c_1 \leq x \leq d_{k+1}$. It follows that

$$A(-x,x) \leq f(x) \quad \text{for all } x \geq c_1.$$

This completes the proof of Theorem 1.

3. Bases with logarithmic growth. In Theorem 1 we constructed unique representation bases whose counting functions tend slowly to infinity. It is natural to ask if there exist unique representation bases that are dense in the sense that their counting functions tend rapidly to infinity. In the following theorem we use the previous algorithm to construct a unique representation basis A whose counting function $A(-x,x)$ has order of magnitude $\log x$.

THEOREM 2. *There exists a unique representation basis A for the integers such that*

$$\frac{2\log x}{\log 5} + 2\left(1 - \frac{\log 3}{\log 5}\right) \leq A(-x,x) \leq \frac{2\log x}{\log 3} + 2 \quad \textit{for all } x \geq 1.$$

Proof. We apply the method of Theorem 1 with

$$c_k = d_k \quad \text{for all } k \geq 1.$$

This is essentially a greedy algorithm construction, since at each iteration we choose the smallest possible value of c_k. It is instructive to compute the first few sets A_k. Since

$$A_1 = \{0,1\}, \quad 2A_1 = \{0,1,2\},$$

we have $b_1 = 1$ and $c_1 = d_1 = 1$. Then

$$A_2 = \{-4,0,1,3\}, \quad 2A_2 = \{-8,-4,-3,-1,0,1,2,3,4,6\},$$

hence $b_2 = 2$, $c_2 = d_2 = 4$. The next iteration of the algorithm produces the sets

$$\begin{aligned} A_3 &= \{-14,-4,0,1,3,12\}, \\ 2A_3 &= \{-28,-18,-14,-13,-11,-8,-4,-3\} \\ &\quad \cup \{-2,-1,0,1,2,3,4,6,8,12,13,15,24\}, \end{aligned}$$

so we obtain $b_3 = 5$, $c_3 = d_3 = 14$, and

$$A_4 = \{-42,-14,-4,0,1,3,12,47\}.$$

We shall compute upper and lower bounds for the counting function $A(-x,x)$. For $k \geq 1$ we have $1 \leq b_k \leq 2c_k - 1$ and $c_{k+1} = 3c_k + b_k$, hence

$$3c_k + 1 \leq c_{k+1} \leq 5c_k - 1.$$

Since $c_1 = 1$, it follows by induction on k that

$$\frac{3^k - 1}{2} \leq c_k \leq \frac{3 \cdot 5^k + 5}{20}$$

and so

$$\frac{\log c_k}{\log 5} + 1 \le \frac{\log((20c_k - 5)/3)}{\log 5} \le k \le \frac{\log(2c_k + 1)}{\log 3} \le \frac{\log c_k}{\log 3} + 1$$

for all $k \ge 1$. We obtain an upper bound for $A(-x, x)$ as follows. If $c_k \le x < 3c_k$, then

$$A(-x, x) = A_k(-x, x) = 2k \le \frac{2\log c_k}{\log 3} + 2 \le \frac{2\log x}{\log 3} + 2.$$

If $3c_k \le x < c_{k+1}$, then

$$\begin{aligned} A(-x, x) = A_{k+1}(-x, x) = 2k + 1 &\le \frac{2\log c_k}{\log 3} + 3 \\ &\le \frac{2\log(x/3)}{\log 3} + 3 \le \frac{2\log x}{\log 3} + 1. \end{aligned}$$

Therefore,

$$A(-x, x) \le \frac{2\log x}{\log 3} + 2 \qquad \text{for all } x \ge 1.$$

We obtain a lower bound for $A(-x, x)$ similarly. If $c_k \le x < 3c_k$, then

$$\begin{aligned} A(-x, x) = 2k &\ge \frac{2\log c_k}{\log 5} + 2 \ge \frac{2\log(x/3)}{\log 5} + 2 \\ &\ge \frac{2\log x}{\log 5} + 2\left(1 - \frac{\log 3}{\log 5}\right) = \frac{2\log x}{\log 5} + 0.63\ldots \end{aligned}$$

If $3c_k \le x < c_{k+1}$, then, since

$$c_{k+1} = d_{k+1} = b_k + 3c_k \le 5c_k - 1,$$

we have

$$A(-x, x) = 2k + 1 \ge \frac{2\log c_k}{\log 5} + 3 > \frac{2\log(x/5)}{\log 5} + 3 = \frac{2\log x}{\log 5} + 1.$$

Therefore,

$$A(-x, x) \ge \frac{2\log x}{\log 5} + 2\left(1 - \frac{\log 3}{\log 5}\right) \qquad \text{for all } x \ge 1.$$

This completes the proof of Theorem 2.

4. Heuristics and open problems. Let A be a set of integers. If A is a unique representation basis for $\mathbb{Z}$, or, more generally, if A is a set of integers with a bounded representation function, then $A(-x, x) \ll \sqrt{x}$. The following simple result gives an explicit upper bound.

THEOREM 3. *Let A be a nonempty set of integers such that the representation function of A is bounded. If $r_A(n) \le r$ for all n, then $A(-x, x) \le \sqrt{8rx}$ for all $x \ge r$.*

Proof. Let $A(-x,x) = k$. The number of ordered pairs $(a,a') \in A \times A$ with $-x \leq a \leq a' \leq x$ is exactly $(k^2+k)/2$. For each of these ordered pairs we have $-2x \leq a+a' \leq 2x$. For each integer $n \in [-2x, 2x]$ there are at most r such pairs (a,a') with $a+a'=n$, and so

$$\frac{k^2+k}{2} \leq r(4x+1).$$

It follows that

$$A(-x,x) = k \leq \sqrt{8rx + \frac{8r+1}{4}} - \frac{1}{2} \leq \sqrt{8rx} \qquad \text{for } x \geq r.$$

This completes the proof.

Theorem 3 has a natural analogue for sets A of nonnegative integers.

THEOREM 4. *Let A be a set of nonnegative integers such that every sufficiently large integer can be represented as the sum of two elements of A. If $r_A(n) \geq 1$ for all $n > n_0$, then*

$$A(0,x) \geq 2\sqrt{x} - 1 \qquad \textit{for all } x \geq n_0^2.$$

If A is a set of nonnegative integers such that $r_A(n) \leq r$ for all $n \geq 0$, then

$$A(0,x) \leq 2\sqrt{rx} \qquad \textit{for all } x \geq 1.$$

Proof. Let $A(0,x) = k$. Suppose that $r_A(n) \geq 1$ for all $n > n_0$. The number of ordered pairs $(a,a') \in A \times A$ with $0 \leq a \leq a' \leq x$ is exactly $(k^2+k)/2$. For each such pair we have $0 \leq a+a' \leq 2x$. For each integer n with $n_0 < n \leq 2x$ there is at least one pair (a,a') with $a+a'=n$, and so

$$\frac{k^2+k}{2} \geq 2x - n_0.$$

It follows that

$$A(0,x) = k \geq \sqrt{4x - 2n_0 + \frac{1}{4}} - \frac{1}{2} \geq 2\sqrt{x} - 1 \qquad \text{for } x \geq n_0^2.$$

Suppose that $r_A(n) \leq r$ for all $n \geq 0$. If $a, a' \in A$ and $0 \leq a \leq a' \leq x$, then $0 \leq a+a' \leq 2x$. Since $r_A(0) \leq 1$ and $r_A(1) \leq 1$, it follows, as in the proof of Theorem 3, that

$$\frac{k^2+k}{2} \leq r(2x-1)+2,$$

and so

$$A(0,x) = k \leq \sqrt{4rx + \frac{17-8r}{4}} - \frac{1}{2} \leq 2\sqrt{rx} \qquad \text{for } x \geq 1.$$

This completes the proof.

A set A of nonnegative integers is called a *basis* (resp. an *asymptotic basis*) if every (resp. every sufficiently large) integer can be represented as the sum of two elements of A. By Theorem 1, there exist arbitrarily sparse sets of integers that are unique representation bases for $\mathbb{Z}$. On the other hand, by Theorem 4, a set A of nonnegative integers that is a basis or asymptotic basis for the set of nonnegative integers must have a counting function $A(0,x)$ that grows at least as fast as $\sqrt{x}$, and if the representation function of A is bounded, then $A(0,x)$ cannot grow faster than a constant multiple of $\sqrt{x}$. This phenomenon can be interpreted as follows: If $0 \le n \le N$, then there are infinitely many pairs (a,a') of integers whose sum is n, and the summands a and a' can be arbitrarily large in absolute value. On the other hand, if a and a' are constrained to be nonnegative integers, then they must be chosen from the finite number of integers in the bounded interval $[0,N]$. If A is an asymptotic basis, then A is forced to contain many numbers in the interval $[0,N]$, and this increases the probability that some number has many representations. This phenomenon may underlie the Erdős–Turán conjecture.

Theorem 1 asserts that a unique representation basis A for the integers can be arbitrarily sparse, while Theorem 3 states that A cannot be too dense, since $A(-x,x) \ll \sqrt{x}$. In Theorem 2 we constructed a unique representation basis such that $A(-x,x) \ge (2/\log 5)\log x + 0.63$. It is not known what functions can be lower bounds for counting functions of unique representation bases. Here are some unsolved problems on this theme.

1. For each real number $c > 2/\log 5$, does there exist a unique representation basis A such that $A(-x,x) \ge c\log x$ for all sufficiently large x?

2. Does there exist a unique representation basis A such that

$$\lim_{x\to\infty} \frac{A(-x,x)}{\log x} = \infty?$$

3. Does there exist a number $\theta > 0$ and a unique representation basis A such that $A(-x,x) \ge x^\theta$ for all sufficiently large x?

4. Does there exist a number $\theta < 1/2$ such that $A(-x,x) \le x^\theta$ for every unique representation basis A and for all sufficiently large x?

Department of Mathematics
Lehman College (CUNY)
Bronx, NY 10468, U.S.A.
E-mail: nathansn@alpha.lehman.cuny.edu

Received on 20.2.2002
and in revised form on 28.10.2002 (4233)