THE INVERSE PROBLEM FOR REPRESENTATION FUNCTIONS FOR GENERAL LINEAR FORMS

Peter Hegarty
Department of Mathematics, Chalmers University of Technology and Gothenburg University, 41296 Gothenburg, Sweden
hegarty@math.chalmers.se

Received: 12/28/07, Revised: 1/17/08, Accepted: 4/8/08, Published: 4/17/08

Abstract

The inverse problem for representation functions takes as input a triple $(\mathbb{X}, f, \mathcal{L})$, where $\mathbb{X}$ is a countable semigroup, $f : \mathbb{X} \to \mathbb{N}_0 \cup \{\infty\}$ a function, $\mathcal{L} : a_1x_1 + \cdots + a_hx_h$ an $\mathbb{X}$-linear form and asks for a subset $A \subseteq \mathbb{X}$ such that there are $f(x)$ solutions (counted appropriately) to $\mathcal{L}(x_1, \ldots, x_h) = x$ for every $x \in \mathbb{X}$, or a proof that no such subset exists.

This paper represents the first systematic study of this problem for arbitrary linear forms when $\mathbb{X} = \mathbb{Z}$, the setting which in many respects is the most natural one. Having first settled on the 'right' way to count representations, we prove that every primitive form has a unique representation basis, i.e.: a set A which represents the function $f \equiv 1$. We also prove that a partition regular form (i.e.: one for which no non-empty subset of the coefficients sums to zero) represents any function f for which $\{f^{-1}(0)\}$ has zero asymptotic density. These two results answer questions recently posed by Nathanson.

The inverse problem for partition irregular forms seems to be more complicated. The simplest example of such a form is $x_1 - x_2$, and for this form we provide some partial results. Several remaining open problems are discussed.

1. Introduction and Definitions

A fundamental notion in additive number theory is that of *basis*. Given a positive integer h, a subset $A \subseteq \mathbb{N}_0$ for which $0 \in A$ is said to be a *basis* for $\mathbb{N}_0$ of *order* h if, for every $n \in \mathbb{N}_0$ the equation

$$x_1 + x_2 + \cdots + x_h = n \tag{1.1}$$

has at least one solution in A. The requirement that $0 \in A$ means that, in words, A is a basis of order h if every positive integer can be written as the sum of at most h positive integers from A.

In classical number theory, we encounter questions of the type: is the following set A a basis for $\mathbb{N}_0$ and, if so, of what order? Famous examples include the cases when A is the set $\mathbb{N}_0^k$ of perfect k:th powers, for some fixed k (Waring's Problem), or the set $\mathbb{P}_{0,1}$ of primes together with 0 and 1 (Goldbach's Problem). In both these cases, it is in fact more natural to consider a slightly weaker notion, namely that of *asymptotic basis*. A subset $A \subseteq \mathbb{N}_0$ is said to be an asymptotic basis of order h if (1.1) has a solution for every $n \gg 0$. For example, in Waring's Problem, if $g(k)$ and $G(k)$ denote the order, resp. asymptotic order, of the set $\mathbb{N}_0^k$, then it is known that $G(k)$ is considerably less than $g(k)$ for large k. Regarding the primes, Vinogradov's Theorem says that $\mathbb{P}_{0,1}$ is an asymptotic basis of order 4, while it remains open as to whether it is actually a basis of even that order. Goldbach's conjecture would imply the much stronger result that $\mathbb{P}_{0,1}$ is a basis of order 3. In this regard, it is well-known that the subset of the positive even integers representable as the sum of two primes has asymptotic density one (see, for example, [9] Theorem 3.7). This motivates a further fairly natural weakening of the notion of basis. In the terminology of [7], we say that $A \subseteq \mathbb{N}_0$ is a *basis of order h for almost all* $\mathbb{N}_0$ if the set of $n \in \mathbb{N}_0$ for which (1.1) has a solution in A has asymptotic density one.

In the terminology commonly used by practitioners of the subject, the above classical problems are illustrations of a *direct problem*, where we are in essence seeking a description of the h-fold sumset of a specified set A. The corresponding *inverse problem* is to construct a set A with a specified so-called *(unordered) representation function* of a certain order. Let $f : \mathbb{N}_0 \to \mathbb{N}_0$ be any function, $h \in \mathbb{N}$ and $A \subseteq \mathbb{N}_0$. We say that f is the (unordered) representation function of A of order h if, for every $n \in \mathbb{N}_0$,

$$f(n) = \#\{(x_1, \ldots, x_h) \in A^h : x_1 \leq x_2 \leq \cdots \leq x_h \text{ and (1.1) holds}\}. \tag{1.2}$$

If (1.2) holds then we write $f = f_{A,h}$. The relationship between bases and representations functions is thus that

- A is a basis of order h if and only if $f_{A,h}^{-1}(0)$ is empty,
- A is an asymptotic basis of order h if and only if $f_{A,h}^{-1}(0)$ is finite,
- A is a basis of order h for almost all $\mathbb{N}_0$ if and only if $d[f_{A,h}^{-1}(0)] = 0$.

The inverse problem for bases/representation functions in $\mathbb{N}_0$ is, in general, very hard. Probably the single most famous illustration of this is the long-standing question of Erdős and Turán [3] as to whether there exists an asymptotic basis of any order h whose representation function is bounded. Not much is known beyond the facts that, on the one hand, $f_{A,2}$ cannot be ultimately constant [1] while, on the other, there exist for every h so-called *thin* bases A_h satisfying $f_{A_h,h}(n) = \Theta(\log n)$ [2].

In seeking a more tractable inverse problem, a natural starting point are the following two observations :

First, the various notions of basis make sense in any additive semigroup, not just $\mathbb{N}_0$.

Second, it is easy to see intuitively why the inverse problem is hard in $\mathbb{N}_0$. Namely, when trying to construct a set A with a given representation function f of a given order h, we cannot use negative numbers to help 'fill in gaps'. More precisely, suppose we try to construct our set A one element at a time and at some point have constructed a finite set A' such that

$$f_{A',h}(n) \leq f(n) \text{ for every } n \in \mathbb{N}_0. \tag{1.3}$$

Assuming $f^{-1}(0)$ is finite, say, there will be a smallest $n = n_1$ for which we have strict inequality in (1.3). We would now like to add some more elements to A' which create a new solution to (1.1) for $n = n_1$ while not violating (1.3). If we could use negative numbers then, as long as $f^{-1}(0)$ is finite, a natural way to do this would be to add to A' exactly h new elements which (a) don't all have the same sign (b) are all much larger in absolute value than anything currently in A' (c) almost cancel each other out in exactly one way, in which case they add up to n_1.

These observations led Nathanson to consider the inverse problem for representation functions in $\mathbb{Z}$ or, more generally, in countable abelian groups. The fundamental result showing that we have a much more tractable problem in this setting is the following :

Theorem 1.1 [5] *Let $f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\}$ be any function for which $f^{-1}(0)$ is a finite set. Then for every $h \in \mathbb{N}_{\geq 2}$ there exists a subset $A \subseteq \mathbb{Z}$ such that $f_{A,h} = f$.*

In particular, the Erdős-Turán question has a positive answer in $\mathbb{Z}$: we can even construct a set A such that $f_{A,h}(n) = 1$ for every n, a so-called *unique representation basis of order h* for $\mathbb{Z}$. Nathanson's proof of Theorem 1.1 follows the idea in the second observation above.

Now that we have a more tractable problem, we can look to push our investigations deeper. One line of enquiry which seems natural is to extend the basic notion of basis further by replacing the left-hand side of (1.1) by an arbitrary linear form $a_1x_1 + \cdots + a_hx_h$. If the a_i are assumed to be integers, then this idea makes sense in any additive semigroup, otherwise one should work in a commutative ring. For the remainder of this paper, though, we shall always be working in $\mathbb{Z}$, but the interested reader is invited to extend the discussion to a more general setting. Note that the various notions of basis are only meaningful if the linear form is *primitive*, i.e.: if the coefficients are relatively prime. This will be assumed throughout.

We now start with a couple of formal definitions.

Definition 1.2 Let $a_1, \ldots, a_h$ be relatively prime non-zero integers and let $\mathcal{L} = \mathcal{L}_{a_1,\ldots,a_h}$ denote the linear form $a_1x_1 + \cdots + a_hx_h$. A subset $A \subseteq \mathbb{Z}$ is said to be an *$\mathcal{L}$-basis* if the equation

$$a_1x_1 + \cdots + a_hx_h = n \tag{1.4}$$

has at least one solution for every $n \in \mathbb{Z}$.

Similarly, we say that A is an *asymptotic $\mathcal{L}$-basis* if (1.4) has a solution for all but finitely many n, and that A is an *$\mathcal{L}$-basis for almost all $\mathbb{Z}$* if those n for which (1.4) has no solution form a set of asymptotic density zero.

Remark 1.3 Recall that a subset $S \subseteq \mathbb{Z}$ is said to have asymptotic density zero if

$$\lim_{n \to +\infty} \frac{|S \cap [-n, n]|}{2n + 1} = 0. \tag{1.5}$$

To generalize the notion of unordered representation function to arbitrary linear forms requires a bit more care. The definition we give below is, we think, the natural one. First we need some terminology. A solution $(x_1, \dots, x_h)$ of (1.4) is said to be a *representation* of n by the form $\mathcal{L} = \mathcal{L}_{a_1,\dots,a_h}$. We say that two representations $(x_1, \dots, x_h)$ and $(y_1, \dots, y_h)$ of the same integer n are *equivalent* if, for every $\xi \in \mathbb{Z}$,

$$\sum_{x_i = \xi} a_i = \sum_{y_i = \xi} a_i. \tag{1.6}$$

For example, for the form $\mathcal{L}_{1,-1}$, any two representations (x, x) and (y, y) of zero are equivalent. As another example, for the form $\mathcal{L}_{2,-3,5}$, the representations (x, y, y) and (y, x, x) of $2x + 2y$ are equivalent.

We now define the *(unordered) $\mathcal{L}$-representation function $f_{A,\mathcal{L}}$* of a subset $A \subseteq \mathbb{Z}$ as

$$f_{A,\mathcal{L}}(n) = \#\{\text{equivalence classes of representations of } n \text{ by } \mathcal{L}\}. \tag{1.7}$$

There are a few existing results on the inverse problem for bases for general linear forms. Indeed in [8] the problem was already raised in the much more difficult setting of $\mathbb{N}_0$, in which case one may assume that the coefficients a_i in (1.4) are positive. No results were proven in that paper, and none of the specific problems the authors posed have, to the best of our knowledge, been settled since. They do make the intriguing observation, though, that for some forms one can construct a unique representation basis for $\mathbb{N}_0$, for example the form $x_1 + ax_2$ for any $a > 1$. Further examples are given in Theorem 3 of [6]. It would be fascinating to have a full classification of the forms for which this is possible. The one result of note we are aware of is Vu's extension [10] of the Erdős-Tetali result on thin bases to general linear forms.

In the setting of $\mathbb{Z}$, one is first and foremost interested in generalizing Theorem 1.1. There are some recent results of Nathanson [6] on binary forms, and in [7] he poses some problems for general forms. Our results answer some of his questions and supersede those in [5].

It should be noted here that in all the papers referenced above, only the ordered representation function is considered, meaning that one distinguishes between equivalent representations of the same number. For Vu's result, this distinction is not important (since his is a Θ-result), but the results we shall prove here have a much more elegant formulation when one works with unordered representations.

We close this section by briefly summarizing the results to follow. In Section 2 we prove that for any primitive form $\mathcal{L}$ there exists a unique representation basis. This generalizes the main result of [4] and answers Problem 16 of [7]. Our method is founded on observation **II** on page 2 and is thus basically the same as that employed in these earlier papers. However, we believe our presentation is much more streamlined, especially when specialized to the forms $x_1 + \cdots + x_h$.

In Section 3 we seek a generalization of Theorem 1.1. We introduce the notion of an *automorphism* of a linear form and show that a form has no non-trivial automorphisms (we will say what 'non-trivial' means) if and only if it is *partition regular* in the sense of Rado, i.e.: no non-empty subset of the coefficients sums to zero. Our main result in this section is that, if $\mathcal{L}$ is partition regular then, for any $f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\}$ such that the set $f^{-1}(0)$ has density zero, there exists $A \subseteq \mathbb{Z}$ for which $f_{A,\mathcal{L}} = f$. Since any form all of whose coefficients have the same sign is partition regular, this result generalizes Theorem 1.1. But it also extends that theorem, since we only require $f^{-1}(0)$ to have density zero, and not necessarily be finite. It thus answers Problem 13 and partly resolves Problem 17 in [7].

Irregular forms seem to be harder to deal with. The simplest such form is $\mathcal{L}_{1,-1} : x_1 - x_2$. In Section 4 we study this form but our results are weaker than those in Section 3. Open problems remain and these are discussed in Section 5.

Finally, note that the methods of proof in Sections 3 and 4 are in essence no different from those in Section 2. The main point here is in identifying the 'right' theorems, but once this is done no really new ideas are needed to carry out the proofs.

2. Unique Representation Bases

Before stating and proving the main result of this section, we introduce some more notation and terminology similar to that in (1.6) above. Let $\mathcal{L} : a_1 x_1 + \cdots + a_h x_h$ be a linear form. Let m, p be positive integers, $(r_1, \ldots, r_m)$ any m-tuple and $(s_1, \ldots, s_p)$ any p-tuple of integers, and

$$\pi_1 : \{1, \ldots, m\} \to \{1, \ldots, h\}, \quad \pi_2 : \{1, \ldots, p\} \to \{1, \ldots, h\}, \tag{2.1}$$

any functions. We say that the sums $\sum_{i=1}^{m} a_{\pi_1(i)} r_i$ and $\sum_{i=1}^{p} a_{\pi_2(i)} s_i$ are *equivalent w.r.t.* $\mathcal{L}$, and write

$$\sum_{i=1}^{m} a_{\pi_1(i)} r_i \equiv_{\mathcal{L}} \sum_{i=1}^{p} a_{\pi_2(i)} s_i \tag{2.2}$$

if, for every $\xi \in \mathbb{Z}$,

$$\sum_{r_i=\xi} a_{\pi_1(i)} = \sum_{s_i=\xi} a_{\pi_2(i)}. \tag{2.3}$$

Note that this generalizes the notion of equivalent representations in Section 1 since the latter corresponds to the special case $m = p = h$, $\pi_1 = \pi_2 = \mathrm{id}$.

The remainder of this section is devoted to the proof of the following result.

Theorem 2.1 *(i) Let $\mathcal{L}$ be a linear form. Then there exists a unique representation $\mathcal{L}$-basis if and only if $\mathcal{L}$ is primitive.*
(ii) Let $\mathcal{L} = \mathcal{L}_{a_1,\dots,a_h}$. The following are equivalent :
(a) $\mathcal{L}$ is primitive and not all a_i have the same sign,
(b) for every $n \in \mathbb{Z}$, there exists a unique representation $\mathcal{L}$-basis $A(n) \subseteq [n, +\infty)$.

Proof. We concentrate on proving part (i) : the proof of part (ii) will then be an immediate consequence of our approach. Clearly there can be no $\mathcal{L}$-basis if $\mathcal{L}$ is imprimitive, so suppose $\mathcal{L}$ is primitive. The result is trivial if $\mathcal{L} = \mathcal{L}_{\pm 1}$, so we may assume that $\mathcal{L}$ is a function of at least two variables. We find it convenient to use the slightly unusual notation $\mathcal{L} = \mathcal{L}_{a_1,\dots,a_{h+1}}$, where $h \geq 1$. Henceforth, we deal with a fixed form $\mathcal{L}$, so h and the coefficients a_i are fixed. Our task is to construct a unique representation basis A for $\mathcal{L}$. Let d_0 be any non-zero integer and put $A_0 := \{d_0\}$. We will construct the set A step-by-step as

$$A = \bigsqcup_{k=0}^{\infty} A_k, \tag{2.4}$$

where, for each $k > 0$, the set A_k will consist of $h+1$ suitably chosen integers, which we denote as

$$A_k = \{d_{k,1}, \dots, d_{k,h}, e_k\}. \tag{2.5}$$

We adopt the following ordering of the integers :

$$0, 1, -1, 2, -2, 3, -3, \dots, \tag{2.6}$$

and denote the ordering by $\mathcal{O}$. For each $k > 0$ the elements of A_k will be chosen so that

(I) A_k represents the least integer t_k in the ordering $\mathcal{O}$ not already represented by $B_{k-1} := \sqcup_{j=0}^{k-1} A_j$,
(II) no integer is represented more than once by B_k.

Since the set $B_0 = A_0$ clearly already satisfies property **(II)**, it is clear that if both **(I)** and **(II)** are satisfied for every $k > 0$, then the set A given by (2.5) will be a unique representation basis.

Since $\mathcal{L}$ is primitive, it represents 1. Fix a choice $(s_1, \dots, s_{h+1})$ of a representation of 1.

Let M be a fixed, very large positive real number (how large M needs to be will become clear in what follows).

Fix $k > 0$. Suppose $A_0, \ldots, A_{k-1}$ have already been chosen in order to satisfy **(I)** and **(II)**. Let t_k be the least integer in the ordering $\mathcal{O}$ not represented by B_{k-1}. First choose any h positive numbers $\delta_{k,1}, \ldots, \delta_{k,h}$ such that

$$\frac{\delta_{k,1}}{d_{k-1,h}} > M, \quad \frac{\delta_{k,i+1}}{\delta_{k,i}} > M, \quad \text{for } i = 1, \ldots, h-1, \tag{2.7}$$

and put

$$\epsilon_k := -\left\lfloor \frac{1}{a_{h+1}} \left(\sum_{i=1}^{h} a_i \delta_{k,i} \right) \right\rfloor. \tag{2.8}$$

Let

$$\sum_{i=1}^{h} a_i \delta_{k,i} + a_{h+1} \epsilon_k := u_k \in [0, a_{h+1}], \tag{2.9}$$

and choose the elements of A_k as

$$(d_{k,1}, \ldots, d_{k,h}, e_k) := (\delta_{k,1}, \ldots, \delta_{k,h}, \epsilon_k) + (t_k - u_k) \cdot (s_1, \ldots, s_{h+1}). \tag{2.10}$$

Our choice immediately guarantees that **(I)** is satisfied. The remainder of the proof is concerned with showing that **(II)** still holds provided the integer M is sufficiently large. This is done by establishing the following two claims :

CLAIM 1 : *Let $(x_1, \ldots, x_{h+1}), (y_1, \ldots, y_{h+1})$ be any two $(h+1)$-tuples of integers in B_k. Then exactly one of the following holds :*

(i)

$$\sum_{x_i \in A_k} a_i x_i \equiv_{\mathcal{L}} \sum_{y_i \in A_k} a_i y_i, \tag{2.11}$$

(ii) the difference

$$\sum_{i=1}^{h+1} a_i x_i - \sum_{i=1}^{h+1} a_i y_i \tag{2.12}$$

is much larger in absolute value than any integer represented by B_{k-1},

(iii)

$$\sum_{x_i \in A_k} a_i x_i - \sum_{y_i \in A_k} a_i y_i \equiv_{\mathcal{L}} \pm \left(\sum_{i=1}^{h} a_i d_{k,i} + a_{h+1} e_k \right). \tag{2.13}$$

CLAIM 2 : *Suppose (iii) holds in Claim 1. Then*

$$\sum_{y_i \notin A_k} a_i y_i - \sum_{x_i \notin A_k} a_i x_i \equiv_{\mathcal{L}} \pm \sum_{i=1}^{h+1} a_i z_i, \tag{2.14}$$

for some $(h+1)$-tuple $(z_1, \dots, z_{h+1})$ of integers in B_{k-1}.

Indeed, suppose that $(x_1, \dots, x_{h+1})$ and $(y_1, \dots, y_{h+1})$ are any two $(h+1)$-tuples in B_k such that

$$\sum_{i=1}^{n+1} a_i x_i = \sum_{i=1}^{n+1} a_i y_i = T, \text{ say.} \tag{2.15}$$

Then either (i) or (iii) in Claim 1 holds. But if (iii) holds then Claim 2 gives the contradiction that the integer t_k is already represented by B_{k-1}. Suppose (i) holds. Let z be any element of B_{k-1} and put

$$x_i' := \begin{cases} z, & \text{if } x_i \in A_k, \\ x_i, & \text{if } x_i \in B_{k-1}, \end{cases} \qquad y_i' := \begin{cases} z, & \text{if } y_i \in A_k, \\ y_i, & \text{if } y_i \in B_{k-1}. \end{cases} \tag{2.16}$$

Then, since B_{k-1} represents every integer at most once, we must have that $(x_1', \dots, x_{h+1}')$ and $(y_1', \dots, y_{h+1}')$ are equivalent representations of T. But then $(x_1, \dots, x_{h+1})$ and $(y_1, \dots, y_{h+1})$ are also equivalent representations of T, so B_k satisfies **(II)** in this case also.

Proof of CLAIM 1 : To simplify notation, put

$$w_i := d_{k,i} \text{ for } i = 1, \dots, h; \quad w_{h+1} := e_k. \tag{2.17}$$

Consider the difference

$$\sum_{x_i \in A_k} a_i x_i - \sum_{y_i \in A_k} a_i y_i := \sum_{i=1}^{h+1} c_i w_i, \tag{2.18}$$

where

$$c_i := \sum_{x_u = w_i} a_u - \sum_{y_v = w_i} a_v. \tag{2.19}$$

Alternative (i) trivially holds if all $c_i = 0$, so so we may assume that some $c_i \neq 0$.

First suppose $c_{h+1} = 0$ and let $j \in [1, h]$ be the largest index for which $c_j \neq 0$. Then, for $M \gg 0$, it is clear that the left-hand side of (2.18) is $\Theta(d_{k,j})$ and hence alternative (ii) holds.

So finally we may suppose that $c_{h+1} \neq 0$. Let $f \in \mathbb{Q}$ be such that $c_{h+1} = f \cdot a_{h+1}$. Then

$$c_{h+1} w_{h+1} = -f a_h w_h + \Psi_h, \tag{2.20}$$

where, for $M \gg 0$, the 'error term' Ψ_h must be much smaller in absolute value than the 'leading term' $-f a_h w_h$. Thus alternative (ii) will hold unless $c_h = f a_h$. But then

$$c_h w_h + c_{h+1} w_{h+1} = -f a_{h-1} w_{h-1} + \Psi_{h-1}, \tag{2.21}$$

where, once again, for $M \gg 0$, the term Ψ_{h-1} must be much smaller in absolute value than $f a_{h-1} w_{h-1}$. Hence alternative (ii) holds unless $c_{h-1} = f a_{h-1}$ and, by iteration of the same argument, unless $c_i = fa_i$ for $i = 1, \ldots, h$. In that case we thus have that

$$\sum_{x_i \in A_k} a_i x_i - \sum_{y_i \in A_k} a_i y_i \equiv_{\mathcal{L}} f\left(\sum_{i=1}^{n+1} a_i w_i\right). \tag{2.22}$$

But $f \cdot a_i \in \mathbb{Z}$ for $i = 1, \ldots, h+1$ and since $\mathcal{L}$ is primitive, this implies that $f \in \mathbb{Z}$. But then it is clear that we must have $|f| = 1$ and hence that alternative (iii) holds. $\diamond$

Proof of CLAIM 2 : Without loss of generality we may assume that

$$\sum_{x_i \in A_k} a_i x_i - \sum_{y_i \in A_k} a_i y_i \equiv_{\mathcal{L}} \sum_{i=1}^{h+1} a_i w_i, \tag{2.23}$$

and now need to construct an $(h+1)$-tuple $(z_1, \ldots, z_{h+1})$ of integers in B_{k-1} such that

$$\sum_{y_i \in B_{k-1}} a_i y_i - \sum_{x_i \in B_{k-1}} a_i x_i \equiv_{\mathcal{L}} \sum_{i=1}^{h+1} a_i z_i. \tag{2.24}$$

Let $i_1 < i_2 < \cdots < i_m$ be the indices for which $x_i \in B_{k-1}$. We shall decompose the index set $\{1, \ldots, h+1\}$ as the disjoint union of $m+1$ subsets $S_1, \ldots, S_{m+1}$ defined as follows :

Fix l with $1 \leq l \leq m$. Set $S_{l,0} := \{i_l\}$. For each $j > 0$ set

$$S_{l,j} := \{i : x_i \in \{w_k\}_{k \in S_{l,j-1}}\}. \tag{2.25}$$

Noting that the sets $S_{l,j}$ are pairwise disjoint for different j and hence empty for all $j \gg 0$, we set

$$S_l := \bigsqcup_j S_{l,j}. \tag{2.26}$$

It is also easy to see that the sets $S_1, \ldots, S_m$ are pairwise disjoint. We define

$$S_{m+1} := \{1, \ldots, h+1\} \backslash \bigsqcup_{l=1}^{m} S_l. \tag{2.27}$$

Note further that the sets $W_1, \ldots, W_{m+1}$ are pairwise disjoint, where

$$W_l := \{w_i : w_i = x_j \text{ or } w_j \text{ for some } j \in S_l\}, \quad l = 1, \ldots, m+1, \tag{2.28}$$

and that

$$A_k = \{w_1, \ldots, w_{h+1}\} = \bigsqcup_{l=1}^{m+1} W_l. \tag{2.29}$$

Let z be any element of B_{k-1}. We are now ready to define the $(h+1)$-tuple $(z_1, \ldots, z_{h+1})$. Let $1 \leq i \leq h+1$. We put

$$z_i := \begin{cases} y_i, & \text{if } y_i \in B_{k-1}, \\ z, & \text{if } y_i \in W_{m+1}, \\ x_{i_l}, & \text{if } y_i \in W_l,\ 1 \leq l \leq m. \end{cases} \tag{2.30}$$

Then (2.24) now follows from (2.23), since the latter implies that

$$\sum_{i \in S_{m+1}} a_i w_i \equiv_{\mathcal{L}} \sum_{i \in S_{m+1}} a_i x_i, \tag{2.31}$$

and, for $1 \leq l \leq m$, that

$$\sum_{y_i \in W_l} a_i y_i \equiv_{\mathcal{L}} \sum_{i \in S_l} a_i x_i - \sum_{i \in S_l} a_i w_i - a_{i_l} x_{i_l}. \tag{2.32}$$

$\diamond$

Since Claim 2 has been proved, the first part of Theorem 2.1 has also been proved. For the second part, it follows immediately from the definitions (2.7), (2.8) and (2.10) that, if the coefficients a_i don't all have the same sign, then the elements of A in (2.4) can all be chosen to lie in any given half-line. So we are done. $\square$

Remark 2.2 If we were instead to work with ordered representations, then it is an immediate corollary of Theorem 2.1 that there exists a unique representation basis for the form $a_1 x_1 + \cdots + a_h x_h$ if and only if there do not exist two distinct subsets I, I' of $\{1, \ldots, h\}$ such that

$$\sum_{i \in I} a_i = \sum_{i \in I'} a_i. \tag{2.33}$$

This resolves Problem 16 in [7].

Remark 2.3 The proof of Theorem 2.1 simplifies considerably if the form $\mathcal{L}$ is partition regular, thus in particular in the case of the forms $x_1 + \cdots + x_h$. See Remark 3.5 below for an explanation. This is why we think our presentation streamlines those in earlier papers.

3. Partition Regular Forms

We wish to use the method of the previous section in order to generalize Theorem 1.1. As in [7], we adopt the following notations :

$$\mathcal{F}_0(\mathbb{Z}) := \{f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\} : f^{-1}(0) \text{ is finite}\}, \tag{3.1}$$

$$\mathcal{F}_\infty(\mathbb{Z}) := \{f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\} : f^{-1}(0) \text{ has asymptotic density zero}\}. \tag{3.2}$$

Theorem 1.1 is a statement about $\mathcal{F}_0(\mathbb{Z})$. Only minor modifications to the method of Section 2 will be required, both to obtain a similar result for general linear forms, and to extend the result to $\mathcal{F}_\infty(\mathbb{Z})$. We prepare the ground for this with a couple of lemmas. First some terminology :

Definition 3.1 An *automorphism* of the linear form $\mathcal{L} : a_1 x_1 + \cdots + a_h x_h$ is a pair of functions (ψ, χ) from the set $\{x_1, \ldots, x_h\}$ of variables to $\{x_1, \ldots, x_h\} \cup \{0\}$ such that the linear form

$$\sum_{i=1}^{h} a_i \psi(x_i) - \sum_{i=1}^{h} a_i \chi(x_i) \tag{3.3}$$

is the same form as $\mathcal{L}$. The automorphism is said to be *trivial* if $\chi \equiv 0$.

For example, a non-trivial automorphism of the form $\mathcal{L}_{1,-1} : x_1 - x_2$ is given by

$$\psi(x_1) = x_1, \quad \psi(x_2) = 0, \qquad \chi(x_1) = x_2, \quad \chi(x_2) = 0. \tag{3.4}$$

Our first lemma is for the purpose of generalizing Theorem 1.1 to other linear forms :

Lemma 3.2 *A linear form $\mathcal{L}$ is partition regular if and only if it possesses no non-trivial automorphisms.*

Proof. Denote $\mathcal{L} : a_1 x_1 + \cdots + a_h x_h$ as usual. First suppose $\mathcal{L}$ is partition regular and thus, without loss of generality, that $h \geq 2$ and $a_1 + \cdots + a_r = 0$ for some $2 \leq r \leq h$. Set

$$\psi(x_1) = 0, \quad \psi(x_i) = x_i, \; i = 2, \ldots, h, \tag{3.5}$$

and

$$\chi(x_1) = 0, \quad \chi(x_i) = x_1, \; i = 2, \ldots, r, \quad \chi(x_i) = 0, \; i = r+1, \ldots, h. \tag{3.6}$$

Then one easily verifies that (ψ, χ) is a non-trivial automorphism of $\mathcal{L}$.

Conversely, let (ψ, χ) be a non-trivial automorphism of $\mathcal{L}$. For each $i = 1, \ldots, h$, (3.3) yields an equation between coeffcients of the form

$$a_i = \sum_{j \in X_i} a_j - \sum_{j \in Y_i} a_j, \tag{3.7}$$

where X_i and Y_i are subsets of $\{1, \ldots, h\}$. The definition of automorphism means that every index $j \in \{1, \ldots, h\}$ occurs in at most one of the X_i and at most one of the Y_i. Non-triviality means that there is at least one i such that $(X_i, Y_i) \neq (\{i\}, \phi)$. Without loss of generality, suppose that $(X_i, Y_i) \neq (\{i\}, \phi)$ for $i = 1, \ldots, r$ only, and some $r \leq h$. Adding together the left and right hand sides of (3.7) for $i = 1, \ldots, r$ yields an equation of the form

$$a_1 + \cdots + a_r = \sum_{j \in X} a_j - \sum_{j \in Y} a_j, \tag{3.8}$$

for some disjoint subsets X and Y of $\{1, \ldots, h\}$, with $X \subseteq \{1, \ldots, r\}$. From (3.8) we can extract a non-empty subset of the coefficients summing to zero, except if $X = \{1, \ldots, r\}$ and $Y = \phi$. But it is easily seen that the latter is impossible when $\chi \not\equiv 0$.

The next lemma is for the purpose of extending our results to $\mathcal{F}_\infty(\mathbb{Z})$:

Lemma 3.3 *Let $S \subseteq \mathbb{Z}$ such that $d(S) = 1$. Let l, m, p be any three positive integers. For each $n \in \mathbb{Z}$ set*

$$X_{l,m,p}(n) := \mathbb{Z} \cap \{\frac{a}{b} n + c : (a,b,c) \in \mathbb{Z}^3 \cap ([-l,l] \times \pm[1,m] \times [-p,p])\}, \tag{3.9}$$

and set

$$S_{l,m,p} := \{n : X_{l,m,p}(n) \subseteq S\}. \tag{3.10}$$

Then $d(S_{l,m,p}) = 1$.

Proof. The proof follows immediately from the following two facts :

(i) the intersection of finitely many sets of asymptotic density one has the same property

(ii) since S has density one, the same is true, for any fixed integers a, b, c, with $b \neq 0$, of the set

$$\{n \in \mathbb{Z} : \frac{a}{b}n + c \in S \cup (\mathbb{Q}\backslash\mathbb{Z})\} \tag{3.11}$$

We are now ready to state the main result of this section :

Theorem 3.4 *Let* $\mathcal{L} : a_1x_1 + \cdots + a_hx_h$ *be any partition regular linear form. Then for any* $f \in \mathcal{F}_\infty(\mathbb{Z})$*, there exists a subset* $A \subseteq \mathbb{Z}$ *such that* $f_{A,\mathcal{L}} = f$.

Proof. Let a partition regular $\mathcal{L} : a_1x_1 + \cdots + a_hx_h$ and $f \in \mathcal{F}_\infty(\mathbb{Z})$ be given : we shall show how to construct $A \subseteq \mathbb{Z}$ with $f_{A,\mathcal{L}} = f$. Let $\mathcal{M}$ be the multisubset of $\mathbb{Z}$ consisting of $f(n)$ repititions of n for every n. The problem amounts to constructing a 'unique representation basis' for $\mathcal{M}$. This is some countable set : let $\mathcal{O} = \{\tau_1, \tau_2, \dots\}$ be any well-ordering of it. We now construct A step-by-step as in (2.4)-(2.10). This time the ordering $\mathcal{O}$ is as just defined above. We'll have $t_k = \tau_{k'}$ for some k' depending on k. Two requirements must be satisfied when we tag on the numbers $d_{k,1}, \dots, d_{k,h}, e_k$ to our set A :

(I) No new representation is created of any number appearing before t_k in the ordering $\mathcal{O}$.

(II) No representation is created of any integer n for which $f(n) = 0$.

To satisfy these requirements, the integer denoted M in (2.7) will now have to depend on k. The first difficulty arises because, when considering **(I)**, since the ordering $\mathcal{O}$ is chosen randomly, we have no control over how quickly the sizes of numbers in this ordering grow as ordinary integers. Clearly, $M = M_k$ can be chosen large enough to take account of this difficulty in the sense that Claim 1 holds as before. We still need to rule out case (iii) of that claim occurring, and it is here that we make use of the assumption that $\mathcal{L}$ is partition regular, for (2.13) describes an automorphism of $\mathcal{L}$, just as long as the numbers $d_{k,1}, \dots, d_{k,h}, e_k$ are distinct.

The only remaining problem is thus **(II)**. Let $S := \mathbb{Z}\backslash f^{-1}(0)$. By assumption $d(S) = 1$. Then it follows from Lemma 3.3 that there exists a choice of a sufficiently large $M = M_k$ which will mean that **(II)** is indeed satisfied. Indeed once $d_{k,1}, \dots, d_{k,h-1}$ have been chosen with due regard to **I**, one just needs to choose $d_{k,h}$ to also lie in a set $S_{l,m,p}$, where l, m, p are fixed integers depending a priori on all of the numbers $d_{k,1}, \dots, d_{k,h-1}, t_k, a_1, \dots, a_{h+1}, s_1, \dots, s_{h+1}$.

Thus there is indeed a choice of M_k that works at each step. □

Remark 3.5 The proof of Theorem 2.1 simplifies for partition regular forms in the same way as in the argument just presented. Namely, we can ignore Case (iii) of Claim 1, and thus don't need the most technical part of the proof, which is the proof of Claim 2.

4. The Form $x_1 - x_2$

We do not know if there exist any partition irregular forms for which Theorem 3.4 still holds. For the simplest such form, namely $x_1 - x_2$, this is clearly not the case. Henceforth we denote this form by $\mathcal{D}$. Specializing our terminology from Section 1 to the form $\mathcal{D}$, the (unordered) representation function $f_{A,\mathcal{D}}$ of a non-empty subset A of $\mathbb{Z}$ is given by

$$f(0) = 1, \tag{4.1}$$

$$f(n) = \{(a_1, a_2) \in A^2 : a_1 - a_2 = n\}, \quad \text{if } n \neq 0. \tag{4.2}$$

Note that, in particular, the unordered and ordered representation functions coincide for this form, except at $n = 0$, where all representations $a - a = 0$ are considered equivalent : see (1.6) and the examples after it. Relation (4.1) imposes an immediate restriction on the functions representable by $\mathcal{D}$. Also, it is clear from (4.2) that any representable function must be even. There is a more serious obstruction, however. Let $f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\}$ and suppose $f(n) \geq 3$ for some n. Suppose $f_{A,\mathcal{D}} = f$ for some $A \subseteq \mathbb{Z}$ and let $a_1, \ldots, a_6 \in A$ be such that

$$a_1 - a_2 = a_3 - a_4 = a_5 - a_6 = n, \tag{4.3}$$

are three pairwise non-equivalent representations of n (i.e.: the numbers a_1, a_3, a_5 are distinct). Then we also have the equalities

$$a_1 - a_3 = a_2 - a_4, \quad a_3 - a_5 = a_4 - a_6, \quad a_1 - a_5 = a_2 - a_6, \tag{4.4}$$

and at least one of these three differences must be different from n. Thus there exists some other number m for which $f(m) \geq 2$.

The following definition captures this kind of condition imposed on a function f representable by $\mathcal{D}$:

Definition 4.1 Let $f : \mathbb{Z} \to \mathbb{N}_0 \cup \{\infty\}$. A sequence (finite or infinite) $s_1, s_2, s_3, \ldots$ of positive integers is said to be *plentiful* for f if, for every pair $l \leq m$ of positive integers, we have

$$f\left(\sum_{i=l}^{m} s_i\right) > 1. \tag{4.5}$$

The main result of this section is the following.

Theorem 4.2 *Let $f \in \mathcal{F}_0(\mathbb{Z})$ be even with $f(0) = 1$.*

(i) If $f^{-1}(\infty) \neq \phi$ then f is representable by $\mathcal{D}$ if and only if there exists an infinite plentiful sequence for f.

(ii) If $f^{-1}(\infty) = \phi$ but f is unbounded, then f is representable by $\mathcal{D}$ if and only if there exist arbitrarily long plentiful sequences for f.

We do not know whether this result can be extended to functions in $\mathcal{F}_\infty(\mathbb{Z})$, nor exactly which bounded functions in $\mathcal{F}_0(\mathbb{Z})$ can be represented by $\mathcal{D}$.

Proof. Throughout this proof, since we are working with a fixed form $\mathcal{D}$, we will write simply f_A for the representation function of a subset A of $\mathbb{Z}$.

We begin with the proof of part (i) and then outline the changes needed to prove part (ii). First suppose f is representable by $\mathcal{D}$ and let $A \subseteq \mathbb{Z}$ be such that $f_A = f$. Suppose $f(n) = \infty$. Let $(x_i, y_i)_{i=1}^{\infty}$ be a sequence of pairs of elements of A such that $x_i - y_i = n$ for each i and such that the sequence (x_i) is either strictly increasing or strictly decreasing. Let

$$s_i := \begin{cases} x_{i+1} - x_i, & \text{if } (x_i) \text{ increasing,} \\ x_i - x_{i+1}, & \text{if } (x_i) \text{ decreasing.} \end{cases} \tag{4.6}$$

Let $1 \leq l \leq m$. Then

$$x_{m+1} - y_{m+1} = x_l - y_l \Rightarrow x_{m+1} - x_l = y_{m+1} - y_l = \pm \sum_{i=l}^{m} s_i, \tag{4.7}$$

and hence

$$f_A\left(\sum_{i=l}^{m} s_j\right) \geq 2. \tag{4.8}$$

Thus the sequence (s_i) is plentiful for f.

Conversely, suppose there exists an infinite plentiful sequence $(s_i)_{i=1}^{\infty}$ for f. We will construct a set A which represents f. Set

$$S := \left\{ \sigma_{l,m} := \sum_{i=l}^{m} s_i \;\middle|\; 1 \leq l \leq m < \infty \right\}. \tag{4.9}$$

Set

$$\mathcal{I} := \{n : f(n) > 1\}, \quad \mathcal{J} := \{n : f(n) = 1\}, \tag{4.10}$$

and note that $S \subseteq \mathcal{I}$. Let $\mathcal{M}$ be the multisubset of $\mathbb{Z}$ consisting of $f(n)$ copies of n for each n, and $\mathcal{O}$ any well-ordering of $\mathcal{M}$. Put $A_0 := \{d_0\}$ for any choice of a non-zero integer d_0. The set A will be constructed step-by-step as in (2.4). At each step $k > 0$ the set A_k will consist of two suitably chosen integers x_k and y_k. As before we set $B_k := \sqcup_{j=0}^{k} A_k$ and denote by t_k the least number in the ordering $\mathcal{O}$ not yet represented by B_{k-1}. We also set

$$\mathcal{U}_k := \{n : f_{B_k}(n) > f_{B_{k-1}}(n)\}, \tag{4.11}$$

Our choices will be made so as to ensure that the following two requirements are satisfied for every k :

(I) $t_k \in \mathcal{U}_k$ and, moreover, for any $n \in \mathcal{U}_k$ it is the case that

$$f_{B_{k-1}}(n) < f(n), \tag{4.12}$$

$$f_{B_k}(n) \le f_{B_{k-1}}(n) + 2, \tag{4.13}$$

and

$$\text{if } f_{B_k}(n) = f_{B_{k-1}}(n) + 2 \text{ then } f_{B_{k-1}}(n) = 0 \text{ and } n \in S. \tag{4.14}$$

(II) Suppose $n \in \mathcal{I}$. Let $p := f_{B_k}(n)$ and

$$a_1 - b_1 = \cdots = a_p - b_p = n, \tag{4.15}$$

be the different representations of n in B_k, where $a_1 < a_2 < \cdots < a_p$. Then there exist integers $0 < m_1 < m_2 < \cdots < m_p$ such that

$$a_{i+1} - a_i = \sigma_{m_i, m_{i+1}}, \qquad i = 1, \ldots, p-1. \tag{4.16}$$

It is clear that if **(I)** and **(II)** are satisfied for every $k \ge 0$, then the set A given by (2.4) represents f. The condition **(II)** will be useful in establishing (4.14). The elements of the different A_k are chosen inductively. Observe that **(I), (II)** are trivially satisfied for $k = 0$, so suppose $k > 0$ and that **(I), (II)** are satisfied for each $k' < k$. We now describe how the elements of A_k may be chosen. Let

$$M_k := \max\{|n| : n \in B_{k-1}\}. \tag{4.17}$$

and note that $f_{B_{k-1}}(n) = 0$ for all $n > 2M_k$.

Case I : $t_k \in \mathcal{J}$.

Then **(II)** will continue to hold no matter what we do. We now choose x_k to be any integer greater than $2|t_k| + 3M_k$, and choose $y_k := x_k - t_k$. This choice of x_k and y_k guarantees that, if $n \in \mathcal{U}_k$, then

(a) $f_{B_{k-1}}(n) = 0$ and

(b) $f_{B_k}(n) = 1$,

hence that **(I)** is satisfied. To verify (a), we observe that if $n \in \mathcal{U}_k$ then either $n = \pm t_k = \pm(x_k - y_k)$ or $|n| > 2M_k$. For (b), we note that if $a, b \in B_{k-1}$ and $x_k - a = y_k - b$, then $x_k - y_k = a - b$, contradicting the assumption that $t_k \in \mathcal{J}$.

Case II : $t_k \in \mathcal{I}$.

Let $p := f_{B_{k-1}}(t_k)$. If $p = 0$ then proceed as in Case I. Otherwise let $(a_i, b_i)_{i=1}^p$ be the different representations of t_k in B_{k-1} and let $m_1, \ldots, m_p$ be the integers for which (4.15) is satisfied (with $k-1$ instead of k). We choose

$$x_k := a_p + \sigma_{m_p, m_{p+1}}, \tag{4.18}$$

for some sufficiently large integer m_{p+1} such that $x_k > 2|t_k| + 3M_k$. Then we take $y_k := x_k - t_k$. Reasoning as in Case I, the size of x_k and y_k guarantee that (4.12) and (4.13) will be satisfied, and the relationships (4.16) and (4.18) will imply (4.14), and thus ensure that **(II)** still holds.

This completes the induction step, and hence the proof of part (i) of the theorem.

Now we briefly outline the proof of part (ii). That representability of f implies the existence of arbitrarily long plentiful sequences is shown in the same way as before. Suppose now such sequences exist. The construction of A such that $f_A = f$ proceeds as above and, with notation as before, the only difference is in the inductive choice of the elements in A_k for $k > 0$. Suppose we have already chosen A'_k for $k' < k$ so that

$$f_{B_{k-1}}(n) \leq f(n) \quad \forall\, n \in \mathbb{Z}. \tag{4.19}$$

Let t_k be the least integer in the ordering $\mathcal{O}$ for which $f_{B_{k-1}}(t_k) < f(t_k)$. Set $\gamma_k := f(t_k) - f_{B_{k-1}}(t_k)$. The set A_k will be the union of $2\gamma_k$ elements $x_{i,k}, y_{i,k}$, $i = 1, \ldots, \gamma_k$.

First, with M_k defined as in (4.17), we choose $x_{1,k}$ to be any integer greater than $2|t_k| + 3M_k$ and take $y_{1,k} := x_{1,k} - t_k$. Since there exist arbitrarily long plentiful sequences, we can find such a sequence

$$0 < a_1 < a_2 < \cdots < a_{\gamma_k - 1} \tag{4.20}$$

such that each of the quotients

$$\frac{a_1}{x_{1,k}}, \quad \frac{a_{i+1}}{a_i}, \quad i = 1, \ldots, \gamma_k - 1 \tag{4.21}$$

is arbitrarily large. We then wish to choose the remaining elements of A_k as

$$x_{i,k} := x_{i-1,k} + a_{i-1}, \qquad y_{i,k} := x_{i,k} - t_k, \qquad i = 2, \ldots, \gamma_k. \tag{4.22}$$

Provided the quotients in (4.21) are all sufficiently large, it is clear that, if $n \in \mathcal{U}_k$, then either

(a) $n = \pm t_k$ and $f_{B_k}(n) = f_{B_{k-1}}(n) + \gamma_k = f(n)$, or

(b) $f_{B_{k-1}}(n) = 0$ and either $f_{B_k}(n) = 1$, or $f_{B_k}(n) = 2$ and $n \in \mathcal{I}$.

Thus (a) and (b) ensure that (4.19) is also satisfied for this value of k, and thus the set A given by (2.4) will satisfy $f_A = f$.

Hence the proof of Theorem 4.2 is complete. □

5. Open Problems

We only mention what are probably the two most glaring issues left unresolved by the investigations above.

1. Which functions $f \in \mathcal{F}_\infty(\mathbb{Z})$ are representable by a partition irregular form ? Does there exist such a form which represents any such function ? For the form $\mathcal{D}$, we want to know if Theorem 4.2 can be extended to $\mathcal{F}_\infty(\mathbb{Z})$, and which bounded functions are representable.

2. The methods employed in this paper to construct sets with given representation functions have, in common with previous similar methods, the obvious weakness that they produce very sparse sets. It is an important unsolved problem to find the maximal possible density of a set with a given representation function : this problem is still unsolved for every possible f and $\mathcal{L}$, though the most natural case to look at is $f \equiv 1$. It should be investigated to what extent existing optimal constructions for the forms $x_1 + \cdots + x_h$ can be extended to general forms.

Acknowledgement I thank Boris Bukh for a very helpful discussion. I also thank the referee for a very careful reading of the paper.

References

[1] G. A. Dirac, *Note on a problem in additive number theory*, J. London Math. Soc. **26** (1951), 312–313.

[2] P. Erdős and P. Tetali, *Representations of integers as the sum of k terms*, Random Structures Algorithms **1** (1990), 245–261.

[3] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215.

[4] M. B. Nathanson, *Unique representation bases for the integers*, Acta Arith. **108** (2003), no. 1, 1–8.

[5] ——, *Every function is the representation function of an additive basis for the integers*, Port. Math. (N.S.) **62** (2005), no. 1, 55–72.

[6] ——, *Representation functions of bases for binary linear forms*, Funct. Approx. Comment. Math. **37** (2007), 341–350.

[7] ——, *Inverse problems for representation functions in additive number theory,* `http://arxiv.org/abs/0712.0408`

[8] A. Sárkőzy and V. T. Sós, *On additive representation functions, in : The Mathematics of Paul Erdős, I,* 129–150, Algorithms Combin. **13**, Springer, Berlin, 1997.

[9] R. C. Vaughan, *The Hardy-Littlewood method (second edition)*, Cambridge Tracts in Mathematics, no. 125, Cambridge University Press, 1997.

[10] V. H. Vu, *On a refinement of Waring's problem*, Duke Math J. **105** (2000), 107–134.