Discrete Comput Geom 19:343–353 (1998)

Discrete & Computational
Geometry

Rectification Principles in Additive Number Theory*

Y. F. Bilu,[1] V. F. Lev,[2] and I. Z. Ruzsa[3]

[1]Forschungsinstitut für Mathematik, ETH-Zentrum, CH-8092 Zurich, Switzerland
yuri@math.ethz.ch

[2]Department of Mathematics, University of Georgia, Athens, GA 30602, USA
seva@math.uga.edu

[3]Mathematical Institute of the Hungarian Academy of Sciences,
Pf. 127, H-1364 Budapest, Hungary
ruzsa@math-inst.hu

Communicated by János Pach

Abstract. We consider two general principles which allow us to reduce certain additive problems for residue classes modulo a prime to the corresponding problems for integers.

1. Introduction

It is well known that additive problems in finite abelian groups are generally more difficult than analogous problems in $\mathbb{Z}$. For instance, consider the following classical problem: given an abelian group $\mathcal{G}$, describe all pairs of finite sets $K, L \subseteq \mathcal{G}$ such that

$$|K + L| < |K| + |L|.$$

When $\mathcal{G} = \mathbb{Z}$ (or a torsion-free abelian group) the answer is almost trivial: K and L must be arithmetic progressions with the same difference. When $\mathcal{G}$ is a cyclic group of prime order, the answer is given by Vosper's theorem [24], which is quite nontrivial. And when $\mathcal{G}$ is an arbitrary finite abelian group, we should turn to the extremely complicated recursive classification of Kemperman [14]. (A few years ago the problem was solved for torsion-free nonabelian groups [6], [13].)

* Research of the first author was supported by the SFB 170 "Geometrie und Analysis" at Göttingen. Research of the third author was supported by Hungarian National Foundation for Scientific Research, Grant No. 17433.

Nevertheless, more than 30 years ago, Freiman [10, Sect. 3.13] discovered that, at least for cyclic groups of prime order, certain additive problems can be reduced to corresponding problems in $\mathbb{Z}$, provided the sets in consideration are "not very large." Seemingly, this important observation did not receive much attention.

On the other hand, it was recently observed in [15] that there is another reduction method. It is weaker than Freiman's method in that it requires the sets to be "very small" instead of "not very large." However, unlike the method of Freiman it imposes no additional restrictions on the sets and handles easily the case of distinct summands, which makes it applicable when Freiman's method fails.

The objective of this paper is to apply Freiman's discovery and the reduction method mentioned above to concrete additive problems.

2. Freiman's Rectification Principle

For simplicity, we consider only the case of equal summands: $K = L$. Using [22, Lemma 3.3] we can extend the results to distinct summands.

We need the concept of F_s-isomorphism [10], [22]. Let $\mathcal{G}, \mathcal{H}$ be abelian groups, and consider subsets $K \subset \mathcal{G}$ and $L \subset \mathcal{H}$. The bijection $\varphi\colon K \to L$ is *Freiman's isomorphism of order s* or, shortly, *F_s-isomorphism*, if for any $a_1, \dots, a_{2s} \in K$,

$$a_1 + \cdots + a_s = a_{s+1} + \cdots + a_{2s}$$

if and only if

$$\varphi(a_1) + \cdots + \varphi(a_s) = \varphi(a_{s+1}) + \cdots + \varphi(a_{2s}).$$

It is easily seen that, if K and L are F_{s+1}-isomorphic, then they are also F_s-isomorphic. Clearly, two sets K and L are F_1-isomorphic if and only if $|K| = |L|$.

Theorem 2.1 (Freiman's Rectification Principle). *For any positive numbers $\sigma \in \mathbb{R}$ and $s \in \mathbb{Z}$ there exists a positive constant $c_1 = c_1(\sigma, s)$ such that the following holds*:

Let p be a prime number and let $K \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfy $|K| \le c_1 p$ and

$$|K + K| < \sigma |K|.$$

Then there exists a set of integers $K' \subset \mathbb{Z}$ such that the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ induces an F_s-isomorphism of K' onto K.

To put it briefly, this theorem asserts that if a set of residues has a small sumset and is not too large itself, then it is F_s-isomorphic to a set of integers.

A proof of Theorem 2.1 (for $s = 2$) is briefly sketched in [10, Sect. 3.12]. Our proof given below is simpler than the original, but requires substantially no new ideas.

The argument is based on the following result of Freiman:

Theorem 2.2 (Freiman). *Let σ be a positive real number and let K be a finite set of integers satisfying*

$$|K + K| < \sigma |K|.$$

Then there exist positive integers $r, b_1, \dots, b_r$ and nonzero integers $g_0, g_1, \dots, g_r$ such that

$$r \le c_2(\sigma), \qquad b_1 \cdots b_r \le c_3(\sigma)|K|,$$

and

$$\begin{aligned} K \subseteq \Pi &= \Pi(g_0; g_1, \dots, g_r; b_1, \dots, b_r) \\ &:= \{g_0 + \beta_1 g_1 + \cdots + \beta_r g_r : \beta_i = 0, \dots, b_i - 1\}, \end{aligned}$$

where $c_2(\sigma)$ and $c_3(\sigma)$ are positive constants, depending only on σ.

There are two different proofs of this theorem. The first is Freiman's original, see [10], [11], and [3]. The second is due to Ruzsa, see [23] and [20].

Proof of Theorem 2.1. We shall see that the theorem holds with

$$c_1 = (2sc_2(2\sigma))^{-c_2(2\sigma)}(c_3(2\sigma))^{-1}.$$

Let $K_0 \subseteq \{0, 1, \dots, p-1\}$ be the preimage of K under the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. Then $|K_0 + K_0| \le 2|K + K| < 2\sigma|K_0|$, whence by Theorem 2.2 we have $K_0 \subseteq \Pi(g_0; g_1, \dots, g_r; b_1, \dots, b_r)$, where $r, b_1, \dots, b_r$ are positive integers satisfying

$$r \le c_2(2\sigma), \qquad b_1 \cdots b_r \le c_3(2\sigma)|K|,$$

and $g_1, \dots, g_r \in \mathbb{Z}$.

Put $\varepsilon_i = (2srb_i)^{-1}$. Then

$$\begin{aligned} p\varepsilon_1 \cdots \varepsilon_r &= p(2sr)^{-r}(b_1 \cdots b_r)^{-1} \\ &\ge p(2sc_2(2\sigma))^{-c_2(2\sigma)}(c_3(2\sigma))^{-1}|K|^{-1} = c_1 p|K|^{-1} \ge 1. \end{aligned}$$

Hence by Minkowski's theorem on linear inequalities [7, App. B, Theorem III], there exists a nonzero vector $(a, a_1, \dots, a_r) \in \mathbb{Z}^{r+1}$ such that

$$\begin{aligned} |a| &< p, \\ \left| \frac{ag_i}{p} - a_i \right| &\le \varepsilon_i \qquad (1 \le i \le r). \end{aligned} \tag{2.1}$$

We have $a \ne 0$, since otherwise it would follow from (2.1) that $a_1 = \cdots = a_r = 0$. Thus, the integer a satisfies

$$\begin{aligned} a &\not\equiv 0 \pmod{p}, \\ \left\| \frac{ag_i}{p} \right\| &\le \varepsilon_i \qquad (1 \le i \le r), \end{aligned}$$

where $\|\cdot\|$ stands for the distance from the nearest integer. This gives

$$\left\| \frac{a(x-g_0)}{p} \right\| \le \sum_{i=1}^{r} b_i \left\| \frac{ag_i}{p} \right\| \le (2s)^{-1}$$

for any $x \in K_0$.

Let

$$K_1 \subseteq \left\{ -\left\lfloor \frac{p}{(2s)} \right\rfloor, -\left\lfloor \frac{p}{(2s)} \right\rfloor + 1, \dots, \left\lfloor \frac{p}{(2s)} \right\rfloor \right\}$$

be the set of integers congruent modulo p to one of the numbers $a(x-g_0)$ (for some $x \in K_0$). Clearly, K_1 is mapped onto K by $x \mapsto ux + g_0 \bmod p$, where u is an arbitrary integer satisfying $au \equiv 1 \pmod p$. It follows that K_1 is F_s-isomorphic to K, for any algebraic sum of $2s$ elements of K_1 which is 0 modulo p is also 0 in $\mathbb{Z}$.

Finally, we define $K' = \{ux + g_0 : x \in K_1\}$. □

Freiman's rectification principle allows us to reduce various additive problems in $\mathbb{Z}/p\mathbb{Z}$ to corresponding problems in $\mathbb{Z}$. Unfortunately, we have to make the restrictive assumption $|K| \le c_1 p$. Restricting to "not very large" sets is the price we have to pay for the use of such a powerful tool as Freiman's Theorem 2.2. For instance, the main results of [4] and [5] also assume that the sets in question are small enough.

It would be nice to find a proof of the rectification principle independent of Theorem 2.2. For $\sigma < 2.4$ and $s = 2$ such a proof is implicit in [10, Sect. 2.3], where the following result is obtained.

Theorem 2.3 [10]. *Let $K \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfy $|K+K| < 2.4|K|$, and suppose, in addition, that $|K| < p/35$. Then K can be covered by a "short" arithmetic progression modulo p: There exist $g_0, g_1 \in \mathbb{Z}/p\mathbb{Z}$ and a positive integer $b \le |K+K| - |K| + 1$ such that*

$$K \subseteq \{g_0 + g_1\beta : \beta = 0, \dots, b-1\}. \tag{2.2}$$

Conjecturally, the assertion of this theorem is true for $|K+K| \le \max\{p-1, 3|K|-4\}$. Using Theorem 2.1, we can easily prove it for sufficiently small K.

Indeed, suppose that $|K| < c_1 p$, where $c_1 = c_1(3,2)$ is the constant of Theorem 2.1. Let $K' \subset \mathbb{Z}$ be the set of integers which is F_2-isomorphic to K and is mapped onto K by the canonical mapping $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. Then K' also satisfies $|K'+K'| \le 3|K'| - 4$ (since $|K'| = |K|$ and $|K'+K'| = |K+K|$ by the definition of an F_s-isomorphism). By another and well-known result of Freiman [10, Theorem 1.9] (which has a relatively easy elementary proof; see also [17]), there exist $a, d \in \mathbb{Z}$ and a positive integer $b \le |K'+K'| - |K'| + 1$ such that

$$K' \subseteq \{a + d\beta : \beta = 0, \dots, b-1\}.$$

Now (2.2) is clearly satisfied if g_0 and g_1 are the elements of $\mathbb{Z}/p\mathbb{Z}$ congruent to a and d, respectively.

This is a first illustrative example which shows how rectification methods can be used to reduce difficult additive problems in $\mathbb{Z}/p\mathbb{Z}$ to easier problems in $\mathbb{Z}$.

3. Direct Rectification

Theorem 2.1 shows that any (not too large) set of residues $K \subseteq \mathbb{Z}/p\mathbb{Z}$ with a small sumset is isomorphic to a set of integers. It turns out that this is true for *any* K, regardless of its sumset, provided that $|K|$ is *very* small. Specifically, it is shown in [15] that any $K \subseteq \mathbb{Z}/p\mathbb{Z}$ of the cardinality $k = |K|$ is contained in an arithmetic progression modulo p of at most

$$2k^{-1/(k-1)}p^{1-1/(k-1)} + 1 \tag{3.1}$$

terms. If $k \le \log_4 p + \log_4 \log_4 p$ (where $\log_4$ is the logarithm base 4), then the number (3.1) is less than $p/2 + 1$, whence K is F_2-isomorphic to a set of integers. Similarly, if $k \le \log_{2s} p + \log_{2s} \log_{2s} p$, then K is F_s-isomorphic to a set of integers. Essentially the same can be obtained by a direct application of the idea we used in the proof of Theorem 2.1.

Theorem 3.1. *Let $K \subseteq \mathbb{Z}/p\mathbb{Z}$, where p is a prime. If $|K| \le \log_{2s} p$, then there exists a set of integers $K' \subset \mathbb{Z}$ such that the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ induces an F_s-isomorphism of K' onto K.*

Proof. Let $K = \{g_1, \dots, g_r\}$, and put $\varepsilon_1 = \cdots = \varepsilon_r = (2s)^{-1}$. Since $r \le \log_{2s} p$, we have $p\varepsilon_1 \cdots \varepsilon_r \ge 1$. Therefore, applying Minkowski's theorem exactly in the same way as in the proof of Theorem 2.1, we find $a \in \mathbb{Z}$ satisfying

$$\begin{aligned} a &\not\equiv 0 \pmod p, \\ \left\| \frac{ag_i}{p} \right\| &\le (2s)^{-1} \qquad (1 \le i \le r). \end{aligned}$$

Now let $m_i \in \{-\lfloor p/(2s)\rfloor, -\lfloor p/(2s)\rfloor + 1, \dots, \lfloor p/(2s)\rfloor\}$ be defined from $m_i \equiv ag_i \pmod p$, and put $K' = \{um_1, \dots, um_k\}$ where u is any integer, inverse to a in $\mathbb{Z}/p\mathbb{Z}$. Then the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ maps K' onto K, and this mapping is an F_s-isomorphism, which follows immediately from $|m_i| < p/(2s)$ (as in the proof of Theorem 2.1). □

This theorem is nearly best possible: here is an example which shows that there exists a set $K \subseteq \mathbb{Z}/p\mathbb{Z}$ of cardinality $|K| \le 2\log_2 p + 1$ which is not F_2-isomorphic to any set of integers. (This example can easily be generalized to produce a set of cardinality at most $2\log_s p + 1$ which is not F_s-isomorphic to any set of integers.)

Put $N = \lfloor \log_2 p \rfloor$, and write $p = 2^{d_1} + 2^{d_2} + \cdots + 2^{d_t}$, where $0 \le d_1 < d_2 < \cdots < d_t$, $t \le N + 1$. We define

$$K = \{0\} \cup \{1, 2, 4, \dots, 2^N\} \cup \{2^{d_1} + 2^{d_2}, 2^{d_1} + 2^{d_2} + 2^{d_3}, \dots, 2^{d_1} + 2^{d_2} + \cdots + 2^{d_{t-1}}\}$$

(all the numbers are considered as residues modulo p), so that $|K| \le 2N + 1$. We assume that K is F_2-isomorphic to a set of integers

$$K' = \{0\} \cup \{a_0, a_1, a_2, \dots, a_N\} \cup \{b_2, b_3, \dots, b_{t-1}\}$$

and obtain a contradiction. Let $\varphi\colon K \to K'$ be the isomorphism. As the notation suggests, we suppose (which does not restrict the generality) that $\varphi(0) = 0$ and that a_i, b_i are the images in K' of the corresponding elements of K. Then for any $i \in [0, N-1]$, the equality in $\mathbb{Z}/p\mathbb{Z}$

$$0 + 2^{i+1} = 2^i + 2^i$$

implies

$$0 + a_{i+1} = a_i + a_i,$$

which yields subsequently

$$a_1 = 2a_0, \quad a_2 = 4a_0, \ldots, \quad a_N = 2^N a_0.$$

Next, from

$$0 + (2^{d_1} + 2^{d_2}) = 2^{d_1} + 2^{d_2}$$

we obtain

$$b_2 = (2^{d_1} + 2^{d_2})a_0,$$

and then from

$$0 + (2^{d_1} + \cdots + 2^{d_i} + 2^{d_{i+1}}) = (2^{d_1} + \cdots + 2^{d_i}) + 2^{d_{i+1}}$$

for $i = 2, \ldots, t-2$ we obtain

$$b_{i+1} = b_i + a_{d_{i+1}},$$

which yields

$$b_3 = (2^{d_1} + 2^{d_2} + 2^{d_3})a_0, \ldots, \quad b_{t-1} = (2^{d_1} + 2^{d_2} + \cdots + 2^{d_{t-1}})a_0.$$

But this is a contradiction in view of

$$\begin{aligned} b_{t-1} + a_{d_t} = a_0 p &\neq 0 + 0 \quad \text{in } \mathbb{Z}, \\ (2^{d_1} + 2^{d_2} + \cdots + 2^{d_{t-1}}) + 2^{d_t} &= 0 + 0 \quad \text{in } \mathbb{Z}/p\mathbb{Z}. \end{aligned}$$

To show how Theorem 3.1 can be applied in the case of distinct summands, consider the following problem: Given s sets $K_1, \ldots, K_s$ in an abelian group $\mathcal{G}$, how many representations of the form

$$x = a_1 + \cdots + a_s; \ a_i \in K_i \qquad (i = 1, \ldots, s) \tag{3.2}$$

can an element $x \in \mathcal{G}$ have? We assume here that the cardinalities $|K_i|$ are preassigned. For $\mathcal{G} = \mathbb{Z}$, we have the following result:

Theorem 3.2 [16, Theorem 1]. *Let $K_1, \ldots, K_s \subseteq \mathbb{Z}$ be a finite sets of integers. Then the number of solutions of* (3.2) *is maximized, when $x = 0$, and K_i are the sets of consecutive integers*

$$K_i = \{\alpha_i, \alpha_i + 1, \ldots, \gamma_i\} \qquad (i = 1, \ldots, s),$$

where integers α_i, γ_i are chosen in such a way that

$$\gamma_i - \alpha_i + 1 = |K_i|, \qquad |\alpha_i + \gamma_i| \le 1 \qquad (i = 1, \dots, s),$$

$$|(\alpha_1 + \gamma_1) + \cdots + (\alpha_s + \gamma_s)| \le 1.$$

Using direct rectification we can easily transfer Theorem 3.2 to small subsets of $\mathbb{Z}/p\mathbb{Z}$.

Theorem 3.3. *Let $K_1, \dots, K_s \subseteq \mathbb{Z}/p\mathbb{Z}$ be sets of residues modulo a prime p, and assume that $|K_1| + \cdots + |K_s| \le \log_{2s} p$. Then the number of solutions of* (3.2) *is maximized, when $x = 0$, and K_i are the sets of consecutive residues*

$$K_i = \{\alpha_i, \alpha_i + 1, \dots, \gamma_i\} \pmod p \qquad (i = 1, \dots, s),$$

where the integers α_i, γ_i are chosen in such a way that

$$\gamma_i - \alpha_i + 1 = |K_i|, \qquad |\alpha_i + \gamma_i| \le 1 \qquad (i = 1, \dots, s), \tag{3.3}$$

$$|(\alpha_1 + \gamma_1) + \cdots + (\alpha_s + \gamma_s)| \le 1. \tag{3.4}$$

Proof. For any abelian group $\mathcal{G}$ and for any $L_1, \dots, L_s \subseteq \mathcal{G}$, $x \in \mathcal{G}$, denote by $N_x(L_1, \dots, L_s)$ the number of solutions of

$$a_1 + \cdots + a_s = x; \qquad a_i \in L_i,$$

and let

$$N(L_1, \dots, L_s) = \max_{x \in \mathcal{G}} N_x(L_1, \dots, L_s).$$

Define $K = K_1 \cup \cdots \cup K_s$. Let $\varphi\colon K' \to K$ be an F_s-isomorphism of a set of integers $K' \subseteq \mathbb{Z}$ onto K, and let K_i' be the preimage of K_i in K' $(i = 1, \dots, s)$. Then evidently an equality

$$a_1' + \cdots + a_s' = a_{s+1}' + \cdots + a_{2s}'$$

with $a_i', a_{s+i}' \in K_i'$ holds if and only if

$$\varphi(a_1') + \cdots + \varphi(a_s') = \varphi(a_{s+1}') + \cdots + \varphi(a_{2s}')$$

holds, and it follows that

$$N(K_1, \dots, K_s) = N(K_1', \dots, K_s').$$

By Theorem 3.2, the right-hand side can only increase if, for all $i = 1, \dots, s$, we replace K_i' by $K_i'' = \{\alpha_i, \alpha_i + 1, \dots, \gamma_i\} \subseteq \mathbb{Z}$, where α_i, γ_i satisfy (3.3) and (3.4):

$$N(K_1', \dots, K_s') \le N(K_1'', \dots, K_s'').$$

Now, let $\overline{K}_i$ be the images of K_i'' under the canonical homomorphism $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. The assertion of the theorem follows from the observation that for any integer $x \in \{-\lfloor p/2 \rfloor, -\lfloor p/2 \rfloor + 1, \dots, \lfloor p/2 \rfloor\}$ with the corresponding residue $\overline{x} \in \mathbb{Z}/p\mathbb{Z}$,

$$N_x(K_1'', \dots, K_s'') = N_{\overline{x}}(\overline{K}_1, \dots, \overline{K}_s),$$

and thus

$$N(K_1'', \dots, K_s'') = N(\overline{K}_1, \dots, \overline{K}_s). \qquad \square$$

Notice, that in the proof above we could not apply Theorem 2.1 not only because the sets K_i are distinct, but also (and mainly) because there are no restrictions on the cardinalities $|2K_i|$. However, Theorem 3.1 works perfectly in this situation.

4. Erdős–Heilbronn Conjecture

Let $h \geq 2$ be an integer and let K be a subset of the set of elements of an abelian group. Denote by $h\widehat{\ }K$ the set of all sums of h distinct elements from K:

$$h\widehat{\ }K = \{a_1 + \cdots + a_h : a_1, \dots, a_h \in K \text{ and } a_i \neq a_j \text{ for } 1 \leq i < j \leq h\}.$$

Let p be a prime. Erdős and Heilbronn (see [9, p. 95]) conjectured that any $K \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfies

$$|2\widehat{\ }K| \geq \min\{2|K| - 3, p\}. \tag{4.1}$$

Note that the inequality $|2\widehat{\ }K| \geq 2|K| - 3$ trivially holds for $K \subseteq \mathbb{Z}$ (and for finite subsets of torsion-free abelian groups). In general, we have

Proposition 4.1 (Folklore).

(a) *For any positive integer h and any finite set $K \subseteq \mathbb{Z}$ we have*

$$|h\widehat{\ }K| \geq h|K| - h^2 + 1. \tag{4.2}$$

(b) *If $|K| \geq \max\{h+2, 5\}$, then equality in* (4.2) *holds if and only if K is an arithmetic progression.*

For a proof see [19, Theorems 1 and 2], [5, App. C], or [20, Theorem 1.10].

After a number of partial results, say, [12], [18], and [21] (see [2] for more references), the Erdős–Heilbronn conjecture (4.1) was finally settled by Dias da Silva and Hamidoune [8]. Another proof was suggested in [1]. Actually, Dias da Silva and Hamidoune proved a more general inequality

$$|h\widehat{\ }K| \geq \min\{h|K| - h^2 + 1, p\} \tag{4.3}$$

for arbitrary $h \geq 2$ and $K \subseteq \mathbb{Z}/p\mathbb{Z}$.

Recently Alon et al. [2] obtained a fairly general additive theorem which contains inequality (4.3) as a particular case.

However, to the best of our knowledge, the problem of when the *equality* in (4.3) holds is still open. Here we obtain an answer for sufficiently small $K \subseteq \mathbb{Z}/p\mathbb{Z}$ as a direct consequence of Proposition 4.1(b) and Theorem 2.1.

Theorem 4.1. *For any $h \geq 2$ there exists a constant $c_4 = c_4(h)$ with the following property. For any prime p and any set of residues $K \subseteq \mathbb{Z}/p\mathbb{Z}$ such that*

$$\max\{h+2, 5\} \leq |K| \leq c_4 p, \tag{4.4}$$

the equality

$$|h\widehat{\ }K| = h|K| - h^2 + 1 \tag{4.5}$$

holds if and only if K is an arithmetic progression.

Proof. Put $c_4(h) = \min\{h^{-1}, c_1(2h, h)\}$, where c_1 is defined in Theorem 2.1. If $K \subseteq \mathbb{Z}/p\mathbb{Z}$ is an arithmetic progression and $|K| \leq p/h$, then (4.5) obviously holds.

Conversely, assume that $K \subseteq \mathbb{Z}/p\mathbb{Z}$ satisfies (4.4) and (4.5). Fix an $(h-2)$-element subset $H \subseteq K$ and denote $L = K \setminus H$. Then $|2\widehat{\ }L| \leq |h\widehat{\ }K| < h|K|$. Therefore

$$|K+K| \leq |2\widehat{\ }K| + |K| \leq |2\widehat{\ }L| + |H+K| + |K| < h|K| + (h-2)|K| + |K| < 2h|K|.$$

By Theorem 2.1, the set K is F_h-isomorphic to a set of integers $K' \subseteq \mathbb{Z}$. Then clearly $|h\widehat{\ }K'| = |h\widehat{\ }K| = h|K'| - h^2 + 1$, and by Proposition 4.1(b), K' is an arithmetic progression. Then so is K. □

Freiman et al. [12] applied a similar "rectification" approach for $h = 2$. Their technique is quite different and is not based on Theorem 2.2, and for $h = 2$ their result is much stronger than Theorem 4.1 above. However, the method of [12] does not extend to $h \geq 3$. See also Rödseth [21].

When the Erdős–Heilbronn conjecture was proved, it had been conjectured by the second author that in fact a much more general result holds. Specifically, let K and L be subsets of an abelian group, such that $|K| \leq |L|$, and let $\tau: K \to L$ be an arbitrary mapping from K to L. Define $K \overset{\tau}{+} L$ to be the set of all the sums $a + b$ (where $a \in K,\ b \in L$) such that $b \neq \tau(a)$:

$$K \overset{\tau}{+} L = \{a + b: a \in K,\ b \in L,\ \text{and } b \neq \tau(a)\}.$$

Conjecture 4.1 (Lev). *Let K and L be subsets of $\mathbb{Z}/p\mathbb{Z}$ satisfying $|K| \leq |L|$, and let $\tau: K \to L$ be an arbitrary mapping from K to L. Then*

$$|K \overset{\tau}{+} L| \geq \min\{|K| + |L| - 3, p\}.$$

Using Theorems 2.1 or 3.1 we are able to prove this for small K, L. First, we need a corresponding result in $\mathbb{Z}$.

Theorem 4.2. *Let K and L be finite subsets of $\mathbb{Z}$ satisfying $|K| \le |L|$, and let $\tau\colon K \to L$ be an arbitrary mapping from K to L. Then*

$$|K \overset{\tau}{+} L| \ge |K| + |L| - 3.$$

Proof. Write down the elements of K and L in ascending order: $K = \{a_1, \dots, a_k\}$ and $L = \{b_1, \dots, b_l\}$, where $a_i < a_j$ and $b_i < b_j$ for $i < j$.

We first assume that $|K| < |L|$. Then there exists $b_j \in L$ which is not an image of an element of K under τ. Therefore among the $k + l - 1$ distinct sums

$$\begin{aligned} a_1 + b_1 < a_1 + b_2 < \cdots < a_1 + b_{j-1} \\ < a_1 + b_j < a_2 + b_j < \cdots < a_k + b_j \\ < a_k + b_{j+1} < a_k + b_{j+2} < \cdots < a_k + b_l, \end{aligned}$$

at most one sum in the first row and at most one sum in the last row are excluded by the condition $b \ne \tau(a)$. At least $k + l - 3$ remaining sums fall into $K \overset{\tau}{+} L$.

Now assume $|K| = |L|$. Then either there exists $b_j \in L$ which has no preimage in K, and we can repeat the argument above; or τ is a bijection, in which case we consider $k + l - 1$ distinct sums

$$a_1 + b_1 < a_1 + b_2 < \cdots < a_1 + b_l < a_2 + b_l < a_3 + b_l < \cdots < a_k + b_l,$$

and observe again, that at most two of them may *not* fall into $K \overset{\tau}{+} L$. □

Theorem 4.3. *The assertion of Conjecture* 4.1 *holds provided that either $L = K$ and $|K| = |L| \le c_5 p$ (with a sufficiently small absolute constant c_5), or $|K| + |L| \le \log_4 p$.*

Proof. In the first case ($L = K$, $|K| = |L| \le c_5 p$) we observe that

$$|K + K| \le |K \overset{\tau}{+} K| + |K| \le 3|K| - 4,$$

assuming $|K \overset{\tau}{+} K| < 2|K| - 3$. Then by Theorem 2.1, K is F_2-isomorphic to a set of integers K'. Let $\tau'\colon K' \to K'$ be the mapping induced by τ. Then K' satisfies $|K' \overset{\tau'}{+} K'| < 2|K'| - 3$, which, as Theorem 4.2 shows, is impossible for $K' \subseteq \mathbb{Z}$.

In the second case ($|K| + |L| \le \log_4 p$), we find, as in Theorem 3.3, a set of integers $M \subseteq \mathbb{Z}$ which is F_2-isomorphic to the union $K \cup L$, define $K', L' \subseteq M$ to be the preimages of K, L, respectively, and define $\tau'\colon K' \to L'$ to be the mapping induced by τ. Then by Theorem 4.2,

$$|K \overset{\tau}{+} L| = |K' \overset{\tau'}{+} L'| \ge |K'| + |L'| - 3 = |K| + |L| - 3.$$ □

Using [22, Lemma 3.3] the last theorem can be extended to all sets K and L such that $\varepsilon|L| \le |K| \le |L| \le c_6(\varepsilon) p$ for any $\varepsilon > 0$.

As a concluding remark, we note that the rectification method can be used not only for the group $\mathbb{Z}/p\mathbb{Z}$: for instance, in [4] it is applied for the torus $\mathbb{R}^m/\mathbb{Z}^m$.

Acknowledgment

We would like to thank Dani Berend and the referee for pointing out some inaccuracies in the manuscript.

References

1. N. Alon, M. B. Nathanson, and I. Z. Ruzsa, Adding distinct congruence classes modulo a prime, *Amer. Math. Monthly*, **102** (1995), 250–255.
2. N. Alon, M. B. Nathanson, and I. Z. Ruzsa, The polynomial method and restricted sums of congruence classes, *J. Number Theory*, **56** (1996), 404–417.
3. Yu. Bilu, Structure of sets with small sumsets, Mathématiques Stochastiques, Univ. Bordeaux 2, Preprint 94-10, Bordeaux, 1994.
4. Yu. Bilu, The $(\alpha + 2\beta)$-inequality on the torus, *J. London Math. Soc*. To appear.
5. Yu. Bilu, Addition of sets of integers of positive density, *J. Number Theory*. To appear.
6. L. V. Brailovski and G. A. Freiman, On a product of finite subsets in a torsion-free group, *J. Algebra*, **130** (1990), 462–476.
7. J. W. S. Cassels, *An Introduction to Diophantine Approximations*, Cambridge Tracts in Mathematics and Mathematical Physics, vol. **45**, Cambridge University Press, Cambridge, 1965.
8. J. A. Dias da Silva and Y. O. Hamidoune, Cyclic spaces for Grassmann derivatives and additive theory, *Bull. London Math. Soc*., **26** (1994), 140–146.
9. P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, L'Enseignement Mathématique, Geneva, 1980.
10. G. A. Freiman, *Foundations of a Structural Theory of Set Addition* (Russian), Kazan', 1966; English translation: Translation of Mathematical Monographs, vol. 37, American Mathematical Society, Providence, RI, 1973.
11. G. A. Freiman, What is the structure of K if $K + K$ is small?, In: *Number Theory, New York* 1984–1985 (D. V. Chudnovsky et al., eds.), Lecture Notes in Mathematics, vol. 1240, Springer-Verlag, New York, 1987, pp. 109–134.
12. G. A. Freiman, L. Low, and J. Pitman, The proof of Paul Erdős' conjecture of the addition of different residue classes modulo a prime number, In: *Structure Theory of Set Addition*, preprint, Tel Aviv–Marseilles, 1992/93.
13. Y. O. Hamidoune, An isoperimetric method in additive theory, *J. Algebra*, **179** (1996), 622–630.
14. I. H. B. Kemperman, On small sumsets in an abelian group, *Acta Math*., **103** (1960), 63-88.
15. V. F. Lev, Simultaneous approximations and covering by arithmetic progressions in $\mathbb{F}_p$. In preparation.
16. V. F. Lev, On the number of solutions of a linear equation over finite sets. Submitted.
17. V. F. Lev and P. Yu. Smeliansky, On addition of two distinct sets of integers, *Acta Arith*., **LXX.1** (1995), 85–91.
18. R. Mansfield, How many slopes in a polygon? *Israel J. Math*., **39** (1981), 265–272.
19. M. B. Nathanson, Inverse theorems for subset sums, *Trans. Amer. Math. Soc*., **347** (1995), 1409–1418.
20. M. B. Nathanson, *Additive Number Theory*: 2. *Inverse Theorems and the Geometry of Sumsets*, Graduate Texts in Mathematics, vol. 165, Springer-Verlag, New York, 1996.
21. Ö. J. Rödseth, Sums of distinct residues mod *p*, *Acta Arith*., **LXV.2** (1993), 181–184.
22. I. Z. Ruzsa, Arithmetical progressions and the number of sums, *Period Math. Hungar.*, **25** (1) (1992), 105–111.
23. I. Z. Ruzsa, Generalized arithmetical progressions and sumsets, *Acta Math. Hungar.*, **65** (4) (1994), 379–388.
24. A. G. Vosper, The critical pairs of subsets of a group of prime order, *J. London Math. Soc*., **31** (1956), 200–205, 280–282.

Received January 21, 1997, *and in revised form April* 9, 1997.