SMALL DOUBLING AND ADDITIVE STRUCTURE
MODULO A PRIME

ØYSTEIN J. RØDETH

Yahya ould Hamidoune (1947–2011) in memoriam

Abstract. Let $\emptyset \neq A, B \subseteq \mathbb{Z}/p\mathbb{Z}$, where $p$ is a prime. The Cauchy-Davenport theorem gives a lower bound for the number of distinct sums $a + b$, where $a \in A$ and $b \in B$. The corresponding inverse theorem, due to Vosper, determines the structure of $A$ and $B$ if the lower bound is attained. A generalization of Vosper’s theorem was conjectured by Hamidoune, Serra, and Zémor in 2006. We prove that this conjecture is indeed correct. This is known to give a fairly good answer to the $3k - 3$ problem in $\mathbb{Z}/p\mathbb{Z}$.

1. Introduction

If nothing else is said, $A$ and $B$ will in this paper denote non-empty subsets of $\mathbb{Z}/p\mathbb{Z}$, where $p$ is a rational prime. The Minkowski sum of $A$ and $B$, or simply the sum-set $A + B$, is defined by

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

The Minkowski sum of more than two sets is defined in a similar way. In particular, we write $2A$ for the sumset $A + A$. We write $|A|$ for the cardinality of the set $A$, while the complement of $A$ in $\mathbb{Z}/p\mathbb{Z}$ is denoted by $\overline{A}$. We say that a subset of $A$ is covered by $A$.

The Cauchy-Davenport theorem gives a lower bound for the number of distinct residue classes in the Minkowski sum $A + B$.

**Theorem 1** (Cauchy-Davenport). If $A + B \neq \mathbb{Z}/p\mathbb{Z}$, we have

$$|A + B| \geq |A| + |B| - 1.$$  

This is a basic result in additive combinatorial number theory. The theorem was proven by Cauchy [4] in 1813 and rediscovered by Davenport [6, 7] in 1935.
In 1955 Freiman [10]–[13] introduced the term “inverse problem” in additive number theory, and proved some nice inverse theorems in $\mathbb{Z}$. Soon after, Vosper [28, 29] found and proved a substantial inverse theorem mod $p$. He determined the structure of the pairs $A, B$ for which the Cauchy-Davenport theorem is valid with equality.

If there exist $a, d \in \mathbb{Z}/p\mathbb{Z}$ such that

$$A = \{a + jd \mid j = 0, 1, \ldots, k - 1\},$$

then $A$ is an arithmetic progression with common difference $d$. We do not distinguish between positive and negative common differences. The number of distinct elements in $A$ is the length of the progression.

The diameter $\text{diam}(A)$ of $A$ is the length of the shortest arithmetic progression which covers $A$. The set $\{xa + y \mid a \in A\}$, where $x \neq 0$ and $y$ are residue classes mod $p$, is an affine image of $A$. Now, the diameter $\text{diam}(A)$ is the smallest positive integer $d$ such that the interval $[0, d-1]$ contains some affine image of $A$. Therefore $\text{diam}(A)$ is also called the affine diameter of $A$. The set $A$ can be covered by a short arithmetic progression if $\text{diam}(A) \leq |2A| - |A| + 1$.

To $A, B$ we make correspond the set

$$(1) \quad C = -(A + B),$$

and we define $r = r(A, B)$ by

$$(2) \quad r = |A + B| - |A| - |B| + 1.$$ Application of the Cauchy-Davenport theorem to (2) shows that $r \geq 0$ if $C \neq \emptyset$; that is, if $A + B \neq \mathbb{Z}/p\mathbb{Z}$. The following theorem is a variant of the conjecture of Hamidoune, Serra, Zémor [17].

**Theorem 2.** Let $A$ and $B$ be subsets of $\mathbb{Z}/p\mathbb{Z}$, and assume that

$$(3) \quad |A| \geq r + 3, \quad |B| \geq r + 3, \quad |C| \geq r + 2.$$ Then there are arithmetic progressions $\mathcal{A} \supseteq A$ and $\mathcal{B} \supseteq B$ covering $A$ and $B$, such that the length of $\mathcal{A}$ is $|\mathcal{A}| = |A| + r$, the length of $\mathcal{B}$ is $|\mathcal{B}| = |B| + r$, and the two arithmetic progressions $\mathcal{A}$ and $\mathcal{B}$ have the same common difference.

This theorem remains valid if we replace the conditions (3) by

$$|A| \geq r + 2, \quad |B| \geq r + 3, \quad |C| \geq r + 3;$$

that is, the conjecture of Hamidoune, Serra, and Zémor [17] is true.

The case $r = 0$ of Theorem 2 is essentially Vosper’s theorem [28, 29]. The case $r = 1$ is due to Hamidoune and Rodseth [16], while Hamidoune, Serra, and Zémor [17] worked their way through the case $r = 2$. 

2
We prefer to use $\mathbb{Z}/p\mathbb{Z}$ as a background for conveying our ideas, as the background noise is then rather moderate. But of course, the ideas can be used in more general settings. We could replace the modulus $p$ by an arbitrary positive integer, as in the Cauchy-Davenport-Chowla theorem \cite{5}; cf.\cite{25}. We could also consider $A$ and $B$ as subsets of a locally compact abelian group with a Haar measure; cf. \cite{19} and \cite{15}.

Let us take a brief look on the structure of this paper. In the next section we state our contribution to the $3k-3$ problem in $\mathbb{Z}/p\mathbb{Z}$. Then we show that the Dias da Silva-Hamidoune theorem, formerly the Erdős-Heilbronn conjecture, is a consequence of our $3k-3$ result in $\mathbb{Z}/p\mathbb{Z}$. In Section 4 we present the Davenport transform, which we use in the proof of Theorem 2. After a brief section on symmetry, we conclude with a few words about the $3k-3$ problem in $\mathbb{Z}/p\mathbb{Z}$.

2. The $3k-3$ Problem in $\mathbb{Z}/p\mathbb{Z}$

The following beautiful result, sometimes called Freiman’s 2.4-theorem, was published by Freiman some 50 years ago.

**Theorem 3** (Freiman). Let $A \subseteq \mathbb{Z}/p\mathbb{Z}$ and $k = |A|$. If $|2A| \leq 2.4k-3$ and if $k < p/35$, then $A$ can be covered by a short progression.

It was shown in \cite{24} that the condition $k < p/35$ can be replaced by the weaker $k < p/10.7$. By Theorem 2 it follows that we can use approximately $k < p/2.8$ instead.

As we understand it, Freiman proved a stronger result; namely that there exists an absolute constant $c$ such that if $|2A| < 3k-3$ and $k < p/c$, then $A$ can be covered by a short progression. A simpler proof was given by Bilu, Lev, and Ruzsa \cite{3}. Later, Green and Ruzsa \cite{14} showed that Freiman’s conjecture is true for $c = 10^{180}$; cf. \cite{17, 27}. By Theorem 2 we see that $c = 4$ suffices, or more precisely, the conjecture is true if the condition $k < p/c$ is replaced by $k < p/4 + 3/2$.

**Theorem 4.** Let $\emptyset \neq A \subseteq \mathbb{Z}/p\mathbb{Z}$. If $|2A| < 3k-3$ and $k < p/4 + 3/2$, then $A$ can be covered by a short arithmetic progression.

3. Restricted Minkowski Sums

Let us demonstrate the strength of Theorem 4 by deducing the Dias da Silva-Hamidoune theorem, formerly known as the Erdős-Heilbronn conjecture.

Let $A = \{a_0, a_1, \ldots, a_{k-1}\} \subseteq \mathbb{Z}/p\mathbb{Z}$ with $k = |A|$. Let $s$ denote the number of distinct residue classes of the form $a_i + a_j$ with $i \neq j$. Early in the 1960s Erdős and Heilbronn conjectured that

$$s \geq \min\{p, 2k-3\}.$$
Thirty years later, the truth of (4) was proven by Dias da Silva and Hamidoune [9], using multilinear algebra and representation theory; see also [8]. Soon after, Alon, Nathanson, and Ruzsa [1, 2] came up with another proof, where they introduced the simple and beautiful “polynomial method”. This method is also presented in [21].

To prove (4) we follow [22] and apply Theorem 4. Form the $k \times k$-matrix $M = (a_i + a_j)$. Let $t$ denote the number of distinct entries in $M$. Then $t = |2A|$. The number $s$ equals the number of distinct entries outside the main diagonal. In particular we have

$$s + k \geq t.$$  

If $t \geq 3k - 3$, then $s \geq 2k - 3$. Therefore we only have to consider the case $t < 3k - 3$. By Theorem 4 if $p \geq 4k - 5$, then $\text{diam}(A) \leq t - k + 1$. Renumbering the $a_i$ if necessary, each $a_i \in A$ has an integer representative $r_i$ such that $0 = r_0 < r_1 < \ldots < r_{k-1}$, where

$$r_{k-1} = \text{diam}(A) - 1 \leq t - k \leq 2k - 4.$$  

Then the $2k - 3$ integers

$$r_0 + r_1 < r_0 + r_2 < \ldots < r_0 + r_{k-1}$$

$$< r_1 + r_{k-1} < r_2 + r_{k-1} < \ldots < r_{k-2} + r_{k-1}$$

are distinct mod $p$, and we are finished.

4. The Davenport Transform

For the application of the Davenport transform, we use a technique which seems to go back to Vosper. This method was used by Yahya and myself to prove Theorem 5 in [16]. It was also employed in [23] to give a short proof of Vosper’s theorem. For a nice exposition of the Davenport transform, see Husbands [18].

Let $\emptyset \neq A, B \subseteq \mathbb{Z}/p\mathbb{Z}$ such that $A + B \neq \mathbb{Z}/p\mathbb{Z}$. Assume that $0 \in B$ and $|B| \geq 2$. We define

$$E = (A + 2B) \cap \overline{(A + B)}.$$  

Then we have

$$A + 2B = (A + B) \cup E,$$

where the union is disjoint. Since $B$ generates $\mathbb{Z}/p\mathbb{Z}$ additively, we have $E \neq \emptyset$.

For $e \in E$, we define

$$B_e = B \cap (e + C) \quad \text{and} \quad B^e = B \cap (e + \overline{C}).$$

We refer to $B_e$ as a Davenport transform of $B$. We have $0 \in B_e$ and $B_e \cup B^e = B$, $B_e \cap B^e$. 

4
Furthermore, for an \( e \in E \), there are \( a \in A \) and \( b, b' \in B \), such that \( e = a + b + b' \). Hence, \( e - (a + b) = b' \in B^e \). Thus we have \( B^e \neq \emptyset \); that is,

\[
(5) \quad 1 \leq |B_e| \leq |B| - 1.
\]

Moreover, we also have

\[
(6) \quad A + B \supseteq (A + B_e) \cup (e - B^e).
\]

Here, the union is disjoint since we (in self-explanatory notation) have that if \( a + b_e = e - b^e \), then \( e - b_e = a + b^e \in A + B \), so that \( b_e \in B_e \cap B^e = \emptyset \), a contradiction. Hence, by (6),

\[
(7) \quad |A + B| - |B| \geq |A + B_e| - |B_e|.
\]

Using (5) and (7), the Cauchy-Davenport theorem follows easily by induction on \( |B| \). This was Davenport’s goal. Let us add a few extra lines.

Let us assume that \( B_e = \{0\} \) for all \( e \in E \). Then \( B^e = B^x \) for all \( e \in E \), where \( B^x = B \setminus \{0\} \). By (5), we have

\[
A + B \supseteq A \cup (E - B^x),
\]

where the union is disjoint. The Cauchy-Davenport theorem gives us

\[
r + 1 \geq |\{0\} + 1| = A + 2B| - |A + B|.
\]

Let us collect these results in a lemma.

**Lemma 1.** If \( B_e = \{0\} \) for all \( e \in E \), then

\[
|B| \leq r + 2 \quad \text{if} \quad A + 2B \neq \mathbb{Z}/p\mathbb{Z}.
\]

\[
|C| \leq r + 1 \quad \text{if} \quad A + 2B = \mathbb{Z}/p\mathbb{Z}.
\]

5. **Vosper’s Theorem**

Since we now have the necessary machinery lined up, it is not much work to prove Vosper’s inverse theorem. But let us first write down a simple, but useful, lemma; cf. [26, p. 205].

**Lemma 2.** The subset \( A \) of \( \mathbb{Z}/p\mathbb{Z} \) with \( |A| \geq 2 \), is an arithmetic progression with common difference \( d \neq 0 \), if

\[
|\{0, d\} + A| \leq 1 + |A|.
\]

**Proof.** An affine transformation shows that it is no restriction to set \( d = 1 \). The result is clear if \( |A| \geq p - 1 \). Suppose that \( |A| \leq p - 2 \). Consider the residue classes \( 0, 1, \ldots, p - 1 \) mod \( p \) as consecutive and equidistant points on the circle. Then we have exactly one element \( a \in A \) with \( a + 1 \notin A \). Hence, the elements of \( A \) form a set of consecutive points
on the circle; that is, $A$ is an arithmetic progression with common difference 1. \hfill \Box

We now show Vosper’s theorem.

**Theorem 5** (Vosper). Let $A$ and $B$ be subsets of $\mathbb{Z}/p\mathbb{Z}$ satisfying $|B| \geq 2$, $|C| \geq 2$. If $r(A, B) = 0$, then $A$ is an arithmetic progression.

**Proof.** We set $r = 0$, and assume that $0 \in B$. For $|B| = 2$, we have $|A + B| = |A| + 1$,

and by Lemma [2], $A$ is an arithmetic progression.

Assume that the result is false for some $B$ with $|B| \geq 2$ minimal. Then $|B| \geq 3$. By the minimality of $|B|$, we have $B_e = \{0\}$. Hence $B_e = \{0\}$ for any $e \in E$. Then we also have $B_e = B^e$. This holds for all $e \in E$.

Moreover, we have

$$A + B \supseteq A \cup (e - B^e) \text{ for any } e \in E,$$

where the union is disjoint. Thus we have

$$A + B \supseteq A \cup (E - B^e),$$

so that

$$|A| + |B| - 1 = |A + B| \geq |A| + |B| - 2 + |E|,$$

and we get $|E| = 1$. In combination with the assumptions $|B| \geq 2$ and $|C| \geq 2$, this gives $|B| = 2$. Now, Lemma [2] shows that $A$ is an arithmetic progression. \hfill \Box

6. **Proof of Theorem 2**

In this section we prove an auxiliary result, and show that this result has Theorem [2] as an easy consequence.

**Theorem 6.** If $|B| \geq r + 3$ and $|C| \geq r + 2$, then

$$\operatorname{diam}(A) \leq |A| + r,$$

where $r = r(A, B)$ is given by (2).

**Proof.** By Vosper’s theorem we have that Theorem [6] holds for $r = 0$. Assume that Theorem [6] is false. Consider the least $r$ for which there is a pair $A, B$ with $0 \in B$ and $|B| \geq 2$, such that

$$(8) \quad \operatorname{diam}(A) \geq |A| + r(A, B) + 1.$$ 

Choose such a pair, where $|B|$ is minimal. By (7) and (8), we get

$$\operatorname{diam}(A) \geq |A| + r(A, B_e) + 1.$$ 

6
Assume that $|B_e| \geq 2$. By the minimality of $r(A, B)$, we then have $r(A, B) = r(A, B_e)$. By the minimality of $|B|$, we have a contradiction. Thus we have $B_e = \{0\}$ for all $e \in E$, and the theorem follows by Lemma 1.

We now prove Theorem 2. By Theorem 6, we only have to show that the two arithmetic progressions $\mathcal{A}$ and $\mathcal{B}$ have the same common difference. We can assume that

$$\mathcal{A} = \{0, 1, 2, \ldots, |A| + r - 1\},$$
$$\mathcal{B} = \{0, b, 2b, \ldots, (|B| + r - 1)b\}.$$

For any integer representative $b$ in the interval $2 \leq b \leq p - 2$, we clearly have $|\mathcal{A} + \{0, b\}| \geq 2 + |\mathcal{A}|$. Hence, by Lemma 2 the common difference of $\mathcal{B}$ is 1. This means that the two arithmetic progressions $\mathcal{A}$ and $\mathcal{B}$ have the same common difference. This concludes the proof of Theorem 2.

7. Symmetry

Let $A, B, C$ be subsets of $\mathbb{Z}/p\mathbb{Z}$ satisfying $A + B + C = (\mathbb{Z}/p\mathbb{Z})^\times$, where $(\mathbb{Z}/p\mathbb{Z})^\times$ denotes the set of non-zero residue classes mod $p$. Now, two of the sets $A, B, C$ determine uniquely the third, as long as the Minkowski sum of the two sets is not equal to the whole of $\mathbb{Z}/p\mathbb{Z}$.

This is easy to see: Let $A$ and $B$ be given. Put

$$C = -(A + B) + X,$$

where $X$ is some unknown non-empty set of residue classes mod $p$. Then

$$(\mathbb{Z}/p\mathbb{Z})^\times = A + B + C + X = (\mathbb{Z}/p\mathbb{Z})^\times + X,$$

so that $|X| = 1$. It follows that $X = \{0\}$. If $A$ and $B$ are given, we thus have that $C$ is uniquely determined as $\{1\}$. The quantity $r = r(A, B)$ is defined by (2). In fact, we have

$$p + 1 - r = |A| + |B| + |C|,$$

which shows that $r$ is symmetric in $A, B, C$.

We have already seen that if $A$ and $B$ are given, we get Theorem 2 as presented in the introduction. Now, if $B$ and $C$ are given, we get the theorem conjectured by Hamidoune, Serra, and Zémor.
8. Finis

There is a conjecture saying that if \( p \) is large, then Theorem 4 is valid without any special upper bound on \( k \). Seva Lev [20] has great expectations to a proof of this conjecture. He says: “A “true” combinatorial proof . . . may result in a real progress in additive combinatorial number theory.”

References

[1] N. Alon, M. B. Nathanson, and I. Z. Ruzsa, Adding distinct congruence classes modulo a prime, *American Math. Monthly* **102** (1995), 250–255.
[2] N. Alon, M. B. Nathanson, and I. Z. Ruzsa, The polynomial method and restricted sums of residue classes, *J. Number Theory* **56** (1996), 404–417.
[3] Y. F. Bilu, V. F. Lev, and I. Z. Ruzsa, Rectification principles in additive number theory, *Discrete Comput. Geom.*, **19** (1998), 343–353.
[4] A. L. Cauchy, Recherches sur les nombres, *J. École Polytech.* **9** (1813), 99–123; also in *Oeuvres*, Série 2, Tome 1, 39–63.
[5] I. Chowla, A theorem on the addition of residue classes: Application to the number \( \Gamma(k) \) in Waring’s problem, *Proc. Indian Acad. Sci., Section A*, **1** (1935), 242–243.
[6] H. Davenport, On the addition of residue classes, *J. London Math. Soc.* **10** (1935), 30–32.
[7] H. Davenport, A historical note, *J. London Math. Soc.* **22** (1947), 100-101.
[8] J. A. Dias da Silva and Y. O. Hamidoune, A note on the minimal polynomial of the Kronecker sum of two linear operators, *Linear Algebra and its Applications* **141** (1990), 283–287.
[9] J. A. Dias da Silva and Y. O. Hamidoune, Cyclic spaces for Grassmann derivatives and additive theory, *J. London Math. Soc.* **26** (1994), 140–146.
[10] G. A. Freiman, Inverse problems of additive number theory (Russian), *Učebn. Zap. Kazan. Univ.* **115** (1955) (14), 109–115.
[11] G. A. Freiman, Inverse problems of the additive theory of numbers (Russian), *Izv. Akad. Nauk SSSR* Ser. Mat. **19** (1955), 175–284.
[12] G. A. Freiman, Inverse problems of additive theory of numbers of finite sets. I. (Russian), *Izv. Vysš. Učebn. Zaved. Matematika*, **13** (6) (1959), 202–213.
[13] G. A. Freiman, *Foundations of a Structural Theory of Set Addition*, Translations of Mathematical Monographs, Vol. 37, American Math. Soc., Providence, R. I. (1973).
[14] B. Green and I. Z. Ruzsa, Sets with small sumset and rectification, *Bull. London Math. Soc.* **38** (2006), no. 1, 43–53.
[15] J. T. Griesmer, An inverse theorem: when \( m(A+B) = m(A) + m(B) \) in locally compact abelian groups, [arXiv:1112.6403v1](https://arxiv.org/abs/1112.6403v1).
[16] Y. O. Hamidoune and O. J. Rödseth, An inverse theorem mod \( p \), *Acta Arith.*, **92** (3) (2000), 251–262.
[17] Y. O. Hamidoune, O. Serra, and G. Zémor, On the critical pair theory in \( \mathbb{Z}/p\mathbb{Z} \), *Acta Arith.* **121** (2) (2006), 99–115.
[18] L. Husbands, *Approximate Groups in Additive Combinatorics: A Review of Methods and Literature*, MSc thesis, University of Bristol 2009, 109 pp.
[19] M. Kneser, Summenmengen in lokalkompakten abelschen Gruppen, Math. Zeit., 66 (1956), 88–110.
[20] V. F. Lev, Web page, [http://math.haifa.ac.il/~seva](http://math.haifa.ac.il/~seva).
[21] M. B. Nathanson, Additive Number Theory: Inverse Problems and the Geometry of Sumsets, Springer, New York, 1996.
[22] Ø. J. Rødseth, Sums of distinct residues mod p, Acta Arith. 65 (1993), 181–184.
[23] Ø. J. Rødseth, Sumsets mod p, Trans. R. Norw. Soc. Sci. Lett. 2006 (4), 1–10.
[24] Ø. J. Rødseth, On Freiman’s 2.4-theorem, Trans. R. Norw. Soc. Sci. Lett. 2006 (4), 11–18.
[25] Ø. J. Rødseth, The Erdős-Heilbronn conjecture to an arbitrary modulus, in preparation.
[26] T. Tao and V. H. Vu, Additive Combinatorics, Camb. Stud. Advanced Math. 105, Cambridge University Press, 2006.
[27] O. Serra and G. Zémor, Large sets with small doubling modulo $p$ are well covered by arithmetic progressions, Ann. L’Institut Fourier 59 (5) (2009), 2043–2060.
[28] A. G. Vosper, The critical pairs of subsets of a group of prime order, J. London Math. Soc. 31 (1956), 200–205.
[29] A. G. Vosper, Addendum to “The critical pairs of subsets of a group of prime order”, J. London Math. Soc. 31, (1956), 280–282.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BERGEN, JOHS. BRUNSGT. 12, N-5008 BERGEN, NORWAY

E-mail address: rodseth@math.uib.no
URL: [http://math.uib.no/folk/nmaoy](http://math.uib.no/folk/nmaoy)