The inverse Erdős-Heilbronn problem

Van H. Vu*

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA

vanvu@math.rutgers.edu

Philip Matchett Wood

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA

matchett@math.rutgers.edu

Submitted: Aug 14, 2008; Accepted: Jul 24, 2009; Published: Aug 7, 2009
Mathematics Subject Classification: 11P70

Abstract

The famous Erdős-Heilbronn conjecture (first proved by Dias da Silva and Hamidoune in 1994) asserts that if A is a subset of $\mathbb{Z}/p\mathbb{Z}$, the cyclic group of the integers modulo a prime p, then $|A \widehat{+} A| \geqslant \min\{2|A| - 3, p\}$. The bound is sharp, as is shown by choosing A to be an arithmetic progression. A natural inverse result was proven by Karolyi in 2005: if $A \subset \mathbb{Z}/p\mathbb{Z}$ contains at least 5 elements and $|A \widehat{+} A| \leqslant 2|A| - 3 < p$, then A must be an arithmetic progression.

We consider a large prime p and investigate the following more general question: what is the structure of sets $A \subset \mathbb{Z}/p\mathbb{Z}$ such that $|A \widehat{+} A| \leqslant (2+\epsilon)|A|$?

Our main result is an asymptotically complete answer to this question: there exists a function $\delta(p) = o(1)$ such that if $200 < |A| \leqslant (1-\epsilon')p/2$ and if $|A \widehat{+} A| \leqslant (2+\epsilon)|A|$, where $\epsilon' - \epsilon \geqslant \delta > 0$, then A is contained in an arithmetic progression of length $|A \widehat{+} A| - |A| + 3$.

With the extra assumption that $|A| \leqslant (\frac{1}{2} - \frac{1}{\log^c p})p$, our main result has Dias da Silva and Hamidoune's theorem and Karolyi's theorem as corollaries, and thus, our main result provides purely combinatorial proofs for the Erdős-Heilbronn conjecture and an inverse Erdős-Heilbronn theorem.

1 Introduction

For A a subset of an abelian group, we define the *sumset* of A to be the set of all sums of two elements in A, namely,

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

*V. Vu is supported by NSF grant DMS-0901216 and DOD grant AFOSAR-FA-9550-09-1-0167.

and we define the *restricted sumset* of A to be the set of all sums of two distinct elements of A, namely,

$$A \widehat{+} A := \{a + b : a, b \in A \text{ and } a \neq b\}.$$

Sumsets in a general abelian group have been extensively studied (see [31] for a survey), and we will focus on sumsets of $\mathbb{Z}/p\mathbb{Z}$, the integers modulo p, where p is a prime (see [29] for a survey). For variations on restricted sumset addition, see [25], [26], and [27].

Cauchy [8] and Davenport [9] proved independently that for every $A \subset \mathbb{Z}/p\mathbb{Z}$ we have $|A + A| \geqslant \min\{p, 2|A| - 1\}$. The problem of finding a lower bound for the cardinality of restricted sumsets in $\mathbb{Z}/p\mathbb{Z}$ is much harder. Erdős and Heilbronn made the following conjecture in 1964, which was proved by Dias da Silva and Hamidoune [10] thirty years later.

Theorem 1.1. [10] *For every $A \subset \mathbb{Z}/p\mathbb{Z}$, we have $|A \widehat{+} A| \geqslant \min\{p, 2|A| - 3\}$.*

The $2|A| - 1$ term in the Cauchy-Davenport theorem and the $2|A| - 3$ term in the Dias da Silva-Hamidoune theorem come from the extremal case when A is an arithmetic progression. For unrestricted sumsets, Vosper [40, 39] showed that an arithmetic progression is indeed the only extremal example:

Theorem 1.2. [40, 39] *For $A \subset \mathbb{Z}/p\mathbb{Z}$, if $|A + A| = 2|A| - 1 < p$, then A is an arithmetic progression.*

Though the situation with restricted sumsets is much more difficult, in 2005, Gyula Károlyi [24] proved a theorem that is just as strong as Vosper's:

Theorem 1.3. [24] *For $A \subset \mathbb{Z}/p\mathbb{Z}$, if $|A \widehat{+} A| = 2|A| - 3 < p$ and $5 \leqslant |A|$, then A is an arithmetic progression.*

Theorem 1.3 is notable in that Károlyi [24] succeeds in using an algebraic approach to prove a structural result, which has the added benefit that using ideas in [21, 22], Károlyi is able extend Theorem 1.3 to an arbitrary abelian group (see [24]).

Our goal is to investigate the following more general question:

Question 1.4. *For a constant $0 \leqslant c \leqslant 1$, classify all subsets $A \subset \mathbb{Z}/p\mathbb{Z}$ for which $|A| < p/(2 + c)$ and $|A \widehat{+} A| \leqslant (2 + c)|A|$.*

The $c = 1$ case of Question 1.4 is similar to a conjectural result suggested by Lev [25, page 29] (see Remark 5.1 for a comparison).

Partial answers for Question 1.4 were given by Bilu, Lev, and Ruzsa [5], by Freiman, Low, and Pitman [13], by Lev [25], and by Schoen [33]. To the best of our knowledge, the most current result is the following from [33]:

Theorem 1.5. [33] *For every $\epsilon > 0$, there exists a constant $n_0 = n_0(\epsilon)$ such that every set $A \subset \mathbb{Z}/p\mathbb{Z}$ satisfying $n_0 \leqslant |A| \leqslant p/35$ and satisfying*

$$|A \widehat{+} A| \leqslant (2.4 - \epsilon)|A|$$

is contained in an arithmetic progression in $\mathbb{Z}/p\mathbb{Z}$ of at most $|A \widehat{+} A| - |A| + 3$ terms.

Our main result is the following:

Theorem 1.6 (main theorem). *There exist absolute constants* $p_0 \geqslant 2^{94}$ *and* $c > 0$ *such that the following holds for all* $p \geqslant p_0$ *and all* $0 \leqslant \epsilon < \epsilon' \leqslant 10^{-4}$ *satisfying* $\epsilon' - \epsilon \geqslant \frac{c(\log\log p)^2}{(\log p)^{2/3}}$. *If* $200 \leqslant |A| \leqslant \frac{p-3}{2(1+\epsilon')}$ *and if*

$$|A \widehat{+} A| \leqslant (2+\epsilon)\,|A|\,,$$

then A *is contained in an arithmetic progression of at most* $|A \widehat{+} A| - |A| + 3$ *terms.*

When $|A| \geqslant (p+3)/2$, it is trivial that $A \widehat{+} A$ is all of $\mathbb{Z}/p\mathbb{Z}$. Thus, Theorem 1.6 provides an asymptotically complete answer to Question 1.4 for small c via combinatorial methods. As corollaries to Theorem 1.6, it is easy to derive asymptotically complete versions of Theorem 1.1 and Theorem 1.3, thus providing alternate proofs for the Erdős-Heilbronn conjecture and an inverse Erdős-Heilbronn theorem, except for those A such that $(1-\delta)p/2 < |A| \leqslant (p+1)/2$ or $|A| < 200$, where δ goes to zero as p increases.

2 A combinatorial approach

There are two previous approaches to proving of the Erdős-Heilbronn conjecture. Dias da Silva and Hamidoune [10] used representation theory of the symmetric group, Young tableau, and exterior algebras in their proof. Later, Alon, Nathanson, and Ruzsa [3, 4] found another proof using the powerful Combinatorial Nullstellensatz (see [1, 2, 23] for surveys). Both proofs have a strong algebraic flavor, and in a remarkable step forward, Károlyi [24] used the Combinatorial Nullstellensatz and careful algebraic analysis to prove Theorem 1.3 ([24] also gives an alternate proof of Theorem 1.2).

A more combinatorial approach to the Erdős-Heilbronn conjecture (Theorem 1.1) is the rectification method, introduced by Freiman [12]. To apply the rectification method, one shows that if $|A \widehat{+} A|$ is sufficiently small then A can be viewed as a set of integers, and then one appeals to a version of Theorem 1.1 for subsets of integers (which is not hard to prove). The rectification method was used by Freiman, Low, and Pitman [13] in 1999 to prove Theorem 1.1 with the additional assumption that $60 \leqslant |A| \leqslant p/50$.

To prove our main result (Theorem 1.6), we will combine ideas from the rectification method with a strong new result due to Serra and Zémor [36] (see Subsection 4.2 for a discussion of the Serra-Zémor result). The first step in our proof, which we will carry out in the next section, is to reduce the study of restricted sumsets to non-restricted sumsets. This approach was first applied to the inverse Erdős-Heilbronn problem by Schoen [33] in 2002.

3 Translating between $A + A$ and $A \widehat{+} A$

Lemma 3.1. *There exists an absolute constant* c_0 *such that if* p *is sufficiently large and* $A \subset \mathbb{Z}/p\mathbb{Z}$, *then*

$$|A \widehat{+} A| > |A + A| - p\left(\frac{c_0(\log\log p)^2}{(\log p)^{2/3}}\right).$$

Proof. We proceed by bounding the cardinality of the set $E := \{z \in A : z + z \notin A \widehat{+} A\}$. Note that by the definition of sumset and restricted sumset, $|A + A| = |A \widehat{+} A| + |E|$. If

$$|E| \geqslant p \left(\frac{c_0 (\log \log p)^2}{(\log p)^{2/3}} \right)$$

for a particular constant c_0, then by [7] and the fact that p is sufficiently large, we have that the set E contains a non-trivial three-term arithmetic progression, say $a, b, c \in E$ such that $a \neq c$ and $a + c = 2b$. But then $b + b = 2b = a + c \in A \widehat{+} A$, a contradiction of the definition of the set E. Thus, we must have that

$$|E| < p \left(\frac{c_0 (\log \log p)^2}{(\log p)^{2/3}} \right).$$

Hence

$$|A + A| = |A \widehat{+} A| + |E| < |A \widehat{+} A| + p \left(\frac{c_0 (\log \log p)^2}{(\log p)^{2/3}} \right),$$

which is the desired inequality. □

Later, we found out that Schoen [33] proved a similar result to the above, using a different argument. Both arguments use results of Bourgain [6, 7] on integer sets containing no arithmetic progressions, and in the case when $|A|/p$ is bounded from below by a constant, our bound compares favorably to [33].

4 Background Results

4.1 Rectification

The rectification approach to sumset problems is to show that a subset $A \subset \mathbb{Z}/p\mathbb{Z}$ must behave the same way as a subset $B \subset \mathbb{Z}$, and then to appeal to a sumset result for the integers. For example, Schoen [33] proved Theorem 1.5 by passing to the integers and then applying a corollary of the following result, which is due to Lev (see [25, Theorem 1]).

Theorem 4.1. [25] *Let B be a set of $n \geqslant 3$ non-negative integers such that* $\gcd(B) = 1$ *and* $0 \in B$. *Then,*

$$\left|B \widehat{+} B\right| \geqslant \begin{cases} \max(B) + |B| - 2 & \text{if } \max(B) \leqslant 2\,|B| - 5, \\ 2.61\,|B| - 6 & \text{if } \max(B) \geqslant 2\,|B| - 4. \end{cases}$$

The rectification method was used by Freiman, Low, and Pitman [13, Theorem 2] to give the first partial answer to Question 1.4, and Lev [25] improved on their result to get the following theorem.

Theorem 4.2. [25] *Let A be a subset of $\mathbb{Z}/p\mathbb{Z}$ where $200 \leqslant |A| \leqslant p/50$. If*

$$|A \widehat{+} A| \leqslant 2.18\,|A| - 6,$$

then A is contained in an arithmetic progression of at most $|A \widehat{+} A| - |A| + 3$ terms.

We will use Theorem 4.2 to prove our main theorem (Theorem 1.6) in the case where A has cardinality $200 \leqslant |A| \leqslant p/50$.

4.2 The isoperimetric method

The isoperimetric method is an alternative to the rectification method, and it is used to indirectly show that a subset $A \subset \mathbb{Z}/p\mathbb{Z}$ behaves like a subset of the integers, typically by studying an extremal set that is constructed using the original set A. The isoperimetric method was introduced by Hamidoune [14] and was developed by the same author [15, 16] along with Serra and Zémor as coauthors [18, 19]. For a survey of the isoperimetric method, see [34].

The following is the main result from the isoperimetric method that we will use, and it was proven by Serra and Zémor [36, Theorem 3].

Theorem 4.3. [36] *There exist positive numbers p_0 and ϵ' such that for all primes $p > p_0$, any subset A of $\mathbb{Z}/p\mathbb{Z}$ such that*

(i) $|A + A| < (2 + \epsilon')\,|A|$ *and*

(ii) $m = |A + A| - 2\,|A|$ *satisfies* $m \leqslant \min\{|A| - 4, p - |A + A| - 3\}$

is contained in an arithmetic progression of at most $|A| + m + 1$ terms. Furthermore, one can take $\epsilon' = 10^{-4}$ and $p_0 = 2^{94}$.

Previous inverse theorems for sumsets focused on making the value of ϵ' as large as possible, even at the expense of requiring $|A|$ to be small. Serra and Zémor [36], on the other hand, proved the above result allowing $|A|$ to be as large as possible, at the expense of requiring ϵ' to be small.

5 Proof of the main theorem (Theorem 1.6)

By Theorem 4.2, we may assume that $|A| > p/50$. By hypothesis $|A \mathbin{\widehat{+}} A| \leqslant (2 + \epsilon)\,|A|$, and so by Lemma 3.1,

$$\begin{aligned}(2+\epsilon)\,|A| \geqslant |A \mathbin{\widehat{+}} A| &> |A + A| - p\left(\frac{c_0(\log\log p)^2}{(\log p)^{2/3}}\right)\\ &\geqslant |A + A|\left(1 - \frac{c'(\log\log p)^2}{(\log p)^{2/3}}\right),\end{aligned}$$

where, say, $c' = 50c_0$.

It is straightforward to verify condition (ii) of Theorem 4.3, and so we need to verify condition (i) by showing

$$\frac{2+\epsilon}{1 - \left(\frac{c'(\log\log p)^2}{(\log p)^{2/3}}\right)} \leqslant 2 + \epsilon'. \tag{1}$$

Setting $c = (2 + 10^{-4})c'$, we see that Inequality (1) is true if $\frac{c(\log\log p)^2}{(\log p)^{2/3}} \leqslant \epsilon' - \epsilon$, which holds by assumption.

Thus, we can apply Theorem 4.3 to show that A is contained in an arithmetic progression with at most $|A + A| - |A| + 1 \leqslant (1 + \epsilon')\,|A| + 1 \leqslant (p-1)/2$ terms. The next step is to show that A is Freiman isomorphic of order 2 to a set integers satisfying the hypotheses of Theorem 4.1, which will allow us to conclude the result (see [38, Chapter 5.3] for a discussion of Freiman isomorphisms).

Let $L := \{a_0 + id \mod p : 0 \leqslant i \leqslant (1+\epsilon')\,|A|\}$ be an arithmetic progression containing A, where i, a_0, and d are integers. Note that L is Freiman isomorphic or order 2 to the set of integers $M = \{0, 1, 2, \ldots, \lfloor (1+\epsilon')\,|A| \rfloor\}$ and that A is Freiman isomorphic of order 2 to the set of integers $B = \{i \in M : a_0 + id \mod p \in A\}$. We may assume (by shifting L if necessary) that $a_0 \mod p \in A$, so that $0 \in B$ and B consists of non-negative integers. Since B is sufficiently dense in the interval M (recall, M contains at most $(1 + \epsilon')\,|B| + 1$ elements), we know that there exist two elements of B that differ by exactly 1, and so $\gcd(B) = 1$. Finally, we have $\left|B \hat{+} B\right| = |A \hat{+} A| \leqslant (2 + \epsilon)\,|A| = (2 + \epsilon)\,|B|$, and so by Theorem 4.1, we have that

$$\max(B) \leqslant \left|B \hat{+} B\right| - |B| + 2 = |A \hat{+} A| - |A| + 2.$$

Thus, B is contained in $M' := \{0, 1, 2, \ldots, |A \hat{+} A| - |A| + 2\}$, and so A is contained in $L' := \{a_0 + id \mod p : 0 \leqslant i \leqslant |A \hat{+} A| - |A| + 2\}$. We have thus shown that A is contained in an arithmetic progression of at most $|A \hat{+} A| - |A| + 3$ terms. □

Remark 5.1. It has been conjectured (see [25, page 29]) that a structure theorem along the lines of Theorem 1.6 may hold for a subset $A \subset \mathbb{Z}/p\mathbb{Z}$ satisfying $|A \hat{+} A| \leqslant 3\,|A| - 7$ and $|A| \leqslant (p-C)/2$, for some relatively small absolute constant C. However, it is possible to randomly construct sets A such that $|A|$ is slightly larger than $p/3$ and such that A has no arithmetic structure. Such a set A automatically satisfies $|A \hat{+} A| \leqslant 3\,|A| - 7$ (since $3\,|A| \geqslant p + 7$) and therefore violates the conjecture. In general, by the same random construction, any structure result derived from the hypothesis $|A \hat{+} A| \leqslant (2 + c)\,|A|$, where $0 \leqslant c \leqslant 1$ is a constant, must also include the hypothesis $|A| \leqslant p/(2 + c)$. For this reason, we include the hypothesis $|A| < p/(2 + c)$ in Question 1.4.

Acknowledgments

We would like to thank the anonymous referee for very helpful comments. Also, the work of the second author was supported by an NSF Graduate Research Fellowship.

References

[1] Noga Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), no. 1-2, 7–29, Recent trends in combinatorics (Mátraháza, 1995).

[2] ———, *Discrete mathematics: methods and challenges*, Proceedings, International Congress of Mathematicians, Vol. I (Beijing), Higher Ed. Press, 2002, pp. 119–135.

[3] Noga Alon, Melvyn B. Nathanson, and Imre Ruzsa, *Adding distinct congruence classes modulo a prime*, Amer. Math. Monthly **102** (1995), no. 3, 250–255.

[4] ———, *The polynomial method and restricted sums of congruence classes*, J. Number Theory **56** (1996), no. 2, 404–417.

[5] Y. F. Bilu, V. F. Lev, and I. Z. Ruzsa, *Rectification principles in additive number theory*.

[6] Jean Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.

[7] Jean Bourgain, *Roth's theorem on progressions revisited*, J. Anal. Math. **104** (2008), no. 1, 155–192.

[8] Augustin Louis Cauchy, *Recherches sur les nombres*, J. École polytech **9** (1813), 99–116.

[9] Harold Davenport, *O the addition of residue classes*, J. London Math. Soc. **10** (1935), 30–32.

[10] J. A. Dias da Silva and Y. O. Hamidoune, *Cyclic spaces for Grassmann derivatives and additive theory*, Bull. London Math. Soc. **26** (1994), no. 2, 140–146.

[11] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique, vol. 28, Université de Genève L'Enseignement Mathématique, Geneva, 1980.

[12] G. A. Freĭman, *Foundations of a structural theory of set addition*, American Mathematical Society, Providence, R. I., 1973, Translated from the Russian, Translations of Mathematical Monographs, Vol 37.

[13] Gregory A. Freiman, Lewis Low, and Jane Pitman, *Sumsets with distinct summands and the Erdős-Heilbronn conjecture on sums of residues*, Astérisque (1999), no. 258, xii–xiii, 163–172, Structure theory of set addition.

[14] Y. O. Hamidoune, *An isoperimetric method in additive theory*, J. Algebra **179** (1996), no. 2, 622–630.

[15] Yahya Ould Hamidoune, *Subsets with small sums in abelian groups. I. The Vosper property*, European J. Combin. **18** (1997), no. 5, 541–556.

[16] ———, *Some results in additive number theory. I. The critical pair theory*, Acta Arith. **96** (2000), no. 2, 97–119.

[17] Yahya Ould Hamidoune and Øystein J. Rødseth, *An inverse theorem mod p*, Acta Arith. **92** (2000), no. 3, 251–262.

[18] Yahya Ould Hamidoune, Oriol Serra, and Gilles Zémor, *On the critical pair theory in $\mathbb{Z}/p\mathbb{Z}$*, Acta Arith. **121** (2006), no. 2, 99–115.

[19] ———, *On the critical pair theory in abelian groups: Beyond Chowla's theorem. to appear in* Combinatorica, arXiv:math/0603478v2 [math.NT] (22 Oct 2007), 23 pages.

[20] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. (2) **35** (1987), no. 3, 385–394.

[21] Gyula Károlyi, *On restricted set addition in abelian groups*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **46** (2003), 47–54 (2004).

[22] ———, *The Erdős-Heilbronn problem in abelian groups*, Israel J. Math. **139** (2004), 349–359.

[23] ———, *A compactness argument in the additive theory and the polynomial method*, Discrete Math. **302** (2005), no. 1-3, 124–144.

[24] ———, *An inverse theorem for the restricted set addition in abelian groups*, J. Algebra **290** (2005), no. 2, 557–593.

[25] Vsevolod F. Lev, *Restricted set addition in groups. I. The classical setting*, J. London Math. Soc. (2) **62** (2000), no. 1, 27–40.

[26] ———, *Restricted set addition in groups. II. A generalization of the Erdős-Heilbronn conjecture*, Electron. J. Combin. **7** (2000), Research Paper 4, 10 pp.

[27] ———, *Restricted set addition in groups. III. Integer sumsets with generic restrictions*, Period. Math. Hungar. **42** (2001), no. 1-2, 89–98.

[28] ———, *Restricted set addition in abelian groups: results and conjectures*, J. Théor. Nombres Bordeaux **17** (2005), no. 1, 181–193.

[29] Øystein J. Rødseth, *Sumsets mod p*, Skr. K. Nor. Vidensk. Selsk. (2006), no. 4, 1–10.

[30] I. Z. Ruzsa, *Arithmetical progressions and the number of sums*, Period. Math. Hungar. **25** (1992), no. 1, 105–111.

[31] Imre Z. Ruzsa, *Sumsets*, European Congress of Mathematics, Eur. Math. Soc., Zürich, 2005, pp. 381–389.

[32] Tom Sanders, *Appendix to 'Roth's theorem on progressions revisited,' by J. Bourgain*, J. Anal. Math. **104** (2008), no. 1, 193–206.

[33] Tomasz Schoen, *The cardinality of restricted sumsets*, J. Number Theory **96** (2002), no. 1, 48–54.

[34] O. Serra, *An isoperimetric method for the small sumset problem*, Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, pp. 119–152.

[35] Oriol Serra and Gilles Zémor, *On a generalization of a theorem by Vosper*, Integers (2000), A10, 10 pp.

[36] Oriol Serra and Gilles Zémor, *Large sets with small doubling modulo p are well covered by an arithmetic progression*, arXiv:0804.0935 [math.NT] (6 Apr 2008), 16 pages.

[37] E. Szemerédi, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. **56** (1990), no. 1-2, 155–158.

[38] Terence Tao and Van Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.

[39] A. G. Vosper, *Addendum to "The critical pairs of subsets of a group of prime order"*, J. London Math. Soc. **31** (1956), 280–282.

[40] ———, *The critical pairs of subsets of a group of prime order*, J. London Math. Soc. **31** (1956), 200–205.