LINEAR ALGEBRAIC METHOD AND THE ERDŐS-HEILBRONN CONJECTURE

GUANZHONG YANG

Abstract. Two foundational results in additive combinatorics are the Cauchy–Davenport theorem and the Erdős–Heilbronn theorem. They give lower bounds for the size of sumsets and restricted sumsets over finite fields modulo primes. In this note, we develop Das's linear algebraic method and give a new elementary proof of the Alon–Nathanson–Ruzsa theorem for restricted sumsets. This directly implies the Erdős–Heilbronn theorem. Compared with the classical polynomial method via Combinatorial Nullstellensatz, our proof uses only basic linear algebra over finite fields, including Vandermonde matrices and solvability of linear systems.

1. Introduction

Additive combinatorics is a central branch of combinatorial number theory, focusing on the additive structure of subsets of integers and finite fields [7]. A foundational problem in this field is to give sharp lower bounds for the cardinality of sumsets, which has wide applications in number theory, combinatorics, and theoretical computer science.

Let p be a prime number, and let $\mathbb{Z}/p\mathbb{Z}$ denote the finite field of integers modulo p. For nonempty subsets A, B of $\mathbb{Z}/p\mathbb{Z}$, the classical sumset is defined as

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

The celebrated Cauchy-Davenport theorem, a cornerstone of additive combinatorics, gives a tight lower bound for the size of such sumsets:

$$|A+B|\geqslant \min\{p,\ |A|+|B|-1\}.$$

In 1964, Erdős and Heilbronn proposed a famous conjecture for the so-called restricted sumset, where the summands are required to be distinct [5]. For nonempty subsets A, B of $\mathbb{Z}/p\mathbb{Z}$, we define the restricted sumset as

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

2020 *Mathematics Subject Classification.* Primary 11B75; Secondary 11P70.
Key words and phrases. sumset, Erdős–Heilbronn conjecture, linear algebra.

The Erdős-Heilbronn conjecture states that for any nonempty subset A of $\mathbb{Z}/p\mathbb{Z}$,

$$|A\dot{+}A| \geqslant \min\{p,\ 2|A|-3\}.$$

This conjecture remained open for 30 years, until it was first proven by Dias da Silva and Hamidoune in 1994, using tools from exterior algebra and representation theory [4]. In 1995, Alon, Nathanson and Ruzsa developed the polynomial method, specifically the Combinatorial Nullstellensatz [2, 1], to give a simpler proof of a more general result: for nonempty subsets A, B of $\mathbb{Z}/p\mathbb{Z}$ with $|A| \neq |B|$,

$$|A\dot{+}B| \geqslant \min\{p,\ |A|+|B|-2\}.$$

This general result directly implies the original Erdős-Heilbronn conjecture as a corollary.

In the decades since, multiple new proofs of the Cauchy-Davenport theorem and the Erdős-Heilbronn conjecture have been developed, using tools ranging from harmonic analysis to the uncertainty principle on finite cyclic groups [6]. In 2004, Das introduced a linear algebraic method to give a new proof of the Cauchy-Davenport theorem, relying only on elementary linear algebra over finite fields [3]. In this note, we extend Das's linear algebraic approach to give a new, elementary proof of the Alon-Nathanson-Ruzsa theorem for restricted sumsets. Compared with existing proofs, our method requires no advanced algebraic or combinatorial machinery, only standard undergraduate-level linear algebra, including Vandermonde matrices and the solvability of linear systems over finite fields.

The rest of this paper is organized as follows. Section 2 collects the preliminary definitions and auxiliary lemmas that form the foundation of our proof. Section 3 states our main results, including the main theorem and its direct corollary, the Erdős-Heilbronn theorem. The complete proof of the main theorem is presented in Section 4. We conclude with some brief remarks and directions for future work in Section 5.

2. Preliminaries

In this section, we introduce the core definitions and auxiliary lemmas used throughout the paper. All sets are assumed to be nonempty subsets of $\mathbb{Z}/p\mathbb{Z}$, where p is a fixed prime number. For a finite set S, we write $|S|$ for the cardinality (number of elements) of S.

Let $A = \{a_1, \ldots, a_m\}$ be a subset of $\mathbb{Z}/p\mathbb{Z}$ with $a_1, \ldots, a_m$ pairwise distinct. Let $w(a_1), \ldots, w(a_m)$ be a sequence in $\mathbb{Z}/p\mathbb{Z}$ satisfying $w(a_\ell)$ is nonzero in $\mathbb{Z}/p\mathbb{Z}$ for some $1 \leqslant \ell \leqslant m$. We shall say a sequence

$u_1, \ldots, u_n$ a nonzero sequence if $u_\ell \neq 0$ for some $1 \leqslant \ell \leqslant n$. We use $e_A(w)$ to denote the smallest natural number i such that $\sum_{j=1}^{m} w(a_j)a_j^i$ is nonzero in $\mathbb{Z}/p\mathbb{Z}$. We have the following conclusion.

Lemma 2.1. *Let A, w and $e_A(w)$ be as above. Then*

$$e_A(w) \leqslant |A| - 1.$$

Proof. It is proved by contradiction. Suppose that $e_A(w) \geqslant |A|$. This means

$$\sum_{j=1}^{m} w(a_j)a_j^i = 0 \tag{2.1}$$

for all $0 \leqslant i \leqslant m-1$.

The (coefficients) matrix M is defined to be

$$M = (a_j^{i-1})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant m}}. \tag{2.2}$$

Note that

$$\det(M) = \prod_{1 \leqslant i < j \leqslant m} (a_j - a_i).$$

Since $a_1, \ldots, a_m$ are pairwise distinct in $\mathbb{Z}/p\mathbb{Z}$, $\det(M)$ is nonzero. Consider the system of linear equations

$$M\mathbf{x} = \mathbf{0}, \tag{2.3}$$

where $\mathbf{x} = (x_1, \ldots, x_m)^T$ and $\mathbf{0} = (0, \ldots, 0)^T$.

On one hand, (2.3) has only the zero solution due to the fact that $\det(M)$ is nonzero. On the other hand, $(w(a_1), \ldots, w(a_m))^T$ is a solution to (2.3) in view of (2.1). This is a contradiction since $w(a_\ell)$ is nonzero in $\mathbb{Z}/p\mathbb{Z}$ for some $1 \leqslant \ell \leqslant m$. This completes the proof. □

The inequality in Lemma 2.1 is sharp since we have the following.

Lemma 2.2. *Let $A = \{a_1, \ldots, a_m\}$ be a subset of $\mathbb{Z}/p\mathbb{Z}$ with $m \geqslant 2$. Then there exists a nonzero sequence $w(a_1), \ldots, w(a_m)$ such that*

$$e_A(w) = |A| - 1.$$

Proof. The proof is similar to that of Lemma 2.1. Let M be as in (2.2). Now consider the linear equations

$$M\mathbf{x} = \mathbf{b}, \tag{2.4}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{b} = (0, \ldots, 0, 1)^T$. Since $\det(M)$ is nonzero, (2.4) has a unique solution $(w_1, \ldots, w_m)^T$. Since $\mathbf{b}$ is a nonzero vector, $(w_1, \ldots, w_m)^T$ is also a nonzero vector. In particular, we have

w_ℓ is nonzero for some $1 \leqslant \ell \leqslant m$. On choosing $w(a_j) = w_j$ for $1 \leqslant j \leqslant m$, we have

$$\sum_{j=1}^{m} w(a_j) a_j^i = 0$$

for all $0 \leqslant i \leqslant m-2$, and

$$\sum_{j=1}^{m} w(a_j) a_j^{m-1} = 1.$$

According to the definition of $e_A(w)$, we have $e_A(w) = m-1 = |A|-1$. This completes the proof. □

3. Main Results

In this section, we state our main results, including the general lower bound for restricted sumsets (the Alon-Nathanson-Ruzsa theorem) and its direct corollary, the Erdős-Heilbronn theorem. We continue to work over the finite field $\mathbb{Z}/p\mathbb{Z}$ for a fixed prime p, with all sumset notation consistent with Section 1.

Theorem 3.1 (Main Theorem)**.** *Let A and B be nonempty subsets of $\mathbb{Z}/p\mathbb{Z}$ with $|A| \neq |B|$. Then the cardinality of the restricted sumset satisfies*

$$|A \dot{+} B| \geqslant \min\{p,\ |A| + |B| - 2\}.$$

This main theorem directly implies the classical Erdős-Heilbronn conjecture, which we state formally as a corollary below.

Corollary 3.1 (Erdős-Heilbronn Theorem)**.** *Let A be a nonempty subset of $\mathbb{Z}/p\mathbb{Z}$. Then*

$$|A \dot{+} A| \geqslant \min\{p,\ 2|A| - 3\}.$$

Proof. Fix any element $a_0 \in A$, and define $A' = A \setminus \{a_0\}$. Then $|A| \neq |A'|$, and it is straightforward to verify that $A \dot{+} A = A \dot{+} A'$: any sum $a + b$ with $a, b \in A$ and $a \neq b$ can be written as a sum of an element from A and an element from A' with distinct summands, and vice versa. Applying the Main Theorem to the pair (A, A'), we get

$$\begin{aligned} |A \dot{+} A| = |A \dot{+} A'| &\geqslant \min\{p,\ |A| + |A'| - 2\} \\ &= \min\{p,\ |A| + (|A| - 1) - 2\} \\ &= \min\{p,\ 2|A| - 3\}, \end{aligned}$$

which completes the proof. □

Our proof of the Main Theorem, presented in the next section, uses only the elementary linear algebraic tools established in Section 2. This gives a more accessible proof of the Erdős-Heilbronn theorem, avoiding the advanced combinatorial and algebraic machinery used in prior arguments.

4. Proof of the Main Theorem

In this section, we give the complete proof of the Main Theorem using the linear algebraic tools from Section 2. We split the proof into several auxiliary lemmas, followed by the final argument for the main result.

Let $A = \{a_1, \dots, a_m\}$ and $B = \{b_1, \dots, b_k\}$ be subsets of $\mathbb{Z}/p\mathbb{Z}$. Let $w_1(a_1), \dots, w_1(a_m)$ and $w_2(b_1), \dots, w_2(b_k)$ be two sequences in $\mathbb{Z}/p\mathbb{Z}$. For nonnegative integer i, we introduce

$$\alpha_i := \alpha_i(A, w_1) = \sum_{j=1}^{m} w_1(a_j) a_j^i \tag{4.5}$$

and

$$\beta_i := \beta_i(B, w_2) = \sum_{j=1}^{k} w_2(b_j) b_j^i. \tag{4.6}$$

Let $C = A \dot{+} B$, and suppose that $C = \{c_1, \dots, c_t\}$. For $c_j \in C$, we define

$$w(c_j) = \sum_{\substack{a \in A, b \in B \\ a+b=c_j}} w_1(a) w_2(b)(a-b).$$

Then we introduce

$$\gamma_i := \gamma_i(C, w) = \sum_{j=1}^{t} w(c_j) c_j^i.$$

Lemma 4.1. *Let A, B, C and $\alpha_i, \beta_i, \gamma_i$ be as above. Then*

$$\gamma_n = \sum_{i=0}^{n} C_n^i \alpha_{i+1} \beta_{n-i} - \sum_{i=0}^{n} C_n^i \alpha_i \beta_{n+1-i}.$$

Proof. Note that

$$\gamma_n = \sum_{j=1}^{t} w(c_j)c_j^n = \sum_{j=1}^{t} c_j^n \sum_{\substack{a\in A, b\in B \\ a+b=c_j}} w_1(a)w_2(b)(a-b)$$
$$= \sum_{a\in A, b\in B} w_1(a)w_2(b)(a-b)(a+b)^n.$$

Since

$$(a+b)^n = \sum_{i=0}^{n} C_n^i a^i b^{n-i},$$

we have

$$\gamma_n = \sum_{a\in A, b\in B} w_1(a)w_2(b)(a-b)\sum_{i=0}^{n} C_n^i a^i b^{n-i}$$
$$= \sum_{a\in A, b\in B} w_1(a)w_2(b)\sum_{i=0}^{n} C_n^i a^{i+1} b^{n-i}$$
$$- \sum_{a\in A, b\in B} w_1(a)w_2(b)\sum_{i=0}^{n} C_n^i a^i b^{n+1-i}.$$

We observe that

$$\sum_{a\in A, b\in B} w_1(a)w_2(b)\sum_{i=0}^{n} C_n^i a^{i+1} b^{n-i}$$
$$= \sum_{i=0}^{n} C_n^i \Big(\sum_{a\in A} w_1(a)a^{i+1}\Big)\Big(\sum_{b\in B} w_2(b)b^{n-i}\Big)$$
$$= \sum_{i=0}^{n} C_n^i \alpha_{i+1}\beta_{n-i}.$$

Similarly, we also have

$$\sum_{a\in A, b\in B} w_1(a)w_2(b)\sum_{i=0}^{n} C_n^i a^i b^{n+1-i}$$
$$= \sum_{i=0}^{n} C_n^i \Big(\sum_{a\in A} w_1(a)a^{i}\Big)\Big(\sum_{b\in B} w_2(b)b^{n+1-i}\Big)$$
$$= \sum_{i=0}^{n} C_n^i \alpha_{i}\beta_{n+1-i}.$$

Now we can conclude from above that

$$\gamma_n = \sum_{i=0}^{n} C_n^i \alpha_{i+1}\beta_{n-i} - \sum_{i=0}^{n} C_n^i \alpha_i \beta_{n+1-i}.$$

The proof of this lemma is finished. □

Lemma 4.2. *Let A, B, C and $\alpha_i, \beta_i, \gamma_i$ be as above. Let r, s be non-negative integers. Assume that $\alpha_i = 0$ for $0 \leqslant i \leqslant r$ and $\beta_i = 0$ for $0 \leqslant i \leqslant s$. Then*

$$\gamma_{r+s+1} = \Big(C_{r+s+1}^r - C_{r+s+1}^s\Big)\alpha_{r+1}\beta_{s+1}$$

and

$$\gamma_n = 0 \quad \textit{for all} \quad 0 \leqslant n \leqslant r+s.$$

Proof. By Lemma 4.1,

$$\gamma_{r+s+1} = \sum_{i=0}^{r+s+1} C_{r+s+1}^i \alpha_{i+1}\beta_{r+s+1-i} - \sum_{i=0}^{r+s+1} C_{r+s+1}^i \alpha_i \beta_{r+s+2-i}.$$

If $i \leqslant r-1$, then $i+1 \leqslant r$ and thus $\alpha_{i+1} = 0$. If $i \geqslant r+1$, then $r+s+1-i \leqslant s$ and thus $\beta_{r+s+1-i} = 0$. Therefore, $\alpha_{i+1}\beta_{r+s+1-i} = 0$ for all $i \neq r$. Then we have

$$\sum_{i=0}^{r+s+1} C_{r+s+1}^i \alpha_{i+1}\beta_{r+s+1-i} = C_{r+s+1}^r \alpha_{r+1}\beta_{s+1}.$$

Similarly, $\alpha_i\beta_{r+s+2-i} = 0$ for all $i \neq r+1$. Then we have

$$\sum_{i=0}^{r+s+1} C_{r+s+1}^i \alpha_i \beta_{r+s+2-i} = C_{r+s+1}^s \alpha_{r+1}\beta_{s+1}.$$

We conclude from above

$$\gamma_{r+s+1} = \Big(C_{r+s+1}^r - C_{r+s+1}^s\Big)\alpha_{r+1}\beta_{s+1}.$$

We have

$$\gamma_n = \sum_{i=0}^{n} C_n^i \alpha_{i+1}\beta_{n-i} - \sum_{i=0}^{n} C_n^i \alpha_i \beta_{n+1-i}.$$

For $n \leqslant r+s$, we have $(i+1) + (n-i) = n+1 \leqslant r+s+1$. Then we have other $i+1 \leqslant r$ or $n-i \leqslant s$. Thus, $\alpha_{i+1}\beta_{n-i} = 0$ for all $0 \leqslant i \leqslant n$. Similarly, $\alpha_i\beta_{n+1-i} = 0$ for all $0 \leqslant i \leqslant n$. Then we conclude that $\gamma_n = 0$ for $n \leqslant r+s$.

The proof of this lemma is finished. □

Lemma 4.3. *Suppose that $|A| = 1$ or $|B| = 1$. Then*

$$|A \dot{+} B| \geqslant \min\{p,\ |A| + |B| - 2\}.$$

Proof. Without loss of generality, we can assume that $|A| = 1$. If $|B| = 1$, then the desired inequality holds trivially (although it is possible that $A\dot{+}B = \emptyset$). Now we consider the case $|B| \geqslant 2$. We write $A = \{a_0\}$ and $|B| = k$. We can find $k-1$ distinct elements $b_1, \ldots, b_{k-1}$ in B such that $b_i \neq a_0$ for $1 \leqslant i \leqslant k-1$.

Note that $a_0 + b_1, \ldots, a_0 + b_{k-1} \in A\dot{+}B$. Thus, $|A\dot{+}B| \geqslant k-1 = |A| + |B| - 2 = \min\{p,\ |A| + |B| - 2\}$. The proof of this lemma is finished.

□

Now we are able to prove the Main Theorem. If $|A|+|B|-2 \geqslant p+1$, we claim that there exist nonempty subsets $A' \subset A$ and $B' \subset B$ such that $|A'| + |B'| - 2 = p$ and $|A'| \neq |B'|$.

Suppose that $|A| + |B| - 2 = p + d$ with $d \geqslant 1$. We write $d_1 = \lfloor d/2 \rfloor$ and $d_2 = \lceil d/2 \rceil$. Clearly, $d_1 + d_2 = d$. Without loss of generality, we assume that $|A| < |B|$. We choose d_2 elements $a'_1, \ldots, a'_{d_2}$ in A and d_1 elements $b'_1, \ldots, b'_{d_1}$ in B. Let $A' = A \setminus \{a'_1, \ldots, a'_{d_2}\}$ and $B' = B \setminus \{b'_1, \ldots, b'_{d_1}\}$. From $|A| + |B| - 2 = p + d$, we can see that $|A| \geqslant d+2 \geqslant d_2+2$. Thus, A' is nonempty and $|A'| = |A|-d_2$. Similarly, B' is nonempty and $|B'| = |B|-d_1$. We have $|A'| < |B'|$ since $|A| < |B|$ and $d_2 \geqslant d_1$. It is easy to check that $|A'|+|B'|-2 = |A|-d_2+|B|-d_1-2 = |A| + |B| - 2 + d = p$. Therefore, the above claim is true.

Now it suffices to prove the Main Theorem in the case $|A|+|B|-2 \leqslant p$, since if $|A|+|B|-2 \geqslant p+1$ then $|A\dot{+}B| \geqslant |A'\dot{+}B'| \geqslant \min\{p,\ |A'|+|B'|-2\} = p = \min\{p,\ |A| + |B| - 2\}$.

From now on, we assume that $|A|+|B|-2 \leqslant p$. Let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_k\}$. We have $m+k \leqslant p+2$. In view of Lemma 4.3, we can assume that $m \geqslant 2$ and $k \geqslant 2$. Let $C = A\dot{+}B = \{c_1, \ldots, c_t\}$. By Lemma 2.2, there exists a nonzero sequence $w_1(a_1), \ldots, w_1(a_m)$ such that

$$e_A(w_1) = |A| - 1 = m - 1. \tag{4.7}$$

Also, there exists a nonzero sequence $w_2(b_1), \ldots, w_2(b_k)$ such that

$$e_B(w_2) = |B| - 1 = k - 1. \tag{4.8}$$

Recalling the definitions of $\alpha_i := \alpha_i(A, w_1)$ and $\beta_i := \beta_i(B, w_2)$, we conclude from (4.7) and (4.8) that $\alpha_i = 0$ for $0 \leqslant i \leqslant m-2$, $\alpha_{m-1} \neq 0$, $\beta_i = 0$ for $0 \leqslant i \leqslant k-2$ and $\beta_{k-1} \neq 0$.

Applying Lemma 4.2 with $r = m-2$ and $s = k-2$, we obtain

$$\gamma_{m+k-3} = \Big(C^{m-2}_{m+k-3} - C^{k-2}_{m+k-3}\Big)\alpha_{m-1}\beta_{k-1}.$$

Since $m+k-3 \leqslant p-1$ and $m \neq k$, we have $C_{m+k-3}^{m-2} - C_{m+k-3}^{k-2} \not\equiv 0 \pmod{p}$. Note that $\alpha_{m-1}\beta_{k-1}$ is nonzero, we have $\gamma_{m+k-3} \neq 0$. By Lemma 4.2, we also have $\gamma_n = 0$ for $0 \leqslant n \leqslant m+k-4$.

Therefore, $m+k-3$ is the smallest natural number i such that $\sum_{j=1}^{t} w(c_j)c_j^i$ is nonzero. According to the definition of $e_C(w)$, we have $e_C(w) = m+k-3$.

We apply Lemma 2.1 to the set C to conclude that $e_C(w) \leqslant |C|-1$. Now we obtain $|C| \geqslant e_C(w)+1 = m+k-2$. This proves $|A \dot{+} B| \geqslant |A|+|B|-2 = \min\{p,\ |A|+|B|-2\}$. The proof of the Main Theorem is finished.

5. Concluding Remarks

In this note, we developed the linear algebraic method initiated by Das to give a new, elementary proof of the Alon-Nathanson-Ruzsa theorem for restricted sumsets, which directly implies the classical Erdős–Heilbronn conjecture. Unlike prior proofs that rely on the Combinatorial Nullstellensatz, exterior algebra, or harmonic analysis, our argument uses only standard undergraduate linear algebra over finite fields, making the result more accessible to a broader audience.

There are several natural directions for future work. First, it would be interesting to extend this linear algebraic approach to more general restricted sumset results, for example the Dias da Silva–Hamidoune theorem for sums of n distinct elements. Second, our method is currently restricted to sumsets over finite fields of prime order. It remains open whether this framework can be adapted to give lower bounds for sumsets over composite moduli or more general abelian groups. Finally, this approach may have applications to other problems in additive combinatorics, such as the Erdős–Szemerédi theorem on sum-product sets, which we leave for future investigation.

References

1. N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), 7–29.
2. N. Alon, M. B. Nathanson, and I. Z. Ruzsa, *Adding Distinct Congruence Classes Modulo a Prime*, Amer. Math. Monthly **102** (1995), 250–255.
3. P. Das, *Value Sets of Polynomials and the Cauchy-Davenport Theorem*, Finite Fields Appl. **10** (2004), 113–122.
4. J. A. Dias da Silva and Y. O. Hamidoune, *Cyclic Spaces for Grassmann Derivatives and Additive Theory*, Bull. Lond. Math. Soc. **26** (1994), 140–146.
5. P. Erdős and H. Heilbronn, *On the Addition of Residue Classes Modulo p*, Acta Arith. **9** (1964), 149–159.

6. T. Tao, *An Uncertainty Principle for Cyclic Groups of Prime Order*, Math. Res. Lett. **12** (2005), 121–127.
7. T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge Univ. Press, Cambridge, 2006.

Department of Mathematics, Imperial College London, UK
Email address: gy625@ic.ac.uk