Sumsets of Sidon sets

by

IMRE Z. RÜZSA (Budapest)

1. Introduction. A Sidon set is a set $A$ of integers with the property that all the sums $a + b$, $a, b \in A, a \leq b$ are distinct. A Sidon set $A \subset [1, N]$ can have as many as $(1 + o(1))\sqrt{N}$ elements, hence $\sim N/2$ sums. The distribution of these sums is far from arbitrary. Erdős, Sárközy and T. Sós [1, 2] established several properties of these sumsets. Among other things, in [2] they prove that $A + A$ cannot contain an interval longer than $C\sqrt{N}$, and give an example that $N^{1/3}$ is possible. In [1] they show that $A + A$ contains gaps longer than $c \log N$, while the maximal gap may be of size $O(\sqrt{N})$.

We improve these bounds. In Section 2, we give an example of $A + A$ containing an interval of length $c\sqrt{N}$; hence in this question the answer is known up to a constant factor. In Section 3, we construct $A$ such that the maximal gap is $\ll N^{1/3}$. In Section 4, we construct $A$ such that the maximal gap of $A + A$ is $O(\log N)$ in a subinterval of length $cN$.

2. Interval in the sumset. The constructions of Sections 2 and 3 are variants of Erdős and Turán’s classical construction of a dense Sidon set (see e.g. [3]). We quote the common idea in the form of a lemma.

**Lemma 2.1.** If $p$ is a prime and $i, j, k, l$ are integers such that

$$i + j \equiv k + l \pmod{p} \quad \text{and} \quad i^2 + j^2 \equiv k^2 + l^2 \pmod{p},$$

then either $i \equiv k$ and $j \equiv l$, or $i \equiv l$ and $j \equiv k$.

**Theorem 2.2.** Let $c$ be a positive number, $c < 1/\sqrt{54}$. For sufficiently large $N$ there is a Sidon set $A \subset [1, N]$ of integers such that $A + A$ contains an interval of length $c\sqrt{N}$.

**Proof.** Let $p$ be the largest prime below $\sqrt{2N/3} - 4$. For an integer $i$ let $a_i$ denote the smallest nonnegative residue of $i^2$ modulo $p$. Write $q = 2[p/4] + 1$. Let

Supported by Hungarian National Foundation for Scientific Research, Grant No. T 017433.
\[ s_i = 2i + qa_i, \quad t_i = N - i - qa_i, \]

\[ A_1 = \{ s_i : p/6 < i < p/3 \}, \quad A_2 = \{ t_i : p/6 < i < p/3 \}. \]

Our set will be \( A = A_1 \cup A_2 \). Clearly \( s_i + t_i = N + i \in A + A \), thus \( A + A \) contains an interval of length

\[ [p/3] - [p/6] = p/6 + O(1) \sim \sqrt{N/54}. \]

It remains to show that \( A \) is a Sidon set.

Suppose that \( A \) contains four numbers that form a nontrivial solution of the equation \( x + y = u + v \). These numbers can be distributed between \( A_1 \) and \( A_2 \) in five ways. Let Case \( m, 0 \leq m \leq 4 \), refer to the possibility that \( m \) are in \( A_1 \) and \( 4 - m \) in \( A_2 \).

**Case 0.** This leads to the equation \( s_i + s_j = s_k + s_l \), or

\[ 2(i + j - k - l) = q(a_k + a_l - a_i - a_j). \]

Since \( q \) is odd, we have

\[ q | i + j - k - l. \]

These numbers satisfy

\[ (p + 1)/6 \leq i, j, k, l \leq (p - 1)/3, \]

hence

\[ |i + j - k - l| < p/3 < q, \]

thus (2.1) implies \( i + j = k + l \), hence also \( a_i + a_j = a_k + a_l \). This implies

\[ i^2 + j^2 \equiv k^2 + l^2 \pmod{p}. \]

We conclude by Lemma 2.1 that \( (i, j) \) is a permutation of \( (k, l) \).

**Case 1.** This leads to the equation \( s_i + s_j = s_k + t_l \). Since \( 0 < s_i < p(q + 1) \) and \( t_l > N - p(q + 1) \), the right side is always larger than the left, as

\[ 3p(q + 1) < 3p \frac{p + 4}{2} < N. \]

**Case 2.** This means either \( s_i + s_j = t_k + t_l \) or \( s_i + t_j = s_k + t_l \). The first is clearly impossible, since the left side is smaller than the right. The second can be rewritten as

\[ 2i - 2k + l - j = q(a_j + a_k - a_i - a_l). \]

By (2.2) we have

\[ |2i - 2k + l - j| \leq (p - 3)/3 < q, \]

thus we conclude that

\[ 2(i - k) = l - j \]
and

\[ a_k - a_i = a_i - a_j. \]

This equation implies

\[ k^2 - i^2 = (k - i)(k + i) \equiv l^2 - j^2 = (l - j)(l + j) \pmod{p}. \]

By substituting \(2(i - k)\) in place of \(l - j\) this is transformed into

\[ (k - i)(2l + 2j - k - i) \equiv 0 \pmod{p}. \]

By (2.2), the second factor satisfies \(0 < 2l + 2j - k - i < p\), thus it is not a multiple of \(p\). Hence \(k \equiv i\), which implies \(k = i\) and we have a trivial solution.

Case 3 is treated like Case 1, and Case 4 like Case 0. ■

3. An ubiquitous sumset. We say that a set \(X\) forms a \(d\)-chain in an interval if every subinterval of length \(d\) contains at least one element of \(X\).

**Theorem 3.1.** For all sufficiently large \(N\) there is a Sidon set \(A \subset [0, N]\) with the property that \(A + A\) forms a \(CN^{1/3}\)-chain in the interval \([0, 2N]\). Here \(C\) is an absolute constant.

**Proof.** Let \(p\) be the smallest prime satisfying \(2p^3 > 3N\). As before, we denote by \(a_i\) the smallest nonnegative residue of \(i^2\) modulo \(p\). Our set will contain the numbers

\[ s_i = a_i + 2ip + 2b_i p^2, \quad 0 \leq i \leq p - 1, \]

with certain integers \(b_i\).

First we show that these numbers form a Sidon set for an arbitrary choice of the integers \(b_i\). Indeed, suppose that \(s_i + s_j = s_k + s_l\), or

\[ a_i + a_j + 2p(i + j) + 2p^2(b_i + b_j) = a_k + a_l + 2p(k + l) + 2p^2(b_k + b_l). \]

By comparing the residues modulo \(2p\) we find that

\[ a_i + a_j \equiv a_k + a_l \pmod{2p}. \]

Since the left and right sides are both in the interval \([0, 2p - 2]\), this congruence implies equality. It also implies that

\[ i^2 + j^2 \equiv k^2 + l^2 \pmod{p}. \]

Now we delete the \(a\)'s from (3.1), divide by \(p\) and find that

\[ i + j \equiv k + l \pmod{p}. \]

From Lemma 2.1 we conclude that \((i, j)\) is a permutation of \((k, l)\).

Now we choose \(b_i\) so that \(A\) lies in \([0, N]\) and \(A + A\) is dense in \([0, 2N]\). Certainly \(s_i \geq 0\) if \(b_i \geq 0\), and \(s_i \leq N\) holds if we require that

\[ i + pb_i \leq \frac{N - p}{2p}. \]
Write
\[ M = \left\lceil \frac{N}{2p^2} \right\rceil - 1. \]

The largest value of \( b_i \) that satisfies (3.2) is either \( M \) or \( M + 1 \); it is \( M + 1 \) for (3.3)
\[ i \leq i_0 = \left\lceil p \left( \frac{N}{2p^2} \right) - \frac{1}{2} \right\rceil, \]
and \( M \) otherwise.

Observe that since \( 3N \leq 2p^3 \), we have \( 3M \leq p - 1 \).

We put \( b_{3r} = r \) for \( 0 \leq r \leq M \), \( b_{3r} = 0 \) for \( M < r < p/3 \), \( b_{3r+1} = 0 \) for all \( r \) and \( b_{3r+2} = M + 1 \) if \( 3r + 2 \leq i_0 \), \( b_{3r+2} = M \) otherwise.

We have to show that the numbers \( s_i + s_j \) appear in any interval of length \( CN^{1/3} \). Since \( 0 \leq a_i < p = O(N^{1/3}) \), we have
\[ s_i + s_j = 2p(i + j + p(b_i + b_j)) + O(N^{1/3}), \]
and it is sufficient to show that the numbers \( i + j + p(b_i + b_j) \) form a C-chain in \([0, N/p]\) with a constant \( C \).

Write
\[
B_0 = \{ a_{3r} + pb_{3r} : 0 \leq r \leq M \}, \\
B_1 = \{ a_{3r+1} + pb_{3r+1} : 0 \leq r \leq (p-2)/3 \}, \\
B_2 = \{ a_{3r+2} + pb_{3r+2} : 0 \leq r \leq (p-3)/3 \}.
\]

The elements of \( B_0 \) are the multiples of \( p + 3 \) from 0 till \((p + 3)\). The elements of \( B_1 \) are the numbers \( \equiv 1 \) (mod 3) between 1 and \( p - 1 \), so they form a 6-chain in \([0, p+3]\). Hence \( B_0 + B_1 \) forms a 6-chain in the interval \([0, (M+1)(p+3)]\).

The elements of \( B_2 \) are the numbers
(3.4) \[ 2 + p(M + 1), 5 + p(M + 1), \ldots, 2 + 3R + p(M + 1), \]
where \( R \) is such that
(3.5) \[ 2 + 3R + p(M + 1) \leq \frac{N-p}{2p} < 2 + 3(R + 1) + p(M + 1), \]
and after these the numbers
(3.6) \[ 2 + 3(R + 1) + pM, \ldots, 2 + 3 \left[ \frac{p-3}{3} \right] + pM. \]

The length of the gaps within a block is 3. By (3.5), the first element of the block in (3.6) is at most \( N/(2p) - p + 3 \), the difference between the last element of (3.6) and the first of (3.4) is at most 6, while the last element of (3.4) is at least \( N/(2p) - 4 \) again by (3.5). Hence \( B_2 \) forms a 6-chain in \([N/(2p) - (p+3), N/(2p)]\). (One of the blocks may be empty; in this case we
easily get the same conclusion.) Consequently, \( B_0 + B_2 \) forms a 6-chain in 
\[ N/(2p) - (p + 3), N/2 + M(p + 3) \].

By the definition of \( M \) we see that
\[ N/(2p) - (p + 3) < (M + 1)(p + 3), \]
thus the intervals overlap and \( B_0 + (B_1 \cup B_2) \) forms a 12-chain in
\[ [0, N/2 + M(p + 3)]. \]

Finally, we consider \( B_2 + B_2 \). It forms a 6-chain in 
\[ [N/p - 2(p + 3), N/p] \]
which overlaps with the previous interval, so together they form a 18-chain
in \([0, N/p]\) as required.

4. With small gaps through a long interval. We show that if instead of the whole interval \([0, 2N]\) we are content with a positive portion, then the \( N^{1/3} \) of the previous theorem can be reduced to \( \log N \).

**Theorem 4.1.** For all \( c < 1/5 \) and sufficiently large \( N \) there is a Sidon set \( A \subset [0, N] \) with the property that \( A + A \) forms a \( C \log N \)-chain in the interval \([N, (1 + c)N]\). Here \( C \) is a positive absolute constant.

The proof of this theorem is based on a different construction of a Sidon set, which we describe below.

Let \( p \) be a prime, \( g \) a primitive root modulo \( p \) and write \( q = p(p - 1) \).
For each \( 1 \leq i \leq p - 1 \) let \( a_i \) denote the solution of the congruence
\[ a_i \equiv i \pmod{p - 1}, \quad a_i \equiv i \pmod{p}, \quad 1 \leq a_i \leq q. \]
The set \( B = \{a_i\} \) forms a Sidon set modulo \( q \), that is, the sums \( a_i + a_j \) have all distinct residues modulo \( q \) [4, Theorem 4.4].

We need the following additional property of \( B \).

**Lemma 4.2.** For a suitable choice of \( g \) no interval of length \( M = \phi(p - 1)^{1/3} \) contains more than two numbers whose residues modulo \( q \) are elements of \( B \).

**Proof.** All elements of \( B \) satisfy \( g^b \equiv b \pmod{p} \). Hence if there are three in an interval of length \( M \), say \( a, a + u, a + v \) with \( 0 < u < v \leq M \), then the congruences
\[ g^a \equiv a, \quad g^{a+u} \equiv a + u, \quad g^{a+v} \equiv a + v \pmod{p} \]
hold. On substituting the first into the others we obtain
\[ a(g^u - 1) \equiv v, \quad a(g^v - 1) \equiv u \pmod{p}, \]
hence (observe that \( a \equiv g^a \neq 0 \))
\[ u(g^u - 1) \equiv v(g^v - 1) \pmod{p}. \]
For fixed $u,v$ this is an equation of degree $v$ in $g$, hence has at most $v$ solutions. By summing this for all pairs $u,v$ we conclude that there are less than $M^3$ values of $g$ for which such triplets exist. Since there are altogether $\phi(p-1) = M^3$ primitive roots, there must be a value of $g$ for which no such triplet exists.

Though it is likely that other dense Sidon sets, constructed via finite fields, also have a similar property, we were unable to establish it.

**Proof of Theorem 4.1.** Let $p$ be the largest prime satisfying $5p(p-1) \leq N$. We consider the set $B$ described above, with a $g$ as provided by Lemma 4.2.

We divide $B$ into three subsets $B_1, B_2, B_3$ randomly, that is, all $3^{p-1}$ partitions are considered with equal probability. We put

$$A = B_1 \cup (B_2 + q) \cup (5q - B_3) \subset [1, 5q] \subset [1, N].$$

First we show that $A$ is a Sidon set for each partition. Suppose that $A$ contains four elements $x, y, u, v$ satisfying $x + y = u + v$. We call $B_1 \cup (B_2 + q)$ the lower half and $5q - B_3$ the upper half of $A$.

If all four are from the lower half or all from the upper half, then this would violate the Sidon property of the residues modulo $q$.

If one is from the lower and three from the upper half, or three from the lower and one from the upper one, then we get a contradiction by comparing the magnitudes.

If two variables come from each half, then there are two possibilities. If $x, y$ are from one half and $u, v$ from the other, then again the magnitude of the sides leads to a contradiction. Assume finally that both sides contain a number from the lower and one from the upper half, say $x, u$ from the lower and $y, v$ from the upper. The residues of $x, u, -y, -v$ are elements of $A$ and they satisfy

$$x + (-v) \equiv (-y) + u \pmod{q},$$

which again contradicts the Sidon property of $A$ modulo $q$.

Now we begin to establish the chain property.

The numbers $a_i - a_j, i \neq j,$ are all incongruent modulo $q$, and none of them is divisible by $p$ or $p-1$. Their number is $(p-1)(p-2)$, which is the same as the total number of residues modulo $q$ that are not divisible by $p$ or $p-1$. Hence for every $u$ such that $p \mid u$ and $p-1 \nmid u$ there is exactly one pair $i, j$ such that

\[ a_i - a_j \equiv u \pmod{q} .\tag{4.1} \]

In particular, if $1 \leq u \leq q$, then there is a pair $i, j$ such that

$$a_i - a_j = u \quad \text{or} \quad a_i - a_j = u - q.$$
If the first case holds, then we have
\[ 5q + u = a_i + (5q - a_j), \]
hence \( 5q + u \in A + A \) if \( a_i \in A_1 \) and \( a_j \in A_3 \). In the second case we have
\[ 5q + u = (a_i + q) + (5q - a_j), \]
hence \( 5q + u \in A + A \) if \( a_i \in A_2 \) and \( a_j \in A_3 \). In both cases
\[ \text{Prob}(5q + u \in A + A) = 1/9. \]

Now take any interval \( (s, s + t] \) of length \( t = [C \log N] \) contained in \([5q, 6q]\). In this interval there may be at most one multiple of \( p \) and one of \( p - 1 \); each other has a chance \( 1/9 \) of being in \( A + A \). These events are not independent; we can claim independence only if the numbers \( a_i, a_j \) used in the representations (4.1) are all distinct. For a fixed \( n = 5q + u \in (s, s + t] \) we have to exclude those numbers that are in \( a_i - B, a_j - B, B - a_i \) or \( B - a_j \) modulo \( q \). By Lemma 4.2 each of these sets has at most 2 elements in an interval of length \( t < M \) (we have \( M > p^{1/3 - \varepsilon} \) by the familiar estimates for the \( \phi \) function). Thus for any \( n \) there are at most 8 other numbers that can spoil the independence. By the greedy algorithm we find \( (t - 2)/9 \) numbers in \( (s, s + t] \), none divisible by \( p \) or \( p - 1 \), such that all the \( a_i, a_j \) in their representations (4.1) are distinct. Hence the probability that none of them is in \( A + A \) is less than \( (8/9)^{(t-2)/9} < 1/N \) if \( C \) is large enough. Consequently, with positive probability this does not happen for any choice of \( s \), which means that \( A + A \) forms a \( C \log N \)-chain in \([5q, 6q] \supset [N, (6/5 - \varepsilon)N] \).

Acknowledgements. This work was stimulated by conversations with Prof. P. Erdős.

References