Canad. Math. Bull. Vol. **33** (3), 1990

SIDON SETS

BY

H. L. ABBOTT

ABSTRACT. Denote by $g(n)$ the largest integer m such that every set of integers of size n contains a subset of size m whose pairwise sums are distinct. It is shown that $g(n) > cn^{1/2}$ for any constant $c < 2/25$ and all sufficiently large n.

A set S of integers is called a Sidon set if its pairwise sums are distinct; that is, if it contains no solutions of the equation $a+b = c+d$ other than the trivial solutions given by $\{a, b\} = \{c, d\}$. S. Sidon [9], [10] was led to consider such sets in connection with certain questions in analysis. We remark that in the literature Sidon sequences are sometimes called B_2-sequences and that the phrase Sidon set is used in the more recent literature on harmonic analysis with a different meaning. See, for example, [12]. We shall be concerned here only with number-theoretic questions.

Denote by $f(n)$ the size of a largest Sidon subset of $\{1, 2, \ldots, n\}$. It follows from results of Bose [2], Bose and Chowla [3], Chowla [4], Erdös and Turán [5], and Singer [11] that

$$f(n) = n^{1/2}(1 + o(1)), \text{ as } n \longrightarrow \infty.$$

A comprehensive survey of these papers may be found in Chapter II of the book of Halberstam and Roth [6]. [6] also gives an account of the main results concerning infinite Sidon sets. See the paper of Ajtai, Komlós and Szemerédi [1] for an important development in this regard.

If A is a finite set of integers we denote by $\|A\|$ the size of a largest Sidon subset of A, and we write $\|\{a_i\}_1^n\|$ instead of $\|\{a_1, a_2, \ldots, a_n\}\|$. Let g be defined by

$$g(n) = \min\{\|A\| : A \text{ is a set of integers of size } n\}.$$

$g(n)$ is thus the largest integer m such that every set of integers of size n contains a Sidon subset of size m. It is clear that $g(n) \leqq f(n)$. Erdös asked whether $\lim_{n\to\infty}(g(n)/f(n)) = 1$. This question has not been answered. For a long time the best lower bound for $g(n)$ was that of Mian and Chowla [8] who showed that $g(n) > cn^{1/3}$ for some positive constant c. Komlós, Sulyok and Szemerédi [7] made striking progress toward answering Erdös' question by showing that there exists a constant $c > 0$ such that

$$g(n) > cn^{1/2} \tag{1}$$

Received by the editors March 20, 1989 and, in revised form August 15, 1989.
AMS (1985) Subject Classification: 11B75.

for all sufficiently large n. In fact, they prove a general theorem of which (1) is a special case. It follows from their theorem that (1) holds for any $c < 2^{-15} = 0.0000305\ldots$. The object of this paper is to prove the following result:

THEOREM *(1) holds for any* $c \leqq 2/25 = 0.08$.

Some of the improvement is obtained by refining the method used in [7] and part of it by exploiting the connection between Sidon sets and a result of Singer [11]. See the remarks at the end of the paper.

We first record four facts that we shall need later. Facts 1, 2 and 4 are given in [7] in some form. Fact 3 is easy to prove. Let $a_1, a_2, \ldots, a_n$ be distinct integers.

FACT 1. If there is a positive integer q and distinct integers $r_1, r_2, \ldots, r_n$ such that $a_i = h_i q + r_i$, $|r_i| < q/4$, then $\|\{a_i\}_1^n\| \geqq \{r_i\}_1^n\|$.

FACT 2. Suppose there exists an odd positive integer q such that $a_i = h_i q + r_i$, $|r_i| < q/4$. Let t be a positive integer and let $b_i = th_i q + r_i$. Then $\|\{a_i\}_1^n\| \geqq \|\{b_i\}_1^n\|$.

FACT 3. Let p be a prime, $p \equiv 1 (\mathrm{mod}\ 4)$. Then there exists an integer λ and a subset A of $\{a_1, a_2, \ldots, a_n\}$ of size $[n/2]$ such that for $a_i \in A$ we have $a_i + \lambda = h_i p + r_i$, where $|r_i| < p/4$.

FACT 4. Let $p \geqq n$ be a prime that does not divide any of the differences $a_i - a_j$, $i \neq j$. Then there exists an integer t, $1 \leqq t \leqq p-1$, and a subset A of $\{a_1, a_2, \ldots, a_n\}$ of size $[n/2] - 2$ such that for each $a_i \in A$ we have $ta_i = h_i p + r_i$ where $|r_i| < p/4$ and the r's are distinct.

LEMMA A. *Let* $0 < a_1 < a_2 < \cdots < a_n$, $a_n > (4^n)^{4^n}$. *Then there exists an integer* $q < a_n$ *and distinct integers* $r_1, r_2, \ldots, r_n$ *such that* $a_i = h_i q + r_i$, $|r_i| < q/4$.

Lemma A is a special case of Lemma $1'$ of [7]. We observe that by Fact 1 and repeated application of Lemma A, an arbitrary set $\{a_1, a_2, \ldots, a_n\}$ of n positive integers may be replaced by a set $\{b_1, b_2, \ldots, b_n\}$, $0 < b_1 < b_2 < \cdots < b_n \leqq (4^n)^{4^n}$, that satisfies $\|\{a_i\}_1^n\| \geqq \|\{b_i\}_1^n\|$. In what follows all o-estimates refer to $n \to \infty$.

LEMMA B. *Let* $0 < a_1 < a_2 < \cdots < a_n \leqq (4^n)^{4^n}$. *There exist integers* $0 < b_1 < b_2 < \cdots < b_m$ *such that*

$$b_m < 4^n n^2 \log n, \quad m = \frac{n}{2}(1 + o(1)) \text{ and } \|\{a_i\}_1^n\| \geqq \|\{b_i\}_1^m\|.$$

PROOF. For primes p satisfying $4^n < p < 4^n n^2 \log n$ let $\psi(p)$ denote the number of pairs (i, j), $1 \leqq i < j \leqq n$, such that $p \mid (a_j - a_i)$ and define T by

$$T = \sum_p \psi(p).$$

Then

$$(4^n)^T < \prod_{1 \leqq i < j \leqq n} (a_j - a_i) < (4^n)^{4^n \binom{n}{2}}$$

https://doi.org/10.4153/CMB-1990-056-6 Published online by Cambridge University Press

from which it follows that

$$T < \left(\frac{1}{2} + o(1)\right) 4^n n^2.$$

Thus, for some prime q satisfying $4^n < q < 4^n n^2 \log n$, we must have

$$\psi(q) < \frac{\left(\frac{1}{2} + o(1)\right) 4^n n^2}{\pi(4^n n^2 \log n) - \pi(4^n)}$$

where $\pi(x)$ denotes the number of primes not exceeding x. It follows from the prime number theorem and some routine calculations that

$$\psi(q) < \left(\frac{\log 4}{2} + o(1)\right) \frac{n}{\log n} < \frac{n}{\log n}.$$

Thus there exists a subset A of $\{a_1, a_2, \ldots, a_n\}$ of size $\nu > n(1 - 1/\log n)$ such that q does not divide the difference of any two members of A. By Fact 4, there exists an integer $t, 1 \leqq t \leqq q - 1$, and a subset A^* of A of size $m = [\nu/2] - 2$ such that for each $a_i \in A^*$ we have $ta_i = h_i q + r_i$ where $|r_i| < q/4$ and the r's are distinct. Choose as $b_1, b_2, \ldots, b_m$ the integers obtained by translating the r's by $[q/4] + 1$. Then

$$0 < b_1 < b_2 < \cdots < b_m < q < 4^n n^2 \log n,$$

$$m = \frac{n}{2}(1 + o(1))$$

and

$$\begin{aligned} \|\{a_i\}_1^n\| = \|\{ta_i\}_1^n\| &\geqq \|\{ta_i : a_i \in A^*\}\| \\ &\geqq \|\{r_i : a_i \in A^*\}\|, \text{ by Fact 1} \\ &= \|\{b_i\}_1^m\|. \end{aligned}$$

LEMMA C. *Let $\{b_1, b_2, \ldots, b_m\}$ be the set whose existence was shown in Lemma B. Let $\delta > (\log 2)/4$. Then there exist integers $0 < c_1 < c_2 < \cdots < c_l$ such that*

$$c_l < \delta n^2, \; l = \frac{n}{4}\left(1 - \frac{\log 2}{4\delta} + o(1)\right) \text{ and } \|\{b_i\}_1^m\| \geqq \|\{c_i\}_1^l\|.$$

PROOF. For primes p satisfying $n^2/\log n < p < 2\delta n^2$ let $\psi(p)$ denote the number of pairs (i, j), $1 \leqq i < j \leqq m$ such that $p \mid (b_j - b_i)$, and define T by

$$T = \sum_p \psi(p).$$

Then

$$\left(\frac{n^2}{\log n}\right)^T < \prod_{1 \leqq i < j \leqq m} (b_j - b_i) < (4^n n^2 \log n)^{\binom{m}{2}}$$

from which it follows that

$$T < \left(\frac{\log 2}{8} + o(1)\right)\frac{n^3}{\log n}.$$

Thus, for some prime q satisfying $n^2/\log n < q < 2\delta n^2$ we must have

$$\psi(q) < \frac{\left(\frac{\log 2}{8} + o(1)\right)\frac{n^3}{\log n}}{\pi(2\delta n^2) - \pi\left(\frac{n^2}{\log n}\right)}.$$

It follows from the prime number theorem and some straightforward calculations that

$$\psi(q) < \left(\frac{\log 2}{8\delta} + o(1)\right)n = \left(\frac{\log 2}{4\delta} + o(1)\right)m.$$

Thus there exists a subset B of $\{b_1, b_2, \ldots, b_m\}$ of size $\nu = (1 - \log 2/4\delta + o(1))m$ such that q does not divide the difference of any two members of B. By Fact 4, there exists an integer t, $1 \leqq t \leqq q-1$, and a subset B^* of B of size $l = [\nu/2] - 2$ such that for each $b_i \in B^*$ we have $tb_i = h_i q + r_i$ where $|r_i| < q/4$ and the r's are distinct. Let $c_1, c_2, \ldots, c_l$ be the numbers obtained by translating the r's by $[q/4] + 1$. Then

$$0 < c_1 < c_2 < \cdots < c_l < \frac{q}{2} < \delta n^2, \quad l = \frac{n}{4}\left(1 - \frac{\log 2}{4\delta} + o(1)\right) \text{ and,}$$

by Fact 1, $\|\{b_i\}_1^m\| \geqq \|\{c_i\}_1^l\|$. □

LEMMA D. *Let $\{c_1, c_2, \ldots, c_l\}$ be the set constructed in Lemma C. Let β be a positive number satisfying $\beta^2 > \delta/2$. Let*

$$\gamma = (256\beta\delta^2 - 64\beta\delta \log 2 - 16\delta^2 + 8\delta \log 2 - (\log 2)^2)/4096\beta\delta^2.$$

Then there exist integers $0 < d_1 < d_2 < \cdots < d_s$ satisfying

$$d_s < \frac{\beta n}{2}, \; s = (\gamma + o(1))n \text{ and } \|\{c_i\}_1^l\| \geqq \|\{d_i\}_1^s\|.$$

PROOF. Let q be the least prime exceeding βn and let p be the largest prime not exceeding $2q$ satisfying $p \equiv 1 (\text{mod } 4)$. Observe that $pq = 2\beta^2 n^2(1 + o(1))$. By Fact 3, there exists an integer λ and a subset C of $\{c_1, c_2, \ldots, c_l\}$ of size $t = [l/2]$ such that for $c_i \in C$ we have $c_i + \lambda = h_i p + r_i$ where $|r_i| < p/4$. There is no loss in assuming (by relabelling, if neccessary) that $C = \{c_1, c_2, \ldots, c_t\}$. For $1 \leqq u \leqq q-1$ and $1 \leqq i \leqq t$ let $e(u, i) = uh_i p + r_i$. By Fact 2, for each u,

$$\|\{e(u, i)\}_1^t\| \geqq \|\{c_i\}_1^t\|.$$

https://doi.org/10.4153/CMB-1990-056-6 Published online by Cambridge University Press

For $1 \leqq j < i \leqq t$ and $1 \leqq u \leqq q-1$ let

$$\psi_u(i, j) = \begin{cases} 1 & \text{if } q \mid e(u, i) - e(u, j) \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that for some pairs (u, v) and (i, j), $1 \leqq u < v \leqq q-1$, $1 \leqq j < i \leqq t$, we have

$$\psi_u(i, j) = \psi_v(i, j) = 1.$$

Then $q \mid p(h_i - h_j)(v-u)$ and this implies that $q \mid h_i - h_j$. Now $h_i = h_j$ implies that $q \mid r_i - r_j$. Since $0 \leqq |r_i - r_j| < p/2 < q$ we must then have $r_i = r_j$ and thus $c_i = c_j$, a contradiction. Thus $h_i \neq h_j$. If follows that $|h_i - h_j| \geqq q$ so that we get

$$|c_i - c_j| = |(c_i + \lambda) - (c_j + \lambda)| = |(h_i - h_j)p + (r_i - r_j)| \geqq 2\beta^2 n^2(1 + o(1)).$$

However, c_i and c_j lie in $[1, \delta n^2]$ so that $|c_i - c_j| < \delta n^2$. Since $2\beta^2 > \delta$ this yields a contradiction. Thus for each pair (i, j), $1 \leqq j < i \leqq t$, there is at most one u, $1 \leqq u \leqq q-1$, such that $\psi_u(i, j) = 1$. Thus

$$\sum_{u=1}^{q-1} \sum_{1 \leqq j < i \leqq t} \psi_u(i, j) = \sum_{1 \leqq j < i \leqq t} \sum_{u=1}^{q-1} \psi_u(i, j) \leqq \binom{t}{2}.$$

It follows that for some u, $1 \leqq u \leqq q-1$,

$$\sum_{1 \leqq j < i \leqq t} \psi_u(i, j) \leqq \frac{1}{q-1} \binom{t}{2}.$$

Let

$$z = \left[\frac{1}{q-1} \binom{t}{2} \right].$$

Then there is a subset S of $\{1, 2, \ldots, t\}$ of size $\nu = t - z$ such that for $(i, j) \in S$, $i > j$, q does not divide $e(u, i) - e(u, j)$. There is no loss in assuming that $S = \{1, 2, \ldots, \nu\}$. By Fact 4, for some w, $1 \leqq w \leqq q-1$, there is a subset E of $\{e(u, 1), e(u, 2), \ldots, e(u, \nu)\}$ of size $s = [\nu/2] - 2$ such that for $e(u, i) \in E$ we have $we(u, i) = h_i' q + r_i'$ where the integers r_i' satisfy $|r_i'| < q/4$ and are distinct. Again, there is no loss of generality in supposing that $E = \{e(u, 1), e(u, 2), \ldots, e(u, s)\}$. Let $0 < d_1 < d_2 < \cdots < d_s$ be the numbers obtained by translating $r_1', r_2', \ldots, r_s'$ by $[q/4] + 1$. Then $d_s < q/2 < \beta n/2$. Some straightforward calculations show that $s = (\gamma + o(1))n$. Furthermore, by Fact 1, $\|\{c_i\}_1^t\| \geqq \|\{d_i\}_1^s\|$.

We need a further lemma; namely, the following result of Singer [11]. See also [6], Chapter II.

LEMMA E. *Let p be a prime. Then there exist $p+1$ Sidon sets, each of size $p+1$, whose union is $\{1, 2, \ldots, p^2 + p + 1\}$.*

https://doi.org/10.4153/CMB-1990-056-6 Published online by Cambridge University Press

The proof of the Theorem may now be completed as follows. Let $S = \{d_1, d_2, \ldots, d_s\}$ be the set constructed in Lemma D. Let p be the least prime such that $p^2+p+1 > d_s$. By Lemma E, there are $p+1$ Sidon sets whose union is $\{1, 2, \ldots, p^2+p+1\}$. One of these sets must contain at least $s/(p+1)$ members of S. Thus

$$g(n) \geqq \frac{s}{p+1} \geqq \left(\gamma\left(\frac{2}{\beta}\right)^{1/2} + o(1)\right) n^{1/2}$$

$$= \frac{\sqrt{2}}{4096}\left\{\frac{256\beta\delta^2 - (64\log 2)\beta\delta - 16\delta^2 + (8\log 2)\delta - (\log 2)^2}{\beta^{3/2}\delta^2} + o(1)\right\}\sqrt{n}.$$

The only restrictions on β and δ are $\delta > \log 2/4$ and $2\beta^2 > \delta$. It is legitimate to choose $\beta = 0.6834$ and $\delta = 0.9340$. We then get $g(n) > 0.0805\sqrt{n}$. This completes the proof of the Theorem. □

Readers familiar with [7] will have noticed that the arguments used in proving Lemmas B, C and D are based very heavily on the ideas and the techniques developed in that paper. Note, however, that it requires three applications of Lemma 2 of [7] to reduce an arbitrary set of size n in $[1, (4^n)^{4^n}]$ to a set (of size $(n/64)(1+o(1))$) in $[1, 4n^2(\log n)^2]$. Lemma 3 of [7] effects a reduction to $[1, n^{3/2}]$ and Lemma 4 a reduction to $[1, cn]$, c a constant. Roughly, we have replaced these five reductions by three. Lemmas B and C effect a reduction from $[1, (4^n)^{4^n}]$ to $[1, \delta n^2]$ and thus accomplish a little more than the three applications of Lemma 2 of [7]. Lemma D takes one from $[1, \delta n^2]$ to $[1, \beta n/2]$.

We conclude with some remarks about the more general question considered in [7].

Let $a_1, a_2, \ldots, a_L$ be integers whose sum is zero. Suppose that not all of the a_i are zero and let $A_1 = \{i : a_i > 0\}$ and $A_2 = \{i : a_i < 0\}$. We call a solution $x_1, x_2, \ldots, x_L$ of $a_1x_1 + a_2x_2 + \cdots + a_Lx_L = 0$ trivial if $\{x_i : i \in A_i\} \cap \{x_i : i \in A_2\} \neq \phi$. Consider the following system of equations with integer coefficients:

$$(\rho) \qquad \sum_{i=1}^{L} a_{ij}x_i = 0, \quad j = 1, 2, \ldots, R.$$

Suppose that $\sum_{i=1}^{L} a_{ij} = 0$ for $j = 1, 2, \ldots, R$. We call a solution $x_1, x_2, \ldots, x_L$ of (ρ) trivial if it is a trivial solution of at least one of the equations of the system. Let $f_\rho(n)$ denote the size of a largest subset of $\{1, 2, \ldots, n\}$ containing only trivial solutions of (ρ) and let $g_\rho(n)$ denote the largest integer m such that every set of integers of size n contains a subset of size m containing only trivial solutions of (ρ). Let

$$\alpha = \max_{1 \leqq j \leqq R} \sum_{i=1}^{L} |a_{ij}|.$$

In [7] it is proved that

$$g_\rho(n) \geqq \frac{1}{8\alpha^6} f_\rho(n).$$

https://doi.org/10.4153/CMB-1990-056-6 Published online by Cambridge University Press

Note that in the case where (ρ) is $x_1 + x_2 - x_3 - x_4 = 0$ we have $\alpha = 4$ and thus (1). We can prove analogues of Lemmas B, C and D. From these and Lemma 6 of [7] it may be deduced that

$$g_\rho(n) > \frac{1}{2\alpha^2}\left\{1 - \frac{2}{\alpha^3} - \frac{4\log\alpha}{\alpha^4}\right\} f_\rho(n).$$

In the case of Sidon sets, this is weaker than the bound given by the Theorem and thus illustrates the effect of Singer's Theorem, which is not available in the general case. The question, raised in [7], as to whether there is an absolute constant c (independent of (ρ)) such that $g_\rho(n) > cf_\rho(n)$ remains open.

References

1. M. Ajtai, J. Komlós, E. Szemerédi, *A dense infinite Sidon sequence*, European Journal of Combinatorics, **2** (1981), 1–11.

2. R. C. Bose, *An affine analogue of Singer's Theorem*, Journal of the Indian Math. Soc., **6** (1942), 1–15.

3. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helvet., **37** (1962-63), 141–147.

4. S. Chowla, *Solution of a problem of Erdös and Turán in additive number theory*, Proc. Nat. Acad. Sci. India, **14** (1944), 1–2.

5. P. Erdös and P. Turán, *On a problem of Sidon in additive number theory and some related problems*, Jour. Lond. Math. Soc., **16** (1941), 212–215. addendum, ibid. **19** (1944), 208.

6. H. Halberstam and K. F. Roth, *Sequences*, Oxford University Press, 1966.

7. J. Komlós, M. Sulyok and E. Szemerédi, *Linear problems in combinatorial number theory*, Acta Math. Acad. Sci. Hung., **26** (1975), 113–121.

8. A. Mian and S. Chowla, *On the B_2-squences of Sidon*, Proc. Nat. Acad. Sci. India, **14** (1944), 3–4.

9. S. Sidon, *Ein Satz uber trigonomietrische Polynome und seine Anwendungen in der Theorie der Fourier-Reihen*, Math. Ann., **106** (1932), 536–539.

10. ———, *Uber die Fourier Konstanten der Funktionen der Klasse L_p für $p > 1$*, Acta Sci. Math. Szeged, **7** (1935), 175–176.

11. J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc., **43** (1938), 377–385.

12. W. Rudin, *Trigonometric series with gaps*, J. Math. Mech. **9** (1960), 203–227.

Department of Mathematics
University of Alberta
Edmonton, Alberta, Canada T6G 2G1