Combinatorics, Probability and Computing (2007) 16, 775–787. © 2007 Cambridge University Press
doi:10.1017/S0963548307008401 First published online 28 February 2007 Printed in the United Kingdom

Difference Covers

TOMASZ SCHOEN*

Department of Discrete Mathematics, Adam Mickiewicz University,
ul. Umultowska 87, 61-614 Poznań, Poland
(e-mail: schoen@amu.edu.pl)

Received 31 May 2005; revised 27 November 2006

Let S be a finite set of integers. We consider a problem of finding $D(S)$, the minimum size of a set A, such that $S \subseteq A - A$. We give a characterization for 'extremal' sets and prove lower and upper bounds on $D(S)$ in terms of additive properties of S.

1. Introduction

Let

$$A - A = \{a - a' : a, a' \in A\}$$

be the *difference set* of a set $A \subseteq \mathbb{N}$ and put $(A - A)_+ = (A - A) \cap \mathbb{N}$. More than 50 years ago Redei and Rényi [8] showed that every set $A \subseteq \mathbb{Z}$ satisfying

$$\{1, \ldots, n\} \subseteq A - A$$

contains at least $(2.43n)^{1/2}$ numbers and they constructed such a set with $((8/3 + o(1))n)^{1/2}$ elements.

In this paper we consider the general setting of the problem. For any set $S \subseteq \{1, \ldots, n\}$ and $A \subseteq \mathbb{Z}$ such that $S \subseteq A - A$, we call A a *difference cover* of S and we put

$$D(S) = \min\{|A| : S \subseteq A - A,\ A \subseteq \mathbb{Z}\}.$$

We have the following trivial bounds:

$$\frac{\sqrt{8|S| + 1} + 1}{2} \leqslant D(S) \leqslant |S| + 1. \tag{1.1}$$

The lower bound follows from the inequality $|(A - A)_+| \leqslant \frac{|A|(|A|-1)}{2}$ and the second one follows from the fact that the set $A = S \cup \{0\}$ is a difference cover of S. We show that

* Research partially supported by KBN Grant 2 P03A 007 24.

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

all the above trivial bounds can be achieved and we give a characterization of extremal sets. Moreover, we estimate $D(S)$ in terms of additive properties of S, such as the size of the difference set $S-S$, the number of solutions to the equation $x_1-x_2=y_1-y_2$ or the maximum number of representations of an integer as a difference of two elements for S. We also give a bound on $D(S)$ for 'typical sets'. It is worth mentioning that difference covers are used in theoretical computer science [4].

2. Extremal sets

In this section we determine all extremal sets, *i.e.*, sets satisfying $D(S)=(\sqrt{8|S|+1}+1)/2$ or $D(S)=|S|+1$. Let us recall that $A\subseteq\mathbb{Z}$ is a *Sidon set* if every natural number has at most one representation in the form $a-a'$, for $a,a'\in A$ or, equivalently, $|A-A|=|A|(|A|-1)+1$. We call a set $S\subseteq\mathbb{N}$ *independent* if all sums

$$\sum_{s\in T} s,\ \ T\subseteq S$$

are distinct (for example, the set $\{1,2,\dots,2^k,\dots\}$ is independent). Clearly, every independent subset of $\{1,\dots,n\}$ has at most $(1+o(1))\log_2 n$ elements.

Theorem 2.1. *$S\subseteq\{1,\dots,n\}$. Then*

(a) *$D(S)=(\sqrt{8|S|+1}+1)/2$ if and only if $S=(A-A)_+$, for some Sidon set A,*
(b) *$D(S)=|S|+1$ if and only if S is an independent set.*

Proof. Let A be a difference cover of S of size $D(S)$. Observe that $D(S)=(\sqrt{8|S|+1}+1)/2$ if and only if every element of S has exactly one representation in the form $a-a'$, $a,a'\in A$ and $(A-A)_+\subseteq S$ which means that A is a Sidon set and $S=(A-A)_+$.

To prove the second part of the theorem suppose that A is a set satisfying $S\subseteq A-A$, where S is an independent set. For every integer $s\in S$ we choose exactly one pair $a_s,a'_s\in A$ with $s=a_s-a'_s$, and define a graph $H=(A,E)$ by

$$E=\{\{a_s,a'_s\}:s\in S\},$$

hence $|E|=|S|$. We show that H does not contain any cycle. Suppose, to the contrary, that there exists a cycle $C=\{a_1,\dots,a_k\}$ in H. By definition of H there are distinct elements $s_1,\dots,s_k\in S$ and a suitable choice of signs $\pm$ such that

$$\sum_{i=1}^{k}\pm s_i=0.$$

Hence there is a partition $U\cup V=\{1,\dots,k\}$ such that

$$\sum_{u\in U} s_u=\sum_{v\in V} s_v,$$

which contradicts that S is an independent set. Thus, H has no cycle, whence

$$|S|=|E|\leqslant|A|-1,$$

as desired.

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

Now assume that S is not an independent set. Hence there are disjoint non-empty subsets $T, T' \subseteq S$ satisfying

$$\sum_{s\in T} s = \sum_{s\in T'} s.$$

We define a set A as follows. Put $T = \{s_{i_1}, \dots, s_{i_{|T|}}\}$ and $T' = \{s_{j_1}, \dots, s_{j_{|T'|}}\}$. Furthermore, let

$$B = \left\{ \sum_{k=1}^{\ell} s_{i_k} : \ell = 1, \dots, |T| \right\}$$

and

$$B' = \left\{ \sum_{k=1}^{\ell} s_{j_k} : \ell = 1, \dots, |T'| \right\}.$$

Observe that $\sum_{s\in T} s = \sum_{s\in T'} s$ implies $|B \cup B'| < |T| + |T'|$. Finally, define

$$A = \{0\} \cup B \cup B' \cup (S \setminus (T \cup T')).$$

Clearly $T \subseteq B - (B \cup \{0\})$ and $T' \subseteq B' - (B' \cup \{0\})$. Since $0 \in A$ it follows that $S \setminus (T \cup T') \subseteq A - A$, so $S \subseteq A - A$ and

$$|A| \leqslant 1 + |B \cup B'| + |S| - |T| - |T'| \leqslant |S|. \qquad \square$$

3. Typical sets

Here we are interested in estimating $D(S)$ for a typical (random) set S. Using a simple counting argument, we show that for a typical set S we have $D(S) = \min(|S|^{1-o(1)}, n^{1/2-o(1)})$. Below, by $\log x$ we mean $\log_e x$.

Theorem 3.1. *Let $\varepsilon \in (0, 1]$ and let $n > n_\varepsilon$ be a natural number. Then, for every $m \in \mathbb{N}$ with $1 \leqslant m \leqslant n$ there are more than $(1-\varepsilon)\binom{n}{m}$ sets $S \subseteq \{1, \dots, n\}$ of cardinality m, satisfying*

$$D(S) > \frac{a}{2},$$

where a is any integer such that

$$a < m\frac{\log(2n/a^2)}{\log(2en/a)} - \frac{\log(a/\varepsilon)}{\log(2en/a)}.$$

Proof. Observe that if A is a difference cover of S then we can find a new difference cover A' of S such that $|A'| \leqslant 2|A|$ and $A' \subseteq \{0, 1, \dots, 2n\}$. Indeed, putting

$$A' = \bigcup_{k\in\mathbb{Z}} ((A \cap \{kn+1, kn+2, \dots, (k+2)n\}) - kn),$$

we see that $(A' - A') \cap \{1, \dots, n\} = (A - A) \cap \{1, \dots, n\}$ and $|A'| \leqslant 2|A|$.

We can assume that $\sqrt{2m} \leqslant a \leqslant 2\sqrt{n}$. Note that there are $\sum_{i=1}^{a}\binom{2n}{i} \leqslant a\binom{2n}{a}$ non-empty subsets of $\{1, \dots, 2n\}$ with at most a elements. Each a-element set has at most $\binom{a}{2}$ differences in $\{1, \dots, n\}$, so that every such a set can be a difference cover for at most $\binom{\binom{a}{2}}{m}$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

m-element subsets of $\{1,\dots,n\}$. Finally the number of m-element subsets contained in at least one such difference set does not exceed $a\binom{2n}{a}\binom{\binom{a}{2}}{m}$. On the other hand we have $\binom{n}{m}$ subsets of $\{1,\dots,n\}$ of cardinality m. Therefore, if

$$a\binom{2n}{a}\binom{\binom{a}{2}}{m} < \varepsilon\binom{n}{m}, \tag{3.1}$$

or equivalently

$$a\binom{2n}{a}\frac{\binom{a}{2}(\binom{a}{2}-1)\dots(\binom{a}{2}-m+1)}{n\dots(n-m+1)} < \varepsilon,$$

then there are more than $(1-\varepsilon)\binom{n}{m}$ sets $S \subseteq \{1,\dots,n\}$ such that $|S| = m$ and $D(S) > a/2$. Since $\binom{n}{k} \leqslant (en/k)^k$ it follows that

$$a\binom{2n}{a} \leqslant a\left(\frac{2en}{a}\right)^a.$$

Note that, if the inequality

$$a\left(\frac{2en}{a}\right)^a\left(\frac{a^2}{2n}\right)^m < \varepsilon \tag{3.2}$$

holds, then (3.1) is also satisfied. Inequality (3.2) is equivalent to

$$a < m\frac{\log(2n/a^2)}{\log(2en/a)} - \frac{\log(a/\varepsilon)}{\log(2en/a)}. \tag{3.3}$$

Thus, (3.1) holds for every a fulfilling the above inequality. □

The next corollary shows the existence of 'large' sets with almost maximal size of difference cover. It provides a less exact but more explicit bound on $D(S)$.

Corollary 3.2. *Let n be a natural number. Then, for every positive integer m with $6 \leqslant m \leqslant \sqrt{n}$ there exists a set $S \subseteq \{1,\dots,n\}$ of cardinality m satisfying*

$$D(S) > \frac{1}{2e}\left(1 - \frac{\log m}{\log n/m}\right)m - \frac{1}{2}.$$

Proof. We apply Theorem 3.1 with $\varepsilon = 1$. Observe that if c is such that $m = n^c$ then (3.3) holds for $a = \frac{1}{e}\frac{1-2c}{1-c}m - 1$. Indeed,

$$\begin{aligned} m\frac{\log(2n/a^2)}{\log(2en/a)} - \frac{\ln a}{\log(2en/a)} &= \left(1 - \frac{\ln ea}{\ln 2en/a}\right)m - \frac{\ln a}{\log(2en/a)} \\ &\geqslant \left(1 - \frac{\log n^c}{\log n^{1-c}}\right)m - \frac{\log(n^c)}{\log(n^{1-c})} \\ &> \left(1 - \frac{c}{1-c}\right)m - 1 \geqslant a. \end{aligned}$$

Hence, by Theorem 3.1 there is a set $S \subseteq \{1,\dots,n\}$ of cardinality m with

$$D(S) > a/2 \geqslant \frac{1}{2e}\left(1 - \frac{\log m}{\log n/m}\right)m - \frac{1}{2},$$

which completes the proof. □

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

4. Lower bounds

We give here some examples of 'dense' sets with large difference covers. Natural candidates are additively independent sets, such as Sidon sets. We show that for every Sidon set S we have $D(S) = \Omega(|S|^{2/3})$ and, slightly surprisingly, in general one cannot improve this bound. We start with a general result providing a lower bound on $D(S)$ in terms of two basic additive parameters:

$$M = M(S) = \max_{t \neq 0} v(t),$$

where $v(t) = v_S(t)$ denotes the number of representations $t = s - s'$, $s, s' \in S$, and

$$L = L(S) = \sum_{t \neq 0} v^2(t),$$

the number of quadruples (x_1, x_2, y_1, y_2) in S^4 such that $x_1 - x_2 = y_1 - y_2 \neq 0$.

Theorem 4.1. *Let S be a set and suppose that $S \subseteq A - A$. Then*

$$|S|^2 - |S||A| \leqslant |A - A||A|M \tag{4.1}$$

and

$$|S|^2 - |S||A| \leqslant |A - A|^{1/2}|A|(L - |S|^2)^{1/2}. \tag{4.2}$$

Furthermore, for every $0 \leqslant \gamma \leqslant 1/2$, $0 \leqslant \delta \leqslant 1$ there exist arbitrarily large sets S, S' with $M = M(S) = (1 + o(1))|S|^{\gamma}$, $L = L(S') \leqslant (1 + o(1))|S|^{1+\delta}$ and a difference cover A of S and S' such that

$$|S|^2 = (1/4 + o(1))|A - A||A|M$$

and

$$|S'|^2 \geqslant (1/12 + o(1))|A - A|^{1/2}|A|L^{1/2},$$

as $|S|, |S'| \to \infty$.

Proof. Let A be any set with $S \subseteq A - A$. For any $a \in A$ define

$$S_a = (a - A) \cap S.$$

Observe that for every $s_1, s_2 \in S_a$ we have $s_1 - s_2 \in A - A$. Indeed, from the definition of S_a it follows that there are numbers $a_1, a_2 \in A$ with $a - a_1 = s_1$ and $a - a_2 = s_2$. Hence $s_1 - s_2 = (a - a_1) - (a - a_2) = a_2 - a_1$, so

$$S_a - S_a \subseteq A - A.$$

Clearly $S = \bigcup_{a \in A} S_a$. Let T_a, for $a \in A$, be any family satisfying $T_a \subseteq S_a$, $S = \bigcup_{a \in A} T_a$ and $T_a \cap T_{a'} = \emptyset$ for $a \neq a'$. Let X be the number of triples (t_1, t_2, a) such that $a \in A$, $t_1, t_2 \in T_a$ and $t_1 \neq t_2$. Clearly

$$X = \sum_{a \in A} |T_a|(|T_a| - 1).$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

On the other hand $t_1 - t_2 \in A - A$ and every integer has at most M representations as a difference $t_1 - t_2$. Hence

$$X \leqslant M|A - A|,$$

and

$$\sum_{a\in A} |T_a|^2 - \sum_{a\in A} |T_a| \leqslant M|A - A|.$$

By the Cauchy–Schwarz inequality we have

$$\frac{(\sum_{a\in A} |T_a|)^2}{|A|} - |S| \leqslant M|A - A|,$$

and thus

$$|S|^2 - |S||A| \leqslant |A - A||A|M.$$

To show the second lower bound we put $R = (A - A) \cap (S - S) \setminus \{0\}$ and observe that

$$\begin{aligned} X &= \sum_{a\in A} \sum_{r\in R} v_{T_a}(r) \\ &\leqslant \sum_{r\in R} v_S(r) \\ &\leqslant \left(\sum_{r\in R} v_S^2(r) \right)^{1/2} |R|^{1/2} \\ &\leqslant \left(\sum_{r\neq 0} v_S^2(r) \right)^{1/2} |A - A|^{1/2} \\ &= (L - |S|^2)^{1/2} |A - A|^{1/2} \end{aligned}$$

and the assertion follows.

In order to prove the last part of the theorem take any B_4 set $A \subseteq \{1, \dots, n\}$ (A is B_4 if every integer can be written in at most one way as a sum of four elements of A). We estimate $h(k)$, the number of representations of $k \in \mathbb{N}$ in the form

$$k = (a_1 - a_2) - (a_3 - a_4), \tag{4.3}$$

where $a_1, a_2, a_3, a_4 \in A$ and $a_1 > a_2, a_3 > a_4$. Suppose that $h(k) > 1$ and $a_1', a_2', a_3', a_4' \in A$ is another representation of k. Then

$$a_1 + a_2' + a_3' + a_4 = a_1' + a_2 + a_3 + a_4',$$

so $\{a_1, a_2', a_3', a_4\} = \{a_1', a_2, a_3, a_4'\}$. If $a_1 \neq a_3$ and $a_2 \neq a_4$ we have $h(k) \leqslant 2$. If, say, $a_1 = a_3$ then $a_2 \neq a_4$, and hence $a_2 \in \{a_2', a_3'\}$ and $a_4 \in \{a_1', a_4'\}$. Thus, all solutions to (4.3) are of the form

$$k = (a - a_2) - (a - a_4), \tag{4.4}$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

or

$$k = (a_4 - a) - (a_2 - a). \tag{4.5}$$

There are $|A| - A(a_4)$ solutions to (4.4)) and $A(a_2) - 1$ solutions to (4.5), where $A(x) = |A \cap [x]|$. Therefore

$$h(k) = |A| - A(a_4) + A(a_2) - 1$$

and, similarly, if $a_2 = a_4$ then $h(k) = |A| - A(a_1) + A(a_3) - 1$.

For fixed $0 < \gamma < 1/2$ and $0 < \delta < 1$ put

$$p = 2^{\frac{\gamma}{\gamma-2}}|A|^{\frac{2\gamma-1}{2-\gamma}}, \qquad q = |A|^{\frac{2(1-\delta)}{3-\delta}}$$

and assume that A is large enough. We define S and S' by

$$\mathbb{P}(a_1 - a_2 \in S) = p, \qquad \mathbb{P}(a_1 - a_2 \in S') = q,$$

independently for every $a_1, a_2 \in A$, where $a_1 > a_2$. Then clearly $D(S) \leqslant |A|$, $D(S') \leqslant |A|$ and $|S| = p(1/2 + o(1))|A|^2$, $|S'| = q(1/2 + o(1))|A|^2$ with high probability. Furthermore, we have

$$\mathbb{E}v_S(k) = p^2 h(k)$$

and

$$\mathbb{E}L(S') \leqslant \sum_{r \in A-A} v_{S'}^2(r) + \sum_{r \notin A-A} v_{S'}^2(r) \leqslant 2q^4|A|^4 + 2q^4|S'|^2 \leqslant 3q^4|A|^4.$$

Hence, by Chernoff's inequality,

$$M(S) = (1 + o(1))p^2|A| = (1 + o(1))2^{1/2}p^{3/2}|S|^{1/2} = (1 + o(1))|S|^{\gamma}$$

with high probability and, by Markov's inequality,

$$L(S') \leqslant 6q^4|A|^4 = (6 + o(1))|S'|^{1+\delta}$$

with probability at least $1/2$. Finally, there exist sets S and S' such that $M(S) = (1 + o(1))|S|^{\gamma}$, $L(S') \leqslant (1 + o(1))|S'|^{1+\delta}$,

$$|S|^2 = (1/4 + o(1))p^2|A|^4 = (1/4 + o(1))|A|^3 M(S) = (1/4 + o(1))|A - A||A|M(S)$$

and

$$|S'|^2 = (1/4 + o(1))q^2|A|^4 \geqslant (1/12 + o(1))L^{1/2}(S')|A|^2 = (1/12 + o(1))L^{1/2}(S')|A||A - A|^{1/2}$$

as $|A| \to \infty$. □

As direct corollaries to Theorem 4.1 we obtain a lower bound for the difference cover of the set of squares and Sidon sets.

Corollary 4.2. *If* $Q_n = \{1^2, 2^2, 3^2, \dots\} \cap \{1, \dots, n\}$ *then*

$$D(Q_n) \geqslant n^{1/3 - o(1)}.$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

Proof. It is well known that any natural number k has $k^{o(1)}$ representations in $Q_n - Q_n$. Therefore, by (4.1) we obtain

$$D(Q_n) \geqslant (|Q_n|/n^{o(1)})^{2/3}. \qquad \square$$

It seems to be an interesting open question to find the most efficient difference cover for squares less than n. The author could not prove any non-trivial lower bound.

Open problem. *Determine the order of magnitude of* $D(Q_n)$.

Corollary 4.3. *If S is a Sidon set then*

$$D(S) \geqslant (1 + o(1))|S|^{2/3}. \qquad \square$$

We show that the constant $2/3$ in Corollary 4.3 can not be replaced by a bigger one. Moreover, (4.2) tells us that $D(S) = \Omega(|S|^{2/3})$ for every set with $L(S) = O(|S|)$. The next theorem also provides an example of a set with $L = 0$ and $D(S) = O(|S|^{2/3})$.

Theorem 4.4. *There is a Sidon set S such that*

$$D(S) \leqslant \big((2 + o(1))|S|\big)^{2/3}.$$

Proof. Let A be a B_4 set. From extremal graph theory (see [3]) we know that for every natural number t there is a C_4-free graph $G_t = (V, E)$ with $|V| = t$, $|E| = (1/2 + o(1))t^{3/2}$. Let $G = G_{|A|}$, and assume that $V(G) = A$. Define

$$S := \{a - a' : \ a > a' \quad \text{and} \quad \{a, a'\} \in E(G)\}.$$

We show that S is a Sidon set. Indeed, otherwise there are numbers $s_1, s_2, s_3, s_4 \in S$, $\{s_1, s_2\} \neq \{s_3, s_4\}$, and $a_1, a_1', a_2, a_2', a_3, a_3', a_4, a_4' \in A$ such that

$$s_1 + s_2 = s_3 + s_4$$

and $s_i = a_i - a_i'$ for every $i \in [4]$. Thus, we have

$$a_1 + a_2 + a_3' + a_4' = a_1' + a_2' + a_3 + a_4.$$

Hence $\{a_1, a_2, a_3', a_4'\} = \{a_1', a_2', a_3, a_4\}$ and it follows that every number a_i has degree 2 in the graph $G[a_1, a_2, a_3', a_4']$, so that $G[a_1, a_2, a_3', a_4'] = C_4$, a contradiction. Finally, S is a Sidon set and clearly A is a difference cover of S. Furthermore, we have $|S| = |E(G)| = (1/2 + o(1))|A|^{3/2}$, so that

$$D(S) \leqslant \big((2 + o(1))|S|\big)^{2/3},$$

and the proof is completed. $\square$

We conclude this section with an observation that a large difference set (or sumset) is too weak to force a large difference cover of S. Even if we assume that every subset of S has a large difference set we can tell nothing about the size of the difference cover. For example, let $S = (A - A)_+$, where A is a B_4 set, and $S' \subseteq S$. We show that

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

$|S'-S'| = \Omega(|S'|^2)$, while $D(S) = O(|S|^{1/2})$. Every pair of distinct elements $s, s' \in S'$ such that $s = a_1 - a_2$, $s' = a_1' - a_2'$, $a_1 \neq a_1'$ and $a_2 \neq a_2'$, gives two differences $s - s'$ and $s' - s$, and each of them has at most two representations in the difference set $S - S$ (for the detailed analysis see the proof of Theorem 4.1). Let $a \in A$ and denote by $r(a)$ the number of elements in S' of the form $a - a'$, with $a' \in A$. Then, we have

$$|S'-S'| \geqslant \frac{1}{2} \sum_{(a,a'):a-a' \in S'} \big(|S'| - r(a) - r(a') + 1\big). \tag{4.6}$$

On the other hand, let $a_1, \ldots, a_{r(a)} \in A$ be numbers such that $a - a_1, \ldots, a - a_{r(a)} \in S'$. Since A is a B_4 set it follows that the set $\{a - a_1, \ldots, a - a_{r(a)}\}$ is B_4 as well, so

$$|S'-S'| \geqslant r(a)(r(a)-1) + 1, \tag{4.7}$$

for every $a \in A$. If $\max r(a) \leqslant |S'|/3$, then by (4.6),

$$|S'-S'| \geqslant \frac{1}{6}|S'|^2,$$

otherwise, by (4.7) we have $|S'-S'| \geqslant |S'|^2/9 - |S'|/3$, as claimed.

5. Upper bounds

It turns out that a small size for $S - S$ (or $S + S$) is strong enough to show that $D(S)$ can be not too large. First we consider sets with very small sumsets and show that $D(S) = O(|S|^{1/2})$. Observe that the set

$$\{1, 2, \ldots, l\} \cup \{l, 2l, \ldots, \lceil k/l \rceil l\},$$

where $l = \lfloor \sqrt{k} \rfloor$, is a difference cover of $\{1, \ldots, k\}$, and hence

$$D(\{1, \ldots, k\}) \leqslant 2\sqrt{k} + 1 \leqslant 3\sqrt{k}.$$

Of course, the same estimate holds for every arithmetic progression of length k.

Theorem 5.1. *For every C there exists a C' such that if $|S + S| \leqslant C|S|$ then*

$$D(S) \leqslant C'|S|^{1/2},$$

provided that $|S| > N(C)$.

Proof. Let $|S + S| \leqslant C|S|$ for some constant C. By the well-known theorem of Freiman [5], if $|S|$ is large enough, then S is subset of a generalized arithmetic progression

$$P = P(q_1, \ldots, q_d; l_1, \ldots, l_d, a) = \left\{ m = a + \sum_{i=1}^{d} x_i q_i, 0 \leqslant x_i \leqslant l_i \right\}$$

such that $|S| \geqslant \alpha \prod_{i=1}^{d} (l_i + 1)$, $\alpha = \alpha(C)$ and $d = d(C)$. We can write $P = a + P_1 + \cdots + P_d$, where $P_i = \{0, q_i, \ldots, l_i q_i\}$. Let A_i be a difference cover of size at most $3\sqrt{|P_i|}$. Then obviously the set

$$A = (A_1 + \cdots + A_d) \cup (a + A_1 + \cdots + A_d)$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

is a difference cover of P. Thus, we have

$$D(S) \leqslant D(P) \leqslant |A| \leqslant 2\prod_{i=1}^{d} |A_i| \leqslant 2\prod_{i=1}^{d}(3(l_i+1)^{1/2}) \leqslant 2\alpha^{-1}3^d|S|^{1/2}$$

and the assertion follows. □

We can provide a slightly weaker, but more general estimate of $D(S)$ depending on the size of $S+S$ (or $S-S$). For that purpose we prove the following result.

Theorem 5.2. *Let $S, T \subseteq \mathbb{Z}$ be finite sets such that $T \neq \emptyset$. Then*

$$D(S) \leqslant \frac{|S+T|\log(e|T|)}{|T|} + |T| + 1.$$

Proof. To begin with, note that, for every set B, the set

$$A = B \cup \{0\} \cup (S \setminus (B-B))$$

is a difference cover of S. In order to construct a small difference cover of S we make use of the set T and a random subset $R \subseteq S+T$ defined by

$$\mathbb{P}(r \in R) = p,$$

independently for every $r \in S+T$, for some $p \in (0,1)$. Let $B = T \cup R$. Observe that for any element $s \in S$, $s \notin B-B$ if and only if $(s+T) \cap R = \emptyset$. It follows that

$$\mathbb{P}(s \notin B-B) \leqslant (1-p)^{|T|} < e^{-p|T|},$$

so

$$\begin{aligned}\mathbb{E}|A| &\leqslant |T| + \mathbb{E}|R| + 1 + \mathbb{E}|S \setminus (B-B)| \\ &\leqslant |T| + p|S+T| + \sum_{s\in S} e^{-p|T|} + 1.\end{aligned}$$

For $p = \ln|T|/|T|$, we obtain

$$\mathbb{E}|A| \leqslant \frac{|S+T|\ln(e|T|)}{|T|} + |T| + 1. \tag{5.1}$$

Thus, by (5.1), there exists a difference cover A' of S satisfying

$$|A'| \leqslant \mathbb{E}|A| \leqslant \frac{|S+T|\log(e|T|)}{|T|} + |T| + 1$$

and the assertion follows. □

Corollary 5.3. *For every finite set $S \subseteq \mathbb{Z}$ we have*

$$D(S) \leqslant 2(|S \pm S|\log(e|S|))^{1/2} + 2.$$

Proof. If $|S+S| \geqslant |S|^2/\log(e|S|)$ (or $|S-S| \geqslant |S|^2/\log(e|S|)$), then

$$D(S) \leqslant |S| + 1 \leqslant (|S+S|\log(e|S|))^{1/2} + 1.$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

Now assume that $|S+S| < |S|^2/\log|S|$. Let T be any subset of S ($-S$ respectively) of size $\lceil(|S+S|\log(e|S|))^{1/2}\rceil$. By Theorem 5.2, we obtain

$$\begin{aligned} D(S) &\leqslant \frac{|S+T|\log(e|T|)}{\lceil(|S+S|\log(e|S|))^{1/2}\rceil} + \lceil(|S+S|\log(e|S|))^{1/2}\rceil + 1 \\ &\leqslant 2(|S+S|\log(e|S|))^{1/2} + 2, \end{aligned}$$

as required. □

The above result is best possible up to $O(\log^{1/2}|S|)$ factor. Indeed, let $S \subseteq \{1,\dots,n\}$ be a random subset with roughly $n^{1/2+\delta}$ elements, $\delta > 0$. Then $|A+A|, |A-A| \sim 2n$ and by Theorem 3.1

$$D(S) = \Omega(n^{1/2}) = \Omega(|A+A|^{1/2}).$$

Observe also that the number of elements in the sumset and the difference set can be essentially different (see [7] and [9]), so the sign $\pm$ in Corollary 5.3 can be useful. It may happen that $D(S)$ is small while $|S+S|$ is big (we have already given an example of such a set in the previous section). For instance, if

$$S = \{1,\dots,k\} \cup \{k, 2k, \dots, k^2\}$$

then clearly $|S+S| = \Omega(|S|^2)$ and $D(S) = O(|S|^{1/2})$. For such sets Corollary 5.3 is useless.

A lower bound on $v(t)$ (an analogue of $M(S)$) would not be useful since it is trivial. It turns out that a lower bound on some average of $v(t)$ is strong enough to force such an upper bound on $D(S)$. We show an upper bound in terms of the following parameters. Define

$$v_s = \frac{1}{|S|}\sum_{t\in S-s} v(t)$$

and put

$$\mu = \mu(S) = \min_{s\in S} v_s.$$

Large $L(S)$ also does not imply small $D(S)$ because S can consist of an arithmetic progression and an arbitrary set. It turns out, however, that if the function L is large then there is a large subset of S with small difference cover. In such a situation one could apply an effective version of the Balog–Szemerédi–Gowers theorem [2, 6] to show that a large subset of S has a small difference cover. In order to obtain a more efficient estimate we use the next theorem.

Theorem 5.4. *Let $S \subseteq \mathbb{Z}$ be a finite set; then*

$$D(S) \leqslant \left(\frac{2\log\mu}{\mu}\right)^{1/2}|S| + \left(\frac{1}{\mu}\right)^{1/2}|S| + 1,$$

where $\mu = \min_{s\in S} v_s$ and $v_s = \frac{1}{|S|}\sum_{t\in S-s} v(t)$. Moreover, if $L = c|S|^3 \geqslant |S|^2$ and S is large enough, there exists a set $T \subseteq S$ such that $|T| \geqslant (c/8)|S|$ and

$$D(T) \leqslant 8(c^{-1}|S|)^{1/2}.$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

Proof. Let S be a finite set. For $0 \leqslant q \leqslant 1$ we define random sets R and C by

$$\mathbb{P}(t \in R) = q\frac{v(t)}{|S|} = p_t \quad \text{and} \quad \mathbb{P}(s \in C) = q,$$

independently for every $t \in S - S$ and every $s \in S$. Put $B = R \cup C$ and again observe that the set

$$A = B \cup \{0\} \cup (S \setminus (B - B))$$

is a difference cover of S. Note that, for every $s \in S$, $s \notin B - B$, if for each $s' \in S$ we have $s' - s \notin R$ or $s' \notin C$. Hence

$$\mathbb{P}(s \notin B - B) \leqslant \prod_{t \in S - s} (1 - qp_t) < e^{-q\sum_{t \in S-s} p_t} = e^{-q^2 v_s},$$

so

$$\begin{aligned}\mathbb{E}|A| &\leqslant \mathbb{E}|R| + \mathbb{E}|C| + \mathbb{E}|S \setminus (B - B)| + 1\\ &\leqslant q\frac{1}{|S|}\sum_{t \in S-S} v(t) + q|S| + \sum_{s \in S} e^{-q^2 v_s} + 1\\ &\leqslant 2q|S| + |S|e^{-q^2\mu} + 1.\end{aligned}$$

If we put $q = (\log \mu/(2\mu))^{1/2}$ then

$$\mathbb{E}|A| \leqslant \left(\frac{2\log\mu}{\mu}\right)^{1/2}|S| + \left(\frac{1}{\mu}\right)^{1/2}|S| + 1 \leqslant 3\left(\frac{\log\mu}{\mu}\right)^{1/2}|S| + 1.$$

Now we prove the second assertion. Suppose that $L = c|S|^3$ and define

$$U = \{s \in S : v_s \geqslant (c/2)|S|\}.$$

Then

$$\begin{aligned}\sum_{s \in U} v_s &= \sum_{s \in S} v_s - \sum_{s \in S \setminus U} v_s \geqslant \frac{1}{|S|}\sum_t v^2(t) - (c/2)|S|^2\\ &= \frac{L}{|S|} - (c/2)|S|^2 \geqslant (c/2)|S|^2,\end{aligned}$$

so

$$|U| \geqslant (c/2)|S|.$$

Putting $q = ((c/2)|S|)^{-1/2}$ and $X = X(B) = |(B - B) \cap S|$, we obtain

$$\begin{aligned}\mathbb{E}(X) &= \sum_{s \in S} \mathbb{P}(s \in B - B) \geqslant |U| - \sum_{s \in U} e^{-q^2 v_s}\\ &\geqslant \left(1 - \frac{1}{e}\right)|U| \geqslant (c/4)|S|.\end{aligned}$$

We have

$$\mathbb{P}\left(X \geqslant \frac{1}{2}\mathbb{E}X\right) \geqslant \frac{\mathbb{E}X}{2|S|} \geqslant \frac{c}{4} \geqslant \frac{1}{4|S|},$$

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press

and by a Chernoff-type inequality (see Theorem A.1.11 in [1])

$$\mathbb{P}(|R| > 2\mathbb{E}|R|) < e^{-\frac{q|S|}{27}} \leqslant \frac{1}{10|S|}, \qquad \mathbb{P}(|C| > 2\mathbb{E}|C|) < e^{-\frac{q|S|}{27}} \leqslant \frac{1}{10|S|}.$$

Thus, there exists a set B with $|B| \leqslant 4q|S| = 8(c^{-1}|S|)^{1/2}$ such that $T = (B - B) \cap S$ consists of at least $\mathbb{E}X/2 \geqslant (c/8)|S|$ elements. □

References

[1] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, Wiley-Interscience, New York.

[2] Balog, A. and Szemerédi, E. (1994) A statistical theorem of set addition. *Combinatorica* **14** 263–268.

[3] Bollobás, B. (1978) *Extremal Graph Theory*, Academic Press, London.

[4] Colbourn, C. J. and Ling, A. C. H. (2000) Quorums from difference covers. *Inform. Processing Letters* **75** 57–59.

[5] Freiman, G. (1973) *Foundations of a Structural Theory of Set Addition*, Vol. 37 of *Translations of Math. Monographs*, AMS, Providence, RI.

[6] Gowers, W. T. (2001) A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** 465–588.

[7] Hennecart, F., Robert, G. and Yudin, A. (1999) On the number of sums and differences. *Astérisque* **258** 173–178.

[8] Redei, L. and Rényi, A. (1949) Über die Darstellung der Zahlen $1, 2, \ldots, N$ mittels Differenzen. (In Russian.) *Mat. Sb., N. Ser.* **24** 385–389.

[9] Ruzsa, I. Z. (1978) On the cardinality of $A + A$ and $A - A$. In *Combinatorics: Proc. Fifth Hungarian Colloq., Keszthely 1976*, Vol. 18 of *Coll. Math. Soc. Bolyai*, North-Holland, pp. 933–938.

https://doi.org/10.1017/S0963548307008401 Published online by Cambridge University Press