ON THE SIZE OF DISSOCIATED BASES

VSEVOLOD F. LEV AND RAPHAEL YUSTER

Abstract. We prove that the sizes of the maximal dissociated subsets of a given finite subset of an abelian group differ by a logarithmic factor at most. On the other hand, we show that the set \( \{0, 1\}^n \subseteq \mathbb{Z}^n \) possesses a dissociated subset of size \( \Omega(n \log n) \); since the standard basis of \( \mathbb{Z}^n \) is a maximal dissociated subset of \( \{0, 1\}^n \) of size \( n \), the result just mentioned is essentially sharp.

Recall, that subset sums of a subset \( \Lambda \) of an abelian group are group elements of the form \( \sum_{b \in B} b \), where \( B \subseteq \Lambda \); thus, a finite set \( \Lambda \) has at most \( 2^{\|\Lambda\|} \) distinct subset sums.

A famous open conjecture of Erdős, first stated about 80 years ago (see [B96] for a relatively recent related result and brief survey), is that if all subset sums of an integer set \( \Lambda \subseteq [1, n] \) are pairwise distinct, then \( |\Lambda| \leq \log_2 n + O(1) \) as \( n \to \infty \); here \( \log_2 \) denotes the base-2 logarithm. Similarly, one can investigate the largest possible size of subsets of other “natural” sets in abelian groups, possessing the property in question; say,

What is the largest possible size of a set \( \Lambda \subseteq \{0, 1\}^n \subseteq \mathbb{Z}^n \) with all subset sums pairwise distinct?

In modern terms, a subset of an abelian group, all of whose subset sums are pairwise distinct, is called dissociated. Such sets proved to be extremely useful due to the fact that if \( \Lambda \) is a maximal dissociated subset of a given set \( A \), then every element of \( A \) is representable (generally speaking, in a non-unique way) as a linear combination of the elements of \( \Lambda \) with the coefficients in \( \{-1, 0, 1\} \). Hence, maximal dissociated subsets of a given set can be considered as its “linear bases over the set \( \{-1, 0, 1\} \)”. This interpretation naturally makes one wonder whether, and to what extent, the size of a maximal dissociated subset of a given set is determined by this set. That is,

Is it true that all maximal dissociated subsets of a given finite set in an abelian group are of about the same size?

In this note we answer the two above-stated questions as follows.
Theorem 1. For a positive integer $n$, the set $\{0, 1\}^n$ (consisting of those vectors in $\mathbb{Z}^n$ with all coordinates being equal to 0 or 1) possesses a dissociated subset of size $(1 + o(1)) n \log_2 n / \log_2 9$ (as $n \to \infty$).

Theorem 2. If $\Lambda$ and $M$ are maximal dissociated subsets of a finite subset $A \subsetneq \{0\}$ of an abelian group, then

$$\frac{|M|}{\log_2 (2|M| + 1)} \leq |\Lambda| < |M| \left( \log_2 (2M) + \log_2 \log_2 (2|M|) + 2 \right).$$

We remark that if a subset $A$ of an abelian group satisfies $A \subseteq \{0\}$, then $A$ has just one dissociated subset; namely, the empty set.

Since the set of all $n$-dimensional vectors with exactly one coordinate equal to 1 and the other $n - 1$ coordinates equal to 0 is a maximal dissociated subset of the set $\{0, 1\}^n$, comparing Theorems 1 and 2 we conclude that the latter is sharp in the sense that the logarithmic factors cannot be dropped or replaced with a slower growing function, and the former is sharp in the sense that $n \log_2 n$ is the true order of magnitude of the size of the largest dissociated subset of the set $\{0, 1\}^n$. At the same time, the bound of Theorem 2 is easy to improve given that the underlying group has bounded exponent.

Theorem 3. Let $A$ be a finite subset of an abelian group $G$ of exponent $e := \exp(G)$. If $r$ denotes the rank of the subgroup $\langle A \rangle$, generated by $A$, then for any maximal dissociated subset $\Lambda \subseteq A$ we have

$$r \leq |\Lambda| \leq r \log_2 e.$$

We now turn to the proofs.

Proof of Theorem 1. We will show that if $n > (2 \log_2 3 + o(1)) m / \log_2 m$, with a suitable choice of the implicit function, then the set $\{0, 1\}^n$ possesses an $m$-element dissociated subset. For this we prove that there exists a set $D \subseteq \{0, 1\}^m$ with $|D| = n$ such that for every non-zero vector $s \in S := \{-1, 0, 1\}^m$ there is an element of $D$, not orthogonal to $s$. Once this is done, we consider the $n \times m$ matrix whose rows are the elements of $D$; the columns of this matrix form then an $m$-element dissociated subset of $\{0, 1\}^n$, as required.

We construct $D$ by choosing at random and independently of each other $n$ vectors from the set $\{0, 1\}^m$, with equal probability for each vector to be chosen. We will show that for every fixed non-zero vector $s \in S$, the probability that all vectors from $D$ are orthogonal to $s$ is very small, and indeed, the sum of these probabilities over
all \( s \in S \setminus \{0\} \) is less than 1. By the union bound, this implies that with positive probability, every vector \( s \in S \setminus \{0\} \) is not orthogonal to some vector from \( D \).

We say that a vector from \( S \) is of type \( (m^+, m^-) \) if it has \( m^+ \) coordinates equal to +1, and \( m^- \) coordinates equal to −1 (so that \( m - m^+ - m^- \) of its coordinates are equal to 0). Suppose that \( s \) is a non-zero vector from \( S \) of type \( (m^+, m^-) \). Clearly, a vector \( d \in \{0, 1\}^m \) is orthogonal to \( s \) if and only if there exists \( j \geq 0 \) such that \( d \) has exactly \( j \) non-zero coordinates in the (+1)-locations of \( s \), and exactly \( j \) non-zero coordinates in the (−1)-locations of \( s \). Hence, the probability for a randomly chosen \( d \in \{0, 1\}^m \) to be orthogonal to \( s \) is

\[
\frac{1}{2^{m^+ + m^-}} \sum_{j=0}^{\min\{m^+, m^-\}} \binom{m^+}{j} \binom{m^-}{j} = \frac{1}{2^{m^+ + m^-}} \binom{m^+ + m^-}{m^+} < \frac{1}{\sqrt{1.5(m^+ + m^-)}}.
\]

It follows that the probability for all elements of our randomly chosen set \( D \) to be simultaneously orthogonal to \( s \) is smaller than \( (1.5(m^+ + m^-))^{-n/2} \).

Since the number of elements of \( S \) of a given type \( (m^+, m^-) \) is \( \binom{m}{m^+} \binom{m^+ + m^-}{m^+} \), to conclude the proof it suffices to estimate the sum

\[
\sum_{1 \leq m^+ + m^- \leq m} \binom{m}{m^+ + m^-} \binom{m^+ + m^-}{m^+} (1.5(m^+ + m^-))^{-n/2}
\]

showing that its value does not exceed 1.

To this end we rewrite this sum as

\[
\sum_{t=1}^{m} \binom{m}{t} (1.5t)^{-n/2} \sum_{m^+=0}^{t} \binom{t}{m^+} = \sum_{t=1}^{m} \binom{m}{t} 2^t (1.5t)^{-n/2}
\]

and split it into two parts, according to whether \( t < T \) or \( t \geq T \), where \( T := m/(\log_2 m)^2 \). Let \( \Sigma_1 \) denote the first part and \( \Sigma_2 \) the second part. Assuming that \( m \) is large enough and

\[
n > 2 \log_2 3 \frac{m}{\log_2 m} (1 + \varphi(m))
\]

with a function \( \varphi \) sufficiently slowly decaying to 0, we have

\[
\Sigma_1 \leq \left( \frac{m}{T} \right) 2^T 1.5^{-n/2} < \left( \frac{9m}{T} \right)^T 1.5^{-n/2} = (3 \log_2 m)^{2T} 1.5^{-n/2},
\]

whence

\[
\log_2 \Sigma_1 < \frac{2m}{(\log_2 m)^2} \log_2 (3 \log_2 m) - \log_2 3 \log_2 1.5 \frac{m}{\log_2 m} (1 + \varphi(m)) < -1,
\]
and therefore $\Sigma_1 < 1/2$. Furthermore,

$$\Sigma_2 \leq T^{-n/2} \sum_{t=1}^{m} \binom{m}{t} 2^t < T^{-n/2} 3^m,$$

implying

$$\log_2 \Sigma_2 < m \log_2 3 - (\log_2 m - 2 \log_2 \log_2 m) \log_2 3 \frac{m}{\log_2 m} \left(1 + \varphi(m)\right)$$

$$= m \log_2 3 \left(\frac{2 \log_2 \log_2 m}{\log_2 m} \left(1 + \varphi(m)\right) - \varphi(m)\right)$$

$$< -1.$$

Thus, $\Sigma_2 < 1/2$; along with the estimate $\Sigma_1 < 1/2$ obtained above, this completes the proof. \qed

Proof of Theorem \[2\] Suppose that $\Lambda, M \subseteq A$ are maximal dissociated subsets of $A$. By maximality of $\Lambda$, every element of $A$, and consequently every element of $M$, is a linear combination of the elements of $\Lambda$ with the coefficients in $\{-1, 0, 1\}$. Hence, every subset sum of $M$ is a linear combination of the elements of $\Lambda$ with the coefficients in $\{-|M|, -|M| + 1, \ldots, |M|\}$. Since there are $2^{|M|}$ subset sums of $M$, all distinct from each other, and $(2^{|M|} + 1)^{|\Lambda|}$ linear combinations of the elements of $\Lambda$ with the coefficients in $\{-|M|, -|M| + 1, \ldots, |M|\}$, we have

$$2^{|M|} \leq (2^{|M|} + 1)^{|\Lambda|},$$

and the lower bound follows.

Notice, that by symmetry we have

$$2^{|\Lambda|} \leq (2^{|\Lambda|} + 1)^{|M|},$$

whence

$$|\Lambda| \leq |M| \log_2 (2^{|\Lambda|} + 1). \quad (\ast)$$

Observing that the upper bound is immediate if $M$ is a singleton (in which case $A \subseteq \{-g, 0, g\}$, where $g$ is the element of $M$, and therefore every maximal dissociated subset of $A$ is a singleton, too), we assume $|M| \geq 2$ below.

Since every element of $\Lambda$ is a linear combination of the elements of $M$ with the coefficients in $\{-1, 0, 1\}$, and since $\Lambda$ contains neither 0, nor two elements adding up to 0, we have $|\Lambda| \leq (3^{|M|} - 1)/2$. Consequently, $2|\Lambda| + 1 \leq 3^{|M|}$, and using $(\ast)$ we get

$$|\Lambda| \leq |M|^2 \log_2 3.$$ 

Hence,

$$2|\Lambda| + 1 < |M|^2 \log_2 9 + 1 < 4|M|^2,$$
and substituting this back into (*) we obtain

$$|\Lambda| < 2|M| \log_2(2|M|).$$

As a next iteration, we conclude that

$$2|\Lambda| + 1 < 5|M| \log_2(2|M|),$$

and therefore, by (*),

$$|\Lambda| \leq |M|(\log_2(2|M|) + \log_2(2|M|) + \log_2(5/2)).$$

□

Proof of Theorem 3. The lower bound follows from the fact that \(\Lambda\) generates \(\langle A \rangle\), the upper bound from the fact that all \(2^{|\Lambda|}\) pairwise distinct subset sums of \(\Lambda\) are contained in \(\langle A \rangle\), whereas \(|\langle A \rangle| \leq e^r\). □

We close our note with an open problem.

For a positive integer \(n\), let \(L_n\) denote the largest size of a dissociated subset of the set \(\{0, 1\}^n \subseteq \mathbb{Z}^n\). What are the limits

$$\liminf_{n \to \infty} \frac{L_n}{n \log_2 n} \text{ and } \limsup_{n \to \infty} \frac{L_n}{n \log_2 n}?$$

Notice, that by Theorems 1 and 2 we have

$$1/\log_2 9 \leq \liminf_{n \to \infty} \frac{L_n}{n \log_2 n} \leq \limsup_{n \to \infty} \frac{L_n}{n \log_2 n} \leq 1.$$