Note on a variant of the Erdős–Ginzburg–Ziv problem

by

CHAO WANG (Tianjin)

1. Introduction. P. Erdős, A. Ginzburg and A. Ziv [3] proved that from any sequence of integers of length $2n - 1$ one can extract a subsequence of length $n$ whose sum is congruent to zero modulo $n$.

A. Bialostocki and P. Dierker [1] proved that if $A = (a_1, \ldots, a_{2n-2})$ is a sequence of integers of length $2n - 2$ and there are no indices $i_1, \ldots, i_n$ belonging to $\{1, \ldots, 2n - 2\}$ such that

$$a_{i_1} + a_{i_2} + \ldots + a_{i_n} \equiv 0 \pmod{n},$$

then there are two residue classes modulo $n$ such that $n - 1$ of the $a_i$’s belong to one of the classes and the remaining $n - 1$ belong to the other class.

In order to study the relation between the number of classes present in a sequence $A = (a_1, \ldots, a_g)$ and the possibility to have a relation like (1), A. Bialostocki and M. Lotspeich [2] introduced the following function.

Definition 1.1 ([2]). Let $n, k$ be positive integers, $1 \leq k \leq n$. We define $f(n, k)$ to be the least integer $g$ for which the following holds: If $A = (a_1, \ldots, a_g)$ is a sequence of integers of length $g$ such that the number of distinct modulo $n$ is equal to $k$, then there are $n$ indices $i_1, \ldots, i_n$ belonging to $\{1, \ldots, g\}$ such that $a_{i_1} + \ldots + a_{i_n} \equiv 0 \pmod{n}$.

The Erdős–Ginzburg–Ziv theorem implies that $f(n, k)$ exists and is not greater than $2n - 1$. It is easy to see that $f(n, 1) = n$, $f(n, 2) = 2n - 1$, $f(n, k) \geq n$, and

$$f(n, k) \leq 2n - 2 \quad \text{for} \ 2 < k \leq n.$$

For given $n$, we will formulate the problem and work in the context of $\mathbb{Z}_n$, the cyclic group of residue classes modulo $n$. Let us define $f(n, k)$ in the following equivalent way.

Definition 1.2 ([4]). Let $n, k$ be positive integers, $1 \leq k \leq n$. Denote by $f(n, k)$ the least integer $g$ for which the following holds: If $A = (a_1, \ldots, a_g)$ is a sequence of elements of $\mathbb{Z}_n$ of length $g$ such that the number of distinct

2000 Mathematics Subject Classification: Primary 11B50.

[53]
a_i’s is equal to k, then there are n indices \( i_1, \ldots, i_n \) belonging to \{1, \ldots, g\} such that \( a_{i_1} + \ldots + a_{i_n} = 0 \).

**Notation.** A sequence \( A = (0, 0, 1, 1, 1, 2, 3, 5) \) will also be denoted by \( A = (0^2, 1^3, 2, 3, 5) \). The elements of \( \mathbb{Z}_n \) will be denoted by 0, 1, \ldots, \( n-1 \).


**Theorem 1.1 ([4]).** Let \( n \) be a positive integer. Then \( f(n, n) = n \) if \( n \) is odd and \( f(n, n) = n + 1 \) if \( n \) is even.

**Theorem 1.2 ([4]).** Let \( n \geq 5 \) and \( 1 + n/2 < k \leq n - 1 \). Then \( f(n, k) = n + 2 \).

In this article, \( k \) and \( n \) will be positive integers. We prove the following theorems.

**Theorem 1.3.** If \( k = 2m + 1 \geq 3 \) is odd and
\[
n \geq \max\{4m^2 - 4, m(m + 3)/2 + 2\},
\]
then
\[
f(n, k) = 2n - m^2 - 1.
\]

**Theorem 1.4.** If \( k = 2m \) is even and
\[
n \geq \max\{4m(m - 1) - 4, m(m + 1)/2 + 1\},
\]
then
\[
f(n, k) = 2n - m(m - 1) - 1.
\]

**2. Proofs.** In order to prove Theorems 1.3 and 1.4, we need some preliminaries that appeared in [5].

**Theorem 2.1 ([5]).** Let \( n \geq 2 \) and \( 2 \leq k \leq \lfloor n/4 \rfloor + 2 \), and let \((a_1, \ldots, a_{2n-k})\) be a sequence of length \( 2n - k \) in \( \mathbb{Z}_n \). Suppose that for any \( n \)-subset \( I \) of \( \{1, \ldots, 2n - k\} \), \( \sum_{i \in I} a_i \neq 0 \). Then one can rearrange the sequence as
\[
(a, \ldots, a, b, \ldots, b, c_1, \ldots, c_{2n-k-u-v}),
\]
where \( u \geq n - 2k + 3 \), \( v \geq n - 2k + 3 \), \( u + v \geq 2n - 2k + 2 \) and \( a-b \) generates \( \mathbb{Z}_n \).

In [5], Weidong Gao introduced the following two definitions.

**Definition 2.1 ([5]).** Let \( S = (a_1, \ldots, a_k) \) be a sequence of elements in \( \mathbb{Z}_n \). For any \( b \in \mathbb{Z}_n \), we denote by \( b + S \) the sequence \((b + a_1, \ldots, b + a_k)\). For any \( 1 \leq r \leq k \), we define \( \sum_r(S) \) to be the set of all elements in \( \mathbb{Z}_n \) which can be expressed as a sum over an \( r \)-term subsequence of \( S \), i.e.,
\[
\sum_r(S) = \{a_{i_1} + \ldots + a_{i_r} \mid 1 \leq i_1 < \ldots < i_r \leq k\}.
\]

**Definition 2.2 ([5]).** Let \( S = (a_1, \ldots, a_m) \) and \( T = (b_1, \ldots, b_m) \) be two sequences of elements in \( \mathbb{Z}_n \) with \( |S| = |T| \). We say that \( S \) is equivalent
to $T$ (written $S \sim T$) if there exist an integer $c$ coprime to $n$, an element $x \in \mathbb{Z}_n$, and a permutation $\delta$ of $\{1, \ldots, m\}$ such that $a_i = c(b_{\delta(i)} - x)$ for every $i = 1, \ldots, m$. Clearly, “$\sim$” is an equivalence relation; and if $S \sim T$, then $0 \in \sum_n(S)$ if and only if $0 \in \sum_n(T)$.

With the above two definitions, Theorem 2.1 is equivalent to

**Lemma 2.2.** Let $n \geq 2$ and $2 \leq k \leq \lfloor n/4 \rfloor + 2$, and let $A = (a_1, \ldots, a_{2n-k})$ be a sequence of length $2n - k$ in $\mathbb{Z}_n$. If $0 \notin \sum_n(A)$, then

$$A \sim (0^u, 1^v, c_1, \ldots, c_{2n-k-u-v}),$$

where $u \geq n - 2k + 3$, $v \geq n - 2k + 3$, $u + v \geq 2n - 2k + 2$.

**Proof of Theorem 1.3.** Since $k = 2m + 1 \geq 3$, we have $m \geq 1$. Consider the sequence

$$E = (0^{n-m(m+3)/2-1}, 1^{n-m(m+1)/2}, 2, 3, \ldots, m, n - m, n - m + 1, \ldots, n - 1),$$

which contains exactly $k = 2m + 1$ distinct elements of $\mathbb{Z}_n$ and has

$$n - m(m+3)/2 - 1 + n - m(m+1)/2 + m - 1 + m = 2n - m^2 - 2$$

terms. Every $n$-term subsequence of $E$ has non-zero sum, so

$$f(n, k) \geq 2n - m^2 - 1.$$  

Suppose $E = (a_1, \ldots, a_{2n-m^2-1})$ is a sequence containing exactly $k$ distinct elements of $\mathbb{Z}_n$. Since $n \geq 4m^2 - 4 = 4(m^2 + 1) - 8$, from Lemma 2.2, we know that

$$E \sim (0^u, 1^v, c_1, \ldots, c_q),$$

where $u \geq n - 2m^2 + 1$, $v \geq n - 2m^2 + 1$, $u + v \geq 2n - 2m^2$, all $c_i \neq 0, 1$.

As $E$ contains $k$ distinct elements of $\mathbb{Z}_n$, we have $q \geq 2m - 1$, $u + v \leq 2n - m^2 - 1 - (2m - 1) = 2n - m(m+2)$.

Let $F = (0^u, 1^v, c_1, \ldots, c_q)$. Suppose $0 \notin \sum_n(E)$. Then $0 \notin \sum_n(F)$. It is easy to verify that $u + v \geq n$, so $n - v \leq u < u + 1$. For each $1 \leq i \leq q$, if $n - v \leq c_i \leq u + 1$, then $(0^{ci-1}, 1^{n-ci}, c_i)$ is an $n$-term subsequence of $F$ which has zero sum, which is impossible, so $c_i > u + 1$ or $c_i < n - v$. Without loss of generality, we can assume that $c_1, \ldots, c_s$ are all greater than $u + 1$, and $c_{s+1}, \ldots, c_q$ are all less than $n - v$.

It is easy to see that $c_i + c_j \geq n + 2$, $1 \leq i \neq j \leq s$. Since

$$2n - c_i - c_j \leq 2n - 2(u + 2) = v + 2n - u - (u + v) - 4$$

$$\leq v + 2n - (n - 2m^2 + 1) - (2n - 2m^2) - 4$$

$$= v - (n - 4m^2 + 4) - 1 < v,$$
it follows that if \( c_i + c_j \leq n + u + 2 \), then \((0^{c_i+c_j-n-2}, 1^{n-c_i-c_j}, c_i, c_j)\) is an \( n \)-term subsequence of \( F \) which has zero sum, so

\[
(2) \quad c_i + c_j > n + u + 2, \quad 1 \leq i \neq j \leq s.
\]

Suppose that for some \( t > 1 \) we have proved

\[
(3) \quad c_{i_1} + \ldots + c_{i_t-1} > (t-2)n + u + (t-1) + 1 + (u+2)
\]

\[
= (t-2)n + 2u + t + 2
\]

\[
\geq (t-2)n + 2(n-2m^2 + 1) + t + 2
\]

\[
= (t-1)n + (n-4m^2 + 4) + t
\]

\[
\geq (t-1)n + t,
\]

and

\[
(5) \quad tn - c_{i_1} - \ldots - c_{i_{t-1}} - c_{i_t}
\]

\[
\leq tn - [(t-2)n + u + (t-1) + 1] - (u + 2)
\]

\[
= 2n - 2u - t - 2
\]

\[
= v + 2n - u - (u + v) - t - 2
\]

\[
\leq v + 2n - (n - 2m^2 + 1) - (2n - 2m^2) - t - 2
\]

\[
= v - (n - 4m^2 + 4) - (t - 1) < v.
\]

If \( c_{i_1} + \ldots + c_{i_{t-1}} + c_{i_t} \leq (t-1)n + u + t \), then (4) and (5) show that

\[
(0^{c_{i_1}+\ldots+c_{i_t-1}-(t-1)n-u-t}, 1^{tn-c_{i_1}-\ldots-c_{i_t}}, c_{i_1}, \ldots, c_{i_t})
\]

is an \( n \)-term subsequence of \( F \) which has zero sum, so

\[
(6) \quad c_{i_1} + \ldots + c_{i_t} > (t-1)n + u + t, \quad 1 \leq i_1, \ldots, i_t \leq s,
\]

\[
i_1, \ldots, i_t \text{ pairwise distinct}.
\]

So we have proved that (6) holds for each \( 1 \leq t \leq s \) by induction. In particular, letting \( t = s \), we have

\[
(7) \quad c_1 + \ldots + c_s > (s-1)n + u + s.
\]

On the other hand, it is easy to see that \( c_{s+i} + c_{s+j} \leq n, 1 \leq i \neq j \leq q-s \). Since

\[
c_i + c_j - 2 \leq 2(n - v - 1) - 2
\]

\[
= u + 2n - v - (u + v) - 4
\]

\[
\leq u + 2n - (n - 2m^2 + 1) - (2n - 2m^2) - 4
\]

\[
= u - (n - 4m^2 + 4) - 1 < u,
\]
it follows that if $c_{s+i} + c_{s+j} \geq n - v$, then $(0^{c_{s+i}+c_{s+j}-2}, 1^{n-c_{s+i}-c_{s+j}}, c_{s+i}, c_{s+j})$ is an $n$-term subsequence of $F$ which has zero sum, so

$$c_{s+i} + c_{s+j} < n - v, \quad 1 \leq i \neq j \leq q - s.$$

Suppose that for some $t > 1$ we have proved

$$(8) \quad c_{s+i_1} + \ldots + c_{s+i_{t-1}} < n - v, \quad 1 \leq i_1, \ldots, i_{t-1} \leq q - s,$$

$i_1, \ldots, i_{t-1}$ pairwise distinct.

Then for every $i_t$ such that $1 \leq i_t \leq q - s$ and $i_t \neq i_j, 1 \leq j \leq t - 1$,

$$(9) \quad c_{s+i_1} + \ldots + c_{s+i_{t-1}} + c_{s+i_t} - t$$

$$\leq (n - v - 1) + (n - v - 1) - t$$

$$= 2n - 2v - t - 2$$

$$= u + 2n - v - (u + v) - t - 2$$

$$\leq u + 2n - (n - 2m^2 + 1) - (2n - 2m^2) - t - 2$$

$$= u - (n - 4m^2 + 4) - (t - 1) < u.$$

If $c_{s+i_1} + \ldots + c_{s+i_{t-1}} + c_{s+i_t} \geq n - v$, then (8) and (9) show that

$$(0^{c_{s+i_1}+\ldots+c_{s+i_{t-1}}-t}, 1^{n-c_{s+i_1}-\ldots-c_{s+i_t}}, c_{s+i_1}, \ldots, c_{s+i_t})$$

is an $n$-term subsequence of $F$ which has zero sum, so

$$(10) \quad c_{s+i_1} + \ldots + c_{s+i_t} < n - v, \quad 1 \leq i_1, \ldots, i_t \leq q - s,$$

$i_1, \ldots, i_t$ pairwise distinct.

So we have proved (10) for each $1 \leq t \leq q - s$ by induction. In particular, letting $t = q - s$, we have

$$c_{s+1} + \ldots + c_q < n - v.$$

The inequality (7) is equivalent to

$$(n - c_1) + (n - c_2) + \ldots + (n - c_s) < n - u - s.$$  

For $1 \leq i \leq s$, let $e_i = n - c_i$. Then $0 < e_i < n - u - 1$ and

$$(11) \quad e_1 + \ldots + e_s \leq n - u - s - 1.$$  

For $1 \leq i \leq q - s$, let $d_i = c_{s+i}$. Then $1 < d_i < n - v$ and

$$(12) \quad d_1 + \ldots + d_{q-s} \leq n - v - 1.$$  

Suppose that $\{e_1, \ldots, e_s\}$ has $w$ distinct elements. Then $\{d_1, \ldots, d_{q-s}\}$ has $2m - 1 - w$ distinct elements. From (11) and (12), we know that

$$e_1 + \ldots + e_s + d_1 + \ldots + d_{q-s} \leq 2n - u - v - s - 2.$$  

But in fact,
\[(e_1 + e_2 + \ldots + e_s + d_1 + d_2 + \ldots + d_{q-s}) - (2n - u - v - s - 2)\]
\[\geq 1 + 2 + 3 + \ldots + w + 1 \cdot (s - w) + 2 + 3 + \ldots + (2m - w)\]
\[+ 2 \cdot (2n - m^2 - 1 - u - v - s - (2m - 1 - w)) - (2n - u - v - s - 2)\]
\[\geq w(w + 1)/2 + s - w + (2m - w - 1)(2m - w + 2)/2 + 2n - 2m^2\]
\[- 4m - u - v - s + 2w + 2\]
\[= 2n - u - v + w^2 - 2mw + w - 3m + 1\]
\[\geq m(m + 2) + w^2 - 2mw + w - 3m + 1\]
\[= (m - w - 1/2)^2 + 3/4 > 0.\]
Contradiction! So \(0 \in \sum_n(E)\), which means \(f(n, k) \leq 2n - m^2 - 1\), and the proof is finished. 

Proof of Theorem 1.4. The proof is similar to that of Theorem 1.3. We leave it to the interested reader.

Letting \(k = 2, 3, 4, 5, 6\), we get the following corollary.

Corollary 2.3.
\[
\begin{align*}
f(n, 2) &= 2n - 1, \quad n \geq 2, \\
f(n, 3) &= 2n - 2, \quad n \geq 4, \\
f(n, 4) &= 2n - 3, \quad n \geq 4, \\
f(n, 5) &= 2n - 5, \quad n \geq 12, \\
f(n, 6) &= 2n - 7, \quad n \geq 20.
\end{align*}
\]

Acknowledgements. This work was done under the auspices of the Ministry of Education of China, the Ministry of Science and Technology, and the National Science Foundation of China.

I would like to thank Professor Gao Weidong and the referee for their helpful suggestions and comments.

References


Center for Combinatorics  
Nankai University  
Tianjin 300071, P.R. China  
E-mail: wch2001@eyou.com

*Received on 20.5.2002  
and in revised form on 22.10.2002* (4287)