Discrete Mathematics 309 (2009) 2658–2662

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

A polynomial-time algorithm for finding zero-sums

Alberto del Lungo [1], Claudio Marini [*], Elisa Mori [2]

Dipartimento di Scienze Matematiche ed Informatiche, Università degli Studi di Siena, Pian dei Mantellini 44, 53100 Siena, Italy

ARTICLE INFO

Article history:
Received 10 September 2007
Received in revised form 20 May 2008
Accepted 9 June 2008
Available online 9 October 2008

Keywords:
Combinatorial problem
Combinatorial group theory
Permutation
Abelian group
Erdös–Ginzburg–Ziv Theorem
Zero-sum Ramsey theory

ABSTRACT

Erdös, Ginzburg and Ziv proved that any sequence of $2n-1$ (not necessary distinct) members of the cyclic group $\mathbb{Z}_n$ contains a subsequence of length n the sum of whose elements is congruent to zero modulo n. There are several proofs of this celebrated theorem which combine combinatorial and algebraic ideas. Our main result is an alternative and constructive proof of this result. From this proof, we deduce a polynomial-time algorithm for finding a zero-sum n-sequence of the given $(2n-1)$-sequence of an abelian group G with n elements (a fortiori for $\mathbb{Z}_n$). To the best of our knowledge, this is the first efficient algorithm for finding zero-sum n-sequences.

1. Introduction

Erdös, Ginzburg and Ziv proved the following theorem in 1961 [7].

Theorem 1.1. *For any sequence $x_1x_2\cdot\cdots\cdot x_{2n-1}$ of (not necessarily distinct) elements of the cyclic group $\mathbb{Z}_n$, there exists a subset I of $\{1,2,\ldots,2n-1\}$ such that $|I|=n$ and $\sum_{i\in I}x_i=0 \bmod n$.*

This beautiful result is the cornerstone of almost all combinatorial research on *zero-sum Ramsey problems*. This theory is a newly established area of combinatorics, and its paradigm can be formulated as follows:

- suppose the elements of the combinatorial structure are mapped into a finite group G. Does there exist a prescribed substructure such that the sum of the weights of its elements is 0 in G?

A survey of zero-sum Ramsey theory with a list of conjectures and open problems is given in [4] and some generalizations of the Erdös–Ginzburg–Ziv Theorem are in [2,5,9,11]. Various researchers have obtained extensions of this result to graph theory (see e.g. [1,3,8,14]).

Quite surprisingly, the Erdös–Ginzburg–Ziv Theorem can be proved in numerous distinct ways: in [7] it is shown that it is sufficient to prove the theorem when n is prime.

Proposition 1.2. *For a prime p and for any sequence $x_1x_2\cdot\cdots\cdot x_{2p-1}$ of (not necessarily distinct) elements of the cyclic group $\mathbb{Z}_p$ there exists a subset of $\{1,2,\ldots,2p-1\}$ such that $|I|=p$ and $\sum_{i\in I}x_i=0$.*

* Corresponding author.
E-mail addresses: marinic@unisi.it (C. Marini), emori@unisi.it, emori@sienabiotech.it (E. Mori).

1 This paper was nearly completed when Alberto suddenly died on June 1, 2003. We present this work as a tribute to his memory.

2 Present address: Siena Biotech SpA, via Torre Fiorentina, 1 53100 Siena, Italy.

0012-365X/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2008.06.018

Our main result deals with the following complexity question:

- how difficult is to find a zero-sum n-sequence of the given $(2n-1)$-sequence of $\mathbb{Z}_n$?

Hamidoune [12] pointed out that there are twelve proofs of the Erdös–Ginzburg–Ziv Theorem, but unfortunately these proofs do not provide efficient algorithms for finding zero-sum sequences. To the best of our knowledge, there is only one polynomial-time algorithm defined by Füredi and Kleitman [8] for finding a spanning tree of a complete graph K_{n+1} such that the sum of the weights of its edges is 0 modulo n.

We tackle this complexity question by giving an alternative and constructive proof of the Erdös–Ginzburg–Ziv Theorem. The approach of this proof is completely different from the previous proofs. Our basic tools, described in Section 3, are:

- a combinatorial theorem on abelian groups proved in 1952 by Marshall Hall [10],
- a "coupled system".

From this proof there follows an algorithm which finds zero-sum n-sequences of the given $(2n-1)$-sequence of an abelian group G, with n elements, in time $\mathcal{O}(n^2)$.

An application of this algorithm to the reconstruction of permutation matrices from their diagonal sums is given in [6].

2. Notation and preliminaries

Let $\mathbb{N}$ denote the set of positive integers and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $a, b \in \mathbb{Z}$, we set $[a, b] = \{x \in \mathbb{Z} | a \leq x \leq b\}$. All abelian groups will be written additively.

Let G be a finite abelian group with n elements, let $\mathcal{F}(G)$ denote the free abelian monoid with basis G, and let $S \in \mathcal{F}(G)$. Then S is called a *sequence over* G with length l, and it will be written in the form

$$S = \prod_{i=1}^{l} g_i = g_1 \cdot \dots \cdot g_l = \prod_{g \in G} g^{\mathsf{v}_g(S)} \quad \text{where all } \mathsf{v}_g(S) \in \mathbb{N}_0.$$

A sequence $T \in \mathcal{F}(G)$ is called a *subsequence* of S, if $\mathsf{v}_g(T) \leq \mathsf{v}_g(S)$ for every $g \in G$. The unit element $1 \in \mathcal{F}(G)$ is called the *empty sequence*. We denote by

- $|S| = l = \sum_{g \in G} \mathsf{v}_g(S) \in \mathbb{N}_0$ the *length* of S,
- $\sigma(S) = \sum_{i=1}^{l} g_i = \sum_{g \in G} \mathsf{v}_g(S) g \in G$ the *sum* of S.

For $g \in G$ and $k \in \mathbb{N}_0$, we introduce the following notation and definitions:

- $M^k = \{S \in \mathcal{F}(G) : |S| = k\}$, denotes the set of sequences S having length k;
- $M_g^k = \{S \in \mathcal{F}(G) : \sigma(S) = g, |S| = k\}$, denotes the set of sequences S having sum g and length k;
- P denotes the group of permutations of G;
- $\mathcal{L} = P \times G$ denotes the set of "labelled permutations" of G;
- for $\pi \in P$ and $S \in M^n$ write
$$\pi \vdash S \quad \text{if } S = \{a + \pi(a) : a \in G\};$$
- for $\pi \in P$, $g \in G$ (i.e., $(\pi, g) \in \mathcal{L}$) and $S \in M^{n-1}$ write
$$(\pi, g) \vdash S \quad \text{if } S = \{a + \pi(a) : a \in G \setminus \{g\}\};$$
- for $\pi \in P$ and $g, g', h \in G$, write
$$s_{\pi,(g,g')}(h) = \begin{cases} \pi(g) & h = g' \\ \pi(g') & h = g \\ \pi(h) & \text{otherwise} \end{cases}$$

(i.e., π and s differ just by exchanging the values taken by g and g', provided $g \neq g'$).

3. An alternative and constructive proof of the Erdös–Ginzburg–Ziv Theorem

In order to prove the Erdös-Ginzburg-Ziv Theorem, we use the following result.

Theorem 3.1 (*M. Hall*). *If G is a finite abelian group with n elements, and $S \in M^n$, then there exists a permutation π of the elements of G such that $S = \{a + \pi(a) : a \in G\}$ if and only if*

$$\sigma(S) = 0, \tag{1}$$

namely $\sum_{g \in G} \mathsf{v}_g(S) g = 0$.

This result was rediscovered by Salzborn and Szekeres in [13], where they proved the following claim, equivalent to Hall's Theorem.

- If $S \in M^{n-1}$, then there exists a permutation π of the elements of G and an element g of G such that $S = \{a + \pi(a) : a \in G \setminus \{g\}\}$.

Their proof is different from Hall's, but it is at the same "level" and "difficulty". These are constructive proofs of this result that give an algorithm to find a permutation π satisfying the statement of the theorem in time $\mathcal{O}(n^2)$.

We point out that the theorem states that condition (1) is a necessary and sufficient condition for the existence of a permutation π of the elements of G such that $S = \{a + \pi(a) : a \in G\}$. Therefore, we can establish the existence of this permutation in linear time. However, if we want to construct such a permutation, then we have to perform the algorithm deduced by one of the constructive proofs of the theorem, and these algorithms run in $\mathcal{O}(n^2)$ time.

Using the previous notation, Hall's Theorem can be formulated as follows:

$$\forall S \in M^n \exists \pi \in P : \pi \vdash S \Leftrightarrow S \in M_0^n, \tag{2}$$

and the equivalent formulation of Salzborn and Szekeres becomes:

$$\forall S \in M^{n-1} \exists (\pi, g) \in \mathcal{L} : (\pi, g) \vdash S. \tag{3}$$

We now define a *transition system* on $\mathcal{L}$:

- for $(\pi, g), (\pi', g') \in \mathcal{L}$ and $(c, d) \in G \times G$ write $(\pi, g) \stackrel{(c,d)}{\rightarrow} (\pi', g')$ if
  (a) $g + \pi(g') = c$,
  (b) $g' + \pi(g') = d$,
  (c) $\pi' = s_{\pi,(g,g')}$
- if $g = g'$ (or equivalently $c = d$ or $\pi = \pi'$), we have a *degenerate* transition.

More precisely, for each pair $((\pi, g), c) \in \mathcal{L} \times G$ there is a unique pair $((\pi', g'), d) \in \mathcal{L} \times G$, such that $(\pi, g) \stackrel{(c,d)}{\rightarrow} (\pi', g')$. Notice that,

- if $(\pi, g) \vdash S$, with $S \in M^{n-1}$ and $(\pi, g) \stackrel{(c,c)}{\rightarrow} (\pi, g)$ (i.e., a degenerate transition), then $\pi \vdash Sc$, with $Sc \in M_0^n$ as consequence that $\sigma(Sc) = 2\sum_{g \in G} g = 0$.

The main tool of the proof is the following concept of a *coupled system*, of which we give an informal description.

Let $\mathcal{A}$, $\mathcal{B}$ be two copies of $\mathcal{L}$ that are interacting by sending each other "messages" (i.e., elements from G) in an alternating way:

- if $\mathcal{A}$ is in state (α, a) and receives c from $\mathcal{B}$, it performs the transition $(\alpha, a) \stackrel{(c,d)}{\rightarrow} (\alpha', a')$ and sends d to $\mathcal{B}$;
- if $\mathcal{B}$ is in state (β, b) and receives d from $\mathcal{A}$, it performs the transition $(\beta, b) \stackrel{(d,c')}{\rightarrow} (\beta', b')$ and sends c' to $\mathcal{A}$;
- initially $\mathcal{A}$ is in state (α_0, a_0) and β is in state (β_0, b_0) and $\mathcal{A}$ receives $c_0 \in G$.

This process generates transition sequences:

$$\begin{array}{lccccccc}
\text{in } \mathcal{A}: & (\alpha_0, a_0) & \xrightarrow{(c_0,d_0)} & (\alpha_1, a_1) & \xrightarrow{(c_1,d_1)} & (\alpha_2, a_2) & \xrightarrow{(c_2,d_2)} & \cdots \\
\text{in } \mathcal{B}: & & (\beta_0, b_0) & \xrightarrow{(d_0,c_1)} & (\beta_1, b_1) & \xrightarrow{(d_1,c_2)} & (\beta_2, b_2) & \cdots
\end{array}$$

More formally, let (α_0, a_0), (β_0, b_0) be in $\mathcal{L}$, and let $c_0 \in G$. The *coupled system* determined by (α_0, a_0), (β_0, b_0) and c_0 is the 4-tuple of sequences

$$(\{(\alpha_j, a_j)\}, \{c_j\}, \{(\beta_j, b_j)\}, \{d_j\}) \quad \text{with } j \geq 0,$$

determined by the conditions

$$(\alpha_j, a_j) \stackrel{(c_j,d_j)}{\rightarrow} (\alpha_{j+1}, a_{j+1}) \quad \text{with } j \geq 0,$$

$$(\beta_j, b_j) \stackrel{(d_j,c_{j+1})}{\rightarrow} (\beta_{j+1}, b_{j+1}) \quad \text{with } j \geq 0.$$

We have the following properties. Let $A_j, B_j \in M^{n-1}$ with $(\alpha_j, a_j) \vdash A_j$ and $(\beta_j, b_j) \vdash B_j$. Then

$$A_{j+1} = d_j^{-1} A_j c_j \qquad B_{j+1} = c_{j+1}^{-1} B_j d_j. \tag{4}$$

From these relations it is easy to deduce that:

$$a_{j+1} = a_j + d_j - c_j, \qquad b_{j+1} = b_j + c_{j+1} - d_j. \tag{5}$$

Define $a_{-1}, b_{-1}, c_{-1}, d_{-1}$ so that (5) holds for $j = -1$ as well. The process ends when it produces a degenerate transition, that is $(\alpha_i, a_i) \stackrel{(c_i,c_i)}{\rightarrow} (\alpha_i, a_i)$ or $(\beta_i, b_i) \stackrel{(d_i,d_i)}{\rightarrow} (\beta_i, b_i)$, namely $A_i c_i \in M_0^n$ or $B_i d_i \in M_0^n$. Hence either the coupled system produces a degenerate transition or the process continues indefinitely. We show that this second alternative cannot arise. Assume that the process continues indefinitely. Since $a_0, \ldots, a_j$ and $b_0, \ldots, b_j$ are chosen from the finite group G, we have two possibilities:

1. there exist i and j such that $j > i$, $a_0, \ldots, a_i, \ldots, a_j$ are all distinct, $b_0, \ldots, b_j$ are all distinct, but $a_{j+1} = a_i$;
2. there exist i and j such that $j > i$, $a_0, \ldots, a_{j+1}$ are all distinct, $b_0, \ldots, b_i, \ldots, b_j$ are all distinct, but $b_{j+1} = b_i$.

As we have assumed there is no degenerate transition, we must have $i < j$. In the first case, by relation (5), $a_{k+1} + b_k - d_k = a_k + b_{k-1} - d_{k-1}$ for all $k \geq 0$, and so by iteratively applying this we obtain $a_{j+1} + b_j - d_j = a_i + b_{i-1} - d_{i-1}$. Since $a_{j+1} = a_i$, we have

$$b_j - d_j = b_{i-1} - d_{i-1}. \tag{6}$$

From $(\alpha_j, a_j) \overset{(c_j,d_j)}{\rightarrow} (\alpha_{j+1}, a_{j+1})$, it follows that $a_{j+1} + \alpha_j(a_{j+1}) = d_j$. By $a_{j+1} = a_i$, we get $a_i + \alpha_j(a_i) = d_j$. The coupled system provides:

$$(\alpha_i, a_i) \xrightarrow{(c_i,d_i)} (\alpha_{i+1}, a_{i+1}) \xrightarrow{(c_{i+1},d_{i+1})} \cdots \xrightarrow{(c_{j-1},d_{j-1})} (\alpha_j, a_j) \xrightarrow{(c_j,d_j)} (\alpha_{j+1}, a_i)$$

and since $a_i, a_{i+1}, \ldots, a_j$ are all distinct, we obtain:

$$c_i = a_i + \alpha_{i+1}(a_i) = a_i + \alpha_{i+2}(a_i) = \cdots = a_i + \alpha_{j-1}(a_i) = a_i + \alpha_j(a_i) = d_j.$$

Thus, by relation (6), $b_j = b_{i-1} + c_i - d_{i-1}$. Finally, by relation (5), $b_i = b_{i-1} + c_i - d_{i-1}$, and so $b_j = b_i$. Since, by hypothesis, $b_0, \ldots, b_i, \ldots b_j$ are all distinct, this is a contradiction. In the second case, by proceeding in the same way, from $b_{j+1} = b_i$ we obtain that $a_{j+1} = a_{i+1}$. But this is a contradiction (since $j > i$). In fact, by hypothesis $a_0, \ldots, a_{i+1}, \ldots, a_{j+1}$ are all distinct. Consequently, the coupled system produces a degenerate transition in i steps, with $i < n$. More precisely:

Lemma 3.2. *Let G be an abelian group with n elements, $(\alpha_0, a_0) \vdash A_0$, $(\beta_0, b_0) \vdash B_0$, $c_0 \in G$ be a starting position of the coupled system. Then, there exists an integer $i < n$ such that either $\alpha_i \vdash A_i c_i$ with $A_i c_i \in M_0^n$, or $\beta_i \vdash B_i d_i$ with $B_i d_i \in M_0^n$.*

This lemma allows us to prove the Erdös–Ginzburg–Ziv Theorem:

Proof. Let $S = \prod_{i=1}^{2n-1} g_i$ be a sequence of (not necessarily distinct) elements of the abelian group G. Let $A_0 = \prod_{i=1}^{n-1} g_i$, $B_0 = \prod_{i=n}^{2n-2} g_i$ and $c_0 = g_{2n-1}$. Since $A_0, B_0 \in M^{n-1}$, by Hall's Theorem, there exist $(\alpha_0, a_0), (\beta_0, b_0) \in \mathcal{L}$ such that $(\alpha_0, a_0) \vdash A_0$ and $(\beta_0, b_0) \vdash B_0$. By Lemma 3.2, there exists $i < n$ such that the coupled system produces a degenerate transition, that is $\alpha_i \vdash A_i c_i$ or $\beta_i \vdash B_i \cup \{d_i\}$. Therefore $A_i c_i$ or $B_i d_i \in M_0^n$ is a zero-sum n-sequence; moreover, in view of (4), we have $S = A_j B_j c_j$ for all j, and thus these are both subsequences of S. ■

4. The algorithm

The previous proof provides an algorithm which finds a zero-sum n-sequence of the given $(2n-1)$-sequence of G. The basic steps of the algorithm are:

Algorithm

Input: A sequence $S = \prod_{i=1}^{2n-1} g_i$ of (not necessarily distinct) elements of the abelian group G with n elements;

Output: a zero-sum n-subsequence $Z = z_1 \cdot \cdots \cdot z_n$ of S;

1. **Let** $A_0 := \prod_{i=1}^{n-1} g_i$, $B_0 := \prod_{i=n}^{2n-2} g_i$ and $c_0 := g_{2n-1}$;
   perform the algorithm deduced by the proof of Hall's Theorem. This step finds $(\alpha_0, a_0), (\beta_0, b_0) \in \mathcal{L}$ such that $(\alpha_0, a_0) \vdash A_0$ and $(\beta_0, b_0) \vdash B_0$;
2. **Let** $(\alpha_0, a_0) \vdash A_0$, $(\beta_0, b_0) \vdash B_0$, c_0 be the starting position of the coupled system;
   perform the coupled system until a degenerate transition is produced;
   if $A_i c_i \in M_0^n$ **then** $Z := A_i c_i$ **else** $Z := B_i d_i$.

The algorithm deduced by the proof of Hall's Theorem performs a system which is similar to our coupled system. We call this tool the *Hall system*, and it is described in [10]. Since the Hall system is performed only once for each element of A_0 and B_0 (i.e., $2(n-1)$ times), every time it generates at most n transitions, and every transition requires a constant time, we have that the computational complexity of the first step of the algorithm is $\mathcal{O}(n^2)$.

From Lemma 3.2, it follows that the coupled system achieves a degenerate transition after at most $2n$ transitions. Since every transition can be generated in constant time by keeping the permutations α_i, β_i and their inverses, we obtain that the computational complexity of the second step is $\mathcal{O}(n)$.

Consequently, the algorithm finds a zero-sum n-sequence of the given $(2n-1)$-sequence $g_1 \cdot \cdots \cdot g_{2n-1}$ in time $\mathcal{O}(n^2)$.

Example. Let 43303525143 be a sequence of elements of the group $G = \mathbb{Z}_6$. We set $A_0 = 43303$, $B_0 = 52514$ and $c_0 = 3$. By performing the first step of the algorithm, we find $(\alpha_0, a_0) \vdash A_0$ and $(\beta_0, b_0) \vdash B_0$, where

$$\alpha_0 = (4, 2, 1, 3, 5, 0), \qquad a_0 = 5, \qquad \beta_0 = (5, 1, 3, 4, 0, 2), \qquad b_0 = 5.$$

Then, we run the second step:

$$\begin{array}{llllll}
(\alpha_0, a_0) \overset{(c_0,d_0)}{\rightarrow} (\alpha_1, a_1), & \text{where} & \alpha_1 = (0, 2, 1, 3, 5, 4), & a_1 = 0,\ d_0 = 4, & A_1 = \{3, 3, 0, 3, 3\} \\
(\beta_0, b_0) \overset{(d_0,c_1)}{\rightarrow} (\beta_1, b_1), & \text{where} & \beta_1 = (2, 1, 3, 4, 0, 5), & b_1 = 0,\ c_1 = 5, & B_1 = \{2, 5, 1, 4, 4\} \\
(\alpha_1, a_1) \overset{(c_1,d_1)}{\rightarrow} (\alpha_2, a_2), & \text{where} & \alpha_2 = (5, 2, 1, 3, 0, 4), & a_2 = 4,\ d_1 = 3, & A_2 = \{5, 3, 3, 0, 3\} \\
(\beta_1, b_1) \overset{(d_1,c_2)}{\rightarrow} (\beta_2, b_2), & \text{where} & \beta_2 = (3, 1, 2, 4, 0, 5), & b_2 = 0,\ c_2 = 5, & B_2 = \{3, 2, 1, 4, 4\} \\
(\alpha_2, a_2) \overset{(c_2,d_2)}{\rightarrow} (\alpha_3, a_3), & \text{where} & \alpha_3 = (5, 2, 0, 3, 1, 4), & a_3 = 2,\ d_2 = 3, & A_3 = \{5, 3, 0, 5, 3\} \\
(\beta_2, b_2) \overset{(d_2,c_3)}{\rightarrow} (\beta_3, b_3), & \text{where} & \beta_3 = (3, 2, 1, 4, 0, 5), & b_3 = 1,\ c_3 = 2, & B_3 = \{3, 3, 1, 4, 4\} \\
(\alpha_3, a_3) \overset{(c_3,d_3)}{\rightarrow} (\alpha_4, a_4), & \text{where} & \alpha_4 = \alpha_3, & a_4 = 2,\ d_3 = 2, & A_4 = A_3.
\end{array}$$

The coupled system provides a degenerate transition system $(\alpha_3, a_3) \overset{(c_3,c_3)}{\rightarrow} (\alpha_3, a_3)$ and the process ends. Therefore, $\alpha_3 \vdash A_3c_3 = 530532 \in M_0^6$, and so $Z = 530532$ is a zero-sum 6-sequence of the given 11-sequence of $\mathbb{Z}_6$.

References

[1] N. Alon, Y. Caro, On three zero-sum Ramsey-type problems, Graph Theory 17 (1993) 177–192.
[2] N. Alon, M. Dubiner, Zero-sum sets of prescribed size, in: Combinatorics, in: Paul Erdös is Eighty, vol. 1, Keszthely, Hungary, 1993, pp. 33–50.
[3] A. Bialostocki, P. Dierker, On the Erdös–Ginzburg–Ziv theorem and the Ramsey numbers for star matching, Discrete Mathematics 110 (1992) 1–8.
[4] Y. Caro, Zero-sum problems—A survey, Discrete Mathematics 152 (1996) 93–113.
[5] Y. Caro, Remarks on a Zero-sum theorem, Journal of Combinatorial Theory, Series A 76 (1996) 315–322.
[6] A. Del Lungo, Reconstructing permutation matrices from diagonal sums, Theoretical Computer Science 281 (2002) 235–249.
[7] P. Erdös, A. Ginzburg, A. Ziv, A theorem in additive number theory, Bull. Res. Council, Israel F 10 (1961) 41–43.
[8] Z. Füredi, D.J. Kleitman, On zero-trees, Journal of Graph Theory 16 (1992) 107–120.
[9] W.D. Gao, A combinatorial problem on finite abelian groups, Journal of Number Theory 58 (1996) 100–103.
[10] M. Hall Jr., A combinatorial problem on abelian groups, Proceedings of American Mathematical Society 3 (1952) 548–587.
[11] Y.O. Hamidoune, On weighted sums in abelian groups, Discrete Mathematics 162 (1996) 127–132.
[12] Y.O. Hamidoune, private communication, 2002.
[13] F. Salzborn, G. Szekeres, A problem in combinatorial group theory, Ars Combinatoria 7 (1979) 3–5.
[14] L. Schrijver, P.D. Seymour, A simpler proof and a generalization of the zero-trees theorem, Journal of Combinatorial Theory Series A 58 (1991) 301–305.