AN ADDITION THEOREM ON THE INTEGERS MODULO n

By

Paul Lemke

and

Daniel Kleitman

IMA Preprint Series # 429

July, 1988

AN ADDITION THEOREM ON THE INTEGERS MODULO n

PAUL LEMKE* AND DANIEL KLEITMAN†

Abstract. We show that, given positive integers n, d, and $a_1, \ldots, a_d$ such that $d|n$ and $a_i|n\ \forall i$, there exists a non-empty subset S of $\{1, \ldots, d\}$ such that $d|\sum_{i\in S} a_i$ and $\sum_{i\in S} a_i \leq n$. This resolves a conjecture of P. Erdös and P. Lemke.

1. Introduction. Let d be a divisor of n and suppose we have d divisors of n, not necessarily distinct, which we write as $\{a_i\}$ for $i = 1, 2, \ldots, d$.

It is obvious that one can find a non-empty subset S of $\{1, 2, \ldots d\}$ such that the sum of the a_i's having index in S is a multiple of d. Erdös and Lemke ([1]) framed the question: can one always find such a multiple that is at most n? In particular, one can consider the case $d = n$, in which the question becomes: Given d divisors of d, not necessarily distinct, can one always find a subset of them that sums to d? This was the problem first considered by Lemke.

We prove a positive answer to the general question below.

We do so by proving the following statement, from which it follows easily:

THEOREM 1. *Given a positive integer d and integers $a_1, a_2, \ldots, a_d$, there exists a non-empty set $Q \subseteq \{1, 2, \ldots, d\}$ such that:*

$$d \mid \sum_{i\in Q} a_i,$$

$$\text{and} \qquad \sum_{i\in Q} \gcd(a_i, d) \leq d.$$

In the next section we present proof of this statement, and of the Erdös-Lemke conjecture from it.

In the final section we describe a number of related open questions.

2. Main Result. We begin by proving the Erdös-Lemke conjecture from Theorem 1:

*Institute for Mathematics and its Applications, 514 Vincent Hall, 206 Church St. S.E., University of Minnesota, Minneapolis, MN 55455

†Department of Mathematics, MIT, Cambridge, Massachusetts 02139. AFOSR-86-0078, NSF 8606225.

COROLLARY. *Given positive integers n, d and $a_1, a_2, \ldots, a_d$ with $d|n$ and $a_i|n$ for $i = 1, 2, \ldots, d$, there is a non-empty subset S of $\{1, 2, \ldots, d\}$ such that*

$$d \mid \sum_{i \in S} a_i$$

$$\text{and} \qquad \sum_{i \in S} a_i \leq n.$$

Proof: From Theorem 1 there is a subset S of $\{1, 2, \ldots, d\}$ for which $d | \sum_{i \in S} a_i$ and $\sum_{i \in S} \gcd(a_i, d) \leq d$. Since both a_i and d divide n for each i, we have

$$a_i = \gcd(a_i, n) \leq \tfrac{n}{d} \gcd(a_i, d).$$

We conclude that $\sum_{i \in S} a_i \leq n$. □ □

We note that the ideas of the Theorem below can be used in a straightforward way to construct S from the a_i's efficiently. One therefore knows not only that such an S exists, but that it can easily be found.

The idea of the proof is as follows. It is trivial if d is a prime, and we procede by induction on the number of prime factors of d.

An argument is presented that handles the case in which d's largest prime factor is 2 or 3 (Case II, below).

Otherwise we use the induction hypothesis on $\frac{n}{p}$, where p is a prime factor of n which is greater than or equal to 5.

In particular if there are enough a_i's that are multiples of p, we can divide them by p, employ that hypothesis and multiply back to obtain the conclusion. (Case I below). This could be done using only divisors relatively prime to n or divisible by p.

Otherwise, we form a construction that only uses integers that are relatively prime to p, again making use of the induction hypothesis (Cases III, IV and V).

We now present the proof.

Proof of Theorem 1: First, we make the inductive assumption that Theorem 1 is true for all values of d smaller than some given integer n, and then prove that it holds for $d = n$. The proof is trivial if n is a prime, so we assume that n is composite.

Now, if p is any prime divisor of n, define:

$$S_p = \{i | p \text{ divides } a_i\},$$
$$T_p = \{i | \gcd(a_i, n) \geq 2\} - S_p,$$
$$\text{and} \quad Z = \{i | \gcd(a_i, n) = 1\}.$$

Then $S_p \cap T_p = S_p \cap Z = T_p \cap Z = \phi$ and $S_p \cup T_p \cup Z = \{1, 2, \ldots, n\}$.

CASE I: $|S_p| + \lfloor \frac{|Z|}{p} \rfloor \geq \frac{n}{p}$ *for some prime divisor* p *of* n.

If $|S_p| \geq \frac{n}{p}$ then we choose a cardinality $\frac{n}{p}$ subset of S_p, call it $\overline{S}_p$, and then use the inductive assumption to apply Theorem 1 with $d = \frac{n}{p}$ and the d integers $\{\frac{a_i}{p}\}$, $i \in \overline{S}_p$, to obtain a non-empty set $Q \subseteq \overline{S}_p$ such that

$$\frac{n}{p} \mid \sum_{i \in Q} \frac{a_i}{p}$$

$$\text{and} \qquad \sum_{i \in Q} \gcd(\tfrac{a_i}{p}, \tfrac{n}{p}) \leq \tfrac{n}{p}.$$

Multiplying each of these through by p gives the desired result.

If $|S_p| < \frac{n}{p}$, then since $\lfloor \frac{|Z|}{p} \rfloor \geq \frac{n}{p} - |S_p|$, we can choose $\frac{n}{p} - |S_p|$ disjoint cardinality p subsets $A_1, A_2, \ldots, A_{\frac{n}{p}-|S_p|} \subseteq Z$, and using Theorem 1 with $d = p$ we can find non-empty sets $Q_1, Q_2, \ldots, Q_{\frac{n}{p}-|S_p|}$ with

$$Q_j \subseteq A_j \quad \text{for} \quad j = 1, 2, \ldots, \tfrac{n}{p} - |S_p|$$

$$\text{and} \qquad p \mid \sum_{i \in Q_j} a_j \quad \text{for} \quad j = 1, 2, \ldots, \tfrac{n}{p} - |S_p|.$$

Letting $b_1, b_2, \ldots, b_{|S_p|}$ be the members of S_p (in arbitrary order), we then define the sets $\{Q_j\}$ for $\frac{n}{p} - |S_p| < j \leq \frac{n}{p}$ by:

$$Q_j = \{b_{j-\frac{n}{p}+|S_p|}\} \quad \text{for} \quad j = \tfrac{n}{p} - |S_p| + 1, \tfrac{n}{p} - |S_p| + 2, \ldots, \tfrac{n}{p}.$$

We again use induction with $d = \frac{n}{p}$ and the integers $\{\frac{1}{p}\sum_{i \in Q_j} a_i\}$, $j = 1, 2, \ldots, \frac{n}{p}$, to obtain a non-empty set $R \subseteq \{1, 2, \ldots, \frac{n}{p}\}$ such that

$$\frac{n}{p} \mid \sum_{j \in R} \frac{1}{p} \sum_{i \in Q_j} a_i$$

$$\text{and} \qquad \sum_{j \in R} \gcd(\tfrac{1}{p} \sum_{i \in Q_j} a_i, \tfrac{n}{p}) \leq \tfrac{n}{p},$$

or, after multiplying through by p,

(1)
$$n \mid \sum_{j \in R} \sum_{i \in Q_j} a_i$$

$$\text{and} \qquad \sum_{j \in R} \gcd(\sum_{i \in Q_j} a_i, n) \leq n.$$

Since $\gcd(a_i, n) = 1$ for $i \in Z$ and since $Q_j \subseteq Z$ for $j = 1, 2, \ldots, \frac{n}{p} - |S_p|$, we have

$$\sum_{i \in Q_j} \gcd(a_i, n) = |Q_j| \leq p$$

for these j, and also since $p \mid \sum_{i\in Q_j} a_i$ for these j, we have

$$\gcd(\sum_{i\in Q_j} a_i, n) \geq p \text{ for } j = 1, 2, \ldots, \tfrac{n}{p} - |S_p|.$$

Also, $|Q_j| = 1$ for $j = \frac{n}{p} - |S_p| + 1, \frac{n}{p} - |S_p| + 2, \ldots, \frac{n}{p}$, so that

$$\gcd(\sum_{i\in Q_j} a_i, n) = \sum_{i\in Q_j} \gcd(a_i, n)$$

for these j. Thus we have

$$\sum_{i\in Q_j} \gcd(a_i, n) \leq \gcd(\sum_{i\in Q_j} a_i, n) \quad \text{for} \quad j = 1, 2, \ldots, \tfrac{n}{p},$$

and thus

$$\sum_{j\in R} \sum_{i\in Q_j} \gcd(a_j, n) \leq \sum_{j\in R} \gcd(\sum_{i\in Q_j} a_i, n) \leq n. \tag{2}$$

Letting $Q = \bigcup_{j\in R} Q_j$, (1) and (2) give the required result.

We will assume from this point on that Case I does not apply.

CASE II: $\sum_{i=1}^{k} \frac{1}{p_i} \leq 1$, *where* $p_1, p_2, \ldots, p_k$ *are the distinct prime divisors of* n.

Since Case I does not apply, we must have $|S_{p_i}| + \left\lfloor \frac{|Z|}{p_i} \right\rfloor \leq \frac{n}{p_i} - 1$ for each i, which implies

$$|S_{p_i}| + \tfrac{|Z|}{p_i} \leq \tfrac{n}{p_i} - \tfrac{1}{p_i}, \quad i = 1, 2, \ldots, k. \tag{3}$$

Summing (3) over all i gives

$$\sum_{i=1}^{k} |S_{p_i}| + |Z| \sum_{i=1}^{k} \tfrac{1}{p_i} \leq n \sum_{i=1}^{k} \tfrac{1}{p_i} - \sum_{i=1}^{k} \tfrac{1}{p_i}. \tag{4}$$

Since $S_{p_1} \cup S_{p_2} \cup \cdots \cup S_{p_k} \cup Z = \{1, 2, \ldots, n\}$ we have

$$\sum_{i=1}^{k} |S_{p_i}| + |Z| \geq n. \tag{5}$$

Multiplying (5) by $\sum_{i=1}^{k} \frac{1}{p_i}$ and subtracting it from (4) gives:

$$\left(1 - \sum_{i=1}^{k} \tfrac{1}{p_i}\right) \sum_{i=1}^{k} |S_{p_i}| \leq -\sum_{i=1}^{k} \tfrac{1}{p_i},$$

which is impossible. We may therefore assume without loss of generality that $\sum_{i=1}^{k} \frac{1}{p_i} > 1$, which implies that n has at least 3 distint prime divisors, and that at least one of them is ≥ 5. We can therefore assume from now on that $p \geq 5$.

CASE III. $|T_p| \geq \frac{3n}{2p}$.

Let A_1 be an arbitrary subset of T_p of cardinality $\frac{n}{p}$. Then we know from the inductive assumption that there exists a non-empty subset Q_1 of A_1 such that

$$\frac{n}{p} \mid \sum_{i \in Q_1} a_i \tag{6}$$

$$\text{and} \qquad \sum_{i \in Q_1} \gcd(a_i, \tfrac{n}{p}) \leq \tfrac{n}{p}. \tag{7}$$

By definition of T_p we know that $p \nmid a_i$ and $\gcd(a_i, n) \geq 2$ for $i \in T_p$. It follows that $\gcd(a_i, \frac{n}{p}) = \gcd(a_i, n) \geq 2$, so that for (7) to hold we must have $|Q_1| \leq \frac{n}{2p}$. We can then choose a set A_2 of cardinality $\frac{n}{p}$ from $T_p - Q_1$ and again use the inductive assumption to extract from it a non-empty subset Q_2 with the same properties (6) and (7) that hold for Q_1, and we will then also have $|Q_2| \leq \frac{n}{2p}$. Since Case I does not apply we know that $|S_p| \leq \frac{n}{p}$, and since we have

$$|S_p| + |T_p| + |Z| = n \tag{8}$$

and $T_p \cap Z = \phi$, we have $|T_p \cup Z - Q_1 - Q_2| \geq \frac{(p-2)n}{p}$. We can therefore select $p-2$ disjoint sets $A_3, A_4, \ldots, A_p$ of cardinality $\frac{n}{p}$ from $T_p \cup Z - Q_1 - Q_2$ and then find respective subsets $Q_3, Q_4, \ldots, Q_p$ of these sets with the properties (6) and (7) that hold for Q_1. Using Theorem 1 with $d = p$ and the d integers $\{\frac{p}{n} \sum_{j \in Q} a_i\}$, $j = 1, 2, \ldots, p$, we obtain a non-empty set $R \subseteq \{1, 2, \ldots, p\}$ such that

$$p \mid \sum_{j \in R} \frac{p}{n} \sum_{i \in Q_j} a_i,$$

and therefore

$$n \mid \sum_{j \in R} \sum_{i \in Q_j} a_i.$$

Also, since $p \nmid a_i$ for $i \in T_p \cup Z$ we have $\gcd(a_i, \frac{n}{p}) = \gcd(a_i, n)$. From (7) we then get

$$\sum_{j \in R} \sum_{i \in Q_j} \gcd(a_i, n) = \sum_{j \in R} \sum_{i \in Q_j} \gcd(a_i, \tfrac{n}{p}) \leq |R| \sum_{i \in Q_j} \gcd(a_i, \tfrac{n}{p}) \leq |R| \tfrac{n}{p} \leq n,$$

so that with $Q = \cup_{j \in R} Q_j$, the conclusion of Theorem 1 is satisfied for Case III.

CASE IV. $\frac{n}{p} \leq |T_p| < \frac{3n}{2p}$.

Define A_1 and Q_1 as in Case III, so that $Q_1 \subseteq T_p$, $|Q_1| \leq \frac{n}{2p}$, and (6) and (7) hold. Then if $|T_p \cup Z - Q_1| \geq \frac{(p-1)n}{p}$, we can define sets $A_2, A_3, \ldots, A_p$ and $Q_2, Q_3, \ldots, Q_p$ in the

same manner as $A_3, A_4, \ldots, A_p$ and $Q_3, Q_4, \ldots, Q_p$ were defined in Case III, so that the conclusion of Theorem 1 is again satisfied with $Q = Q_1 \cup Q_2 \cup \cdots \cup Q_p$. The alternative is if $|T_p \cup Z - Q_1| < \frac{(p-1)n}{p}$, which implies

$$|T_p| + |Z| < \tfrac{(2p-1)n}{2p}, \tag{9}$$

since $T_p \cap Z = \phi$ and $|Q_1| \leq \frac{n}{2p}$. However, if we multiply (8) by p and subtract (9) times $p-1$, and then subtract (3) times p, we get:

$$|T_p| > pn - \tfrac{(p-1)(2p-1)n}{2p} - n + 1 = \tfrac{(p-1)n}{2p} + 1,$$

which is a contradiction to the premise of Case IV when $p \geq 4$. We are therefore left with:

CASE V. $|T_p| < \frac{n}{p}$.

Here, we choose A_1 to be a subset of $T_p \cup Z$ of cardinality $\frac{n}{p}$ containing T_p. We then find $Q_1 \subseteq A_1$ as in Case III. We have:

$$\tfrac{n}{p} \geq \sum_{i \in Q_1} \gcd(a_i, \tfrac{n}{p}) = \sum_{i \in Q_1 \cap T_p} \gcd(a_i, \tfrac{n}{p}) + \sum_{i \in Q_1 \cap Z} \gcd(a_i, \tfrac{n}{p}) \geq 2|Q_1 \cap T_p| + |Q_1 \cap Z|$$

$$\text{and} \quad |Q_1 \cap Z| \leq |A_1 \cap Z| = |A_1| - |A_1 \cap T_p| = \tfrac{n}{p} - |T_p|.$$

We may then add $\frac{n}{p} \geq 2|Q_1 \cap T_p| + |Q_1 \cap Z|$ to $\frac{n}{p} - |T_p| \geq |Q_1 \cap Z|$ and divide the result by 2 to obtain:

$$\tfrac{n}{p} - \tfrac{1}{2}|T_p| \geq |Q_1 \cap T_p| + |Q_1 \cap Z| = |Q_1|. \tag{10}$$

If $|T_p \cup Z - Q_1| \geq \frac{(p-1)n}{p}$, then as in Case IV, the conclusion of Theorem 1 follows. Otherwise, we have

$$|T_p| + |Z| - |Q_1| < \tfrac{(p-1)n}{p},$$

which, together with (10), implies $|T_p| + |Z| < \frac{(p-1)n}{p} + \frac{n}{p} - \frac{1}{2}|T_p|$, or

$$\tfrac{3}{2}|T_p| + |Z| < n. \tag{11}$$

However, if we add twice (11) to p times (3) and then subtract 3 times (8) we get:

$$(p-3)|S_p| < -1,$$

which is impossible for $p \geq 3$, so the proof of Theorem 1 is complete. □ □

3. Open Problems. The following questions are raised by these results.

Question 1. What happens to the Corollary in the Main Result if we drop the requirement that $d|n$, asking instead for the minimum value $\frac{1}{n}\sum_{i\in S} a_i$ may have, subject to the constraints $d|\sum_{i\in S} a_i$, $S \neq \phi$? J. Selfridge has conjectured ([2]) that if we define $f(d)$ to be the maximum value of that minimum over all n and all sequences $a_1, a_2, \ldots a_n$, then

$$f(d) = \sum_{i=0}^{d-1} \frac{1}{id+1}.$$

(It is easy to show that $f(d) \geq \sum_{i=0}^{d-1} \frac{1}{id+1}$).

Question 2. Theorem 1 may be generalized to the following conjecture on groups.

CONJECTURE: Any sequence of $|G|$ elements (not necessarily distinct) of the finite group G contains a non-empty subsequence $g_1, g_2, \ldots, g_k$ such that:

$$g_1 g_2 \cdots g_k = e$$

and

$$\sum_{i=1}^{k} \frac{1}{|g_i|} \leq 1.$$

It may be seen that Theorem 1 proves the special case of the above conjecture where G is a cyclic group. There is also a simple proof of the above conjecture when $G = (\mathbf{Z}_p)^n$, p a prime, or when G is a dihedral or dicyclic group, and we have verified it for all groups G with $|G| \leq 15$.

Question 2 may be generalized even further to the following:

CONJECTURE: If G is a finite group and H is a subgroup of G, then every sequence of $|G|$ elements of G contains a subsequence $g_1, g_2, \ldots, g_k$, with $1 \leq k \leq \frac{|G|}{|H|}$, such that

$$g_1 g_2 \cdots g_k \in H,$$

and

$$\sum_{i=1}^{k} \frac{1}{|g_i|} \leq \frac{1}{|g_1 g_2 \cdots g_k|}.$$

We have verified this conjecture for $|G| \leq 11$. If $H = \{e\}$, then it reduces to the previous conjecture.

If $G = \mathbf{Z}_n$ and $H = \{0, d, 2d, \ldots, (\frac{n}{d} - 1)d\}$, then the above conjecture reduces to:

CONJECTURE: If n and d are positive integers such that $d|n$, and if $a_1, a_2, \ldots, a_n$ are integers, then there exists $S \subseteq \{1, 2, \ldots, n\}$ such that:

$$d|\sum_{i\in S} a_i,$$

$$\sum_{i\in S} \gcd(a_i, n) \leq \gcd(\sum_{i\in S} a_i, n),$$

and $\quad 1 \leq |S| \leq d.$

This number theoretical conjecture can then be generalized further to the following conjecture, which has no analogue for groups:

CONJECTURE: Given positive integers n and d with $d|n$, and integers $a_1, a_2, \ldots, a_n$ not divisible by d, there exists an integer m relatively prime to n and a subset S of $\{1, 2, \ldots, n\}$ such that:

$$d|\sum_{i\in S} \mathrm{mod}(ma_i, n),$$

and $\quad \sum_{i\in S} \mathrm{mod}(ma_i, n)|n,$

where "mod (k, n)" denotes the least non-negative residue of k modulo n.

This conjecture has been verified for $n \leq 11$. It is not trivial even when $d = n$ and n is a prime, and, in fact, we have no proof even for this case.

Another set of potentially interesting related questions is exemplified by the following: "How small can one reduce the number of a_i's to in Theorem 1, or, more generally, in the group theory problem of Question 2, given that k of them are distinct?" For example, Eggleton and Erdös ([3]) have shown that any sequence of n elements of a finite Albelian group G, exactly k of which are distinct, contains a non-empty subsequence $g_1, g_2, \ldots, g_\ell$ such that

$$g_1 g_2 \ldots g_\ell = e, \tag{12}$$

provided that $n \geq \max(|G| - \binom{k}{2}, k\binom{k}{2})$. This result may well also hold for our case, where we also require that

$$\sum_{i=1}^{\ell} \frac{1}{|g_i|} \leq 1. \tag{13}$$

(It also appears quite likely to be true if G is non-albelian.) They also show in [3] that $n \geq |G| - k + 1$ is also sufficient to imply the existence of a subsequence satisfying (12). This

result does not hold in our case for general groups because, for example, we need $n = 2^m$ to find a subsequence satisfying both (12) and (13) if $G = (\mathbf{Z}_2)^m$, even if $k = 2^m - 1$. However, $n \geq |G| - k + 1$ may well be sufficient in our case for cyclic groups.

If the n elements in the original sequence are *all* distinct, it follows immediately from a result of Olson ([4]) that there exists a non-empty sequence satisfying (12) and (13) if $n > (4|G| - 3)^{\frac{1}{2}}$ when G is a cyclic group of prime order. In [5] Szemerédi shows that a solution to (12) exists if G is an Abelian group and $n > c|G|^{\frac{1}{2}}$, where c is an absolute constant. While this is not true in general if we require a sub-sequence that satisfies both (12) *and* (13) (because of the example $G = (\mathbf{Z}_2)^m$ given above), it may still hold when G is a cyclic group.

Acknowledgement.

We would like to thank Paul Erdös for his encouragement on this problem. We would also like to thank the Institute for Mathematics and its Applications for its hospitality during our collaboration on this paper.

REFERENCES

[1] P. Erdös, personal communication to P. Lemke, May 1987.
[2] J. Selfridge, personal communication to P. Erdös, December 1987.
[3] R.B. Eggleton and P. Erdös, *Two Combinatorial Problems in Group Theory*, Acta Arithmetica, 21 (1972), pp. 111-116.
[4] John E. Olson, *An Addition Theorem Modulo p*, Journal of Combinatorial Theory, 5 (1968), pp. 45–52.
[5] E. Szemerédi, *On a Conjecture of Erdös and Heilbronn*, Acta Arithmetica, 17 (1970), pp. 227–229.