PACIFIC JOURNAL OF MATHEMATICS
Vol. 133, No. 2, 1988

ON DIVISORS OF SUMS OF INTEGERS, III

CARL POMERANCE, A. SÁRKÖZY AND C. L. STEWART

In this paper we show that if $A_1, A_2, \ldots, A_k$ are "dense" sets of integers, then there is a sum $a_1 + a_2 + \cdots + a_k$ with $a_1 \in A_1$, $a_2 \in A_2, \ldots, a_k \in A_k$ that is divisible by a "small" prime.

1. Let $P(n)$ and $p(n)$ denote the greatest and smallest prime factor of n, respectively. Recently in several papers, Balog, Erdös, Maier, Sárközy, and Stewart have studied problems of the following type: if $A_1, \ldots, A_k$ are "dense" sets of positive integers, then what can be said about the arithmetical properties of the sums $a_1 + \cdots + a_k$ with $a_1 \in A_1, \ldots, a_k \in A_k$? In particular, Balog and Sárközy proved that there is a sum a_1+a_2 ($a_1 \in A_1$, $a_2 \in A_2$) for which $P(a_1+a_2)$ is "small", i.e., all the prime factors of a_1+a_2 are small. On the other hand, Balog and Sárközy and Sárközy and Stewart studied the existence of a sum $a_1 + \cdots + a_k$ for which $P(a_1 + \cdots + a_k)$ is large.

In this paper we study $p(a_1 + \cdots + a_k)$. Our goal is to show that if $A_1, \ldots, A_k$ are sets of positive integers then there exists a sum $a_1 + \cdots + a_k$ with $a_1 \in A_1, \ldots, a_k \in A_k$ that is divisible by a "small" prime. In the most interesting special case, namely $A_1 = \cdots = A_k$, there are sums $a_1 + \cdots + a_k$ divisible by k, so that $p(a_1 + \cdots + a_k) \leq k$. In order to exclude such trivial cases, we shall ask that the "small" prime factor of $a_1 + \cdots + a_k$ also exceeds some prescribed bound V.

In §3 we will study the case when the geometric mean of the cardinalities of the sets $A_i \subset \{1, \ldots, N\}$ is between $\sqrt{N}$ and N. The crucial tool will be the large sieve. In §4 we will extend the range (when $k > 2$) by studying the case $\min_i |A_i| > N^{1/k+\varepsilon}$. Here Gallagher's larger sieve will be used. The results in §§3 and 4 do not give especially good results when the sets $A_1, \ldots, A_k$ are very "dense". In §5, we will give an essentially best possible result for the small prime factors of the sums $a_1 + \cdots + a_k$ in the case when

$$(|A_1| \cdots |A_k|)^{1/k} > N \exp(-c \log k \log N / \log\log N)$$

for a certain positive constant c. Finally in §§6 and 7 we will construct sets so that none of the sums $a_1 + \cdots + a_k$ has a small prime factor. In particular, in §6 we will discuss the conjecture of Ostmann [**6**] that

there do not exist infinite sets of positive integers A_1, A_2 such that $A_1 + A_2$ differs from the set of primes by at most a finite set.

The constants $c_1, c_2, \ldots$ appearing in different sections are independent from each other, so that, for example, c_1 in §3 and c_1 in §4 are not necessarily the same.

2. In this section, for the convenience of the reader, we collect the lemmas of [8]. Primary references for the first three lemmas and a proof of the fourth may be found there. Further lemmas in this paper will be presented as needed in subsequent sections.

LEMMA 1 (*Cauchy-Davenport*). *Let p be a prime number and let A, B be subsets of $\mathbf{Z}/p\mathbf{Z}$. Then*

$$|A + B| \geq \min\{|A| + |B| - 1, p\}.$$

LEMMA 2 (*Large sieve*). *Let A be a set of integers in the interval $[M + 1, M + N]$. For each prime p let $\nu(p)$ denote the number of residue classes* mod p *which contain a member of A. Then for any positive number Q we have*

$$|A| \leq \frac{N + Q^2}{L}$$

where $L = \sum'_{q \leq Q} \prod_{p|q} (p - \nu(p))/\nu(p)$, the dash indicating the sum is over square-free positive integers q.

LEMMA 3 (*Gallagher's larger sieve*). *With the same notation as in Lemma 2, if S is a finite set of primes, then*

$$|A| \leq \left(-\log N + \sum_{p \in S} \log p\right) \Big/ \left(-\log N + \sum_{p \in S} \frac{\log p}{\nu(p)}\right),$$

provided the denominator is positive.

LEMMA 4. *Let p, k be integers with $k \geq 2$ and $p \geq 1 + (k-1)^k$. Then*

$$\min_D \prod_{i=1}^{k} \left(\frac{1}{x_i} - 1\right) = \left(\frac{k}{1 + (k-2)/p} - 1\right)^k,$$

$$\min_D \sum_{i=1}^{k} \frac{1}{x_i} = \frac{k^2}{1 + (k-2)/p},$$

where D is the set of $\vec{x} \in \mathbf{R}^k$ with

$$x_1 + \cdots + x_k \leq 1 + \frac{k-2}{p} \quad \text{and} \quad \frac{1}{p} \leq x_i \leq 1 - \frac{1}{p} \quad \text{for } i = 1, \ldots, k.$$

3. In this section we study the situation when the geometric mean of the sets of integers considered exceeds $\sqrt{N}$.

THEOREM 1. *Let N be a positive integer, let A_1, A_2 be non-empty subsets of $\{1, \ldots, N\}$ and let $T = (|A_1||A_2|)^{1/2}$. Let S be a set of prime numbers, let Q be a positive integer, and let J denote the number of square-free integers up to Q divisible only by the primes in S. If*

$$TJ > N + Q^2, \tag{3.1}$$

then there is some prime $p \in S$ such that each residue class mod p *contains a member of the sum set $A_1 + A_2$.*

Proof. For $i = 1, 2$ and for each prime p let $\nu_i(p)$ denote the number of residue classes mod p that contain an element of A_i. For each $Q \geq 1$, Lemma 2 implies

$$|A_i| \leq (N + Q^2) / \sum_{q \leq Q}{}' \prod_{p \mid q} \frac{p - \nu_i(p)}{\nu_i(p)}$$

for $i = 1, 2$ where the dash indicates the sum is over square-free positive integers q. Thus

$$T \leq (N + Q^2)/H, \tag{3.2}$$

where

$$H = \left(\prod_{i=1}^{2} \sum_{q \leq Q}{}' \prod_{p \mid q} \left(\frac{p}{\nu_i(p)} - 1 \right) \right)^{1/2}.$$

From the Cauchy-Schwarz inequality we have

$$H \geq \sum_{q \in Q}{}' \left(\prod_{p \mid q} \left(\frac{p}{\nu_1(p)} - 1 \right) \left(\frac{p}{\nu_2(p)} - 1 \right) \right)^{1/2}. \tag{3.3}$$

Assume now that for every prime $p \in S$ there is a residue class mod p that contains no member of $A_1 + A_2$. Then from Lemma 1 we have

$$\nu_1(p) + \nu_2(p) \leq p$$

for every prime $p \in S$. We now apply Lemma 4 with $k = 2$ and $x_i = \nu_i(p)/p$. If q is divisible only by primes in S we thus have

$$\left(\prod_{p|q}\left(\frac{p}{\nu_1(p)}-1\right)\left(\frac{p}{\nu_2(p)}-1\right)\right)^{1/2} \geq 1.$$

Using this in (3.3) we get $H \geq J$, so that (3.2) contradicts (3.1). Thus there is some prime $p \in S$ such that each residue class mod p contains a member of $A_1 + A_2$.

COROLLARY 1. *If* $V \geq 5.5$ *and*

$$T > \frac{2N \log V}{V} + 8V \log V, \tag{3.4}$$

then there is some prime p *with* $V < p \leq 2V$ *such that every residue class* mod p *contains a member of* $A_1 + A_2$.

Proof. We apply Theorem 1 with S the set of primes in $(V, 2V]$ and $Q = 2V$. Then

$$J = \pi(2V) - \pi(V) > \frac{V}{2 \log V},$$

by Rosser and Schoenfeld [6]. Thus (3.4) implies (3.1).

REMARKS. Clearly (3.4) implies $T \gg \sqrt{N} \log N$, so the corollary is only applicable in this case. It should be noted that the corollary generalizes Theorem 2 of Balog and Sárközy [3].

COROLLARY 2. *Let* V *be a positive integer. There is a positive number* c_1, *effectively computable in terms of* V, *such that if*

$$T > c_1 N^{1/2} \tag{3.5}$$

then there is a prime p *with*

$$V < p < c_1 N/T$$

such that every residue class mod p *contains a member of* $A_1 + A_2$.

Proof. There are positive constants c_2, c_3 effectively computable in terms of V such that if $Q > c_2$ and if S is the set of primes between V and Q, then $J > c_3 Q$ (where J is as defined in the theorem). Let

$$c_1 = \max\{c_2, 2/c_3\}, \quad Q = c_1 N/T.$$

Then (3.5) implies $Q^2 < N$, so that

$$TJ > c_3TQ = c_1c_3N \geq 2N > N + Q^2.$$

Thus the theorem implies the corollary.

We now generalize the situation of Theorem 1 to where more than two sets of integers are involved. Rather than giving a general result, we content ourselves with a result analogous to Corollary 2.

THEOREM 2. *Let $0 < \varepsilon < 1$, let V and N be positive integers, let $A_1, \ldots, A_k$ be non-empty subsets of $\{1, \ldots, N\}$ where $k \geq 3$, and let $T = (|A_1| \cdots |A_k|)^{1/k}$. There exist positive numbers c_4, c_5, c_6 effectively computable in terms of V, k, ε such that if $N > c_4$ and*

$$T > c_5 N^{1/2}/(\log N)^{k-2-\varepsilon} \tag{3.6}$$

then there is a prime p with

$$V < p < c_6(N/T)/(\log(2N/T))^{k-2-\varepsilon}$$

such that each residue class mod p *contains a member of the sum set $A_1 + \cdots + A_k$.*

Proof. For $i = 1, \ldots, k$ and each prime p, let $\nu_i(p)$ denote the number of residue classes mod p that contain an element of A_i. Thus, as in the proof of Theorem 1, we have (3.2) for any positive number Q, where now

$$H = \left(\prod_{i=1}^{k} \sum_{q \leq Q}{}' \prod_{p \mid q} \left(\frac{p}{\nu_i(p)} - 1 \right) \right)^{1/k}.$$

Using Hölder's inequality $k - 1$ times, we get

$$\begin{aligned} H &\geq \sum_{q \leq Q}{}' \left(\prod_{i=1}^{k} \prod_{p \mid q} \left(\frac{p}{\nu_i(p)} - 1 \right) \right)^{1/k} \\ &= \sum_{q \leq Q}{}' \prod_{p \mid q} \left(\prod_{i=1}^{k} \left(\frac{p}{\nu_i(p)} - 1 \right) \right)^{1/k}. \end{aligned} \tag{3.7}$$

Suppose now that for each prime p with $V < p < Q$ there is some residue class mod p that contains no member of $A_1 + \cdots + A_k$. Then by an iteration of Lemma 1, for each prime p with $V < p < Q$ we have

$$\nu_1(p) + \cdots + \nu_k(p) \leq p + k - 2.$$

Thus with Lemma 4 applied to $x_i = \nu_i(p)/p$ for $i = 1, \ldots, k$ we have

$$\left(\prod_{i=1}^{k}\left(\frac{p}{\nu_i(p)} - 1\right)\right)^{1/k} \geq \frac{k}{1+(k-2)/p} - 1 > k - 1 - \varepsilon,$$

where the last inequality requires $p \geq c_7 = c_7(k, \varepsilon)$. Thus from (3.7) and our assumption, we have

$$H \geq \sum_{q \leq Q}^{*} \prod_{p \mid q}(k - 1 - \varepsilon) = \sum_{q \leq Q}^{*}(k - 1 - \varepsilon)^{\omega(q)}, \tag{3.8}$$

where the star indicates a sum over square-free integers not divisible by any prime smaller than the maximum of V and c_7. Note that $\omega(q)$ denotes the number of prime factors of q.

To estimate the last sum in (3.8) we use the following lemma.

LEMMA 5. *Let $2 \leq y \leq \exp((\log x)^{2/5})$, let $r > 0$, and let z be a complex number with $|z| \leq r$. If * denotes a sum over square-free integers n free of prime factors below y, then*

$$\sum_{n \leq x}^{*} z^{\omega(n)} = g(y, z)x(\log x)^{z-1} + O_r(x|(\log x)^{z-2}(\log y)^{1-z}|)$$

where

$$g(y, z) = \Gamma(z)^{-1}\left(\prod_{p}\left(1 + \frac{z}{p}\right)\left(1 - \frac{1}{p}\right)^{z}\right)\prod_{p<y}\left(1 + \frac{z}{p}\right)^{-1}.$$

Proof. The proof follows from the method of Alladi [1] and Selberg [10]. In Alladi, the same sum is estimated but without the restriction that n be square-free. In Selberg, the same sum is estimated but without the restriction that n be free of prime factors below y.

We now apply Lemma 5 with $r = k - 1$, $z = k - 1 - \varepsilon$, $x = Q$, and $y = \max\{V, c_7\}$. We thus deduce from (3.8) that there are positive constants $c_8 = c_8(V, k, \varepsilon)$ and $c_9 = c_9(V, k, \varepsilon)$ such that if $Q \geq c_8$ we have

$$H \geq c_9 Q(\log Q)^{k-2-\varepsilon}. \tag{3.9}$$

We now choose $c_6 \geq 2^k/c_9$ so large that if

$$Q = c_6 \frac{N}{T} \Big/ \left(\log \frac{2N}{T}\right)^{k-2-\varepsilon}, \tag{3.10}$$

we have

$$Q \geq c_8 \quad \text{and} \quad \log Q > \frac{1}{2}\log\frac{2N}{T}. \tag{3.11}$$

Thus from (3.9), (3.10), and (3.11) we have

$$\frac{N}{H} \leq \frac{N}{c_9 Q(\log Q)^{k-2-\varepsilon}} < \frac{2^{k-2-\varepsilon}}{c_6 c_9} T < \frac{1}{2}T.$$

Putting this last inequality into (3.2) we get (using (3.9), (3.10), and (3.11))

$$T < \frac{2Q^2}{H} \leq \frac{2Q}{c_9(\log Q)^{k-2-\varepsilon}} \leq \frac{2^{k-1-\varepsilon}c_6 N}{c_9 T(\log(2N/T))^{2(k-2-\varepsilon)}}.$$

Thus since $c_6 \geq 2^k/c_9$,

$$T < c_6 N^{1/2}/(\log(2N/T))^{k-2-\varepsilon} < N^{2/3}$$

for $N \geq c_4 = c_4(V,k,\varepsilon)$. Hence $2N/T \geq N^{1/3}$, so that

$$T < c_6 3^k N^{1/2}/(\log N)^{k-2-\varepsilon}. \tag{3.12}$$

Thus if we let $c_5 = c_6 3^k$, then (3.12) contradicts (3.6) which proves the theorem.

4. Theorems 1 and 2 covered the case when the cardinalities of the given sets are greater than $N^{1/2}$. In this section we are going to study the case when all the cardinalities are greater than $N^{1/k+\varepsilon}$ (only when $k \geq 3$ will we obtain something new). Perhaps the conclusion of Theorem 3 holds when we assume only that all the cardinalities are greater than N^{ε}, but we have not been able to prove this. On the other hand, as will be shown in §6, it is not enough to assume that the cardinalities are greater than $c \log N/\log\log N$ for some positive constant c.

THEOREM 3. *Let ε be a positive real number, let V and N be positive integers and let $A_1, \ldots, A_k$ be non-empty subsets of $\{1, \ldots, N\}$. There exists a positive number c_1 effectively computable in terms of V, k, and ε such that if $N > c_1$ and*

$$\min |A_i| > N^{1/k+2\varepsilon}, \tag{4.1}$$

then there is a prime p with

$$V < p < N^{1/k+\varepsilon} \tag{4.2}$$

such that every residue class mod p *contains a member of $A_1 + \cdots + A_k$.*

Proof. From Lemma 3, for any $Q \geq 1$

$$|A_i| \leq \frac{-\log N + \sum_{V<p<Q} \log p}{-\log N + \sum_{V<p<Q} \log p/\nu_i(p)} \tag{4.3}$$

provided the denominator is positive and where $\nu_i(p)$ has the same meaning as in the previous section. Assume that for each prime p in the range (4.2), there is some residue class mod p that is not represented in the sum set $A_1 + \cdots + A_k$. Thus from Lemma 1,

$$\nu_1(p) + \cdots + \nu_k(p) \leq p + k - 2$$

for every prime p in the range (4.2). Thus from Lemma 4 we have

$$\frac{1}{k} \sum_{i=1}^{k} \frac{\log p}{\nu_i(p)} = \frac{\log p}{kp} \sum_{i=1}^{k} \frac{1}{\nu_i(p)/p} \geq \frac{\log p}{p} \cdot \frac{k}{1 + (k-2)/p}.$$

For $p > c_2 = c_2(V, k, \varepsilon)$ we have $p > V$ and

$$\frac{k}{1 + (k-2)/p} > k - \varepsilon/2.$$

Thus

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^{k} \left(-\log N + \sum_{V<p<Q} \frac{\log p}{\nu_i(p)} \right) &= -\log N + \sum_{V<p<Q} \frac{1}{k} \sum_{i=1}^{k} \frac{\log p}{\nu_i(p)} \\ &\geq -\log N + \sum_{c_2<p<Q} (k - \varepsilon/2) \frac{\log p}{p} \\ &\geq -\log N + (k - \varepsilon/2)(\log Q - c_3) \end{aligned} \tag{4.4}$$

where $c_3 = c_3(V, k, \varepsilon)$ and we use Theorem 425 in [5].

Choose $Q = N^{1/k+\varepsilon}$ and note that we may assume that $0 < \varepsilon < 1$. Then for $N > c_4 = c_4(V, k, \varepsilon)$, (4.4) implies that the average of the denominators in (4.3) is at least $(\varepsilon k/2) \log N$. Thus there exists some i such that the denominator in (4.3) is at least $(\varepsilon k/2) \log N$ and for this i we have

$$|A_i| \leq \frac{2}{\varepsilon k \log N} \left(-\log N + \sum_{V<p<Q} \log p \right) \leq \frac{4N^{1/k+\varepsilon}}{\varepsilon k \log N}$$

for $N > c_5(V, k, \varepsilon)$, using the prime number theorem. This estimate contradicts (4.1) for $N > c_1(V, k, \varepsilon)$ and thus proves the theorem.

5. If the sets $A_1, \ldots, A_k$ are "dense", the previous results are not very sharp as the following result shows. Later in this section we shall show that Theorem 4 is essentially the best possible result for "dense" sets.

THEOREM 4. *Let ε be a positive real number, let V and N be positive integers and let $A_1, \ldots, A_k$ be non-empty subsets of $\{1, \ldots, N\}$. There exist positive numbers c_1 and c_2 which are effectively computable in terms of V, k, and ε such that if $N > c_1$ and*

$$T = (|A_1| \ldots |A_k|)^{1/k} > N \exp\{-(1-\varepsilon)\log k \log N / \log\log N\}, \tag{5.1}$$

then there is some prime p with

$$V < p \leq c_2 + \frac{1+\varepsilon}{\log k} \log(2N/T) \log\log(2N/T)$$

such that each residue class mod p *has a member in $A_1 + \cdots + A_k$.*

Proof. Let Q be a positive number. With $\nu_i(p)$ as before, the Chinese Remainder Theorem gives that

$$|A_i| < N \prod_{V<p\leq Q} \frac{\nu_i(p)}{p} + \prod_{V<p\leq Q} \nu_i(p) \tag{5.2}$$

for $i = 1, \ldots, k$. Put

$$Q = \lambda + \frac{1+\varepsilon}{\log k} \log(2N/T) \log\log(2N/T) \tag{5.3}$$

where λ is a positive constant which shall be chosen later. In the following, all numbered constants $c_3, c_4, \ldots$ depend effectively on V, k, and ε, only. Plainly we may assume $0 < \varepsilon < 1$.

From (5.1),

$$\frac{N}{T} < \exp\{(1-\varepsilon)\log k \log N / \log\log N\}$$

and so for $N > c_3$ we have

$$Q < \lambda + (1 - \varepsilon^2/2) \log N.$$

Therefore by the prime number theorem, for $N > c_4 \cdot \exp(3\lambda/\varepsilon^2)$,

$$\prod_{V<p\leq Q} p < N,$$

and so by (5.2)

$$|A_i| < 2N \prod_{V<p\leq Q} \frac{\nu_i(p)}{p}$$

for $i = 1, \ldots, k$. Thus

$$T < 2N \prod_{V<p\le Q} \left(\prod_{i=1}^{k} \frac{\nu_i(p)}{p} \right)^{1/k}. \tag{5.4}$$

Suppose now that for each prime p with $V < p \le Q$, there is a residue class mod p that contains no member of $A_1 + \cdots + A_k$. We shall show that this assumption leads to a contradiction for λ sufficiently large. From Lemma 1, for each prime p with $V < p \le Q$ we have

$$\nu_1(p) + \cdots + \nu_k(p) \le p + k - 2.$$

Thus (5.4) leads, via the arithmetic mean-geometric mean inequality and Mertens' theorem to

$$\begin{aligned} T &< 2N \prod_{V<p\le Q} \frac{1}{k} \left(1 + \frac{k-2}{p}\right) \\ &\le 2Nk^{-\pi(Q)+\pi(V)} \left(\prod_{p\le Q} \left(1 - \frac{1}{p}\right) \right)^{-k+2} \\ &\le c_5 N k^{-\pi(Q)+\pi(V)} (\log Q)^{k-2}. \end{aligned} \tag{5.5}$$

Solving (5.5) for $\pi(Q)$ we get

$$\pi(Q) \le \frac{1}{\log k} \log(2N/T) + c_6 + k \log\log Q.$$

Recall from (5.3) that if λ is large, then so is Q. Thus if $\lambda \ge c_7$, we have

$$\pi(Q) \le \frac{1 + \varepsilon/2}{\log k} \log(2N/T)$$

and so the prime number theorem gives

$$Q \le \frac{1+\varepsilon}{\log k} \log(2N/T) \log\log(2N/T)$$

for $\lambda \ge c_8 > c_7$. Since this inequality is incompatible with (5.3), our assumption that each prime p with $V < p \le Q$ has a residue class containing no member of $A_1 + \cdots + A_k$ must be false. Thus the theorem is proved with $c_2 = c_8$.

THEOREM 5. *Let ε be a positive real number and let $k \ge 2$ be an integer. There exist positive numbers c_9, c_{10}, which are effectively computable in terms of k and ε, such that if N is a positive integer with $N > c_9$ and t is a real number with*

$$N \exp(-(1-\varepsilon) \log k \log N / \log\log N) < t < N, \tag{5.6}$$

then there exists a set $A \subset \{1, \dots, N\}$ *such that*

$$t/2 < |A| < 3tk \tag{5.7}$$

and none of the sums $a_1 + \cdots + a_k$ (*where* $a_1, \dots, a_k \in A$) *has a prime factor* p *with*

$$k < p < \frac{1-\varepsilon}{\log k} \log \frac{2N}{|A|} \log\log \frac{2N}{|A|} - c_{10}. \tag{5.8}$$

Proof. Let $r(n, q)$ denote the least non-negative residue of $n \bmod q$. Define the positive integer Q by

$$N \prod_{k<p<Q} \frac{[p/k]}{p} > t \geq N \prod_{k<p<Q+1} \frac{[p/k]}{p} \tag{5.9}$$

and let

$$A = \{a \leq N : 0 < r(a, p) < \frac{p}{k} \text{ for all } p \text{ with } k < p < Q\}.$$

Thus for each prime p with $k < p < Q$, A has members in at most $[p/k]$ residue classes mod p. By the Chinese Remainder Theorem, we have

$$\left| |A| - N \prod_{k<p<Q} \frac{[p/k]}{p} \right| < \prod_{k<p<Q} [p/k]. \tag{5.10}$$

From (5.6) and (5.9) we have

$$k^{\pi(Q)-\pi(k)} \leq \prod_{k<p<Q} \frac{p}{[p/k]} < \frac{N}{t} < \exp\left\{ \frac{(1-\varepsilon)\log k \log N}{\log\log N} \right\},$$

so that

$$\pi(Q) - \pi(k) < (1-\varepsilon)\log N / \log\log N.$$

Thus if $N > c_{11} = c_{11}(k, \varepsilon)$, the prime number theorem gives

$$\prod_{k<p<Q} p < N^{1-\varepsilon/2} < \frac{1}{2}N.$$

Thus from (5.10),

$$\frac{1}{2}N \prod_{k<p<Q} \frac{[p/k]}{p} < |A| < \frac{3}{2}N \prod_{k<p<Q} \frac{[p/k]}{p},$$

so that (5.7) follows from (5.9).

Clearly, by the definition of A, no member of kA is divisible by any prime p with $k < p < Q$. Thus it suffices to show that Q is greater

than the right side of (5.8) for an appropriate choice of c_{10}. By (5.7) it suffices to show that

$$Q > \frac{1-\varepsilon/2}{\log k}\log\frac{2N}{t}\log\log\frac{2N}{t} - c_{12}. \tag{5.11}$$

But by (5.9) we have

$$\begin{aligned}\log\frac{2N}{t} &\le \log\left(2\prod_{k<p<Q+1}\frac{p}{[p/k]}\right)\\ &< \log\left(2\prod_{k<p<Q+1}k\left(1-\frac{k}{p}\right)^{-1}\right)\\ &\le \log(c_{13}k^{\pi(Q)-\pi(k)}\log^k Q)\\ &\le \pi(Q)\log k + k\log\log Q + c_{14}.\end{aligned} \tag{5.12}$$

Using the prime number theorem, (5.11) can be easily derived from (5.12) by separately treating the cases $2N/t < c_{15}$, $2N/t \ge c_{15}$.

6. Note that Corollary 2 of Theorem 1 implies that if A, B are subsets of $\{1,\dots,N\}$ and every member of $A+B$ is prime, then $|A|\,|B| = O(N)$. If we take $A = \{1\}$, $B = \{p-1 : p < N \text{ prime}\}$, then $|A|\,|B| \sim N/\log N$, so that this result is close to best possible. There is a related conjecture due to Ostmann [**6**]. He conjectured that there do not exist infinite sets A, B of positive integers such that $A+B$ differs from the set of primes by only a finite set. Of course if such sets exist then for all large N,

$$|A(N)|\,|B(N)| > N/\log N \tag{6.1}$$

where $A(N) = A \cap [1,N]$, $B(N) = B \cap [1,N]$.

In this connection we are able to prove the following result.

THEOREM 6. *Let N and l be positive integers with $l < \log N$. There is an effectively computable constant c_1 such that if $N > c_1$, then there exist $A, B \subset \{1,\dots,N\}$ with $|B| = l$,*

$$|A| = k > \frac{N}{l(\log N)^l}$$

and every member of $A+B$ is prime.

Proof. Let c_1 be so large that if $N > c_1$ then

$$\pi(N) - \log N > \frac{N}{\log N}.$$

Note then that

$$\binom{\pi(N)}{l}\Big/\binom{N-1}{l-1} = \frac{\pi(N)(\pi(N)-1)\cdots(\pi(N)-l+1)}{l(N-1)(N-2)\cdots(N-l+1)}$$
$$\geq \frac{(N/\log N)^l}{lN^{l-1}} = \frac{N}{l(\log N)^l}.$$

Thus the theorem will follow from the following lemma. Indeed we take P the set of primes in $[1, N]$, A the set in the lemma shifted down by 1 and B the set in the lemma shifted up by 1. (Before the shifts we only know $A \subset [2, N], B \subset [0, N-2]$, so the shifts put A, B in $[1, N]$.)

LEMMA 6. *Let N be a positive integer and let P be a non-empty subset of $\{1, \ldots, N\}$. Let l be an integer with $1 \leq l \leq |P|$. Then there is a set $A \subset P$ and a set of non-negative integers B such that*

$$A + B \subset P, \qquad |A| \geq \binom{|P|}{l}\Big/\binom{N-1}{l-1}, \qquad |B| = l.$$

This lemma can be found in [4]. For the convenience of the reader, we repeat the proof here.

Proof of the lemma. There are $\binom{|P|}{l}$ l-element subsets of P. To each such subset $\{p_1, \ldots, p_l\}$ with $p_1 < \cdots < p_l$, associate the $l-1$-element subset $\{p_2 - p_1, \ldots, p_l - p_1\}$ of $\{1, \ldots, N-1\}$. Thus there is some $l-1$-element subset $\{h_1, \ldots, h_{l-1}\}$ associated to at least $k \geq \binom{|P|}{l}/\binom{N-1}{l-1}$ l-element subsets of P. Let $a_1, \ldots, a_k$ denote the least elements of these k different l-element subsets associated to $\{h_1, \ldots, h_{l-1}\}$. Thus $a_1, \ldots, a_k$ are distinct members of P. The lemma follows with $A = \{a_1, \ldots, a_k\}$, $B = \{0, h_1, \ldots, h_{l-1}\}$.

REMARKS. We expect that Theorem 6 cannot be substantially improved. In particular if $\min\{k, l\} \geq 2$, then we conjecture that for every $\varepsilon > 0$, $kl < N/(\log N)^{2-\varepsilon}$ for N large in terms of ε. Perhaps even $kl = O(N/(\log N)^2)$. Such results, as a comparison with (6.1) reveals, would imply the truth of the Ostmann conjecture. Further, we expect that if $\min\{k, l\} > \log N$ (or indeed even $2 \log N/\log\log N$) then there are members of $A + B$ with arbitrarily many distinct prime factors as $N \to \infty$.

On the other hand, if $0 < \varepsilon < 1$ and we choose

$$l = [(1-\varepsilon)\log N/\log\log N],$$

then Theorem 6 implies there are sets $A, B \subset \{1, \ldots, N\}$ with $|A| \geq |B| \geq l$ and such that every member of $A + B$ is prime, provided that N is sufficiently large in terms of ε.

Finally note that in the theorem it is possible to also require $A+B \subset [N/2, N]$, but then we must replace the lower bound for $|A|$ with, say, $|A| = k > N/l(2\log N)^l$. To see this, choose P in Lemma 6 as the set of primes in $[N/2, N]$.

7. In the last section we showed there exist sets $A, B, \subset \{1, \dots, N\}$ such that $|A| \geq |B| \gg \log N/\log\log N$ and all members of $A + B$ are prime. In this section we consider the case $A = B$. Now, of course, we cannot prohibit even members of the sum set $2A$, so instead we look for an example where the members of $2A$ are twice a prime. The following result almost achieves this goal. The proof uses very little about the primes—only that they are fairly numerous.

THEOREM 7. *For all large N there is a set $A \subset \{[N/4], \dots, N\}$ with*

$$|A| > \log\log N \tag{7.1}$$

and each sum $a + a'$ with $a, a' \in A$ and $a \neq a'$ is of the form $2p$ where p is prime.

This theorem follows easily from the following lemma, which is a slight sharpening of a lemma in Szemerédi [**11**]. This lemma of Szemerédi has become known as the "Cube Lemma", see page 44 of the Graham, Rothschild, Spencer book "Ramsey Theory". Thanks are due to Paul Erdös and Joel Spencer for the latter reference.

LEMMA 7. *If N is large, $B \subset \{1, \dots, N\}$ with*

$$|B| \geq N^{4/5}, \tag{7.2}$$

and we put

$$l = \left[\frac{11}{10}\log\frac{\log 3N}{\log(3N/|B|)}\right], \tag{7.3}$$

then there exist positive integers $y, x_1, \dots, x_l$ with $x_i \neq x_j$ for $i \neq j$ and

$$\{y\} + \{0, x_1\} + \cdots + \{0, x_l\} \subset B. \tag{7.4}$$

To derive Theorem 7 from Lemma 7, we choose B to be the set of integers of the form $2p$ where p is prime and $N/2 \leq 2p \leq N$. From (7.4), $y \in B$ so y is even. Put

$$A = \left\{\frac{y}{2} + x_1, \dots, \frac{y}{2} + x_l\right\}.$$

Then (7.1) follows easily from (7.3) and a crude estimate for the number of primes between $N/4$ and $N/2$, while $a + a' = 2p$ for $a, a' \in A$, $a \neq a'$ holds by (7.4).

Proof of Lemma 7. It suffices to show the existence of sets $B_0, \ldots, B_l$ and distinct positive integers $x_1, \ldots, x_l$ such that

$$B_0 = B, \tag{7.5}$$

$$B_k + \{0, x_k\} \subset B_{k-1} \quad \text{for } k = 1, \ldots, l, \tag{7.6}$$

$$|B_k| \geq |B|^{2^k}/(3N)^{2^k-1} \quad \text{for } k = 1, \ldots, l. \tag{7.7}$$

In fact, if $B_0, \ldots, B_l, x_1, \ldots, x_l$ satisfy these conditions, then by (7.5) and (7.6), (7.4) holds for any $y \in B_l$, while (7.3) and (7.7) imply B_l is not empty. This then will complete the proof of Lemma 7.

We are going to construct $B_0, \ldots, B_l, x_1, \ldots, x_l$ recursively. Let $B_0 = B$. Assume now that $0 \leq k \leq l - 1$ and that $B_0, \ldots, B_k$ and, in the case $k > 0$, $x_1, \ldots, x_k$ have already been defined. For $1 \leq d \leq N - 1$ let $f(B_k, d)$ denote the number of solutions of

$$b - b' = d, \quad \text{where } b, b' \in B_k.$$

Then in order to define B_{k+1} and x_{k+1}, we need an estimate for

$$M = \max f(B_k, d) \tag{7.8}$$

where the maximum is over all d with $1 \leq d \leq N-1$, $d \notin \{x_1, \ldots, x_k\}$.

Clearly, for all d we have $f(B_k, d) \leq |B_k|$. Also

$$\sum_{d=1}^{N-1} f(B_k, d) = \binom{|B_k|}{2} \tag{7.9}$$

since $b - b' \in \{1, \ldots, N-1\}$ for any pair $b, b' \in B_k$ with $b > b'$. If we majorize $f(B_k, d)$ by $|B_k|$ for $d \in \{x_1, \ldots, x_k\}$ and by M otherwise, (7.9) implies

$$\binom{|B_k|}{2} \leq k|B_k| + (N - 1 - k)M < k|B_k| + NM,$$

so that

$$M > \frac{1}{2N}((|B_k|^2 - |B_k|) - 2k|B_k|) = \frac{|B_k|}{3N}\left(\frac{3}{2}|B_k| - \frac{3}{2} - 3k\right). \tag{7.10}$$

From (7.3), (7.7), and a simple calculation, we have (for N larger than some absolute and computable constant)

$$|B_k| > 3 + 6l > 3 + 6k,$$

so that (7.10) and (7.7) imply

$$M > \frac{|B_k|^2}{3N} \geq \frac{1}{3N}\left(\frac{|B|^{2^k}}{(3N)^{2^k-1}}\right)^2 = \frac{|B|^{2^{k+1}}}{(3N)^{2^{k+1}-1}}. \tag{7.11}$$

Let $x_{k+1} \in \{1, \ldots, N-1\} - \{x_1, \ldots, x_k\}$ denote an integer for which the maximum in (7.8) is attained and let

$$B_{k+1} = \{b \in B_k : b + x_{k+1} \in B_k\}.$$

Thus (7.6) holds for $k+1$ and since $|B_{k+1}| = M$, (7.11) implies (7.7) holds for $k+1$. This completes the proof of the existence of $B_0, \ldots, B_l$, $x_1, \ldots, x_l$ with the desired properties, so that Lemma 7 is proved.

REMARK. It is possible to show there is a set of primes P^* with no three in arithmetic progression and such that the number of members of P^* up to x is greater than $x/e^{c\sqrt{\log x}}$ for a certain positive constant c. The proof of Theorem 7 shows that for every large N there is a set of integers $A \subset [N/4, N]$ with $|A| > \log\log N$ and such that if $a, a' \in A$ with $a \neq a'$, then $a + a' = 2p$ for some prime $p \in P^*$. However, since no three members of P^* are in arithmetic progression, it follows that either $2a$ or $2a'$ is not of the form $2q$ for $q \in P^*$. Thus there is at most one $a \in A$ that is also in P^*. We conclude that the seemingly mild restriction in Theorem 7 that $a \neq a'$ will probably be difficult to remove. At least, an attempt to remove it must use more properties of the set of all primes than we have used.

It is probable that the circle method can be used to prove that for every fixed k there are distinct primes $p_1, \ldots, p_k$ such that each $p_i + p_j$ is twice a prime. We expect that the largest set of primes in $[N/4, N]$ with each double sum twice a prime is of order of magnitude $\log N/\log\log N$, but we do not expect this will be proved anytime soon.

REFERENCES

[1] K. Alladi, *Distribution of $\nu(n)$ in the sieve of Eratosthenes*, Quart. J. Math. (Oxford Ser.), **33** (1982), 129–148.

[2] A. Balog and A. Sárközy, *On sums of sequences of integers*, I, Acta Arith., **44** (1984), 73–86.

[3] ———, *On sums of sequences of integers*, II, Acta Math. Acad. Sci. Hungar., **44** (1984), 169–179.

[4] P. Erdös, C. L. Stewart, and R. Tijdeman, *Some Diophantine equations with many solutions*, Compositio Math., to appear.

[5] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5-th ed., Oxford, 1979.
[6] H. Ostmann, *Additive Zahlentheorie*, Springer Verlag, 1956.
[7] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math., **6** (1962), 64–94.
[8] A. Sárközy and C. L. Stewart, *On divisors of sums of integers*, I, Acta Math. Acad. Sci. Hungar., **48** (1986), 147–154.
[9] ———, *On divisors of sums of integers*, II, J. Reine Angew. Math., **365** (1986), 171–191.
[10] A. Selberg, *Note on a paper by L. G. Sathe*, J. Indian Math. Soc., **18** (1954), 83–87.
[11] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar., **20** (1969), 89–104.

Received November 24, 1986 and in revised form June 2, 1987. The first and second authors' research was supported in part by an NSF grant. The third author's research was supported in part by an NSERC grant.

UNIVERSITY OF GEORGIA
ATHENS, GA 30602

MATHEMATICS INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCE
BUDAPEST, HUNGARY

AND

UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO, CANADA N2L 3G1