COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA 23 (2) (2003) 263–281

MAXIMAL COLLECTIONS OF INTERSECTING ARITHMETIC PROGRESSIONS

KEVIN FORD*

Received October 29, 1999

Let $N_t(k)$ be the maximum number of k-term arithmetic progressions of real numbers, any two of which have t points in common. We determine $N_2(k)$ for prime k and all large k, and give upper and lower bounds for $N_t(k)$ when $t \geq 3$.

Recent work by R. Howard, G. Károlyi and L. Székely [6] on the Erdős–Ko–Rado intersection theorems ([3],[7],[11]) have led to consideration of the following related problem:

For $t \geq 2$, what is the maximum number of distinct arithmetic progressions of k real numbers, any pair of which have t common members?

We will denote the maximum by $N_t(k)$. In this note we determine the exact value of $N_2(k)$ for large k and give some tools for bounding $N_t(k)$ for $t \geq 3$. For brevity, we say a configuration of arithmetic progressions (APs) is *t-intersecting* if every pair of APs have at least t points in common.

By the *difference* of an AP we mean the common difference between consecutive elements of the AP. We use (a,b) to denote the greatest common divisor of a and b, the open interval (a,b), or the ordered pair (a,b), depending on the context. Likewise $[a,b]$ denotes the least common multiple of a

Mathematics Subject Classification (2000): 05D05, 11B75, 11B25; 11A05, 11N05, 11N37

* Research supported in part by NSF grant DMS-0070618.

and b, or the closed interval $[a,b]$. The notation $\lfloor x \rfloor$ will denote the greatest integer $\le x$.

For a configuration of k-term t-intersecting APs, let $D=\{d_1,\dots,d_l\}$ denote the set of distinct differences that the arithmetic progressions (APs) have. For each i let b_i be the number of APs with difference d_i. Clearly the ratios d_i/d_j must be rational, thus we may assume without loss of generality that the progressions consist of integers. We may also assume that the numbers d_i have no common prime factor. Since the APs of difference d_i must intersect pairwise in at least t elements, it follows that

$$b_i \le k - t + 1 \tag{1.1}$$

for every i. Also, $[d_i,d_j]$ is the distance between elements in the intersection of an AP of difference d_i and one of difference d_j. Thus we must have $[d_i,d_j]\le d_i(k-1)/(t-1)$ for every i,j. In other words,

$$\frac{d_j}{(d_i,d_j)} \le \frac{k-1}{t-1} \qquad \forall i,j. \tag{1.2}$$

By a theorem of Balasubramanian–Soundararajan [1] (formerly a conjecture of R. L. Graham [5]), $l\le (k-1)/(t-1)$. It follows from (1.1) that

$$N_t(k) \le (k-t+1)\left\lfloor \frac{k-1}{t-1} \right\rfloor, \tag{1.3}$$

which we refer to as the trivial upper bound.

In sections 2, 3 and 4 we work with the case $t=2$, proving an exact formula for $N_2(k)$ that holds for all large k and "most" smaller k. Section 5 deals with the case $t\ge 3$, and here we prove less precise bounds for $N_t(k)$.

2. Determining $N_2(k)$ when k is prime

We begin with an example of a large configuration of 2-intersecting APs.

Example 1. For $1\le i<j\le k$, let B_{ij} be the AP whose ith element is 0 and whose jth element is $k!$. This configuration of APs is clearly 2-intersecting, and shows that

$$N_2(k) \ge \frac{k(k-1)}{2}. \tag{2.1}$$

This lower bound is roughly a factor 2 smaller than the trivial upper bound (1.1).

We shall show that in fact $N_2(k) = \frac{1}{2}k(k-1)$ for all large k, as well as show that all configurations of $\frac{1}{2}k(k-1)$ k-term 2-intersecting APs are equivalent to the configuration in Example 1, i.e. they are equivalent modulo translations and dilations.

We begin with a self-contained simple proof that $N_2(k)=\frac{1}{2}k(k-1)$ when k is a prime. In particular, we do not require the theorem from [1]. The proof of this bound for general k requires the application of some powerful results, in particular a strong form of the Balasubramanian–Soundararajan theorem and explicit bounds for the number of primes in short intervals.

For each i, let D_i be the set of the b_i APs with difference d_i, and let $P_i = \cup_{A\in D_i} A$. In particular, P_i is itself an AP of $\le 2k-t$ numbers with difference d_i. To simplify the analysis, we shall suppose that

$$b_1 \ge b_2 \ge \cdots \ge b_l. \tag{2.2}$$

Our improvements to (1.1) all stem from an analysis of how the APs with two distinct differences may be configured.

Lemma 2.1. *For every i, j we have*

$$b_i \le 2\left(k - \frac{d_j}{(d_i,d_j)}\right) - 1.$$

Proof. Assume that $d_j/(d_i,d_j) = [d_i,d_j]/d_i > k/2$, else the lemma follows from (1.1). Let y denote the least member of P_i, and let $S=P_i\cap P_j$. Then S is itself an AP with difference $[d_i,d_j]$ and must contain 2 numbers in the interval $[y, y+(k-1)d_i]$. Hence the smallest member x of S is $\le y+(k-1)d_i-[d_i,d_j]$. Each AP of difference d_i contains two points of S and starts at one of the points $y+md_i$, $0\le m\le k-2$. However, an AP of difference d_i starting at one of the points $x+d_i,\dots,x+m_0d_i$ $(m_0=2[d_i,d_j]/d_i-k)$ only contains one point in S, namely $x+[d_i,d_j]$. Thus

$$b_i \le k-1-m_0 = k-1-\frac{2d_j}{(d_i,d_j)}+k. \qquad \blacksquare$$

Remark. Lemma 2.1 is best possible, in the sense that for any two numbers d_i, d_j with $[d_i,d_j]/d_i > k/2$, there is a configuration of k-term 2-intersecting APs with differences d_i, d_j and with $b_i = 2(k-d_j/(d_i,d_j))-1$. For example, take the single AP of difference d_j and starting at $x=(k-1)d_i-[d_i,d_j]$, together with the APs of difference d_i starting at the points hd_i for $0\le h\le k-2$ and $hd_i \notin \{x+d_i,\dots,x+m_0d_i\}$.

Proposition A. *If there are two points common to all APs* $\in D_i$ *and the distance between them is* md_i*, then* $b_i \le k-m$. *This holds for any* $t\ge 2$.

Proof. Let x be the smaller of the two common points. Each AP $\in D_i$ must start at one of the points $x-rd_i$, $0\le r\le k-m-1$. ∎

Lemma 2.2. *For every pair* i,j*, either*

$$b_i \le k - \frac{d_j}{(d_i,d_j)} \qquad \text{or} \qquad b_j \le k - 2\frac{d_i}{(d_i,d_j)}.$$

In particular, if $d_i/(d_i,d_j) > (k-1)/2$*, then*

$$b_i \le k - \frac{d_j}{(d_i,d_j)}.$$

Proof. Let $S=P_i\cap P_j=\{x_1,\dots,x_r\}$. Note that $r\ge 2$ and $x_{h+1}-x_h=[d_i,d_j]$ for every $h<r$. Each AP $\in D_j$ contains at least 2 numbers in S. If some AP $\in D_j$ contains exactly two numbers in S, say x_h and x_{h+1}, then every AP $\in D_i$ must contain these two points. Then Proposition A implies $b_i\le k–d_j/(d_i,d_j)$.

Next assume that $r\ge 3$ and every AP $\in D_j$ contains at least 3 points in S. If there are three points in S that are common to all APs $\in D_j$, then by Proposition A, $b_j\le k-2d_i/(d_i,d_j)$. Otherwise $r\ge 4$ and the intersection of all AP $\in D_j$ contains $w\le 2$ points in S. Thus, for some h there is an AP $\in D_j$ containing x_h,x_{h+1} but not x_{h+2} and another AP $\in D_j$ containing x_{h+3-w},x_{h+2-w} but not x_{h+1-w} (Here if $m<1$ or $m>r$ we define $x_m = x_1+(x_2-x_1)(m-1)$). Therefore every AP $\in D_i$ contains at least two points x_g $(g\le h+1)$ and two points x_g $(g\ge h+2-w)$. In particular, every AP $\in D_i$ contains x_h and x_{h+1}, which gives $b_i\le k-d_j/(d_i,d_j)$ by Proposition A. ∎

By combining Lemmas 2.1 and 2.2, we obtain a bound on b_n+b_m for any pair m,n.

Lemma 2.3. *For every* m,n *we have*

$$b_m + b_n \le 2k - g_{m,n} + \max(0, k-g_{m,n}-1),$$

where

$$g_{m,n} = \max_{\substack{1\le i\le n\\ 1\le j\le m}} \frac{d_i+d_j}{(d_i,d_j)}.$$

Proof. For some $i\le n$, $j\le m$ we have $g_{m,n}=e_i+e_j$, where

$$e_i = \frac{d_i}{(d_i,d_j)}, \quad e_j = \frac{d_j}{(d_i,d_j)}.$$

Let $g=g_{m,n}$ and $c=3k-b_n-2g-1$. First, if $e_i \geq k-\frac{c+1}{2}$, then Lemma 2.1 gives $b_j \leq c$. Otherwise, suppose that $e_j > k-b_n$. By (2.2) and Lemma 2.1, $e_j \leq k-\frac{1}{2}(b_i+1) \leq k-\frac{1}{2}(b_n+1)$. Also by (2.2), $b_i \geq b_n > k-e_j$, so Lemma 2.2 gives

$$\begin{aligned} b_j &\leq k-2e_i = k-2(g-e_j) \\ &\leq 3k-2g-b_n-1 = c. \end{aligned}$$

Finally, if $e_i < k-\frac{c+1}{2}$ and $e_j \leq k-b_n$, we have

$$b_i \geq b_n = 3k-2g-c-1 > k-2g+2e_i = k-2e_j,$$

whence by Lemma 2.2

$$b_j \leq k-e_i = k-g+e_j \leq 2k-g-b_n.$$

Therefore,

$$b_m \leq b_j \leq \max(c, 2k-g-b_n),$$

and the lemma follows. ∎

Lemma 2.4. *If $p \geq k$ is a prime and $m+n \geq p$, then*

$$b_m + b_n \leq 2k-p.$$

Proof. By Lemma 2.3, it suffices to show that $g_{m,n} \geq p$. Without loss of generality, assume $\gcd(d_1, d_2, \ldots, d_{\max(m,n)}) = 1$. For brevity, let $c_{i,j} = d_i/(d_i, d_j)$ for each i, j. If $p | d_i$ for some $i \leq \max(m,n)$, then $p \nmid d_j$ for some $j \leq \max(m,n)$ and then $c_{i,j} \geq p \geq k$, contradicting (1.2). Thus, two of the numbers $d_1, \ldots, d_n, -d_1, \ldots, -d_m$ are congruent modulo p (and not congruent to 0). If d_i and d_j are congruent, then $|c_{i,j} - c_{j,i}| \geq p \geq k$, which implies that $c_{i,j} \geq k$ or $c_{j,i} \geq k$, which is again impossible. Therefore, d_i and $-d_j$ are congruent for some $i \leq n$ and $j \leq m$. It follows that $c_{i,j} + c_{j,i} = p$. ∎

Theorem 1. *If k is prime, then $N_2(k) = \frac{1}{2}k(k-1)$.*

Proof. The cases $k=2$ and $k=3$ are trivial. Suppose $k=p \geq 5$. First suppose that $l \leq k/2$. By Lemma 2.2, at most one of the numbers b_i can equal $k-1$, so that

$$N := \sum_{i=1}^{l} b_i \leq (k/2)(k-2) + 1 < \frac{1}{2}k(k-1).$$

Next, suppose $l > k/2$. By Lemma 2.4, we have

$$b_m + b_{k-m} \le 2k - p = k \qquad (k - l \le m \le l).$$

We next show that

$$l \le k-1. \tag{2.3}$$

First, if $p|d_i$ for some i, then $p \nmid d_j$ for some j and then $d_i/(d_i,d_j) \ge p = k$, which contradicts (1.2). Likewise, if $d_i \equiv d_j \pmod p$, then

$$\left| \frac{d_i}{(d_i,d_j)} - \frac{d_j}{(d_i,d_j)} \right| \ge p,$$

which implies one of the quotients is $\ge p$, again a contradiction. Thus the numbers d_i are distinct modulo p and not divisible by p, which proves (2.3). Applying (2.3) and the trivial bound (1.1) for b_m $(m < p - l)$, we obtain

$$\begin{aligned} N &\le (k - p/2)(2l - p + 1) + (p - l - 1)(k - 1) \\ &= \frac{1}{2}k(k-1) - (k - l) + 1 \le \frac{1}{2}k(k-1). \end{aligned}$$

Combined with the lower bound (2.1), the theorem is established. ∎

3. The theorem for general k

We first state several results related to Graham's Conjecture that we shall require. For a set $A = \{a_1, \dots, a_n\}$ of positive integers, we define

$$A^* = \left\{ \frac{L}{a_1}, \frac{L}{a_2}, \dots, \frac{L}{a_n} \right\}, \qquad L = \operatorname{lcm}[a_1, a_2, \dots, a_n],$$

which refer to as the dual of A.

Lemma 3.1 (Balasubramanian–Soundararajan). *If $A = \{a_1, \dots, a_n\}$ is a set of positive integers, then for some i, j we have*

$$\frac{a_i}{(a_i, a_j)} \ge n.$$

Furthermore, if $n \ge 5$, $(a_1, \dots, a_n) = 1$ and $a_i/(a_i,a_j) \le n$ for all i,j, then either $A = \{1, \dots, n\}$ or $A^* = \{1, \dots, n\}$.

Remarks. This theorem was a conjecture of R. L. Graham [5] and had been previously proved for all large n independently by Szegedy [10] and

Zaharescu [12]. It follows immediately from (1.2) that $l \le k-1$, where l is the number of distinct differences of the APs in the configuration.

From now on, we suppose that $\gcd(a : a \in A) = 1$. For brevity, we define $G(A)$ to be the maximum of $a/(a,a')$ over all pairs of elements a, a' belonging to a set A. We let $f(N)$ be the largest number f so that the following holds:

> For every set of positive integers A with $|A| = M$, $N - f \le M \le N$ and $G(A) \le N$, either A or A^* is contained in $\{1, 2, \dots, N\}$.

Borrowing ideas from [1], we will prove the following in section 4.

Lemma 3.2. *If $N \ge e^{10000}$, then*

$$f(N) \ge \frac{0.156N}{\log^3 N}.$$

Remarks. In [1], the authors claim that their methods yield

$$f(N) \ge \frac{cN}{\log N \log\log N}$$

for some positive constant c, but this appears to be too optimistic. In a separate paper [4] we will show that the methods of [1] can be used to prove that

$$f(N) \ge \frac{cN \log\log N}{\log^2 N}.$$

We also have need of upper bounds on the gaps between consecutive primes. We use explicit bounds for the error term in the Prime Number Theorem given by Rosser and Schoenfeld (Theorem 11 of [9], see also [8]).

Lemma 3.3. *Let $\theta(x) = \sum_{p \le x} \log p$, the sum being over primes p. For $x \ge$ 101,*

$$|\theta(x) - x| \le \varepsilon(x)x, \quad \varepsilon(x) = 0.21962(\log x)^{1/4} e^{-0.321979\sqrt{\log x}}.$$

Lemma 3.4. *For $k \ge e^{1000}$ there is a prime in the interval $[k, k+a]$, where*

$$a = 0.44k(\log k)^{1/4} e^{-0.321979\sqrt{\log k}}.$$

Proof. For $x \ge e^{1000}$, $\varepsilon(x) < 0.001$, so by Lemma 3.3 we have

$$\begin{aligned}\theta(x + 2.003x\varepsilon(x)) - \theta(x) &> (x + 2.003x\varepsilon(x))(1 - \varepsilon(x)) - x(1 + \varepsilon(x)) \\ &= x\left(0.003\varepsilon(x) - 2.003\varepsilon^2(x)\right) > 0.\end{aligned}$$ ∎

The last tool we need is a method of bounding b_i non-trivially when Lemma 2.4 does not apply.

Lemma 3.5. *When* $k \ge 10$ *and* $1 \le h \le l$ *we have*

$$\sum_{i=1}^{h} b_i \le \min\left(kh-2h+1, kh-\frac{h^2}{2\log k}\right).$$

Proof. The bound $kh-2h+1$ follows for all h since at most one of the b_i can equal $k-1$ (from Lemma 2.2). When $h<2.5\log k$, $2h>h^2/(2\log k)$, so the second bound follows as well. Next, suppose $h \ge 2.5\log k$. The numbers d_i all lie in some interval of the form $[B,(k-1)B]$, since otherwise there would be two of them with $d_i/(d_i,d_j) \ge d_i/d_j > k-1$. If I is a collection of m indices with $m \ge \frac{3}{2}\log k$, then by Dirichlet's box principle, there are two indices $i,j \in I$ such that

$$1 < \frac{d_i}{d_j} < k^{1/m} < 1 + \frac{\log k}{m-\frac{1}{2}\log k} \le 2.$$

Therefore,

$$(d_i,d_j) \le d_i - d_j < d_j \frac{\log k}{m-\frac{1}{2}\log k} < d_i \frac{\log k}{m-\frac{1}{2}\log k}.$$

By Lemma 2.2 it follows that

$$\min(b_i,b_j) < k - \frac{m}{\log k} + \frac{1}{2}.$$

Applying this argument successively with $m=h, m=h-1, \ldots$, we find a sequence of indices $\{i_m : \frac{3}{2}\log k \le m \le h\}$ such that for every m, $i_m \le h$ and

$$b_{i_m} \le k - \left\lfloor \frac{m}{\log k} + \frac{1}{2} \right\rfloor.$$

Now write $s = \lfloor h/\log k + 0.5 \rfloor$ and note that $\log k$ is irrational. Thus, since $b_i = k-1$ for at most one i,

$$\begin{aligned}
\sum_{i\le h} b_i &\le kh - 2\lfloor 2.5\log k\rfloor + 1 - \sum_{2.5\log k < m \le h} \left\lfloor \frac{m}{\log k} + \frac{1}{2} \right\rfloor \\
&= kh - 2\lfloor 2.5\log k\rfloor + 1 - \sum_{3\le r\le s-1} r\Big(\lfloor (r+\tfrac{1}{2})\log k\rfloor - \lfloor (r-\tfrac{1}{2})\log k\rfloor\Big) \\
&\qquad -s(h - \lfloor (s-\tfrac{1}{2})\log k\rfloor) \\
&= kh - sh + 1 + \sum_{2\le r\le s-1} \lfloor (r+\tfrac{1}{2})\log k\rfloor
\end{aligned}$$

$$\leq kh - sh + 1 + (\log k) \sum_{r=2}^{s-1} (r + 1/2)$$
$$\leq kh - \frac{15}{8} \log k + 1 - \frac{h^2}{2 \log k}. \qquad \blacksquare$$

Theorem 2. *For each k let $a(k)=p-k$ where p is the smallest prime $\geq k$. If $k \geq 26$ and*

$$\text{(3.1)} \qquad \min(f(k-1)+2, \pi(k-1) - \pi(2k/3)+1) \geq 2a(k) \log k,$$

the following holds: $N_2(k) = \frac{k(k-1)}{2}$ and any configuration of $\frac{k(k-1)}{2}$ 2-intersecting k-term APs is equivalent to the configuration in Example 1. Here $\pi(x)$ denotes the number of primes $\leq x$.

Proof. Let $N = \sum b_i$, let p be the smallest prime $\geq k$ and $a = p - k$. By Lemma 3.1, $l \leq k-1$. By Theorems 1 and 2 of [8], we have

$$\text{(3.2)} \qquad \frac{x}{\log x - 1/2} < \pi(x) < \frac{x}{\log x} + \frac{3x}{2 \log^2 x} \qquad (x > 67).$$

Therefore, for $k \geq 400$ we have

$$\pi(k-1) - \pi(2k/3) + 1 \leq \frac{k}{\log k} \left(1 + \frac{3}{2 \log k} - \frac{2/3}{\log k - 0.9} + \frac{\log k}{k}\right) < \frac{k}{2 \log k}.$$

By a short computation, $\pi(k–1)–\pi(2k/3)+1 < k/(2 \log k)$ for $26 \leq k \leq 399$, and therefore the hypothesis implies $a < \frac{k}{4 \log^2 k}$. Hence, if $l < p/2$ then Lemma 3.5 gives

$$N \leq kl - \frac{l^2}{2 \log k}$$
$$\leq \frac{1}{2} k(k-1) + \frac{ak}{2} - \frac{(k+a-1)^2}{8 \log k} < \frac{k(k-1)}{2}.$$

Next assume $l > p/2$. For $p-l \leq m \leq l$, Lemma 2.4 gives $b_m + b_{p-m} \leq 2k-p$, hence

$$2 \sum_{m=p-l}^{l} b_m \leq (2k-p)(2l-p+1).$$

Applying Lemma 3.5 to bound $b_1+\cdots+b_{p-l-1}$, we deduce that

$$\begin{aligned} N &\le (2k-p)(l-(p-1)/2)+k(p-l-1)-\frac{(p-l-1)^2}{2\log k} \\ &= \frac{k(k-1)}{2}+\frac{a^2-a}{2}+a(k-l)-\frac{(k-l+a-1)^2}{2\log k}. \end{aligned}$$

If $a=0$ and $l\le k-2$ we obtain $N<\frac{1}{2}k(k-1)$. Likewise, if $a\ge 1$ and $l<k-2a\log k$, then

$$\begin{aligned} N &\le \frac{k(k-1)}{2}+\frac{a^2-a}{2}+(k-l)\left(a-\frac{k-l+2a-2}{2\log k}\right) \\ &< \frac{k(k-1)}{2}-\frac{3}{2}(a^2-a)\le \frac{k(k-1)}{2}. \end{aligned}$$

Lastly, we have to consider the two cases $a=0, l=k-1$ and $l>k-2a\log k$. For the first case, Lemma 3.1 implies that either the set $D=\{d_1,\dots,d_l\}$ or D^* is equal to $\{1,2,\dots,k-1\}$. Furthermore, by (3.2) and a short computation, when $k\ge 26$ there is at least one prime in $(2k/3,k-1]$ and at least two primes in $(k/2,k-1]$. In the second case, by hypothesis and the fact that $\pi(2k/3)\ge\pi(k/2)+1$ for $k\ge 17$ (by (3.2) and a short computation), we have

(3.3)
$$|D|=l\ge k-1-\min(f(k-1),\pi(k-1)-\pi(2k/3)-1,\pi(k-1)-\pi(k/2)-2).$$

Since $G(D)\le k-1$, by the definition of $f(k-1)$, either D or D^* is a subset of $\{1,2,\dots,k\text{–}1\}$. Furthermore, (3.3) also implies that D (or D^*, as appropriate) contains at least one prime in $(2k/3,k-1]$ and two primes in $(k/2,k-1]$.

Suppose first that D is a subset of $\{1,2,\dots,k-1\}$. Suppose that $d_j=p$ where p is a prime $>2k/3$. For $i\ne j$, $(d_i,d_j)=1$ so by Lemma 2.2 $b_j\le k-d_i$, whence $d_i\le k-b_j$ for all $i\ne j$. If $d_i>\frac{1}{2}(k-b_j)$, Lemma 2.2 implies $b_i\le k-d_j=k-p<k/3$. If $d_i\le\frac{1}{2}(k-b_j)$ then Lemma 2.1 gives $b_i\le 2(k-p)-1<2k/3-1$. Therefore

$$\begin{aligned} N &\le \frac{k-b_j}{2}\left(\frac{2k-4}{3}\right)+\frac{k-b_j}{2}\left(\frac{k-1}{3}\right)+b_j \\ &= \frac{k^2}{2}-\frac{5k}{6}+b_j\left(\frac{11}{6}-\frac{k}{2}\right)<\frac{k(k-1)}{2}. \end{aligned}$$

In the case that D^* is a subset of $\{1,2,\dots,k-1\}$, D^* contains two primes p,q larger than $k/2$. Let L be the least common multiple of the numbers d_i, and suppose that $d_j = L/p$. For every $i \neq j$, we have $p|d_i$, thus

$$\frac{d_i}{(d_i,d_j)} = p, \qquad \frac{d_j}{(d_i,d_j)} = L/d_i.$$

Lemma 2.2 implies that $b_i \le k - L/d_i$. Likewise, if $d_h = L/q$ then $b_j \le k - L/d_j$. Thus

$$N \le \sum_{i=1}^{l}(k - L/d_i) \le lk - \sum_{i=1}^{l} i = l(k-(l+1)/2) \le k(k-1)/2.$$

Furthermore, $N = \frac{1}{2}k(k-1)$ precisely when $l = k-1$, $D^* = \{1,2,\dots,k-1\}$ and $b_i = k-i$ for each i. In particular, there are $k-1$ APs of difference L, with just two numbers, say 0 and L, which are common to all of them. One of the APs, A_1, contains 0 and L but not $2L$, while another, A_2, contains 0 and L but not $-L$. Every AP with difference $d_i \neq L$ must have two points in common with each of A_1, A_2. Since $d_i|L$, it follows that the AP must contain both 0 and L. Since every AP contains both 0 and L, it follows immediately that the configuration of APs is equivalent to Example 1. ∎

Theorem 3. *The conclusion of Theorem 2 holds for* $k \ge 10^{8000}$.

Proof. By (3.2), $\pi(k-1) - \pi(2k/3) + 1 > k/(4\log k)$. For $k \ge 10^{8000} > e^{18420}$, Lemmas 3.2 and 3.4 imply

$$a(k) \le 0.44k(\log k)^{1/4} e^{-0.321979\sqrt{\log k}} < \frac{0.063k}{\log^4 k} < \frac{f(k-1)+2}{2\log k}.$$

Therefore (3.1) holds for such k and the conclusion of Theorem 2 follows. ∎

In [4] much better lower bounds are proven for $f(k)$, but these are still insufficient to prove (3.1) for all k (or even $k \ge 10^{100}$). The barrier is our lack of adequate bounds for primes in short intervals (e.g. $[x, x + x/(6\log^3 x)]$) in the range $e^{100} \le x \le e^{9000}$.

4. A lower bound for $f(N)$

In this section we prove Lemma 3.2. By much longer arguments we can improve the bounds roughly by a factor of $\log N$ ([4]). Our first lemma is a variant of a lemma proved independently by Boyle [2] and Szegedy [10].

Lemma 4.1. *Suppose* $|A|=M$, $M\le N<2(M-2)$ *and* $G(A)\le N$. *If some but not all elements of* A *are divisible by a prime* $q>N/2$, *then either* $q\in A, A\subset[1,N]$ *or* $q\in A^*, A^*\subset[1,N]$.

Proof. Without loss of generality, suppose q divides $a_1,\dots,a_s$ and does not divide $a_{s+1},\dots,a_M$. We may also assume that $1\le s\le M/2$, else replace A with A^*. First, we have $q\le N$, else $a_1/(a_1,a_M)\ge q>N$, contradicting our hypothesis. Next, for $1\le i\le s$, $s+1\le j\le M$ we have $(a_i/q)|a_j$, for otherwise

$$\frac{a_i}{(a_i,a_j)}=q\frac{a_i/q}{(a_i/q,a_j)}\ge 2q>N,$$

which is impossible. Let $b_i=a_i/q$ for $i\le s$ and set $B=\mathrm{lcm}[b_1,\dots,b_s]$. Also let $C=\gcd(a_{s+1},\dots,a_M)$. Then it follows that $B|C$. Next define indices k and t by

$$\frac{B}{b_k}=\max_i\frac{B}{b_i},\qquad \frac{a_t}{C}=\max_j\frac{a_j}{C}.$$

As the numbers B/b_i are distinct positive integers, $B/b_k\ge s$. Likewise $a_t/C\ge M-s$. It follows that

$$N\ge\frac{a_t}{(a_t,a_k)}=\frac{a_t}{(a_t,b_k)}=\frac{a_t}{b_k}=\frac{a_t}{C}\frac{B}{b_k}\frac{C}{B}\ge s(M-s)\frac{C}{B}.$$

Since $C/B\ge 1$, this forces $s=1$. Thus $B|a_i$ for all $1\le i\le M$, which forces $B=1$ and hence $C=1$. It follows that $a_1=q$ and for $j\ge 2$, $a_j=\frac{a_j}{(a_j,a_1)}\le N$, as required. ∎

The example $A=\{a\le N:(6,N)>1\}\cup\{6\lfloor N/3\rfloor\}$ shows that $f(N)\le N/3-1$ for $N\ge 5$. From now on, we assume that $|A|=M$ with $M\ge 2N/3+1$. In particular, when $N\ge 7$, $2(M-2)>N$.

We need to introduce some of the notation from [1]. Suppose $A=\{a_1,\dots,a_M\}$. If p is a prime in $(1.5N,2N)$ and $p-N\le m\le N$, define

$$r_p(m)=\left|\left\{\text{pairs }(a_i,a_j): a_i<a_j, \frac{a_i}{(a_i,a_j)}=m, \frac{a_j}{(a_i,a_j)}=p-m\right\}\right|.$$

Lemma 4.2. *If* $N\ge 7$, $G(A)\le N$ *and* $|A|=M$, *then for each prime* $p\in(N,2N)$ *we have*

$$\sum_{\substack{\frac{p+1}{2}\le m\le N\\ r_p(m)\ge 2}}(r_p(m)-1)\ge\sum_{\substack{\frac{p+1}{2}\le m\le N\\ r_p(m)=0}}1-(N-M).$$

Proof. We first claim that the numbers in A are all coprime to p and incongruent modulo p. If $p|a_i$, then $p\nmid a_j$ for some j, whence $a_i/(a_i,a_j)\ge$

$p > N$, a contradiction. Similarly, if $a_i \equiv a_j (\bmod p)$, $a_i > a_j$, and neither is divisible by p, then

$$\frac{a_i}{(a_i, a_j)} - \frac{a_j}{(a_i, a_j)} \geq p > N,$$

which also contradicts $G(A) \leq N$. This proves the claim. Since there are $\frac{p-1}{2}$ quadratic residues modulo p, by the box principle there are $\geq M - \frac{p-1}{2}$ distinct pairs with $a_i < a_j$ and $a_i^2 \equiv a_j^2 \pmod p$. Since $a_i \not\equiv a_j \pmod p$, $p | \frac{a_i + a_j}{(a_i, a_j)}$. But $\frac{a_i+a_j}{(a_i,a_j)} \leq 2N < 2p$, so $\frac{a_i+a_j}{(a_i,a_j)} = p$. Thus the pair is counted once in $r_p(m)$ with $m = a_j/(a_i, a_j)$. This means

$$\sum_{\frac{p+1}{2} \leq m \leq N} r_p(m) \geq M - \frac{p-1}{2} = \left(N - \frac{p-1}{2}\right) - (N - M)$$
$$= \sum_{\frac{p+1}{2} \leq m \leq N} 1 - (N - M),$$

and the lemma follows. ∎

If $G(A) \leq N$, $3N/2 < p < 2N$, $|A| = M$ and neither A nor A^* lies in $\{1, 2, \ldots, N\}$, Lemma 4.1 implies that $r_p(m) = 0$ whenever m is prime or $p - m$ is prime. By Lemma 4.2, for each prime $p \in (1.5N, 2N)$ we have

$$\sum_{\substack{\frac{p+1}{2} \leq m \leq N \\ r_p(m) \geq 2}} (r_p(m) - 1) \geq \pi(N) - \pi(p - N) - (N - M). \tag{4.1}$$

From Lemmas 2.3–2.5 of [1], it follows that

$$r_p(m) \leq (K(m) + 1)(K(p - m) + 1), \tag{4.2}$$

where

$$K(m) = K_N(m) = \left|\{ab | m : 1 < b/a \leq N/m\}\right|. \tag{4.3}$$

Actually in [1] a stronger bound is proved, but (4.2) suffices for our purposes. Putting these tools together gives

Lemma 4.3. *Let $\mathcal{P}$ be a subset of the primes in $(1.5N, 2N)$. Then*

$$f(N) \geq$$
$$-1 + \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \left[\pi(N) - \pi(p - N) - \sum_{\frac{p+1}{2} \leq m \leq N} ((K(m) + 1)(K(p - m) + 1) - 1) \right].$$

Proof. Suppose $|A|=M$, $G(A)\le N$ and neither A nor A^* lies in $\{1,\dots,N\}$. By (4.1) and (4.2) we obtain

$$N-M\ge \pi(N)-\pi(p-N)-\sum_{\frac{p+1}{2}\le m\le N}((K(m)+1)(K(p-m)+1)-1)$$

for each $p\in\mathcal{P}$. Averaging over $p\in\mathcal{P}$ gives the result. ∎

Lemma 4.4. *If* $6N^{2/3}\le\lambda\le N/5$, *then*

$$\sum_{N-\lambda\le m\le N} K(m)\le \frac{\lambda^2\log N}{3(N-\lambda)}.$$

Proof. The left side counts the number of triples (a,b,c) with

$$N-\lambda\le abc\le N, \qquad 1<\frac{b}{a}\le\frac{N}{abc}.$$

This implies

$$\frac{N-\lambda}{ab}\le c\le\frac{N}{b^2},\quad a\ge\frac{N}{\lambda}-1\ge 4,\quad b\le(1+\beta)a,\quad \beta=\frac{\lambda}{N-\lambda}.$$

We divide the triples into two classes. Let T_1 be the number of triples with $a\le N^{1/3}-1$ and T_2 be the number of remaining triples. For each pair (a,b), the number of c is at most

$$\frac{N}{b^2}-\frac{N-\lambda}{ab}+1=\frac{N}{b}\left(\frac{1}{b}-\frac{1}{a(1+\beta)}\right)+1.$$

The function on the right is a decreasing function of b and is positive for $b<(1+\beta)a$, so

$$\begin{aligned}
T_1 &\le \sum_{a,b}1+\sum_a\int_0^{a\beta}\frac{N}{(a+t)^2}-\frac{N-\lambda}{a(a+t)}\,dt\\
&=\sum_{a,b}1+\sum_a\frac{1}{a}\left(\frac{\beta N}{1+\beta}-(N-\lambda)\log(1+\beta)\right)\\
&\le\sum_a \beta a+\frac{1}{a}\left(\lambda-(N-\lambda)\left(\beta-\frac{1}{2}\beta^2\right)\right)\\
&\le\frac{\beta N^{2/3}}{2}+\frac{\lambda^2}{2(N-\lambda)}\sum_a\frac{1}{a}\\
&\le\frac{\beta N^{2/3}}{2}+\frac{\lambda^2}{2(N-\lambda)}\left(\frac{1}{3}\log N-\log 3\right).
\end{aligned}$$

To bound T_2, note that $c \le N/b^2 \le N^{1/3}$. For each c, both a and b lie in $[\frac{N-\lambda}{\sqrt{Nc}}, \frac{N}{\sqrt{Nc}}]$, and the number of pairs (a,b) is therefore at most

$$\frac{1}{2}\left(\frac{\lambda^2}{Nc} + \frac{\lambda}{\sqrt{Nc}}\right).$$

Summing on c and using

$$\sum_{n\le x} \frac{1}{n} \le \log x + 0.58, \qquad \sum_{n\le x} \frac{1}{\sqrt{n}} \le 1 + \int_1^x \frac{dt}{\sqrt{t}} < 2\sqrt{x}$$

gives

$$T_2 \le \frac{\lambda^2}{2N}\left(\frac{1}{3}\log N + 0.58\right) + \frac{\lambda}{2\sqrt{N}}(2N^{1/6}).$$

Applying the hypotheses on λ, we obtain

$$T_1 + T_2 \le \frac{\lambda^2}{2(N-\lambda)}\left(\frac{2}{3}\log N - 0.51 + \frac{3N^{2/3}}{\lambda}\right) \le \frac{\lambda^2 \log N}{3(N-\lambda)}. \qquad \blacksquare$$

To prove Lemma 3.2, take $\mathcal{P} = \mathcal{P}_B$, the set of primes in $[2N-2B, 2N-B]$ for some parameter $B < N/4$. By Lemma 3.3,

$$\begin{aligned}\sum_{p\in\mathcal{P}_B} \pi(N) - \pi(p-N) &\ge |\mathcal{P}_B|\frac{\theta(N)-\theta(N-B)}{\log N}\\ &\ge |\mathcal{P}_B|\frac{B - 2N\varepsilon(N/2)}{\log N}.\end{aligned}$$

Also, since the right side of (4.2) is invariant if m is replaced by $p-m$, we have

$$\begin{aligned}&\sum_{\substack{p\in\mathcal{P}_B\\ \frac{p+1}{2} \le m \le N}} ((K(m)+1)(K(p-m)+1)-1) =\\ &\frac{1}{2}\sum_{\substack{p\in\mathcal{P}_B\\ p-N\le m\le N}} (K(m)K(p-m) + 2K(m)) \le\\ &\frac{1}{2}\left(\sum_{N-2B\le m\le N} K(m)\right)^2 + |\mathcal{P}_B| \sum_{N-2B\le m\le N} K(m).\end{aligned}$$

For simplicity, we have essentially ignored the fact that p is prime in the sum over $K(m)K(p-m)$. Therefore, if $3N^{2/3}<B<N/10$, Lemmas 4.3 and 4.4 give

$$f(N) \ge -1 + \frac{B-2N\varepsilon(N/2)}{\log N} - \frac{4B^2\log N}{3(N-2B)} - |\mathcal{P}_B|^{-1}\frac{8B^4\log^2 N}{9(N-2B)^2}.$$

Also from Lemma 3.3 we obtain

$$|\mathcal{P}_B| \ge \frac{B-4N\varepsilon(N)}{\log 2N}.$$

Now suppose that $N\ge e^{10000}$. Then $\varepsilon(N)\le\varepsilon(N/2)\le 0.023(\log N)^{-3}$. We take $B=0.29N(\log N)^{-2}$ and readily obtain

$$f(N) \ge -1 + \frac{0.289984N}{\log^3 N} - \frac{0.112133N}{\log^3 N} - \frac{0.021681N}{\log^3 N} \ge \frac{0.156N}{\log^3 N}. \qquad \blacksquare$$

5. The bounds when $t\ge 3$

We begin with an example of a large configuration of t-intersecting APs, which is a generalization of Example 1.

Example 2. For $1\le i<j\le k$ and $(t-1)|(j-i)$, let B_{ij} denote the AP whose ith element is 0 and whose jth element is $(t-1)k!$. Writing $\theta=\frac{k}{t-1}-\lfloor\frac{k}{t-1}\rfloor$, we have

$$N_t(k) \ge \sum_{i=1}^{k-t+1}\left\lfloor\frac{k-i}{t-1}\right\rfloor = \frac{k^2-(t-1)k}{2t-2}+\frac{t-1}{2}(\theta-\theta^2). \tag{5.1}$$

Conjecture. *For every k,*

$$N_t(k) = \frac{k^2-(t-1)k}{2t-2}+\frac{t-1}{2}(\theta-\theta^2),$$

and every configuration of $N_t(k)$ t-intersecting APs is equivalent to the configuration in Example 2.

In short, the conjectured upper bound for $N_t(k)$ occurs with a configuration of APs with t points common to all of them. Unfortunately we cannot prove this for any $t\ge 3$, even for sufficiently large k. However, we can improve on the trivial bound (1.3) by means of analogs of Lemmas 2.1 and 2.2.

We adopt the notation of sections 1 and 2 and suppose that

$$b_1 \ge \cdots \ge b_l.$$

The next two lemmas are direct analogs of Lemmas 2.1 and 2.2. It is primarily the weakness of Lemma 5.1, which is non-trivial only when $d_j/(d_i,d_j)\ge k/t$, that prevents us from proving the Conjecture.

Lemma 5.1. *For $t \geq 2$ and every i,j we have*

$$b_i \leq t\left(k-(t-1)\frac{d_j}{(d_i,d_j)}\right)-1.$$

Proof. Assume that $d_j/(d_i,d_j) > k/t$, else the lemma follows from (1.1). Denote by y the smallest member of P_i and let $S = P_i \cap P_j$. S is itself an AP with difference $[d_i,d_j]$ and must contain t numbers in $[y, y+(k-1)d_i]$, so the smallest member x of S is $\leq y+(k-1)d_i-(t-1)[d_i,d_j]$. By (1.1), each AP of difference d_i contains t points of S and starts at one of the points $y+md_i$, $0 \leq m \leq k-t$. However, an AP of difference d_i starting at one of the points $x+h[d_i,d_j]+md_i$, $(0 \leq h \leq t-2, 1 \leq m \leq td_j/(d_i,d_j)-k$, $x+h[d_i,d_j]+md_i \leq y+(k-t)d_i)$ only contain $t-1$ points in S. By (1.2), $md_i \leq [d_i,d_j]-d_i$, so these starting points are distinct. The number of these starting points is

$$\geq (t-1)\left(t\frac{d_j}{(d_i,d_j)}-k\right)-(t-2),$$

thus

$$b_i \leq k-t+1-\left[(t-1)\left(t\frac{d_j}{(d_i,d_j)}-k\right)-(t-2)\right]. \qquad \blacksquare$$

Lemma 5.2. *Suppose $t \geq 2$. For every pair i,j, either*

$$b_i \leq k-(t-1)\frac{d_j}{(d_i,d_j)} \qquad \text{or} \qquad b_j \leq k-t\frac{d_i}{(d_i,d_j)}.$$

In particular, if $d_i/(d_i,d_j) > (k-1)/t$, then

$$b_i \leq k-(t-1)\frac{d_j}{(d_i,d_j)}.$$

Proof. Let $S = P_i \cap P_j = \{x_1,\dots,x_r\}$. Note that $r \geq t$ and $x_{h+1}-x_h = [d_i,d_j]$ for every $h < r$. Each AP $\in D_j$ contains at least t numbers in S. If some AP $\in D_j$ contains exactly t numbers in S, say $x_h,\dots,x_{h+t-1}$, then every AP $\in D_i$ must contain these t points. By Proposition A, $b_i \leq k-(t-1)d_j/(d_i,d_j)$. Next assume that $r \geq t+1$ and every AP $\in D_j$ contains at least $t+1$ points in S. If there are $t+1$ points in S that are common to all AP $\in D_j$, then by Proposition A, $b_j \leq k-td_i/(d_i,d_j)$. Otherwise $r \geq t+2$ and the intersection of all AP $\in D_j$ contains $w \leq t$ points in S. Thus, for some h there is an AP $\in D_j$ containing $x_h,\dots,x_{h+t-1}$ but not x_{h+t} and another AP $\in D_j$ containing

$x_{h+2t-1-w}, \dots, x_{h+t-w}$ but not $x_{h+t-1-w}$. Therefore every AP $\in D_i$ contains at least t points x_g $(g \le h+t-1)$ and t points x_g $(g \ge h+t-w)$. In particular, every AP $\in D_i$ contains x_h and x_{h+t-1}, which gives $b_i \le k-(t-1)d_j/(d_i,d_j)$ as before. ∎

There is, unfortunately, no direct analog of Lemma 2.3. Lemmas 5.1 and 5.2 together do, however, provide significant improvements over the trivial bounds (1.3).

Theorem 4. *Suppose $t \ge 3$. We have $N_t(k) \le (A_t + o(1))k^2$ as $k \to \infty$, where*

$$A_t = \frac{3t^2 - 2t + 1}{4t^2(t-1)}.$$

Proof. Suppose $1 \le m \le l \le \frac{k-1}{t-1}$ and define p_m to be the largest prime $\le 2m$. For each i,j let $e_{i,j} = d_i/(d_i,d_j)$. Suppose m is such that $p_m \ge k/t$, i.e. $k/t \le p_m \le \frac{2k-2}{t-1}$. Let

$$c = \max\left(k - \frac{t-1}{2}p_m, \frac{2t-1}{t}k - (t-1)p_m\right)$$

and note that $1 \le c \le k$. If $e_{i,j} \ge \frac{k}{t-1} - \frac{c}{t(t-1)}$ for some $i \le m, j \le m$, then Lemma 5.1 gives $b_j < c$. Otherwise for all $i \le m$, $j \le m$ we have

$$e_{i,j} < \frac{k}{t-1} - \frac{c}{t(t-1)} \le \frac{t-1}{t^2}k + \frac{p_m}{t} \le p_m.$$

Since $2m \ge p_m$, the proof of Lemma 2.4 implies that for some $i \le m$, $j \le m$ we have $e_{i,j} + e_{j,i} = p_m$. Assume without loss of generality that $e_{i,j} \ge e_{j,i}$. Lemma 5.2 implies that either

$$b_j \le k - (t-1)e_{i,j} < k - \frac{t-1}{2}p_m \le c$$

or

$$b_i \le k - te_{j,i} = k - p_m t + te_{i,j} < \frac{2t-1}{t-1}k - p_m t - \frac{c}{t-1} \le c.$$

In all cases $b_m \le \min(b_i, b_j) < c$, provided that $p_m \ge k/t$. Since $p_m = 2m - o(m)$, we have

$$\begin{aligned}\sum_{m=1}^{l} b_m &\le \sum_{\substack{m \le l \\ p_m \ge 2k/t}} \left(k - \frac{t-1}{2}p_m\right) + \sum_{k/t \le p_m < 2k/t} \left(\frac{2t-1}{t}k - p_m(t-1)\right) + \sum_{p_m < k/t} k \\ &\le \left(\frac{1}{2t^2(t-1)} + o(1)\right)k^2 + \left(\frac{t+1}{4t^2} + o(1)\right)k^2 + \left(\frac{1}{2t} + o(1)\right)k^2 \\ &= (A_t + o(1))k^2.\end{aligned}$$

∎

Remarks. The $o(1)$ can be removed with Lemmas 3.2 and 3.5 as in the proof of Theorem 2. Furthermore, by Lemma 3.2, it can easily be shown that $\sum b_i$ is at most the conjectured bound when $l \geq \frac{k-1}{t-1} - f(\frac{k-1}{t-1})$.

Thanks. The author thanks László Székely for posing the problem of bounding $N_t(k)$ and for helpful discussions, and thanks the referees for careful reading of the paper and pointing out several misprints in the original version.

References

[1] R. Balasubramanian and K. Soundararajan: On a conjecture of R. L. Graham, *Acta Arith.,* **75** (1996), 1–38.
[2] R. D. Boyle: On a problem of R. L. Graham, *Acta Arith.,* **34** (1978), 163–177.
[3] P. Erdős and C. Ko, R. Rado: Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser. 2,* **12** (1961), 313–318.
[4] K. Ford: A strong form of a problem of R. L. Graham (preprint).
[5] R. L. Graham: Problem 5749, *Amer. Math. Monthly,* **77** (1970), 775.
[6] R. Howard, G. Károlyi and L. Székely: Towards a Katona-type proof for the 2-intersecting Erdős–Ko–Rado theorem, *Electronic J. Combinatorics,* **8(1)** (2001), R31 (8 pages).
[7] G. O. H. Katona: A simple proof of the Erdős–Chao Ko–Rado theorem, *J. Combinatorial Theory Ser. B,* **13** (1972), 183–184.
[8] J. B. Rosser and L. Schoenfeld: Approximate formulas for some functions of prime numbers, *Illinois J. Math.,* **6** (1962), 64–94.
[9] L. Schoenfeld: Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$, *Math. Comp.,* **30** (1976), 337–360.
[10] M. Szegedy: The solution of Graham's greatest common divisor problem, *Combinatorica,* **6** (1986), 67–71.
[11] R. M. Wilson: The exact bound in the Erdős–Ko–Rado theorem, *Combinatorica,* **4** (1984), 247–257.
[12] A. Zaharescu: On a conjecture of Graham, *J. Number Theory,* **27** (1987), 33–40.

Kevin Ford

University of South Carolina,
Columbia, SC, USA.