On the Union of Arithmetic Progressions

Shoni Gilboa* Rom Pinchasi†

August 2, 2014

Abstract

We show that for any integer $n \geq 1$ and real $\epsilon > 0$, the union of n arithmetic progressions with pairwise distinct differences, each of length n, contains at least $c(\epsilon)n^{2-\epsilon}$ elements, where $c(\epsilon)$ is a positive constant depending only in ϵ. This estimate is sharp in the sense that the assertion becomes invalid for $\epsilon = 0$. We also obtain estimates for the "asymmetric case" where the number of progressions is distinct from their lengths.

1 Introduction

A finite arithmetic progression is one of the most fundamental objects in number theory. Formally it is a sequence of numbers $a_1, a_2, \ldots, a_n$ such that for every $1 < i < n$ we have $a_{i+1} - a_i = a_i - a_{i-1}$. The difference $a_i - a_{i-1}$ is called the *difference* of the arithmetic progression.

For integers $n > 1$ and $\ell > 1$ let $u_\ell(n)$ be the minimum possible cardinality of a union of n arithmetic progressions, each of length ℓ, with pairwise distinct differences.

Clearly, $u_\ell(n) \leq n\,\ell$, but this inequality is not tight in general, not even up to a multiplicative constant.

It is not hard to see, for instance, that $u_2(n) = \lceil \frac{1}{2} + \sqrt{2n + \frac{1}{4}} \rceil$. This is because any two numbers form an arithmetic progression of length 2 and therefore any set of m numbers such that no two of their differences are the same (for example $1, 2, 2^2, \ldots, 2^{m-1}$) is a union of $\binom{m}{2}$ arithmetic progressions with pairwise distinct differences. The following example, given in [9], shows that $u_3(n) = O\left(n^{1-\frac{\log 4}{\log 27}}\right)$: Let k be the minimal integer such that $\frac{1}{9k}3^{3k} \geq n$. For any disjoint $A, B \subset \{1, 2, \ldots, 3k\}$ of cardinality k, consider the three term arithmetic progression $\{2\sum_{i\in A} 3^i, \sum_{i\in A} 3^i + \sum_{i\in B} 3^i, 2\sum_{i\in B} 3^i\}$. We get $\frac{1}{2} \cdot \frac{(3k)!}{(k!)^3} \geq \frac{1}{9k}3^{3k} \geq n$ three term arithmetic progressions with pairwise distinct differences, whose union is of size $2\binom{3k}{k} \leq \frac{16}{9}\left(\frac{27}{4}\right)^k = O\left(n^{1-\frac{\log 4}{\log 27}}\right)$. We note that a lower bound $u_3(n) \geq n^{6/11}$ directly follows from a result of Katz and Tao [8].

*Mathematics Dept., The Open University of Israel, Raanana 43107 , Israel. tipshoni@gmail.com.

†Mathematics Dept., Technion—Israel Institute of Technology, Haifa 32000, Israel. room@math.technion.ac.il.
Supported by ISF grant (grant No. 1357/12).

The trivial upper bound $u_\ell(n) \le n\,\ell$ is not tight for large values of ℓ as well. In particular, consider the symmetric case where $\ell = n$. It turns out that $u_n(n) = o(n^2)$. To see this take (perhaps the most natural choice of) arithmetic progressions: $A_j = \{i\,j \mid i \in [n]\}$ for $j = 1, \ldots, n$. Here, and in the sequel, the notation $[n]$ stands for the set $\{1, 2, \ldots, n\}$. Clearly, the union $\bigcup_{j=1}^{n} A_j$ is precisely the set $\{i\,j \mid i, j \in [n]\}$. It was shown already in 1955 by Erdős [5] that

$$|\{i\,j \mid i, j \in [n]\}| = o\left(n^2/(\log n)^\alpha\right)$$

for some $\alpha > 0$. The exact assymptotics

$$|\{i\,j \mid i, j \in [n]\}| \sim \frac{n^2}{(\log n)^{1-\frac{1+\log\log 2}{\log 2}}(\log\log n)^{\frac{3}{2}}}$$

was given in 2008 by Ford [6]. Consequently, we obtain the desired improved upper bound for $u_n(n)$.

In this paper we show that $u_\ell(n)$ cannot be much smaller than $n\,\ell$, provided ℓ is not much smaller than n, as captured in the following theorem, giving a lower bound for $u_\ell(n)$ for smaller values of ℓ as well.

Theorem 1.1. *For every $\varepsilon > 0$ there is a positive constant $c_1(\varepsilon)$, depending only on ε, such that for any positive integers n and ℓ*

$$u_\ell(n) \ge \begin{cases} c_1(\varepsilon)\, n^{\frac{1}{2}-\varepsilon}\,\ell & \text{for } \ell \le n^{\frac{1}{2}-\varepsilon} \\ c_1(\varepsilon)\,\ell^2 & \text{for } n^{\frac{1}{2}-\varepsilon} \le \ell \le n^{1-\varepsilon} \\ c_1(\varepsilon)\, n^{1-\varepsilon}\,\ell & \text{for } n^{1-\varepsilon} \le \ell. \end{cases}$$

The proof of Theorem 1.1 relies upon upper bounds for the following two functions, that are of independent interest,

$$f_d(m,n) = \max_{\substack{A,B\subset(0,\infty)\\ |A|\le m\ ,\ |B|\le n}} \left|\left\{(a,b) \in A\times B \mid \frac{a}{b} \in [d]\right\}\right|,$$

$$g_d(n) = \max_{\substack{B\subset(0,\infty)\\ |B|\le n}} \left|\left\{(b_1,b_2) \in B^2 \mid b_1 < b_2,\ \exists p,q \in [d] : \frac{b_1}{b_2} = \frac{p}{q}\right\}\right|.$$

The paper is organized as follows. In Section 2 we provide an upper bound for the function g_d above. Using this bound, we provide an upper bound for the function f_d in Section 3. Theorem 1.1 is proved in Section 4. In Section 5 we give a number-theory application of our upper bound for the function g_d.

Addendum: After this paper was completed and before it was published it was brought to our attention that the main tools in our proof as well as the main lemma in our paper, which is

Proposition 2.2 in Section 2, appear in a different context in [1]. Lemma 3.9 in [1] gives only a slightly weaker bound than the one in Proposition 2.2 in this paper. Lemma 3.10 in [1] is a consequence of Lemma 3.9 in [1] in the same way that Theorem 5.1 in this paper is a consequence of Proposition 2.2. In fact, Lemma 3.10 in [1] is stated in a stronger form than Theorem 5.1 that allows us to get an explicit subquadratic lower bound for $u_\ell(n)$ in Theorem 1.1 in terms of n and ℓ without introducing $\epsilon > 0$. We comment about this in Section 5.

2 Rational quotients with bounded numerator and denominator

For positive integer d define

$$R_d = \left\{ \frac{k}{\ell} \mid k, \ell \in [d] \right\}.$$

Definition 2.1. *For a positive integer d and a finite set B of positive real numbers define*

$$\mathcal{G}_d(B) = \left\{ (b_1, b_2) \subseteq B \mid b_1 < b_2, \quad \frac{b_1}{b_2} \in R_d \right\}.$$

For positive integers n and d define

$$g_d(n) = \max_{\substack{B \subset (0,\infty) \\ |B| \leq n}} |\mathcal{G}_d(B)|.$$

Clearly,

$$g_d(n) \leq (n-1)|R_d| < n\, d^2. \tag{1}$$

This bound is useful when d is small. For large values of d we have the following improved upper bound whose proof is the main goal of this section:

Proposition 2.2. *For any positive integer k, there is a positive constant $c_2(k)$, depending only on k, such that for any positive integer n and any integer $d > c_2(k)$*

$$g_d(n) < (200k+1) n^{1+\frac{1}{k}} d^{1-\frac{1}{2k}}. \tag{2}$$

The proof of Proposition 2.2 will be as follows: Given a set B of n positive real numbers, $\mathcal{G}_d(B)$ can be viewed as the set of edges of a graph whose vertices are the numbers in B. If the number of edges in this graph is large, then the Bondy-Simonovits Theorem [3] implies that this graph must contain many cycles of length $2k$. However, the number of such cycles is easy to bound in terms of the number of solutions to $r_1 r_2 \cdots r_{2k} = 1$, where each of the r_i's is a quotient of two natural numbers that are smaller than or equal to d. This number of solutions can be estimated using standard estimates for the number of divisors function.

Hence, given a finite set of positive real numbers B we define

$$\mathcal{C}_{d,2k}(B) = \left\{(b_1, b_2, \ldots, b_{2k}) \in B^{2k} \mid \frac{b_1}{b_2}, \frac{b_2}{b_3}, \ldots, \frac{b_{2k-1}}{b_{2k}}, \frac{b_{2k}}{b_1} \in R_d, \forall 1 \leq i < j \leq 2k : b_i \neq b_j\right\}.$$

Notice that $\mathcal{C}_{d,2k}(B)$ corresponds to the set of cycles of length $2k$ in the graph corresponding to $\mathcal{G}_d(B)$. We start with bounding the cardinality of $\mathcal{C}_{d,2k}(B)$ from below in terms of $|\mathcal{G}_d(B)|$. We first get a basic lower bound for $|\mathcal{C}_{d,2k}(B)|$ using the Bondy-Simonovits Theorem, which states that a graph with n vertices and no simple cycles of length $2k$ has no more than $100k \cdot n^{1+\frac{1}{k}}$ edges. Later, we enhance this basic lower bound for $|\mathcal{C}_{d,2k}(B)|$ in the case where $|\mathcal{G}_d(B)|$ is large.

Lemma 2.3. *For any positive integers k and d, and for any finite set B of n positive real numbers,*

$$\frac{1}{4k} |\mathcal{C}_{d,2k}(B)| \geq |\mathcal{G}_d(B)| - 100k |B|^{1+\frac{1}{k}}.$$

Proof. Form a graph on the vertex set B, by connecting two distinct vertices $b_1, b_2 \in B$ if and only if $\frac{b_1}{b_2} \in R_d$. This graph obviously has $|B|$ vertices, $|\mathcal{G}_d(B)|$ edges, and precisely $\frac{1}{2\cdot 2k} |\mathcal{C}_{d,2k}(B)|$ simple cycles of length $2k$. Now remove an edge from every simple cycle of length $2k$ in this graph. We get a graph with $|B|$ vertices and at least

$$|\mathcal{G}_d(B)| - \frac{1}{4k} |\mathcal{C}_{d,2k}(B)|$$

edges. The resulting graph has no simple cycle of length $2k$. The result now follows directly from the Bondy-Simonovits Theorem stated above. □

Next, we use a standard probabilistic argument to enhance the lower bound of Lemma 2.3.

Lemma 2.4. *For any positive integers $k \geq 2$ and d, and for any finite set B of positive real numbers such that $|B| > d^{1+\frac{1}{2(k-1)}}$, we have the following inequality:*

$$|\mathcal{G}_d(B)| < \frac{1}{4k} |\mathcal{C}_{d,2k}(B)| \, d^{-(2k-1)} + 200k \, |B|^{1+\frac{1}{k}} d^{1-\frac{1}{2k}}.$$

Proof. Let $p := d^{-1-\frac{1}{2(k-1)}}$, and let B_p be a random subset of B obtained by choosing each element independently with probability p. By Lemma 2.3,

$$\frac{1}{4k} |\mathcal{C}_{d,2k}(B_p)| \geq |\mathcal{G}_d(B_p)| - 100k \, |B_p|^{1+\frac{1}{k}}.$$

Taking expectations, we get

$$\frac{1}{4k} E \, |\mathcal{C}_{d,2k}(B_p)| \geq E|\mathcal{G}_d(B_p)| - 100k \, E\left(|B_p|^{1+\frac{1}{k}}\right). \tag{3}$$

Notice that from the linearity of expectation we have:

$$E \, |\mathcal{C}_{d,2k}(B_p)| = |\mathcal{C}_{d,2k}(B)| \, p^{2k} \tag{4}$$

and

$$E|\mathcal{G}_d(B_p)| = |\mathcal{G}_d(B)|p^2. \tag{5}$$

As for $E\left(|B_p|^{1+\frac{1}{k}}\right)$, note that $E|B_p| = |B|p$ and $V|B_p| = |B|p(1-p)$. Therefore, since $1 < |B|p$,

$$E\left(|B_p|^2\right) = V|B_p| + (E|B_p|)^2 = |B|p(1-p) + (|B|p)^2 < 2|B|^2p^2.$$

Now, by Jensen's inequality,

$$E\left(|B_p|^{1+\frac{1}{k}}\right) \le \left(E\left(|B_p|^2\right)\right)^{\frac{1}{2}+\frac{1}{2k}} < 2|B|^{1+\frac{1}{k}}p^{1+\frac{1}{k}}. \tag{6}$$

Plugging (4), (5) and (6) in (3) we get

$$\frac{1}{4k}\,|\mathcal{C}_{d,2k}(B)|\,p^{2k} > |\mathcal{G}_d(B)|p^2 - 200k|B|^{1+\frac{1}{k}}p^{1+\frac{1}{k}},$$

hence

$$|\mathcal{G}_d(B)| < \frac{1}{4k}\,|\mathcal{C}_{d,2k}(B)|\,p^{2k-2} + 200k\,|B|^{1+\frac{1}{k}}p^{-1+\frac{1}{k}} = \frac{1}{4k}\,|\mathcal{C}_{d,2k}(B)|\,d^{-(2k-1)} + 200k\,|B|^{1+\frac{1}{k}}d^{1-\frac{1}{2k}}. \quad \square$$

We now approach the task of bounding $|\mathcal{C}_{d,2k}(B)|$ from above. We start with the following well known number-theoretic bound on the number of divisors $\tau(m)$ of an integer m.

Lemma 2.5. *For any $\delta > 0$ there is a positive constant $c_3(\delta)$ depending only on δ, such that for any positive integer m,*

$$\tau(m) < c_3(\delta)m^{\delta}.$$

Proof. Let $m = \prod_{i=1}^k p_i^{r_i}$ be the prime factorization of m. Notice that m has $\tau(m) = \prod_{i=1}^k (1+r_i)$ divisors. For any $1 \le i \le k$,

$$(p_i^{r_i})^{\delta} = e^{\delta r_i \ln p_i} > 1 + \delta r_i \ln p_i.$$

Therefore,

$$\frac{\tau(m)}{m^{\delta}} = \prod_{i=1}^k \frac{1+r_i}{(p_i^{r_i})^{\delta}} < \prod_{i=1}^k \frac{1+r_i}{1+\delta r_i \ln p_i} \le \prod_{i=1}^k \frac{1}{\min\{1, \delta \ln p_i\}} = \prod_{\substack{1\le i\le k \\ \ln p_i \le 1/\delta}} \frac{1}{\delta \ln p_i} \le \prod_{\substack{p \text{ prime} \\ p \le e^{1/\delta}}} \frac{1}{\delta \ln p}.$$

Hence, $\tau(m) < c_3(\delta)m^{\delta}$, where $c_3(\delta) := \prod_{\substack{p \text{ prime} \\ p \le e^{1/\delta}}} \frac{1}{\delta \ln p}$. $\square$

Lemma 2.6. *For any positive integer k there is a positive constant $c_4(k)$, depending only on k, such that for any positive integer d and any finite set B of positive real numbers we have*

$$|\mathcal{C}_{d,2k}(B)| < c_4(k)\,|B|\,d^{2k+\frac{1}{4k}}.$$

Proof. We notice that

$$\begin{aligned}
|\mathcal{C}_{d,2k}(B)| &= \left\{(b_1,b_2,\ldots,b_{2k})\in B^{2k} \mid \frac{b_1}{b_2},\frac{b_2}{b_3},\ldots,\frac{b_{2k-1}}{b_{2k}},\frac{b_{2k}}{b_1}\in R_d,\ \forall 1\le i<j\le 2k: b_i\ne b_j\right\} \le \\
&\le \left|\left\{(b_1,b_2,\ldots,b_{2k})\in B^{2k} \mid \forall 1\le i\le 2k-1: \frac{b_i}{b_{i+1}}\in R_d,\ \frac{b_{2k}}{b_1}\in R_d\right\}\right| \le \\
&\le |B|\cdot|\{(r_1,r_2,\ldots,r_{2k}) \mid \forall 1\le i\le 2k: r_i\in R_d,\ r_1r_2\cdots r_{2k}=1\}| \le \\
&\le |B|\cdot\left|\left\{((p_1,p_2,\ldots,p_{2k}),(q_1,q_2,\ldots,q_{2k}))\in \left([d]^{2k}\right)^2 \mid p_1p_2\cdots p_{2k}=q_1q_2\cdots q_{2k}\right\}\right| \le \\
&\le |B|\cdot\sum_{m=1}^{d^{2k}}\left|\left\{(p_1,p_2,\ldots,p_{2k})\in[d]^{2k} \mid p_1p_2\cdots p_{2k}=m\right\}\right|^2 \le |B|\cdot\sum_{m=1}^{d^{2k}}\tau(m)^{4k}.
\end{aligned}$$

By Lemma 2.5, $\tau(m) < c_3(1/32k^3)m^{1/32k^3}$ for any m, and we get

$$|\mathcal{C}_{d,2k}(B)| < |B|\cdot\sum_{m=1}^{d^{2k}}\left(c_3(1/32k^3)m^{1/32k^3}\right)^{4k} \le \left(c_3(1/32k^3)\right)^{4k}|B|\,d^{2k+\frac{1}{4k}}.$$

This completes the proof with $c_4(k) := \left(c_3(1/32k^3)\right)^{4k}$. □

We are now prepared for proving Proposition 2.2.

Proof of Proposition 2.2. If $n \le d^{1+\frac{1}{2(k-1)}}$, then (2) holds because

$$g_d(n) \le \binom{n}{2} < n^2 = n^{1+\frac{1}{k}}n^{1-\frac{1}{k}} \le n^{1+\frac{1}{k}}\left(d^{1+\frac{1}{2(k-1)}}\right)^{1-\frac{1}{k}} = n^{1+\frac{1}{k}}d^{1-\frac{1}{2k}}.$$

We therefore assume $n > d^{1+\frac{1}{2(k-1)}}$. Let B be a set of n positive real numbers. By Lemma 2.4,

$$|\mathcal{G}_d(B)| < 200k\,n^{1+\frac{1}{k}}d^{1-\frac{1}{2k}} + \frac{1}{4k}|\mathcal{C}_{d,2k}(B)|\,d^{-(2k-1)}. \tag{7}$$

By Lemma 2.6,

$$|\mathcal{C}_{d,2k}(B)| < c_4(k)n\,d^{2k+\frac{1}{4k}}.$$

Hence, for $d \ge c_2(k) := (c_4(k)/4k)^{4k}$,

$$|\mathcal{C}_{d,2k}(B)| < 4k\,n\,d^{2k+\frac{1}{2k}}. \tag{8}$$

Plugging (8) in (7) and using our assumption that $n > d^{1+\frac{1}{2(k-1)}} \ge d$, we get that for $d \ge c_2(k)$,

$$|\mathcal{G}_d(B)| < 200k\,n^{1+\frac{1}{k}}d^{1-\frac{1}{2k}} + n\,d^{1+\frac{1}{2k}} < (200k+1)n^{1+\frac{1}{k}}d^{1-\frac{1}{2k}}.$$ □

3 Bounded integer quotients

Definition 3.1. *For positive integers* m, n*, and* d *define*

$$f_d(m,n) = \max_{\substack{A,B\subset(0,\infty) \\ |A|\le m\ ,\ |B|\le n}} |\{(a,b)\in A\times B \mid \frac{a}{b}\in[d]\}|.$$

Remark 3.2. *Notice that* $f_d(m,n) = f_d(n,m)$*. This is because given sets* A *and* B *of positive real numbers such that* $|A| = m$*,* $|B| = n$*, and* $f_d(m,n) = |\{(a,b)\in A\times B \mid \frac{a}{b}\in[d]\}|$*, the sets* $A' = \{\frac{1}{b} \mid b\in B\}$ *and* $B' = \{\frac{1}{a} \mid a\in A\}$ *show that* $f_d(n,m) \ge f_d(m,n)$*. Consequently,* $f_d(m,n) = f_d(n,m)$*. Therefore, we may assume, if needed, with no loss of generality that* $m \le n$*, or that* $m \ge n$*.*

The main goal of this section is to prove the following proposition.

Proposition 3.3. *For any* $\varepsilon > 0$ *there is a positive constant* $c_6(\varepsilon)$*, depending only on* ε*, such that for any positive integers* m, n*, and* d

$$f_d(m,n) < c_5(\varepsilon)\min\{n^\varepsilon, m^\varepsilon\}\sqrt{m\,n\,d}.$$

Proof. With no loss of generality (see Remark 3.2) assume that $n \le m$. We may also assume that $m < n\,d$, because if $m \ge n\,d$, then $f_d(m,n) \le n\,d = \sqrt{(n\,d)n\,d} \le \sqrt{m\,n\,d}$ (see Proposition 3.5 for further discussion).

Let A and B be finite sets of positive real numbers such that $|A| \le m, |B| \le n$ and $f_d(m,n) = |\{(a,b)\in A\times B \mid \frac{a}{b}\in[d]\}|$. The proposition will follow by comparing lower and upper bounds for the cardinality of the set

$$W = \{(a,b_1,b_2)\in A\times B^2 \mid \frac{a}{b_1},\frac{a}{b_2}\in[d],\quad b_1<b_2\}.$$

We first establish an upper bound for $|W|$. For convenience define

$$S_d = \{(p,q) \mid p,q\in[d],\quad p<q,\quad \gcd(p,q)=1\}.$$

We have:

$$\begin{aligned}
|W| &= \left|\{(b_1,b_2,k_1,k_2)\in B^2\times[d]^2 \mid k_1b_1=k_2b_2\in A,\quad b_1<b_2\}\right| \leq \\
&\leq \left|\{(b_1,b_2,k_1,k_2)\in B^2\times[d]^2 \mid k_1b_1=k_2b_2,\quad b_1<b_2\}\right| = \\
&= \sum_{(p,q)\in S_d}\left|\left\{(b_1,b_2,k_1,k_2)\in B^2\times[d]^2 \mid \frac{b_1}{b_2}=\frac{k_2}{k_1}=\frac{p}{q}\right\}\right| = \\
&= \sum_{(p,q)\in S_d}\left|\left\{(b_1,b_2)\in B^2 \mid \frac{b_1}{b_2}=\frac{p}{q}\right\}\right|\cdot\left|\left\{(k_1,k_2)\in [d]^2 \mid \frac{k_2}{k_1}=\frac{p}{q}\right\}\right| = \\
&= \sum_{(p,q)\in S_d}\left|\left\{(b_1,b_2)\in B^2 \mid \frac{b_1}{b_2}=\frac{p}{q}\right\}\right|\cdot\lfloor\frac{d}{q}\rfloor \leq \\
&\leq \sum_{q=2}^{d}\left(|\mathcal{G}_q(B)|-|\mathcal{G}_{q-1}(B)|\right)\frac{d}{q} = |\mathcal{G}_d(B)|+\sum_{q=2}^{d-1}|\mathcal{G}_q(B)|\left(\frac{d}{q}-\frac{d}{q+1}\right) = \\
&< |\mathcal{G}_d(B)|+d\sum_{q=2}^{d-1}q^{-2}|\mathcal{G}_q(B)|.
\end{aligned}$$

Let $k := \lceil 1/2\varepsilon\rceil$. By Proposition 2.2, there is a positive constant $c_2(k)$, depending only on k, such that for any $c_2(k) < q \leq d$,

$$|\mathcal{G}_q(B)| < (200k+1)n^{1+\frac{1}{k}}q^{1-\frac{1}{2k}}.$$

For any q, we have by (1) that $|\mathcal{G}_q(B)| < n\,q^2$.

Therefore,

$$\begin{aligned}
|W| &\leq |\mathcal{G}_d(B)|+d\sum_{q=2}^{c_2(k)}q^{-2}|\mathcal{G}_q(B)|+d\sum_{q=c_2(k)+1}^{d-1}q^{-2}|\mathcal{G}_q(B)| < \\
&< (200k+1)n^{1+\frac{1}{k}}d^{1-\frac{1}{2k}}+(c_2(k)-1)n\,d+(200k+1)n^{1+\frac{1}{k}}d\sum_{q=c_2(k)+1}^{d-1}\frac{1}{q^{1+\frac{1}{2k}}}.
\end{aligned}$$

Hence,

$$|W| < c(\varepsilon)n^{1+2\varepsilon}d, \tag{9}$$

where $c(\varepsilon) := (200k+1)+(c_2(k)-1)+(200k+1)\displaystyle\sum_{q=c_2(k)+1}^{\infty}\frac{1}{q^{1+\frac{1}{2k}}}$.

To get a lower bound for $|W|$, we define $r(a) = |\{b\in B \mid \frac{a}{b}\in[d]\}|$ for any $a\in A$. Then, by the convexity of the function $\binom{x}{2}=\frac{x(x-1)}{2}$ (or what is sometime referred to as Jensen's inequality):

$$|W| = \sum_{a\in A}\binom{r(a)}{2} \geq m\binom{\frac{1}{m}\sum_{a\in A}r(a)}{2}. \tag{10}$$

Combining the upper and lower bounds for $|W|$, namely, (9) and (10), we get

$$m\binom{\frac{1}{m}\sum_{a\in A} r(a)}{2} < c(\varepsilon)n^{1+2\varepsilon}d.$$

Now, we deduce

$$f_d(m,n) = |\{(a,b)\in A\times B \mid \frac{a}{b}\in[d]\}| = \sum_{a\in A} r(a) < \frac{m}{2} + \sqrt{\frac{m^2}{4} + 2c(\varepsilon)m\,n^{1+2\varepsilon}d.}$$

This implies the desired result, as $n \le m < n\,d$. □

3.1 Tightness of Proposition 3.3

In this section we deviate from the flow of the argument to address the question whether Proposition 3.3 is tight. The results of this section will not be used elsewhere.

We will show that the upper bound in Proposition 3.3 for $f_d(n,m)$ is essentially tight (see Proposition 3.4 below), provided each of the parameters m, n, and d is (sufficiently) smaller than the product of the other two. When one of m, n, and d is considerably larger than the product of the other two, the upper bound in Proposition 3.3 is no longer tight, as follows from Proposition 3.5 below, in which the exact values of $f_d(m,n)$ in those cases are determined.

Proposition 3.4. *If $m \le \frac{1}{4}nd$, $n \le \frac{1}{4}md$, and $d \le \frac{1}{4}mn$, then $f_d(m,n) \ge \frac{1}{8}\sqrt{m\,n\,d}$.*

Proof. Set $k = \lfloor\sqrt{m\,d/n}\rfloor$, $\ell = \lfloor\sqrt{n\,d/m}\rfloor$, and $t = \lfloor\sqrt{m\,n/d}\rfloor$. Consider the sets

$$A = \{(k+\ell)^r\, i\}_{r\in[t],i\in[k]}\,, \quad \text{and} \quad B = \{(k+\ell)^r/j\}_{r\in[t],j\in[\ell]}\,.$$

Then

$$|A| = t\,k \le \sqrt{\frac{m\,n}{d}}\cdot\sqrt{\frac{m\,d}{n}} = m \quad \text{and}$$

$$|B| = t\,\ell \le \sqrt{\frac{m\,n}{d}}\cdot\sqrt{\frac{n\,d}{m}} = n.$$

Notice that

$$\left|\left\{(a,b)\in A\times B \mid \frac{a}{b}\in[d]\right\}\right| \ge t\,k\,\ell \ge \frac{1}{2}\sqrt{\frac{m\,n}{d}}\cdot\frac{1}{2}\sqrt{\frac{m\,d}{n}}\cdot\frac{1}{2}\sqrt{\frac{n\,d}{m}} = \frac{1}{8}\sqrt{m\,n\,d}. \quad □$$

Proposition 3.5.
1. *If $d \ge m\,n$ then $f_d(m,n) = m\,n$.*
2. *If $n \ge m\,d$ then $f_d(m,n) = m\,d$.*
3. *If $m \ge n\,d$ then $f_d(m,n) = n\,d$.*

Proof. 1. For any $A, B \subset (0, \infty)$ with $|A| \le m, |B| \le n$ we obviously have

$$|\{(a,b) \in A \times B \mid \frac{a}{b} \in [d]\}| \le |A \times B| = |A| \cdot |B| \le m\,n.$$

To see that this upper bound can actually be attained, consider, for instance, the sets $A = \{1/i\}_{i \in [m]}$ and $B = [n]$.

2. For any $A, B \subset (0, \infty)$ with $|A| \le m, |B| \le n$ we have

$$|\{(a,b) \in A \times B \mid \frac{a}{b} \in [d]\}| = |\{(a,k) \in A \times [d] \mid \frac{a}{k} \in B\}| \le m\,d.$$

This upper bound can indeed be attained, for example by taking $A = \{(d+1)^i\}_{i \in [m]}$ and $B = \{(d+1)^i/k\}_{i \in [m], k \in [d]}$.

3. For any $A, B \subset (0, \infty)$ with $|A| \le m, |B| \le n$ we have

$$|\{(a,b) \in A \times B \mid \frac{a}{b} \in [d]\}| = |\{(b,k) \in B \times [d] \mid k \cdot b \in A\}| \le n\,d.$$

Equality is attained, for example, by taking $A = \{(d+1)^j k\}_{j \in [n], k \in [d]}$ and $B = \{(d+1)^j\}_{j \in [n]}$.

□

4 Union of arithmetic progressions

In this Section we prove Theorem 1.1.

Recall that for integers $n > 1$ and $\ell > 1$, $u_\ell(n)$ is the minimum possible cardinality of a union of n arithmetic progressions, each of length ℓ, with pairwise distinct differences.

As a consequence of Proposition 3.3 we prove the first estimate of Theorem 1.1.

Proposition 4.1. *For any $\varepsilon > 0$ there is a positive constant $c_6(\varepsilon)$, depending only on ε, such that for any positive integers n and ℓ*

$$u_\ell(n) > c_6(\varepsilon) n^{\frac{1}{2} - \varepsilon} \ell.$$

Proof. Take n arithmetic progressions, each of length ℓ, with pairwise distinct differences, and let U be their union. If each $x \in U$ belongs to less than $\sqrt{n}$ of the progressions, then $n\,\ell < |U|\sqrt{n}$ and consequently $|U| > \sqrt{n}\,\ell > n^{\frac{1}{2} - \varepsilon} \ell$.

Therefore, assume there is $x \in U$ which belongs to at least $\sqrt{n}$ progressions. In any such progression at least $d := \lceil \frac{\ell - 1}{2} \rceil$ of the terms are on the same side of x (that is, are either all smaller, or all larger than x). Therefore, in at least $\sqrt{n}/2$ progressions there are at least d terms on the same side of x and without loss of generality we assume they are larger than x in these progressions. We now concentrate only on these progressions. Let B be the set of differences of these arithmetic progressions, and let $A = \{i\,b \mid i \in [d], \;\; b \in B\}$. Proposition 3.3 implies

$$d\,|B| = |\{(a,b) \in A \times B \mid \frac{a}{b} \in [d]\}| \le f_d(|A|, |B|) < c_5(\varepsilon)|B|^\varepsilon \sqrt{|A|\,|B|\,d},$$

hence

$$|U| \geq |\{x+a \mid a \in A\}| = |A| > \frac{1}{c_5(\varepsilon)^2}|B|^{1-2\varepsilon}d \geq \frac{1}{c_5(\varepsilon)^2}\left(\frac{\sqrt{n}}{2}\right)^{1-2\varepsilon}\frac{\ell-1}{2} \geq \frac{1}{c_5(\varepsilon)^2\, 2^{3-2\varepsilon}}n^{\frac{1}{2}-\varepsilon}\ell.$$

This completes the proof with $c_6(\varepsilon) := \dfrac{1}{c_5(\varepsilon)^2\, 2^{3-2\varepsilon}}$. □

The lower bounds in Theorem 1.1 in the regime $n^{\frac{1}{2}-\varepsilon} \leq \ell$ are established in Proposition 4.3 below. The proof of Proposition 4.3 uses Proposition 2.2, ideas similar to those appearing in the proof of Proposition 3.3, and the following lemma (recall the definition of R_d from Section 2).

Lemma 4.2. *Suppose that the increasing arithmetic progressions* $(a_1+(j-1)b_1)_{j=1}^{\ell}$ *and* $(a_2+(j-1)b_2)_{j=1}^{\ell}$ *have at least* $r \geq 2$ *common elements, then*

$$\frac{b_1}{b_2} \in R_{\lfloor \frac{\ell-1}{r-1} \rfloor}.$$

Proof. Since $(a_1+(j-1)b_1)_{j=1}^{\ell}$ and $(a_2+(j-1)b_2)_{j=1}^{\ell}$ have at least two common elements, $\frac{b_1}{b_2}$ is necessarily rational, and with no loss of generality we may assume b_1, b_2 are both integers. The intersection of the two progressions under consideration is then itself an arithmetic progression with the difference $\operatorname{lcm}(b_1,b_2)$ and the diameter (which is the difference between the largest and the smallest of its elements) at least $(r-1)\operatorname{lcm}(b_1,b_2)$. Consequently, $(r-1)\operatorname{lcm}(b_1,b_2) \leq (\ell-1)b_i$ and therefore $b_i/\gcd(b_1,b_2) \leq \frac{\ell-1}{r-1}$ $(i \in \{1,2\})$, which is equivalent to the assertion of the lemma. □

To complete the proof of Theorem 1.1, we establish

Proposition 4.3. *For any* $\varepsilon > 0$ *there is a positive constant* $c_7(\varepsilon)$*, depending only on* ε*, such that for any positive integers* n *and* ℓ

$$u_\ell(n) > c_7(\varepsilon)\min\left\{n^{1-\varepsilon}\,\ell,\, \ell^2\right\}.$$

Proof. Let $P_1, P_2, \ldots, P_n$ be n arithmetic progressions, each of length ℓ, with pairwise distinct differences. We will use the following well known estimate of Dawson and Sankoff [4] on the cardinality of the union of sets via the cardinalities of their pairwise intersections,

$$|\bigcup_{i=1}^{n} P_i| \geq \frac{(\sum_{i=1}^{n}|P_i|)^2}{\sum_{1\leq i,j\leq n}|P_i \cap P_j|}. \tag{11}$$

Hence, we examine $I := \sum_{1\leq i_1<i_2\leq n}|P_{i_1}\cap P_{i_2}|$.

Let $b_1, \ldots, b_n$ be the differences of the progressions $P_1, \ldots, P_n$, respectively and let $B = \{b_1, \ldots, b_n\}$. Clearly

$$I = \sum_{r=1}^{\ell-1}\left|\left\{(i_1,i_2) \in [n]^2 \mid b_{i_1} < b_{i_2},\ \ |P_{i_1}\cap P_{i_2}| \geq r\right\}\right|.$$

Trivially,

$$\left|\left\{(i_1,i_2)\in[n]^2 \mid b_{i_1}<b_{i_2},\quad |P_{i_1}\cap P_{i_2}|\ge 1\right\}\right| \le \binom{n}{2}.$$

For $r\ge 2$ we use Lemma 4.2 to obtain

$$\left|\left\{(i_1,i_2)\in[n]^2 \mid b_{i_1}<b_{i_2},\quad |P_{i_1}\cap P_{i_2}|\ge r\right\}\right| \le \left|\mathcal{G}_{\lfloor\frac{\ell-1}{r-1}\rfloor}(B)\right| \le g_{\lfloor\frac{\ell-1}{r-1}\rfloor}(n).$$

Hence,

$$I \le \binom{n}{2} + \sum_{r=2}^{\ell-1} g_{\lfloor\frac{\ell-1}{r-1}\rfloor}(n).$$

Let $k:=\lceil 1/\varepsilon\rceil$. By Proposition 2.2, there is a constant $c_2(k)$ such that for any $2\le r\le \lfloor\frac{\ell-1}{c_2(k)+1}\rfloor+1$ we have

$$g_{\lfloor\frac{\ell-1}{r-1}\rfloor}(n) < (200k+1)n^{1+\frac{1}{k}}\left(\left\lfloor\frac{\ell-1}{r-1}\right\rfloor\right)^{1-\frac{1}{2k}} \le (200k+1)n^{1+\frac{1}{k}}(\ell-1)^{1-\frac{1}{2k}}\frac{1}{(r-1)^{1-\frac{1}{2k}}}.$$

For $\lfloor\frac{\ell-1}{c_2(k)+1}\rfloor+2\le r\le \ell-1$, we use the simpler estimate (1) to get

$$g_{\lfloor\frac{\ell-1}{r-1}\rfloor}(n) < n\left(\left\lfloor\frac{\ell-1}{r-1}\right\rfloor\right)^2 < (c_2(k)+1)^2\,n.$$

Therefore,

$$I \le \binom{n}{2} + \sum_{r=2}^{\lfloor\frac{\ell-1}{c_2(k)+1}\rfloor+1}(200k+1)n^{1+\frac{1}{k}}(\ell-1)^{1-\frac{1}{2k}}\frac{1}{(r-1)^{1-\frac{1}{2k}}} + \sum_{r=\lfloor\frac{\ell-1}{c_2(k)+1}\rfloor+2}^{\ell-1}(c_2(k)+1)^2\,n.$$

Hence

$$I < c(\varepsilon)\,n\,\ell\,\max\{n/\ell, n^{1/k}\} \le c(\varepsilon)\,n\,\ell\,\max\{n/\ell, n^{\varepsilon}\}, \tag{12}$$

for some positive constant $c(\varepsilon)$ depending only on ε.

The proposition follows by plugging (12) in (11). □

5 An application: Graham's conjecture on average

In this section we draw a number theoretical application to our upper bounds for the function g_d in Section 2. This application, apart from providing an alternative presentation for the proof of Proposition 4.3, is directly related to a famous conjecture of Graham [7].

Theorem 5.1. *For every $\varepsilon>0$ there exists $c(\varepsilon)>0$ with the following property. Let $a_1<\ldots<a_n$ be n natural numbers. Then*

$$\sum_{1\le i<j\le n} \frac{\gcd(a_i,a_j)}{a_j} < c(\varepsilon) n^{1+\varepsilon}. \tag{13}$$

Proof. Denote $A = \{a_1,\dots,a_n\}$. Notice that every summand on the left-hand side of (13) is of the form $\frac{1}{d}$ for some positive integer d. The simple but crucial observation is that if $1 \le i < j \le n$ such that $\frac{\gcd(a_i,a_j)}{a_j} = \frac{1}{d}$, then $\frac{a_i}{a_j} \in R_d$. Therefore, $\frac{\gcd(a_i,a_j)}{a_j} = \frac{1}{d}$, for $1 \le i < j \le n$, if and only if $(a_i,a_j) \in \mathcal{G}_d(A) \setminus \mathcal{G}_{d-1}(A)$. (Recall the definition of R_d and $\mathcal{G}_d(A)$ in Section 2.)

Fix a positive integer k, to be determined later. By Proposition 2.2, there exists $c_2(k) > 0$ such that for every $d > c_2(k)$

$$|\mathcal{G}_d(A)| \le g_d(n) < (200k+1) n^{1+\frac{1}{k}} d^{1-\frac{1}{2k}}. \tag{14}$$

For every d, $|\mathcal{G}_d(A)| \le g_d(n) < nd^2$, by (1). We get that

$$\begin{aligned}
\sum_{1\le i<j\le n} \frac{\gcd(a_i,a_j)}{a_j} &= \sum_{d\ge 2} \left| \left\{ (i,j) \in [n]^2 \mid i<j,\ \frac{\gcd(a_i,a_j)}{a_j} = \frac{1}{d} \right\} \right| \cdot \frac{1}{d} = \\
&= \sum_{d\ge 2} \left(|\mathcal{G}_d(A)| - |\mathcal{G}_{d-1}(A)|\right) \frac{1}{d} = \sum_{d\ge 2} |\mathcal{G}_d(A)| \left(\frac{1}{d} - \frac{1}{d+1} \right) \le \\
&\le \sum_{2\le d\le c_2(k)} d^{-2} |\mathcal{G}_d(A)| + \sum_{d>c_2(k)} d^{-2} |\mathcal{G}_d(A)| \le \\
&\le c_2(k) n + (200k+1) n^{1+\frac{1}{k}} \sum_{d>c_2(k)} \frac{1}{d^{1+\frac{1}{2k}}}.
\end{aligned}$$

Take k to be a positive integer such that $\frac{1}{k} < \varepsilon$ and let $c(\varepsilon) = c_2(k) + (200k+1) \sum_{d>c_2(k)} \frac{1}{d^{1/2k}(d+1)}$ to get the desired result. □

Remark. *It is not hard to verify that the bound in Theorem 5.1 cannot be improved to be linear in n. This can be seen for example by taking $a_1,\dots,a_n$ to be $1,\dots,n$, respectively. Then a direct computation, using some classical number theory estimates, shows that in this case the left-hand side of (13) is of the order of magnitude of $n \log n$, up to some absolute multiplicative constant.*

Theorem 5.1 allows us to write in a slightly different way the proof of Proposition 4.3.

Indeed, suppose we wish to bound from below the union of n arithmetic progressions, $P_1,\dots,P_n$, each of length ℓ, with pairwise distinct differences $a_1,\dots,a_n$, respectively. With no loss of generality we may assume that $a_1 < \ldots < a_n$ and that they are all positive integers. We will again use (11). Hence, we examine the cardinalities of the pairwise intersections of the progressions.

Consider two progressions of length ℓ: $\{p+(j-1)q\}_{j=1}^{\ell}$ and $\{p'+(j-1)q'\}_{j=1}^{\ell}$, where q,q' are positive integers. Their intersection is in itself an arithmetic progression and it is not hard to see that the difference of this progression (assuming it has at least two elements) is equal to the smallest number divisible by both q and q'. It follows that the size of the intersection of the two progressions is less than or equal to $1 + \frac{\min(\ell q, \ell q')}{\mathrm{lcm}(q,q')} = 1 + \ell\frac{\gcd(q,q')}{\max(q,q')}$.

It follows from (11) and the above discussion that the union $|\bigcup_{i=1}^{n} P_i|$ is bounded from below by

$$\frac{(n\ell)^2}{n\ell + n^2 + 2\ell\sum_{1\leq i<j\leq n}\frac{\gcd(a_i,a_j)}{a_j}}.$$

In view of Theorem 5.1, this expression is greater than $\min(\frac{1}{3c(\varepsilon)}n^{1-\varepsilon}\ell, \frac{1}{2}\ell^2)$.

It is interesting to note the relation of Theorem 5.1 to a well known conjecture of Graham [7]. Graham conjectured that given any n positive integers $a_1 < \ldots < a_n$, there are two of them that satisfy $\frac{a_j}{\gcd(a_i,a_j)} \geq n$. This conjecture has a long history with many contributions. It was finally completely (that is, for all values of n) solved in [2], where one can also find more details on the history and references related to this conjecture.

From (13) it follows that there is a pair of indices $1 \leq i < j \leq n$ such that $\frac{\gcd(a_i,a_j)}{a_j} < \frac{c(\varepsilon)n^{1+\varepsilon}}{\binom{n}{2}}$. This implies $\frac{a_j}{\gcd(a_i,a_j)} > \frac{1}{2c(\varepsilon)}n^{1-\varepsilon}$. This lower bound is indeed much weaker than the desired one in the conjecture of Graham, but on the other hand this argument shows that "on average" $\frac{a_j}{\gcd(a_i,a_j)}$ is quite large.

Addendum: As was mentioned in the introduction of this paper, Theorem 5.1 appears already in [1] (as lemma 3.10 there). In fact, Lemma 3.10 in [1] is stated in a stronger form:

Lemma 5.2 (Lemma 3.10 in [1])**.** *There exists an absolute constant $c > 0$ such that for any positive integers $a_1 < \ldots < a_n$, we have*

$$\sum_{1\leq i<j\leq n}\frac{gcd(a_i,a_j)}{a_j} < ne^{c\sqrt{\log n \log\log n}}. \tag{15}$$

As we have seen above $u_\ell(n)$ is bounded from below by

$$\frac{(n\ell)^2}{n\ell + n^2 + 2\ell\sum_{1\leq i<j\leq n}\frac{\gcd(a_i,a_j)}{a_j}}.$$

Plugging here the upper bound from Lemma 3.10 from [1] for $\sum_{1\leq i<j\leq n}\frac{\gcd(a_i,a_j)}{a_j}$, we get the following lower bound for $u_\ell(n)$:

$$u_\ell(n) \geq \frac{(n\ell)^2}{n\ell + n^2 + 2\ell n e^{c\sqrt{\log n \log\log n}}},$$

where $c > 0$ is an absolute positive constant independent of ℓ and n.

Acknowledgments

We thank Vsevolod F. Lev for interesting discussions about the problem and for pointing out the relation of Theorem 5.1 to the conjecture of Graham. We thank Noga Alon for bringing to our attention the close relation between [1] and this paper.

References

[1] N. Alon, I. Z. Ruzsa, Non-averaging subsets and non-vanishing transversals. *J. Combin. Theory Ser. A* **86** (1999), no. 1, 1–13.

[2] R. Balasubramanian, K. Soundararajan, On a conjecture of R. L. Graham. *Acta Arith.* **75** (1996), no. 1, 1–38.

[3] J. A. Bondy and M. Simonovits, Cycles of even length in graphs, J. Combinatorial Theory Ser. B **16** (1974), 97–105.

[4] D. A. Dawson and D. Sankoff, An inequality for probabilities, Proc. Amer. Math. Soc. **18** (1967), 504–507.

[5] P. Erdös, Some remarks on number theory, Riveon Lematematika **9** (1955), 45–48.

[6] K. Ford, The distribution of integers with a divisor in a given interval, Ann. of Math. (2) **168** (2008), no. 2, 367–433.

[7] R. L. Graham, Advanced Problem 5749*, *Amer. Math. Monthly* **77** (1970), 775.

[8] N. H. Katz and T. Tao, Bounds on arithmetic projections, and applications to the Kakeya conjecture, Math. Res. Lett. **6** (1999), no. 5-6, 625–630.

[9] I. Z. Ruzsa, Sumsets, in *European Congress of Mathematics*, 381–389, Eur. Math. Soc., Zürich.