A STRONG FORM OF A PROBLEM OF R. L. GRAHAM

KEVIN FORD

ABSTRACT. If A is a set of M positive integers, let $G(A)$ be the maximum of $a_i/\gcd(a_i, a_j)$ over $a_i, a_j \in A$. We show that if $G(A)$ is not too much larger than M, then A must have a special structure.

1. Introduction

In 1970, R. L. Graham [3] conjectured that for any set of n positive integers, there are two of them, say a and b, such that $a/(a,b) \ge n$. Here (a,b) is the greatest common divisor of a and b. Graham's conjecture was proved for all large n independently by Zaharescu [5] and Szegedy [4] in the mid-1980s. Introducing several new ideas, and making use of explicit bounds for prime number counting functions, Balasubramanian and Soundararajan [1] recently proved the conjecture for all n. They also noted that their method of proof could be used to prove a stronger form of Graham's conjecture, but gave no details.

For a set $A = \{a_1, \dots, a_n\}$ of positive integers, define

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

which we refer to as the dual of A. Let $G(A)$ be the maximum over all i, j of $\dfrac{a_i}{(a_i, a_j)}$. We will confine our discussion to sets with $\gcd(a_1, \dots, a_n) = 1$, since $G(A) = G(dA)$, where $dA = \{da_1, \dots, da_n\}$. Also, since $\dfrac{a_i}{(a_i, a_j)} = \dfrac{L/a_j}{(L/a_i, L/a_j)}$ for all i, j, it follows that $G(A) = G(A^*)$.

Theorem BS (Balasubramanian-Soundararajan [1]) *Let $n > 4$. For every set A of n positive integers, $G(A) \ge n$. Furthermore, if $G(A) = n$ then either A or A^* is equal to $\{1, 2, \dots, n\}$.*

The strengthening of Graham's conjecture which we are concerned with is an extension of the second part of the conjecture. We show that if A is

---

2000 *Mathematics Subject Classification.* 11A05.

Research supported in part by National Science Foundation grant DMS-0070618.

a set of M positive integers and $G(A) = N$ with N "not too much larger" than M, then either A or A^* lies in $\{1, 2, \dots, N\}$.

Definition. Let $f(N)$ denote the largest number R so that the following holds: for every set A of M positive integers with $N - R \le M \le N$ and $G(A) \le N$, either A or A^* lies in $\{1, 2, \dots, N\}$.

Theorem 1. *We have* $f(N) \ge \dfrac{cN \log\log N}{\log^2 N}$ *for large* N*, where* $c > 0$ *is an absolute constant.*

Lower bounds for $f(N)$ have an application to a problem of determining the maximum number of k-term arithmetic progressions of real numbers one can have, any two of which have two elements in common (see [2]). This in fact was the motivation for this work (In [2] a crude bound $f(N) \ge 0.156\frac{N}{\log^3 N}$ for $N \ge e^{10000}$ is proved). In this paper we concentrate only on the behavior of the bound for large N, as a totally explicit version of Theorem 2 would require a great deal of extra computation. By Theorem BS, $f(N) \ge 0$ for $N \ge 5$. A natural question is to determine the smallest X so that $f(N) \ge 1$ for $N \ge X$. The example $A = \{2, 3, 4, 6, 8, 9, 10, 12, 18\}$ shows that $f(10) = 0$. Perhaps one can prove that $f(N) \ge 1$ for $N \ge 11$ using the methods in [1].

Remark. Balasubramanian and Soundararajan claim that their method yields $f(N) \ge \dfrac{cN}{\log N \log\log N}$, but this appears to be too optimistic.

We can also show a non-trivial upper bound on $f(N)$.

Theorem 2. *We have* $f(N) = O\left(\dfrac{N}{\log\log N}\right)$.

Proof. Suppose that N is large, set $L = \frac{1}{2}\log N$ and let H be the product of the primes $\le L$. By the Prime Number Theorem, $N^{2/5} \le H \le N^{3/5}$ for large N. Let $N_0 = H\lfloor N/H \rfloor$ so that $N \ge N_0 \ge N - H \ge N - N^{3/5}$. Here $\lfloor x \rfloor$ denotes the largest integer $\le x$. Let

$$A = \{m \le N_0 : (m, H) > 1\} \cup \{2N_0\}.$$

It is clear that $G(A) \le N$ and neither A nor A^* is a subset of $\{1, 2, \dots, N\}$. Also

$$\begin{aligned} |A| &= N_0 + 1 - \phi(N_0) = N_0 + 1 - N_0 \prod_{p \le L} (1 - 1/p) \\ &\ge N_0 - \frac{c_1 N_0}{\log L} \ge N - \frac{c_2 N}{\log\log N}. \end{aligned}$$

Here c_1, c_2 are positive absolute constants. □

2. General lower bounds

We first need to introduce some of the notation from [1]. Suppose $A = \{a_1, \dots, a_M\}$, $\gcd(a_1, \cdots, a_M) = 1$, $N \ge 7$ and $G(A) \le N$. If p is a prime in $(1.5N, 2N)$ and $p - N \le m \le N$, define

$$R_p(m) = \left\{ \text{pairs } (a_i, a_j) : \frac{a_i}{(a_i, a_j)} = m, \frac{a_j}{(a_i, a_j)} = p - m \right\} \tag{2.1}$$

and put $r_p(m) = |R_p(m)|$. Our proof is based on upper and lower bounds for averages of $r_p(m)$. Suppose that neither A nor A^* lies in $\{1, 2, \dots, N\}$, and that $N/2 + 2 < M \le N$. We need not consider M outside this range, since the set $A = \{a \le N : (6, N) > 1\} \cup \{6\lfloor N/3 \rfloor\}$ shows that $f(N) \le N/3 - 1 < N/2 - 2$ for $N \ge 7$. By Lemmas 4.1 and 4.2 of [2],

$$\begin{aligned} \sum_{\substack{\frac{p+1}{2} \le m \le N \\ r_p(m) \ge 2}} (r_p(m) - 1) &\ge \sum_{\substack{\frac{p+1}{2} \le m \le N \\ r_p(m) = 0}} 1 - (N - M) \\ &\ge \pi(N) - \pi(p - N - 1) - (N - M), \end{aligned} \tag{2.2}$$

where $\pi(x)$ denotes the number of primes $\le x$. Let

$$K_{D,N}(m) = \left| \{ m = abc : 1 < a < b \le D, (a, b) = 1, \tfrac{b}{a} \le \tfrac{N}{m} \} \right|. \tag{2.3}$$

For any triple (a, b, c) counted in $K_{D,N}(m)$, we have

$$\frac{m}{N - m} \le a \le D - 1, \quad a + 1 \le b \le \frac{N}{m} a, \quad c \le \frac{N}{b^2}. \tag{2.4}$$

In particular, $b \le \sqrt{N}$, so

$$K_{D,N}(m) = K_{\sqrt{N},N}(m) \qquad (D \ge \sqrt{N}). \tag{2.5}$$

Let

$$D(p, A) = \max_{p - N \le m \le N} \max_{(a_i, a_j), (a_{i'}, a_{j'}) \in R_p(m)} \left\{ \frac{\gcd(a_i, a_j)}{\gcd(a_i, a_j, a_{i'}, a_{j'})} \right\}.$$

Lemma 2.1. *If $D = D(p, A)$, then*

$$D = 1 \quad \text{or} \quad \frac{N}{2N - p} \le D \le N,$$

and for $\frac{p+1}{2} \le m \le N$ we have

$$r_p(m) \le (K_{D,N}(m) + 1)(K_{D,N}(p - m) + 1).$$

Proof. This follows from Lemmas 2.3, 2.4 and 2.5 of [1]. □

It follows from Lemma 2.1 and the definition of $D(p, A)$ that $r_p(m) \le 1$ for all m if and only if $D(p, A) = 1$.

The next lemma, a slightly weaker form of Lemma 4.1 of [1], shows that A cannot contain many elements divisible by primes $> 2ND^{-1/3}$.

Lemma 2.2. *Suppose p is a prime in $(1.5N, 2N-\sqrt{N})$ and $D = D(p,A) > 1$. With the possible exception of two primes, no prime $q > 2ND^{-1/3}$ can divide an element of A.*

A version of the Prime Number Theorem with crude error term will also be needed:

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O\left(\frac{x}{\log^{10} x}\right). \tag{2.6}$$

We may now state the fundamental lower bound for $f(N)$. Here $P^+(n)$ denotes the largest prime factor of n.

Theorem 3. *Suppose N is large and let $\mathscr{P}$ be a subset of the primes in $(1.5N, 2N-\sqrt{N})$. Then*

$$\begin{aligned} f(N) \ge -1 + \min\Bigg(&\frac{N}{3(\log N)^{3/2}}, \\ &|\mathscr{P}|^{-1} \min_{\sqrt{\log N} \le D \le \sqrt{N}} \sum_{p\in\mathscr{P}} \{S_1(p,N,D) - S_2(p,N,D)\}\Bigg), \end{aligned} \tag{2.7}$$

where

$$\begin{aligned} S_1(p,N,D) &= \left|\{m \in [p-N, N] : P^+(m) > 2ND^{-1/3}\}\right| - \frac{2N-p}{ND^{-1/3}} - 2, \\ S_2(p,N,D) &= \sum_{\frac{p+1}{2} \le m \le N} ((K_{D,N}(m)+1)(K_{D,N}(p-m)+1) - 1). \end{aligned}$$

Proof. Suppose $|A| = M$, $G(A) \le N$ and neither A nor A^* is contained in $\{1, 2, \ldots, N\}$. Let

$$D_0 = \max_{1.5N < p < 2N-\sqrt{N}} D(p,A).$$

If $D_0 = 1$, let p be the smallest prime $> 1.5N$. By (2.6), $p \le 1.6N$. Since $D(p,A) = 1$, $r_p(m) \le 1$ for $p - N \le m \le N$ and thus by (2.2) and (2.6),

$$N - M \ge \pi(N) - \pi(p-N-1) \ge \frac{N}{3\log N}.$$

If $1 < D_0 \le \sqrt{\log N}$, let p be the smallest prime $> 2N - N/D_0$. By (2.6), $p \le 2N - N/(2D_0)$. By Lemma 2.1, $D(p,A) = 1$ and we similarly obtain from (2.2) and (2.6) the bound

$$N - M \ge \pi(N) - \pi(p-N-1) \ge \frac{N}{3D_0 \log N} \ge \frac{N}{3(\log N)^{3/2}}.$$

Lastly, if $D_0 > \sqrt{\log N}$, then we apply Lemma 2.2. Let $q_1, \ldots, q_s$ be the primes in the interval $(2ND^{-1/3}, N]$. Since $2ND^{-1/3} \ge N^{2/3}$, each number $m \le N$ is divisible by at most one prime q_i. Fix $p \in \mathscr{P}$ and let R_i be the number of $m \in [p-N, N]$ divisible by q_i. Then $r_p(m) = 0$ for at least $S_0(p, N, D_0)$ values of $m \in [\frac{p+1}{2}, N]$, where

$$S_0(p,N,D_0) = \min_{\substack{T \subset \{1,2,\ldots,s\} \\ |T| = s-2}} \sum_{i \in T} R_i.$$

By (2.2), Lemma 2.1 and the fact that $K_{D',N}(m) \le K_{D,N}(m)$ if $D' \le D$,

$$N - M \ge S_0(p,N,D_0) - S_2(p,N,D(p,A)) \ge S_0(p,N,D_0) - S_2(p,N,D_0).$$

Averaging over $p \in \mathscr{P}$ gives

$$N - M \ge |\mathscr{P}|^{-1} \sum_{p \in \mathscr{P}} \big[S_0(p,N,D_0) - S_2(p,N,D_0) \big]. \tag{2.8}$$

Since $S_0(p, N, D)$ is an increasing function of D and $S_2(p, N, D)$ is constant for $N^{1/2} \le D \le N$ by (2.5), the minimum over D_0 of the right side of (2.8) occurs for some $D_0 \le N^{1/2}$. Finally, $S_0(p, N, D) \ge S_1(p, N, D)$ and this completes the proof. □

3. Lower bounds for S_1

Lemma 3.1. *Suppose N is large, $1.98N \le p \le 2N - N/\log^5 N$ and $100 \le D \le \sqrt{N}$. Then*

$$S_1(p,N,D) \ge \frac{(2N-p)\log D}{6 \log N}.$$

Proof. Let $R = \frac{p-N}{2ND^{-1/3}}$ and note that $R \le \frac{1}{2}N^{1/6}$. At most one prime $q > 2ND^{-1/3}$ can divide any number in $[p-N, N]$, so

$$S_1(p,N,D) \ge \sum_{1 \le r \le R} \left(\pi\left(\frac{N}{r}\right) - \pi\left(\frac{p-N}{r}\right)\right) - \frac{2N-p}{N^{5/6}} - 2.$$

By hypothesis, $2 \le 2(2N-p)N^{-5/6}$. Thus, by (2.6),

$$S_1(p,N,D) \ge \sum_{1 \le r \le R} \left(\int_{(p-N)/r}^{N/r} \frac{dt}{\log t} - O\left(\frac{N}{r \log^{10} N}\right)\right) - \frac{3(2N-p)}{N^{5/6}}.$$

The integral is

$$\ge \frac{2N-p}{r \log(N/r)} \ge \frac{2N-p}{r \log N}$$

and

$$\sum_{r \le R} \frac{1}{r} \ge \int_1^{R+1} \frac{dt}{t} \ge \log\left(\frac{0.98}{2} D^{1/3}\right) \ge 0.178 \log D$$

since $D \ge 100$. For large N the result follows. □

4. Upper bounds for S_2

Lemma 4.1. *If* $\max(N/D, 2N^{5/6}) \le \lambda \le N/50$, *then*

$$\sum_{N-\lambda \le m \le N} K_{D,N}(m) \le \frac{\lambda^2}{2(N-\lambda)} \left(\log\left(\frac{\lambda D}{N}\right) + 4\frac{\lambda}{N} + \frac{D^2}{2(N-\lambda)} \right).$$

Remark 1. When $\lambda < N/D$, the left side is zero by (2.4).

Proof. Ignoring the condition $(a,b)=1$ in (2.3), the left side in the lemma is at most the number of triples (a,b,c) with

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

By (2.4),

$$\frac{N-\lambda}{ab} \le c \le \frac{N}{b^2}, \quad a \ge \frac{N-\lambda}{\lambda}, \quad b \le (1+\beta)a, \quad \beta = \frac{\lambda}{N-\lambda}. \tag{4.1}$$

Let E be a parameter in $[\frac{N}{\lambda}, D]$, let T_1 be the number of triples with $a \le E-1$ and T_2 be the number of remaining triples.

We first estimate T_1. For each pair (a,b), the number of c is at most

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

This is a decreasing function of b and is positive for $b \le (1+\beta)a$, so for each a, the number of pairs (b,c) is

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

Thus

$$\begin{aligned}
T_1 &\le \frac{\beta E^2}{2} + \frac{\lambda^2}{2(N-\lambda)} \int_{N/\lambda-2}^{E} \frac{dt}{t} \\
&= \frac{\lambda^2}{2(N-\lambda)} \left(\log\left(\frac{E\lambda}{N}\right) - \log(1-2\lambda/N) + \frac{E^2}{\lambda} \right) \\
&\le \frac{\lambda^2}{2(N-\lambda)} \left(\log\left(\frac{E\lambda}{N}\right) + 2.1\frac{\lambda}{N} + \frac{E^2}{\lambda} \right).
\end{aligned}$$

When $D \le N^{1/3}$, we take $E = D$, so that $T_2 = 0$ and

$$\frac{E^2}{\lambda} \le \frac{N^{2/3}}{\lambda} \le \frac{\lambda}{4N}.$$

This gives the lemma in this case. Next assume $D > N^{1/3}$. To bound T_2, note that

$$\frac{N-\lambda}{D^2} \le c < \frac{N}{E^2}.$$

For fixed c, we count the number of (a,b) for which

$$\frac{N-\lambda}{\sqrt{Nc}} \le a < b \le \frac{N}{\sqrt{Nc}}, \quad b \ge \frac{N-\lambda}{ac}.$$

By symmetry (counting solutions with $b < a$ also), this is

$$\begin{aligned}
&= \frac{1}{2}\left[|\{\tfrac{N-\lambda}{\sqrt{Nc}} \le a \le \tfrac{N}{\sqrt{Nc}}, \tfrac{N-\lambda}{ac} \le b \le \tfrac{N}{\sqrt{Nc}}\}| - |\{\sqrt{\tfrac{N-\lambda}{c}} \le a \le \tfrac{N}{\sqrt{Nc}}\}| \right] \\
&\le \frac{1}{2}\left[\sum_a \left(\frac{N}{\sqrt{Nc}} - \frac{N-\lambda}{ac} + 1 \right) - \left(\frac{N}{\sqrt{Nc}} - \sqrt{\frac{N-\lambda}{c}} - 1 \right) \right] \\
&\le \frac{1}{2}\left[\int_{\frac{N-\lambda}{\sqrt{Nc}}}^{\frac{N}{\sqrt{Nc}}} \frac{N}{\sqrt{Nc}} - \frac{N-\lambda}{ac} + 1 \, da + \frac{\lambda}{\sqrt{Nc}} + 2 - \frac{\sqrt{N} - \sqrt{N-\lambda}}{\sqrt{c}} \right] \\
&\le \frac{1}{2}\left[\frac{\lambda}{c} + \frac{2\lambda}{\sqrt{Nc}} - \frac{N-\lambda}{c} \log \frac{N}{N-\lambda} + 2 - \frac{\lambda}{2\sqrt{Nc}} \right] \\
&\le 1 + \frac{3\lambda}{4\sqrt{Nc}} + \frac{\lambda^2}{4c(N-\lambda)}.
\end{aligned}$$

Next we sum over c, using for $0 < x < y$ the bounds

$$\sum_{x \le c \le y} c^{-1/2} < 2\sqrt{y}, \qquad \sum_{x \le c \le y} c^{-1} \le 1/x + \log(y/x).$$

We conclude that

$$\begin{aligned}
T_2 &\le \frac{N}{E^2} + \frac{3\lambda}{2E} + \frac{\lambda^2}{4(N-\lambda)} \left(\frac{D^2}{N-\lambda} + \log \frac{ND^2}{(N-\lambda)E^2} \right) \\
&\le \frac{\lambda^2}{2(N-\lambda)} \left(\frac{D^2}{2(N-\lambda)} + 0.51\frac{\lambda}{N} + \log \frac{D}{E} + \frac{3N}{E\lambda} + \frac{2N^2}{\lambda^2 E^2} \right).
\end{aligned}$$

Take $E = N^{1/3}$, which is close to optimal. Then combine the bounds for T_1 and T_2, using the bound $\lambda \ge 2N^{5/6}$ to simplify the expression. This completes the proof for $D > N^{1/3}$. □

Lemma 4.2. *Uniformly in $x \ge y \ge 2$ we have*

$$\sum_{y \le n \le x} \frac{1}{\phi(n)} \ll \log \frac{x}{y} + \frac{\log x}{y}.$$

Proof. Start with the identity

$$\frac{n}{\phi(n)} = \sum_{d|n} \frac{\mu^2(d)}{\phi(d)}.$$

Then

$$\begin{aligned}\sum_{y\le n\le x} \frac{1}{\phi(n)} &= \sum_{d\le x} \frac{\mu^2(d)}{d\phi(d)} \sum_{y/d\le m\le x/d} \frac{1}{m} \\ &\le \sum_{d\le x} \frac{\mu^2(d)}{d\phi(d)} \left(\frac{d}{y} + \log\frac{x}{y}\right) \\ &\le \frac{1}{y}\prod_{p\le x}\left(1+\frac{1}{p-1}\right) + \left(\log\frac{x}{y}\right)\sum_{d=1}^{\infty}\frac{\mu^2(d)}{d\phi(d)} \\ &\ll \log\frac{x}{y} + \frac{\log x}{y}.\end{aligned}$$

□

Let $\mathscr{P}_B$ be the set of primes in $[N-2B, N-B]$, where $2N^{5/6} < B \le N/100$. Making the substitution $m \to p-m$ in the definition of S_2, we see that

$$\sum_{p\in\mathscr{P}_B} S_2(p,N,D) = \sum_{\substack{p\in\mathscr{P}_B \\ p-N\le m\le N}} \frac{1}{2}K_{D,N}(m)K_{D,N}(p-m) + K_{D,N}(m). \tag{4.2}$$

Suppose $\frac{N}{2B} \le D \le \sqrt{N}$. By (2.6) and Lemma 4.1,

$$\sum_{\substack{p\in\mathscr{P}_B \\ p-N\le m\le N}} K_{D,N}(m) \ll \frac{B^3}{N\log N}\left(\log\frac{2BD}{N} + \frac{B+D^2}{N}\right). \tag{4.3}$$

Lemma 4.3. *Suppose $2N^{5/6} \le B \le N/100$, $N/(2B) \le D \le \sqrt{N}$ and $N-2B \le m \le N$. Then*

$$\sum_{2N-2B\le p\le N+m} K_{D,N}(p-m) \ll \frac{B^2\log D}{N\log N}.$$

Proof. By (2.4) and (4.1), the left side is at most the number of triples (a,b,c) with $abc+m$ prime and

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

Put $E = \min(D, N^{1/3})$, let T_1 be the number of triples with $a \le E-1$ and T_2 be the number of remaining triples. For fixed $a < b \le E$, the number of c is at most the number of primes in $[2N-2B, 2N]$ which are $\equiv m \pmod{ab}$. This is

$$\ll \frac{B}{\phi(ab)\log(2B/(ab))} \ll \frac{B}{\phi(a)\phi(b)\log N}$$

by the Brun-Titchmarsh inequality and the inequality $\phi(ab) \ge \phi(a)\phi(b)$. By Lemma 4.2,

$$\begin{aligned}\sum_a \frac{1}{\phi(a)} \sum_b \frac{1}{\phi(b)} &\ll \sum_a \frac{1}{\phi(a)} \left(\log(1+\beta) + \frac{\log a}{a} \right) \\ &\ll \frac{B}{N}\left(\log \frac{2BE}{N-2B} + \frac{B \log E}{N} \right) + \sum_{a \ge N/(3B)} \frac{\log a}{a\phi(a)} \\ &\ll \frac{B}{N}\left(\log \frac{2BE}{N} + \log E + \log \frac{N}{B} \right) \\ &\ll \frac{B \log E}{N}.\end{aligned}$$

Therefore,

$$T_1 \ll \frac{B^2 \log E}{N \log N}. \tag{4.4}$$

If $D \le N^{1/3}$, then $T_2 = 0$ and the lemma follows from (4.4). Otherwise we bound T_2 starting with the inequalities

$$\frac{N-2B}{D^2} \le c \le \frac{N}{E^2} = N^{1/3}$$

and

$$\frac{N-2B}{\sqrt{Nc}} \le a < b \le \frac{N}{\sqrt{Nc}}.$$

In particular, $bc \le N^{2/3}$. For fixed b, c the number of a is at most the number of primes in $[2N-2B, 2N]$ which are $\equiv m \pmod{bc}$. By the Brun-Titchmarsh inequality, this is

$$\ll \frac{B}{\phi(bc)\log(\frac{2B}{bc})} \ll \frac{B}{\phi(b)\phi(c)\log N}.$$

By Lemma 4.2 again,

$$\begin{aligned}\sum_{b,c} \frac{1}{\phi(b)\phi(c)} &\ll \sum_c \frac{1}{\phi(c)} \left(\log \frac{N}{N-2B} + \frac{\log N}{E} \right) \\ &\ll \frac{B}{N} \sum_c \frac{1}{\phi(c)} \\ &\ll \frac{B}{N} \left(\log \frac{D}{E} + \frac{B}{N} + \frac{D^2 \log N}{N} \right).\end{aligned}$$

Therefore

$$T_2 \ll \frac{B^2}{N\log N}\left(\log\frac{D}{E} + \frac{B}{N} + \frac{D^2\log N}{N}\right). \tag{4.5}$$

Together, (4.4) and (4.5) give the lemma in the case $D > N^{1/3}$ because $\log D \gg \log N \gg (D^2/N)\log N$. $\square$

If $D < \frac{N}{2B}$, then the left side of (4.2) is zero. Otherwise, putting together (4.2), (4.3) and Lemmas 4.1 and 4.3 gives the following.

Lemma 4.4. *If* $2N^{5/6} \le B \le N/100$ *and* $1 \le D \le \sqrt{N}$, *then*

$$\sum_{p\in\mathscr{P}_B} S_2(p,N,D) \ll \frac{B^4\log^2 D}{N^2\log N} + \frac{B^3\log D}{N\log N}.$$

5. Proof of Theorem 1

Take $B = \frac{c_1 N}{\log N}$, where c_1 is a sufficiently small positive constant, and put $\mathscr{P} = \mathscr{P}_B$. By (2.6), $|\mathscr{P}_B| \gg B/\log N$. Consequently, by Lemma 3.1,

$$\sum_{p\in\mathscr{P}_B} S_1(p,N,D) \gg \frac{B^2\log D}{\log^2 N}.$$

By Theorem 3 and Lemma 4.4, there are absolute constants c_2, c_3, so that when N is large we have

$$f(N) \ge \min\Bigg(\frac{N}{3(\log N)^{3/2}},\\ \min_{\sqrt{\log N}\le D\le\sqrt{N}}\left[\frac{c_2 B\log D}{\log N} - c_3\left(\frac{B^2\log D}{N} + \frac{B^3\log^2 D}{N^2}\right)\right]\Bigg).$$

The minimum of the inner expression occurs at $D = \sqrt{\log N}$ if c_1 is small enough, and this completes the proof.

References

[1] R. Balasubramanian and K. Soundararajan, *On a conjecture of R. L. Graham*, Acta Arith. **75** (1996), 1–38.

[2] K. Ford, *Maximal collections of intersecting arithmetic progressions*, Combinatorica **23** (2) (2003), 263–281.

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

[4] M. Szegedy, *The solution of Graham's greatest common divisor problem*, Combinatorica **6** (1986), 67–71.

[5] A. Zaharescu, *On a conjecture of Graham*, J. Number Theory **27** (1987), 33–40.

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA