ACTA ARITHMETICA
LXXV.1 (1996)

On a conjecture of R. L. Graham

by

R. Balasubramanian (Madras)
and K. Soundararajan (Princeton, N.J.)

Dedicated to Professor K. Ramachandra
on the occasion of his sixtieth birthday

1. Introduction. Let N be an integer and $\mathbf{A} = \{a_1, \ldots, a_N\}$, $a_1 < \ldots < a_N$, be a set of N integers. A well-known conjecture of Graham [5] states that there exist $a_i, a_j \in \mathbf{A}$ with $a_i/(a_i, a_j) \ge N$.

If we let $M = \mathrm{lcm}[a_1, \ldots, a_N]$ and put $\mathbf{A}^* = \{M/a_1, \ldots, M/a_N\}$ then it is an easily verified observation of Winterle [11] that

$$\frac{M/a_i}{(M/a_i, M/a_j)} = \frac{a_j}{(a_i, a_j)}.$$

We call $\mathbf{A}^*$ the *reciprocal set* of $\mathbf{A}$. It is clear from the above remark that the set $\{1, \ldots, N\}$ and its reciprocal set form extremal examples to the above conjecture. For the case $N = 4$ we have a third extremal set $\{2, 3, 4, 6\}$.

Thus a stronger version of Graham's conjecture is that if $\mathbf{A} = \{a_1, \ldots, a_N\}$ is such that $(a_1, \ldots, a_N) = 1$ and for all i, j, $a_i/(a_i, a_j) \le N$ then either one of $\mathbf{A}$ and $\mathbf{A}^*$ is the set $\{1, \ldots, N\}$ or $N = 4$ and $\mathbf{A} = \{2, 3, 4, 6\}$.

The conjecture in its weaker form was proved in a variety of special cases. For example Winterle [11] showed the conjecture in the case where a_1 is a prime. Vélez [10] established the conjecture for $N = p + 1$ (p a prime) and also gave a proof (due to Szemerédi) for $N = p$. Boyle [1] extended these results to establish the conjecture when $N = p, p + 1, p + 2, p + 3, p^2$. For a more complete account of the history of the problem (until 1980) see pages 78 and 79 of Erdős and Graham [4]. Significant progress was made towards the conjecture in its weaker form by Szegedy [9] and Zaharescu [12], who independently established it for all large N. In a later paper, Cobeli, Vâjâitu and Zaharescu [3] established the weaker Graham conjecture for all $N \ge 10^{70}$ under the assumption of the Riemann Hypothesis. They also commented that their method could be pushed to yield $N \ge 10^{60}$ but would not yield

$N \geq 10^{50}$. Recently, F. Y. Cheng and C. Pomerance [2] have shown the stronger Graham conjecture when $N > 10^{50\,000}$. The purpose of this paper is to establish Graham's conjecture, in its stronger form, unconditionally. Since the conjecture is trivial for $N \leq 4$ we restrict our attention to $N \geq 5$.

THEOREM 1.1. *Let $N \geq 5$ be an integer and $\mathbf{A} = \{a_1, \ldots, a_N\}$, where $a_1 < \ldots < a_N$ and $(a_1, \ldots, a_N) = 1$, be a set of N integers. Then there exist $a_i, a_j \in \mathbf{A}$ with*

$$a_i/(a_i, a_j) \geq N.$$

The inequality is strict if both $\mathbf{A}$ and $\mathbf{A}^*$ are different from $\{1, \ldots, N\}$.

Let p be a prime "close" to $2N$ and α an integer in $[(p+1)/2, N]$. The starting point of our investigations is the function

$$r_p(\alpha) = |\{d : \alpha d, (p-\alpha)d \in \mathbf{A}\}|.$$

This function is tacitly present in the work of both Szegedy and Zaharescu. The motivation for considering $r_p(\alpha)$ lies in the fact that if either $\mathbf{A}$ or $\mathbf{A}^*$ is $\{1, \ldots, N\}$ then $r_p(\alpha) = 1$ for all α. On the other hand, if $\mathbf{A}$ is a set not satisfying Graham's conjecture and neither $\mathbf{A}$ nor $\mathbf{A}^*$ is $\{1, \ldots, N\}$ then one can find a p and an α with $r_p(\alpha) \neq 1$. Thus there is the hope that producing α's with $r_p(\alpha) \neq 1$ would lead to a contradiction and thereby prove the conjecture. Indeed, for a set $\mathbf{A}$ not satisfying the weaker Graham conjecture, Zaharescu shows the existence of α with $r_p(\alpha) \geq 2$. Then he exhibits "lots" of β's such that $(\alpha, p-\beta) = (p-\alpha, \beta) = 1$ and $r_p(\beta) \geq 1$. This leads to a contradiction provided p is close enough to $2N$. Here he utilises a well-known result of Huxley that for every $\varepsilon > 0$ and large x there is a prime in the interval $(x, x + x^{7/12+\varepsilon})$. While Szegedy exhibits lots of α's with $r_p(\alpha) \geq 2$, he too requires results of the type

$$\pi(x + x^{9/14}) - \pi(x) \gg x^{9/14}/\log x.$$

By the nature of these tools the bound on N is extremely weak. Indeed one needs $N \geq e^{10^6}$, say, to ensure the validity of the results of Szegedy and Zaharescu. Since the Riemann Hypothesis implies the existence of primes in intervals as short as $(x, x + cx^{1/2}\log x)$ for a reasonably small constant c, Cobeli, Vâjâitu and Zaharescu were able to refine Zaharescu's argument to establish the weaker conjecture (conditionally) for $N \geq 10^{70}$.

From computer calculations on gaps between prime numbers (see [7]), we will establish (in §3) Theorem 1.1 for all $N \leq 2.22 \cdot 10^{12}$. If $N \geq 2.22 \cdot 10^{12}$, we consider the quantity

$$Q = \sum_{p \in [2N-2G(N), 2N-G(N)]} \sum_{\substack{\alpha \in [(p+1)/2, N] \\ r_p(\alpha) \geq 2}} (r_p(\alpha) - 1),$$

where $G(N)$ will be fixed later (in §6). In Section 4 we will derive a lower bound for Q. This is achieved by exhibiting many α for which $r_p(\alpha) = 0$. We then bound $r_p(\alpha)$ in terms of a more tractable function $k_D(n)$ (see Lemma 2.5). Applying the Montgomery–Vaughan version of the Brun–Titchmarsh theorem (see [6]) we obtain (in §5) an upper bound for Q. The upper and lower bounds are shown to yield a contradiction if (roughly) there are "lots" of primes in intervals of the form $(x, x + G(x))$. If $N \le e^{2000}$ then we will choose $G(N) \asymp N/\log N$ and for larger N, $G(N) \asymp N/\log^2 N$. The results of Rosser and Schoenfeld [8] are now sufficient for our purposes.

A modification of our arguments can be used to establish the more general result: if $\mathbf{A}$ and $\mathbf{B}$ are two N element sets then there exists $a \in \mathbf{A}$ and $b \in \mathbf{B}$ with $\max(a/(a,b), b/(a,b)) \ge N$. The inequality is strict unless $\mathbf{A} = \mathbf{B} = \{1, \ldots, N\}$ or $\mathbf{A} = \mathbf{B} = \{1, \ldots, N\}^*$. This may be proved by considering the function

$$\widehat{r}_p(\alpha) = \{d : \alpha d \in \mathbf{A},\ (p - \alpha)d \in \mathbf{B}\}.$$

The proof of Theorem 1.1 goes through *mutatis mutandis*.

Our methods will also show that there exists a positive constant c such that if neither $\mathbf{A}$ nor $\mathbf{A}^*$ is contained in $[1, N + cN/(\log N \log\log N)]$ then there exist $a_i, a_j \in \mathbf{A}$ with

$$\frac{a_i}{(a_i, a_j)} \ge N + \frac{cN}{\log N \log\log N}.$$

We are grateful to the referee for his valuable comments which have improved our exposition. The referee also drew our attention to [2] and corrected an oversight in Lemma 4.2. One of us (K. S.) is grateful to Professor T. D. Wooley for his help and encouragement.

2. Preliminary lemmata. Throughout this paper bold-face, uppercase letters (e.g. $\mathbf{U}$) will denote sets of integers. For a set $\mathbf{U}$ the cardinality of $\mathbf{U}$ will be denoted by $|\mathbf{U}|$. The expression (a, b) can mean either $\gcd(a, b)$ or an element of $\mathbb{N}^2$ or the open interval (a, b). The intended meaning will usually be clear from the context. In cases of possible ambiguity we have indicated our meaning explicitly. The same holds for the symbols $[a, b]$.

In the sequel $\mathbf{A} = \{a_1, \ldots, a_N\}$ is a set of N integers with $(a_1, \ldots, a_N) = 1$, $a_1 < \ldots < a_N$, and for all i, j, $a_i/(a_i, a_j) \le N$. Further, we also assume that neither $\mathbf{A}$ nor $\mathbf{A}^*$ is the set $\{1, \ldots, N\}$. $G(N)$ will denote a function (to be chosen later) satisfying $0 \le G(N) \le (1 - 1/\sqrt{2})N$. Let $\mathbf{P}$ denote the set of primes in the interval $[2N - 2G(N), 2N - G(N)]$. Let p denote a generic prime in $[2N - 2G(N), 2N]$. Let $\mathbf{J}_p = \{(p + 1)/2, \ldots, N\}$ and for $\alpha \in \mathbf{J}_p$ define

$$r_p(\alpha) = |\{d : \alpha d, (p - \alpha)d \in \mathbf{A}\}|.$$

Lemma 2.1. *With notations as above*,

$$\sum_{\alpha\in\mathbf{J}_p} r_p(\alpha) \ge |\mathbf{J}_p|.$$

Thus

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 1}} (r_p(\alpha)-1) \ge \sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1.$$

Proof. Since $p > N$ and $(a_1,\dots,a_N)=1$, we see that p cannot divide any element of $\mathbf{A}$. Further if $i\ne j$ then $a_i \not\equiv a_j \pmod p$ since otherwise we would have

$$\max\left(\frac{a_i}{(a_i,a_j)},\frac{a_j}{(a_i,a_j)}\right) \ge \frac{|a_i-a_j|}{(a_i,a_j)} \ge p > N.$$

The numbers a_i^2 $(i=1,\dots,N)$ lie in at most $(p-1)/2$ residue classes $\bmod p$ and no residue class contains more than two values of a_i^2. Hence

$$|\{(i,j): i<j,\ a_i^2\equiv a_j^2 \pmod p\}| \ge N-\frac{p-1}{2} = |\mathbf{J}_p|.$$

By our previous remark this implies that

$$|\{(i,j): i<j,\ a_i\equiv -a_j \pmod p\}| \ge |\mathbf{J}_p|.$$

If $i<j$ and $a_i\equiv -a_j \pmod p$ then for some k,

$$2N \ge \frac{a_i+a_j}{(a_i,a_j)} = pk.$$

Since $p>N$ it follows that $k=1$ and so $a_i+a_j=p(a_i,a_j)$. Writing $a_i=(p-\alpha)(a_i,a_j)$, $a_j=\alpha(a_i,a_j)$ we see that since $a_i<a_j$, α is in $\mathbf{J}_p$. Thus this would be a contributor to $\sum_{\alpha\in\mathbf{J}_p} r_p(\alpha)$. The proof follows at once.

Lemma 2.2. *If $q>N/2$ is prime then $q\nmid a_i$ for any $a_i\in\mathbf{A}$.*

Proof. This is essentially Theorem 1 of Boyle [1]. While Boyle is concerned with only the weaker conjecture of Graham, his proof requires only the obvious modification of replacing $N-1$'s by N's to yield the lemma. An almost identical proof may also be found in Szegedy [9].

From Lemmata 2.1 and 2.2 we see that

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 1}} (r_p(\alpha)-1) \ge \sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \sum_{\substack{\alpha\in\mathbf{J}_p\\ \alpha\text{ prime}}} 1 + \sum_{\substack{\alpha\in\mathbf{J}_p\\ p-\alpha\text{ prime}}} 1$$
$$= \pi(N)-\pi(p-N-1).$$

So if there are primes in $[p-N,N]$ then we are assured of the existence of $\alpha\in\mathbf{J}_p$ with $r_p(\alpha)\ge 2$.

Suppose $\alpha \in \mathbf{J}_p$ and $r_p(\alpha) \geq 2$. Thus there exist integers d_1, d_2, A with $(d_1, d_2) = 1$, $d_1 \neq d_2$ such that

$$\alpha d_1 A, (p-\alpha) d_1 A, \alpha d_2 A, (p-\alpha) d_2 A \in \mathbf{A}.$$

The next two lemmata shed some light on this situation.

LEMMA 2.3. *We may write*

$$d_i = X_i Y_i \qquad (i = 1, 2),$$

where $X_i = (d_i, \alpha)$ *and* $Y_i = (d_i, p-\alpha)$. *Further*,

$$\max\left(\frac{X_1}{X_2}, \frac{X_2}{X_1}\right) \leq \frac{N}{\alpha}$$

and similarly

$$\max\left(\frac{Y_1}{Y_2}, \frac{Y_2}{Y_1}\right) \leq \frac{N}{p-\alpha},$$

whence

$$\max\left(\frac{d_1}{d_2}, \frac{d_2}{d_1}\right) \leq \frac{N^2}{\alpha(p-\alpha)}.$$

Consequently,

$$\frac{N-\alpha}{\alpha} \min(X_1, X_2) \geq |X_1 - X_2|$$

and

$$\frac{N+\alpha-p}{p-\alpha} \min(Y_1, Y_2) \geq |Y_1 - Y_2|.$$

Finally, *it is not possible to have* $X_1 = X_2 = Y_1 = Y_2 = 1$.

Proof. Clearly $X_i Y_i \mid d_i$. Suppose $d_i = X_i Y_i Z_i$. Now

$$\frac{\alpha d_1 A}{(\alpha d_1 A, (p-\alpha) d_2 A)} = \frac{\alpha d_1}{(\alpha d_1, (p-\alpha) d_2)} = \frac{\alpha d_1}{(\alpha, d_2)(p-\alpha, d_1)}$$
$$= \frac{\alpha X_1 Z_1}{X_2} \leq N.$$

Similarly $\alpha X_2 Z_2 / X_1 \leq N$ whence $\alpha^2 Z_1 Z_2 \leq N^2$. Since $\alpha > p/2 > N/\sqrt{2}$ this implies $Z_1 = Z_2 = 1$. This proves our first assertion; the remaining statements follow similarly or are trivial.

LEMMA 2.4. *Let* $r_p(\alpha) = s \geq 2$ *and let* $\alpha d_i A, (p-\alpha) d_i A$ $(i = 1, \ldots, s)$ *be elements of* $\mathbf{A}$ *with* $(d_1, \ldots, d_s) = 1$. *Then for all* $1 \leq i \leq s$, $d_i \mid \alpha(p-\alpha)$.

Proof. The case $s = 2$ is a consequence of Lemma 2.3. Suppose $q^t \parallel d_i$ where q is a prime and $t \geq 1$. There exists $1 \leq j \leq s$ with $(d_j, q) = 1$. We argue with the elements

$$\alpha d_i^* AB, \quad (p-\alpha) d_i^* AB, \quad \alpha d_j^* AB, \quad (p-\alpha) d_j^* AB,$$

where $B = (d_i, d_j)$ and $d_i^* = d_i/B$, $d_j^* = d_j/B$. The proof of Lemma 2.3 shows $d_i^* \mid \alpha(p-\alpha)$. Clearly $q^t \parallel d_i^*$ as $(B, q) = (d_j, q) = 1$. Hence $q^t \mid \alpha(p-\alpha)$. Since this holds for all d_i and all primes q the lemma is proved.

Define the set $\mathbf{S} \subset \mathbb{N}^2$ by

$$\mathbf{S} = \{(d_1, d_2) : \gcd(d_1, d_2) = 1,\ \exists p \in [2N - 2G(N), 2N - G(N)], \alpha \in \mathbf{J}_p \text{ such that for some integer } A,\ \alpha d_i A, (p-\alpha) d_i A \in \mathbf{A}\ (i = 1, 2)\}.$$

Let D be the least integer with the property

$$\mathbf{S} \cap [1, D]^2 = \mathbf{S}.$$

Let $(d_1, d_2) \in \mathbf{S}$ and p, α, A have their natural meanings. Suppose $d_1 < d_2$. Then, by Lemma 2.3, $d_2 \le d_1 N^2/(\alpha(p-\alpha))$. Also $d_1 d_2 \mid \alpha(p-\alpha)$. So

$$d_2^2 \le \frac{d_1 d_2 N^2}{\alpha(p-\alpha)} \le N^2.$$

It follows that $D \le N$.

For $n \in [N - 2G(N), N]$, let

$$k_D(n) = |\{(\lambda, X_1, X_2) : n = \lambda X_1 X_2,\ 1 < X_1 < X_2 \le D,\ X_2/X_1 \le N/n\}|.$$

We will now obtain a bound for $r_p(\alpha)$ in terms of $k_D(\alpha)$ and $k_D(p-\alpha)$. This will enable us later to obtain upper bounds for the average value of $r_p(\alpha)$.

Lemma 2.5. *With the above notations, for all primes* $p \in [2N - 2G(N), 2N - G(N)]$ *and* $\alpha \in \mathbf{J}_p$,

$$r_p(\alpha) \le (k_D(\alpha) + 1)(k_D(p-\alpha) + 1).$$

Proof. Suppose $r_p(\alpha) = s$ and that $\alpha d_i A$, $(p-\alpha) d_i A$ $(i = 1, \ldots, s)$ are elements of $\mathbf{A}$ with $(d_1, \ldots, d_s) = 1$. From Lemma 2.4 we may write $d_i = u_i v_i$, where $u_i = (d_i, \alpha)$ and $v_i = (d_i, p-\alpha)$. We split the s integers d_i into k sets T_j $(j = 1, \ldots, k)$ such that if $d_l, d_m \in T_j$ then $u_l = u_m$ and if $d_l \in T_j$, $d_m \in T_w$ $(j < w)$ then $u_l < u_m$. It is of course permissible for the T_j's to be singletons.

Consider a generic set T_j. Suppose $T_j = \{d_{j1}, \ldots, d_{jx}\}$, where $d_{j1} < \ldots < d_{jx}$. If $2 \le y \le x$, we may write

$$p - \alpha = \mu \frac{v_{j1}}{(v_{j1}, v_{jy})} \cdot \frac{v_{jy}}{(v_{j1}, v_{jy})}$$

for some integer μ. Appealing to Lemma 2.3 we see that $v_{jy}/v_{j1} \le N/(p-\alpha)$ and we also know that $v_{jy}/(v_{j1}, v_{jy}) \le D$ so that the above would be one of the solutions counted in $k_D(p-\alpha)$. Thus $x - 1 \le k_D(p-\alpha)$ or $|T_j| \le k_D(p-\alpha) + 1$.

It remains to show that the number k of sets T_j is less than $k_D(\alpha) + 1$. For this, we choose a representative from each set T_j. Since now the "u

values" are distinct the earlier argument gives us solutions to $\alpha = \lambda X_1 X_2$. The proof follows at once.

We now state a version of the Brun–Titchmarsh theorem (due to Montgomery and Vaughan) which will be used in Section 5.

LEMMA 2.6. *Let x, y be positive real numbers and let k, l be coprime integers. Let $\pi(x; k, l)$ be the number of primes $p \le x$ with $p \equiv l \pmod{k}$. Then, if $y > k$,*

$$\pi(x+y; k, l) - \pi(x; k, l) \le \frac{2y}{\phi(k)\log(y/k)}.$$

Proof. This is Theorem 2 of Montgomery and Vaughan [6].

Finally, we recall some elementary inequalities which will prove useful later.

LEMMA 2.7. *Let $2 \le y < x$ be real numbers. Then*

$$\sum_{y<n\le x} \frac{1}{n} \le \log \frac{[x]}{[y]}$$

and

$$\sum_{1\le n\le x} \frac{1}{n} \ge \log\left(\frac{e([x]+1)}{2}\right).$$

Let $\Lambda(n)$ denote, as usual, the von Mangoldt function and let $\psi(t) = \sum_{n\le t} \Lambda(n)$. If $t \ge 100$, then

$$\psi(t) \le 1.1t \quad \textit{and} \quad \pi(t) \le 1.1\frac{t}{\log t}.$$

Finally, if $y \ge 1000$,

$$\sum_{y<p\le x} \frac{1}{p} \ge \frac{\log(x/y)}{\log x} - \frac{1.2}{\log x}.$$

Proof. Clearly,

$$\sum_{y<n\le x} \frac{1}{n} \le \sum_{y<n\le x} \int\limits_{n-1}^{n} \frac{dt}{t} \le \int\limits_{[y]}^{[x]} \frac{dt}{t} = \log \frac{[x]}{[y]}.$$

Similarly,

$$\sum_{n\le x} \frac{1}{n} = 1 + \sum_{2\le n\le x} \frac{1}{n} \ge 1 + \sum_{2\le n\le x} \int\limits_{n}^{n+1} \frac{dt}{t} = \log\left(\frac{e([x]+1)}{2}\right).$$

The inequalities for $\psi(t)$ and $\pi(t)$ follow from Theorem 6 of Rosser and Schoenfeld [8].

From the definition of the von Mangoldt function, we see that $\log n = \sum_{d|n} \Lambda(d)$. Now

$$\log([x]!) = \sum_{n\le[x]} \log n = \sum_{n\le[x]} \sum_{d|n} \Lambda(d) = \sum_{d\le[x]} \Lambda(d) \left[\frac{[x]}{d}\right] \le [x] \sum_{d\le x} \frac{\Lambda(d)}{d}.$$

Also,

$$\log([x]!) = \sum_{2\le n\le[x]} \log n \ge \sum_{2\le n\le[x]} \int_{n-1}^{n} \log t\,dt = [x]\log[x] - [x] + 1$$

and so

$$\sum_{n\le x} \frac{\Lambda(n)}{n} \ge \log[x] - 1 + \frac{1}{[x]} \ge \log x - 1.$$

Next,

$$\log([y]!) = \sum_{n\le[y]} \Lambda(n) \left[\frac{y}{n}\right] \ge y \sum_{n\le[y]} \frac{\Lambda(n)}{n} - \psi(y).$$

Again

$$\log([y]!) = \sum_{n\le[y]} \log n \le \sum_{n\le[y]} \int_{n}^{n+1} \log t\,dt = [y+1]\log[y+1] - [y]$$

and so

$$\sum_{n\le y} \frac{\Lambda(n)}{n} \le \frac{[y+1]}{y} \log[y+1] - \frac{[y]}{y} + \frac{\psi(y)}{y} \le \log y + 0.12,$$

since $y \ge 1000$ and $\psi(y) \le 1.1y$. Thus

$$\sum_{y<n\le x} \frac{\Lambda(n)}{n} \ge \log\frac{x}{y} - 1.12.$$

Observe that

$$\sum_{y<n\le x} \frac{\Lambda(n)}{n} = \sum_{y<p\le x} \frac{\log p}{p} + \sum_{\substack{p,m\ge2\\ y<p^m\le x}} \frac{\log p}{p^m} \le \log x \sum_{y<p\le x} \frac{1}{p} + \sum_{\substack{p,m\ge2\\ p^m\ge100}} \frac{\log p}{p^m}.$$

An easy calculation shows that

$$\sum_{\substack{p,m\ge2\\ p^m\ge1000}} \frac{\log p}{p^m} \le 0.08.$$

Thus

$$\log x \sum_{y<p\le x} \frac{1}{p} \ge \log\frac{x}{y} - 1.12 - 0.08 = \log\frac{x}{y} - 1.2.$$

The proof follows at once.

LEMMA 2.8. *Let $l \mid 6$ and x, y be real numbers with $1000 \le y \le x$. Let*

$$C = \sum_{n=1}^{\infty} \frac{\mu(n)^2}{n\phi(n)} = 1.94\ldots$$

Let $h(1) = 1$ and for $l > 1$ let $h(l) = \prod_{p\mid l}(p-1)^2/(p^2-p+1)$. Then

$$\sum_{\substack{n\le x\\(n,l)=1}} \frac{1}{\phi(n)} \le Ch(l)\log\frac{ex}{l}$$

and

$$\sum_{\substack{y<n\le x\\(n,l)=1}} \frac{1}{\phi(n)} \le Ch(l)\frac{(x-y+\phi(l)\log(ex/l))}{y}.$$

Proof. Observe that $n/\phi(n) = \sum_{d\mid n}\mu(d)^2/\phi(d)$ and so, using Lemma 2.7,

$$\begin{aligned}
\sum_{\substack{n\le x\\(n,l)=1}} \frac{1}{\phi(n)} &= \sum_{\substack{n\le x\\(n,l)=1}} \frac{1}{n}\sum_{d\mid n}\frac{\mu(d)^2}{\phi(d)} = \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{\phi(d)} \sum_{\substack{n\le x,\, d\mid n\\(n,l)=1}} \frac{1}{n}\\
&= \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)} \sum_{\substack{n\le x/d\\(n,l)=1}} \frac{1}{n} \le \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)} \sum_{k\le x/dl} \sum_{\substack{r=1\\(r,l)=1}}^{l-1} \frac{1}{kl+r}\\
&\le \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)}\cdot\frac{\phi(l)}{l}\log\frac{ex}{dl} \le \frac{\phi(l)}{l}\log\frac{ex}{l} \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)}\\
&\le \frac{\phi(l)}{l}\left(\log\frac{ex}{l}\right) C\prod_{p\mid l}\left(1+\frac{1}{p(p-1)}\right)^{-1}\\
&\le C\prod_{p\mid l}\frac{(p-1)^2}{p^2-p+1}\log\frac{ex}{l}.
\end{aligned}$$

Similarly, using the result just proved,

$$\begin{aligned}
\sum_{\substack{y<n\le x\\(n,l)=1}} \frac{1}{\phi(n)} &\le \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)} \sum_{\substack{y/d<n\le x/d\\(n,l)=1}} \frac{1}{n}\\
&\le \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)}\cdot\frac{d}{y} \sum_{\substack{y/d<n\le x/d\\(n,l)=1}} 1
\end{aligned}$$

$$\le \sum_{\substack{d\le x\\(d,l)=1}} \frac{\mu(d)^2}{d\phi(d)}\cdot\frac{d}{y}\left(\frac{\phi(l)}{l}\cdot\frac{x-y}{d}+\phi(l)\right)$$

$$\le \frac{\phi(l)}{l}\cdot\frac{x-y}{y}\sum_{\substack{d\le x\\(d,l)=1}}\frac{\mu(d)^2}{d\phi(d)}+\frac{\phi(l)}{y}\sum_{\substack{d\le x\\(d,l)=1}}\frac{1}{\phi(d)}$$

$$\le C\prod_{p\mid l}\frac{(p-1)^2}{p^2-p+1}\cdot\frac{(x-y+\phi(l)\log(ex/l))}{y}.$$

3. Graham's conjecture for $N \le 2.22\cdot 10^{12}$

LEMMA 3.1. *If $N\ge 10$ and $2N\ge p\ge 2N-2\sqrt{N}$ then $r_p(\alpha)\le 2$ for all $\alpha\in \mathbf{J}_p$.*

Proof. Suppose $r_p(\alpha)\ge 3$. Then there exist d_1,d_2,d_3,A with $d_1<d_2<d_3$ and $(d_1,d_2,d_3)=1$ such that $\alpha d_iA,(p-\alpha)d_iA\in\mathbf{A}$. By Lemma 2.4 we may write $d_i=X_iY_i$, where $X_i=(d_i,\alpha)$, $Y_i=(d_i,p-\alpha)$.

By Lemma 2.3,

$$\frac{N-\alpha}{N}\min\left(\frac{X_1}{(X_1,X_2)},\frac{X_2}{(X_1,X_2)}\right)\ge\frac{|X_1-X_2|}{(X_1,X_2)}.$$

Since X_1 and X_2 divide α, $X_1X_2/(X_1,X_2)^2\mid\alpha$ and so $X_1X_2/(X_1,X_2)^2\le\alpha\le N$. Thus

$$\left|\frac{X_1}{(X_1,X_2)}-\frac{X_2}{(X_1,X_2)}\right|^2\le\frac{(N-\alpha)^2}{N^2}\cdot\frac{X_1X_2}{(X_1,X_2)^2}\le\frac{(N-\alpha)^2}{N}<1.$$

It follows that

$$\frac{X_1}{(X_1,X_2)}=\frac{X_2}{(X_1,X_2)}=1$$

and similarly that

$$\frac{X_1}{(X_1,X_3)}=\frac{X_3}{(X_1,X_3)}=1.$$

Hence $(X_1,X_2)=(X_1,X_3)=((X_1,X_2),(X_1,X_3))=(X_1,X_2,X_3)=1$ and so $X_1=X_2=X_3=1$.

Since $d_1<d_2<d_3$ we must have $Y_1<Y_2<Y_3$. By Lemma 2.3,

$$\frac{N+\alpha-p}{p-\alpha}\min\left(\frac{Y_1}{(Y_1,Y_2)},\frac{Y_2}{(Y_1,Y_2)}\right)\ge\frac{Y_2-Y_1}{(Y_1,Y_2)}.$$

Since Y_1 and Y_2 divide $p-\alpha$, $Y_1Y_2/(Y_1,Y_2)^2\mid p-\alpha$ and so $Y_1Y_2/(Y_1,Y_2)^2$

$\le p-\alpha$. Hence

$$\begin{aligned} p-\alpha &\ge \frac{Y_1Y_2}{(Y_1,Y_2)^2} \\ &\ge \frac{(p-\alpha)(Y_2-Y_1)}{(N+\alpha-p)(Y_1,Y_2)}\left(\frac{Y_2-Y_1}{(Y_1,Y_2)}+\frac{(p-\alpha)(Y_2-Y_1)}{(N+\alpha-p)(Y_1,Y_2)}\right) \\ &\ge \frac{(p-\alpha)N(Y_2-Y_1)^2}{(N+\alpha-p)^2(Y_1,Y_2)^2} > \frac{p-\alpha}{4}\cdot\frac{(Y_2-Y_1)^2}{(Y_1,Y_2)^2} \end{aligned}$$

and so $Y_2 = Y_1 + (Y_1, Y_2)$. Arguing similarly we obtain

$$Y_3 = Y_1 + (Y_1, Y_3), \qquad Y_3 = Y_2 + (Y_2, Y_3)$$

so that $(Y_1, Y_3) = (Y_1, Y_2) + (Y_2, Y_3)$. Further (again from Lemma 2.3),

$$\begin{aligned} Y_1 &\ge (Y_1,Y_3)\frac{p-\alpha}{N+\alpha-p}, \\ Y_3 = Y_1 + (Y_1,Y_3) &\ge (Y_1,Y_3)\frac{N}{N+\alpha-p} \end{aligned}$$

and

$$Y_2 = Y_3 - (Y_2,Y_3) \ge (Y_1,Y_3)\frac{N}{N+\alpha-p} - (Y_2,Y_3).$$

Since Y_1, Y_2, Y_3 are divisors of $p-\alpha$ we see, by using the above inequalities, that

$$\begin{aligned} p-\alpha \ge \operatorname{lcm}[Y_1,Y_2,Y_3] &= \frac{Y_1Y_2Y_3}{(Y_1,Y_2)(Y_1,Y_3)(Y_2,Y_3)} \\ &\ge \frac{p-\alpha}{N-p+\alpha}\left(\frac{(Y_1,Y_3)N}{(Y_2,Y_3)(N-p+\alpha)}-1\right)\frac{N}{N-p+\alpha}\cdot\frac{(Y_1,Y_3)}{(Y_1,Y_2)}. \end{aligned}$$

Consequently,

$$4 \ge \frac{(N-p+\alpha)^2}{N} \ge \left(\frac{(Y_1,Y_3)N}{(Y_2,Y_3)(N-p+\alpha)}-1\right)\frac{(Y_1,Y_3)}{(Y_1,Y_2)}.$$

If we put $x = (Y_2,Y_3)/(Y_3,Y_1) \in (0,1)$ then we must have $4(1-x) \ge (\sqrt{N}/(2x)) - 1$ or $2x(5-4x) \ge \sqrt{N}$. This is false since $2x(5-4x) \le 25/8 \le \sqrt{10}$.

LEMMA 3.2. *Theorem* 1.1 *holds in the range* $7000 \le N \le 2.22\cdot 10^{12}$.

Proof. An inspection of the table on page 85 of [7] reveals the existence of a prime in the interval $(x, x+\sqrt{x}/(1+\sqrt{2}))$ for $6900 \le x \le 4.44\cdot 10^{12}$. Thus for our range of N we can find p in $[2N-\sqrt{N}, 2N-\sqrt{N}/(1+\sqrt{2})]$ with $\pi(N)-\pi(p-N)\ge 1$. Thus from Section 2 there exists $\alpha\in\mathbf{J}_p$ with $r_p(\alpha)\ge 2$.

From Lemma 2.3 it follows that either $X_1X_2 \ge \alpha^2/(N-\alpha)^2$ or $Y_1Y_2 \ge (p-\alpha)^2/(N+\alpha-p)^2$, where X_i, Y_i have their usual meanings. Since $X_1X_2 \mid \alpha$

and $Y_1Y_2 \mid (p-\alpha)$ we must have either $\alpha \le (N-\alpha)^2$ or $p-\alpha \le (N+\alpha-p)^2$. Neither of the above can hold. This completes the proof.

LEMMA 3.3. *If for any* $j = 0, 1, 2$ *or* 3 *there exists a prime* p *in* $[2N - \sqrt{(j+1)N}, 2N - \sqrt{jN}]$ *such that* $\pi(N) - \pi(p-N-1) \ge r+1$, *where*

$$r = |\{n : n = \lambda Y(Y+1), 1 \le \lambda \le j, 1 < Y, p-N \le n \le (p+1)/2\}|,$$

then Theorem 1.1 *holds for this value of* N. *Consequently*, *Theorem* 1.1 *is valid in the range* $10 \le N \le 7000$ *with the possible exceptions* $N = 27$, 65.

Proof. Suppose $p \in [2N - \sqrt{(j+1)N}, 2N - \sqrt{jN}\,]$ and $\alpha \in \mathbf{J}_p$ is such that $r_p(\alpha) \ge 2$. Then arguing as in Lemma 3.2 we see that $p-\alpha = \lambda Y(Y+1)$ for some $1 \le \lambda \le j$ and Y an integer. Thus the number of integers α with $r_p(\alpha) \ge 2$ is at most r. On the other hand, from Lemmata 2.1 and 2.2,

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) \ge \pi(N) - \pi(p-N-1) \ge r+1.$$

By Lemma 3.1,

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) = \sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)= 2}} 1.$$

This is a contradiction. An easy computer calculation verifies the truth of our second assertion.

LEMMA 3.4. *Theorem* 1.1 *holds for* $N = 5$, 6, 7, 8 *and* 9.

Proof. By Lemma 2.2 we may assume that all elements of $\mathbf{A}$ are composed of the primes 2 and 3. If $2^j3^k \in \mathbf{A}$ then since there exists $a \in \mathbf{A}$ with $(a, 2) = 1$ it follows that $0 \le j \le \log N/\log 2$ and similarly $0 \le k \le \log N/\log 3$.

Thus if $N = 5, 6$ or 7 then $\mathbf{A} \subseteq \{1, 2, 3, 4, 6, 12\}$. Clearly at most one of 1 and 12 can be in $\mathbf{A}$. This establishes the cases $N = 6, 7$. If $N = 5$ then observe that $1 \in \mathbf{A}$ implies $6, 12 \notin \mathbf{A}$ and $12 \in \mathbf{A}$ implies $1, 2 \notin \mathbf{A}$. Thus neither 1 nor 12 can be in $\mathbf{A}$, which is also a contradiction.

If $N = 8$ then $\mathbf{A} \subseteq \{1, 2, 3, 4, 6, 8, 12, 24\}$. Since at most one of 1 and 24 can be in $\mathbf{A}$, Theorem 1.1 holds in this case.

Finally, if $N = 9$ then $\mathbf{A} \subset \{1, 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, 72\} = \mathbf{D}$, say. Let j denote the least element of $\mathbf{A}$. Clearly, no integer exceeding $10j$ can be in $\mathbf{A}$. Thus the set

$$\{1, \ldots, j-1\} \cup (\mathbf{D} \cap \{n : n \in \mathbb{N},\ n \ge 10j\})$$

is disjoint from $\mathbf{A}$. If $j \le 4$, it is easily verified that $j-1+|\mathbf{D}\cap\{n \ge 10j\}| \ge 4$ and so $|\mathbf{A}| \le |\mathbf{D}| - 4 \le 8$, which is a contradiction. Thus $j \ge 5$ and once again $|\mathbf{A}| \le |\mathbf{D}| - 4 \le 8$, which is impossible. The lemma follows.

LEMMA 3.5. *Theorem* 1.1 *holds for the cases* $N = 27$, 65.

Proof. Suppose $N = 27$. We argue with the prime $p = 43$. Suppose $\alpha \in [22, 27]$ is such that $r_p(\alpha) \geq 2$. Then we have elements $\alpha d_i A, (p-\alpha) d_i A$, $i = 1, 2$; $(d_1, d_2) = 1$, $d_1 < d_2$. We may write $d_i = X_i Y_i$, where $X_i = (d_i, \alpha)$ and $Y_i = (d_i, p - \alpha)$. It may easily be verified from the inequality $\min(X_1, X_2) \geq \alpha |X_1 - X_2|/(N - \alpha)$ that $X_1 = X_2 = 1$. The corresponding inequality for Y_i shows that $Y_2 = Y_1 + 1$ and that $(p - \alpha)/(Y_1(Y_1 + 1)) \leq 4$. Thus the only possibilities for α are 23 and 25. It can be easily verified that $r_p(\alpha) \leq 2$ for these values of α. On the other hand,

$$\sum_{r_p(\alpha) \geq 2} (r_p(\alpha) - 1) \geq \pi(27) - \pi(15) = 3.$$

A contradiction ensues.

Suppose $N = 65$. We argue with $p = 113$. Suppose $\alpha \in [57, 65]$ is such that $r_p(\alpha) \geq 2$. The preceding arguments would show that the only possibilities are $\alpha = 57, 65$. In both these cases it is easily seen that $r_p(\alpha) \leq 2$. However,

$$\sum_{r_p(\alpha) \geq 2} (r_p(\alpha) - 1) \geq \pi(65) - \pi(48) = 3.$$

This contradiction completes the proof.

Lemmata 3.2 through 3.5 prove Theorem 1.1 for all $5 \leq N \leq 2.22 \cdot 10^{12}$. Henceforth we will assume that $N \geq 2.22 \cdot 10^{12}$.

4. Lower bounds for the average value of $r_p(\alpha) - 1$**.** In this section we are interested in obtaining lower bounds for the sum

$$\sum_{p \in [2N - 2G(N), 2N - G(N)]} \sum_{\substack{\alpha \in \mathbf{J}_p \\ r_p(\alpha) \geq 2}} (r_p(\alpha) - 1).$$

From Lemma 2.1 we know that this is

$$\geq \sum_{p \in [2N - 2G(N), 2N - G(N)]} \sum_{\substack{\alpha \in \mathbf{J}_p \\ r_p(\alpha) = 0}} 1.$$

Further, from Lemma 2.2 we concluded that if $\alpha(p - \alpha)$ contains a prime factor $> N/2$ then $r_p(\alpha) = 0$. We now extend Lemma 2.2 to show that $r_p(\alpha) = 0$ if $\alpha(p - \alpha)$ contains a prime factor $> (N + G(N))/D^{1/3}$ (recall that D was defined in the paragraph following Lemma 2.4). We also recall that $D \leq N$.

Throughout the rest of the paper $G(N)$ will satisfy the bound $(5N)^{2/3} \leq G(N) \leq N/1000$. Our choice of $G(N)$ (in §6) will be consistent with this assumption.

LEMMA 4.1. *Suppose* $\mathbf{P}$ *is not empty. There is a prime* $p \in \mathbf{P}$ *and an integer* $\alpha \in \mathbf{J}_p$ *such that each prime* q *dividing an element of* $\mathbf{A}$, *with* q *not dividing* $\alpha(p-\alpha)$, *satisfies* $q \le (N+G(N))/D^{1/3}$. *Consequently*, *with the possible exception of at most two primes*, *no prime* q *greater than* $(N+G(N))/D^{1/3}$ *can divide an element of* $\mathbf{A}$.

Proof. From the definition of D, there exist integers $d_1 < d_2 = D$ with $(d_1, d_2) = 1$, a prime p in $\mathbf{P}$, an integer $\alpha \in \mathbf{J}_p$ and an integer A with

$$\alpha d_i A, (p-\alpha) d_i A \in \mathbf{A} \qquad (i = 1, 2).$$

From Lemma 2.3 we may write $d_i = X_i Y_i$, where $X_i = (\alpha, d_i)$ and $Y_i = (p-\alpha, d_i)$. We also know that $d_1 d_2 \mid \alpha(p-\alpha)$ and $d_1 \ge \alpha(p-\alpha)D/N^2$.

Let $q \nmid \alpha(p-\alpha)$ be a prime $\ge N(1+G(N)/N)/D^{1/3}$. We will first establish that $q \nmid A$. Then, arguing with the reciprocal set $\mathbf{A}^*$ we will obtain the lemma.

Suppose $q \mid A$. Clearly there exists $a \in \mathbf{A}$ with $q \nmid a$. Put $G = (a, A)$ and let

$$B_1 = \left(\frac{a}{G}, X_1\right), \quad B_2 = \left(\frac{a}{G}, X_2\right), \quad B_3 = \left(\frac{a}{G}, Y_1\right), \quad B_4 = \left(\frac{a}{G}, Y_2\right).$$

Let $B = \prod_i B_i$. Finally, let

$$F_1 = \left(\frac{a}{BG}, \frac{\alpha}{X_2}\right), \quad F_3 = \left(\frac{a}{BG}, \frac{p-\alpha}{Y_2}\right),$$
$$F_2 = \left(\frac{a}{BG}, \frac{\alpha}{X_1}\right), \quad F_4 = \left(\frac{a}{BG}, \frac{p-\alpha}{Y_1}\right)$$

and put $F = \prod_i F_i$.

Observe that $GB \operatorname{lcm}[F_1, F_2, F_3, F_4] \mid a$ and so

$$\begin{aligned} a &\ge GB \operatorname{lcm}[F_1, F_2, F_3, F_4] = BG[F_1, F_2][F_3, F_4] \\ &= \frac{BGF}{(F_1, F_2)(F_3, F_4)} \ge \frac{BGFX_1X_2Y_1Y_2}{\alpha(p-\alpha)} = \frac{BGFd_1d_2}{\alpha(p-\alpha)}. \end{aligned}$$

Note that $N \ge \alpha d_1 A/(a, \alpha d_1 A)$ and $N \ge a/(a, \alpha d_1 A)$. So

$$(a, \alpha d_1 A) \ge \max\left(\frac{a}{N}, \frac{\alpha d_1 A}{N}\right).$$

Now, since $(\alpha, B_4) = 1$,

$$\begin{aligned} (a, \alpha d_1 A) &= G\left(\frac{a}{G}, \alpha d_1\right) = GB_1B_3\left(\frac{a}{GB_1B_3}, \alpha\right) = GB_1B_3B_2\left(\frac{a}{BG}, \frac{\alpha}{X_2}\right) \\ &= GB_1B_2B_3F_1. \end{aligned}$$

Similarly,

$$GB_1B_2B_4F_2 = (a, \alpha d_2 A) \ge \max\left(\frac{a}{N}, \frac{\alpha d_2 A}{N}\right),$$

$$GB_1B_3B_4F_3 = (a, (p-\alpha)d_1 A) \ge \max\left(\frac{a}{N}, \frac{(p-\alpha)d_1 A}{N}\right)$$

and

$$GB_2B_3B_4F_4 = (a, (p-\alpha)d_2 A) \ge \max\left(\frac{a}{N}, \frac{(p-\alpha)d_2 A}{N}\right).$$

Taking the product of these four inequalities, we obtain

$$\begin{aligned} G^4B^3F &\ge \max\left(\left(\frac{a}{N}\right)^4, \frac{\alpha^2(p-\alpha)^2 d_1^2 d_2^2 A^4}{N^4}\right) \\ &\ge \frac{a}{N} \cdot \frac{(\alpha(p-\alpha)d_1d_2)^{3/2}A^3}{N^3} \\ &\ge \frac{GBF(\alpha(p-\alpha))^{1/2}(d_1d_2)^{5/2}A^3}{N^4}. \end{aligned}$$

Hence $G^3A^{-3}B^2(d_1d_2)^{-2} \ge (\alpha(p-\alpha))^{1/2}(d_1d_2)^{1/2}N^{-4}$. Since $q \mid A$ but $q \nmid G$, $G/A \le q^{-1}$. Also $B \le d_1d_2$. Thus

$$q^{-3} \ge (\alpha(p-\alpha))^{1/2}(d_1d_2)^{1/2}N^{-4}.$$

Since $d_2 = D$ and $d_1 \ge \alpha(p-\alpha)d_2/N^2 = \alpha(p-\alpha)D/N^2$ by Lemma 2.3, we see that

$$\begin{aligned} q^{-3} &\ge (\alpha(p-\alpha))^{1/2}\left(\frac{D^2\alpha(p-\alpha)}{N^2}\right)^{1/2} N^{-4} = \alpha(p-\alpha)DN^{-5} \\ &\ge N(p-N)DN^{-5} \ge (N-2G(N))DN^{-4}. \end{aligned}$$

So

$$q^3 \le \frac{N^4}{D(N-2G(N))} \le N^3D^{-1}\left(1-2\frac{G(N)}{N}\right)^{-1}$$

whence

$$q \le ND^{-1/3}\left(1-2\frac{G(N)}{N}\right)^{-1/3} \le (N+G(N))D^{-1/3},$$

since $G(N)/N \le 10^{-3}$. This is a contradiction.

Suppose $q \mid a \in \mathbf{A}$. Let $M = \operatorname{lcm}[a_1, \dots, a_N]$. Then $q \mid M$. Also $\alpha d_1 A \mid M$ and $\alpha d_2 A \mid M$ and so $\alpha d_1 d_2 A \mid M$ and similarly $(p-\alpha)d_1d_2A \mid M$. So $\alpha(p-\alpha)d_1d_2A \mid M$. Thus the elements of $\mathbf{A}^*$ corresponding to $\alpha d_1 A$, $\alpha d_2 A$, $(p-\alpha)d_1A$, $(p-\alpha)d_2A$ are of the form $(p-\alpha)d_2A^*$, $(p-\alpha)d_1A^*$, $\alpha d_2 A^*$, $\alpha d_1 A^*$ (respectively), where $q \mid A^*$. Our earlier argument again yields a contradiction. This completes the proof of our first assertion.

Since $D \le N$, $(N+G(N))/D^{1/3} > N^{2/3}$ and so $\alpha(p-\alpha)$ $(< N^2)$ can have at most two prime divisors greater than $(N+G(N))/D^{1/3}$. Clearly these are the only possible exceptional primes which might divide elements of $\mathbf{A}$. The second assertion follows.

Let θ be such that

$$\text{(i)}\qquad \pi(2N-G(N))-\pi(2N-2G(N)) \ge \frac{\theta G(N)}{\log N}$$

and, for any $y \ge G(N)$, if $N/G(N) \ge r \ge 1$, then

$$\text{(ii)}\qquad \pi\left(\frac{N}{r}\right)-\pi\left(\frac{N-y}{r}\right) \ge \frac{\theta y}{r\log N}.$$

We assume that $G(N)$ is such that θ exceeds $1/4$: this assumption will be vindicated by our choice of $G(N)$ in Section 6.

Lemma 4.2. *Let p be any element of $\mathbf{P}$ and suppose*

$$D \le ((N+G(N))/G(N))^3.$$

Then

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) \ge \frac{\theta(2N-p)}{6\log N}\log(0.94(e/2)^3D).$$

If $D > ((N+G(N))/G(N))^3$, then

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) \ge \frac{1}{2}\cdot\frac{2N-p}{\log N}\log\left(\frac{(0.94e)^\theta}{e^{2.3}2^\theta}\left(\frac{N+G(N)}{G(N)}\right)^{\theta-1}D^{1/3}\right).$$

Consequently,

$$\sum_{p\in\mathbf{P}}\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) \ge \frac{\theta}{8}\cdot\frac{P_1\log D}{\log N},$$

where $P_1 = \sum_{p\in\mathbf{P}}(2N-p)$.

Proof. From Lemma 2.1 we see that

$$\sum_{p}\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}} (r_p(\alpha)-1) \ge \sum_{p}\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1.$$

From Lemma 4.1, $r_p(\alpha)=0$ if α contains a prime factor greater than $(N+G(N))/D^{1/3}$ (unless the prime divisor happened to be one of the two possible exceptions). Further, since $D < N$, $(N+G(N))/D^{1/3} > N^{2/3}$ and so α can have at most one prime divisor greater than $(N+G(N))/D^{1/3}$. Thus

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \sum_{q>(N+G(N))/D^{1/3}} \sum_{\substack{\alpha\in\mathbf{J}_p\\ q\mid\alpha}} 1 - 2\left(\frac{G(N)D^{1/3}}{N}+1\right).$$

The second term in the right hand side is a consequence of the two possible exceptional primes which, clearly, can divide at most $1+G(N)D^{1/3}/N$ elements of $\mathbf{J}_p$.

Also by Lemma 4.1, $r_p(\alpha)=0$ if $p-\alpha$ contains a prime factor greater than $(N+G(N))/D^{1/3}$ (again with two possible exceptional primes). Arguing as above we see that

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \sum_{q>(N+G(N))/D^{1/3}} \sum_{\substack{\alpha\in\mathbf{J}_p\\ q\mid p-\alpha}} 1 - 2\left(\frac{G(N)D^{1/3}}{N}+1\right).$$

From these two inequalities we deduce that

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \frac{1}{2}\sum_{q>(N+G(N))/D^{1/3}} \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1 - 2\left(\frac{G(N)D^{1/3}}{N}+1\right).$$

Since $D\le N$, $N\ge 10^{12}$ and $G(N)\ge(5N)^{2/3}$ we see that

$$\begin{aligned} 2\left(\frac{G(N)D^{1/2}}{N}+1\right) &\le \frac{G(N)}{\log N}\left(\frac{2\log N}{N^{2/3}}+\frac{2\log N}{(5N)^{2/3}}\right)\\ &\le \frac{G(N)}{\log N}\cdot\frac{24\log 10}{10^8}\left(1+\frac{1}{5^{2/3}}\right)\le 10^{-6}\frac{G(N)}{\log N}. \end{aligned}$$

Suppose $D\le(N+G(N))^3/G(N)^3$. If $\alpha\in[p-N,N]$ has a prime divisor larger than $(N+G(N))/D^{1/3}$, then we may write $\alpha=qr$, where $r\le D^{1/3}\alpha/(N+G(N))$ and q is prime. Thus

$$\begin{aligned} \sum_{q>(N+G(N))/D^{1/3}} \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1 &\ge \sum_{r\le D^{1/3}(N-2G(N))/(N+G(N))} \sum_{(p-N)/r\le q\le N/r} 1\\ &\ge \sum_{r\le 0.99D^{1/3}}\left(\pi\left(\frac{N}{r}\right)-\pi\left(\frac{p-N}{r}\right)\right). \end{aligned}$$

Since $N-(p-N)=2N-p\ge G(N)$, we see that

$$\pi\left(\frac{N}{r}\right)-\pi\left(\frac{p-N}{r}\right)\ge\frac{\theta(2N-p)}{r\log N}.$$

Hence, using Lemma 2.7 and since $\theta \ge 1/4$,

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \frac{1}{2}\cdot\frac{\theta(2N-p)}{\log N}\sum_{r\le 0.99D^{1/3}}\frac{1}{r} - 2\Big(\frac{G(N)D^{1/3}}{N}+1\Big)$$
$$\ge \frac{\theta(2N-p)}{6\log N}\log(0.95(e/2)^3D) - 10^{-6}\frac{G(N)}{\log N}$$
$$\ge \frac{\theta(2N-p)}{6\log N}(\log(0.95(e/2)^3D) - 24\cdot 10^{-6})$$
$$\ge \frac{\theta(2N-p)}{6\log N}\log(0.94(e/2)^3D).$$

We now turn to the case $D \ge (N+G(N))^3/G(N)^3$. Clearly,

$$\sum_{q>(N+G(N))/D^{1/3}}\ \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1$$
$$= \sum_{q>G(N)}\ \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1 + \sum_{G(N)\ge q>(N+G(N))/D^{1/3}}\ \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1.$$

By the result just proved, the first sum on the right hand side is

$$\ge \frac{\theta(2N-p)}{6\log N}\log\Big(\frac{0.95(e/2)^3(N+G(N))^3}{G(N)^3}\Big).$$

As for the second sum, using Lemma 2.7, we see that

$$\sum_{G(N)\ge q>(N+G(N))/D^{1/3}}\ \sum_{\substack{\alpha\in[p-N,N]\\ q\mid\alpha}} 1$$
$$\ge \sum_{G(N)\ge q>(N+G(N))/D^{1/3}}\Big(\frac{2N-p}{q}-1\Big)$$
$$\ge (2N-p)\sum_{G(N)\ge q>(N+G(N))/D^{1/3}}\frac{1}{q} - \pi(G(N))$$
$$\ge \frac{2N-p}{\log G(N)}\log\frac{G(N)D^{1/3}}{N+G(N)} - 1.2\,\frac{2N-p}{\log G(N)} - 1.1\,\frac{G(N)}{\log G(N)}$$
$$\ge \frac{2N-p}{\log N}\Big(\log\frac{G(N)D^{1/3}}{N+G(N)} - 2.3\Big).$$

Since $\theta \ge 1/4$,

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \frac{1}{2}\cdot\frac{2N-p}{\log N}\log\biggl(\frac{(0.95e)^\theta}{e^{2.3}2^\theta}\biggl(\frac{N+G(N)}{G(N)}\biggr)^\theta\frac{G(N)}{N+G(N)}D^{1/3}\biggr) - 10^{-6}\frac{G(N)}{\log N}$$
$$\ge \frac{1}{2}\cdot\frac{2N-p}{\log N}\log\biggl(\frac{(0.94e)^\theta}{e^{2.3}2^\theta}\biggl(\frac{N+G(N)}{G(N)}\biggr)^{\theta-1}D^{1/3}\biggr).$$

Our last assertion is a trivial consequence of the above inequalities in the range $D\le (N+G(N))^3/G(N)^3$. In the range $(N+G(N))^3/G(N)^3\le D\le (N+G(N))^4/G(N)^4$ we note that from the preceding paragraph it is immediate that

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \frac{\theta(2N-p)}{6\log N}\log\biggl(\frac{0.94(e/2)^3(N+G(N))^3}{G(N)^3}\biggr)$$
$$\ge \frac{\theta(2N-p)}{6\log N}\cdot\frac{3\log D}{4}\ge\frac{\theta}{8}(2N-p)\frac{\log D}{\log N};$$

summing over p our assertion follows in this case. Lastly, if $D\ge (N+G(N))^4/G(N)^4$ then, using $N/G(N)\ge 1000\ge e^{6.9}$,

$$\sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)=0}} 1 \ge \frac{2N-p}{2\log N}\log\biggl(\frac{(0.94e)^\theta}{e^{2.3}2^\theta}\biggl(\frac{N+G(N)}{G(N)}\biggr)^{\theta-1}D^{1/3}\biggr)$$
$$\ge \frac{2N-p}{2\log N}\log\biggl(\frac{(0.94e)^\theta}{e^{2.3}2^\theta}D^{\theta/4}\biggl(\frac{N+G(N)}{G(N)}\biggr)^{1/3}\biggr)$$
$$\ge\frac{\theta}{8}(2N-p)\frac{\log D}{\log N};$$

summing over p we obtain the desired conclusion. The proof is complete.

5. Upper bounds for the average value of $r_p(\alpha)-1$. We recall that, as stated in the preceding section, $G(N)$ satisfies the bounds $(5N)^{2/3}\le G(N)\le N/1000$.

From Lemma 2.5 we see that

$$\sum_{p\in[2N-2G(N),2N-G(N)]}\ \sum_{\substack{\alpha\in\mathbf{J}_p\\ r_p(\alpha)\ge 2}}(r_p(\alpha)-1)$$
$$\le\sum_p\sum_{\alpha\in\mathbf{J}_p}(k_D(\alpha)+k_D(p-\alpha)+k_D(\alpha)k_D(p-\alpha))$$
$$=\sum_p\sum_{\alpha\in[p-N,N]}k_D(\alpha)+\sum_p\sum_{\alpha\in\mathbf{J}_p}k_D(\alpha)k_D(p-\alpha).$$

Lemma 5.1. *Let* $T = \max(N, N^2/D^2)$. *Let* $l \mid 6$ *and* $x \in [N - 2G(N), N - G(N)]$. *Then*

$$\sum_{\substack{\alpha\in[x,N]\\(\alpha,l)=1}} k_D(\alpha)$$

$$\le \frac{1.01}{4}\left(\frac{\phi(l)}{l}\right)^3 \frac{(N-x)^2}{N}\left(\log\frac{3(N-x)^2}{4T} + l\right) + \frac{(N-x)^2}{4N}\left(3.1\frac{\phi(l)}{l} + \frac{1.02(5l+2)\phi(l)^3}{2l^2} + \frac{1.05(7l+2)\phi(l)^3}{l^3}\right).$$

In particular, *for any* $p \in \mathbf{P}$,

$$\sum_{\alpha\in[p-N,N]} k_D(\alpha) \le \frac{(2N-p)G(N)}{2N}\log\left(\frac{40G(N)^2}{T}\right).$$

Also,

$$\sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,6)=1}} k_D(\alpha) \le \frac{1.01}{108}\cdot\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T} + 0.51\frac{G(N)^2}{N}.$$

Proof. Clearly if $\alpha = N$ then $k_D(\alpha) = 0$. So we may suppose that $\alpha < N$. Observe that if $\alpha = \lambda X_1 X_2$ is a solution counted in $k_D(\alpha)$ then, since $X_1 < X_2 \le X_1 N/\alpha$,

$$\frac{\alpha}{\lambda} = X_1X_2 \ge X_2^2\frac{\alpha}{N}$$

whence $X_2 \le \sqrt{N/\lambda}$. Similarly,

$$\frac{\alpha}{\lambda} = X_1X_2 \le X_1^2\frac{N}{\alpha}$$

whence $X_1 \ge \alpha/\sqrt{N\lambda} \ge x/\sqrt{N\lambda}$. Since $X_1, X_2 \le D$, clearly $\lambda \ge (N - 2G(N))/D^2 \ge 0.99N/D^2$. Also, trivially, $\lambda \ge 1$ and so $\lambda \ge 0.99\max(N/D^2, 1) = 0.99T/N$. Note that

$$\frac{N}{\alpha} \ge \frac{X_2}{X_1} = 1 + \frac{X_2 - X_1}{X_1} \ge 1 + \frac{1}{X_1}$$

whence $X_1 \ge \alpha/(N-\alpha)$. Hence $\alpha \ge \lambda X_1(X_1+1) \ge \lambda\alpha N/(N-\alpha)^2$ and so $\lambda \le (N-\alpha)^2/N \le (N-x)^2/N$. Finally, $X_2 = \alpha/(\lambda X_1)$ and so

$$X_2 \ge \max\left(X_1 + 1, \frac{x}{\lambda X_1}\right).$$

Thus

$$\sum_{\substack{\alpha\in[x,N]\\(\alpha,l)=1}} k_D(\alpha)$$

$$\leq \sum_{\substack{0.99T/N\leq\lambda\leq 4G(N)^2/N\\(\lambda,l)=1}} \sum_{\substack{x/\sqrt{N\lambda}\leq X_1\leq\sqrt{N/\lambda}-1\\(X_1,l)=1}} \sum_{\substack{\max(X_1+1,\,x/(\lambda X_1)\leq X_2\leq\sqrt{N/\lambda}\\(X_2,l)=1}} 1.$$

Suppose $\lambda \geq 1.01(N-x)^2/(4N)$. We claim that there is at most one choice for X_1 and X_2. The interval $[x/\sqrt{N\lambda}, \sqrt{N/\lambda}]$ can contain at most three integers, say $a-1$, a and $a+1$. If X_1 and X_2 differ by more than 1 we would obtain a contradiction to $N/\alpha \geq X_2/X_1 \geq 1+(2/X_1)$. Thus there are at most two possibilities: $X_1 = a-1$, $X_2 = a$ and $X_1 = a$, $X_2 = a+1$. Since $\lambda(a(a+1)-a(a-1)) = 2\lambda a > N-x$ at most one of these possibilities can occur. Hence X_1 and X_2 have at most one choice. Hence

$$\sum_{0.99T/N\leq\lambda\leq(N-x)^2/N} \sum_{X_1} \sum_{\substack{X_2\\(\lambda X_1X_2,l)=1}} 1$$

$$\leq \sum_{0.99T/N\leq\lambda\leq(N-x)^2/(4N)} \sum_{X_1} \sum_{\substack{X_2\\(\lambda X_1X_2,l)=1}} 1 + \frac{3(N-x)^2}{4N}\cdot\frac{\phi(l)}{l} + \phi(l)$$

$$\leq \sum_{\substack{0.99T/N\leq\lambda\\\lambda\leq(N-x)^2/(4N)}} \sum_{\substack{X_1\\(\lambda X_1,l)=1}} \left(\left(\left[\sqrt{\frac{N}{\lambda}}\right] - \max\left(X_1, \frac{x}{\lambda X_1}-1\right)\right)\frac{\phi(l)}{l} + \phi(l)\right)$$

$$+\frac{3.1\phi(l)}{4l}\cdot\frac{(N-x)^2}{N}.$$

Now,

$$\sum_{\substack{\sqrt{x/\lambda}\leq X_1\leq\sqrt{N/\lambda}-1\\(X_1,l)=1}} \left(\left[\sqrt{\frac{N}{\lambda}}\right] - X_1\right)$$

$$\leq \sum_{\substack{[\sqrt{x}/(l\sqrt{\lambda})]<k\\k\leq[\sqrt{N}/(l\sqrt{\lambda})]}} \sum_{\substack{\beta=1\\(\beta,l)=1}}^{l} \left(\left[\sqrt{\frac{N}{\lambda}}\right] - (lk+\beta)\right)$$

$$\leq \phi(l)\sum_k\left(\left[\sqrt{\frac{N}{\lambda}}\right] - lk\right)$$

$$\leq \phi(l)\left(\frac{1}{l}\sqrt{\frac{N}{\lambda}} - \frac{\sqrt{x}}{l\sqrt{\lambda}} + 1\right)\left(\sqrt{\frac{N}{\lambda}} - \frac{l}{2}\left(\frac{\sqrt{N}}{l\sqrt{\lambda}} + \frac{\sqrt{x}}{l\sqrt{\lambda}} - 1\right)\right)$$

$$= \frac{\phi(l)}{2l}\Big(\frac{\sqrt{N}-\sqrt{x}}{\sqrt{\lambda}}+l\Big)^2 = \frac{\phi(l)}{2l}\Big(\frac{N-x}{(\sqrt{N}+\sqrt{x})\sqrt{\lambda}}+l\Big)^2$$
$$\le 1.01\frac{\phi(l)}{2l}\Big(\frac{N-x}{2\sqrt{N\lambda}}+l\Big)^2.$$

Similarly,

$$\sum_{\substack{x/\sqrt{N\lambda}\le X_1\le\sqrt{x/\lambda}\\ (X_1,l)=1}}\Big(\Big[\sqrt{\frac{N}{\lambda}}\Big]-\frac{x}{X_1\lambda}+1\Big)$$
$$= \sum_{X_1}\Big(\Big[\sqrt{\frac{N}{\lambda}}\Big]+1-\frac{x}{\lambda}\Big(\frac{x}{\sqrt{N\lambda}}+X_1-\frac{x}{\sqrt{N\lambda}}\Big)^{-1}\Big)$$
$$= \sum_{X_1}\Big(\Big[\sqrt{\frac{N}{\lambda}}\Big]+1-\sqrt{\frac{N}{\lambda}}\Big(1+\frac{\sqrt{N\lambda}}{x}\Big(X_1-\frac{x}{\sqrt{N\lambda}}\Big)\Big)^{-1}\Big)$$
$$\le \sum_{X_1}\Big(\Big[\sqrt{\frac{N}{\lambda}}\Big]+1-\sqrt{\frac{N}{\lambda}}\Big(1-\frac{\sqrt{N\lambda}}{x}\Big(X_1-\frac{x}{\sqrt{N\lambda}}\Big)\Big)\Big)$$
$$= \sum_{X_1}\Big(1+\frac{N}{x}\Big(X_1-\frac{x}{\sqrt{N\lambda}}\Big)\Big)$$
$$\le \frac{\phi(l)}{l}\Big(\sqrt{\frac{x}{\lambda}}-\frac{x}{\sqrt{N\lambda}}+l\Big)$$
$$+\frac{N}{x}\cdot\frac{\phi(l)}{2l}\Big(\sqrt{\frac{x}{\lambda}}-\frac{x}{\sqrt{N\lambda}}+l\Big)\Big(\sqrt{\frac{x}{\lambda}}-\frac{x}{\sqrt{N\lambda}}+2l\Big)$$
$$\le 1.01\frac{\phi(l)}{2l}\Big(\frac{N-x}{2\sqrt{N\lambda}}+l\Big)\Big(\frac{N-x}{2\sqrt{N\lambda}}+2l+2\Big).$$

Using these estimates, we see that

$$\sum_{\substack{\alpha\in[x,N]\\ (\alpha,l)=1}} k_D(\alpha)-\frac{3.1\phi(l)}{4l}\cdot\frac{(N-x)^2}{N}$$
$$\le \sum_{\substack{0.99T/N\le\lambda\le(N-x)^2/(4N)\\ (\lambda,l)=1}}\Big(\frac{1.01}{2}\Big(\frac{\phi(l)}{l}\Big)^2\Big(\frac{N-x}{2\sqrt{N\lambda}}+l\Big)\Big(\frac{N-x}{\sqrt{N\lambda}}+3l+2\Big)$$
$$+1.01\frac{\phi(l)^2}{l}\Big(\frac{N-x}{2\sqrt{N\lambda}}+l\Big)\Big)$$

$$= \sum_{\substack{0.99T/N \le \lambda \le (N-x)^2/(4N) \\ (\lambda,l)=1}} \frac{1.01}{2}\left(\frac{\phi(l)}{l}\right)^2\left(\frac{N-x}{2\sqrt{N\lambda}}+l\right)\left(\frac{N-x}{\sqrt{N\lambda}}+5l+2\right).$$

Clearly,

$$\begin{aligned}
\sum_{\substack{0.99T/N \le \lambda \le (N-x)^2/(4N) \\ (\lambda,l)=1}} \frac{1}{\lambda}
&\le \phi(l) + \sum_{\substack{l+0.99T/N \le \lambda \le (N-x)^2/N \\ (\lambda,l)=1}} \frac{1}{\lambda} \\
&\le \phi(l) + \sum_{1+[0.99T/(lN)] \le k \le [(N-x)^2/(4lN)]} \sum_{\substack{\beta=1 \\ (\beta,l)=1}}^{l} \frac{1}{kl+\beta} \\
&\le \phi(l) + \frac{\phi(l)}{l} \sum_{1+[0.99T/(lN)] \le k \le [(N-x)^2/(4lN)]} \frac{1}{k} \\
&\le \frac{\phi(l)}{l}\left(\log\frac{3(N-x)^2}{4T}+l\right).
\end{aligned}$$

Similarly, we deduce that

$$\sum_{\substack{0.99T/N \le \lambda \le (N-x)^2/(4N) \\ (\lambda,l)=1}} \frac{1}{\sqrt{\lambda}} \le \phi(l) + \frac{\phi(l)}{l}\cdot\frac{N-x}{\sqrt{N}}.$$

Since $G(N)/\sqrt{N} \ge 100$ and $l \le 6$, we see that

$$\begin{aligned}
\sum_{\substack{\alpha\in[x,N] \\ (\alpha,l)=1}} k_D(\alpha)
&\le \frac{3.1\phi(l)}{l}\cdot\frac{(N-x)^2}{4N} + \frac{1.01}{4}\left(\frac{\phi(l)}{l}\right)^3\frac{(N-x)^2}{N}\left(\log\frac{3(N-x)^2}{4T}+l\right) \\
&\quad + \frac{1.01(7l+2)}{4}\left(\frac{\phi(l)}{l}\right)^3\frac{(N-x)}{\sqrt{N}}\left(\frac{N-x}{\sqrt{N}}+l\right) \\
&\quad + \frac{1.01(5l+2)}{2}\cdot\frac{\phi(l)^3}{l^2}\left(\frac{(N-x)^2}{4N}+l\right) \\
&\le \frac{1.01}{4}\left(\frac{\phi(l)}{l}\right)^3\frac{(N-x)^2}{N}\left(\log\frac{3(N-x)^2}{4T}+l\right) \\
&\quad + \frac{(N-x)^2}{4N}\left(3.1\frac{\phi(l)}{l}+\frac{1.02(5l+2)\phi(l)^3}{2l^2}+\frac{1.05(7l+2)\phi(l)^3}{l^3}\right).
\end{aligned}$$

This proves our first assertion.

The other two statements may be proved in much the same way. We only need to make minor modifications to obtain the improved constants. These changes are obvious when dealing with the sum $\sum_{\alpha\in[p-N,N]} k_D(\alpha)$ and since there is no need to split the various sums into blocks of length l we obtain better constants. For the final sum we argue a bit more carefully. If X_1 and X_2 are coprime to 6 then they must differ at least by 2. This gives us the improved bound for λ: $\lambda \le G(N)^2/(4N)$. Also if $\lambda \ge G(N)^2/(25N)$, the interval $[(N-G(N))/\sqrt{N\lambda}, \sqrt{N/\lambda}]$ has at most six integers and at most two of these can be coprime to 6. Thus if $\lambda \ge G(N)^2/(25N)$ then X_1 and X_2 have at most one choice. This additional information leads to the better constant in this case.

Lemma 5.2. *Let $\alpha \in [N - G(N), N]$ and put $\alpha' = (\alpha, 6)$. Then*

$$\sum_{\substack{2N-2G(N)\le p\\ p\le \min(2\alpha,2N-G(N))}} k_D(p-\alpha) \le 2.01C^2 h(\alpha')^2 \frac{G(N)^2}{N}\log(4\log G(N))L(\alpha'),$$

where

$$\begin{aligned} L(\alpha') = 1 &+ 1.03\phi(\alpha')\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)}\cdot\frac{\log(7N/G(N))}{\log(4\log G(N))}\\ &+ \frac{10\phi(\alpha')N}{G(N)^{3/2}\log 99}\cdot\frac{\log^2(\sqrt{G(N)}/10)}{\log(4\log G(N))}\\ &+ \frac{0.3}{\log(4\log G(N))}\log\frac{100e^{\alpha'+1}N}{\alpha' G(N)}\\ &+ 13\frac{\alpha' N}{G(N)^{3/2}\log(4\log G(N))} + 23\frac{(\alpha')^2N^2}{G(N)^3\log(4\log G(N))}. \end{aligned}$$

Proof. Suppose, first, that $D \le \sqrt{G(N)}/10$. If $p-\alpha = \mu Y_1Y_2$ is a solution counted in $k_D(p-\alpha)$ then we must have Y_1 and Y_2 less than D and

$$Y_1 < Y_2 \le Y_1\frac{N}{p-\alpha} \le Y_1\frac{N}{2N-2G(N)-\alpha} \le Y_1\frac{N}{N-2G(N)}.$$

If the interval $[Y_1+1, Y_1N/(p-\alpha)]$ is to contain an integer then $Y_1N/(p-\alpha) - Y_1 \ge 1$ and so

$$Y_1 \ge \frac{p-\alpha}{N-p+\alpha} \ge \frac{p-N}{2N-p} \ge \frac{N-2G(N)}{2G(N)} \ge \frac{0.99N}{2G(N)}.$$

Also, clearly, $\mu = (p-\alpha)/(Y_1Y_2)$ satisfies

$$\frac{2N-2G(N)-\alpha}{Y_1Y_2} \le \mu \le \frac{2N-G(N)-\alpha}{Y_1Y_2}.$$

Thus

$$\sum_p k_D(p-\alpha) \le \sum_{\substack{D\ge Y_1\\ Y_1\ge 0.49N/G(N)}} \sum_{\substack{Y_1N/(N-2G(N))\ge Y_2\\ Y_2>Y_1}} \sideset{}{^*}\sum_{\substack{(2N-G(N)-\alpha)/(Y_1Y_2)\ge\mu\\ \mu\ge(2N-2G(N)-\alpha)/(Y_1Y_2)}} 1$$

where the $*$ in the third sum indicates that $\alpha+\mu Y_1Y_2$ is to be prime.

If $(\alpha,\mu Y_1)>1$ then the inner sum is 0. If $(\alpha,\mu Y_1)=1$ then, by the Brun–Titchmarsh theorem (our Lemma 2.6), the inner sum is bounded by

$$\frac{2G(N)}{\phi(\mu Y_1)}\left(\log\frac{G(N)}{Y_1Y_2}\right)^{-1} \le \frac{2G(N)}{\phi(\mu)\phi(Y_1)}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}$$

(since $\phi(\mu Y_1)\ge\phi(\mu)\phi(Y_1)$). Hence, using Lemma 2.8,

$$\begin{aligned}\sum_p k_D(p-\alpha) &\le 2G(N)\sum_{\substack{Y_1\\ (Y_1,\alpha')=1}}\frac{1}{\phi(Y_1)}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}\sum_{\substack{Y_2\\ (Y_2,\alpha')=1}}\frac{1}{\phi(Y_2)}\\ &\le 2G(N)\sum_{\substack{Y_1\\ (Y_1,\alpha')=1}}\frac{1}{\phi(Y_1)}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}\frac{Ch(\alpha')}{Y_1}\\ &\quad\times\left(\frac{2G(N)Y_1}{N-2G(N)}+\phi(\alpha')\log\frac{1.01eY_1}{\alpha'}\right).\end{aligned}$$

Now,

$$\begin{aligned}&\sum_{\substack{0.49N/G(N)\le Y_1\le D\\ (Y_1,\alpha')=1}}\frac{1}{\phi(Y_1)}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}\\ &\quad=\sum_{\substack{Y_1\\ (Y_1,\alpha')=1}}\frac{1}{Y_1}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}\sum_{d\mid Y_1}\frac{\mu(d)^2}{\phi(d)}\\ &\quad=\sum_{\substack{d\le D\\ (d,\alpha')=1}}\frac{\mu(d)^2}{d\phi(d)}\sum_{\substack{0.49N/(dG(N))\le y\le D/d\\ (y,\alpha')=1}}\frac{1}{y}\left(\log\frac{0.99G(N)}{y^2d^2}\right)^{-1}.\end{aligned}$$

Since $G(N)/D^2\ge 100$, we see that

$$\sum_{\substack{0.49N/(dG(N))\le y\le D/d\\ (y,\alpha')=1}}\frac{1}{y}\left(\log\frac{0.99G(N)}{y^2d^2}\right)^{-1}$$

$$\le \sum_{\substack{y\le D/d\\ (y,\alpha')=1}} \frac{1}{y}\Big(\log \frac{0.99G(N)}{y^2d^2}\Big)^{-1}$$

$$\le \frac{\phi(\alpha')}{\log 99} + \sum_{\substack{\alpha'<y\le D/d\\ (y,\alpha')=1}} \frac{1}{y}\Big(\log \frac{0.99G(N)}{y^2d^2}\Big)^{-1}.$$

Next,

$$\sum_{\substack{\alpha'<y\le D/d\\ (y,\alpha')=1}} \frac{1}{y}\Big(\log \frac{0.99G(N)}{y^2d^2}\Big)^{-1}$$

$$= \sum_{D/(\alpha'd)\ge k\ge 1} \sum_{\substack{l=0\\ (l,\alpha')=1}}^{\alpha'-1} \frac{1}{k\alpha'+l}\Big(\log \frac{0.99G(N)}{(k\alpha'+l)^2d^2}\Big)^{-1}$$

$$\le \frac{\phi(\alpha')}{\alpha'} \sum_{k\le D/(\alpha'd)} \frac{1}{k}\Big(\log \frac{0.99G(N)}{(k+1)^2(\alpha'd)^2}\Big)^{-1}$$

$$\le \frac{\phi(\alpha')}{\alpha'\log 25} + \frac{\phi(\alpha')}{\alpha'} \sum_{2\le k\le D/(\alpha'd)} \frac{1}{k}\Big(\log \frac{3.96G(N)}{9(\alpha'dk)^2}\Big)^{-1}$$

(since $(k+1)^2 \le 9k^2/4$ when $k \ge 2$). Since $\log(0.44G(N)/(\alpha'dk)^2)/k$ is a decreasing function of k in the range $2 \le k \le D/(\alpha'd)$, it follows that

$$\sum_{2\le k\le D/(\alpha'd)} \frac{1}{k}\Big(\log \frac{0.44G(N)}{(\alpha'dk)^2}\Big)^{-1}$$

$$\le \frac{1}{2} \sum_{2\le k\le D/(\alpha'd)} \int_{k-1}^{k} \Big(\log \frac{\sqrt{0.44G(N)}}{\alpha'td}\Big)^{-1} \frac{dt}{t}$$

$$\le \frac{1}{2} \int_{1}^{D/(\alpha'd)} \Big(\log \frac{\sqrt{0.44G(N)}}{\alpha'td}\Big)^{-1} \frac{dt}{t}$$

$$= \frac{1}{2}\Big[-\log\log \frac{\sqrt{0.44G(N)}}{\alpha'td}\Big]_1^{D/(\alpha'd)}$$

$$\le \frac{1}{2}\log\log\sqrt{G(N)}.$$

Piecing these observations together (and since $\alpha' \le 6$) we deduce that

$$\sum_{\substack{0.49N/G(N)\le Y_1\le D\\(Y_1,\alpha')=1}} \frac{1}{\phi(Y_1)}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}$$

$$\le \frac{1}{2}\cdot\frac{\phi(\alpha')}{\alpha'}\sum_{\substack{d\le D\\(d,\alpha')=1}}\frac{\mu(d)^2}{d\phi(d)}\left(\frac{\alpha'}{\log 99}+\frac{1}{\log 25}+\log\log\sqrt{G(N)}\right)$$

$$\le \frac{1}{2}\cdot\frac{\phi(\alpha')}{\alpha'}\sum_{\substack{d=1\\(d,\alpha')=1}}^{\infty}\frac{\mu(d)^2}{d\phi(d)}\log(4\log G(N))$$

$$= \frac{Ch(\alpha')}{2}\log(4\log G(N)).$$

Thus

$$\sum_p k_D(p-\alpha)\le \frac{2.01G(N)^2}{N}C^2h^2(\alpha')\log(4\log G(N))$$
$$+\,2G(N)\phi(\alpha')Ch(\alpha')$$
$$\times \sum_{\substack{Y_1\\(Y_1,\alpha')=1}}\frac{1}{Y_1\phi(Y_1)}\log\frac{1.01eY_1}{\alpha'}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}.$$

Let

$$A(t)=\sum_{\substack{0.49N/G(N)<Y_1\le t\\(Y_1,\alpha')=1}}\frac{1}{\phi(Y_1)}.$$

Then $A(0.49N/G(N))=0$ and, by Lemma 2.8, $A(t)\le Ch(\alpha')\log t$. Clearly,

$$\sum_{\substack{Y_1\\(Y_1,\alpha')=1}}\frac{1}{Y_1\phi(Y_1)}\log\frac{1.01eY_1}{\alpha'}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}$$

$$=\int_{0.49N/G(N)}^{D}\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\,dA(t)$$

$$\le\int_{0.49N/G(N)}^{\sqrt{G(N)}/10}\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\,dA(t).$$

Integrating by parts (and since $A(0.49N/G(N))=0$), we obtain

$$\int_{0.49N/G(N)}^{\sqrt{G(N)}/10} \frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\,dA(t)$$
$$= \frac{10}{\sqrt{G(N)}\log 99}\log\frac{1.01e\sqrt{G(N)}}{10\alpha'}A\Big(\frac{\sqrt{G(N)}}{10}\Big) - \int_{0.49N/G(N)}^{\sqrt{G(N)}/10} A(t)\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)'dt$$
$$\le \frac{10Ch(\alpha')}{\sqrt{G(N)}\log 99}\Big(\log\frac{\sqrt{G(N)}}{3}\Big)^2 - \int_{0.49N/G(N)}^{\sqrt{G(N)}/10} A(t)\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)'dt.$$

Since

$$\Big(\frac{1}{t}\log\frac{1.01et}{\alpha'}\Big(\log\frac{0.99G(N)}{t^2}\Big)^{-1}\Big)'$$

is negative in the range $2 \le t \le \sqrt{G(N)}/10$, we deduce that

$$-\int_{0.49N/G(N)}^{\sqrt{G(N)}/10} A(t)\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)'dt \le -Ch(\alpha')\int_{0.49N/G(N)}^{\sqrt{G(N)}/10} \log t\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)'dt.$$

Note that, since $N/G(N) \ge 1000$,

$$-\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)' = \frac{1}{t}\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big) \times\Big(1-\frac{\alpha'}{1.01e}\Big(\log\frac{1.01et}{\alpha'}\Big)^{-1}-\frac{1}{\log(0.99G(N)/t^2)}\Big)$$
$$\ge \frac{1}{t}\Big(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\Big)$$

$$\times\left(1-\frac{\alpha'}{1.01e\log(500e/\alpha')}-\frac{1}{\log 99}\right)$$

$$\geq \frac{0.4}{t}\left(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\right).$$

Hence

$$-\int_{0.49N/G(N)}^{\sqrt{G(N)}/10} \log t\left(\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\right)' dt$$

$$= \left[-\log t\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\right]_{0.49N/G(N)}^{\sqrt{G(N)}/10}$$

$$+ \int_{0.49N/G(N)}^{\sqrt{G(N)}/10} \frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\frac{dt}{t}$$

$$\leq 2.05\frac{G(N)}{N}\left(\log\frac{0.49N}{G(N)}\right)\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)}$$

$$+ 2.5\int_{0.49N/G(N)}^{\sqrt{G(N)}/10} \left(-\frac{\log(1.01et/\alpha')}{t\log(0.99G(N)/t^2)}\right)' dt$$

$$\leq 2.05\frac{G(N)}{N}\cdot\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)}\log\frac{0.49Ne^{2.5}}{G(N)}$$

$$\leq 2.05\frac{G(N)}{N}\cdot\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)}\log\frac{7N}{G(N)}.$$

Putting these remarks together, we see that

$$\sum_{\substack{Y_1\\(Y_1,\alpha')=1}} \frac{1}{Y_1\phi(Y_1)}\log\frac{1.01eY_1}{\alpha'}\left(\log\frac{0.99G(N)}{Y_1^2}\right)^{-1}$$

$$\leq \frac{10Ch(\alpha')}{\sqrt{G(N)}\log 99}\left(\log\frac{\sqrt{G(N)}}{3}\right)^2$$

$$+ 2.05Ch(\alpha')\frac{G(N)}{N}\cdot\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)}\log\frac{7N}{G(N)}.$$

Consequently,

$$\sum_p k_D(p-\alpha) \le \frac{2.01G(N)^2}{N} C^2 h^2(\alpha') \log(4\log G(N)) \times \Big(1 + 1.03\phi(\alpha')\frac{\log(eN/(2\alpha' G(N)))}{\log(3.9G(N)^3/N^2)} \cdot \frac{\log(7N/G(N))}{\log(4\log G(N))} + \frac{10\phi(\alpha')N}{G(N)^{3/2}\log 99} \cdot \frac{\log^2(\sqrt{G(N)}/10)}{\log(4\log G(N))}\Big).$$

This proves the lemma when $D \le \sqrt{G(N)}/10$.

We now turn to the case $D > \sqrt{G(N)}/10$. As noted earlier,

$$\sum_p k_D(p-\alpha) \le \sum_{\substack{D\ge Y_1\\ Y_1\ge 0.49N/G(N)}} \sum_{\substack{Y_1N/(N-2G(N))\ge Y_2\\ Y_2>Y_1}} \sideset{}{^*}\sum_{\substack{(2N-G(N)-\alpha)/(Y_1Y_2)\ge\mu\\ \mu\ge(2N-2G(N)-\alpha)/(Y_1Y_2)}} 1$$
$$\le \sum_{\substack{\sqrt{G(N)}/10\ge Y_1\\ Y_1\ge 0.49N/G(N)}} \sum_{\substack{Y_1N/(N-2G(N))\ge Y_2\\ Y_2>Y_1}} \sideset{}{^*}\sum_{\substack{(2N-G(N)-\alpha)/(Y_1Y_2)\ge\mu\\ \mu\ge(2N-2G(N)-\alpha)/(Y_1Y_2)}} 1 + \sum_{Y_1\ge\sqrt{G(N)}/10} \sum_{\substack{Y_1N/(N-2G(N))\ge Y_2\\ Y_2>Y_1}} \sideset{}{^*}\sum_{\substack{(2N-G(N)-\alpha)/(Y_1Y_2)\ge\mu\\ \mu\ge(2N-2G(N)-\alpha)/(Y_1Y_2)}} 1$$

where, as before, the $*$ indicates that $\alpha + \mu Y_1Y_2$ is to be prime. The result just established takes care of the first sum. It is easy to see that the second sum is bounded by

$$\sum_{\mu\le 100N/G(N)} \sum_{\substack{(N-2G(N))/\sqrt{N\mu}\le Y_1\\ Y_1\le\sqrt{N/\mu}}} \sum_{\substack{Y_1<Y_2\le\sqrt{N/\mu}\\ \alpha+\mu Y_1Y_2 \text{ prime}}} 1.$$

If $\alpha+\mu Y_1Y_2$ is to be prime then $(\mu Y_1Y_2,\alpha)=1$ and, in particular, $(\mu Y_1Y_2,\alpha') = 1$. Thus

$$\sum_{\mu\le 100N/G(N)} \sum_{Y_1} \sum_{\substack{Y_2\\ \alpha+\mu Y_1Y_2 \text{ prime}}} 1 \le \sum_{\substack{\mu\\ (\mu,\alpha')=1}} \sum_{\substack{Y_1\\ (Y_1,\alpha')=1}} \Big(\frac{\phi(\alpha')}{\alpha'}\Big(\Big[\sqrt{\frac{N}{\mu}}\Big] - Y_1\Big) + \phi(\alpha')\Big).$$

Arguing as in Lemma 5.1, we easily see that

$$\sum_{\substack{(N-2G(N))/\sqrt{N\mu}\le Y_1\le\sqrt{N/\mu}\\ (Y_1,\alpha')=1}} \Big(\Big[\sqrt{\frac{N}{\mu}}\Big] - Y_1\Big) \le 2\frac{\phi(\alpha')}{\alpha'}\Big(\frac{G(N)}{\sqrt{N\mu}} + \frac{\alpha'}{2}\Big)^2.$$

Using this, we see that

$$\sum_{\mu}\sum_{Y_1}\sum_{\substack{Y_2\\ \alpha+\mu Y_1Y_2 \text{ prime}}} 1$$

$$\le \sum_{\substack{\mu\le 100N/G(N)\\ (\mu,\alpha')=1}} \left(2\left(\frac{\phi(\alpha')}{\alpha'}\right)^2\left(\frac{G(N)}{\sqrt{N\mu}}+\frac{\alpha'}{2}\right)^2 + \phi(\alpha')\left(\frac{\phi(\alpha')}{\alpha'}\cdot\frac{2G(N)}{\sqrt{N\mu}}+\phi(\alpha')\right)\right)$$

$$= 2\left(\frac{\phi(\alpha')}{\alpha'}\right)^2 \sum_{\substack{\mu\le 100N/G(N)\\ (\mu,\alpha')=1}} \left(\frac{G(N)}{\sqrt{N\mu}}+\frac{\alpha'}{2}\right)\left(\frac{G(N)}{\sqrt{N\mu}}+\frac{3\alpha'}{2}\right).$$

Again, arguing as in Lemma 5.1, we may conclude that

$$\sum_{\substack{\mu\le 100N/G(N)\\ (\mu,\alpha')=1}} \frac{1}{\mu} \le \frac{\phi(\alpha')}{\alpha'}\left(\log\frac{100eN}{\alpha' G(N)}+\alpha'\right)$$

and

$$\sum_{\substack{\mu\le 100N/G(N)\\ (\mu,\alpha')=1}} \frac{1}{\sqrt{\mu}} \le \phi(\alpha') + 2\frac{\phi(\alpha')}{\alpha'}\sqrt{\frac{100N}{G(N)}} = \frac{\phi(\alpha')}{\alpha'}\left(20\sqrt{\frac{N}{G(N)}}+\alpha'\right).$$

Thus

$$\sum_{\mu\le 100N/G(N)}\sum_{Y_1}\sum_{\substack{Y_2\\ \alpha+\mu Y_1Y_2 \text{ prime}}} 1$$

$$\le 2\left(\frac{\phi(\alpha')}{\alpha'}\right)^3\frac{G(N)^2}{N}\left(\log\frac{100eN}{\alpha' G(N)}+\alpha'\right)$$

$$+ 4\alpha'\left(\frac{\phi(\alpha')}{\alpha'}\right)^3\frac{G(N)}{\sqrt{N}}\left(20\sqrt{\frac{N}{G(N)}}+\alpha'\right) + \frac{3}{2}\cdot\frac{\phi(\alpha')^3}{\alpha'}\left(100\frac{N}{G(N)}+\alpha'\right)$$

$$\le 2\left(\frac{\phi(\alpha')}{\alpha'}\right)^3\frac{G(N)^2}{N}\left(\log\frac{100e^{\alpha'+1}N}{\alpha' G(N)} + 42\frac{\alpha' N}{G(N)^{3/2}} + 76\frac{(\alpha')^2N^2}{G(N)^3}\right).$$

It is trivial to verify that

$$\left(\frac{\phi(\alpha')}{\alpha'}\right)^3 h(\alpha')^{-2} \le \frac{9}{8}.$$

Since $9/(8C^2)\le 0.3$ it follows that

$$\sum_{\mu\le 100N/G(N)}\sum_{Y_1}\sum_{\substack{Y_2\\ \alpha+\mu Y_1Y_2\ \text{prime}}} 1 \le 2C^2h(\alpha')^2\frac{G(N)^2}{N}\biggl(0.3\log\frac{100e^{\alpha'+1}N}{\alpha'G(N)}+13\frac{\alpha'N}{G(N)^{3/2}}+23\frac{(\alpha')^2N^2}{G(N)^3}\biggr).$$

Combining this with our earlier estimate for the terms with $Y_1\le\sqrt{G(N)}/10$, we see that

$$\sum_p k_D(p-\alpha)\le 2.01C^2h(\alpha')^2\frac{G(N)^2}{N}\log(4\log G(N))L(\alpha'),$$

where

$$\begin{aligned}L(\alpha')={}&1+1.03\phi(\alpha')\frac{\log(eN/(2\alpha'G(N)))}{\log(3.9G(N)^3/N^2)}\cdot\frac{\log(7N/G(N))}{\log(4\log G(N))}\\&+\frac{10\phi(\alpha')N}{G(N)^{3/2}\log 99}\cdot\frac{\log^2(\sqrt{G(N)}/10)}{\log(4\log G(N))}\\&+\frac{0.3}{\log(4\log G(N))}\log\frac{100e^{\alpha'+1}N}{\alpha'G(N)}\\&+13\frac{\alpha'N}{G(N)^{3/2}\log(4\log G(N))}+23\frac{(\alpha')^2N^2}{G(N)^3\log(4\log G(N))}.\end{aligned}$$

This completes the proof.

LEMMA 5.3. *Let* $P_1=\sum_{p\in\mathbf{P}}(2N-p)$ (*as in Lemma* 4.2) *and put*

$$L_1=0.51L(1)+0.45L(2)+0.27L(3)$$

and

$$L_2=0.31L(1)+0.25L(2)+0.24L(3)+0.19L(6).$$

Then

$$\begin{aligned}\sum_p\sum_\alpha(r_p(\alpha)-1)&\le P_1\frac{G(N)}{2N}\biggl(\log\frac{40G(N)^2}{T}+\frac{16}{\theta}L_1\frac{G(N)\log N}{N}\log(4\log G(N))\\&\quad+\frac{L_2}{2\theta}\cdot\frac{G(N)\log N}{N}\log(4\log G(N))\log\frac{3G(N)^2}{4T}\biggr).\end{aligned}$$

Proof. As noted at the beginning of this section,

$$\sum_{p\in\mathbf{P}}\sum_{\alpha\in\mathbf{J}_p}(r_p(\alpha)-1)\le\sum_p\sum_{\alpha\in[p-N,N]}k_D(\alpha)+\sum_p\sum_{\alpha\in\mathbf{J}_p}k_D(\alpha)k_D(p-\alpha).$$

By Lemma 5.1, we see that

$$\sum_p \sum_{\alpha\in[p-N,N]} k_D(\alpha) \le P_1 \frac{G(N)}{2N} \log\left(\frac{40G(N)^2}{T}\right).$$

Now, by Lemma 5.2,

$$\sum_p \sum_{\alpha\in\mathbf{J}_p} k_D(\alpha)k_D(p-\alpha)$$
$$= \sum_{\alpha\in[N-G(N),N]} k_D(\alpha) \sum_{\substack{2N-2G(N)\le p\\ p\le\min(2\alpha,\,2N-G(N))}} k_D(p-\alpha)$$
$$\le 2.01C^2\frac{G(N)^2}{N}\log(4\log G(N))\sum_\alpha k_D(\alpha)h(\alpha')^2L(\alpha').$$

It is trivial to verify the identity

$$\sum_\alpha k_D(\alpha)h(\alpha')^2L(\alpha')$$
$$= \sum_{\alpha\in[N-G(N),N]} L(\alpha)k_D(\alpha)h(6)^2L(6)$$
$$+ \sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,2)=1}} L(\alpha)k_D(\alpha)(h(3)^2L(3)-h(6)^2L(6))$$
$$+ \sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,3)=1}} L(\alpha)k_D(\alpha)(h(2)^2L(2)-h(6)^2L(6))$$
$$+ \sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,6)=1}} L(\alpha)k_D(\alpha)(h(1)^2L(1)-h(2)^2L(2)$$
$$-h(3)^2L(3)+h(6)^2L(6)).$$

From Lemma 5.1, we see that

$$\sum_{\alpha\in[N-G(N),N]} k_D(\alpha) \le \frac{1.01}{4}\cdot\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T} + 4.01\frac{G(N)^2}{N},$$
$$\sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,2)=1}} k_D(\alpha) \le \frac{1.01}{32}\cdot\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T} + 1.3\frac{G(N)^2}{N},$$
$$\sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,3)=1}} k_D(\alpha) \le \frac{2.02}{27}\cdot\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T} + 4.5\frac{G(N)^2}{N}$$

and

$$\sum_{\substack{\alpha\in[N-G(N),N]\\(\alpha,6)=1}} k_D(\alpha) \le \frac{1.01}{108}\cdot\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T} + 0.51\frac{G(N)^2}{N}.$$

From these inequalities and the identity above, we see that

$$\begin{aligned}
&\sum_\alpha k_D(\alpha)L(\alpha')h(\alpha')^2 \\
&\quad\le \frac{G(N)^2}{N}(0.51h(1)^2L(1) + 4h(2)^2L(2) + 0.8h(3)^2L(3) - 1.28h(6)^2L(6)) \\
&\qquad + 1.01\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T}\biggl(\frac{h(1)^2L(1)}{108} + \Bigl(\frac{2}{27}-\frac{1}{108}\Bigr)h(2)^2L(2) \\
&\qquad + \Bigl(\frac{1}{32}-\frac{1}{108}\Bigr)h(3)^2L(3) + \Bigl(\frac14+\frac{1}{108}-\frac{1}{32}-\frac{2}{27}\Bigr)h(6)^2L(6)\biggr) \\
&\quad\le L_1\frac{G(N)^2}{N} + 0.03\frac{G(N)^2}{N}\log\frac{3G(N)^2}{T}L_2,
\end{aligned}$$

where L_1 and L_2 were defined in the statement of the lemma. Hence

$$\begin{aligned}
\sum_p\sum_\alpha(r_p(\alpha)-1) &\le P_1\frac{G(N)}{2N}\log\frac{40G(N)^2}{T} + 8L_1\frac{G(N)^4}{N^2}\log(4\log G(N)) \\
&\qquad + 0.25L_2\frac{G(N)^4}{N^2}\log(4\log G(N))\log\frac{3G(N)^2}{4T}.
\end{aligned}$$

Since $P_1 = \sum_{p\in\mathbf{P}}(2N-p) \ge \theta\frac{G(N)^2}{\log N}$, it follows that

$$\begin{aligned}
&\sum_p\sum_\alpha(r_p(\alpha)-1) \\
&\qquad\le P_1\frac{G(N)}{2N}\biggl(\log\frac{40G(N)^2}{T} + \frac{16}{\theta}L_1\frac{G(N)\log N}{N}\log(4\log G(N)) \\
&\qquad\qquad + \frac{L_2}{2\theta}\cdot\frac{G(N)\log N}{N}\log(4\log G(N))\log\frac{3G(N)^2}{4T}\biggr).
\end{aligned}$$

This proves the lemma.

6. Completion of the proof. We first make our choice of $G(N)$ and record the value of θ permitted by this choice. (Recall that θ was defined in the paragraph preceding Lemma 4.2.)

LEMMA 6.1. *If $e^{28} \le N < e^{50}$ and $G(N) = N/(5(\log N)^2)$, then $\theta = 1/3$ is permissible. If $e^{50} \le N < e^{2000}$ and $G(N) = N/(30\log N)$ then $\theta = 1/2$ is permissible. If $N \ge e^{2000}$ and $G(N) = 40N/(\log N)^2$ then $\theta = 1/2$ is* *permissible. Finally, our choices for $G(N)$ and θ satisfy the requirements $(5N)^{2/3} \le G(N) \le N/1000$ and $\theta \ge 1/4$.*

Proof. Let $\varepsilon(x)$ be such that $|\psi(x) - x| \le x\varepsilon(x)$. From Theorem 6 of Rosser and Schoenfeld [8], we see that $|\psi(x)-\theta(x)| \le 1.1\sqrt{x}$ when $x \ge 1000$. (Here, as usual, $\theta(x) = \sum_{p\le x} \log p$.) Thus if $x \ge y > 1000$, it follows that

$$\begin{aligned}\pi(x) - \pi(y) &\ge \frac{\theta(x) - \theta(y)}{\log x} \ge \frac{\psi(x) - \psi(y) - 2.2\sqrt{x}}{\log x}\\ &\ge \frac{x - y - x\varepsilon(x) - y\varepsilon(y) - 2.2\sqrt{x}}{\log x}.\end{aligned}$$

The table on page 267 of Rosser and Schoenfeld [8] gives permissible values for $\varepsilon(x)$ when $e^{18} \le x \le e^{4000}$. From this we may easily deduce the lemma in the range $e^{28} \le N \le e^{2000}$.

For the range $N \ge e^{2000}$, we note that Theorem 8 of Rosser and Schoenfeld [8] enables us to take $\varepsilon(x) = 8.6853x/\log^2 x$. The remaining assertions are trivial.

LEMMA 6.2. *Theorem* 1.1 *holds in the range* $e^{28} \le N < e^{50}$.

Proof. In this range we know that $G(N) = N/(5\log^2 N)$ and $\theta = 1/3$. From the definition of L_1, we easily see that

$$\begin{aligned}&\log(4\log G(N))L_1\\ &\quad\le 1.23\log(4\log G(N)) + 1.51\log(35(\log N)^2)\frac{\log(5e(\log N)^2/2)}{\log(3.9N/(125\log^6 N))}\\ &\qquad + 15\frac{5^{1.5}\log^3 N}{\sqrt{N}\log 99}\log^2\frac{\sqrt{G(N)}}{10} + 0.4\log(500e^2\log^2 N)\\ &\qquad + 39.39\frac{5^{1.5}\log^3 N}{\sqrt{N}} + 165\frac{5^3\log^6 N}{N}\\ &\quad\le 1.23\log(4\log G(N)) + 64 \le 71.\end{aligned}$$

Similarly we may verify that $\log(4\log G(N))L_2 \le 71$. From Lemma 5.3, it follows that

$$\begin{aligned}\sum_p\sum_\alpha(r_p(\alpha) - 1) &\le P_1\frac{G(N)}{2N}\biggl(\log\frac{40G(N)^2}{T} + 48\cdot 71\frac{G(N)\log N}{N}\\ &\qquad + 1.5\cdot 71\frac{G(N)\log N}{N}\log\frac{3G(N)^2}{4T}\biggr)\\ &\le \frac{P_1}{10\log^2 N}\biggl(\log\frac{40G(N)^2}{T} + 25 + 0.8\log\frac{3G(N)^2}{4T}\biggr)\end{aligned}$$

$$\le \frac{P_1}{10\log^2 N}\left(\log\frac{40G(N)^2D^2}{N^2}+25+0.8\log\frac{3G(N)^2D^2}{4N^2}\right)$$
$$\le \frac{P_1}{10\log^2 N}(3.6\log D).$$

However, from Lemma 4.2, we see that

$$\sum_p\sum_\alpha(r_p(\alpha)-1)\ge\frac{\theta P_1}{8\log N}\log D=\frac{P_1\log D}{24\log N}.$$

This is clearly a contradiction which proves the lemma.

Lemma 6.3. *Theorem* 1.1 *holds in the range* $e^{50}\le N<e^{2000}$.

Proof. In this range of N, $G(N)=N/(30\log N)$ and $\theta=1/2$. As in Lemma 6.2 we see that

$$\begin{aligned}\log(4\log G(N))L_1 &\le 1.23\log(4\log G(N))+1.51\log(210\log N)\frac{\log(15e\log N)}{\log(3.9N/(30^3\log^3 N))}\\ &\quad+15\frac{30^{1.5}\log^{1.5}N}{\sqrt{N}\log 99}\log^2\frac{\sqrt{G(N)}}{10}+0.4\log(3000e^2\log N)\\ &\quad+39.39\frac{30^{1.5}\log^{1.5}N}{\sqrt{N}}+165\frac{30^3\log^3 N}{N}\\ &\le 1.23\log(4\log G(N))+10.7\le 22.\end{aligned}$$

Similarly we see that $\log(4\log G(N))L_2\le 22$. From Lemma 5.3 it follows that

$$\begin{aligned}\sum_p\sum_\alpha(r_p(\alpha)-1) &\le\frac{P_1}{60\log N}\left(\log\frac{40G(N)^2}{T}+32\frac{22}{30}+\frac{22}{30}\log\frac{3G(N)^2}{4T}\right)\\ &\le\frac{P_1}{60\log N}\left(1.74\log\frac{G(N)^2}{T}+27.2\right)\\ &\le\frac{P_1}{60\log N}\left(1.74\log\frac{G(N)^2D^2}{N^2}+27.2\right)\\ &\le\frac{P_1}{60\log N}(3.48\log D+1.75)\le\frac{P_1}{60\log N}(3.72\log D)\end{aligned}$$

(the last inequality holds since $D\ge N/G(N)$). However, by Lemma 4.2,

$$\sum_p\sum_\alpha(r_p(\alpha)-1)\ge\frac{\theta P_1}{8\log N}\log D=\frac{P_1\log D}{16\log N},$$

which is a contradiction.

Lemma 6.4. *Theorem* 1.1 *holds in the range* $N>e^{2000}$.

P r o o f. In this range of N, $G(N) = 40N/\log^2 N$ and $\theta = 1/2$. Arguing as in the previous two lemmata we see that $\log(4\log G(N))L_1 \le 2.03\log(4\log N) + 4$ and $\log(4\log G(N))L_2 \le 2.03\log(4\log N) + 4$. Hence by Lemma 5.3,

$$\begin{aligned}\sum_p \sum_\alpha (r_p(\alpha) - 1) &\le \frac{20P_1}{\log^2 N}\bigg(\log\frac{40G(N)^2}{T} + 32(2.03\log(4\log N)+4)\frac{G(N)\log N}{N} \\ &\quad + (2.03\log(4\log N)+4)\frac{G(N)\log N}{N}\log\frac{3G(N)^2}{4T}\bigg) \\ &\le \frac{20P_1}{\log^2 N}\bigg(\log\frac{40G(N)^2}{T} + 14.25 + 0.45\log G(N)^2T\bigg) \\ &\le \frac{20P_1}{\log^2 N}\bigg(1.45\log\frac{G(N)^2}{T} + 18\bigg) \\ &\le \frac{20P_1}{\log^2 N}\bigg(1.45\log\frac{G(N)^2D^2}{N^2} + 18\bigg) \\ &\le \frac{20P_1}{\log^2 N}(2.9\log D).\end{aligned}$$

By Lemma 4.2,

$$\sum_p \sum_\alpha (r_p(\alpha) - 1) \ge \frac{\theta P_1\log D}{8\log N} = \frac{P_1\log D}{16\log N},$$

which is a contradiction. This proves the lemma.

Since $2.22\cdot 10^{12} \ge e^{28}$, we see that the above lemmata and the results of Section 3 cover all the values of N. This completes the proof of Theorem 1.1.

References

[1] R. D. Boyle, *On a problem of R. L. Graham*, Acta Arith. 34 (1978), 163–177.

[2] F. Y. Cheng and C. Pomerance, *On a conjecture of R. L. Graham*, Rocky Mountain J. 24 (1994), 961–975.

[3] C. Cobeli, M. Vâjâitu and A. Zaharescu, *Graham's conjecture under Riemann hypothesis*, J. Number Theory 31 (1989), 80–87.

[4] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monographie No. 28 de L'Enseignement Mathématique, Genève, 1980.

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

[6] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika 20 (1973), 119–134.

[7] H. Riesel, *Prime Numbers and Computer Methods for Factorization*, Progr. Math. 57, Birkhäuser, 1985.
[8] J. B. Rosser and L. Schoenfeld, *Sharper bounds for the Chebyshev functions* $\theta(x)$ *and* $\psi(x)$, Math. Comp. 29 (1975), 243–269.
[9] M. Szegedy, *The solution of Graham's greatest common divisor problem*, Combinatorica 6 (1986), 67–71.
[10] W. Y. Vélez, *Some remarks on a number theoretic problem of Graham*, Acta Arith. 32 (1977), 233–238.
[11] R. Winterle, *A problem of R. L. Graham in Combinatorial Number Theory*, in: Proc. of the Louisiana Conference on Combinatorics, Baton Rouge, 1970, 357–361.
[12] A. Zaharescu, *On a conjecture of Graham*, J. Number Theory 27 (1987), 33–40.

Institute for Mathematical Sciences
Tharamani P. O.
Madras 600 113, India

Department of Mathematics
Princeton University
Princeton, New Jersey 08544
U.S.A.

Received on 27.10.1993
and in revised form on 24.9.1995 (2508)