Unit fractions

Thomas F. Bloom and Bhavik Mehta

June 13, 2022

Introduction

This is an interactive blueprint to help with the formalisation of the main result of (`https://arxiv.org/abs/2112.03726`): that if $A \subset \mathbb{N}$ has positive density then there are distinct $n_1, \dots, n_k \in A$ such that $\frac{1}{n_1} + \dots + \frac{1}{n_k} = 1$.

(And other more precise results formulated in that paper.) For context, background, references, and so on, we refer to the original paper (`https://arxiv.org/abs/2112.03726`). This blueprint will (once finished) be a complete *mathematical* guide to the entire proof, and indeed the proofs will be in many places expanded and explained more fully, to help the formalisation process.

Nonetheless, we will be sparing with non-mathematical remarks, and will not usually trouble to explain the context of a particular lemma, or what it's 'really saying', and so on – we will present everything necessary to formalise the proof in Lean, no more and no less.

The actual Lean code can be found at (`https://github.com/leanprover-community/mathlib/blob/unit-fractions/src/number_theory/unit_fractions.lean`). If you'd like to contribute in any way, or have any questions about this project, please email me at bloom@maths.ox.ac.uk.

This blueprint is adapted from the blueprint created by Patrick Massot for the Sphere Eversion project (`https://github.com/leanprover-community/sphere-eversion`).

This blueprint uses Patrick Massot's leanblueprint plugin (`https://github.com/PatrickMassot/leanblueprint`) for plasTeX (`http://plastex.github.io/plastex/`).

Chapter 1

Definitions

Some definitions, local to this paper, that occur frequently.

Definition 1.1. *For any finite $A \subset \mathbb{N}$*

$$R(A) = \sum_{n \in A} \frac{1}{n}.$$

Definition 1.2. *For any finite $A \subset \mathbb{N}$ and prime power q we define*

$$A_q = \{n \in A : q \mid n \text{ and } (q, n/q) = 1\}$$

and let $\mathcal{Q}_A$ be the set of all prime powers q such that A_q is non-empty (i.e. those p^r such that $p^r \| n$ for some $n \in A$).

Definition 1.3. *For any finite $A \subset \mathbb{N}$ and prime power $q \in \mathcal{Q}_A$ we define*

$$R(A; q) = \sum_{n \in A_q} \frac{q}{n}.$$

Definition 1.4. *For any finite set $A \subset \mathbb{N}$, $K \in \mathbb{R}$ and interval I, we define $\mathcal{D}_I(A; K)$ to be the set of those $q \in \mathcal{Q}_A$ such that*

$$\#\{n \in A_q : \text{no element of } I \text{ is divisible by } n\} \ < \ K/q.$$

Chapter 2

Basic Estimates

This section contains standard estimates from analytic number theory that will be required.

Lemma 2.1. *For any $X \geq 3$*

$$\sum_{n \leq X} \omega(n) = X \log\log X + O(X).$$

Proof. Since $\omega(n) = \sum_{p \leq n} 1_{p|n}$, the left-hand side equals, after a change in the order of summation,

$$\sum_{p \leq X} \sum_{n \leq X} 1_{p|n} = \sum_{p \leq X} \left\lfloor \frac{X}{p} \right\rfloor.$$

Since $\lfloor x \rfloor = x + O(1)$, this is equal to

$$X \sum_{p \leq X} \frac{1}{p} + O(\pi(X)) = X \log\log X + O(X),$$

using Lemma 2.7 and the trivial estimate $\pi(X) \ll X$. $\square$

Lemma 2.2. *For any $X \geq 3$*

$$\sum_{n \leq X} \omega(n)^2 \leq X(\log\log X)^2 + O(X \log\log X).$$

Proof. Since $\omega(n) = \sum_{p \leq n} 1_{p|n}$, the left-hand side is equal to, after expanding the sum and rearranging,

$$\sum_{p,q \leq X} \sum_{n \leq X} 1_{p|n} 1_{q|n}.$$

The part of the sum where $p = q$ is

$$\sum_{p \leq X} \sum_{n \leq X} \lfloor X/p \rfloor = X \log\log X + O(X),$$

as in the proof of Lemma 2.1. If $p \neq q$, then $p \mid n$ and $q \mid n$ if and only if $pq \mid n$, and so the sum over n is bounded above by

$$\sum_{n \leq X} 1_{pq|n} = \lfloor X/pq \rfloor \leq X/pq.$$

Therefore

$$\sum_{\substack{p,q\leq X\\ p\neq q}}\sum_{n\leq X}1_{p|n}1_{q|n}\leq X\sum_{\substack{p,q\leq X\\ p\neq q}}\frac{1}{pq}\leq X\sum_{p,q\leq X}\frac{1}{pq}\leq X\left(\sum_{p\leq X}\frac{1}{p}\right)^2.$$

By Lemma 2.7 this is $X(\log\log X+O(1))^2=X(\log\log X)^2+O(X\log\log X)$. The lemma follows by combining the estimates on the two parts of the sum. □

Lemma 2.3 (Turán's estimate)**.** *For any $X\geq 3$*

$$\sum_{n\leq X}(\omega(n)-\log\log X)^2\ll X\log\log X.$$

Proof. The left-hand side equals

$$\sum_{n\leq X}\omega(n)^2-2\log\log X\sum_{n\leq X}\omega(n)+\lfloor X\rfloor(\log\log X)^2.$$

The first summand is at most, by Lemma 2.2,

$$X(\log\log X)^2+O(X\log\log X).$$

The second summand is equal to, by Lemma 2.1,

$$-2\log\log X(X\log\log X+O(X))=-2X(\log\log X)^2+O(X\log\log X).$$

The third summand is equal to

$$(X+O(1))(\log\log X)^2=X(\log\log X)^2+O(X\log\log X).$$

Therefore the main terms cancel, and

$$\sum_{n\leq X}(\omega(n)-\log\log X)^2\leq O(X\log\log X)$$

as required. □

Lemma 2.4 (Chebyshevs' estimate)**.** *For any $X\geq 3$*

$$\pi(X)\ll\frac{X}{\log X}.$$

Proof. □

Lemma 2.5 (Divisor bound)**.** *For any ϵ such that $0<\epsilon\leq 1$, if n is sufficiently large depending on ϵ, then*

$$\tau(n)\leq n^{(1+\epsilon)\frac{\log 2}{\log\log n}}.$$

Proof. We first show that, for any real $K\geq 2$,

$$\tau(n)\leq n^{1/K}K^{2^K}.$$

Write n as the product of unique prime powers $n=p_1^{k_1}\cdots p_r^{k_r}$, so that

$$\frac{\tau(n)}{n^{1/K}}=\frac{\prod_{i=1}^r(k_i+1)}{\prod_{i=1}^r p_i^{k_i/K}}=\prod_{i=1}^r\frac{k_i+1}{p_i^{k_i/K}}.$$

If $p_i > 2^K$ then

$$\frac{k_i+1}{p_i^{k_i/K}} \le \frac{k_i+1}{2^{k_i}} \le 1,$$

since $1+k \le 2^k$ for all integer $k \ge 0$ by Bernoulli's inequality. Therefore

$$\prod_{i=1}^{r} \frac{k_i+1}{p_i^{k_i/K}} \le \prod_{\substack{1\le i\le r\\ p_i<2^K}} \frac{k_i+1}{p_i^{k_i/K}}.$$

If $p_i < 2^K$ then, since $p_i \ge 2$,

$$\frac{k_i+1}{p^{k_i/K}} \le \frac{k_i+1}{2^{k_i/K}} \le \frac{k_i+1}{k_i/K+1/2},$$

using the fact that $x+1/2 \le 2^x$ for all $x \ge 0$. Since $K \ge 2$ the denominator here is $\ge (1+k_i)/K$, and so $\frac{k_i+1}{p^{k_i/K}} \le K$. Therefore

$$\frac{\tau(n)}{n^{1/K}} \le \prod_{\substack{1\le i\le r\\ p_i<2^K}} K \le K^{\pi(2^K)} \le K^{2^K},$$

as required.

The second part of the proof is to apply the first part with

$$K = (1+\epsilon/2)^{-1}\frac{\log\log n}{\log 2}.$$

(The right-hand side tends to ∞ as $n \to \infty$, so for sufficiently large n we have $K \ge 2$ as required.)

Note that $2^K = (\log n)^{\frac{2}{2+\epsilon}} \le (\log n)^{1-\epsilon}$ (since $1-\epsilon > \frac{2}{2+\epsilon}$, since $\epsilon \le 1$) and

$$\log K \le \log\log\log n.$$

Taking logarithms of the inequality in the first part,

$$\log\tau(n) \le \frac{\log n}{K} + 2^K \log K \le \log n\left(\frac{1}{K} + \frac{\log\log\log n}{(\log n)^{\epsilon}}\right)$$

$$= \log n \frac{\log 2}{\log\log n}\left(1+\epsilon/2+\frac{(\log\log\log n)(\log\log n)}{(\log 2)(\log n)^{\epsilon}}\right).$$

For any fixed $\epsilon > 0$, the function $\frac{(\log\log\log n)(\log\log n)}{(\log 2)(\log n)^{\epsilon}} \to 0$ as $n \to \infty$, so for sufficiently large n it is $\le \epsilon/2$, and hence

$$\log\tau(n) \le \log n \frac{\log 2}{\log\log n}(1+\epsilon)$$

as required.

□

Lemma 2.6 (Mertens' estimate)**.** *There exists a constant c such that*

$$\sum_{q\le X} \frac{1}{q} = \log\log X + c + O(1/\log X),$$

where the sum is restricted to prime powers.

Proof. □

Lemma 2.7 (Mertens' estimate, just for primes)**.** *There exists a constant c such that*

$$\sum_{p \leq X} \frac{1}{p} = \log\log X + c + O(1/\log X),$$

where the sum is restricted to primes.

Proof. □

Lemma 2.8 (Mertens' product estimate)**.** *For any $X \geq 2$,*

$$\prod_{p \leq X} \left(1 - \frac{1}{p}\right)^{-1} \asymp \log X.$$

Proof. □

Lemma 2.9 (Sieve of Eratosthenes-Legendre)**.** *For any $x, y \geq 0$ and $u \geq v \geq 1$*

$$\#\{n \in (x, x+y] : p \mid n \implies p \notin [u, v]\} = y \prod_{u \leq p \leq v} \left(1 - \frac{1}{p}\right) + O(2^v).$$

Proof. Let $P = \prod_{u \leq p \leq v} p$. The left-hand side of the estimate in the lemma can be written as

$$\sum_{x \leq n < x+y} 1_{(n,P)=1}.$$

Using the identity $\sum_{d|m} \mu(d) = 1$ if $m = 1$ and 0 otherwise, the fact that $d \mid (n, P)$ if and only if $d \mid n$ and $d \mid P$, and that $\sum_{n \leq z} 1_{d|n} = \lfloor \frac{z}{d} \rfloor$ for any $z \geq 0$,

$$\begin{aligned}
\sum_{x < n \leq x+y} 1_{(n,P)=1} &= \sum_{x < n \leq x+y} \sum_{d|(n,P)} \mu(d) \\
&= \sum_{x < n \leq x+y} \sum_{\substack{d|n \\ d|P}} \mu(d) \\
&= \sum_{d|P} \mu(d) \sum_{x < n \leq x+y} 1_{d|n} \\
&= \sum_{d|P} \mu(d) \left(\left\lfloor \frac{x+y}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor \right).
\end{aligned}$$

Since $\lfloor X \rfloor = X + O(1)$ for any X, we have

$$\left\lfloor \frac{x+y}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor = \frac{y}{d} + O(1)$$

for any x, y, d, and hence

$$\sum_{x < n \leq x+y} 1_{(n,P)=1} = y \sum_{d|P} \frac{\mu(d)}{d} + O(\sum_{d|P} 1).$$

We have $\sum_{d|P} 1 = \tau(P) = 2^k$ where k is the number of primes in $[u, v]$, which is at most v, so the error term here is $O(2^v)$. Finally, expanding out the product shows that

$$\sum_{d|P} \frac{\mu(d)}{d} = \prod_{p \in [u,v]} \left(1 - \frac{1}{p}\right).$$

Inserting this into the above finishes the proof. $\square$

Chapter 3

Deduction of the main results

This section contains the deductions of the headline results from the main technical proposition, Propostion 6.6.

Theorem 3.1 (Solution in sets of positive density)**.** *If $A \subset \mathbb{N}$ has positive upper density then there is a finite $S \subset A$ such that $\sum_{n \in S} \frac{1}{n} = 1$.*

Proof. Suppose $A \subset \mathbb{N}$ has upper density $\delta > 0$. Let $y = C_1/\delta$ and $z = \delta^{-C_2\delta^{-2}}$, where C_1, C_2 are two absolute constants to be determined later. It suffices to show that there is some $d \in [y, z]$ and finite $S \subset A$ such that $R(S) = 1/d$. Indeed, given such an S we can remove it from A and still have an infinite set of upper density δ, so we can find another $S' \subset A \backslash S$ with $R(S') = 1/d'$ for some $d' \in [y, z]$, and so on. After repeating this process at least $\lceil z - y \rceil^2$ times there must be some $d \in [y, z]$ with at least d disjoint $S_1, \dots, S_d \subset A$ with $R(S_i) = 1/d$. Taking $S = S_1 \cup \cdots \cup S_d$ yields $R(S) = 1$ as required.

By definition of the upper density, there exist arbitrarily large N such that $|A \cap [1, N]| \geq \frac{\delta}{2} N$. The number of $n \in [1, N]$ divisible by some prime power $q \geq N^{1-6/\log\log N}$ is

$$\ll N \sum_{N^{1-6/\log\log N} < q \leq N} \frac{1}{q} \ll \frac{N}{\log\log N}$$

by Mertens' estimate Lemma 2.6. Further, by Turán's estimate Lemma 2.3

$$\sum_{n \leq N} (\omega(n) - \log\log N)^2 \ll N \log\log N,$$

the number of $n \in [1, N]$ that do not satisfy

$$\tfrac{99}{100} \log\log N \leq \omega(n) \leq 2 \log\log N \tag{3.1}$$

is $\ll N/\log\log N$. Finally, provided we choose C_2 sufficiently large in the definition of z, Lemma 3.5 ensures that the proportion of all $n \in \{1, \dots, N\}$ not divisible by at least two distinct primes $p_1, p_2 \in [y, z]$ with $4p_1 < p_2$ is at most $\frac{\delta}{8} N$, say.

In particular, provided N is chosen sufficiently large (depending only on δ), we may assume that $|A_N| \geq \frac{\delta}{4} N$, where $A_N \subset A$ is the set of those $n \in A \cap [N^{1-1/\log\log N}, N]$ which satisfy conditions (2)-(4) of Proposition 6.6. Since $|A_N| \geq \frac{\delta}{4} N$,

$$R(A_N) \gg -\log(1 - \delta/4) \gg \delta.$$

In particular, since $y = C_1/\delta$ for some suitably large constant $C_1 > 0$, we have that $R(A_N) \geq 4/y$, say. All of the conditions of Proposition 6.6 are now satisfied (provided N is chosen sufficiently large in terms of δ), and hence there is some $S \subset A_N \subset A$ such that $R(S) = 1/d$ for some $d \in [y, z]$, which suffices as discussed above. $\square$

Theorem 3.2 (Solution in sets of positive logarithmic density, quantitative version)**.** *There is a constant $C > 0$ such that the following holds. If $A \subset \{1, \ldots, N\}$ and*

$$\sum_{n \in A} \frac{1}{n} \geq C \frac{\log\log\log N}{\log\log N} \log N$$

then there is an $S \subset A$ such that $\sum_{n \in S} \frac{1}{n} = 1$.

Proof. Let $C \geq 2$ be an absolute constant to be chosen shortly, and for brevity let $\epsilon = \log\log\log N/\log\log N$, so that we may assume that $R(A) \geq C\epsilon \log N$. Since $\sum_{n \leq X} \frac{1}{n} \ll \log X$, if $A' = A \cap [N^\epsilon, N]$ we have (assuming C is sufficiently large) $R(A') \geq \frac{C}{2}\epsilon \log N$.

Let X be those integers $n \in [1, N]$ not divisible by any prime $p \in [5, (\log N)^{1/1200}]$. Lemma 3.4 implies that, for any $x \geq \exp(\sqrt{\log N})$,

$$|X \cap [x, 2x)| \ll \frac{x}{\log\log N}$$

and hence, by partial summation,

$$\sum_{\substack{n \in X \\ n \in [\exp(\sqrt{\log N}), N]}} \frac{1}{n} \ll \frac{\log N}{\log\log N}.$$

Similarly, if Y is the set of those $N \in [1, N]$ such that $\omega(n) < \frac{99}{100}\log\log N$ or $\omega(n) \geq \frac{101}{100}\log\log N$ then Turán's estimate Lemma 2.3

$$\sum_{n \leq x} (\omega(n) - \log\log n)^2 \ll x \log\log x$$

implies that $|Y \cap [x, 2x)| \ll x/\log\log N$ for any $N \geq x \geq \exp(\sqrt{\log N})$, and so

$$\sum_{\substack{n \in Y \\ n \in [\exp(\sqrt{\log N}), N]}} \frac{1}{n} \ll \frac{\log N}{\log\log N}.$$

In particular, provided we take C sufficiently large, we can assume that $R(A' \backslash (X \cup Y)) \geq \frac{C}{4}\epsilon \log N$, say.

Let $\delta = 1 - 1/\log\log N$, and let $N_i = N^{\delta^i}$, and $A_i = (A' \backslash (X \cup Y)) \cap [N_{i+1}, N_i]$. Since $N_i \leq N^{e^{-i/\log\log N}}$ and A' is supported on $n \geq N^\epsilon$, the set A_i is empty for $i > \log(1/\epsilon)\log\log N$, and hence by the pigeonhole principle there is some i such that

$$R(A_i) \geq \frac{C}{8} \frac{\epsilon \log N}{(\log\log N)\log(1/\epsilon)}.$$

By construction, $A_i \subset [N_{i+1}, N_i] \subset [N_i^{1-1/\log\log N_i}, N_i]$, and every $n \in A_i$ is divisible by some prime p with $5 \leq p \leq (\log N)^{1/1200} \leq (\log N_i)^{1/500}$. Furthermore, every $n \in A_i$ satisfies $\omega(n) \geq \frac{99}{100}\log\log N \geq \frac{99}{100}\log\log N_i$ and $\omega(n) \leq \frac{101}{99}\log\log N \leq 2\log\log N_i$.

Finally, it remains to discard the contribution of those $n \in A_i$ divisible by some large prime power $q > N_i^{1-6/\log\log N_i}$. The contribution to $R(A_i)$ of all such n is at most

$$\sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \sum_{\substack{n \le N_i \\ q|n}} \frac{1}{n} \ll \sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \frac{\log(N_i/q)}{q}$$

$$\ll \frac{\log N_i}{\log\log N_i} \sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \frac{1}{q} \ll \frac{\log N}{(\log\log N)^2},$$

using Lemma 2.6. Provided we choose C sufficiently large, this is $\le R(A_i)/2$, and hence, if $A_i' \subset A_i$ is the set of those n divisible only by prime powers $q \le N_i^{1-6/\log\log N_i}$, then $R(A_i') \ge (\log N)^{1/200}$, say. All of the conditions of Corollary 3.3 are now met, and hence there is some $S \subset A_i' \subset A$ such that $R(S) = 1$, as required. □

Corollary 3.3 (Useful Technical Corollary)**.** *Suppose N is sufficiently large and $A \subset [N^{1-1/\log\log N}, N]$ is such that*

1. *$R(A) \ge 2(\log N)^{1/500}$,*
2. *every $n \in A$ is divisible by some prime p satisfying $5 \le p \le (\log N)^{1/500}$,*
3. *every prime power q dividing some $n \in A$ satisfies $q \le N^{1-6/\log\log N}$, and*
4. *every $n \in A$ satisfies*
$$\tfrac{99}{100} \log\log N \le \omega(n) \le 2\log\log N.$$

There is some $S \subset A$ such that $R(S) = 1$.

Proof. Let k be maximal such that there are disjoint $S_1, \ldots, S_k \subset A$ where, for each $1 \le i \le k$, there exists some $d_i \in [1, (\log N)^{1/500}]$ such that $R(S_i) = 1/d_i$. Let $t(d)$ be the number of S_i such that $d_i = d$. If there is any d with $t(d) \ge d$ then we are done, taking S to be the union of any d disjoint S_j with $R(S_j) = 1/d$. Otherwise,

$$\sum_i R(S_i) = \sum_{1 \le d \le (\log N)^{1/500}} \frac{t(d)}{d} \le (\log N)^{1/500},$$

and hence if $A' = A \backslash (S_1 \cup \cdots \cup S_k)$ then $R(A') \ge (\log N)^{1/500}$.

We may now apply Proposition 6.6 with $y = 1$ and $z = (\log N)^{1/500}$ – note that condition (2) of Proposition 6.6 follows from condition (2) of the hypotheses with $d_1 = 1$ and $d_2 = p \in [5, (\log N)^{1/500}]$ some suitable prime divisor. Thus there exists some $S' \subset A'$ such that $R(S') = 1/d$ for some $d \in [1, (\log N)^{1/500}]$, contradicting the maximality of k. □

3.1 Sieve Lemmas

Lemma 3.4 (Sieve Estimate 1)**.** *Let N be sufficiently large and z, y be two parameters such that $\log N \ge z > y \ge 3$. If X is the set of all those integers not divisible by any prime in $p \in [y, z]$ then*

$$|X \cap [N, 2N)| \ll \frac{\log y}{\log z} N.$$

Proof. Lemma 2.9 yields

$$|X \cap [N, 2N)| = \prod_{y \le p \le z} \left(1 - \frac{1}{p}\right) N + O(2^z).$$

Mertens' estimate 2.8 yields

$$\prod_{p \le z} \left(1 - \frac{1}{p}\right)^{-1} \gg \log z$$

and

$$\prod_{p \le y} \left(1 - \frac{1}{p}\right)^{-1} \ll \log y,$$

whence

$$\prod_{y \le p \le z} \left(1 - \frac{1}{p}\right)^{-1} \gg \frac{\log z}{\log y},$$

and hence

$$\prod_{y \le p \le z} \left(1 - \frac{1}{p}\right) \ll \frac{\log y}{\log z}.$$

Therefore the first term above is $\ll \frac{\log y}{\log z} N$. The second term is

$$\ll 2^z \le 2^{\log N} = N^{\log 2} \ll \frac{N}{\log N} \ll \frac{\log y}{\log z} N,$$

and the result follows. $\square$

Lemma 3.5 (Sieve Estimate 2)**.** *Let N be sufficiently large and z, y be two parameters such that $(\log N)^{1/2} \ge z > 4y \ge 8$. If $Y \subset [1, N]$ is the set of all those integers divisible by at least two distinct primes $p_1, p_2 \in [y, z]$ where $4p_1 < p_2$ then*

$$|\{1, \dots, N\} \backslash Y| \ll \left(\frac{\log y}{\log z}\right)^{1/2} N.$$

Proof. Let $w \in (4y, z)$ be some parameter to be chosen later. Lemma 3.4 implies that the number of $n \in \{1, \dots, N\}$ not divisible by any prime $p \in [w, z]$ is $\ll \frac{\log w}{\log z} N$.

Similarly, for any $p \in [w, z]$, the number of those $n \in [1, N]$ divisible by p and no prime $q \in [y, p/4)$ is

$$\ll \frac{\log y}{\log p} \frac{N}{p}.$$

It follows that the number of $n \in \{1, \dots, N\} \backslash Y$ is

$$\ll \left(\frac{\log w}{\log z} + \log y \sum_{p \ge w} \frac{1}{p \log p} \right) N.$$

By partial summation, $\sum_{p \ge w} \frac{1}{p \log p} \ll 1/\log w$, and hence

$$|\{1, \dots, N\} \backslash Y| \ll \left(\frac{\log w}{\log z} + \frac{\log y}{\log w} \right) N.$$

Choosing $w = \exp\left(\sqrt{(\log y)(\log z)}\right)$ completes the proof. $\square$

Chapter 4

Fourier Analysis

This section contains the part of the proof that uses Fourier analysis to reduce finding solutions to a combinatorial problem involving divisors.

In this section, we write $[A]$ for the lowest common multiple of A (if A is a finite set of naturals).

4.1 Local definitions

Definition 4.1. *For any finite set A of natural numbers and integer $k \geq 1$ we write $F(A;k)$ for the count of the number of subsets $S \subset A$ such that $kR(S)$ is an integer.*

Definition 4.2. *Let $J(A) = (-[A]/2, [A]/2] \cap \mathbb{Z} \backslash \{0\}$.*

Definition 4.3. *For any finite set of naturals B and integer t, we define $C(B;t) = \prod_{n \in B} |\cos(\pi t/n)|$.*

Definition 4.4. *For any finite set of naturals A and integer $k \geq 1$, and real $K > 0$, we define the 'major arc' corresponding to $t \in \mathbb{Z}$ as*

$$\mathfrak{M}(t; A, k, K) = \{h \in J(A) : |h - t[A]/k| \leq K/2k\}$$

Let $\mathfrak{M}(A, k, K) = \cup_{t \in \mathbb{Z}} \mathfrak{M}(t; A, k, K)$.

Definition 4.5. *Let $I_h(K, k)$ be the interval of length K centred at kh, and let $\mathfrak{m}_1(A, k, K, \delta)$ be those $h \in J(A) \backslash \mathfrak{M}(A, k, K)$ such that*

$$\#\{n \in A : \textit{no element of } I_h(K, k) \textit{ is divisible by } n\} \geq \delta.$$

Let $\mathfrak{m}_2(A, k, K, \delta)$ be the rest of $J(A) \backslash \mathfrak{M}(A, k, K)$.

4.2 Precursor general lemmas

Lemma 4.6. *For any $n, m \in \mathbb{N}$, if $e(x) = e^{2\pi i x}$, then if I is any set of m consecutive integers then*

$$1_{m|n} = \frac{1}{m} \sum_{h \in I} e(hn/m).$$

Proof. If $m \mid n$ then n/m is an integer, so $e(hn/m) = 1$ for all $h \in \mathbb{Z}$, so the right-hand side is 1.

If $m \nmid n$ then $e(n/m) \neq 1$. Let $S = \sum_{h\in I} e(hn/m)$, so it suffices to show that $S = 0$. Using $e(x + y) = e(x)e(y)$ we have

$$e(n/m)S = \sum_{h\in I} e((h + 1)n/m).$$

If $r = \min(I)$ and $s = \max(I)$ then the right-hand side is $S + e((s+1)n/m) - e(rn/m)$. But since I is a set of m consecutive integers we know that $s = r + m - 1$, and so

$$e((s + 1)n/m) - e(rn/m) = e(rn/m + n) - e(rn/m) = e(rn/m)e(n) - e(rn/m) = 0,$$

since $e(n) = 1$. Therefore $e(n/m)S = S$, and hence since $e(n/m) \neq 1$, this forces $S = 0$ as required. $\square$

Lemma 4.7. *Let A be a finite set of natural numbers not containing 0. If*

$$\mathcal{P}_A = \{p \text{ prime} : \exists n \in A : p \mid n\}$$

and for all $p \in \mathcal{P}_A$ then $r_p \geq 0$ is the greatest integer such that p^{r_p} divides some $n \in A$, then

$$[A] = \prod_{p\in\mathcal{P}_A} p^{r_p}.$$

(NOTE: This is a generally useful fact when working with lowest common multiples, perhaps this should be in mathlib somewhere.)

Proof. We first note that if $p \mid [A]$ then $p \in \mathcal{P}_A$. If not, suppose $p \notin \mathcal{P}_A$ and $p \mid [A]$. Let $M = [A]/p$. We claim every $n \in A$ divides M, contradicting the definition of lowest common multiple. It suffices to show that if a prime power $q^r \mid n$ then $q \mid M$. But we know $q^r \mid [A] = Mp$, and $(q^r, p) = 1$, so $q \mid M$ as required.

By the fundamental theorem of arithmetic, we can write $[A] = \prod_{p\in\mathcal{P}_A} p^{s_p}$ for some integers $s_p \geq 0$. It remains to show that $s_p = r_p$.

That $s_p \geq r_p$ follows from the fact that $p^{r_p} \mid n$ for some $n \in A$ by definition, hence $p^{r_p} \mid [A]$, hence $r_p \leq s_p$.

Suppose that $s_p > r_p$, and as above consider $M = [A]/p$. We claim every $n \in A$ divides M, the required contradiction. If a prime power $q^r \mid n$ with $q \neq p$ then $q \mid Mp$ hence $q \mid M$. If $p^r \mid n$ then $r \leq r_p$, so $p^r \mid p^{r_p} \mid p^{s_p-1} \mid M$, as required. $\square$

Lemma 4.8. *If A is a finite set of natural numbers not containing 0 such that if $q \in \mathcal{Q}_A$ then $q \leq X$, then $[A] \leq e^{O(X)}$.*

Proof. We have $[A] = \prod_{p\in\mathcal{P}_A} p^{r_p}$, where $p^{r_p} \in \mathcal{Q}_A$, by Lemma 4.7. By hypothesis $p^{r_p} \leq X$ and so $[A] \leq X^{|\mathcal{P}_A|}$. The set $\mathcal{P}_A$ is a subset of all primes $\leq X$, and so $|\mathcal{P}_A| \leq \pi(X) \ll X/\log X$ by Chebyshev's estimate Lemma 2.4. Therefore

$$[A] \leq X^{O(X/\log X)} = e^{O(X)}.$$

$\square$

Lemma 4.9. *If $x \in [0, 1/2]$ then*

$$\cos(\pi x) \leq e^{-2x^2}.$$

Proof. We have Jordan's inequality, which says $\sin(\pi x) \geq 2x$ for all $x \in [0, 1/2]$. Therefore

$$\cos(\pi x)^2 = 1 - \sin(\pi x)^2 \leq 1 - 4x^2.$$

Since $1 - y \leq e^{-y}$ for all $y \geq 0$, the right-hand side is $\leq e^{-4x^2}$. Taking square roots, and using $\cos(\pi x) \geq 0$ for all $x \in [0, 1/2]$, yields

$$\cos(\pi x) \leq e^{-2x^2}.$$

$\square$

Lemma 4.10. *Let A be a finite set of naturals not containing 0. For any $n \in A$, let $\mathcal{Q}_A(n)$ denote all those $q \in \mathcal{Q}_A$ such that $n \in A_q$, then*

$$|\mathcal{Q}_A(n)| \leq \tfrac{1}{\log 2} \log n.$$

Proof. The first step is to show that $\prod_{q \in \mathcal{Q}(n)} q \mid n$ (in fact they're equal, but all we need is the one direction). This can be shown by showing that every $q \in \mathcal{Q}(n)$ divides n, since $n \in A_q$ implies $q \mid n$, and then noting that any two distinct $q_1, q_2 \in \mathcal{Q}(n)$ are coprime. If not, then since they are prime powers, they must both be powers of the same prime, say $q_1 = p^{r_1} < q_2 = p^{r_2}$. Since $n \in A_{q_2}$ we have $p^{r_2} \mid n$, but then $p \mid n/q_1$, so $p \mid (q_1, n/q_1)$, and so $n \notin A_{q_1}$.

Therefore $\prod_{q \in \mathcal{Q}(n)} q \leq n$. Since all prime powers are ≥ 2 it follows that $2^{|\mathcal{Q}(n)|} \leq n$, and the result follows taking logarithms. $\square$

4.3 Towards the main proposition

Lemma 4.11. *If A is a finite set of natural numbers not containing 0 and $k \geq 1$ is an integer then*

$$F(A; k) = \frac{1}{[A]} \sum_{-[A]/2 < h \leq [A]/2} \prod_{n \in A} (1 + e(kh/n)).$$

Proof. For any $S \subset A$, $k \sum_{n \in S} \frac{1}{n} \in \mathbb{Z}$ if and only if $k \sum_{n \in S} \frac{[A]}{n} \in [A] \cdot \mathbb{Z}$. By definition $n \mid [A]$ for all $n \in A$, and so $k \sum_{n \in S} \frac{[A]}{n}$ is an integer. It is ≥ 0 since each summand is. Therefore by Lemma 4.6, if I is any set of $[A]$ consecutive integers

$$1_{k \sum_{n \in S} \frac{1}{n} \in \mathbb{Z}} = 1_{[A] \mid k \sum_{n \in S} \frac{[A]}{n}} = \frac{1}{[A]} \sum_{h \in I} e(kh \sum_{n \in S} \frac{1}{n}).$$

Therefore, changing summation,

$$F(A; k) = \sum_{S \subset A} 1_{k \sum_{n \in S} \frac{1}{n} \in \mathbb{Z}} = \frac{1}{[A]} \sum_{h \in I} \sum_{S \subset A} \prod_{n \in S} e(kh/n).$$

The lemma now follows choosing $I = (-[A]/2, [A]/2] \cap \mathbb{Z}$, and using the general fact that for any indexed set of complex numbers $(x_i)_{i \in I}$

$$\sum_{J \subset I} \prod_{j \in J} x_j = \prod_{i \in I} (1 + x_i).$$

$\square$

Lemma 4.12. *If $k \geq 1$ is an integer and A is a finite set of natural numbers such that there is no $S \subset A$ such that $R(S) = 1/k$, and $R(A) < 2/k$, then*

$$\sum_{-[A]/2<h\leq[A]/2} \Re\left(\prod_{n\in A}(1+e(kh/n))\right) = [A].$$

Proof. For any $S \subset A$ we have $k\sum_{n\in S}\frac{1}{n} \leq kR(A) < 2$, and therefore if $k\sum_{n\in S}\frac{1}{n} \in \mathbb{N}$ then $k\sum_{n\in S}\frac{1}{n} = 0$ or $= 1$. The latter can't happen by assumption, so $k\sum_{n\in S}\frac{1}{n} = 0$. A non-empty sum of > 0 summands is > 0, so if $k\sum_{n\in S}\frac{1}{n} \in \mathbb{Z}$ then $S = \varnothing$. Therefore $F(A;k) = 1$.

By Lemma 4.11 therefore

$$1 = \frac{1}{[A]}\sum_{-[A]/2<h\leq[A]/2}\prod_{n\in A}(1+e(kh/n)).$$

The conclusion follows multiplying both sides by $[A]$ and taking real parts of both sides. □

Lemma 4.13. *If $k \geq 1$ is an integer and A is a finite set of natural numbers such that there is no $S \subset A$ such that $R(S) = 1/k$, and $R(A) < 2/k$, and $[A] \leq 2^{|A|-1}$ then*

$$\sum_{h\in J(A)} \Re\left(\prod_{n\in A}(1+e(kh/n))\right) \leq -2^{|A|-1}.$$

Proof. By Lemma 4.12

$$\sum_{-[A]/2<h\leq[A]/2} \Re\left(\prod_{n\in A}(1+e(kh/n))\right) = [A].$$

By assumption the right-hand side is $\leq 2^{|A|-1}$.

When $h = 0$

$$\Re\left(\prod_{n\in A}(1+e(kh/n))\right) = \Re\left(\prod_{n\in A}2\right) = 2^{|A|}.$$

Therefore

$$\sum_{h\in J(A)} \Re\left(\prod_{n\in A}(1+e(kh/n))\right) + 2^{|A|} \leq 2^{|A|-1},$$

and the result follows after rearranging. □

Lemma 4.14. *If A is a finite set of integers and K is a real such that $[A] > K$ then, for any integer $k \geq 1$, the sets $\mathfrak{M}(t; A, k, K)$ are disjoint for distinct $t \in \mathbb{Z}$.*

Proof. Suppose not and $h \in \mathfrak{M}(t_1) \cap \mathfrak{M}(t_2)$. By definition,

$$|hk - t_1[A]| \leq K/2$$

and

$$|hk - t_2[A]| \leq K/2,$$

and so by the triangle inequality $[A]\,|t_1 - t_2| \leq K$. Since $t_1 \neq t_2$, we know $|t_1 - t_2| \geq 1$, and so $[A] \leq K$, contradicting the assumption. □

Lemma 4.15. *For any finite set of natural numbers A and $\theta \in \mathbb{R}$*

$$\Re\left(\prod_{n\in A}(1+e(\theta/n))\right) = 2^{|A|}\cos(\pi\theta R(A))\prod_{n\in A}\cos(\pi\theta/n).$$

Proof. Rewrite each factor in the product using $1+e(\theta/n) = 2e(\theta/2n)\cos(\pi\theta/n)$, so

$$\Re\left(\prod_{n\in A}(1+e(\theta/n))\right) = \Re\left(2^{|A|}e(\theta R(A)/2)\prod_{n\in A}\cos(\pi\theta/n)\right).$$

Taking out real factors, this is $2^{|A|}\prod_{n\in A}\cos(\pi\theta/n)\Re e(\theta R(A)/2)$, and the claim follows. □

Lemma 4.16. *Let $M \geq 1$ and A a finite set of naturals such that $n \geq M$ for all $n \in A$. Let K be a real such that $K < M$. Let $k \geq 1$ be an integer which divides $[A]$. Suppose that $kR(A) \in [2-k/M, 2)$.*

$$\sum_{h\in\mathfrak{M}(A,k,K)}\Re\left(\prod_{n\in A}(1+e(kh/n))\right) \geq 0.$$

Proof. Since k divides $[A]$, we know that $t[A]/k$ is an integer for any $t \in \mathbb{Z}$, and hence by definition of $\mathfrak{M}(t)$ we can write $h = t[A]/k + r$, where r is an integer satisfying $|r| \leq K/2k$. Therefore, letting

$$J_t = [-K/2k, K/2k] \cap (J(A) - t[A]/k),$$

then

$$\sum_{h\in\mathfrak{M}(t)}\Re\left(\prod_{n\in A}(1+e(kh/n))\right) = \sum_{r\in J_t}\Re\left(\prod_{n\in A}(1+e((t[A]+rk)/n))\right) = \sum_{r\in J_t}\Re\left(\prod_{n\in A}(1+e(rk/n))\right),$$

since $t[A]/n$ is always an integer, by definition of $[A]$.

Using Lemma 4.15, this is

$$2^{[A]}\sum_{h\in\mathfrak{M}(t)}\cos(\pi krR(A))\prod_{n\in A}\cos(\pi kr/n).$$

Since $[A] \geq \min(A) \geq M > K$, the hypotheses of Lemma 4.14 are all met, and so $\mathfrak{M}$ is the disjoint union of $\mathfrak{M}(t)$ as t ranges over $t \in \mathbb{Z}$. Therefore $1_{h\in\mathfrak{M}} = \sum_t 1_{h\in\mathfrak{M}(t)}$, and therefore using the above and rearranging the sum,

$$\sum_{h\in\mathfrak{M}(A,k,K)}\Re\left(\prod_{n\in A}(1+e(kh/n))\right) = \sum_{r\in[-K/2k,K/2k]\cap\mathbb{Z}}\left(\sum_{t\in\mathbb{Z}}1_{r\in J_t}\right)\cos(\pi krR(A))\prod_{n\in A}\cos(\pi kr/n).$$

Since for all $n \in A$ we have $n \geq M > K$ we have $|kr/n| < 1/2$ for all r with $|r| \leq K/2k$, and hence $\cos(\pi kr/n) \geq 0$ for all such n and r.

Furthermore, writing $kR(A) = 2-\epsilon$ for some $0 < \epsilon \leq k/M$, we have (since r is an integer)

$$\cos(\pi krR(A)) = \cos(-\pi r\epsilon) \geq 0,$$

since $|r\epsilon| \leq K/2M < 1/2$ for all $|r| \leq K/2k$. It follows that

$$\left(\sum_{t\in\mathbb{Z}}1_{r\in J_t}\right)\cos(\pi krR(A))\prod_{n\in A}\cos(\pi kr/n) \geq 0$$

for all $r \in [-K/2k, K/2k]\cap\mathbb{Z}$, and hence as the sum of non-negative summands the original sum is non-negative as required. □

Lemma 4.17. *Let $M \geq 1$ and A a finite set of naturals such that $n \geq M$ for all $n \in A$. Let K be a real such that $K < M$. Let $k \geq 1$ be an integer which divides $[A]$. Suppose that $kR(A) \in [2-k/M, 2)$, and there is no $S \subset A$ such that $R(S) = 1/k$, and $[A] \leq 2^{|A|-1}$. Then*

$$\sum_{h \in J(A) \backslash \mathfrak{M}(A,k,K)} C(A; hk) \geq 1/2.$$

Proof. By Lemma 4.13

$$\sum_{h \in J(A)} \Re \left(\prod_{n \in A} (1 + e(kh/n)) \right) \leq -2^{|A|-1}.$$

By Lemma 4.16

$$\sum_{h \in \mathfrak{M}(A,k,K)} \Re \left(\prod_{n \in A} (1 + e(kh/n)) \right) \geq 0.$$

Therefore

$$\sum_{h \in J(A) \backslash \mathfrak{M}(A,k,K)} \Re \left(\prod_{n \in A} (1 + e(kh/n)) \right) \leq -2^{|A|-1}.$$

By the triangle inequality, using $|\Re z| \leq |z|$ and $|1 + e(\theta)| = 2\cos(\pi\theta)$,

$$\sum_{h \in J(A) \backslash \mathfrak{M}(A,k,K)} |\cos(\pi kh/n)| \geq 1/2$$

as required. □

Lemma 4.18. *For any finite set A such that $n \leq N$ for all $n \in A$, and integer t, if $t \equiv t_n \pmod{n}$ for $|t_n| \leq n/2$ for all $n \in A$, then*

$$C(A; t) \leq \exp\left(-\frac{2}{N^2} \sum_{n \in A} t_n^2 \right).$$

Proof. We first note that $|\cos(\pi t/n)| = |\cos(\pi t_n/n)|$ for all $n \in A$, by periodicity of cosine. By Lemma 4.9, therefore

$$|\cos(\pi t/n)| \leq \exp(-2t_n^2/n^2) \leq \exp(-\tfrac{2}{N^2} t_n^2),$$

and the lemma follows taking the product over all $n \in A$. □

Lemma 4.19. *Suppose that N is sufficiently large and $M \geq 8$. Let $T > 0$ be a real and $k \geq 1$ be an integer. Let $A \subset [M, N]$ be a set of integers such that if $q \in \mathcal{Q}_A$ then $q \leq \frac{TK^2}{N^2 \log N}$. Then*

$$\sum_{h \in \mathfrak{m}_1(A,k,K,T)} C(A; hk) \leq 1/8.$$

Proof. We show in fact that for any $h \in \mathfrak{m}_1(A, k, K, T)$ we have

$$C(A; hk) \leq 1/[A]^2.$$

The result then immediately follows since $|\mathfrak{m}_1(A, k, K, T)| \leq |J(A)| \leq [A]$, assuming $[A] \geq 8$, which is true since $[A] \geq \min(A) \geq M \geq 8$.

By Lemma 4.18,

$$C(A;hk) \le \exp\left(-\frac{2}{N^2}\sum_{n\in A} h_n^2\right),$$

where $kh \equiv h_n \pmod n$ and $|h_n| \le n/2$. Let I_h be the interval of length K centred around kh. If no element of I_h is divisible by n then $|h_n| > K/2$. Therefore, by definition of $\mathfrak{m}_1$, $|h_n| > K/2$ for at least T many $n \in A$, and hence $\sum_{n\in A} h_n^2 \ge K^2T/4$, and so

$$C(A;h) \le \exp\left(-\frac{K^2T}{2N^2}\right).$$

It remains to note that by Lemma 4.8

$$[A] \le \exp\left(O\left(\frac{K^2T}{N^2\log N}\right)\right) \le \exp\left(\frac{K^2T}{4N^2}\right)$$

assuming N is sufficiently large. □

Lemma 4.20. *Suppose that $N \ge 4$. Let $A \subset [1,N]$ be a finite set of integers and t an integer. Let $K, L > 0$ be reals and suppose that $q \le \frac{1}{16}LK^2/N^2(\log N)^2$ for all $q \in \mathcal{Q}_A$. Let $\mathcal{D} = \mathcal{D}_I(A;L)$ where I is the interval of length K centred at t. Then*

$$C(A;t) \le N^{-4|\mathcal{Q}_A\backslash\mathcal{D}|}.$$

Proof. For any $n \in A$, let $\mathcal{Q}(n)$ denote all those $q \in \mathcal{Q}$ such that $n \in A_q$. Therefore, for any real $x_n \ge 0$,

$$\prod_{n\in A} x_n = \prod_{n\in A}\prod_{q\in\mathcal{Q}(n)} x_n^{1/|\mathcal{Q}(n)|} = \prod_{q\in\mathcal{Q}}\prod_{n\in A_q} x_n^{1/|\mathcal{Q}(n)|}.$$

By Lemma 4.10 for any $n \in A$ we have $|\mathcal{Q}(n)| \le \frac{1}{\log 2}\log n \le 2\log N$, and so if $0 \le x_n \le 1$,

$$\prod_{n\in A} x_n \le \prod_{q\in\mathcal{Q}}\prod_{n\in A_q} x_n^{1/2\log N} = \prod_{q\in\mathcal{Q}}\left(\prod_{n\in A_q} x_n\right)^{1/2\log N}.$$

In particular,

$$C(A;t) \le \prod_{q\in\mathcal{Q}} C(A_q;t)^{/2\log N}.$$

Using the trivial bound $C(A_q;t) \le 1$, to prove the lemma it therefore suffices to show that for every $q \in \mathcal{Q}\backslash\mathcal{D}_h$ we have $C(A_q;t) \le N^{-8\log N}$.

For any $n \in A$ let $t \equiv t_n \pmod n$, where $|t_n| \le n/2$. For any $q \in \mathcal{Q}\backslash\mathcal{D}$ there are, by definition of $\mathcal{D}$, at least L/q many $n \in A_q$ such that n divides no element of the interval of length K centred at t.

Recall that t_n is the integer in $(-n/2, n/2]$ such that $t \equiv t_n \pmod n$, so that n divides $t - t_n$. If $|t_n| \le K/2$ then $t - t_n$ is in the interval of length K centred at t, which contradicts the previous paragraph. Therefore $|t_n| > K/2$. Hence by Lemma 4.18

$$C(A_q;t) \le \exp\left(-\frac{2}{N^2}\cdot\frac{L}{q}\cdot\frac{K^2}{4}\right).$$

By assumption, $q \le LK^2/16N^2(\log N)^2$, and the proof is complete. □

Lemma 4.21. *Let N be a sufficiently large integer. Let $K, L > 0$ be reals such that $0 < K \leq N$. Let $T > 0$ be any real. Let k be an integer such that $1 \leq k \leq N/64$. Let $A \subset [1, N]$ be a finite set of integers such that*

1. *if $q \in \mathcal{Q}_A$ then $q \leq \frac{1}{16} \frac{LK^2}{N^2 (\log N)^2}$, and*
2. *for any interval I of length K, either*

   (a)
   $$\#\{n \in A : \text{no element of } I \text{ is divisible by } n\} \geq T,$$
   or

   (b) there is some $x \in I$ divisible by all $q \in \mathcal{D}_I(A; L)$.

Then

$$\sum_{h \in \mathfrak{m}_2(A,k,K,T)} C(A; hk) \leq 1/8.$$

Proof. For any $h \in \mathfrak{m}_2$, let I_h be the interval of length K centred at kh. Since $h \notin \mathfrak{m}_1$, condition 2(a) cannot hold, and therefore there is some $x \in I_h$ divisible by all $q \in \mathcal{D}_{I_h}(A; L)$. In particular, kh is distance at most $K/2$ from some multiple of $[\mathcal{D}_{I_h}(A; L)]$. Since $h \in \mathfrak{m}(A, k, K, T)$, we know $h \notin \mathfrak{M}(A, k, K)$, and hence kh is distance greater than $K/2$ from any multiple of $[A]$. Since $[\mathcal{Q}] = [A]$, it follows that $\mathcal{D}_{I_h}(A; L) \neq \mathcal{Q}$.

Furthermore, for any $\mathcal{D} \subset \mathcal{Q}$, the number of $h \in \mathfrak{m}_2$ with $\mathcal{D}_{I_h}(A; L) = \mathcal{D}$ is at most K times the number of multiples of $[\mathcal{D}]$ in $[1, k[A] + K]$. The number of multiples of $[\mathcal{D}]$ in $[1, k[A]+K]$ is at most $(k[A]/[\mathcal{D}])+K$, and hence the number of $h \in \mathfrak{m}_2$ with $\mathcal{D}_{I_h}(A; L) = \mathcal{D}$ is at most

$$Kk\frac{[A]}{[\mathcal{D}]} + K^2 \leq Kk \prod_{q \in \mathcal{Q} \setminus \mathcal{D}} q + K^2 \leq kN^{|\mathcal{Q} \setminus \mathcal{D}|+1} + K^2.$$

In particular, when $\mathcal{D} \neq \mathcal{Q}$, the right-hand side is at most $2kN^{|\mathcal{Q} \setminus \mathcal{D}|+1}$, that is,

$$\sum_{h \in \mathfrak{m}_2} 1_{\mathcal{D}_{I_h}(A;L)=\mathcal{D}} \leq 2kN^{|\mathcal{Q} \setminus \mathcal{D}|+1}.$$

(Here we have used the fact that $[A] \leq [D] \prod_{q \in \mathcal{Q} \setminus \mathcal{D}} q$, which is true because every prime power dividing $[A]$ is in $\mathcal{Q}$, and hence either divides $[D]$ or divides $\prod_{q \in \mathcal{Q} \setminus \mathcal{D}} q$.)

Therefore, for any $\mathcal{D} \subset \mathcal{Q}$, using Lemma 4.20,

$$\sum_{h \in \mathfrak{m}_2} 1_{\mathcal{D}_{I_h}(A;L)=\mathcal{D}} C(A; hk) \leq 2kN^{1-3|\mathcal{Q}_A \setminus \mathcal{D}|}.$$

Since $\mathcal{D} \neq \mathcal{Q}$, this is at most $2kN^{-1-|\mathcal{Q} \setminus \mathcal{D}|}$.

Therefore (using the trivial estimate $|\mathcal{Q}| \leq N$)

$$\begin{aligned}
\sum_{h \in \mathfrak{m}_2} C(A; h, k) &\leq \frac{2k}{N} \sum_{\mathcal{D} \subseteq \mathcal{Q}} N^{-|\mathcal{Q} \setminus \mathcal{D}|} \\
&= \frac{2k}{N}(1 + 1/N)^{|\mathcal{Q}|} \\
&\leq 8k/N \leq 1/8.
\end{aligned}$$

□

Proposition 4.22. *There exists a constant $c > 0$ such that the following holds. Suppose that N is sufficiently large. Let K, L, M, T be reals such that $T, L > 0$ and $8 \leq K < M \leq N$, and let k be an integer such that $1 \leq k \leq M/64$. Let $A \subset [M, N]$ be a set of integers such that*

1. *$R(A) \in [2/k - 1/M, 2/k)$,*
2. *k divides $[A]$,*
3. *if $q \in \mathcal{Q}_A$ then $q \leq c \min\left(M/k, \frac{LK^2}{N^2(\log N)^2}, \frac{TK^2}{N^2 \log N}\right)$, and*
4. *for any interval I of length K, either*

   (a)
   $$\#\{n \in A : \text{no element of } I \text{ is divisible by } n\} \geq T,$$
   or

   (b) there is some $x \in I$ divisible by all $q \in \mathcal{D}_I(A; L)$.

There is some $S \subset A$ such that $R(S) = 1/k$.

Proof. If the conclusion fails, there is an immediate contradiction by combining Lemmas 4.17, 4.19 and 4.21. The only hypothesis that is not immediate to check is the upper bound $[A] < 2^{|A|-1}$. This follows since $|A| \geq MR(A) \geq \frac{2}{k}M - 1 \geq M/k$, and by Lemma 4.8 we have $[A] \leq e^{O(cM/k)}$, which is at most $2^{|A|/2}$ assuming c is sufficiently small. $\square$

Chapter 5

Technical Lemmas

Lemma 5.1. *If* $0 < |n_1 - n_2| \leq N$ *then*

$$\sum_{q|(n_1,n_2)} \frac{1}{q} \ll \log\log\log N,$$

where the summation is restricted to prime powers.

Proof. If $q \mid (n_1, n_2)$ then q divides $|n_1 - n_2|$, and hence in particular $q \leq N$. The contribution of all prime powers p^r with $r \geq 2$ is $O(1)$, and hence it suffices to show that $\sum_{p||n_1-n_2|} \frac{1}{p} \ll \log\log\log N$. Any integer $\leq N$ is trivially divisible by $O(\log N)$ many primes. Clearly summing $1/p$ over $O(\log N)$ many primes is maximised summing over the smallest $O(\log N)$ primes. Since there are $\gg (\log N)^{3/2}$ many primes $\leq (\log N)^2$, we have

$$\sum_{p||n_1-n_2|} \frac{1}{p} \ll \sum_{p \leq (\log N)^2} \frac{1}{p} \ll \log\log\log N$$

by Mertens' estimate 2.6. □

Lemma 5.2. *Let* $1/2 > \epsilon > 0$ *and* N *be sufficiently large, depending on* ϵ. *If* A *is a finite set of integers such that* $R(A) \geq (\log N)^{-\epsilon/2}$ *and* $(1-\epsilon)\log\log N \leq \omega(n) \leq 2\log\log N$ *for all* $n \in A$ *then*

$$\sum_{q \in \mathcal{Q}_A} \frac{1}{q} \geq (1-2\epsilon)e^{-1}\log\log N.$$

Proof. Since, by definition, every integer $n \in A$ can be written uniquely as $q_1 \cdots q_t$ for $q_i \in \mathcal{Q}_A$ for some $t \in I = [(1-\epsilon)\log\log N, 2\log\log N]$, we have that, since $t! \geq (t/e)^t$,

$$R(A) \leq \sum_{t \in I} \frac{\left(\sum_{q \in \mathcal{Q}_A} \frac{1}{q}\right)^t}{t!} \leq \sum_{t \in I} \left(\frac{e}{t} \sum_{q \in \mathcal{Q}_A} \frac{1}{q}\right)^t.$$

Since $(ex/t)^t$ is decreasing in t for $x < t$, either $\sum_{q \in \mathcal{Q}_A} \frac{1}{q} \geq (1-\epsilon)\log\log N$ (and we are done), or the summand is decreasing in t, and hence we have

$$(\log N)^{-\epsilon/2} \leq R(A) \leq 2\log\log N \left(\frac{\sum_{q \in \mathcal{Q}_A} \frac{1}{q}}{(1-\epsilon)e^{-1}\log\log N}\right)^{(1-\epsilon)\log\log N}.$$

The claimed bound follows, using the fact that $e^{-\frac{\epsilon}{2(1-\epsilon)}} \geq 1-\epsilon$ for $\epsilon \in (0,1/2)$, choosing N large enough such that $(2\log\log N)^{2/\log\log N} \leq 1+\epsilon^2$, say.

To verify the inequality $e^{-\frac{\epsilon}{2(1-\epsilon)}} \geq 1-\epsilon$ for $\epsilon \in (0,1/2)$, note that $\frac{\epsilon}{2(1-\epsilon)} \leq \epsilon$ for all $\epsilon \in (0,1/2)$, and hence using $1+x \leq e^x$ for all $x \geq -1$, we have

$$e^{-\frac{\epsilon}{2(1-\epsilon)}} \geq e^{-\epsilon} \geq 1-\epsilon$$

as required. □

Lemma 5.3. *There is a constant $C > 0$ such that the following holds. Let $0 < y < N$ be some reals, and suppose that D is a finite set of integers such that if $d \in D$ and $p^r \| d$ then $y < p^r \leq N$. Then, for any real $k \geq 1$,*

$$\sum_{d \in D} \frac{k^{\omega(d)}}{d} \leq \left(C \frac{\log N}{\log y} \right)^k .$$

Proof. We can write every $d \in D$ uniquely as a product of prime powers $d = p_1^{r_1} \cdots p_k^{r_k}$, where for each i either $r_i = 1$ and $y < p_i \leq N$, or $r_i \geq 2$ and $p_i \leq N$. Therefore the left-hand side is at most

$$\prod_{y<p\leq N} \left(1+\frac{k}{p}\right) \prod_{p \leq N} \left(1+\frac{k}{p^2}+\cdots\right).$$

The second product is, using $1+kx \leq (1+x)^k$,

$$= \prod_{p \leq N} \left(1+\frac{k}{p(p-1)}\right) \leq \left(\prod_{p\leq N}\left(1+\frac{1}{p(p-1)}\right)\right)^k \leq C_1^k$$

for some constant $C_1 > 0$, since the product converges. Similarly,

$$\prod_{y<p\leq N}\left(1+\frac{k}{p}\right) \leq \left(\prod_{y<p\leq N}\left(1+\frac{1}{p}\right)\right)^k,$$

and the product here is $\ll (\log N/\log y)$ by Mertens' bound 2.6. Combining these yields the required estimate. □

Lemma 5.4. *There is a constant $c > 0$ such that the following holds. Let $N \geq M \geq N^{1/2}$ be sufficiently large, and suppose that k is a real number such that $1 \leq k \leq c\log\log N$. Suppose that $A \subset [M,N]$ is a set of integers such that $\omega(n) \leq (\log N)^{1/k}$ for all $n \in A$.*

For all q such that $R(A;q) \geq 1/\log N$ there exists d such that

1. *$qd > M\exp(-(\log N)^{1-1/k})$,*
2. *$\omega(d) \leq \frac{5}{\log k}\log\log N$, and*
3. $$\sum_{\substack{n \in A_q \\ qd \mid n \\ (qd,n/qd)=1}} \frac{qd}{n} \gg \frac{R(A;q)}{(\log N)^{2/k}}.$$

Proof. Fix some q with $R(A;q) \geq (\log N)^{-1/2}$. Let $y = \exp((\log N)^{1-2/k})$. Let D be the set of all d such that if r is a prime power such that $r \mid d$ and $(r, d/r) = 1$ then $y < r \leq N$, and furthermore

$$qd \in (M \exp(-(\log N)^{1-1/k}), N].$$

We first claim that every $n \in A_q$ is divisible by some qd with $d \in D$, such that $(qd, n/qd) = 1$. Let Q_n be the set of prime powers r dividing n such that $(r, n/r) = 1$, so that $n = \prod_{r \in Q_n} r$. Let

$$Q'_n = \{r \in Q_n : r \neq q \text{ and } r > y\}.$$

We let $d = \prod_{r \in Q'_n} r$. We need to check that $d \in D$ and $qd \mid n$. It is immediate from the definitions that if a prime power r divides d with $(r, d/r) = 1$ then $r \in Q'_n$, whence $r > y$. It is also straightforward to check, by comparing prime powers, that $qd \mid n$, and hence $qd \leq n \leq N$. Furthermore,

$$\frac{n}{qd} = \prod_{r \in Q_n \setminus Q'_n \cup \{q\}} r \leq y^{|Q_n|} = y^{\omega(n)} \leq \exp((\log N)^{1-1/k}),$$

by definition of y and the hypothesis that $\omega(n) \leq (\log N)^{1/k}$. It follows that $d \in D$ as required.

If we let $A_q(d) = \{n \in A_q : qd \mid n \text{ and } (qd, n/qd) = 1\}$ then we have shown that $A_q \subseteq \bigcup_{d \in D} A_q(d)$. Therefore

$$R(A;q) = \sum_{n \in A_q} \frac{q}{n} \leq \sum_{d \in D} \sum_{n \in A_q(d)} \frac{q}{n}.$$

We will control the contribution from those d with $\omega(d) > \omega_0 = \frac{5}{\log k} \log\log N$ with the trivial bound

$$\sum_{A_q(d)} \frac{q}{n} \leq \sum_{\substack{n \leq N \\ qd \mid n}} \frac{q}{n} \leq \sum_{m \leq N} \frac{q}{dqm} \leq 2 \frac{\log N}{d},$$

using the harmonic sum bound $\sum_{m \leq N} \frac{1}{m} \leq \log N + 1 \leq 2 \log N$. Therefore

$$\begin{aligned} \sum_{\substack{d \in D \\ \omega(d) > \omega_0}} \sum_{n \in A_q(d)} \frac{q}{n} &\leq 2 \log N \sum_{\substack{d \in D \\ \omega(d) > \omega_0}} \frac{1}{d} \\ &\leq 2k^{-\omega_0} \log N \sum_{d \in D} \frac{k^{\omega(d)}}{d}. \end{aligned}$$

By 5.3 we have

$$\sum_{d \in D} \frac{k^{\omega(d)}}{d} \leq \left(C \frac{\log N}{\log y} \right)^k.$$

and therefore

$$\sum_{\substack{d \in D \\ \omega(d) > \omega_0}} \sum_{n \in A_q(d)} \frac{q}{n} \leq 2k^{-\omega_0} \log N \left(C \frac{\log N}{\log y} \right)^k.$$

Recalling $\omega_0 = \frac{5}{\log k} \log\log N$ and $y = \exp((\log N)^{1-2/k})$ the right-hand side here is at most

$$2(\log N)^{-5} C^k \leq 1/2 \log N \leq \tfrac{1}{2} R(A;q),$$

for sufficiently large N, assuming $k \leq c \log\log N$ for sufficiently small c.

It follows that

$$\tfrac{1}{2}R(A;q) \leq \sum_{\substack{d\in D\\ \omega(d)\leq\omega_0}} \sum_{d\in D} \frac{q}{n}.$$

The result follows by averaging, since by 5.3

$$\sum_{d\in D} \frac{1}{d} \ll \frac{\log N}{\log y} \ll (\log N)^{2/k}.$$

□

Lemma 5.5. *Let N be sufficiently large, and let $\epsilon > 0$ and $A \subset [1, N]$. There exists $B \subset A$ such that*

$$R(B) \geq R(A) - 2\epsilon \log\log N$$

and $\epsilon < R(B; q)$ for all $q \in \mathcal{Q}_B$.

Proof. We will prove the following claim, valid for any finite set $A \subset \mathbb{N}$ with $0 \notin A$ and $\epsilon > 0$, via induction: for all $i \geq 0$ there exists $A_i \subseteq A$ and $Q_i \subseteq \mathcal{Q}_A$ such that

1. the sets $\mathcal{Q}_{A_i}$ and Q_i are disjoint,
2. $R(A_i) \geq R(A) - \epsilon \sum_{q\in Q_i} \frac{1}{q}$,
3. either $i \leq |A\backslash A_i|$ or, for all $q' \in \mathcal{Q}_{A_i}$, $\epsilon < R(A_i; q')$.

Given this claim, the stated lemma follows by choosing $B = A_{|A|+1}$, noting that in the 5th point above the first alternative cannot hold, and furthermore

$$\epsilon \sum_{q\in Q_{|A|+1}} \frac{1}{q} \leq \epsilon \sum_{q\leq N} \frac{1}{q} \leq 2\epsilon \log\log N,$$

by Lemma 2.6, assuming N is sufficiently large.

To prove the inductive claim, we begin by choosing $Q_0 = \varnothing$, and $A_0 = A$. All of the properties are trivial to verify.

Suppose the proposition is true for $i \geq 0$, with associated A_i, Q_i. We will now prove it for $i + 1$. If it is true that, for all $q' \in \mathcal{Q}_{A_i}$, we have $\epsilon < R(A_i; q')$, then we let $A_{i+1} = A_i$ and $Q_{i+1} = Q_i$.

Otherwise, let $q' \in \mathcal{Q}_{A_i}$ be such that $R(A_i; q') \leq \epsilon$, set $A_{i+1} = A_i\backslash\{n \in A_i : q' \mid n \text{ and } (q', n/q') = 1\}$, and $Q_{i+1} = Q_i \cup \{q'\}$.

It remains to verify the required properties. That $Q_{i+1} \subseteq \mathcal{Q}_A$ follows from $Q_i \subseteq \mathcal{Q}_A$ and $q' \in \mathcal{Q}_{A_i} \subseteq \mathcal{Q}_A$ (using the general fact that if $B \subseteq A$ then $\mathcal{Q}_B \subseteq \mathcal{Q}_A$).

For point (1), since $A_{i+1} \subseteq A_i$, and hence $\mathcal{Q}_{A_{i+1}} \subseteq \mathcal{Q}_{A_i}$, it suffices to show that $q' \notin \mathcal{Q}_{A_{i+1}}$. If this were not true, there would be some $n \in A_{i+1}$ such that $q' \mid n$ and $(q', n/q') = 1$, contradiction to the definition of A_{i+1}.

For point (2), we note that

$$R(A_{i+1}) = R(A_i) - \sum_{\substack{n\in A_i\\ q'|n\\ (q',n/q')=1}} \frac{1}{n} = R(A_i) - \frac{1}{q'}R(A_i; q') \geq R(A_i) - \epsilon\frac{1}{q'},$$

and point (2) follows from induction and the observation that

$$\frac{1}{q'} + \sum_{q \in Q_i} \frac{1}{q} = \sum_{q \in Q_{i+1}} \frac{1}{q},$$

since $q' \in \mathcal{Q}_{A_i}$ and hence $q' \notin Q_i$.

Finally, for point (3), we note that since we assume that it is not true that for all $q' \in \mathcal{Q}_{A_i}$, we have $\epsilon < R(A_i; q')$, it must be true that $i \leq |A \backslash A_i|$. We now claim that $i+1 \leq |A \backslash A_{i+1}|$, for which it suffices to show that $A_{i+1} \subsetneq A_i$. This is true since $q' \in \mathcal{Q}_{A_i}$, and hence the set $\{n \in A_i : q' \mid n \text{ and } (q', n/q') = 1\}$ is not empty. $\square$

Lemma 5.6. *Suppose that N is sufficiently large and $N > M$. Let ϵ, α be reals such that $\alpha > 4\epsilon \log\log N$ and $A \subset [M, N]$ be a set of integers such that*

$$R(A) \geq \alpha + 2\epsilon \log\log N$$

and if $q \in \mathcal{Q}_A$ then $q \leq \epsilon M$.

There is a subset $B \subset A$ such that $R(B) \in [\alpha - 1/M, \alpha)$ and, for all $q \in \mathcal{Q}_B$, we have $\epsilon < R(B; q)$.

Proof. We will prove the following, for all N sufficiently large, real $0 < M < N$, reals $\epsilon > 0$ and $\alpha > 4\epsilon \log\log N$ and finite sets $A \subseteq [M, N]$ such that $R(A) \geq \alpha$ and $R(A; q) > \epsilon$ for all $q \in \mathcal{Q}_A$, by induction: for all $i \geq 0$ there exists $A_i \subseteq A$ such that

1. $R(A_i) \geq \alpha - 1/M$,
2. $R(A_i; q) > \epsilon$ for all $q \in \mathcal{Q}_{A_i}$, and
3. either $i \leq |A \backslash A_i|$ or $R(A_i) < \alpha$.

Given this claim, we prove the lemma as follows. Let $A' \subseteq A$ be as given by Lemma 5.5, so that the hypotheses of the inductive claim are satisfied for A'. We now apply the inductive claim, and choose $B = A_{|A|+1}$, noting that the first alternative of point (3) cannot hold.

It remains to prove the inductive claim. The base case $i = 0$ is immediate by hypotheses, choosing $A_0 = A$.

We now fix some $i \geq 0$, with associated A_i as above, and prove the claim for $i+1$. If $R(A_i) < \alpha$ then we set $A_{i+1} = A_i$. Otherwise, we have $R(A_i) \geq \alpha$. By Lemma 5.5 there exists $B \subset A_i$ such that

$$R(B) \geq R(A_i) - 4\epsilon \log\log N \geq \alpha - 4\epsilon \log\log N > 0$$

and $2\epsilon < R(B; q)$ for all $q \in \mathcal{Q}_B$. In particular, since $R(B) > 0$, the set B is not empty. Let $x \in B$ be arbitrary, and set $A_{i+1} = A_i \backslash \{x\}$.

We note that since $x \in B \subseteq A$, we have $x \geq M$, and hence $R(A_{i+1}) = R(A_i) - 1/x \geq \alpha - 1/M$, so point (1) is satisfied. Furthermore, since $R(A_i) \geq \alpha$, we have $i \leq |A \backslash A_i|$, and hence $i + 1 \leq |A \backslash A_{i+1}|$ and point (3) is satisfied.

For point (2), let $q \in \mathcal{Q}_{A_{i+1}} \subseteq \mathcal{Q}_{A_i}$. If it is not true that $q \mid x$ and $(q, x/q) = 1$ then

$$R(A_{i+1}; q) = R(A_i; q) > \epsilon.$$

Otherwise, if $q \mid x$ and $(q, x/q) = 1$, then $q \in \mathcal{Q}_B$. It follows that, since $B \subseteq A_i$, we have

$$R(A_{i+1}; q) \geq R(B; q) - \frac{q}{x} > 2\epsilon - \frac{q}{M} \geq \epsilon,$$

since $q \leq \epsilon M$. The inductive claim now follows. $\square$

Chapter 6

Deduction of main technical proposition

Lemma 6.1. *Suppose N is sufficiently large and $N \geq M \geq N^{1/2}$, and suppose that $A \subset [M, N]$ is a set of integers such that*

$$\tfrac{99}{100} \log\log N \leq \omega(n) \leq 2 \log\log N \quad \textit{for all} \quad n \in A.$$

Then for any interval of length $\leq MN^{-2/(\log\log N)}$, if A_I is the set of those $n \in A$ which divide some element of I, for all $q \in \mathcal{Q}_A$ such that $R(A_I; q) > 1/2(\log N)^{1/100}$ there exists some integer $x_q \in I$ such that $q \mid x_q$ and

$$\sum_{\substack{r \mid x_q \\ r \in \mathcal{Q}_A}} \frac{1}{r} \geq 0.35 \log\log N.$$

Proof. Let I be any interval of length $\leq MN^{-2/(\log\log N)}$, and let $q \in \mathcal{Q}_A$ such that $R(A_I; q) > 1/2(\log N)^{1/100}$.

Apply Lemma 5.4 with A replaced by A_I, and $k = \frac{\log\log N}{\log 2 + \log\log\log N}$, so that $(\log N)^{1/k} = 2 \log\log N$. We choose N large enough such that $k \leq c \log\log N$ (where c is the constant in the statement of Lemma 5.4) and such that $\log k \geq 5000$.

This produces some d_q such that $qd_q > |I|$ and $\omega(d_q) < \frac{1}{1000} \log\log N$, and

$$\sum_{\substack{n \in A_I \\ qd_q \mid n \\ (qd_q, n/qd_q)=1}} \frac{qd_q}{n} \gg \frac{1}{(\log N)^{1/100} (\log\log N)^2} \geq \frac{1}{(\log N)^{1/99}},$$

assuming N is sufficiently large. Let

$$A_I^{(q)} = \{n/qd_q : n \in A_I \text{ with } qd_q \mid n \text{ and } (qd_q, n/qd_q) = 1\}$$

so that, the above inequality states that $R(A_I^{(q)}) \geq (\log N)^{-1/99}$.

We now claim that $2 \log\log N \geq \omega(m) \geq \frac{97}{99} \log\log N$ for all $m \in A_I^{(q)}$. This follows from the trivial $\omega(mqd_q) \geq \omega(m) \geq \omega(mqd_q) - \omega(qd_q)$, and that $2 \log\log N \geq \omega(mqd_q) \geq \frac{99}{100} \log\log N$ (since $mqd_q \in A$) and $\omega(qd_q) \leq 1 + \omega(d_q) < \frac{1}{500} \log\log N$.

We now apply Lemma 5.2 with $\epsilon = 2/99$. This implies that

$$\sum_{r \in \mathcal{Q}_{A_I^{(q)}}} \frac{1}{r} \geq \frac{95}{99} e^{-1} \log\log N \geq 0.35 \log\log N.$$

By definition of A_I, every $n \in A_I$ with $qd_q \mid n$ must divide some $x \in I$, say $x(n)$. We claim that $x(n) = x(m)$ for all $n, m \in A_I$ with $qd_q \mid n$ and $qd_q \mid m$. Otherwise $|x(q,n) - x(q,m)|$ is a non-zero multiple of qd_q, so $|I| < qd_q \leq |x(q,n) - x(q,m)|$, which is an integer in $[1, |I|)$, a contradiction.

Let $x_q \in I$ be such that $x(q,n) = x_q$ for all $n \in A_I$ with $qd_q \mid n$. We claim that $\mathcal{Q}_{A_I^{(q)}} \subset \{r \in \mathcal{Q}_A : r \mid x_q\}$, which concludes the proof using the above bound.

If $r \in \mathcal{Q}_{A_I^{(q)}}$ then there exists some $m \in A_I^{(q)}$ such that $r \mid m$ and $(r, m/r) = 1$. Let $m = n/qd_q$ where $n \in A$ with $qd_q \mid n$ and $(qd_q, n/qd_q) = (qd_q, m) = 1$. Certainly $r \mid n = mqd_q$, since $r \mid m$. Furthermore, r is coprime to qd_q since r divides m. Therefore $(r, n/r) = (r, (m/r)qd_q) = (r, m/r) = 1$, and hence $r \in \mathcal{Q}_A$.

Finally, we need to show that $r \mid x_q$. Let $m \in A_I^{(q)}$ be such that $r \mid m$. Let $n \in A_I$ be such that $qd_q \mid n$ and $m = n/qd_q$. Then $r \mid n = mqd_q$, and hence since $n \mid x_q$ we have $r \mid x_q$ as required. $\square$

Lemma 6.2. *Suppose N is an integer and $\delta, M \in \mathbb{R}$. Suppose that $A \subset [M, N]$ is a set of integers such that, for all $q \in \mathcal{Q}_A$,*

$$R(A; q) \geq 2\delta,$$

Then for any finite set of integers I, for all $q \in \mathcal{D}_I(A; \delta M)$, if A_I is the set of those $n \in A$ that divide some element of I,

$$R(A_I; q) > \delta.$$

Proof. For every $q \in \mathcal{D}_I(A; \delta M)$, by definition,

$$\#\{n \in A_q : n \notin A_I\} < \frac{\delta M}{q}.$$

Therefore

$$R(A_I; q) = \sum_{\substack{n \in A_q \\ n \in A_I}} \frac{q}{n} > \sum_{n \in A_q} \frac{q}{n} - \frac{q}{M}\left(\frac{\delta M}{q}\right).$$
$$= R(A; q) - \delta \geq \delta.$$

$\square$

Proposition 6.3. *Suppose N is sufficiently large and $N \geq M \geq N^{1/2}$, and suppose that $A \subset [M, N]$ is a set of integers such that*

$$\tfrac{99}{100} \log\log N \leq \omega(n) \leq 2 \log\log N \quad \textit{for all} \quad n \in A,$$

$$R(A) \geq (\log N)^{-1/101}$$

and, for all $q \in \mathcal{Q}_A$,

$$R(A; q) \geq (\log N)^{-1/100}.$$

Then either

1. *there is some $B \subset A$ such that $R(B) \geq \frac{1}{3}R(A)$ and*

$$\sum_{q \in \mathcal{Q}_B} \frac{1}{q} \leq \frac{2}{3} \log\log N,$$

or

2. *for any interval of length $\leq MN^{-2/(\log\log N)}$, either*

   (a)
   $$\#\{n \in A : \textit{no element of } I \textit{ is divisible by } n\} \geq M/\log N,$$

   or

   (b) there is some $x \in I$ divisible by all $q \in \mathcal{D}_I(A; M/2(\log N)^{1/100})$.

Proof. Let I be any interval of length $\leq MN^{-2/(\log\log N)}$, and let A_I be those $n \in A$ that divide some element of I, and $\mathcal{D} = \mathcal{D}_I(A; M/2(\log N)^{1/100})$. We may assume that $|A \backslash A_I| < M/\log N$.

Let $\mathcal{E}$ be the set of those $q \in \mathcal{Q}_A$ such that $R(A_I; q) > 1/2(\log N)^{1/100}$, so that by Lemma 6.2 we have $\mathcal{D} \subset \mathcal{E}$. By Lemma 6.1 for each $q \in \mathcal{E}$ there exists $x_q \in I$ such that $q \mid x_q$ and

$$\sum_{\substack{r \mid x_q \\ r \in \mathcal{Q}_A}} \frac{1}{r} \geq 0.35 \log\log N.$$

Let $X = \{x_q : q \in \mathcal{E}\}$. Suppose that $|X| \geq 3$, and say $x, y, z \in X$ are three distinct elements. Then

$$\sum_{q \in \mathcal{Q}_A} \frac{1}{q} \geq \sum_{\substack{q \mid x \\ q \in \mathcal{Q}_A}} \frac{1}{q} + \sum_{\substack{q \mid y \\ q \in \mathcal{Q}_A}} \frac{1}{q} + \sum_{\substack{q \mid z \\ q \in \mathcal{Q}_A}} \frac{1}{q} - \sum_{q \mid (x,y)} \frac{1}{q} - \sum_{q \mid (y,z)} \frac{1}{q} - \sum_{q \mid (x,z)} \frac{1}{q},$$

where the last three sums are restricted to prime powers.

Since $\sum_{q \in \mathcal{Q}_A} \frac{1}{q} \leq (1 + o(1)) \log\log N \leq 1.01 \log\log N$, say, by Lemma 2.6, it follows that

$$\sum_{q \mid (x,y)} \frac{1}{q} + \sum_{q \mid (y,z)} \frac{1}{q} + \sum_{q \mid (x,z)} \frac{1}{q} \geq 0.04 \log\log N.$$

But since $x, y \in I$ we have, by Lemma 5.1,

$$\sum_{q \mid (x,y)} \frac{1}{q} \ll \log\log\log N,$$

and similarly for the other sums, and so the left-hand side is $O(\log\log\log N)$, which is a contradiction for large enough N.

Therefore $|X| \leq 2$. If $\mathcal{D}$ is empty the second conclusion trivially holds, and therefore we may assume there is some $q \in \mathcal{E}$ and hence some $x_q \in X$, so $|X| \geq 1$. If $|X| = 1$ then we are in the second conclusion, letting x be the unique element of X, since by construction $x = x_q$ for all $q \in \mathcal{D}$, and hence in particular $q \mid x$ for all $q \in \mathcal{D}_I$ as required.

Finally, we deal with the case $|X| = 2$. Say $X = \{w_1, w_2\}$ with $w_1 \neq w_2$ (and both are elements of I). Let $A^{(i)} = \{n \in A : n \mid w_i\}$ and $A^{(0)} = A \backslash (A^{(1)} \cup A^{(2)})$. Without loss of generality, label the w_i such that $R(A^{(1)}) \geq R(A^{(2)})$. Suppose first that $R(A^{(0)}) <$

$R(A)/3$. In this case, since $R(A^{(0)})+R(A^{(1)})+R(A^{(2)}) \geq R(A)$, we have $R(A^{(1)}) \geq R(A)/3$. Furthermore, note that

$$\sum_{\substack{q|w_1\\ q\in\mathcal{Q}_A}} \frac{1}{q} \leq \sum_{q\leq N} \frac{1}{q} - \sum_{\substack{q|w_2\\ q\in\mathcal{Q}_A}} \frac{1}{q} + \sum_{q|(w_1,w_2)} \frac{1}{q}.$$

As above, by Lemma 2.6, the assumed lower bound on the sum over $q \mid w_2$, and Lemma 5.1, for sufficiently large N the right-hand side is

$$\leq (1.01 - 0.35 + 0.01) \log\log N \leq \tfrac{2}{3} \log\log N.$$

Every $q \mid \mathcal{Q}_{A^{(1)}}$ is in $\mathcal{Q}_A$ and divides w_1, and hence

$$\sum_{q\in\mathcal{Q}_{A^{(1)}}} \frac{1}{q} \leq \frac{2}{3} \log\log N,$$

and we are in case (1) with $B = A^{(1)}$.

Finally, suppose that $R(A^{(0)}) \geq R(A)/3$. We will derive a contradiction, assuming N is large enough. Let $A' \subset A^{(0)}$ be the set of those $n \in A_I \cap A^{(0)}$ such that if $n \in A_q$ then $q \in \mathcal{E}$. By definition of $\mathcal{E}$ and Mertens' estimate 2.6, and the earlier assumption that $|A\backslash A_I| < M/\log N$,

$$R(A^{(0)}\backslash A') \leq \frac{|A\backslash A_I|}{M} + \sum_{q\in\mathcal{Q}_A\backslash\mathcal{E}} \frac{1}{q} R(A_I; q) \ll \frac{\log\log N}{(\log N)^{1/100}},$$

and so in particular, since $R(A) \geq (\log N)^{-1/101}$, we have $R(A') \gg (\log N)^{-1/101}$.

In particular, $|A'| \gg M/(\log N)^{-1/101}$. Every $n \in A'$ divides some $x \in I$, and there are at most $MN^{-2/\log\log N}$ many $x \in I$, so by the pigeonhole principle there must exist some $x \in I$ (necessarily $x \neq w_1$ and $x \neq w_2$ since $A' \subset A^{(0)}$) such that, if $A'' = \{n \in A' : n \mid x\}$, then

$$|A''| \gg N^{2/\log\log N} (\log N)^{-1/101},$$

and hence $|A''| \geq N^{3/2\log\log N}$, say, assuming N is sufficiently large.

However, if $n \in A''$ then both $n \mid x$ and $n \mid w_1 w_2$ (since every q with $n \in A_q$ is in $\mathcal{E}$ and so divides either w_1 or w_2), and hence n divides

$$(x, w_1 w_2) \leq (x, w_1)(x, w_2) \leq |x - w_1|\,|x - w_2| \leq N^2.$$

Therefore the size of A'' is at most the number of divisors of some fixed integer $m \leq N^2$, which is at most $N^{(1+o(1))2\log 2/\log\log N}$ by Lemma 2.5, and hence we have a contradiction for large enough N, since $2\log 2 < 3/2$. $\square$

Proposition 6.4. *Suppose N is sufficiently large and $N \geq M \geq N^{1/2}$, and suppose that $A \subset [M, N]$ is a set of integers such that*

$$\tfrac{99}{100} \log\log N \leq \omega(n) \leq 2 \log\log N \quad \textit{for all} \quad n \in A,$$

and, for all $q \in \mathcal{Q}_A$,

$$R(A; q) \geq (\log N)^{-1/100},$$

and $\sum_{q\in\mathcal{Q}_A} \frac{1}{q} \leq \frac{2}{3} \log\log N$. Then for any interval of length $\leq MN^{-2/(\log\log N)}$, either

1.
$$\#\{n\in A : \textit{no element of } I \textit{ is divisible by } n\}\geq M/\log N,$$

or

2. *there is some* $x\in I$ *divisible by all* $q\in\mathcal{D}_I(A;M/2(\log N)^{1/100})$.

Proof. Let I be any interval of length $\leq MN^{-2/(\log\log N)}$, and let A_I be those $n\in A$ that divide some element of I, and $\mathcal{D}=\mathcal{D}_I(A;M/2q(\log N)^{1/00})$. We may assume that $|A\backslash A_I|<M/\log N$.

By Lemmas 6.1 and 6.2 for each $q\in\mathcal{D}$ there exists $x_q\in I$ such that $q\mid x_q$ and

$$\sum_{\substack{r|x_q\\ r\in\mathcal{Q}_A}}\frac{1}{r}\geq 0.35\log\log N.$$

Let $X=\{x_q : q\in\mathcal{D}\}$. Suppose that $|X|\geq 2$, and say $x,y\in X$ are two distinct elements. Then

$$\sum_{q\in\mathcal{Q}_A}\frac{1}{q}\geq\sum_{\substack{q|x\\ q\in\mathcal{Q}_A}}\frac{1}{q}+\sum_{\substack{q|y\\ q\in\mathcal{Q}_A}}\frac{1}{q}-\sum_{q|(x,y)}\frac{1}{q},$$

where the third sum is restricted to prime powers. Using our assumed bounds, this implies

$$\sum_{q|(x,y)}\frac{1}{q}\geq\left(0.7-\frac{2}{3}\right)\log\log N\geq\frac{1}{100}\log\log N,$$

say.

But since $x,y\in I$ we have, by Lemma 5.1,

$$\sum_{q|(x,y)}\frac{1}{q}\ll\log\log\log N,$$

which is a contradiction for large enough N.

Therefore $|X|\leq 1$. If $\mathcal{D}$ is empty then the second conclusion trivially holds, for any $x\in I$. Therefore there exists some $q\in\mathcal{D}$, therefore there exists some $x_q\in X$, and hence $|X|=1$. Let $x\in I$ be the unique element of X. By construction $x=x_q$ for all $q\in\mathcal{D}$, and hence in particular $q\mid x$ for all $q\in\mathcal{D}_I$ as required. □

Lemma 6.5. *Let N be sufficiently large. Let ϵ,y,w,M be reals such that $M<N$ and $2\leq\lceil y\rceil\leq\lfloor w\rfloor$ and $1/M<\epsilon\log\log N$ and $\frac{2}{w^2}\geq 3\epsilon\log\log N$ and let $A\subset[M,N]$ be such that*

1. $R(A)\geq 2/y+2\epsilon\log\log N$,
2. *every* $n\in A$ *is divisible by some* $d\in[y,w]$,
3. *if* $q\in\mathcal{Q}_A$ *then* $q\leq\epsilon M$.

Then there exists some $\emptyset\neq A'\subseteq A$ and $d\in[y,w]$ such that $R(A')\in[2/d-1/M,2/d)$, for all $q\in\mathcal{Q}_{A'}$ we have $R(A';q)>\epsilon$, and there is a multiple of d in A', and there are no multiples of any $m\in[y,d)$ in A.

Proof. In the proof below, we write $t_i = \min(\lceil y\rceil + i, \lfloor w\rfloor)$ for all integer $i \geq 0$ for simplicity. It is convenient to note at the outset that $t_0 = \lceil y\rceil$, $t_i \leq w$ for all i, for all $i \geq 0$ we have $t_{i+1} \in [t_i, t_i + 1]$, and if $t_i < \lfloor w\rfloor$ then $t_{i+1} = t_i + 1$. Furthermore, if $i \geq \lfloor w\rfloor - \lceil y\rceil$, then $t_i = \lfloor w\rfloor$.

We will prove by induction on i that, under the same hypotheses as in the lemma, for all $0 \leq i$ there exists $A_i \subseteq A$ and integer d_i with $y \leq d_i \leq t_i$ such that

1. $R(A_i) \in [2/d_i - 1/M, 2/d_i)$,
2. for all $q \in \mathcal{Q}_{A_i}$ we have $R(A_i; q) > \epsilon$,
3. there are no multiples of any $m \in [y, d_i)$ in A, and
4. either there exists a multiple of d_i in A_i, or nothing in A_i is divisible by any $d \in [y, t_i]$.

Given this inductive claim, the lemma follows by letting $A' = A_{\lfloor w\rfloor - \lceil y\rceil}$ and $d = d_{\lfloor w\rfloor - \lceil y\rceil}$, so that then $t_{\lfloor w\rfloor - \lceil y\rceil} = \lfloor w\rfloor$, noting that the assumptions imply that $2/d - 1/M \geq 2/w - 1/M > 0$ and hence $R(A') > 0$ and so $A' \neq \varnothing$, and furthermore by assumption the second alternative of point (3) cannot hold. It remains to prove the inductive claim.

For the base case $i = 0$, we set $d_0 = \lceil y\rceil$ and apply Lemma 5.6 with $\alpha = 2/d_0$ to find some $A_0 \subseteq A$ such that $R(A_0) \in [2/d_0 - 1/M, 2/d_0)$ and for all $q \in \mathcal{Q}_{A_0}$ we have $\epsilon < R(A_0; q)$. (To see that point (4) holds, note that d_0 is the unique integer in $[y, t_0]$.)

For the inductive step, suppose that A_i, d_i are as given for $i \geq 0$. If there exists a multiple of d_i in A_i, then we set $d_{i+1} = d_i$ and $A_{i+1} = A_i$.

If not, then nothing in A_i is divisible by any $d \in [y, t_i]$. We now set $d_{i+1} = t_{i+1}$ and apply Lemma 5.6 to A_i with $\alpha = 2/d_{i+1}$. To be able to apply this, we need to verify that

$$\alpha = 2/d_{i+1} > 4\epsilon \log\log N$$

and

$$R(A_i) \geq \alpha + 2\epsilon \log\log N = 2/d_{i+1} + 2\epsilon \log\log N.$$

For the first, we note that $d_{i+1} \leq w$, and hence we are done since $2/w > 4/w^2 \geq 6\epsilon \log\log N$.

For the second, we note that by assumption, every $n \in A$ is divisible by some $d \in [y, w]$, and hence since we are assuming that nothing in A_i is divisible by any $d \in [y, t_i]$, we must have $t_i < \lfloor w\rfloor$, and hence $d_{i+1} = t_{i+1} = t_i + 1 \geq d_i + 1$. Therefore

$$R(A_i) \geq \frac{2}{d_i} - \frac{1}{M} \geq \frac{2}{d_{i+1} - 1} - \frac{1}{M} \geq \frac{2}{d_{i+1} - 1} - \epsilon \log\log N,$$

since $1/M \leq \epsilon \log\log N$, and furthermore

$$\frac{2}{d_{i+1} - 1} - \frac{2}{d_{i+1}} = \frac{2}{d_{i+1}(d_{i+1} - 1)} \geq \frac{2}{d_{i+1}^2} \geq \frac{2}{w^2} \geq 3\epsilon \log\log N,$$

and hence $R(A_i) \geq \alpha + 2\epsilon \log\log N$ as required.

Thus we can apply Lemma 5.6, to produce some $A_{i+1} \subseteq A_i$ such that $R(A_{i+1}) \in [2/d_{i+1} - 1/M, 2/d_{i+1})$, such that for all $q \in \mathcal{Q}_{A_{i+1}}$ we have $R(A_{i+1}; q) > \epsilon$. Furthermore, if there is something in A_{i+1} divisible by some $d \in [y, t_{i+1}]$, then since $A_{i+1} \subseteq A_i$ there must be something in A_{i+1} divisible by some $d \in [y, t_{i+1}] \backslash [y, t_i]$. Since $t_{i+1} \in [t_i, t_i + 1]$, the only integer that could be in $[y, t_{i+1}] \backslash [y, t_i]$ is $t_{i+1} = d_{i+1}$, and the proof is complete. $\square$

Proposition 6.6 (Main Technical Proposition)**.** *Let N be sufficiently large. Suppose $A \subset [N^{1-1/\log\log N}, N]$ and $1 \leq y \leq z \leq (\log N)^{1/500}$ are such that*

1. *$R(A) \geq 2/y + (\log N)^{-1/200}$,*
2. *every $n \in A$ is divisible by some d_1 and d_2 where $y \leq d_1$ and $4d_1 \leq d_2 \leq z$,*
3. *every prime power q dividing some $n \in A$ satisfies $q \leq N^{1-6/\log\log N}$, and*
4. *every $n \in A$ satisfies*
$$\tfrac{99}{100} \log\log N \leq \omega(n) \leq 2 \log\log N.$$

There is some $S \subset A$ such that $R(S) = 1/d$ for some $d \in [y, z]$.

Proof. Let $M = N^{1-1/\log\log N}$. Apply Lemma 6.5 with $w = z/4$ and $\epsilon = N^{-5/\log\log N}$ to find some corresponding A' and d.

Suppose first that case (2) of Proposition 6.3 holds for A'. The hypotheses of Proposition 4.22 are now met with $k = d$, $\eta = 1/2(\log N)^{1/100}$, and $K = MN^{-2/\log\log N}$. This yields some $S \subset A' \subset A$ such that $R(S) = 1/d_j$ as required.

Otherwise, Proposition 6.3 yields some $B \subset A'$ such that

$$R(B) \geq 2/3d_j - 1/M \geq 1/2d_j + (\log N)^{-1/200}$$

and where $\sum_{q \in \mathcal{Q}_B} \frac{1}{q} \leq \frac{2}{3} \log\log N$.

We apply Lemma 6.5 again, with $y = 4d$ and $w = z$, to produce some corresponding $B' \subset B$ and $d' \in [4d, z]$. Since $\sum_{q \in \mathcal{Q}_{B'}} \frac{1}{q} \leq \sum_{q \in \mathcal{Q}_B} \frac{1}{q} \leq \frac{2}{3} \log\log N$ we can apply Proposition 6.4. The hypotheses of Proposition 4.22 are now met with $k = d'$ and η, K as above, and thus there is some $S \subset B' \subset A$ such that $R(S) = 1/d'$ as required. $\square$