Unit fractions

Thomas F. Bloom and Bhavik Mehta

January 29, 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 $n_1, \ldots, n_k \in A$ such that $\frac{1}{n_1} + \cdots + \frac{1}{n_k} = 1$.

(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`).

Chapter 1

Definitions

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

For any finite $A \subset \mathbb{N}$
$$R(A) = \sum_{n \in A} \frac{1}{n}.$$

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 $n \in A$).

For any finite $A \subset \mathbb{N}$ and prime power $q \in \mathcal{Q}_A$ we define
$$R(A; q) = \sum_{n \in A_q} \frac{q}{n}.$$

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

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

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

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 $X \ge 3$*

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

Proof. The left-hand side equals

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

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

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

as required. □

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

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

Proof. □

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

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

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

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

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

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

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

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

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

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

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

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

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 $K \geq 2$ as required.)

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

$$\log K \leq \log \log \log n.$$

Taking logarithms of the inequality in the first part,

$$\begin{aligned} \log \tau(n) &\leq \frac{\log n}{K} + 2^K \log K \leq \log n \left(\frac{1}{K} + \frac{\log \log \log n}{(\log n)^\epsilon} \right) \\ &= \log n \frac{\log 2}{\log \log n} \left(1 + \epsilon/2 + \frac{(\log \log \log n)(\log \log n)}{(\log 2)(\log n)^\epsilon} \right). \end{aligned}$$

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 $\leq \epsilon/2$, and hence

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

as required.

□

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

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

where the sum is restricted to prime powers.

Proof. □

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

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

where the sum is restricted to primes.

Proof. □

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

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

Proof. □

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

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

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

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

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 $\sum_{n \le z} 1_{d|n} = \lfloor \frac{z}{d} \rfloor$ for any $z \ge 0$,

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

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

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

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 6.2.

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 $S_1, \ldots, S_d \subset A$ 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 $n \in \{1, \ldots, N\}$ 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 6.2. 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 6.2 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. □

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

$$\begin{aligned}\sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \sum_{\substack{n \le N_i \\ q \mid n}} \frac{1}{n} &\ll \sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \frac{\log(N_i/q)}{q} \\ &\ll \frac{\log N_i}{\log\log N_i} \sum_{N_i^{1-6/\log\log N_i} < q \le N_i} \frac{1}{q} \\ &\ll \frac{\log N}{(\log\log N)^2},\end{aligned}$$

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 $S_1, \dots, S_k \subset A$ 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 6.2 with $y = 1$ and $z = (\log N)^{1/500}$ – note that condition (2) of Proposition 6.2 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.

4.1 Local definitions

In this section, we write $[A]$ for the lowest common multiple of A, and 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.

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

Let $C(B;h,k) = \prod_{n\in B} |\cos(\pi kh/n)|$.

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

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

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

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 4.1. *For any $n, m \in \mathbb{N}$, if $e(x) = e^{2\pi ix}$, 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. □

Lemma 4.2. *Let A be a finite set of natural numbers. If*

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

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

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

Proof. By definition it suffices to show that for all $n \in A$ we have $n \mid \prod_{p \in \mathcal{P}_A} p^{r_p}$. By the fundamental theorem of arithmetic, it suffices to show that for each prime power $q^r \mid n$, $q \in \mathcal{P}_A$ and $r \leq r_q$. Both are true by definition of $\mathcal{P}_A$ and r_q respectively. □

Lemma 4.3. *If A is a finite set of natural numbers such that if $q \in \mathcal{Q}_A$ then $q \leq X$, then $[A] \leq e^{O(X)}$.*

Proof. We have $[A] = \prod_{\in \mathcal{P}_A} p^{r_p}$, where $p^{r_p} \in \mathcal{Q}_A$, by Lemma 4.2. 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

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

□

Lemma 4.4. *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

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

□

4.3 Towards the main proposition

Lemma 4.5. *If $A \subset \mathbb{N}$ is finite 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 4.1, if I is any set of $[A]$ consecutive integers

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

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 $(x_i)_{i \in I}$

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

□

Lemma 4.6. *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 $S = \emptyset$. Therefore $F(A;k) = 1$.

By Lemma 4.5 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 4.7. *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 4.6

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

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 4.8. *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 4.9. *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 4.10. *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. 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 4.9, 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 4.8 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. $\square$

Lemma 4.11. *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. 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*

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

Proof. By Lemma 4.7

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

By Lemma 4.10

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

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 $|1 + e(\theta)| = 2\cos(\pi\theta)$,

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

as required. □

Lemma 4.12. *For any finite set A such that $n \leq N$ for all $n \in A$, and integers k, h, if $kh \equiv h_n \pmod{n}$ for $|h_n| \leq n/2$ for all $n \in A$, then*

$$C(A; h, k) \leq \exp\left(-\frac{2}{N^2}\sum_{n \in A} h_n^2\right).$$

Proof. We first note that $|\cos(\pi kh/n)| = |\cos(\pi h_n/n)|$ for all $n \in A$, by periodicity of cosine. By Lemma 4.4, therefore

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

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

Lemma 4.13. *Suppose that N is sufficiently large and $M \geq 8$. Let $k \geq 1$ be an integer. Let $A \subset [M, N]$ be a set of integers such that if $q \in \mathcal{Q}_A$ then $q \leq \frac{MK^2}{N^2(\log N)^2}$. Then*

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

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

$$C(A; h, k) \leq 1/[A]^2.$$

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

By Lemma 4.12,

$$C(A; h, k) \leq \exp\left(-\frac{2}{N^2}\sum_{n \in A} h_n^2\right).$$

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 $M/\log N$ many $n \in A$, and hence $\sum_{n \in A} h_n^2 \geq K^2 M/4\log N$, and so

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

It remains to note that by Lemma 4.3

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

assuming N is sufficiently large. □

Lemma 4.14. *There is a constant $c > 0$ such that the following holds. Suppose that $N \geq 4$. Let $A \subset [1, N]$ be a finite set of integers and h an integer. Let $K, L > 0$ be reals and suppose that $q \leq cLK^2/N^2(\log N)^2$ for all $q \in \mathcal{Q}_A$. Let $\mathcal{D} = \mathcal{D}_I(A; L/q)$ where I is the interval of length K centred at h. Then*

$$C(A; h, k) \leq N^{-4|\mathcal{Q}_A \setminus \mathcal{D}|}.$$

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 $x_n \geq 0$,

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

Using the trivial estimate $\omega(n) \ll \log n$, there is a constant $C > 0$ such that for any $n \in A$ we have $|\mathcal{Q}(n)| \leq C\log N$, and so if $0 \leq x_n \leq 1$,

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

In particular,

$$C(A;h,k) \leq \prod_{q\in\mathcal{Q}} C(A_q;h,k)^{1/C\log N}.$$

Using the trivial bound $C(A_q;h,k) \leq 1$, to prove the lemma it therefore suffices to show that for every $q \in \mathcal{Q}\backslash\mathcal{D}_h$ we have $C(A_q;h,k) \leq N^{-4C\log N}$.

For any $n \in A$ let $kh \equiv h_n \pmod{n}$, where $|h_n| \leq n/2$. 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 $|h_n| > K/2$ and hence by Lemma 4.12

$$C(A_q;h,k) \leq \exp\left(-\frac{2}{N^2}\cdot\frac{L}{q}\cdot\frac{K^2}{4}\right).$$

By assumption, $q \leq LK^2/8CN^2(\log N)^2$, and the proof is complete. $\square$

Lemma 4.15. *There is a constant $c > 0$ such that the following holds. Let N be a sufficiently large integer. Let $K, L > 0$ be reals. Let k be an integer such that $1 \leq k \leq N/2$. Let $A \subset [1,N]$ be a finite set of integers such that*

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

   (a)

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

   or

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

Then

$$\sum_{h\in\mathfrak{m}_2(A,k,K,M/\log N)}\prod_{n\in A}|\cos(\pi kh/n)| \leq 1/8.$$

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 $q \in \mathcal{D}_{I_h}(A;L/q)$. In particular, kh is distance at most $K/2 \leq M/2$ from some multiple of $[\mathcal{D}_{I_h}(A;L/q)]$.

Therefore, for any $\mathcal{D} \subset \mathcal{Q}$, the number of $h \in \mathfrak{m}_2$ with $\mathcal{D}_{I_h}(A;L/q) = \mathcal{D}$ is at most M times the number of multiples of $[\mathcal{D}]$ in $[1,k[A]]$, which is at most

$$Mk\frac{[A]}{[\mathcal{D}]} \leq Mk\prod_{q\in\mathcal{Q}\backslash\mathcal{D}} q \leq kN^{|\mathcal{Q}\backslash\mathcal{D}|+1}.$$

Therefore, for any $\mathcal{D} \subset \mathcal{Q}$, using Lemma 4.14,

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

By definition of $\mathfrak{m}$, if $h \in \mathfrak{m}_2$ then kh is distance greater than $K/2$ from any multiple of $[A]$, and hence $\mathcal{D}_{I_h}(A;L/q) \neq \mathcal{Q}$. Therefore (using the trivial estimate $|\mathcal{Q}| \leq N$)

$$\begin{aligned}\sum_{h\in\mathfrak{m}_2} C(A;h,k) &\leq kN \sum_{\mathcal{D}\subsetneq\mathcal{Q}} N^{-3|\mathcal{Q}\setminus\mathcal{D}|}\\ &\leq \frac{k}{N}(1+1/N)^{|\mathcal{Q}|}\\ &\leq 4k/N \leq 1/8.\end{aligned}$$

□

Proposition 4.16. *There is a constant $c > 0$ such that the following holds. Suppose that N is sufficiently large. Let K, L, M be reals and k an integer such that $1 \leq k \leq N/2$ and $M > K$. 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 the lowest common multiple of A,*
3. *if $q \in \mathcal{Q}_A$ then $q \leq c\min\left(|A|, \frac{LK^2}{N^2(\log N)^2}\right)$, and*
4. *for any interval I of length K, either*
   - *(a)*
     $$\#\{n \in A : \text{no element of } I \text{ is divisible by } n\} \geq M/\log N,$$
     or
   - *(b) there is some $x \in I$ divisible by all $q \in \mathcal{D}_I(A;L/q)$.*

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 4.11, 4.13 and 4.15 (and the required upper bound $[A] < 2^{|A|-1}$ comes from Lemma 4.3).

□

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. $\square$

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. $\square$

Lemma 5.3. *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 $1 \leq k \leq c\log\log N$. 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 $R(A;q) \geq (\log N)^{-1/2}$ 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 D be the set of all d such that if p is a prime and $p^r \| d$ then
$$p^r > y = \exp((\log N)^{1-2/k})$$
and
$$qd \in (M\exp(-(\log N)^{1-1/k}), N].$$
We first claim that every $n \in A_q$ is divisible by some qd with $d \in D$, such that $(qd, n/qd) = 1$. This can be done greedily, just removing from n/q all those prime power divisors $p^r \| n/q$ such that $p^r \leq y$, which removes at most
$$y^{\omega(n)} \leq \exp((\log N)^{1-1/k}).$$
We can therefore bound
$$R(A;q) \leq \sum_{d\in D}\frac{1}{d}\sum_{\substack{n\in A_q\\ qd\mid n\\ (qd,n/qd)=1}}\frac{qd}{n}.$$

We will control the contribution from those d with $\omega(d) > \omega_0 = \frac{5}{\log k}\log\log N$ with the trivial bound
$$\sum_{\substack{n\in A_q\\ qd\mid n\\ (qd,n/qd)=1}}\frac{qd}{n} \leq \sum_{\substack{n\leq N\\ qd\mid n}}\frac{qd}{n} \ll \log N$$

and Mertens' bound 2.6. Together these imply

$$\sum_{\substack{d\in D\\ \omega(d)>\omega_0}}\frac{1}{d}\sum_{\substack{n\in A_q\\ qd|n}}\frac{qd}{n}\ll \log N\sum_{\substack{d\\ p^r\|d\implies y<p^r\leq N\\ \omega(d)\geq\omega_0}}\frac{1}{d}$$

$$\ll k^{-\omega_0}\log N\sum_{\substack{d\\ p^r\|d\implies y<p^r\leq N}}\frac{k^{\omega(d)}}{d}$$

$$\ll C_1^k k^{-\omega_0}\log N\prod_{y<p\leq N}(1+\frac{k}{p-1})$$

$$\leq k^{-\omega_0}\log N\left(C_2\frac{\log N}{\log y}\right)^k$$

for some absolute constants $C_1, C_2 > 0$. Recalling the definitions of y and ω_0, this is

$$\leq C_2^k k^{-\omega_0}(\log N)^3\leq 1/\log N,$$

say, for N sufficiently large. It follows that

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

The result follows since

$$\sum_{d\in D}\frac{1}{d}\leq\sum_{\substack{d\\ p^r\|d\implies y<p^r\leq N}}\frac{1}{d}\ll\prod_{y<p\leq N}\left(1-\frac{1}{p-1}\right)^{-1}\ll\frac{\log N}{\log y}\ll(\log N)^{2/k}.$$

□

Lemma 5.4. *Let N be sufficiently large and $A\subset[1,N]$. There exists $B\subset A$ such that*

$$R(B)\geq R(A)-\frac{1}{(\log N)^{1/200}}$$

and $R(B;q)\geq 2/(\log N)^{1/100}$ for all $q\in\mathcal{Q}_B$.

Proof. We construct a sequence of decreasing sets $A=A_0\supsetneq A_1\supsetneq\cdots\supsetneq A_i$ as follows. Given some A_i, if there is a prime power $q_i\in\mathcal{Q}_{A_i}$ such that

$$R(A_i;q_i)<\frac{2}{(\log N)^{1/100}},$$

then we let $A_{i+1}=A_i\backslash(A_i)_{q_i}$. If no such q_i exists then we halt the construction. This process must obviously terminate in some finite time (since some non-empty amount of A_i is being removed at each step). Suppose that it halts at $A_j=B$, say. The amount lost from $R(A)$ at step i is

$$\sum_{n\in(A_i)_{q_i}}\frac{1}{n}=\frac{1}{q_i}R(A_i;q_i)<\frac{2}{q_i(\log N)^{1/100}},$$

and furthermore each $q \le N$ can appear as at most one such q_i, since after removing $(A_i)_{q_i}$ anything left in A_i cannot have q_i as a coprime divisor. It follows that

$$R(B) > R(A) - \frac{2}{(\log N)^{1/100}} \sum_{q \le N} \frac{1}{q} \ge R(A) - \frac{1}{(\log N)^{1/200}},$$

since $\sum_{q \le N} \frac{1}{q} \ll \log\log N$. □

Lemma 5.5. *Suppose that N is sufficiently large and $N \ge M \ge N^{1/2}$. Let $\alpha > 2/(\log N)^{1/200}$ and $A \subset [M, N]$ be a set of integers such that*

$$R(A) \ge \alpha + \frac{1}{(\log N)^{1/200}}$$

and if $q \in \mathcal{Q}_A$ then $q \le M/(\log N)^{1/100}$.

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

$$R(B; q) \ge \frac{1}{(\log N)^{1/100}}.$$

Proof. We first apply Lemma 5.4 to produce some $A' \subset A$ such that $R(A') \ge \alpha$ and $R(A'; q) \ge 2/(\log N)^{1/100}$ for all $q \in \mathcal{Q}_{A'}$.

We now argue that whenever D is such that $R(D) \ge \alpha$ and $R(D; q) \ge (\log N)^{-1/100}$ for all $q \in \mathcal{Q}_D$ there exists some $x \in D$ such that $R(D \backslash \{x\}; q) \ge (\log N)^{-1/100}$ for all $q \in \mathcal{Q}_D$. Given this, the lemma immediately follows, since we can continue removing such elements from A' one at time until $R(B)$ falls in the required interval.

To see why the above fact holds, apply Lemma 5.4 to obtain some $B \subset D$ (such that $R(B) \ge (\log N)^{-1/200}$, and hence in particular B is non-empty), and let x be any element of B. If $x \notin D_q$ then by definition $R(D \backslash \{x\}; q) = R(D; q) \ge (\log N)^{-1/100}$. If $x \in D_q$ then $x \in B_q$, and so

$$R(D \backslash \{x\}; q) \ge R(B; q) - \frac{q}{x} \ge \frac{2}{(\log N)^{1/100}} - \frac{q}{M} \ge \frac{1}{(\log N)^{1/100}}$$

as required. □

Chapter 6

Deduction of main technical proposition

Proposition 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*

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

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

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 $q \in \mathcal{D}_I(A; M/2q(\log N)^{1/100})$.

If $\sum_{q \in \mathcal{Q}_A} \frac{1}{q} \leq \frac{2}{3} \log\log N$ then case (2) is guaranteed.

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. We may assume that $|A \backslash A_I| < M/\log N$ (or else 2(a) holds), and we need to show that either there is some $x \in I$ divisible by all $q \in \mathcal{D}_I$, or the first case holds.

Let $\mathcal{E}_I$ be the set of those $q \in \mathcal{Q}_A$ such that $R(A_I;q) > 1/2(\log N)^{1/100}$. For every $q \in \mathcal{D}_I$, by definition,

$$R(A_I;q) \geq R(A;q) - \left(\frac{M}{2q(\log N)^{1/100}}\right)\frac{q}{M} > \frac{1}{2(\log N)^{1/100}},$$

and hence in particular $\mathcal{D}_I \subset \mathcal{E}_I$.

For any $q \in \mathcal{E}_I$ we may therefore apply Lemma 5.3 with A replaced by A_I, and k chosen such that $(\log N)^{1/k} = 2\log\log N$. This produces some d_q such that $qd_q > |I|$ and $\omega(d_q) < \frac{1}{500}\log\log N$ (provided N is sufficiently large), and

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

By definition of A_I, every $n \in A_I$ with $qd_q \mid n$ must divide some $x \in I$ – in fact, they must all divide the same $x \in I$ (call this $x_q \in I$, say), since all such x are in particular divisible by $qd_q > |I|$, which can divide at most one element in I.

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, assuming N is sufficiently large, $R(A_I^{(q)}) \geq (\log N)^{-1/99}$, say. We may therefore apply Lemma 5.2 with $\epsilon = 2/99$ (note that since $\omega(n) \geq \frac{99}{100}\log\log N$ for $n \in A$ and $\omega(d_q) < \frac{1}{500}\log\log N$, we must have $\omega(m) \geq \frac{97}{99}\log\log N$ for all $m \in A_I^{(q)}$). This implies that

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

Trivially, $\mathcal{Q}_{A_I^{(q)}} \subset \mathcal{Q}_A$, and further by choice of x_q, all $r \in \mathcal{Q}_{A_I^{(q)}}$ divide x_q, and hence

$$\sum_{\substack{r|x_q\\ r\in\mathcal{Q}_A}} \frac{1}{r} \geq \frac{95}{99}e^{-1}\log\log N \geq 0.35\log\log N.$$

For any two $n_1 \neq n_2 \in I$, we have

$$\sum_{q|(n_1,n_2)} \frac{1}{q} \ll \log\log\log N \leq 0.01\log\log N$$

for N sufficiently large, by Lemma **??**. It follows that if $\sum_{q\in\mathcal{Q}_A}\frac{1}{q} \leq \frac{2}{3}\log\log N$ then there can be at most one such possible value for $x_q \in I$ as q ranges over $\mathcal{E}_I$, and hence this common shared value of x_q is an $x \in I$ divisible by all $q \in \mathcal{E}_I$, and hence certainly by all $q \in \mathcal{D}_I$, as required.

Furthermore, since $\sum_{q\in\mathcal{Q}_A}\frac{1}{q} \leq (1+o(1))\log\log N \leq 1.01\log\log N$, say, there must always be at most two distinct values of $x_q \in I$ as q ranges over $\mathcal{E}_I$. If there is no $x \in I$ divisible by all $q \in \mathcal{D}_I$, there must be exactly two such values, say w_1 and w_2.

Let $A^{(i)} = \{n \in A : n \mid w_i\}$ and $A^{(0)} = A\backslash(A^{(1)} \cup A^{(2)})$. Since every $q \in \mathcal{Q}_{A^{(1)}}$ divides w_1,

$$\sum_{q\in\mathcal{Q}_{A^{(1)}}} \frac{1}{q} \leq \sum_{q\leq N}\frac{1}{q} - \sum_{q|w_2}\frac{1}{q} + \sum_{q|(w_1,w_2)}\frac{1}{q} \leq (1 - \tfrac{95}{99}e^{-1} + o(1))\log\log N.$$

For large enough N, the right-hand side is $\leq \frac{2}{3}\log\log N$, and similarly for $A^{(2)}$. Since $R(A^{(0)})+R(A^{(1)})+R(A^{(2)}) \geq R(A)$, we are in the first case choosing $B = A^{(1)}$ or $B = A^{(2)}$, unless $R(A^{(0)}) \geq R(A)/3$. In this latter case we will derive a contradiction.

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 $q \in \mathcal{E}_I$. By definition of $\mathcal{E}_I$ and Mertens' estimate **??**,

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

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

In particular, $|A'| \gg M/(\log N)^{-1/101}$. Therefore 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.

However, if $n \in A''$ then both $n \mid x$ and $n \mid w_1w_2$ (since every q with $n \in A_q$ is in $\mathcal{E}_I$ and so divides either w_1 or w_2), and hence n divides

$$(x, w_1w_2) \leq (x, w_1)(x, w_2) \leq |x - w_1|\,|x - w_2| \leq N^2.$$

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

Proposition 6.2 (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 $M = N^{1-1/\log\log N}$ and $d_i = \lceil y \rceil + i$. By repeated applications of Lemma 5.5 we can find a sequence $A \supset A_0 \supset A_1 \supset \cdots \supset A_t$, where $d_t = \lceil z/4 \rceil - 1$, such that

$$R(A_i) \in [2/d_i - 1/M, 2/d_i) \quad \text{and} \quad R(A_i;q) \geq (\log N)^{-1/100} \text{ for all } q \in \mathcal{Q}_{A_i}.$$

(Note that the hypotheses of Lemma 5.5 continue to hold since

$$\frac{2}{d_i} - \frac{1}{M} \geq \frac{2}{d_i+1} + \frac{1}{(\log N)^{1/200}} \geq \frac{3}{(\log N)^{1/200}}$$

for all $0 \leq i \leq t$.) Let $0 \leq j \leq t$ be minimal such that there is a multiple of d_j in A_j. Such a j exists by assumption, since every $n \in A$ is divisible by some $d \in [y, z/4)$.

Suppose first that case (2) of Proposition 6.1 holds for A_j. The hypotheses of Proposition 4.16 are now met with $k = d_j$, $\eta = 1/2(\log N)^{1/100}$, 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 6.1 yields some $B \subset A_j$ such that

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

and where $\sum_{q \in \mathcal{Q}_B} \frac{1}{q} \leq \frac{2}{3} \log\log N$. Let $e_i = 4d_j + i$ and, once again, repeatedly apply Lemma 5.5 to find a sequence $B \supset B_0 \supset \cdots \supset B_r$, where $e_r = \lfloor z \rfloor$, such that

$$R(B_i) \in [2/e_i - 1/M, 2/e_i) \quad \text{and} \quad R(B_i; q) \geq (\log N)^{-1/100} \text{ for all } q \in \mathcal{Q}_{B_i}.$$

By minimality of j, no $d \in [y, d_j)$ divides any element of A_j, and hence every $n \in A_j$ is divisible by some $e \in [4d_j, z]$. In particular, there must exist some $0 \leq s \leq r$ such that B_s contains a multiple of e_s. Furthermore, since $\sum_{q \in \mathcal{Q}_{B_s}} \frac{1}{q} \leq \sum_{q \in \mathcal{Q}_B} \frac{1}{q} \leq \frac{2}{3} \log\log N$ we must be in the second case of Proposition 6.1. The hypotheses of Proposition 4.16 are now met with $k = e_s$ and η, K as above, and thus there is some $S \subset B_j \subset A$ such that $R(S) = 1/e_s$. $\square$