On a combinatorial method for counting smooth numbers in sets of integers

Ernie Croot *

February 2, 2007

Abstract

In this paper we develop a method for determining the number of integers without large prime factors lying in a given set S. We will apply it to give an easy proof that certain sufficiently dense sets A and B always produce the expected number of "smooth" sums $a+b$, $a \in A$, $b \in B$. The proof of this result is completely combinatorial and elementary.

1 Introduction

Given a set S, a common question one tries to answer is whether S contains the expected number of "y-smooth" integers, which are those integers n where $P(n) \leq y$, where $P(n)$ denotes the largest prime factor of n. We denote the number of integers in S with this property by $\Psi(S, y)$; and for a number $N > 0$, we denote the set of all y-smooths positive integers $\leq N$ by $\Psi(N, y)$. So,

$$\Psi([N], y) \ = \ \Psi(N, y),$$

where here $[N]$ denotes the set of integers $\{1, 2, ..., N\}$.

If $S \subseteq [N]$ is "typical", then one would expect that

$$\frac{\Psi(S, y)}{|S|} \ \sim \ \frac{\Psi(N, y)}{N}. \tag{1}$$

For example, fix a real number $0 < \theta \leq 1$ and an integer $a \neq 0$, and let S be the set of numbers of the form $p + a \leq N$, where p is prime. S is often

*Supported by an NSF grant.

called a set of "shifted primes". It is conjectured that

$$\Psi(S, N^\theta) \ \sim\ \frac{\pi(N)\Psi(N, N^\theta)}{N} \ \sim\ \rho(\theta^{-1})\pi(N), \tag{2}$$

where

$$\rho(u) \ =\ \lim_{N\to\infty} \frac{\Psi(N, N^{1/u})}{N}.$$

This function ρ is called Dickman's function, and it was proved in [7] that the limit exists. Unfortunately, proving (2) remains a difficult, open problem; however, in [9] J. B. Friedlander gave a beautiful proof that for $\theta > (2\sqrt{e})^{-1} = 0.30326...$,

$$\Psi(S, N^\theta) \gg \pi(N),$$

and in [1], R. Baker and G. Harman proved that for $\theta \geq 0.2961$,

$$\Psi(S, N^\theta) \ >\ \frac{N}{\log^\alpha N},$$

for some $\alpha > 1$ and $N > N_0(a)$.

There are several methods for attacking the general question of proving that (1) holds for a particular set S. One such method involves exponential sums and the circle method, and another uses a Buchstab or inclusion-exclusion identity in combination with a sieve method (such as the Large Sieve). For example, one way that one could count the number of y-smooth integers in a set S is via the following inclusion-exclusion identity:

$$\Psi(S, y) \ =\ |S| - \sum_{\substack{y < p \leq N \\ p \text{ prime}}} |S_p| + \sum_{\substack{y \leq p_1 < p_2 \leq N \\ p_1, p_2 \text{ prime}}} |S_{p_1 p_2}| - \sum_{\substack{y \leq p_1 < p_2 < p_3 \leq N \\ p_1, p_2, p_3 \text{ prime}}} |S_{p_1 p_2 p_3}| + \cdots,$$

where S_d denotes the set of elements of S divisible by d. One problem that one immediately faces is that these prime products $d = p_1 p_2 \cdots p_k$ can be very close to N, and when that is the case, in many applications one does not have good estimates for the size of S_d. That problem can be fixed if one knows that most of the elements of S have a divisor of size between, say, N^δ and $N^{1-\delta}$ that is a product of primes $< y$, because no such integer could have a divisor $d > N^{1-\delta}$ that is a product of primes $\geq y$. One would also need asymptotic estimates for the sizes of S_d for all $d < N^{1-\delta}$ in order to make this approach work; in particular, something like

$$|S_d| \ =\ \frac{|S|(1 + o(1))}{d}, \text{ for all } d < N^{1-\delta}. \tag{3}$$

Actually, all one needs is that this holds for "most $d < N^{1-\delta}$ ", in some appropriate sense. Often, one only has such estimates for $d < \sqrt{N}$ (as is the case for shifted primes). In the smooth sieve method that we will present, asymptotic estimates for S_d will not be needed; all that will be needed are good lower bounds on the size of S_d in a certain average sense. Furthermore, our method works when different positive weightings of the naturals are applied (besides the usual weighting, which assigns value 1 to every natural number).

1.1 Local-Global Sets and the Main Theorem

The key structure that we use in the development of our smooth sieve method is a "Local-Global Set", which we abbreviate as "LG set", and define as follows:

Definition. Let $\epsilon, c \in (0,1)$ be parameters. Then, $\mathcal{A} \subseteq [N]$ is an LG-set with parameters ϵ, c if the following conditions hold:

- If $q_1, q_2 \in \mathcal{A}$ are distinct then $[q_1, q_2] > N$;
- $|\{n \leq N \ : \ q|n \text{ for some } q \in \mathcal{A} \text{ with } q \leq N^c\}| \geq (1-\epsilon)N$.

Notes: This second condition is saying that all but an ϵ proportion of the integers $n \leq N$ are divisible by some element of $\mathcal{A}$ of size at most N^c. The first condition tells us that

$$\sum_{q \in \mathcal{A}} \frac{1}{q} \ \leq \ 1 + O(|\mathcal{A}|/N)$$

(consider the number of $n \leq N$ which are divisible by some $q \in \mathcal{A}$). The sets $\mathcal{A}$ that we construct in this paper will have $|\mathcal{A}| = o(N)$, and so we will have

$$\sum_{q \in \mathcal{A}} \frac{1}{q} \ \leq \ 1 + o(1).$$

On the other hand, the second condition implies that

$$\sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \frac{1}{q} \ \geq \ \frac{1}{N} \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \left\lfloor \frac{N}{q} \right\rfloor \ \geq \ 1 - \epsilon \tag{4}$$

The last inequality here follows from the fact that the second sum counts the number of positive integers $n \leq N$ that are divisible by an element $q \in \mathcal{A}$

satisfying $q \le N^c$; and, such n can be divisible by at most one such q by the first property of LG sets. So, the LG sets we will be considering will have that the sum of reciprocals of elements of $\mathcal{A}$ that are $\le N^c$ are within ϵ of 1.

A central result of the paper, proved in section 3, is the following existence theorem:

Theorem 1 *For every $0 < \epsilon < \epsilon_0$, for every*

$$c \ \ge \ 1 - \epsilon^{2250},$$

and for every $N > N_0(\epsilon)$, there exists an LG set of integers $\mathcal{A} \subseteq [N]$ with parameters ϵ and c.

The origin of our name "Local-Global Set" comes from the following observation: Suppose that $\mathcal{A} \subseteq [N]$ is an LG set with parameters ϵ and c. Suppose that $S \subseteq [N]$, and let $S(a;q)$ denote the number of integers in S that are congruent to a modulo q. Fix a non-negative integer $a \le N$, and suppose that we can prove that

$$\text{For every } q \in \mathcal{A},\ S(a;q) \ > \ \frac{(1-\epsilon)|S|}{q}. \tag{5}$$

If S were equidistributed amongst the residue classes modulo q, then $S(a;q) = |S|/q + O(1)$; so, we are assuming a lower bound that is off from the expected amount by a factor of $1-\epsilon$. Now, since $\mathcal{A}$ is an LG set we have that each element of $S - a$ can be divisible by at most one member of $\mathcal{A}$; so,

$$\sum_{q \in \mathcal{A}} \left(S(a;q) - \frac{|S|}{q} \right) \ \le \ |S| \left(1 - \sum_{q \in \mathcal{A}} \frac{1}{q} \right) \ \le \ \epsilon |S|.$$

If we let $\mathcal{A}_0$ be those $q \in \mathcal{A}$ such that $S(a;q) < |S|/q$, then we have from (5) that for every $q \in \mathcal{A}_0$,

$$\left| S(a;q) - \frac{|S|}{q} \right| \ < \ \frac{\epsilon |S|}{q}.$$

Thus,

$$\begin{aligned} \sum_{q \in \mathcal{A}} \left| S(a;q) - \frac{|S|}{q} \right| \ &= \ \sum_{q \in \mathcal{A}} \left(S(a;q) - \frac{|S|}{q} \right) \ + \ 2 \sum_{q \in \mathcal{A}_0} \left| S(a;q) - \frac{|S|}{q} \right| \\ &< \ 4\epsilon |S|. \end{aligned}$$

Thus, a good uniform lower bound (5) implies equidistribution of the elements of S in residue classes $a \pmod q$ in some average sense. One can think of this as some type of local-to-global phenomenon.

1.2 The Local-Global Sieve for Smooth Numbers

Our smooth sieve result below will be stated in terms of weight functions, rather than sets, as they are more general and more flexible than sets. So, we suppose that $w(n) \geq 0$ is defined for natural numbers, and then our problem is to give good estimates for the size of

$$W(N,\theta) \ := \sum_{\substack{n \leq N \\ P(n) \leq N^\theta}} w(n).$$

Letting

$$\Sigma \ := \sum_{n \leq N} w(n),$$

we note that since the sum $W(N,\theta)$ has only a fraction $\rho(\theta^{-1})$ times as many terms as Σ, we expect that

$$W(N,\theta) \ \sim \ \rho(1/\theta)\Sigma.$$

Our smooth sieve, which is proved in section 4, gives conditions for when this is the case:

Theorem 2 *Suppose* $0 < \epsilon < \epsilon_0$*, and let* $\mathcal{A}$ *be the LG set with parameters* ϵ *and* $c = 1 - \epsilon^{2250}$ *given by Theorem 1.* [1] *Let*

$$\mathcal{A}_1 \ = \ \{q \in \mathcal{A}, q \leq N^c \ : \ P(q) \leq N^\theta\}, \text{ and } \mathcal{A}_2 \ = \ \{q \in \mathcal{A}, q \leq N^c \ : \ P(q) > N^\theta\}.$$

Define the constants

$$\rho_1 \ = \ \sum_{q \in \mathcal{A}_1} \frac{1}{q}, \text{ and } \rho_2 \ = \ \sum_{q \in \mathcal{A}_2} \frac{1}{q}.$$

[1] This theorem makes use of the structure of the LG set given by theorem 1, which is described in section 3; and so, the theorem doesn't hold for just *any* LG set with parameters ϵ and c.

Suppose

$$\sum_{q\in\mathcal{A}_1}\sum_{\substack{n\le N\\ q|n}} w(n) \ > \ (\rho_1-\epsilon)\Sigma; \text{ and,} \tag{6}$$

$$\sum_{q\in\mathcal{A}_2}\sum_{\substack{n\le N\\ q|n}} w(n) \ > \ (\rho_2-\epsilon)\Sigma. \tag{7}$$

Then, we have that

$$(\rho_1-\epsilon)\Sigma \ < \ W(N,\theta) \ < \ (\rho_1+2\epsilon)\Sigma, \tag{8}$$

and

$$(\rho(1/\theta)-3\epsilon)\Sigma \ < \ W(N,\theta) \ < \ (\rho(1/\theta)+4\epsilon)\Sigma. \tag{9}$$

Moreover, if one is only able to show (6), then one can still deduce the lower bound

$$W(N,\theta) \ > \ (\rho_1-\epsilon)\Sigma \ > \ (\rho(1/\theta)-3\epsilon)\Sigma. \tag{10}$$

To see that the assumptions in this theorem are weaker than (3), let $w(n)$ be the indicator function for a set $S\subseteq[N]$. Then, (6) and (7) are just saying that we have a lower bound of the general form

$$S_d \ > \ \frac{(1-\epsilon)|S|}{d}$$

on average.

We will close this section by giving a simple application of the above theorem, and in the next section we will give a more sophisticated application: Suppose that $S\subseteq[N]$ and $|S|=N^{1-o(1)}$. How many ordered pairs $(s_1,s_2)\in S\times S$ have the property that s_1-s_2 is N^θ-smooth? We expect there to be about $\rho(1/\theta)|S|^2$ such pairs. To prove this using the above theorem, define $w(n)$ to be the number of ways of writing $n=s_1-s_2$, $s_1,s_2\in S$. Then, the sum over all $w(n)$ with $n\ge 1$ divisible by q is

$$\sum_{a=0}^{q-1}\binom{S(a;q)}{2},$$

This expression is minimized if the elements of S are as equidistributed amongst the residue classes modulo q as is possible; and so, this expression can be shown to be at least $\sim|S|^2/2q$ in size for $q=o(|S|)$. So, for any $\epsilon>0$

we have that for N sufficiently large both (6) and (7) hold. It follows that the number of pairs (s_1, s_2), $s_1 > s_2$ such that $P(s_1 - s_2) \le N^\theta$ is $\sim \rho(1/\theta)|S|^2/2$; so, there are $\sim \rho(1/\theta)|S|^2$ pairs (s_1, s_2) with $P(s_1 - s_2) \le N^\theta$.

The remainder of the paper is organized as follows. In the next section we will prove that under certain conditions a set S has about $\rho(1/\theta)|S|^2$ ordered pairs $(s_1, s_2) \in S \times S$ such that $P(s_1 + s_2) \le N^\theta$; and, in the process of proving this, we will present a large sieve inequality that follows easily from properties of LG sets. In the final sections we will give proofs of Theorems 1 and 2.

2 An Application of Theorems 1 and 2

Given sets of integers $A, B \subseteq [N]$ having $\gg N$ elements each, it is an interesting and studied question to determine the number of ordered pairs $(a, b) \in A \times B$ such that $P(a + b) \le y$. There are several ways of attacking this sort of problem, one of which is to use the circle method and exponential sums over smooth numbers, and another is to use the large sieve. We could also ask how $\tau(a + b)$ (the number of divisors of $a + b$) is distributed, or how large $P(a + b)$ can be. These types of questions were given a thorough treatment in a series of beautiful papers by A. Balog and A. Sárközy [2], [3], [4], and [5]; P. Erdős, H. Maier, and A. Sárközy [8]; A. Sárközy and C. L. Stewart [11], [12], [13], [14]; C. Pomerance, A. Sárközy, and C. L. Stewart [10]; and R. de la Bretèche [6]. The paper by de la Bretèche is more relevant to the main result of this section, and we give here one of his theorems:

Theorem 3 *Suppose that $A, B \subseteq [N]$. For a given integer $y \le N$, let $u = (\log N)/\log y$. Then, uniformly for $N \ge 3$, $\exp((\log N)^{2/3+\epsilon}) < y \le N$ we have*

$$\begin{aligned} &\#\{a \in A,\ b \in B\ :\ P(a + b) \le y\} \\ &\qquad = |A||B|\rho(u)\left(1 + O\left(\frac{N}{\sqrt{|A||B|}}\frac{\log(u + 1)}{\log y}\right)\right). \end{aligned}$$

R. de la Bretèche used estimates for exponential sums and the circle method to prove this result. Notice that if $|A||B| \ll (N/\log N)^2$, then his result fails to prove that there are the expected number of sums that are y-smooth for any $y < N$, because in this case the big-Oh term is $\gg 1$. What makes his theorem so powerful is the fact that the parameter y is allowed to go all the way down to $\exp((\log N)^{2/3+\epsilon})$.

Let us now consider what happens in the case when $y = N^{\theta}$ (and so $u = 1/\theta$): Is it possible to show that if $|A|, |B| > N^c$, for some $0 < c < 1$, then we get the expected number of sums $a + b$ being y-smooth? It is easy to see that the answer is no, no matter how close to 1 we take c to be. For example, if $\theta + c > 1$ we could take A and B to both be the set of integers $\leq N$ that are divisible by some prime number $p \sim N^{1-c}$. Notice here that $|A| \sim N^c$. The sums $a + b$, $a, b \in A$ are numbers of the form pk, where $k < 2N^c$, and such a sum is N^{θ}-smooth if and only if k is N^{θ}-smooth. Thus, one would expect (and can show) that

$$\frac{\#\{a, b \in A \ : \ P(a + b) \leq N^{\theta}\}}{|A||B|} \ \sim \ \frac{\Psi(2N^c, N^{\theta})}{2N^c} \ \sim \ \rho(c/\theta).$$

On the other hand, the proportion of N^{θ} smooths $\leq N$ is $\sim \rho(1/\theta)$, which is not $\rho(c/\theta)$. So, the type of result we might try to prove is the following:

Theorem 4 *Given $0 < \theta \leq 1$, and $0 < \gamma < \gamma_0$ if $A, B \subseteq [N]$ satisfy $|A|, |B| > (8/\gamma)N^{1-(\gamma/8)^{2250}}$, then for N sufficiently large*

$$\#\{a \in A,\ b \in B \ : \ P(a+b) \leq (2N)^{\theta}\} \ = \ (\rho(1/\theta)+\delta)|A||B|, \text{ where } |\delta| < \gamma.$$

The same result holds for differences $a - b$.

In section 2.2 we will give a short proof of this theorem using Theorem 2 and a version of the Large Sieve.

2.1 A Local-Global Large Sieve

For our proof of Theorem 4 we will require a form of the Large Sieve, which can be proved via modifying the usual proofs of the large sieve; however, we will prove it here (perhaps surprisingly) through a brief and elementary application of LG sets.

Theorem 5 *Given $\epsilon > 0$ and N sufficiently large, let $\mathcal{A} \subseteq [N]$ be any LG set for parameters ϵ and c. Suppose that $C \subseteq [N]$, and let $C(a, q)$ denote the number of elements of C that are congruent to a modulo q. Then, we have that*

$$\sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} \left(C(a, q) - \frac{|C|}{q} \right)^2 \ < \ |C|(\epsilon|C| + N^c).$$

Proof. We note that if $b, c \in C$, and $b \neq c$, then if $q \in \mathcal{A}$ divides $b - c$, we must have that q is unique; otherwise, if $q' \in A$ also divides $b - c$, then $[q, q'] > N$ and $[q, q'] | (b - c)$.

Thus, we have that

$$\begin{aligned} |C|^2 &> \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \#\{(b,c) \in C^2 \ : \ b \neq c, \ q | (b-c)\} \\ &= \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} \left(C(a,q)^2 - C(a,q) \right) \\ &\geq -N^c |C| \ + \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} C(a,q)^2. \end{aligned}$$

Thus,

$$|C|(N^c + |C|) \ < \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} C(a,q)^2.$$

From this and (4) it follows that

$$\begin{aligned} \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} \left(C(a,q) - \frac{|C|}{q} \right)^2 &= \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \sum_{a=0}^{q-1} C(a,q)^2 - |C|^2 \sum_{\substack{q \in \mathcal{A} \\ q \leq N^c}} \frac{1}{q} \\ &< (1 - (1-\epsilon))|C|^2 + N^c |C| \\ &= |C|(\epsilon |C| + N^c). \qquad \blacksquare \end{aligned}$$

2.2 Proof of Theorem 4.

We let $\epsilon = \gamma/8$, and then for $c = 1 - \epsilon^{2250}$ we have from Theorem 1 that for N sufficiently large, there is an LG set $\mathcal{A} \subseteq [2N]$ with parameters ϵ, c. We let $\mathcal{A}_1$ and $\mathcal{A}_2$ be as in Theorem 2 (with N replaced with $2N$).

Let α, β be the indicator functions for the sets A and B, respectively; let

$$A(a,q) \ = \sum_{\substack{n \leq N \\ n \equiv a \pmod q}} \alpha(n), \ \text{and} \ B(a,q) \ = \sum_{\substack{n \leq N \\ n \equiv a \pmod q}} \beta(n);$$

and define the weight function

$$w(n) \ = \ (\alpha * \beta)(n) \ = \sum_{a+b=n} \alpha(a)\beta(b).$$

Then, for $\mathcal{A}' = \mathcal{A}_1$ or $\mathcal{A}_2$ we get

$$\begin{aligned}\sum_{q\in\mathcal{A}'}\sum_{\substack{n\le 2N\\ q|n}} w(n) &= \sum_{q\in\mathcal{A}'}\sum_{a=0}^{q-1} A(a,q)B(q-a,q)\\ &= \sum_{q\in\mathcal{A}'}\sum_{a=0}^{q-1}(A(a,q)-|A|/q)(B(q-a,q)-|B|/q)\\ &\qquad + |A||B|\sum_{q\in\mathcal{A}'}\frac{1}{q}. \end{aligned} \tag{11}$$

Applying the Cauchy-Schwarz inequality and Theorem 5 with $C = A$ and $C = B$ to this last double sum we deduce for $|A|, |B| > \epsilon^{-1}N^c$ that

$$\sum_{q\in\mathcal{A}'}\sum_{a=0}^{q-1}|A(a,q)-|A|/q||B(a,q)-|B|/q| \ < \ 2\epsilon|A||B|. \tag{12}$$

Combining this with (11) we deduce that

$$\sum_{q\in\mathcal{A}'}\sum_{\substack{n\le 2N\\ q|n}} w(n) \ > \ |A||B|\left(-\frac{\gamma}{4}+\sum_{q\in\mathcal{A}'}\frac{1}{q}\right).$$

Thus, the conditions of Theorem 2 are met for ϵ replaced with $\gamma/4$, and we deduce that since $\Sigma = |A||B|$ in our case, then

$$\left|\#\{a\in A,\ b\in B\ :\ P(a+b)\le(2N)^\theta\} - \rho(1/\theta)|A||B|\right| \ < \ \gamma|A||B|. \quad \blacksquare$$

3 Proof of Theorem 1

In the course of our proof we will need to make use of the following consequence of Brun's upper bound sieve:

Theorem 6 *Suppose that $\mathcal{P}$ is a subset of the primes $\le N$. The number of integers $\le N$ not divisible by any prime in $\mathcal{P}$ is*

$$\ll \ N\prod_{p\in\mathcal{P}}\left(1-\frac{1}{p}\right).$$

Define $A(k,\delta)$ to be the set of all integers $q \le N$ of the form $p_1 \cdots p_k$, $p_1 > p_2 > \cdots > p_k > N^\delta$, where

$$\text{For } i = 1,2,...,k-1,\ \frac{N}{p_1 \cdots p_i} \ge p_i;\ \text{while},\ \frac{N}{p_1 \cdots p_k} < p_k. \tag{13}$$

We will show that for every $0 < \epsilon < \epsilon_0$, there exists $\delta > 0$ so that

$$\mathcal{A} = \bigcup_{k \ge 1} A(k,\delta) \tag{14}$$

is an LG set with parameters ϵ and $c = 1 - \epsilon^{2250}$.

3.1 The Set $\mathcal{A}$ satisfies the First Property for an LG Set

First, we show that the set $\mathcal{A}$ described in the previous subsection satisfies the first condition for being an LG set, namely that for any distinct pair of integers $n_1, n_2 \in \mathcal{A}$, we have $\mathrm{lcm}(n_1,n_2) > N$: Say $n_1 \in A(k,\delta)$ and $n_2 \in A(\ell,\delta)$; so, we have the prime factorizations

$$\begin{aligned} n_1 &= p_1 \cdots p_k,\ p_1 > p_2 > \cdots > p_k;\ \text{and} \\ n_2 &= q_1 \cdots q_\ell,\ q_1 > q_2 > \cdots > q_\ell. \end{aligned}$$

Without loss of generality, we can assume that $p_k \le q_\ell$.

Now, if there is some prime q_i which is distinct from the primes $p_1,...,p_k$, then we would have that the lcm of n_1 and n_2 is divisible by the product of primes $q_i p_1 \cdots p_k$, and this product exceeds N, because from (13)

$$q_i p_1 \cdots p_k \ > \ q_i \frac{N}{p_k} \ \ge \ q_\ell \frac{N}{p_k} \ \ge \ N.$$

So we are left to consider what happens when the q_i's are a subset of the p_i's. We break this case into two sub-cases, with the first one where $p_k = q_\ell$, and the second where $p_k < q_\ell$.

In the case $p_k = q_\ell$, we must have that there exists one of the primes $p_i > q_\ell$ such that p_i is distinct from $q_1,...,q_\ell$, since otherwise we would have $n_1 = n_2$. But now our assumption gives

$$\frac{N}{p_1 \cdots p_k} \ \le \ \frac{N}{p_i q_1 \cdots q_\ell} \ < \ \frac{q_\ell}{p_i} \ \le \ 1,$$

which is impossible.

So, we may assume $p_k < q_\ell$. For this case, let $j < k$ be the index where $p_j = q_\ell$ (which exists since q_i's are a subset of the p_i's). Then, we have

$$p_1 \cdots p_j \ \geq \ q_1 \cdots q_\ell.$$

From (13) this gives

$$q_\ell \ = \ p_j \ \leq \ \frac{N}{p_1 \cdots p_j} \ \leq \ \frac{N}{q_1 \cdots q_\ell} \ < \ q_\ell,$$

which is impossible. So, we conclude that the set $\mathcal{A}$ given by (14) satisfies the first condition for being an LG set.

3.2 The Set $\mathcal{A}$ Satisfies the Second Property for an LG Set

Define F to be the set of all integers n satisfying

$$\frac{N}{\log N} \ < \ n \ < \ N,$$

such that

$$\text{If } n = u^2 v, \ v \text{ squarefree, then } u < \log N.$$

We have that the number of $n \leq N$ that are not contained in F is at most

$$N \sum_{u \geq \log N} \frac{1}{u^2} \ \ll \ \frac{N}{\log N}.$$

So,

$$|F| \ = \ N - O(N/\log N).$$

Now define G to be the integers $n \in F$ that are not divisible by any element of $\mathcal{A}$. Our goal will be to show that $|G|$ is "small".

We first observe that one can deduce from (13) that every $n \in G$ has the property that for all $N^\delta < x \leq N$,

$$\frac{N}{\prod_{\substack{p|n, \ p \text{ prime} \\ p \geq x}} p} \ \geq \ x. \tag{15}$$

For if q is the largest x for which this inequality fails (such q exists), then q is prime, and we would have

$$\frac{N}{\prod_{\substack{p|n, \ p \text{ prime} \\ p \geq q}} p} \ < \ q,$$

which would imply $\prod_{p \ge q,\ p|n} p$ is an element of $\mathcal{A}$ dividing $n \in G$, contradiction.

We conclude that for $n \in G$,

$$\prod_{\substack{p|n \\ p<x}} p \ > \ \frac{n}{(\log^2 N) \prod_{\substack{p|n \\ p \ge x}} p} \ > \ \frac{N}{(\log^3 N) \prod_{\substack{p|n \\ p \ge x}} p} \ \ge \ \frac{x}{\log^3 N}. \qquad (16)$$

This brings us now to the conceptual heart of the proof of Theorem 1: Given $x \le N$, the average of $\sum_{p|n, p<x} \log p$ over all $n \le N$ is $\log x + O(1)$; however, for "most" n, as we vary x, this sum should fluctuate about $\log x$, meaning that for "most" n there is an $N^\delta < x \le N$ for which (16) fails to hold. The rest of the proof amounts to formalizing this intuition.

Let $R = 5 \times 10^8$, and let

$$J \ = \ J(\delta) \ = \ \left\lfloor \frac{\log(1/2\delta)}{\log(R)} \right\rfloor. \qquad (17)$$

Note that $R^J \delta \ \le \ 1/2$.

Let $G(h)$ be the set of all $n \in G$ for which exactly h integers j among $1, 2, ..., J$ satisfy

$$\sum_{\substack{p|n,\ p \text{ prime} \\ p \le N^{\delta R^j}}} \log p \ > \ \frac{3\delta R^j \log N}{2}.$$

Write

$$G \ = \ G_1 \ \cup \ G_2,$$

where

$$G_1 \ = \bigcup_{0 \le h \le J/14} G(h), \text{ and } G_2 \ = \bigcup_{J/14 < h \le J} G(h).$$

To show that $|G|$ is "small" we will show that both $|G_1|$ and $|G_2|$ are "small".

3.2.1 $|G_2|$ is "small"

Define

$$z(n) \ = \ z(n, \delta, R) \ := \ \prod_{j=1}^{J} \left(\sum_{\substack{p|n,\ p \text{ prime} \\ p \le N^{\delta R^j}}} \log p \right).$$

Since every $n \in G(h)$ also lies in G, from (16) we have that for $j = 1, ..., J$,

$$\sum_{\substack{p|n,\ p \text{ prime} \\ p \le N^{\delta R^j}}} \log p \ > \ \delta R^j \log N - 3 \log\log N.$$

So, for $n \in G(h)$ we will have

$$z(n) \ > \ (1 + o(1))(\delta \log N)^J R^{J(J+1)/2} (3/2)^h. \tag{18}$$

Lemma 1 *We have that*

$$\sum_{n \le N} z(n) \ \ll \ N(\delta \log N)^J R^{J(J+1)/2} \exp(2J/\sqrt{R}).$$

Proof of the Lemma. We have that

$$\begin{aligned}
\sum_{n \le N} z(n) \quad &= \quad \sum_{\substack{p_1, \ldots, p_J \text{ prime} \\ p_j \le N^{\delta R^j}}} (\log p_1) \cdots (\log p_J) \left\lfloor \frac{N}{[p_1, ..., p_J]} \right\rfloor \\
&< \quad N \sum_{\substack{p_1, \ldots, p_J \text{ prime} \\ p_j \le N^{\delta R^j}}} \frac{(\log p_1) \cdots (\log p_J)}{[p_1, ..., p_J]} \\
&\le \quad N \sum_{\substack{a_1, \ldots, a_J \\ 0 \le a_j \le J - j + 1 \\ j \le a_1 + \cdots + a_j \le J}} \prod_{\substack{1 \le j \le J \\ a_j \ne 0}} \left(\sum_{\substack{p \le N^{\delta R^j} \\ p \text{ prime}}} \frac{\log^{a_j} p}{p} \right) \\
&\ll \quad N(\delta \log N)^J \sum_{\substack{a_1, \ldots, a_J \\ 0 \le a_j \le J - j + 1 \\ j \le a_1 + \cdots + a_j \le J}} \prod_{\substack{1 \le j \le J \\ a_j \ne 0}} \frac{R^{j a_j}}{a_j}.
\end{aligned} \tag{19}$$

Here we have used the basic estimate for $a \ge 1$

$$\sum_{\substack{p \le M \\ p \text{ prime}}} \frac{\log^a p}{p} \ \sim \ \int_2^M \frac{(\log x)^{a-1} dx}{x} \ \sim \ \frac{\log^a M}{a}.$$

Now, suppose that for every $j \le J$ we have

$$j \ \le \ a_1 + \cdots + a_j \ \le \ J, \text{ where } 0 \le a_j \le J - j + 1,$$

and that exactly k of the a_is are non-zero. Then,

$$\prod_{1\le j\le J} R^{ja_j} \ \le \ R^{J(J+1)/2-J+k}.$$

Thus, the final expression of (19) is

$$\begin{aligned}
&\ll \quad N(\delta\log N)^J R^{J(J+1)/2-J} \sum_{1\le k\le J} \binom{J}{k} \sum_{\substack{b_1+\cdots+b_k=J\\ b_i\ge 1}} R^k \\
&= \quad N(\delta\log N)^J R^{J(J+1)/2-J} \sum_{1\le k\le J} R^k \binom{J}{k}\binom{J-1}{k-1} \\
&\ll \quad N(\delta\log N)^J R^{J(J+1)/2-J+1/2} \sum_{1\le k\le J} R^{k-1/2} \binom{2J-1}{2k-1} \\
&< \quad N(\delta\log N)^J R^{J(J+1)/2} \left(1+\frac{1}{\sqrt{R}}\right)^{2J-1} \\
&\ll \quad N(\delta\log N)^J R^{J(J+1)/2} \exp(2J/\sqrt{R}). \quad \blacksquare
\end{aligned}$$

With our choice of $R = 5\times 10^8$ we will have from Lemma 1 that for $\delta > 0$ sufficiently small,

$$\sum_{n\in G} z(n) \ < \ \frac{N}{2}(1.0001\ \delta\log N)^J R^{J(J+1)/2}.$$

Combining this with (18) we deduce

$$|G(h)| \ < \ (1.0001)^J (2/3)^h N.$$

It follows that for N sufficienty large,

$$|G_2| \ < \ 3(1.0001)^J (2/3)^{J/14} N \ < \ \frac{N(0.98)^J}{3}.$$

3.2.2 $|G_1|$ is "small"

To bound $|G_1|$ from above, we will bound $|G(h)|$ from above for $h \le J/14$: Given $h \le J/14$ we split $G(h)$ into smaller subsets as follows

$$G(h) \ = \ \bigcup_{\substack{B\subseteq[J]\\ |B|=h}} G(h;B),$$

where $G(h;B)$ is the set of all $n \in G$ with the property that

$$\sum_{\substack{p|n,\ p \text{ prime} \\ p \leq N^{\delta R^j}}} \log p \ \geq\ \frac{3\delta R^j \log N}{2} \quad \text{if and only if} \quad j \in B.$$

Given B, we let $B' = [J] \setminus B$. We note that every $n \in G(h;B)$ has no prime divisors p lying in any of the intervals

$$[N^{3\delta R^j/4} \log^2 N,\ N^{\delta R^j}],\ \text{where } j \in B'. \tag{20}$$

For if some $n \in G(h;B)$ had such a prime divisor p, then we would have that

$$\prod_{\substack{q|n,\ q \text{ prime} \\ q<p}} q \ \leq\ \frac{1}{p} \prod_{\substack{q|n,\ q \text{ prime} \\ q \leq N^{\delta R^j}}} q \ <\ \frac{N^{3\delta R^j/4}}{\log^2 N} \ \leq\ \frac{p}{\log^4 N},$$

which would contradict (16), and therefore we would have that $n \notin G$.

From Theorem 6, together with the fact that $n \in G(h;B)$ has no prime divisors lying in the intervals (20), we deduce that

$$|G(h;B)| \ \ll\ N \prod_{j \in B'} \prod_{\substack{N^{3\delta R^j/4} \log^2 N < p < N^{\delta R^j} \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right) \ \ll\ (3/4)^{|B'|} N.$$

Thus, for $h \leq J/14$ and for $\delta > 0$ sufficiently small, we will have from Stirling's formula that

$$|G(h)| \ \ll\ N \binom{J}{h} (3/4)^{J-h} \ <\ \frac{N(0.991)^J}{3J}.$$

Thus, for $\delta > 0$ sufficiently small and N sufficiently large,

$$|G_1| \ <\ \frac{N(0.991)^J}{3}.$$

3.3 G is "small", and the Conclusion of the Proof of the Theorem

From the upper bounds on $|G_1|$ and $|G_2|$ we deduce for $R = 5 \times 10^8$ and $\delta > 0$ sufficiently small and N sufficiently large,

$$|G| \ \leq\ |G_1| + |G_2| \ \leq\ \frac{2N(0.991)^J}{3};$$

and so, the number of integers $n \leq N$ not divisible by any element from the set $\mathcal{A}$ is at most

$$(N - |F|) + |G| \ < \ N(0.991)^J.$$

Thus, we have that

$$\frac{1}{N} \sum_{a \in \mathcal{A}} \left\lfloor \frac{N}{a} \right\rfloor \ > \ 1 - (0.991)^J.$$

Now, since the prime divisors of the elements of $\mathcal{A}$ exceed N^δ, we will have from Theorem 6 and Mertens' Theorem that for $N^{1/2} < M < N$ there are

$$\ll \ M \prod_{\substack{p \leq N^\delta \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right) \ \ll \ \frac{M}{\delta \log N}$$

elements of $\mathcal{A}$ that are less than M. Thus, for $1/2 < c \leq 1$,

$$\sum_{\substack{a \in \mathcal{A} \\ N^c < a \leq N}} \frac{1}{a} \ \ll \ \frac{\log(1/c)}{\delta}.$$

It follows that that the number of integers $n \leq N$ divisible by some element of $\mathcal{A}$ that is $\leq N^c$ is

$$\sum_{\substack{a \in \mathcal{A} \\ a \leq N^c}} \left\lfloor \frac{N}{a} \right\rfloor \ > \ N(1 - (0.991)^J - C\delta^{-1} \log(1/c)),$$

where $C > 0$ is some constant.

This will exceed $N(1 - \epsilon)$ provided

$$(0.991)^J \ < \ \frac{\epsilon}{2}, \text{ and } C\delta^{-1} \log(1/c) \ < \ \frac{\epsilon}{2}. \tag{21}$$

From (17) we have that for $\epsilon < 1/2$ the first inequality holds provided

$$\delta \ < \ (\epsilon/2)^{1 - \log(R)/\log(0.991)},$$

The second inequality of (21) holds provided

$$\log(c) \ > \ \frac{-\epsilon\delta}{2C} \ > \ \frac{-(\epsilon/2)^{2 - \log(R)/\log 0.991}}{C}.$$

So, for $\epsilon > 0$ sufficiently small and

$$c \ > \ 1 - \epsilon^{2250} \ > \ 1 - \frac{(\epsilon/2)^{2 - (\log R)/\log 0.991}}{2C},$$

all but at most $N(1 - \epsilon)$ of the integers $n \leq N$ will be divisible by some element of $\mathcal{A}$. ∎

4 Proof of Theorem 2.

First, we remark that

$$1-\epsilon \ \le \ \sum_{\substack{n\in\mathcal{A}\\ n\le N^c}} \frac{1}{n} \ = \ \rho_1+\rho_2 \ < \ 1+o(1). \tag{22}$$

From the properties of the set $\mathcal{A}$ defined at the beginning of section 3 we have that if $q|n$, $n \le N$, and $q \in \mathcal{A}_1$, then $n/q < q_0 \le N^\theta$, where q_0 is the smallest prime divisor of q. Thus, n is N^θ-smooth. Also, if $q|n$ where $q \in \mathcal{A}_2$, then n is obviously not N^θ-smooth, since q has a prime factor exceeding N^θ in this case.

We deduce from this and (6) that

$$W(N,\theta) \ \ge \ \sum_{q\in\mathcal{A}_1} \sum_{\substack{n\le N\\ q|n}} w(n) \ \ge \ (\rho_1-\epsilon)\Sigma.$$

Also,

$$W(N,\theta) \ \le \ \Sigma - \sum_{q\in\mathcal{A}_2} \sum_{\substack{n\le N\\ q|n}} w(n),$$

which, together with (6), (7) and (22), implies

$$W(N,\theta) \ \le \ (1-\rho_2+\epsilon)\Sigma \ \le \ (\rho_1+2\epsilon)\Sigma.$$

Thus, we have proved (8) and the first inequality in (10).

To prove (9) and the second inequality in (10) we must relate $\rho(1/\theta)$ and ρ_1 and ρ_2. To do this we observe that since

$$\rho(1/\theta) \ \ge \ \frac{(1+o(1))\Psi(N,N^\theta)}{N} \ \ge \ \frac{(1+o(1))}{N} \sum_{q\in\mathcal{A}_1} \left\lfloor \frac{N}{q} \right\rfloor \ = \ \rho_1+o(1),$$

and

$$1-\rho(1/\theta) \ \ge \ 1+o(1)-\frac{\Psi(N,N^\theta)}{N} \ \ge \ \frac{(1+o(1))}{N} \sum_{q\in\mathcal{A}_2} \left\lfloor \frac{N}{q} \right\rfloor \ = \ \rho_2+o(1),$$

we deduce from (22) that

$$o(1) \ < \ \rho(1/\theta)-\rho_1 \ < \ \epsilon+o(1).$$

Thus, (9) follows.

5 Acknowledgements

I would like to thank the referee for the many excellent suggestions.

References

[1] R. C. Baker and G. Harman, *Shifted Primes without Large Prime Factors* Acta Arith. **83** (1998), 331-361.

[2] A. Balog and A. Sárközy, *On sums of Integers Having Small Prime Factors. I, II*, Studia Sci. Math. Hungar. **19** (1984), 35-47.

[3] ——————————-, *On Sums of Sequences of Integers. I*, Acta Arith. **44** (1984), 73-86.

[4] ——————————-, *On Sums of Sequences of Integers. II*, Acta Math. Hungar. **44** (1984), 169-179.

[5] ——————————-, *On Sums of Sequences of Integers. III*, Acta Math. Hungar. **44** (1984), 339-349.

[6] R. de la Bretèche, *Sommes sans Grand Facteur Premier [Sums without Large Prime Factors]* Acta Arith. **88** (1999), 1-14.

[7] Dickman, *On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude* Ark. Mat. Astr. Fys **22** (1930), 1-14.

[8] P. Erdős, H. Maier, and A. Sárközy, *On the Distribution of the Number of Prime Factors of Sums* $a + b$ Trans. Amer. Math. Soc. **302** (1987), 269-280.

[9] J. B. Friedlander, *Shifted Primes without Large Prime Factors*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 265, Kluwer Acad. Publ., Dordrecht, 1989.

[10] C. Pomerance, A. Sárközy, and C. L. Stewart, *On Divisors of Sums of Integers. III.* Pacific J. Math. **133** (1988), 363-379.

[11] A. Sárközy and C. L. Stewart, *On Divisors of Sums of Integers. I* Acta Math. Hungar. **48** (1986), 147-154.

[12] ————————————, *On Divisors of Sums of Integers. II.*, J. Reine Angew. Math. **365** (1986), 171-191.

[13] A. Sárközy and C. L. Stewart, *On Divisors of Sums of Integers. IV.* Canad. J. Math. **40** (1988), 788-816.

[14] ————————————, *On Divisors of Sums of Integers. V.* Pacific J. Math. **166** (1994), 373-384.