# WHEN THE SIEVE WORKS

ANDREW GRANVILLE, DIMITRIS KOUKOULOPOULOS AND KAISA MATOMÄKI

May 1, 2012

ABSTRACT. We are interested in classifying those sets of primes $\mathcal{P}$ such that when we sieve out the integers up to $x$ by the primes in $\mathcal{P}^c$ we are left with *roughly* the expected number of unsieved integers. In particular, we obtain the first general results for sieving an interval of length $x$ with primes including some in $(\sqrt{x}, x]$, using methods motivated by additive combinatorics.

## 1. INTRODUCTION AND MOTIVATION

Let $\mathcal{E}$ be a given subset of the primes $\leq x$. The simplest sieve problem asks for estimates of $\mathcal{S}(T, T+x; \mathcal{E})$, the number of integers $n$ in an interval $(T, T+x]$ which have no prime factors in the set $\mathcal{E}$ (we write $(n, \mathcal{E}) = 1$ for convenience). By a simple inclusion-exclusion argument one expects that the number of such integers is about

$$x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right).$$

This is provably always an upper bound, up to a constant:

$$\mathcal{S}(T, T+x; \mathcal{E}) \ll x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

and one gets the analogous lower bound

$$\mathcal{S}(T, T+x; \mathcal{E}) \gg x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

if $\mathcal{E}$ is a subset of the primes up to $x^{1/2 - o(1)}$ (see [3]). There seems to be little hope, in this generality, of increasing the exponent "1/2" without major new ideas. Moreover, one

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

1

needs to have careful hypotheses: if, for instance, $\mathcal{E} = \{p \leq \sqrt{x}\} \cup \{x/\log x < p \leq x\}$ then $\mathcal{S}(0, x; \mathcal{E}) = \pi(x/\log x) - \pi(\sqrt{x}) + 1 \sim x/\log^2 x$, whereas our prediction was $\asymp x/\log x$.

In this article we will prove lower bounds on $\mathcal{S}(0, x; \mathcal{E})$ in certain cases not covered by classical sieve theory, and use this evidence to guess at lower bounds in more generality. Let us first introduce some notation. Let $\mathcal{P}$ be a given subset of the primes $\leq x$, and $\mathcal{E} = \{p \leq x : p \notin \mathcal{P}\}$, so that $\mathcal{E} \cup \mathcal{P}$ is a partition of the primes $\leq x$. Let $\Psi(x; \mathcal{P})$ denote the number of integers up to $x$ all of whose prime factors are in $\mathcal{P}$, that is

$$\Psi(x; \mathcal{P}) := \sum_{\substack{n \leq x \\ p|n \implies p \in \mathcal{P}}} 1 = \sum_{\substack{n \leq x \\ (n, \mathcal{E}) = 1}} 1 = \mathcal{S}(0, x; \mathcal{E}).$$

The inclusion-exclusion argument predicts that

$$\Psi(x; \mathcal{P}) \approx x/u_{\mathcal{P}}, \quad \text{where} \quad u_{\mathcal{P}} := \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right)^{-1}.$$

Hall [5] proved that $\Psi(x; \mathcal{P}) \lesssim (e^{\gamma}/u_{\mathcal{P}}) \, x$, where $\gamma = 0.5772156649\ldots$ is the Euler-Mascheroni constant. Subsequently, the authors of [4] improved this to $\Psi(x; \mathcal{P}) \lesssim (e^{\gamma}/u_{\mathcal{P}} - 1/u_{\mathcal{P}}^{2+o(1)}) \, x$, and showed that this is "best possible" other than being more precise than the "$o(1)$" (which tends to 0 when $u_{\mathcal{P}} \to \infty$). However, in this paper, we are interested in lower bounds on $\Psi(x; \mathcal{P})$.

Hildebrand [6] showed that, for a given value of $u_{\mathcal{P}}$, $\Psi(x; \mathcal{P})$ is asymptotically smallest when $\mathcal{P}$ is the set of primes up to $x^{1/u_{\mathcal{P}}}$. Specifically, he showed that

$$\frac{\Psi(x; \mathcal{P})}{x} \gtrsim \rho(u_{\mathcal{P}});$$

here $\rho(u)$ is the Dickman–de Bruijn function, defined by $\rho(u) = 1$ for $0 \leq u \leq 1$, and $u\rho'(u) = -\rho(u-1)$ for all $u \geq 1$. One can show that $\rho(u) = 1/u^{u+o(u)}$, which is a lot smaller than the expected $1/u$. (See [4] for a different proof of Hildebrand's result.)

This last example is very special in that $\mathcal{P}$ contains no large primes, and it is expected that other than for certain other extraordinary sets $\mathcal{P}$ one has

$$\Psi(x; \mathcal{P}) \asymp x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right).$$

However this question has not really been studied in detail and there are other examples that must be taken into account: Let

$$\mathcal{P} = \bigcup_{1 \leq m \leq N-1} \{p \text{ prime} : x^{m/(N+1)} < p < x^{m/N}\}.$$

Any product of primes in these intervals lies in some interval of the same form. If now $n = p_1 p_2 \cdots p_k \leq x$, where $x^{m_j/(N+1)} < p_j < x^{m_j/N}$, then $m_1 + \cdots + m_k \leq N$. The number of such integers $n$ with $m_1 + \cdots + m_k \leq N - 1$ is $\leq x^{1-1/N}$. On the other hand, if $m_1 + \cdots + m_k = N$, then $k \geq 2$ (since each $m_j \leq N - 1$) and therefore $\Psi(x; \mathcal{P}) \ll_N x/(\log x)^2$, far smaller than the expected $x/u_{\mathcal{P}} \asymp x/\log x$. A key thing to notice is that, in this example,

$$\sum_{p \in \mathcal{P}} \frac{1}{p} \sim (N-1) \log(1 + 1/N) = 1 - \frac{3/2 + o(1)}{N} < 1,$$

by the prime number theorem, as $N \to \infty$. Hence we see that we can obtain a very small $\Psi(x; \mathcal{P})$ when $\sum_{p \in \mathcal{P}} \frac{1}{p} < 1$. One might guess that the issue in the last example is that there are no small primes in $\mathcal{P}$. However, if we let

$$\mathcal{P} = \{p \text{ prime}: \ p \leq x^{1/N^2}\} \cup \bigcup_{1 \leq m \leq N-1} \{p \text{ prime}: \ x^{m/(N+1)} < p < x^{m/N}\},$$

then we would also find that $\Psi(x; \mathcal{P})$ is far smaller than expected if $N \to \infty$ slowly enough.

One might also guess that the main issue in the above examples is that there are no large primes in $\mathcal{P}$. However if we let

$$\mathcal{P} = \{p \text{ prime}: \ x^{1/v} < p \leq x^{1/u}\} \cup \{p \text{ prime}: \ x^{1-1/v} < p \leq x\},$$

then $\Psi(x; \mathcal{P}) = \Psi(x; \mathcal{Q}) + \pi(x) - \pi(x^{1-1/v})$ where $\mathcal{Q} = \{p: \ x^{1/v} < p \leq x^{1/u}\}$. Friedlander [2] established that $\Psi(x; \mathcal{Q}) \sim \sigma(u, v) x/\log(x^{1/v})$ where $\sigma(u, v) = e^{-\gamma} \rho(u) + O(1/v \log v)$. Hence if $v > u^{(1+\epsilon)u}$ then $\Psi(x; \mathcal{P}) \sim e^{-\gamma} v \rho(u) x/\log x$ as $u \to \infty$, whereas the "expected" value is $x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \sim (e^{-\gamma} v^2/u(v-1)) x/\log x$. Hence the ratio $\Psi(x; \mathcal{P})/\text{"Expected value"} \sim u \rho(u)(1 - 1/v)$, which goes to 0 rapidly as $u \to \infty$.

So we see that the size of the primes in $\mathcal{P}$ does not seem to determine whether $\Psi(x; \mathcal{P})$ is close to its "Expected value". Rather, we believe that the most important quantity in determining whether the sieve will work somewhat as expected is the largest $y$ for which $\sum_{p \in \mathcal{P}, \ p \geq y} \frac{1}{p} > 1$.

**Conjecture 1.** *Fix $\epsilon > 0$. There exists a positive constant $c$ such that if $\mathcal{P}$ is a subset of the primes $\leq x$ for which there is some $v \leq c\sqrt{\log x}$ with*

$$\sum_{\substack{p \in \mathcal{P} \\ x^{1/ev} < p \leq x}} \frac{1}{p} \geq 1 + \epsilon,$$

*then*

$$\frac{\Psi(x; \mathcal{P})}{x} \geq \kappa_v \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

*where $\kappa_v$ is a constant with $\kappa_v = v^{-v(1+o_\epsilon(1))}$ as $v \to \infty$.*

*Remark 1.1.* Note that if $\mathcal{P} = \{p \le x^{1/u}\}$, then $\Psi(x; \mathcal{P})/x \sim \rho(u)$. Also, taking $v \sim u$ in the above conjecture yields $\Psi(x; \mathcal{P})/x \gtrsim \kappa_v/u$ and so $\kappa_v \lesssim u\rho(u) = v^{-v(1+o(1))}$. One can even make the bolder guess that $\kappa_v \sim v\rho(v)$.

*Remark 1.2.* Proposition 2.3 below implies that if $\sum_{p\in\mathcal{P}, \ x^{1/u}<p\le x} \frac{1}{p}$ is very small then $\Psi(t; \mathcal{P})$ is indeed substantially smaller than expected for some $t \in [\sqrt{x}, x]$. Hence it is certainly true that such quantities are an important consideration. It would be interesting to obtain the strongest possible uniform version of Proposition 2.3.

The main result in this paper is a weak form of Conjecture 1:

**Theorem 1.** *There exist positive constants $\lambda$ and $c$ such that if $\mathcal{P}$ is a subset of the primes $\le x$ for which there is some $v \le c\sqrt{\log x}$ with*

$$\sum_{\substack{p\in\mathcal{P} \\ x^{1/ev}<p\le x}} \frac{1}{p} \ge 1 + \lambda,$$

*then*

$$\frac{\Psi(x; \mathcal{P})}{x} \gg \frac{1}{v^{O(v)}} \prod_{p\in\mathcal{E}} \left(1 - \frac{1}{p}\right).$$

*Remark 1.3.* One can superficially make Theorem 1 appear stronger. For, if the "$O(v)$" in our lower bound is short for "$Cv$", then, for any $\epsilon \in (0, C)$, we can replace $\lambda$ in the hypothesis by $\Lambda := \lambda + \ln(C/\epsilon)$ so that

$$\sum_{\substack{p\in\mathcal{P} \\ x^{C/\epsilon ev}<p\le x}} \frac{1}{p} \ge \sum_{\substack{p\in\mathcal{P} \\ x^{1/ev}<p\le x}} \frac{1}{p} - \sum_{\substack{p\in\mathcal{P} \\ x^{1/ev}<p\le x^{C/\epsilon ev}}} \frac{1}{p} \ge 1 + \Lambda - \ln(C/\epsilon) = 1 + \lambda,$$

and hence Theorem 1 implies, for $V = \epsilon v/C$, that

$$\frac{\Psi(x; \mathcal{P})}{x} \gg \frac{1}{V^{CV}} \prod_{p\in\mathcal{E}} \left(1 - \frac{1}{p}\right) \ge \frac{1}{v^{\epsilon v}} \prod_{p\in\mathcal{E}} \left(1 - \frac{1}{p}\right).$$

*Remark 1.4.* It will be clear from the proof of Theorem 1 that, under the same assumptions, one gets also the stronger conclusion

$$\frac{\Psi(x; \mathcal{P}) - \Psi(x - y; \mathcal{P})}{y} \gg \frac{1}{v^{O(v)}} \prod_{p\in\mathcal{E}} \left(1 - \frac{1}{p}\right)$$

for every $y > x^{1-1/(3ev)}$. Furthermore the conclusion of Theorem 1 continues to hold if, more generally,

$$\sum_{\substack{p \in \mathcal{P} \\ x^{1/ev} < p \le x^{1/u}}} \frac{1}{p} \ge \frac{1+\lambda}{u}$$

for some $1 \le u \le v \le c\sqrt{\log x}$, where $\lambda$ and $c$ are absolute constants.

The proof of our theorem involves a hodge-podge of techniques, from classical analytic number theory and sieve theory (as one might expect) to additive combinatorics, which seems to be new in this context (though [8] contains some related work as will be explained in Remark 4.1). Our starting point is the following result of Bleichenbacher [1] which may be viewed as a result in continuous additive combinatorics:

**Bleichenbacher's Theorem.** *If $u \ge 1$ and $T$ is an open subset of $(0, \frac{1}{u})$ for which*

$$\int_{t \in T} \frac{dt}{t} > \frac{1}{u} ,$$

*then there exist $t_1, t_2, \ldots, t_k \in T$ for which $t_1 + t_2 + \cdots + t_k = 1$.*

Note that this is "best possible" since for the set $T_N = \bigcup_{j=1}^{N}(\frac{j}{N+1}, \frac{j}{N})$ there is no solution to $t_1 + t_2 + \cdots + t_k = 1$ (as any sum of elements in intervals of the form $(\frac{j}{N+1}, \frac{j}{N})$ is also in an interval of this form), yet $\int_{T_N \cap (0,1/u)} \frac{dt}{t} \ge [N/u]\log(1 + 1/N) > 1/u - 2/N$, which tends to $1/u$ from below as $N \to \infty$. One sees an analogy between this example and the first sieve example that we gave above.

The key idea in this paper is to reduce estimates for $\Psi(x; \mathcal{P})$ to quantitative questions of the type addressed in Bleichenbacher's theorem. Accordingly we make the following conjecture:

**Hypothesis T.** *There exists a constant $\lambda_3 > 0$ such that if $1 \le u \le v$ and $T$ is an open subset of $(\frac{1}{ev}, \frac{1}{u})$ for which*

$$\int_{t \in T} \frac{dt}{t} \ge \frac{1+\lambda_3}{u} ,$$

*then there exists an integer $k \in [u, ev]$ and an absolute constant $\tau_v > 0$ such that*

$$\int \cdots \int_{\substack{t_1+t_2+\cdots+t_k=1 \\ t_1,t_2,\ldots,t_k \in T}} \frac{dt_1 dt_2 \cdots dt_{k-1}}{t_1 t_2 \cdots t_k} \ge \tau_v \left( \int_{t \in T} \frac{dt}{t} \right)^k .$$

By analogy with Bleichenbacher's Theorem, we conjecture that Hypothesis T holds for any choice of $\lambda_3 > 0$. The importance of Hypothesis T can be seen in the following consequence, which will be proven in Section 5.

**Proposition 1.** *If Hypothesis T is true for any fixed $\lambda_3 > 0$ and $\tau_v = v^{-O(v)}$, then Conjecture 1 holds for any fixed $\epsilon > 0$ with $\kappa_v = v^{-O(v)}$.*

We will prove a version of Hypothesis T for some sufficiently large $\lambda_3 > 0$ (see Theorem 7.1) that allows us to deduce Theorem 1 as desired. The value of $\lambda_3$ can be determined explicitly from the proof, but it will certainly not yield Hypothesis T for *every* $\lambda_3 > 0$. See also Remark 7.1 for more discussion about attainable $\lambda$-values.

The layout of this paper is as follows: In Section 2 we explore when there are far fewer integers left unsieved than expected. In Section 4, we make some technical reductions to our original problem, and then in Section 5 we develop equivalent questions in continuous and discrete combinatorics. In Section 6 we develop some tools of discrete additive combinatorics. Finally, in Section 7 we use these tools to establish a discrete combinatorial version of our problem which implies Theorem 1 (as shown in Section 5).

Our results leave us wondering whether Conjecture 1 and Theorem 1 might be an indication of the truth for the more general problem of sieving intervals. Could it be that when we sieve an arbitrary interval of length $x$, with a not-too-large subset of the primes up to $x$, then the number of integers left unsieved is predictable? There are only two types of examples known where we can do accurate calculations to better understand sieving: random intervals and intervals where most of the progressions sieved are $0 \pmod p$, and as far as we know the latter are where most extreme examples come from. Since we have now obtained some understanding of this usual source of extreme examples, we can speculate that this sort of criteria is the main issue, in general.

## 2. Sieving with logarithmic weights

If we introduce the weight $1/n$ at each integer $n$ (the so-called "logarithmic weights"), then we simplify the problem enormously:

**Lemma 2.1.** *If $\mathcal{P}$ is a subset of the primes $\leq x$ and $\mathcal{E} = \{p \leq x\} \setminus \mathcal{P}$, then*

$$\prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \lesssim \frac{1}{\log x} \sum_{\substack{n \leq x \\ p|n \implies p \in \mathcal{P}}} \frac{1}{n} \lesssim e^{\gamma} \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

*where $\gamma$ is the Euler-Mascheroni constant.*

*Proof.* Let us first prove the lower bound. We have

$$\prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right)^{-1} \geq \sum_{\substack{\ell \leq x \\ p|\ell \implies p \in \mathcal{E}}} \frac{1}{\ell},$$

so that

$$\sum_{\substack{m \leq x \\ p|m \implies p \in \mathcal{P}}} \frac{1}{m} \geq \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \sum_{\substack{\ell \leq x \\ p|\ell \implies p \in \mathcal{E}}} \frac{1}{\ell} \sum_{\substack{m \leq x \\ p|m \implies p \in \mathcal{P}}} \frac{1}{m} \geq \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \sum_{n \leq x} \frac{1}{n},$$

since every integer $n \leq x$ may be written as $\ell m$.

On the other hand we have the upper bound

$$\sum_{\substack{m \leq x \\ p|n \implies p \in \mathcal{P}}} \frac{1}{n} \leq \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \sim e^{\gamma} \log x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

by Mertens' theorem. $\qquad\qquad\square$

*Remark 2.1.* Note that

$$(2.1) \qquad\qquad \sum_{\substack{n \leq x \\ p|n \implies p \in \mathcal{P}}} \frac{1}{n} = \frac{\Psi(x; \mathcal{P})}{x} + \int_1^x \frac{\Psi(t; \mathcal{P})}{t^2} dt,$$

so that Lemma 2.1 can be re-phrased as a weighted mean of $\Psi(t; \mathcal{P})$-values:

$$\prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \lesssim \int_1^x \frac{\Psi(t; \mathcal{P})}{t} \cdot \frac{dt}{t} \bigg/ \int_1^x \frac{dt}{t} \lesssim e^{\gamma} \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

since $\Psi(x; \mathcal{P})/x \ll \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right)$ by classical sieve theory (as discussed in the introduction).

We can use Lemma 2.1 to prove a first lower bound in the direction of Theorem 1, though with a different emphasis.

**Corollary 2.2.** *Fix $\epsilon \in (0, 1)$. Suppose that $\mathcal{P}$ is a subset of the primes $\leq x$, and $u \in [1, \log x]$ is such that*

$$\sum_{\substack{p \in \mathcal{P} \\ x^{1/u} < p \leq x}} \frac{1}{p} > \epsilon.$$

*Then there exists $t \in [x^{1/u}, x]$ such that*

$$(2.2) \qquad\qquad \frac{\Psi(t; \mathcal{P})}{t} \gg \frac{\epsilon \min\{1, \epsilon u\}}{\log u} \prod_{\substack{p \in \mathcal{E} \\ p \leq t}} \left(1 - \frac{1}{p}\right).$$

*Proof.* By (2.1)

$$\int_{x^{1/u}}^x \frac{\Psi(t; \mathcal{P})}{t^2} dt \geq \sum_{\substack{x^{1/u} \leq n \leq x \\ p|n \implies p \in \mathcal{P}}} \left(\frac{1}{n} - \frac{1}{x}\right) \geq \frac{1}{2} \sum_{\substack{x^{1/u} \leq n \leq x/2 \\ p|n \implies p \in \mathcal{P}}} \frac{1}{n}.$$

Writing here $n = ab$, where prime factors of $a$ are $\leq x^{1/u}$ and prime factors of $b$ are $> x^{1/u}$, and discarding some $n$, we see that

$$\int_{x^{1/u}}^{x} \frac{\Psi(t; \mathcal{P})}{t^2} dt \geq \frac{1}{2} \sum_{\substack{a \leq x^{\epsilon/2}/2 \\ p|a \implies p \in \mathcal{P} \cap [1, x^{1/u}]}} \frac{1}{a} \sum_{\substack{1 < b \leq x^{1-\epsilon/2} \\ p|b \implies p \in \mathcal{P} \cap (x^{1/u}, x^{1-\epsilon/2}]}} \frac{1}{b}.$$

Here

$$S := \sum_{\substack{1 < b \leq x^{1-\epsilon/2} \\ p|b \implies p \in \mathcal{P} \cap (x^{1/u}, x^{1-\epsilon/2}]}} \frac{1}{b} \geq \sum_{\substack{p \in \mathcal{P} \\ x^{1/u} < p \leq x^{1-\epsilon/2}}} \frac{1}{p} > \epsilon + \log(1 - \epsilon/2) + o(1) \geq \frac{\epsilon}{4},$$

hence $S \geq \frac{\epsilon}{8}(1 + S)$ and so

$$\int_{x^{1/u}}^{x} \frac{\Psi(t; \mathcal{P})}{t^2} dt \geq \frac{\epsilon}{8} \sum_{\substack{a \leq x^{\epsilon/2}/2 \\ p|a \implies p \in \mathcal{P} \cap [1, x^{1/u}]}} \frac{1}{a} \sum_{\substack{b \leq x^{1-\epsilon/2} \\ p|b \implies p \in \mathcal{P} \cap (x^{1/u}, x^{1-\epsilon/2}]}} \frac{1}{b}.$$

Consequently, Lemma 2.1 implies that

$$\int_{x^{1/u}}^{x} \frac{\Psi(t; \mathcal{P})}{t^2} dt \gg \epsilon \prod_{\substack{p \leq x^{\epsilon/2} \\ p \in \mathcal{P} \cap [1, x^{1/u}]}} \left(1 + \frac{1}{p}\right) \prod_{p \in \mathcal{P} \cap (x^{1/u}, x^{1-\epsilon/2}]} \left(1 + \frac{1}{p}\right)$$

$$\gg \epsilon \min\{1, \epsilon u\} \prod_{p \in \mathcal{P}} \left(1 + \frac{1}{p}\right) \asymp \epsilon \min\{1, \epsilon u\} \log x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right).$$

If now for every $t \in [x^{1/u}, x]$ we have that

$$\frac{\Psi(t; \mathcal{P})}{t} \leq \eta \cdot \frac{\epsilon \min\{1, \epsilon u\}}{\log u} \prod_{\substack{p \in \mathcal{E} \\ p \leq t}} \left(1 - \frac{1}{p}\right) \ll \eta \cdot \frac{\epsilon \min\{1, \epsilon u\}}{\log u} \frac{\log x}{\log t} \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right),$$

then

$$\int_{x^{1/u}}^{x} \frac{\Psi(t; \mathcal{P})}{t^2} \ll \eta \cdot \frac{\epsilon \min\{1, \epsilon u\}}{\log u} \log x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right) \int_{x^{1/u}}^{x} \frac{dt}{t \log t}$$

$$\ll \eta \cdot \epsilon \min\{1, \epsilon u\} \log x \prod_{p \in \mathcal{E}} \left(1 - \frac{1}{p}\right).$$

Choosing $\eta$ small enough, we arrive at a contradiction. So the claimed result follows.   □

*Remark 2.2.* Let $T$ be the set of $t \in [x^{1/u}, x]$ for which (2.2) holds. By the same proof, and the usual sieve upper bound, we obtain $\int_{t \in T} dt/(t \log t) \gg \epsilon \min\{1, \epsilon u\}$

The lower bound (2.2) obtained here is much better than the lower bound in Theorem 1, but it only works for some values of $t$. One cannot essentially improve Corollary 2.2 in general, at least when $1/\epsilon \ll u \ll \sqrt{\log x}$: Take a prime $q \asymp (\log u)/\epsilon$ and let $\mathcal{P}$ be the set of primes which are $\equiv 1 \pmod q$. Then the classical sieve yields

$$\frac{\Psi(t, \mathcal{P})}{t} = \frac{1}{t} |\{n = qk + 1 \le t : p \mid n \implies p \notin \mathcal{E} \setminus \{q\}\}|$$

$$\ll \frac{1}{q} \prod_{\substack{p \in \mathcal{E} \setminus \{q\} \\ p \le (t/q)^{1/2}}} \left(1 - \frac{1}{p}\right) \asymp \frac{\epsilon}{\log u} \prod_{\substack{p \in \mathcal{E} \\ p \le t}} \left(1 - \frac{1}{p}\right)$$

for every $t \in [x^{1/u}, x]$.

**Proposition 2.3.** *Suppose that $\mathcal{P}$ is a subset of the primes $\le x$, and $u \in [1, \log x]$ is such that*

$$\sum_{\substack{p \in \mathcal{P} \\ x^{1/u} < p \le x}} \frac{1}{p} := \kappa \ll 1.$$

*There exists $t \in [x^{1/2}, x]$ such that*

$$\frac{\Psi(t; \mathcal{P})}{t} \ll \left(\kappa + x^{-1/6} + u^{-u/6}\right) \prod_{\substack{p \in \mathcal{E} \\ p \le t}} \left(1 - \frac{1}{p}\right).$$

*In particular, if $\kappa = o(1)$ as $u \to \infty$ and $x \to \infty$, then this upper bound is much smaller than what is expected.*

*Proof.* We may assume that $u$ is large enough and $u = o(\log x)$. Define $V$ by $e^V/V = u$, so that $V = \log u + \log \log u + O(1)$. Set $y = x^{1/u}$ and $z = y^{1/v}$ with $v = \min\{\log y, V\}$. Notice that

$$\sum_{\substack{b \ge 1 \\ p \mid b \implies p \in \mathcal{P} \cap (y, x]}} \frac{1}{b} \le \prod_{\substack{y < p \le x \\ p \in \mathcal{P}}} \left(1 - \frac{1}{p}\right)^{-1} \le e^{O(\kappa)} \ll 1.$$

We wish to estimate the sum of $1/n$, summed over all $n$ in the range $\sqrt{x} < n \le x$ for which $p \mid n \implies p \in \mathcal{P}$. Writing $n = ab$ with $a$ having prime factors $\le y$ and $b$ having prime

factors $> y$, and adding some extra $ab$, we see that

$$
\sum_{\substack{\sqrt{x}<n\leq x \\ p|n \implies p\in\mathcal{P}}} \frac{1}{n} \leq \sum_{\substack{a\leq\sqrt{x} \\ p|a \implies p\in\mathcal{P}\cap[1,y]}} \frac{1}{a} \cdot \sum_{\substack{b>1 \\ p|b \implies p\in\mathcal{P}\cap(y,x]}} \frac{1}{b} + \sum_{\substack{\sqrt{x}<a\leq x \\ p|a \implies p\in\mathcal{P}\cap[1,y]}} \frac{1}{a} \sum_{\substack{b\geq 1 \\ p|b \implies p\in\mathcal{P}\cap(y,x]}} \frac{1}{b}
$$

$$
\leq \sum_{\substack{a\geq 1 \\ p|a \implies p\in\mathcal{P}\cap[1,y]}} \frac{1}{a} \cdot (e^{O(\kappa)}-1) + \sum_{\substack{\sqrt{x}<a\leq x \\ p|a \implies p\in\overline{\mathcal{P}}\cap[1,y]}} \frac{1}{a} \cdot e^{O(\kappa)}
$$

$$
\ll \kappa \prod_{\substack{p\leq y \\ p\in\mathcal{P}}} \left(1-\frac{1}{p}\right)^{-1} + \sum_{\substack{\sqrt{x}<a\leq x \\ p|a \implies p\in\overline{\mathcal{P}}\cap[1,y]}} \frac{1}{a}\left(\frac{a}{\sqrt{x}}\right)^{\frac{1}{\log z}}
$$

$$
\ll \prod_{\substack{p\leq y \\ p\in\mathcal{P}}} \left(1-\frac{1}{p}\right)^{-1} \left( \kappa + e^{-uv/2} \prod_{\substack{p\leq y \\ p\in\mathcal{P}}} \left(1-\frac{1}{p}\right) \left(1-\frac{1}{p^{1-\frac{1}{\log z}}}\right)^{-1} \right).
$$

The logarithm of this last Euler product is equal to

$$
\sum_{\substack{p\leq y \\ p\in\mathcal{P}}} \left\{ \log\left(1-\frac{1}{p}\right) - \log\left(1-\frac{1}{p^{1-\frac{1}{\log z}}}\right) \right\} \ll \sum_{\substack{p\leq y \\ p\in\mathcal{P}}} \frac{p^{\frac{1}{\log z}}-1}{p}
$$

$$
\ll \frac{e^v}{\log y} + \int_2^y \frac{\alpha^{\frac{1}{\log z}}-1}{\alpha\log\alpha}d\alpha,
$$

by the Brun-Titchmarsch inequality. Here $e^v/\log y = e^v/(v\log z) \leq u/\log z$ and

$$
\int_2^y \frac{\alpha^{\frac{1}{\log z}}-1}{\alpha\log\alpha}d\alpha \ll \int_2^z \frac{1}{\alpha\log z}d\alpha + \int_z^y \frac{\alpha^{\frac{1}{\log z}}-1}{\log\alpha}d\alpha \leq 1 + \int_1^V \frac{e^w}{w}dw \ll \frac{e^V}{V} = u,
$$

by setting $\alpha = z^w$ in the second integral and noticing that $v\leq V$. Therefore

$$
\sum_{\substack{\sqrt{x}<n\leq x \\ p|n \implies p\in\mathcal{P}}} \frac{1}{n} \ll \prod_{\substack{p\leq x \\ p\in\mathcal{P}}} \left(1-\frac{1}{p}\right)^{-1} \left(\kappa + e^{-uv/2+O(u)}\right) \ll \prod_{\substack{p\leq x \\ p\in\mathcal{E}}} \left(1-\frac{1}{p}\right) \left(\kappa + e^{-uv/3}\right)\log x
$$

$$
\ll \prod_{\substack{p\leq x \\ p\in\mathcal{E}}} \left(1-\frac{1}{p}\right) \left(\kappa + x^{-1/3} + u^{-u/3}\right)\log x,
$$

provided that $u$ is large enough, since $e^{-uv} \leq y^{-u} + e^{-uV} = x^{-1} + e^{-O(u)}(u\log u)^{-u}$. We deduce that

$$
\Psi(x;\mathcal{P}) \leq \sqrt{x} + \sum_{\substack{\sqrt{x}<n\leq x \\ p|n \implies p\in\mathcal{P}}} \frac{x}{n} \ll x\prod_{\substack{p\leq x \\ p\in\mathcal{E}}} \left(1-\frac{1}{p}\right) \left(\kappa + x^{-1/3} + u^{-u/3}\right)\log x.
$$

Integrating by parts and then applying this with $x$ replaced by $\sqrt{x}$, we have

$$\int_{\sqrt{x}}^{x} \frac{\Psi(t;\mathcal{P})}{t}\frac{dt}{t} + \frac{\Psi(x;\mathcal{P})}{x} = \sum_{\substack{\sqrt{x} < n \leq x \\ p|n \implies p \in \mathcal{P}}} \frac{1}{n} + \frac{\Psi(\sqrt{x};\mathcal{P})}{\sqrt{x}}$$

$$\ll \prod_{\substack{p \leq x \\ p \in \mathcal{E}}} \left(1 - \frac{1}{p}\right)\left(\kappa + x^{-1/6} + u^{-u/6}\right)\log x.$$

The result follows, by taking the minimum of $\Psi(t;\mathcal{P})$ on the left side. $\qquad\square$

This last estimate is much smaller than one might have guessed given Lemma 2.1.

We have now seen that if there are very few large primes in $\mathcal{P}$ then one can improve the sieve upper bounds for some values of $t$. However, the conclusion of Proposition 2.3 does not hold, in general, for a given value of $t$. To see this when $t = x$, take $\mathcal{P} = \{p \leq x^\epsilon\} \cup \{p: \ x^{1-\epsilon} < p \leq x\}$. Here $\Psi(x,\mathcal{P})$ contains all integers of the form $mp$ where $p$ is a prime in the range $x^{1-\epsilon} < p \leq x$, and $m \leq x/p$. Hence

$$\Psi(x;\mathcal{P}) \geq \sum_{x^{1-\epsilon} < p \leq x} \frac{x}{p} \gg \epsilon x \asymp x \prod_{\substack{p \in \mathcal{E} \\ p \leq x}} \left(1 - \frac{1}{p}\right).$$

## 3. Some combinatorial lemmas

We devote this section to proving some combinatorial lemmas we will need later. First, we note without proof the following lemma, leaving it as an amusing exercise for the reader.

**Lemma 3.0.** *Let $B$ be a finite subset of the numbers in $(y, z]$. For any $x \geq z$ there exists a positive integer $k$ such that the number of $k$-tuples $b_1, \ldots, b_k \in B$ for which $b_1 + \cdots + b_k \in (x - z, x]$ is $\geq |B|^k/(x/y)$.*

Lemma 3.0 also follows from the following weighted version.

**Lemma 3.1.** *Let $B$ be a finite subset of the numbers in $(y, z]$ and associate to each $b \in B$ a positive weight $w(b)$. For any $x \geq z$ there exists a positive integer $k \leq x/y$ such that*

$$\sum_{\substack{b_1,\ldots,b_k \in B \\ b_1 + \cdots + b_k \in (x-z,x]}} w(b_1)w(b_2)\cdots w(b_k) \geq \frac{1}{x/y}\left(\sum_{b \in B} w(b)\right)^k.$$

*Proof.* Note that if $b_1 + \cdots + b_k \in (x - z, x]$ for some $b_1, \ldots, b_k \in B$, then $ky < b_1 + \cdots + b_k \leq$

$x$ and so $1 \leq k < x/y$. Therefore if we let $K = [x/y]$, then

$$
\begin{aligned}
\beta_k &:= \sum_{\substack{b_1,\ldots,b_k \in B \\ b_1+\cdots+b_k \in (x-z,x]}} w(b_1)w(b_2)\cdots w(b_k) \Bigg/ \left(\sum_{b \in B} w(b)\right)^k \\
&= \sum_{\substack{b_1,\ldots,b_K \in B \\ b_1+\cdots+b_k \in (x-z,x]}} w(b_1)w(b_2)\cdots w(b_K) \Bigg/ \left(\sum_{b \in B} w(b)\right)^K .
\end{aligned}
$$

Consequently, we find that

$$
\sum_{k=1}^{K} \beta_k = \left(\sum_{b \in B} w(b)\right)^{-K} \sum_{b_1,\ldots,b_K \in B} w(b_1)w(b_2)\cdots w(b_K) \sum_{\substack{1 \leq k \leq K \\ b_1+\cdots+b_k \in (x-z,x]}} 1 \geq 1,
$$

since the differences in each sequence $b_1, b_1+b_2, \ldots, b_1+\cdots+b_K$ are $\leq z$, whereas $b_1 \leq z \leq x$ and $b_1 + \cdots + b_K > yK > x - y > x - z$. Taking the maximum of the $\beta_k$ then yields the desired result. $\qquad\square$

**Corollary 3.2.** *Let $\mathcal{P}$ be a subset of the primes in $(x^{1/ev}, x^{1/u}]$ for some $1 \leq u \leq v \leq (\log x)/e$. For any $X \geq x^{1/u}$ there exists a positive integer $\ell \leq K := \frac{ev \log X}{\log x}$ such that*

$$
\sum_{\substack{q_1,\ldots,q_\ell \in \mathcal{P} \\ Xx^{-1/u} < q_1 q_2 \cdots q_\ell \leq X}} \frac{1}{q_1 q_2 \cdots q_\ell} \geq \frac{1}{K} \left(\sum_{q \in \mathcal{P}} \frac{1}{q}\right)^\ell .
$$

*Proof.* Apply Lemma 3.1 with $B = \{\log q : q \in \mathcal{P}\} \subset ((1/ev)\log x, (1/u)\log x]$ and $w(b) = e^{-b} = 1/q$, and then take $\ell = k$ as obtained in that lemma. $\qquad\square$

## 4. TECHNICAL REDUCTIONS

The hypothesis in Theorem 1 relies on there being a reasonable density of "large" primes in $\mathcal{P}$. More generally, we may ask what happens when some interval $(x^{1/ev}, x^{1/u}]$ contains lots of primes of $\mathcal{P}$. Reducing to the analogous problem where $\mathcal{P}$ is now restricted to be a subset of the primes in this interval, we formulate

**Hypothesis P.** *There exist constants $\lambda_1 > 0$ and $C_1 > 1$ such that if $v^2 \leq \lambda_1 \log x/C_1$, $1 \leq u \leq v$ and $\mathcal{P}$ is a subset of the primes in $(x^{1/ev}, x^{1/u}]$ for which*

$$
(4.1) \qquad\qquad \sum_{p \in \mathcal{P}} \frac{1}{p} \geq \frac{1+\lambda_1}{u} ,
$$

*then for any $\delta \in [x^{-1/(3ev)}, 1/2]$ there exists an integer $k \in [u, ev]$ and an absolute constant $\pi_v > 0$ such that*

$$\sum_{\substack{(p_1, \ldots, p_k) \in \mathcal{P}^k \\ (1-\delta)x < p_1 \cdots p_k < x}} \frac{1}{p_1 \cdots p_k} \geq \frac{\delta \pi_v}{\log x} \cdot \left( \sum_{p \in \mathcal{P}} \frac{1}{p} \right)^k .$$

*Remark 4.1.* In [8] Bleichenbacher's theorem is used to prove a result like Hypothesis P but with a logarithmic loss $(\log x)^{-O(v)}$ in the obtained lower bound — see Theorem 4 and Proposition 10.1 there.

*Proof that Hypothesis P with $\pi_v = v^{-O(v)}$ implies Theorem 1 with $\lambda = \lambda_1 + \epsilon$ and $c = \sqrt{\lambda_1/(2C_1)}$.* Set $\eta = \min\{\epsilon, 1\}/3$. Let $\mathcal{A} = \mathcal{P} \cap [1, x^{1/ev}]$ and $\mathcal{B} = \mathcal{P} \cap (x^{1/ev}, x]$ so that

$$\Psi(x; \mathcal{P}) \geq \sum_{\substack{a \leq x^\eta \\ p \mid a \implies p \in \mathcal{A}}} \Psi(x/a; \mathcal{B}),$$

since we can write any $n$ composed only of prime factors from $\mathcal{P}$ as $n = ab$ where $a$ and $b$ are composed only of prime factors from $\mathcal{A}$ and $\mathcal{B}$, respectively. For each $a \leq x^\eta$, we have that

$$\sum_{\substack{p \in \mathcal{B} \\ (x/a)^{1/ev} < p \leq x/a}} \frac{1}{p} \geq \sum_{\substack{p \in \mathcal{P} \\ x^{1/ev} < p \leq x}} \frac{1}{p} + \log(1 - \eta) + o(1) \geq 1 + \lambda_1 + \epsilon - \frac{\eta}{1 - \eta} + o(1) > 1 + \lambda_1,$$

as $x \to \infty$. Applying Hypothesis P with $u = 1$ and $\delta = 1/2$ to the set $\mathcal{B}$ yields

$$\Psi(x/a; \mathcal{B}) \geq \sum_{\substack{x/2a < n \leq x/a \\ p \mid n \implies p \in \mathcal{B}}} \frac{x/2a}{n} \gg \frac{1}{v^{O(v)}} \cdot \frac{x}{a \log x}$$

and consequently

$$\frac{\Psi(x; \mathcal{P})}{x} \gg \frac{1}{v^{O(v)}} \frac{1}{\log x} \sum_{\substack{a \leq x^\eta \\ p \mid a \implies p \in \mathcal{A}}} \frac{1}{a} \gg \frac{\eta}{v^{O(v)}} \prod_{\substack{p \leq x^\eta \\ p \in \mathcal{E}}} \left( 1 - \frac{1}{p} \right) \gg \frac{\eta}{v^{O(v)}} \prod_{p \in \mathcal{E}} \left( 1 - \frac{1}{p} \right),$$

by Lemma 2.1, which completes the proof of Theorem 1.     $\square$

So we have shown that in order to prove Theorem 1 we must prove the more convenient Hypothesis P with certain choices of the parameters therein.

## 5. EQUIVALENT PROBLEMS IN COMBINATORICS

In our sieve question we are seeking to sieve the integers up to $x$ by a given set of primes $\mathcal{E}$ which is, as discussed in the introduction, the same thing as counting the number of integers up to $x$ that are composed of primes from a given set $\mathcal{P}$. This makes this a rather special case of sieving an interval, since the problem can now be approached as a question of counting lattice points: If $p_1 \cdots p_k \leq x$, then

$$\log p_1 + \log p_2 + \cdots + \log p_k \leq \log x$$

and there are various techniques for attacking this problem. However they are not really effective, since here we have an enormous dimension compared to the volume of our region, even when restricting the primes in $\mathcal{P}$ to an interval $[x^{1/ev}, x^{1/u})$. We can however cut the dimension of the problem significantly by taking approximations that do not greatly effect the answer. For example, if we replace each $\log p$ by $[\log p]$ and take $N$ to be an integer close to $\log x$, then we can count integer solutions to $a_1 + a_2 + \cdots + a_m \leq N$, and weight each $a_i$ by the number of primes $p$ in $\mathcal{P}$ for which $[\log p] = a$. However even this problem is of rather high dimension to directly use lattice point counting results, so instead we attack this as a question in combinatorics.

**Hypothesis A.** *There exist constants $\lambda_2 > 0$ and $C_2 > 1$ such that if $v^2 \leq \lambda_2 N/C_2$, $1 \leq u \leq v$ and $A$ is a subset of the integers in $(\frac{N}{ev}, \frac{N}{u}]$ such that*

$$\sum_{a \in A} \frac{1}{a} \geq \frac{1 + \lambda_2}{u},$$

*then there exists an integer $k \in [u, ev]$, an absolute constant $\alpha_v > 0$ and an integer $n \in [N-k, N]$ such that*

$$\sum_{\substack{(a_1, \ldots, a_k) \in A^k \\ a_1 + \cdots + a_k = n}} \frac{1}{a_1 \cdots a_k} \geq \frac{\alpha_v}{N} \left( \sum_{a \in A} \frac{1}{a} \right)^k.$$

**Proposition 5.1(i).** *Hypotheses P and A are equivalent, with $\lambda_2 \asymp \lambda_1$ and $\alpha_v e^{O(v)} \gg \pi_v \gg \alpha_v \min\{1, \lambda_1^v\} v^{-O(v)}$.*

We will prove this at the end of this section. We first note reasons for some of the conditions in Hypothesis A:

• If $A$ is the set of integers in $(N/(k+1), N/k-1)$ then there are no sums of elements of $A$ in the interval $[N-k, N]$ and $\sum_{n \in A} 1/n \sim \log(1 + 1/k) = 1/k + O(1/k^2)$. Hence we must have $\lambda_2 \geq 0$. However, we do believe that Hypothesis A holds for any $\lambda_2 > 0$.

• If $A$ is the set of integers $\equiv 0 \pmod{d}$ in $(\frac{N}{ev}, \frac{N}{v}]$, then $\sum_{a \in A} 1/a \sim 1/d$ as $N \to \infty$, and there are no solutions to $a_1 + \cdots + a_k = n$ for any $n$ in an interval $md < n < (m+1)d$.

Hence $n$ must be chosen from an interval of length $\geq d$. This explains the length of the interval for $n$ in Hypothesis A.

We will eventually prove Hypothesis A with $\alpha_v = 1/v^{O(v)}$ and big enough constant $\lambda_2$. By Proposition 5.1(i) this yields Hypothesis P with $\pi_v = 1/v^{O(v)}$, and therefore Theorem 1 as was shown in Section 4.

Since Hypothesis A involves so many integers, one might think to approximate the set of integers $A$ by a continuous variable; for instance, by considering very short intervals around each $a/N$, so as to obtain Hypothesis T (which is stated in the introduction):

**Proposition 5.1(ii).** *Hypotheses A and T are equivalent, with $\lambda_2 \asymp \lambda_3$ and $\tau_v e^{O(v)} \gg \alpha_v \gg \tau_v / e^{O(v)}$*

Combining this with Proposition 5.1(i) and the result from Section 4 we can deduce Proposition 1 (also stated in the introduction of the paper).

**Proof of Proposition 5.1.** We conclude this section with the proof that our three hypotheses, A, P and T, are equivalent.

*Proof that Hypothesis A implies Hypothesis T.* Assume that $T$ is an open subset of $(1/ev, 1/u)$ such that $\int_T dt/t \geq (1+\lambda_3)/u$. An open subset of the reals is a union of disjoint open intervals, and the number of intervals in the union is countable (as may be seen by labeling each interval with some rational it contains). Hence we may write $T = \bigcup_{i \geq 1}(\alpha_i, \beta_i)$. But then there exists an integer $m$ such that if $S = \bigcup_{i=1}^{m}(\alpha_i, \beta_i)$, then $\int_{t \in S} dt/t \geq (1 + 2\lambda_3/3)/u$. By replacing $T$ in our assumption by $S$ and $\lambda_3$ by $2\lambda_3/3$, we may assume that $T$ is a finite union of open intervals. We select $N$ to be much larger than $\max_i\{v/|\alpha_i - \beta_i|\}$, $mv^3/\lambda_3$ and $mv^4$. Let

$$A = \bigcup_{i=1}^{m}\{a \in \mathbb{Z} : \ \alpha_i N + 2ev < a < \beta_i N - 2ev\}.$$

Since $\int_{t=a/N}^{(a+1)/N} dt/t = 1/a + O(1/a^2)$, we deduce that

$$\sum_{a \in A}\frac{1}{a} = \int_{t \in T}\frac{dt}{t} + O\left(\frac{mv^2}{N}\right) = \left(1 + O\left(\frac{mv^3}{N}\right)\right)\int_{t \in T}\frac{dt}{t} \geq \frac{1 + \lambda_3/2}{u}.$$

Now, if $\lambda_3 \geq 2\lambda_2$, then Hypothesis A implies that there exists an integer $k \in [u, ev]$ and an integer $n \in [N-k, N]$ such that

$$\sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ a_1 + \cdots + a_k = n}}\frac{1}{a_1 \cdots a_k} \geq \frac{\alpha_v}{N}\left(\sum_{a \in A}\frac{1}{a}\right)^k \geq \frac{\alpha_v e^{-O(k)}}{N}\left(\int_{t \in T}\frac{dt}{t}\right)^k.$$

For each $k$-tuple $(a_1, \ldots, a_k) \in A^k$ with $a_1 + \cdots + a_k = n$, consider $t_i \in (a_i/N, (a_i+1)/N) \subset T$ for $1 \le i \le k-1$ and define $t_k = 1 - (t_1 + \cdots + t_{k-1})$. Then we have that

$$\left| t_k - \frac{a_k}{N} \right| = \left| \frac{N-n}{N} + \sum_{i=1}^{k-1} \left( \frac{a_i}{N} - t_i \right) \right| \le \frac{2k-1}{N} \le \frac{2ev}{N}$$

and consequently $t_k \in T$. Hence

$$\int \cdots \int_{\substack{t_1+t_2+\cdots+t_k=1 \\ t_1,t_2,\ldots,t_k \in T}} \frac{dt_1 dt_2 \cdots dt_{k-1}}{t_1 t_2 \cdots t_k} \gg \sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ a_1+\cdots+a_k=n}} \frac{N}{a_k} \int_{\substack{a_i/N < t_i < (a_i+1)/N \\ 1 \le i \le k-1}} \frac{dt_1 \cdots dt_{k-1}}{t_1 \cdots t_{k-1}}$$

$$\gg \frac{1}{e^{O(k)}} \sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ a_1+\cdots+a_k=n}} \frac{N}{a_1 \cdots a_k} \ge \frac{\alpha_v}{e^{O(v)}} \left( \int_{t \in T} \frac{dt}{t} \right)^k,$$

as desired. Hence we can take $\lambda_3 = 3\lambda_2$ and $\tau_v \gg \alpha_v e^{-O(v)}$.

*Proof that Hypothesis T implies Hypothesis A.* Let $N \ge C_2 v^2/\lambda_2$ and $A \subset (N/ev, N/u]$ with $\sum_{a \in A} 1/a \ge (1 + \lambda_2)/u$. Set $T = \bigcup_{a \in A} (a/N, (a+1)/N)$, so that

$$\int_T \frac{dt}{t} = \sum_{a \in A} \frac{1}{a} + O\left( \frac{1}{N/v} \right) = \left( 1 + O\left( \frac{v^2}{N} \right) \right) \sum_{a \in A} \frac{1}{a} \ge \frac{1 + \lambda_2/2}{u},$$

provided that $C_2$ is large enough. If $t_1 + t_2 + \cdots + t_k = 1$ and $a_i = [Nt_i]$, then $N - k \le a_1 + \cdots + a_k \le N$. Now, we have that

$$\frac{1}{a_1 \cdots a_k} \ge \int \cdots \int_{a_i/N < t_i < (a_i+1)/N \, \forall i} \frac{dt_1 dt_2 \cdots dt_{k-1}}{t_1 t_2 \cdots t_{k-1} \cdot (Nt_k)},$$

and so

$$\sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ N-k \le a_1+\cdots+a_k \le N}} \frac{1}{a_1 \cdots a_k} \ge \int \cdots \int_{\substack{t_1+t_2+\cdots+t_k=1 \\ t_1,t_2,\ldots,t_k \in T}} \frac{dt_1 dt_2 \cdots dt_{k-1}}{Nt_1 t_2 \cdots t_k} \gg \frac{\tau_v}{N} \left( \int_{t \in T} \frac{dt}{t} \right)^k,$$

provided that $\lambda_2 \ge 2\lambda_3$. The result follows by averaging over the subsums with $a_1 + \cdots + a_k = n$ for each integer $n$ in the interval $N - k \le n \le N$. Hence we can take $\lambda_2 = 2\lambda_3$ and $\alpha_v \gg \tau_v/e^{O(v)}$.

*Proof that Hypothesis P implies Hypothesis A.* Given our set $A$, let $\mathcal{P}$ be the set of primes in $\bigcup_{a \in A}(e^a, e^{a+1})$, and let $x = e^{[N]+1}$. Then

$$\sum_{p \in \mathcal{P}} \frac{1}{p} = \sum_{a \in A} \frac{1}{a} + O\left( \frac{1}{N/v} \right) = \left( 1 + O\left( \frac{v^2}{N} \right) \right) \sum_{a \in A} \frac{1}{a} \ge \frac{1 + \lambda_2/2}{u},$$

provided that $C_2$ is large enough. So, if we choose $\lambda_2 = 2\lambda_1$, then we can apply Hypothesis P. Now for each $k$-tuple of primes $(p_1, \ldots, p_k) \in \mathcal{P}^k$ such that $x/2 < p_1 \cdots p_k < x$, let $a_j = [\log p_j]$ for all $j$, so that

$$\sum_{j=1}^{k} a_j > \sum_{j=1}^{k} (\log p_j - 1) \geq \log(x/2) - k = [N] + 1 - \log 2 - k$$

and

$$\sum_{j=1}^{k} a_j \leq \sum_{j=1}^{k} \log p_j < \log x = [N] + 1.$$

Since $a_j$ are integers, this implies that $N - k \leq \sum_{j=1}^{k} a_j \leq N$. Hence we deduce that

$$\sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ N-k \leq a_1 + \cdots + a_k \leq N}} \frac{1}{a_1 \cdots a_k} \gg e^{-O(k)} \sum_{\substack{(p_1,\ldots,p_k) \in \mathcal{P}^k \\ x/2 < p_1 \cdots p_k < x}} \frac{1}{p_1 \cdots p_k}$$

$$\gg \frac{e^{-O(k)} \pi_v}{\log x} \cdot \left( \sum_{p \in \mathcal{P}} \frac{1}{p} \right)^k \gg \frac{\pi_v}{e^{O(k)}} \cdot \frac{1}{N} \left( \sum_{a \in A} \frac{1}{a} \right)^k.$$

Hence we can take $\lambda_2 = 2\lambda_1$ and $\alpha_v \gg \pi_v e^{-O(v)}$.

*Proof that Hypothesis A implies Hypothesis P.* Let $\rho = 1 + \delta/(2ev), \eta = \min\{1, \lambda_2\}$ and $N = \log_\rho x - ev$. For each integer $a \in [N/ev + 1, N/u]$ define

$$S_a = \sum_{\substack{p \in \mathcal{P} \\ \rho^a \leq p < \rho^{a+1}}} \frac{1}{p}.$$

Huxley's prime number theorem for short intervals (see Theorem 10.5 in [7] and the subsequent discussion]) yields $|\{y \leq p \leq y + h\}| \sim h/\log y$ for $y^{7/12+\epsilon} \leq h \leq y$. This implies that

(5.1) $$S_a \leq \frac{1 + \eta/10}{a} \ll \frac{v}{N} \asymp \frac{\delta}{\log x}$$

when $C_1$, and thus $x^{1/ev}$, is large enough.

Let $A$ be the set of integers $a$ for which $S_a \geq \frac{\eta}{4a \log(ev)} \sum_{p \in \mathcal{P}} \frac{1}{p}$. Then

$$\sum_{\substack{p \in \mathcal{P} \\ [\log_\rho p] \notin A}} \frac{1}{p} \leq O\left( \frac{\delta}{\log x} \right) + \sum_{\substack{a \notin A \\ \frac{N}{ev} + 1 \leq a \leq \frac{N}{u}}} \frac{\eta}{4a \log(ev)} \sum_{p \in \mathcal{P}} \frac{1}{p} \leq \left( \frac{\eta}{4} + O\left( \frac{\delta v}{\log x} \right) \right) \sum_{p \in \mathcal{P}} \frac{1}{p}$$

$$\leq \frac{\eta}{3} \sum_{p \in \mathcal{P}} \frac{1}{p},$$

provided that $C_1$ is large enough and $\lambda_1 \leq 100\lambda_2$. Using (5.1) we find that

$$\sum_{a \in A} \frac{1}{a} \geq \left(1 - \frac{\eta}{10}\right) \sum_{a \in A} S_a = \left(1 - \frac{\eta}{10}\right) \sum_{\substack{p \in \mathcal{P} \\ [\log_\rho p] \in A}} \frac{1}{p} \geq \left(1 - \frac{\eta}{2}\right) \sum_{p \in \mathcal{P}} \frac{1}{p}.$$

So setting $\lambda_1 = 4\lambda_2$ allows us to apply Hypothesis A. Now for each solution to $a_1 + \cdots + a_k = n \in [N - k, N]$ with $a_1, \ldots, a_k \in A$, consider the primes $p_j \in \mathcal{P}$ with $[\log_\rho p_j] = a_j$. Note that $a_j \leq \log_\rho p_j < a_j + 1$ and so $\log_\rho x - ev - k \leq n \leq \log_\rho(p_1 p_2 \cdots p_k) < n + k \leq \log_\rho x$, which implies that $x > p_1 \cdots p_k > x\rho^{-2ev} = x(1 + \delta/(2ev))^{-2ev} \geq (1 - \delta)x$. Hence

$$\sum_{\substack{(p_1, \ldots, p_k) \in \mathcal{P}^k \\ (1-\delta)x < p_1 \cdots p_k < x}} \frac{1}{p_1 \cdots p_k} \geq \sum_{\substack{(a_1, \ldots, a_k) \in A^k \\ a_1 + \cdots + a_k = n}} S_{a_1} \cdots S_{a_k}$$

$$\geq \left(\frac{\eta}{4\log(ev)} \sum_{p \in \mathcal{P}} \frac{1}{p}\right)^k \sum_{\substack{(a_1, \ldots, a_k) \in A^k \\ a_1 + \cdots + a_k = n}} \frac{1}{a_1 \cdots a_k}$$

$$\geq \left(\frac{\eta(1 + \lambda_1)}{4u\log(ev)}\right)^k \cdot \frac{\alpha_v}{N} \left(\sum_{a \in A} \frac{1}{a}\right)^k \gg \frac{\eta^v \alpha_v}{v^{O(v)}} \cdot \frac{\delta}{\log x} \left(\sum_{p \in \mathcal{P}} \frac{1}{p}\right)^k,$$

which proves the desired result.

## 6. LEMMAS IN ADDITIVE COMBINATORICS

Let us first introduce some notation. Given two additive sets $A$ and $B$, define the sum set $A + B = \{a + b : a \in A, b \in B\}$, the $k$-fold sum set $kA = \{a_1 + \cdots + a_k : a_i \in A\}$, and for any $E \subseteq A \times B$ define the restricted sum set

$$A \overset{E}{+} B = \{a + b : (a, b) \in E\}.$$

Write also

$$r_{kA}(n) = \#\{(a_1, \ldots, a_k) \in A^k : a_1 + \cdots + a_k = n\}$$

for the number of representations. Finally, a set of the form

$$P = \{x_0 + l_1 x_1 + \cdots + l_d x_d : 0 \leq l_j \leq L_j \text{ for all } j\}$$

is called a *generalized arithmetic progression of rank d*.

We need three lemmas from additive combinatorics. The first one lets us pass from a restricted sum set to a regular sum set.

**Lemma 6.1.** *Let $(G, +)$ be an abelian group. If $E \subseteq A \times A$ satisfies*

$$|E| \geq (1 - \delta^2)|A|^2 \quad and \quad |A \overset{E}{+} A| \leq K|A|,$$

*then there exists a set $A' \subseteq A$ such that*

$$|A'| \geq (1 - 2\delta)|A| \quad and \quad |A' - A'| \leq \frac{K^2}{1 - 2\delta}|A|.$$

*Proof.* This is a variant of the Balog-Szemerédi-Gowers theorem (see [10, Theorem 2.29]), more precisely a direct consequence of [10, Exercise 2.5.4] which can be proved by incorporating the hint there to the proof of the Balog-Szemerédi-Gowers theorem in [10, Section 6.4]. $\qquad\square$

The second lemma shows that if $3A$ is small, then we can find a popular large generalized arithmetic progression inside it.

**Lemma 6.2.** *Let $K \geq 1$ and let $A$ be a finite subset of the integers such that $|3A| \leq K|A|$. Then there is a generalized arithmetic progression $P \subseteq 3A$ of rank $O_K(1)$ such that $|P| \gg_K |A|$ and $r_{3A}(n) \gg_K |A|^2$ for all $n \in P$.*

*Proof.* This is a variant of the Ruzsa-Chang theorem (see [10, Theorem 5.30]). That theorem concerns the set $2A - 2A$ instead of $3A$, but replacing the application of [10, Theorem 4.42] in the proof of [10, Theorem 5.30] by an application of [10, Theorem 4.43] handles this. Furthermore, we have added a requirement $r_{3A}(n) \gg_K |A|^2$ for all $n \in P$ but that this is immediate from the proof of [10, Theorem 4.43] (replacing $1/(2\pi K)$ by $1/(4\pi K)$ all over the proof lets one replace statements of the type "$T \subseteq 3A$" by statements of the type "$r_{3A}(n) \gg |A|^2$ for every $n \in T$"). $\qquad\square$

The third lemma shows that sum sets of generalized arithmetic progressions have large popular subsets.

**Lemma 6.3.** *Let $L_j$ be positive integers for $j = 1, \ldots, d$ and let*

$$P = \{x_0 + l_1 x_1 + \cdots + l_d x_d : |l_j| \leq L_j \text{ for all } j\}$$

*be a generalized arithmetic progression. Let $k \in \mathbb{N}$, $\delta \in (0, 1/6^d)$, $\rho = 1 - 3\delta^{1/d} \geq 1/2$ and*

$$Q_k = \{kx_0 + l_1 x_1 + \cdots + l_d x_d : |l_j| \leq \rho k L_j \text{ for all } j\}.$$

*Then, for any $n \in Q_k$, one has $r_{kP}(n) \geq (\delta|P|)^{k-1}$.*

*Proof.* We proceed by induction on $k$. The case $k = 1$ is trivial, so we assume that the claim holds for some $k \in \mathbb{N}$. Let $n \in Q_{k+1}$, so that $n = (k + 1)x_0 + l_1 x_1 + \cdots + l_d x_d$ with $|l_j| \leq \rho(k + 1)L_j$ for all $j$. Now

$$r_{(k+1)P}(n) = \sum_{|i_j| \leq L_j \, \forall j} r_{kP}(kx_0 + (l_1 - i_1)x_1 + \cdots + (l_d - i_d)x_d)$$

$$\geq (\delta|P|)^{k-1} \cdot \#\{(i_1, \ldots, i_d) : |i_j| \leq L_j \text{ and } |l_j - i_j| \leq \rho k L_j \text{ for all } j\}.$$

The right hand side is smallest when $l_j = [\rho(k+1)L_j]$ for all $j \in \{1, \ldots, d\}$, in which case

$$\#\{i_j \in [-L_j, L_j] : |l_j - i_j| \leq \rho k L_j\}$$
$$= \#\{i_j : [\rho(k+1)L_j] - \rho k L_j \leq i_j \leq \min\{L_j, [\rho(k+1)L_j] + \rho k L_j\}\}.$$

Since $2\rho k \geq 2\rho \geq 1$, the minimum above is $L_j$ and hence the number of counted $i_j$ is at least $(1-\rho)L_j$. Hence, since $|P| \leq \prod_{j=1}^{d}(2L_j + 1) \leq \prod_{j=1}^{d}(3L_j)$, we have that

$$r_{(k+1)P}(n) \geq (\delta|P|)^{k-1}(1-\rho)^d \prod_{j=1}^{d} L_j = (\delta|P|)^{k-1}\delta \prod_{j=1}^{d}(3L_j) \geq (\delta|P|)^k.$$

$\square$

## 7. The proof of Hypothesis A

Our additive combinatorial tools do not involve logarithmic weights, so instead of Hypothesis A we apply them to prove the following variant.

**Theorem 7.1.** *There exists a constant $c_0 > 1$ such that if $1 \leq u \leq v$ and $B$ is a subset of the integers in $(\frac{N}{ev}, \frac{N}{u}]$ for which*

$$|B| > \frac{c_0 N}{u^2},$$

*then there is an integer $k \in [u, ev]$ such that*

$$|\{(b_1, \ldots, b_k) \in B^k : N - k \leq b_1 + \cdots + b_k \leq N\}| \gg \frac{1}{e^{O(k)}k^k} \frac{|B|^k}{N}.$$

*Proof that Theorem 7.1 implies Hypothesis A with $\lambda_2 = 2c_0 - 1$ and $\alpha_v = v^{-4v}$.* Let $A \subset [N/(ev), N/u]$ as in Hypothesis A. We claim that there must exist $t \in [u, ev]$ such that

$$(7.1) \qquad \sum_{\substack{a \in A \\ N/(ev) < a \leq N/t}} 1 \geq \frac{c_0 N}{t^2}.$$

Indeed, if this is not the case, then

$$\sum_{\substack{a \in A \\ N/(ev) < a \leq N/u}} \frac{1}{a} = -\int_u^{ev} \frac{t}{N} d \left( \sum_{\substack{a \in A \\ N/(ev) < a \leq N/t}} 1 \right)$$

$$= \frac{u}{N} \sum_{\substack{a \in A \\ N/(ev) < a \leq N/u}} 1 + \int_u^{ev} \left( \sum_{\substack{a \in A \\ N/(ev) < a \leq N/t}} 1 \right) \frac{dt}{N}$$

$$< \frac{u}{N} \cdot \frac{c_0 N}{u^2} + c_0 \int_u^{ev} \frac{1}{t^2} dt < 2 \cdot \frac{c_0}{u} = \frac{1+\lambda}{u}$$

which is a contradiction. So there is some $t \in [u, ev]$ for which (7.1) holds.

Now, set $B = A \cap (N/(ev), N/t]$, so that the hypothesis of Theorem 7.1 is satisfied with $t$ in place of $u$ and some $v' \le ev$ in place of $v$. Let $k$ be as in the conclusion of Theorem 7.1, which necessarily lies in $[u, ev]$. Let $n$ be an integer in $[N - k, N]$ whose number of representations as $b_1 + \cdots + b_k$ is maximal. So $n$ has at least $\gg e^{-O(k)} k^{-k} |B|^k / N$ such representations by Theorem 7.1. Since each $b \in B$ satisfies $1/b \ge u/N \ge 1/N$, and as $|B| \ge c_0 N/t^2 \ge c_0 N/(ev)^2$, we deduce that

$$\sum_{\substack{(b_1,\ldots,b_k) \in B^k \\ b_1+\cdots+b_k=n}} \frac{1}{b_1 \cdots b_k} \gg \frac{1}{N} \left( \frac{e^{-O(1)}|B|}{Nk} \right)^k \gg \frac{e^{-O(v)}}{v^{3v} \cdot N} \ge \frac{\alpha_v}{N} \left( \sum_{a \in A} \frac{1}{a} \right)^k$$

with $\alpha_v \gg v^{-4v}$, since $\sum_{a \in A} \frac{1}{a} \ll \log(ev)$. $\qquad\square$

*Proof of Theorem 7.1.* Let $c_0$ be a large positive constant to be determined later. Notice first that if $|B| \le c_0$, then we only need to find one sum in the interval $[N - k, N]$. In this case the elements of $B$ have size $\le N/u \le u$ by the lower bound for $|B|$, so the claim follows trivially.

From now on we can assume that $|B| > c_0$. We claim that if $|B| > c_0 N/u^2$, then there exists $k$ such that

$$|\{(b_1, \ldots, b_k) \in B^k : N - k \le b_1 + \cdots + b_k \le N\}| \ge c \frac{\delta^{2k}|B|^k}{Nk^k}$$

for some appropriate small positive constants $c$ and $\delta$. If $1 \le u \le c_0$, our claim is trivial, since there are no sets $B \subset (N/ev, N/u]$ with $|B| > c_0 N/u^2$.

Now we prove that if the claim holds when $2^{j-1} c_0 \le u \le 2^j c_0$ for some $j \ge 0$, then it holds when $2^j c_0 \le u \le 2^{j+1} c_0$. Take

$$E = \left\{ (b_1, b_2) \in B \times B : r_{2B}(b_1 + b_2) \ge \delta^2 \frac{|B|^2}{|2B|} \right\},$$

so that

$$|E| \ge |B|^2 - |2B| \cdot \delta^2 \frac{|B|^2}{|2B|} = (1 - \delta^2)|B|^2.$$

Write $C = B \overset{E}{+} B \subseteq (2N/(ev), 2N/u]$. We split the rest of the argument into two cases according to whether $|C| > 4|B|$ or not.

Consider first the case $|C| > 4|B|$. Then

$$|C| > \frac{c_0 N}{(u/2)^2},$$

and by induction hypothesis there is an integer $k/2 \in [u/2, ev/2]$ such that

$$|\{(c_1, \ldots, c_{k/2}) \in C^{k/2} : N - k/2 \leq c_1 + \cdots + c_{k/2} \leq N\}| \geq c \frac{\delta^k |C|^{k/2}}{N(k/2)^{k/2}}.$$

Hence by the definition of $C$ we have

$$|\{(b_1, \ldots, b_k) \in B^k : N - k/2 \leq b_1 + \cdots + b_k \leq N\}|$$
$$\geq c \frac{\delta^k |C|^{k/2}}{N(k/2)^{k/2}} \cdot \left(\delta^2 \frac{|B|^2}{|2B|}\right)^{k/2} = c \frac{\delta^{2k} |B|^k}{Nk^{k/2}} \cdot \left(\frac{2|C|}{|2B|}\right)^{k/2}.$$

The claim now follows if $|C| > |2B|/2k$; this easily follows, since $|C| > 4|B|$ and $|2B| \leq 2N/u \leq 2|B|u/c_0 \leq 2|B|k$.

On the other hand, if $|C| \leq 4|B|$, then by Lemma 6.1 there is $B' \subseteq B$ such that $|B'| \geq |B|/2$ and $|B' - B'| \leq 20|B'|$. Then by [10, Proposition 2.26] we have $|3B'| \ll |B'|$ and hence Lemma 6.2 implies that there is a generalized arithmetic progression $P$ of rank $d \ll 1$ such that $|P| \gg |B|$ and $r_{3B}(n) \gg |B|^2$ for all $n \in P$. "Centralizing" $P$ we can assume that it is of form

$$P = \{x_0 + l_1 x_1 + \cdots + l_d x_d : |l_j| \leq L_j \text{ for all } j\}$$

for some positive integers $L_j$ (doing this reduces the size of $P$ at most by a factor of $1/3^d \gg 1$). Set $\rho = 1 - (3\delta)^{1/d}$ and

$$Q = \{x_0 + l_1 x_1 + \cdots + l_d x_d : |l_j| \leq \rho L_j \text{ for all } j\} \subset P.$$

If $\delta < 1/(6^d)$, then Lemma 6.3 implies that $r_{kP}(n) \geq (\delta|P|)^{k-1}$ for all $k \geq 1$ and $n \in kQ$. Moreover, if $\delta$ is small enough, then $|Q| \geq |P|/3^d \gg |B| > c_0 N/u^2$, so that if $c_0$ is large enough, then $|Q| \geq 15N/u^2$ and, because

$$Q \subseteq P \subseteq 3B \subseteq (3N/(ev), 3N/u],$$

we have

$$\sum_{q \in Q} \frac{1}{q} \geq \frac{15N}{u^2} \cdot \frac{u}{3N} \geq \frac{5}{u}.$$

From here we argue much as in the proof of Proposition 5.1(ii). We begin by removing $[3N/u]$ from $Q$ if it is an element so that $Q \subseteq (3N/(ev), 3N/u - 1]$. Let $T = \bigcup_{q \in Q} \left(\frac{q}{N}, \frac{q+1}{N}\right) \subseteq \left(\frac{3}{ev}, \frac{3}{u}\right)$, so that

$$\int_T \frac{dt}{t} \geq \sum_{q \in Q} \left(1 - \frac{2}{q}\right) \frac{1}{q} \geq \left(1 - \frac{ev}{6N}\right) \sum_{q \in Q} \frac{1}{q} \geq \frac{3}{u}$$

since $N \geq ev$. Hence, Bleichenbacher's theorem implies that there exists an integer $k/3 \in [u/3, ev/3]$ and $t_1, \ldots, t_{k/3} \in T$ for which $t_1 + \ldots + t_{k/3} = 1$. If $t_i \in \left(\frac{q_i}{N}, \frac{q_i+1}{N}\right)$ for each $i$ then $n := q_1 + \ldots + q_{k/3} \in \frac{k}{3}Q \cap [N - k/3, N]$. Recalling that $r_{3B}(m) \gg |B|^2$ for every $m \in P$ and $r_{\frac{k}{3}P}(\ell) \geq (\delta|P|)^{k/3-1}$ for every $\ell \in \frac{k}{3}Q$, we get that

$$
\begin{aligned}
&\#\{(b_1, \ldots, b_k) \in B^k : N - k/3 \leq b_1 + \cdots + b_k \leq N\} \\
&\geq \#\{(b_1, \ldots, b_k) \in B^k : b_1 + \cdots + b_k = n\} \\
&\geq \sum_{\substack{m_1+\ldots+m_{k/3}=n \\ m_1,\ldots,m_{k/3}\in 3B}} \prod_{j=1}^{k/3} r_{3B}(m_j) \geq r_{\frac{k}{3}\cdot P}(n) \cdot (e^{-O(1)}|B|^2)^{k/3} \geq (\delta|P|)^{k/3-1} \cdot e^{-O(k)}|B|^{2k/3} \\
&\geq e^{-O(k)}(\delta|B|)^{k-1} \geq ue^{-O(k)}(\delta|B|)^k/N,
\end{aligned}
$$

as $P \subseteq 3B$. The result follows, since $ue^{-O(k)} \geq (\delta/k)^k$ if $\delta$ is sufficiently small. $\qquad\square$

*Remark 7.1.* One could compute the constant $c_0$ explicitly and thereby the constants $\lambda_i$ in hypothesis P, A and T and, eventually, $\lambda$ in Theorem 1. However, $c_0$ will be relatively large, since the implied constants in Lemma 6.2 are rather large. If one is interested in optimizing $\lambda$, one could, instead of Lemma 6.2, use a result of Lev [9] to show that if the number of "popular" elements in $|2A|$ is at most $K|A|$ for some $K < (3 + \sqrt{5})/2$, then $2A$ contains a "popular" arithmetic progression (of rank 1). Modifying the above arguments, this would lead to Theorem 1 with a smaller and more easily calculable $\lambda < 21$. However, this argument would not yield Hypotheses P, A and T when $u$ is not close to 1 and, in particular, not the latter conclusion in Remark 1.4. By applying Bleichenbacher's theorem in a different way one could probably improve $\lambda$ further, but not to an arbitrarily small constant, as desired.

## 8. Further remarks on Hypotheses P, A and T.

We conclude our paper with an investigation of some other consequences of Hypotheses P, A and T.

**Proposition 8.1.** *Suppose Hypothesis P holds with $\pi_v = 1/v^{O(v)}$. If $\epsilon > 0$, $v^2 \leq \lambda_1 \log x/C_1$, $v \geq 1$ and $\mathcal{P}$ is a subset of the primes in $(x^{1/ev}, x^{1/v}]$ for which $\sigma := \sum_{p\in\mathcal{P}} \frac{1}{p} \geq \frac{1+\lambda_1+\epsilon}{v}$, then there exists an integer $k \in [v, ev]$ such that*

$$
\Psi(x; \mathcal{P}) \gg_\epsilon \frac{x}{\log x} \cdot \frac{\sigma^{k+O(1/\sigma)}}{k \cdot k!}.
$$

*Proof.* We can clearly assume that $\epsilon$ is small. Also, since $\Psi(x; \mathcal{P}) \geq 1$, we may assume that $x$ is large. If $v < 2$ or $\frac{1+\lambda_1}{v-1} > \sigma \geq \frac{1+\lambda_1+\epsilon}{v}$, then $v < \max\{2, \frac{1+\lambda_1+\epsilon}{\epsilon}\} = \frac{1+\lambda_1+\epsilon}{\epsilon}$ and

the proposition follows by Hypothesis P. So we may impose the additional assumptions that $v \geq 2$ and $\sigma \geq \frac{1+\lambda_1}{v-1}$.

Select $w = (1 + \lambda_1)/\sigma$ and note that $w \in [1, v-1]$, since $\sigma \leq \sum_{x^{1/(ev)} < p \leq x^{1/v}} 1/p \lesssim 1$. We begin by applying Corollary 3.2 with $X = x^{1-\frac{w}{v}} \geq x^{1/v}$. So there is $\ell \ll v$ such that

$$\sum_{\substack{q_1,\ldots,q_\ell \in \mathcal{P} \\ x^{1-\frac{w+1}{v}} < q_1 q_2 \cdots q_\ell \leq x^{1-\frac{w}{v}}}} \frac{1}{q_1 q_2 \cdots q_\ell} \gg \frac{\sigma^\ell}{v}.$$

For each $q_1 \cdots q_\ell$ in the above sum we apply Hypothesis P with $x$ replaced by $x/(q_1 \cdots q_\ell)$, and both $u$ and $v$ replaced by $V = \log(x/(q_1 \cdots q_\ell))/\log(x^{1/v})$, which is possible since $V \in [w, w+1]$. Consequently,

$$\sum_{\substack{(q_1,\ldots,q_\ell,p_1,\ldots,p_m) \in \mathcal{P}^{\ell+m} \\ x/2 < q_1 \cdots q_\ell p_1 \cdots p_m < x}} \frac{1}{q_1 \cdots q_\ell p_1 \cdots p_m} \gg \frac{\sigma^{\ell+m}}{w^{O(w)} v \log x}.$$

Finally, note that

$$\begin{aligned}
\Psi(x; \mathcal{P}) &\geq \frac{1}{(\ell+m)!} \sum_{\substack{(q_1,\ldots,q_\ell,p_1,\ldots,p_m) \in \mathcal{P}^{\ell+m} \\ q_1 \cdots q_\ell p_1 \cdots p_m \leq x}} 1 \\
&\geq \frac{x}{2(\ell+m)!} \sum_{\substack{(q_1,\ldots,q_\ell,p_1,\ldots,p_m) \in \mathcal{P}^{\ell+m} \\ x/2 < q_1 \cdots q_\ell p_1 \cdots p_m < x}} \frac{1}{q_1 \cdots q_\ell p_1 \cdots p_m} \\
&\gg \frac{x}{\log x} \cdot \frac{\sigma^{\ell+m}}{w^{O(w)} v (\ell+m)!}.
\end{aligned}$$

Letting $k = \ell + m$ and observing that, necessarily, $k \in [v, ev]$ and $w \log w \ll \sigma^{-1} \log(1/\sigma)$ completes the proof of the proposition.     $\square$

**Corollary 8.2.** *There exist constants $c > 1$ and $c' > 0$ such that if $1 \leq v \leq c'\sqrt{\log x}$ and $\mathcal{P}$ is a subset of the primes in $(x^{1/ev}, x^{1/v}]$ for which*

$$\sum_{p \in \mathcal{P}} \frac{1}{p} \geq \frac{\max\{c, \log v\}}{v},$$

*then there exists an integer $k \in [v, ev]$ such that*

$$\Psi(x; \mathcal{P}) \geq \Psi_k(x; \mathcal{P}) \geq \frac{1}{k!}\left(e^{-O(1)} \sum_{p \in \mathcal{P}} \frac{1}{p}\right)^k \cdot \frac{x}{\log x},$$

*as $x \to \infty$, where $\Psi_k(x; \mathcal{P})$ denotes the number of integers $n \leq x$ such that $n$ is squarefree, $n$ has exactly $k$ prime factors, and all of the prime factors of $n$ come from $\mathcal{P}$.*

We conjecture that Corollary 8.2 holds under the weaker assumption that $\sum_{p \in \mathcal{P}} \frac{1}{p} \geq c/v$ for any $c > 1$, with the implied constant depending at most on $c$.

Consider more generally $\Psi_k(x, \mathcal{P})$ for any $\mathcal{P} \subset \{p \leq x\}$. If $n \in (\sqrt{x}, x]$ is counted by $\Psi_k(x; \mathcal{P})$, then we can uniquely write $n = mp$ with $p$ prime and $m$ composed of primes $< p$. Note that $p^k \geq n \geq \sqrt{x}$. So $m$ has $k - 1$ prime factors and it is $\leq x^{1-1/(2k)}$. Now for each $m$ the number of such primes $p$ is $\lesssim (x/m)/\log(x/m)$ as $x \to \infty$, and therefore

$$\Psi_k(x; \mathcal{P}) \lesssim \sqrt{x} + \frac{2kx}{\log x} \sum_{\substack{p|m \implies p \in \mathcal{P} \\ \omega(m) = k-1}} \frac{\mu^2(m)}{m} \leq \sqrt{x} + \frac{2k}{(k-1)!} \left( \sum_{p \in \mathcal{P}} \frac{1}{p} \right)^{k-1} \cdot \frac{x}{\log x},$$

as $x \to \infty$. In particular one cannot significantly improve the lower bound in Corollary 8.2.

It is not difficult to prove corollaries of Hypotheses A and T that are analogous to Proposition 4.1. Thus we have

**Proposition 8.3.** *Suppose that Hypothesis A holds for some $\lambda_2 > 0$ and $C_2 > 1$ with $\alpha_v = 1/v^{O(v)}$. If $\epsilon > 0$, $v^2 \leq \lambda_2 N/C_2$, $v \geq 1$ and $A$ is a subset of the integers in $(\frac{N}{ev}, \frac{N}{v}]$ such that $\alpha := \sum_{a \in A} 1/a \geq (1 + \lambda_2 + \epsilon)/v$, then there exists an integer $k \in [v, ev]$, and an integer $n$ in the range $N - k \leq n \leq N$, such that*

$$\sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ a_1 + \cdots + a_k = n}} \frac{1}{a_1 \cdots a_k} \gg_\epsilon \frac{1}{N} \cdot \frac{\alpha^{k+O(1/\alpha)}}{k}.$$

Similarly, we have the following result.

**Proposition 8.4.** *Suppose that Hypothesis T holds for some $\lambda_3 > 0$ with $\tau_v = 1/v^{O(v)}$. If $\epsilon > 0$, $v \geq 1$ and $T$ is an open subset of $(\frac{1}{ev}, \frac{1}{v}]$ for which $\tau := \int_{t \in T} dt/t \geq (1 + \lambda_3 + \epsilon)/v$, then there exists an integer $k \in [v, ev]$ such that*

$$\int \cdots \int_{\substack{t_1 + t_2 + \cdots + t_k = 1 \\ t_1, t_2, \ldots, t_k \in T}} \frac{dt_1 dt_2 \cdots dt_{k-1}}{t_1 t_2 \cdots t_k} \gg_\epsilon \frac{\tau^{k+O(1/\tau)}}{k}.$$

In the proofs of Propositions 8.3 and 8.4 we need appropriate analogues to Corollary 3.2. The needed result for the proof of Proposition 8.3 follows in a straightforward way from Lemma 3.1. For the proof of Proposition 8.4 we make note, without proof, of the appropriate result:

**Lemma 8.5.** *Let $T$ be an open subset of $(\frac{1}{ev}, \frac{1}{u}]$. For any $w \geq 1/v$ there exists a positive integer $\ell \leq evw$ such that*

$$\int \cdots \int_{\substack{w-1/v < t_1 + t_2 + \cdots + t_\ell \leq w \\ t_1, t_2, \ldots, t_\ell \in T}} \frac{dt_1 dt_2 \cdots dt_\ell}{t_1 t_2 \cdots t_\ell} \geq \frac{1}{evw} \left( \int_{t \in T} \frac{dt}{t} \right)^\ell.$$

## References

1. D. Bleichenbacher, *The continuous postage stamp problem*, unpublished manuscript (2003.).

2. J.B. Friedlander, *Integers free from large and small primes*, Proc. London Math. Soc. **33** (1976), 565-576.

3. J.B. Friedlander and H. Iwaniec, *Opera de Cribro* (American Mathematical Society Colloqium Publications), vol. 57, American Mathematical Society, Providence, RI, 2010.

4. A. Granville and K. Soundararajan, *The number of unsieved integers up to x*, Acta Arith. **115** (2004), 305-328.

5. R.R. Hall, *Halving an estimate obtained from Selberg's upper bound method*, Acta Arith. **25** (1974), 347-351.

6. A. Hildebrand, *Quantitative mean value theorems for nonnegative multiplicative functions II*, Acta Arith. **48** (1987), 209-260.

7. H. Iwaniec and E. Kowalski, *Analytic Number Theory* (American Mathematical Society Colloqium Publications), vol. 53, American Mathematical Society, Providence, RI, 2004.

8. H.W. Lenstra jr. and Carl Pomerance, *Primality testing with Gaussian periods* (to appear, preprint available at http://www.math.dartmouth.edu/∼carlp/aks041411.pdf).

9. V. F. Lev, *Restricted set addition in groups. III. Integer sumsets with generic restrictions*, Period. Math. Hungar. **42** (2001), 89-98.

10. T. Tao and V. H. Vu, *Additive combinatorics* (Cambridge Studies in Advanced Mathematics), vol. 105, Cambridge University Press, Cambridge, 2006.

Département de mathématiques et de statistiques, Université de Montréal, CP 6128 succ. Centre-Ville, Montréal QC H3C 3J7, Canada.

*E-mail address*: `andrew@dms.umontreal.ca`

Centre de recherches mathématiques, Université de Montréal, CP 6128 succ. Centre-Ville, Montréal, QC H3C 3J7, Canada.

*E-mail address*: `koukoulo@crm.umontreal.ca`

Department of Mathematics, University of Turku, 20014 Turku, Finland.

*E-mail address*: `ksmato@utu.fi`