WHEN THE SIEVE WORKS II

KAISA MATOMÄKI AND XUANCHENG SHAO

ABSTRACT. For a set of primes $\mathcal{P}$, let $\Psi(x;\mathcal{P})$ be the number of positive integers $n \le x$ all of whose prime factors lie in $\mathcal{P}$. In this paper we classify the sets of primes $\mathcal{P}$ such that $\Psi(x;\mathcal{P})$ is within a constant factor of its expected value. This task was recently initiated by Granville, Koukoulopoulos and Matomäki [6] and their main conjecture is proved in this paper. In particular our main theorem implies that, if not too many large primes are sieved out in the sense that

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

for some $\varepsilon > 0$ and $v \ge u \ge 1$, then

$$\Psi(x;\mathcal{P}) \gg_{\varepsilon,v} x \prod_{\substack{p\le x\\ p\notin\mathcal{P}}} \left(1-\frac{1}{p}\right).$$

1. INTRODUCTION

Let $\mathbb{P}$ be the set of all primes and let $\mathcal{P} \subseteq \mathbb{P}$ be a subset of the primes $\le x$. We study the most basic sieving problem, wishing to estimate

$$\Psi(x;\mathcal{P}) := \left|\{n \le x \colon p \mid n \implies p \in \mathcal{P}\}\right|.$$

In other words we sieve the integers in $[1,x]$ by the primes in $\mathcal{P}^c = (\mathbb{P}\cap[1,x]) \setminus \mathcal{P}$. A simple inclusion-exclusion argument suggests that $\Psi(x;\mathcal{P})$ should be approximated by

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

This is always an upper bound, up to a constant, and a lower bound, up to a constant, if $\mathcal{P}$ contains all the primes larger than $x^{1/2-o(1)}$ (see [5, Theorem 11.13] noticing that the sieving limit $\beta = 2$ for $\kappa = 1$). On the other hand there are examples where $\Psi(x;\mathcal{P})$ is much smaller than the expected lower bound. For instance if one fixes $u \ge 1$ and lets $\mathcal{P}$ consist of all the primes up to $x^{1/u}$, then the prediction is about x/u whereas, by an estimate for the number of smooth numbers, we know that $\Psi(x;\mathcal{P}) = \rho(u)x$ with $\rho(u) = u^{-u(1+o(1))}$ as $u \to \infty$, which is much smaller for large u.

KM was supported by Academy of Finland grants no. 137883 and 138522.

XS is supported by a Glasstone Research Fellowship.

The first ones to study what happens if one also sieves out some primes from $[x^{1/2}, x]$ were Granville, Koukoulopoulos and Matomäki [6]. They conjectured that the critical issue is what is the largest y such that

$$\sum_{\substack{p\in\mathcal{P}\\ y\le p\le x^{1/u}}} \frac{1}{p} \ge \frac{1+\varepsilon}{u}. \tag{1.1}$$

More precisely, they conjectured that when this inequality holds, the sieve works about as expected. On the other hand they gave examples with

$$\sum_{y\le p\le x^{1/u}} \frac{1}{p} = \frac{1-\varepsilon}{u}$$

such that $\Psi(x;\mathcal{P})$ is much smaller than expected.

Here we continue this study and show that the conjecture indeed holds.

Theorem 1.1. *Fix $\varepsilon > 0$. If x is large and $\mathcal{P}$ is a subset of the primes $\le x$ for which there are some $1 \le u \le v \le \frac{\log x}{1000\log\log x}$ with*

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

then

$$\frac{\Psi(x;\mathcal{P})}{x} \ge A_v \prod_{p\in\mathcal{P}^c}\left(1-\frac{1}{p}\right),$$

where A_v is a constant with $A_v = v^{-v(1+o_\varepsilon(1))}$ as $v\to\infty$. If u is fixed, one can take $A_v = v^{-e^{-1/u}v(1+o_\varepsilon(1))}$ as $v\to\infty$.

This establishes the main conjecture of [6]. Notice that when $\mathcal{P}$ consists of all the primes $\le x^{e^{(1+2\varepsilon)/u}/v}$, the conditions of the theorem are satisfied and an estimate for smooth numbers shows that

$$\frac{\Psi(x;\mathcal{P})}{x} = v^{-v(e^{-(1+2\varepsilon)/u}+o(1))}$$

and hence the dependence of the constant A_v on v is close to best possible. More refined questions about the asymptotic behavior of A_v may be asked, but we will not address the issue here. One may also ask the same question for sieving in short intervals (or more general sets), but unfortunately our methods seem pretty specific to the case of long intervals (see also [6, Remark 1.4]).

Granville, Koukoulopoulos and Matomäki [6, Sections 3–4] have reduced a slightly weaker form of the conjecture to an additive combinatorial problem similar to the following hypothesis. We will deduce Theorem 1.1 from Hypothesis A in Section 2.1.

Hypothesis A. *Fix $\lambda \in (0,1)$. Let $N \geq v \geq u \geq 1$ be such that $N \geq (100v/\lambda)^2$. If A is a subset of the integers in $(\frac{N}{v}, \frac{N}{u}]$ such that*

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

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

$$|\{(a_1, \dots, a_k) \in A^k : N - k < a_1 + \cdots + a_k \leq N\}| \geq \alpha_v \frac{|A|^k}{N},$$

where α_v is a constant with $\alpha_v = v^{-o_\lambda(v)}$ as $v \to \infty$. If u is fixed and $v \geq 1000u^2/\lambda^2$, one can take $k \leq e^{-1/u}v$.

Furthermore Granville, Koukoulopoulos and Matomäki [6] proved (a slight variant of) Hypothesis A for some large constant λ and $\alpha_v = v^{-O(v)}$ which implies Theorem 1.1 for some large constant ε. Here we will prove Hypothesis A for every $\lambda > 0$ which implies Theorem 1.1 for every $\varepsilon > 0$.

A crucial ingredient is the following result of Bleichenbacher [1] (see [9, Section 9] for the proof) which may be viewed as a qualitative continuous variant of Hypothesis A.

Bleichenbacher's Theorem. *If $u \geq 1$ 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 $t_1, t_2, \dots, t_k \in T$ for which $t_1 + t_2 + \cdots + t_k = 1$.

Actually we will use the following discrete variant of Bleichenbacher's theorem which is a qualitative variant of Hypothesis A.

Proposition 1.2 (Discrete Bleichenbacher)**.** *Let $N \geq u \geq 1$ and let $A \subseteq \{1, \dots, \lfloor N/u \rfloor\}$ be such that*

$$\sum_{a \in A} \frac{1}{a} > \frac{1}{u} + \frac{1}{\sqrt{N} - 1}.$$

Then there exists $a_1, \dots, a_k \in A$ such that $N - k < a_1 + \cdots + a_k \leq N$.

Proof. Notice first that the claim follows trivially if there is $a \in A$ such that $a < \sqrt{N}$ since in this case there is $k \geq \sqrt{N}$ such that $N - \sqrt{N} < ka \leq N$. Hence we can assume that $A \subseteq \{\lceil \sqrt{N} \rceil, \lceil \sqrt{N} \rceil + 1, \dots, \lfloor N/u \rfloor\}$.

Define $T = \bigcup_{a \in A} \left(\frac{a}{N}, \frac{a+1}{N}\right)$ so that

$$\begin{aligned}\int_{t \in T} \frac{dt}{t} &= \sum_{a \in A} \int_{a/N}^{(a+1)/N} \frac{dt}{t} = \sum_{a \in A} \log\left(1 + \frac{1}{a}\right) \geq \sum_{a \in A} \left(\frac{1}{a} - \frac{1}{a^2}\right) \\ &> \frac{1}{u} + \frac{1}{\sqrt{N} - 1} - \sum_{a > \lceil \sqrt{N} \rceil} \frac{1}{a^2} > \frac{1}{u}.\end{aligned}$$

Then Bleichenbacher's theorem implies that there are $t_1, \dots, t_k \in T$ such that $t_1 + \cdots + t_k = 1$. For each j there is a_{i_j} such that $t_j \in (a_{i_j}/N, (a_{i_j}+1)/N)$. But then $N - k < a_{i_1} + \cdots + a_{i_k} < N$. □

The proof of Hypothesis A splits into two cases according to whether much of the set A is contained in $[N/u_0, N/u]$ for some $u_0 = O(1)$ or not. In the first case Hypothesis A follows from an arithmetic removal lemma in a straightforward way, whereas in the second case we develop an analogue of the arithmetic removal lemma with a growing number of variables (see Theorem 3.4 below), which could be of independent interest.

Acknowledgements. This work started when both authors were visiting CRM in Montreal during the analytic part of the thematic year in number theory in Fall 2014, whose hospitality is greatly appreciated. Thanks also to Ben Green for helpful discussions.

2. Some initial reductions

2.1. **Deduction of Theorem 1.1 from Hypothesis A.** As in [6], we first reduce proving Theorem 1.1 to proving a variant of Hypothesis A for the primes called Hypothesis P, and then show that Hypothesis A implies Hypothesis P.

Hypothesis P. *Fix $\lambda \in (0,1)$. If x is large, $1 \le u \le v \le \frac{\log x}{999 \log\log x}$ and $\mathcal{P}$ is a subset of the primes in $(x^{1/v}, x^{1/u}]$ for which*

$$\sum_{p \in \mathcal{P}} \frac{1}{p} \ge \frac{1+\lambda}{u},$$

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

$$\left| \left\{ (p_1, \dots, p_k) \in \mathcal{P}^k : \frac{x}{2} \le p_1 \cdots p_k \le x \right\} \right| \ge \pi_v \cdot \frac{x}{v^k \log x},$$

where π_v is a constant with $\pi_v = v^{-o_\lambda(v)}$ for $v \to \infty$. If u is fixed and $v \ge 1000u^2/\lambda^2$, one can take $k \le e^{-1/u} v$.

Proof that Hypothesis P implies Theorem 1.1. We can clearly assume that $\varepsilon < 1/1000$. Let $\mathcal{A} = \mathcal{P} \cap [1, x^{1/v}]$ and $\mathcal{B} = \mathcal{P} \cap (x^{1/v}, x^{1/u}]$ so that

$$\Psi(x; \mathcal{P}) \ge \Psi(x; \mathcal{P} \cap [1, x^{1/u}]) \ge \sum_{\substack{a \le x^{\varepsilon/(5v)} \\ p \mid a \implies p \in \mathcal{A}}} \Psi(x/a; \mathcal{B}),$$

since we can write any n composed only of prime factors from $\mathcal{P} \cap [1, x^{1/u}]$ as $n = ab$ where a and b are composed only of prime factors from $\mathcal{A}$ and $\mathcal{B}$, respectively. For each $a \le x^{\varepsilon/(5v)}$, we have that

$$\sum_{\substack{p \in \mathcal{B} \\ (x/a)^{1/v} < p \le (x/a)^{1/u}}} \frac{1}{p} \ge \sum_{\substack{p \in \mathcal{P} \\ x^{1/v} < p \le x^{1/u}}} \frac{1}{p} - \sum_{\substack{p \in \mathbb{P} \\ (x/a)^{1/u} < p \le x^{1/u}}} \frac{1}{p}$$

If $(x/a)^{1/u} > x^{1/u} - x^{2/(3u)}$, a trivial estimate gives

$$\sum_{\substack{p\in\mathbb{P}\\ (x/a)^{1/u}<p\le x^{1/u}}} \frac{1}{p} \le \frac{2x^{2/(3u)}}{x^{1/u}/2} = 4x^{-1/3u} \le \frac{\varepsilon}{2u}.$$

Otherwise, Huxley's prime number theorem for short intervals (see e.g. Theorem 10.5 in [8] and the subsequent discussion) yields, once x is large enough,

$$\sum_{\substack{p\in\mathbb{P}\\ (x/a)^{1/u}<p\le x^{1/u}}} \frac{1}{p} \le 2\log\frac{\log x^{1/u}}{\log(x/a)^{1/u}} \le -2\log\Big(1-\frac{\varepsilon}{5v}\Big) \le \frac{\varepsilon}{2u}.$$

Hence, in any case,

$$\sum_{\substack{p\in\mathcal{B}\\ (x/a)^{1/v}<p\le(x/a)^{1/u}}} \frac{1}{p} \ge \frac{1+\varepsilon/2}{u},$$

and applying Hypothesis P to the set $\mathcal{B}$ yields that there exists $k\in[u,v]$ such that

$$\Psi(x/a;\mathcal{B}) \ge \pi_v \cdot \frac{x}{av^k\log x},$$

where $\pi_v = v^{-o_\varepsilon(v)}$ for $v\to\infty$. Consequently

$$\frac{\Psi(x;\mathcal{P})}{x} \gg \pi_v \frac{1}{v^k\log x} \sum_{\substack{a\le x^{\varepsilon/5v}\\ p|a\implies p\in\mathcal{A}}} \frac{1}{a} \gg \frac{\pi_v}{v^k} \prod_{\substack{p\le x^{\varepsilon/5v}\\ p\in\mathcal{P}^c}} \Big(1-\frac{1}{p}\Big) \gg \frac{\pi_v}{v^k} \prod_{p\in\mathcal{P}^c}\Big(1-\frac{1}{p}\Big)$$

by [6, Lemma 2.1]. This gives the desired lower bound since $k\le v$, and in case u is fixed, $k\le e^{-1/u}v$ for large enough v. □

Proof that Hypothesis A implies Hypothesis P. Let $\rho = 1+\big(\frac{\lambda}{1000v}\big)^2$ and $N = \log_\rho x - v$. Define, for $j\ge 0$,

$$A_j = \Big\{a\in(N/v,N/u]\colon \sum_{\substack{p\in\mathcal{P}\\ \rho^a\le p<\rho^{a+1}}} \frac{1}{p} \ge \frac{e^{-j}}{a}\Big\}.$$

Let $J_0 = \log\frac{20v\log v}{\lambda}$ and let j_0 be the smallest integer $j\ge 0$ for which

$$\sum_{a\in A_j} \sum_{\substack{p\in\mathcal{P}\\ \rho^a\le p<\rho^{a+1}}} \frac{1}{p} \ge \frac{1+\frac{\lambda}{3}+\frac{\lambda}{3}\cdot\frac{j}{J_0}}{u}. \tag{2.1}$$

Notice that, since

$$\sum_{a\in A_{J_0}} \sum_{\substack{p\in\mathcal{P}\\ \rho^a\le p<\rho^{a+1}}} \frac{1}{p} \ge \sum_{\substack{p\in\mathcal{P}\\ x^{1/v}<p\le x^{1/u}}} \frac{1}{p} - \sum_{N/v<a\le N/u} \frac{e^{-J_0}}{a} - \sum_{x^{1/u}\rho^{-v/u}\le p\le x^{1/u}} \frac{1}{p} \ge \frac{1+2\lambda/3}{u},$$

necessarily $j_0 \le J_0$. Write $A = A_{j_0}$. Then, by Huxley's prime number theorem in short intervals,

$$\sum_{a\in A}\frac{1}{a} \ge \sum_{a\in A}\log\left(1+\frac{1}{a}\right) \ge (1-\lambda/100)\sum_{a\in A_{j_0}}\sum_{\substack{p\in\mathcal{P}\\ \rho^a\le p<\rho^{a+1}}}\frac{1}{p} \ge \frac{1+\frac{\lambda}{4}}{u}. \tag{2.2}$$

Furthermore, since j_0 was chosen to be the smallest integer for which (2.1) holds, we get that

$$\sum_{a\in A_{j_0}}\frac{e^{-j_0+1}}{a} \ge \sum_{a\in A_{j_0}\setminus A_{j_0-1}}\sum_{\substack{p\in\mathcal{P}\\ \rho^a\le p<\rho^{a+1}}}\frac{1}{p} \ge \frac{\lambda}{3J_0u},$$

so that

$$|A|\cdot e^{-j_0} \ge \frac{N}{ev}\cdot\frac{\lambda}{3J_0u} \gg \frac{\lambda}{v\log v}\cdot\frac{N}{u}.$$

By (2.2), we can apply Hypothesis A to the set A which gives that, for some $k \le v$ (or in case u is fixed and $v \ge 1000u^2/\lambda^2$, $k \le e^{-1/u}v$),

$$|\{(a_1,\dots,a_k)\in A^k\colon N-k<a_1+\cdots+a_k\le N\}| \ge \alpha_v\frac{|A|^k}{N},$$

where $\alpha_v = v^{-o_\lambda(v)}$.

Now for each solution to $a_1+\cdots+a_k \in (N-k,N]$ with $a_1,\dots,a_k\in A$, consider the primes $p_j\in\mathcal{P}$ with $\rho^{a_j}\le p_j<\rho^{a_j+1}$. Note that for such primes $p_1\cdots p_k \le \rho^{N+k}\le x$ and

$$p_1\cdots p_k \ge \rho^{N-k} \ge x\rho^{-2v} = x\left(1+\left(\frac{\lambda}{1000v}\right)^2\right)^{-2v} \ge x/2.$$

Hence

$$\begin{aligned}
&\frac{1}{x}|\{(p_1,\dots,p_k)\in\mathcal{P}^k\colon x/2\le p_1\cdots p_k\le x\}|\\
&\ge \frac{1}{2}\sum_{\substack{(p_1,\dots,p_k)\in\mathcal{P}^k\\ x/2\le p_1\cdots p_k\le x}}\frac{1}{p_1\cdots p_k} \ge \frac{1}{2}\sum_{\substack{(a_1,\dots,a_k)\in A^k\\ N-k<a_1+\cdots+a_k\le N}}\prod_{i=1}^{k}\sum_{\substack{p\in\mathcal{P}\\ \rho^{a_i}\le p<\rho^{a_i+1}}}\frac{1}{p}\\
&\ge \frac{1}{2}\sum_{\substack{(a_1,\dots,a_k)\in A^k\\ N-k<a_1+\cdots+a_k\le N}}\frac{e^{-kj_0}}{a_1\cdots a_k} \ge \frac{1}{2}\frac{e^{-kj_0}}{(N/u)^k}\sum_{\substack{(a_1,\dots,a_k)\in A^k\\ N-k<a_1+\cdots+a_k\le N}}1\\
&\ge \frac{\alpha_v}{2N}\cdot\left(\frac{e^{-j_0}|A|}{N/u}\right)^k \ge \frac{\alpha_v}{v^2\log x}\cdot\frac{1}{e^{O_\lambda(k)}(v\log v)^k} \ge \frac{\pi_v}{v^k\log x},
\end{aligned}$$

where $\pi_v = \frac{\alpha_v}{v^2e^{O_\lambda(k)}(\log v)^v} = v^{-o_\lambda(v)}$ as $v\to\infty$. □

2.2. **Reduction of Hypothesis A to Hypothesis A*.** In this section we reduce Hypothesis A (except for the last claim concerning the case u is fixed) into a variant where $u \asymp v$. Let $u_0 = 3/\lambda$. We claim that, under the assumptions of Hypothesis A, there is some j such that

$$\sum_{\substack{a\in A\\ N/(u_0^{j+1}u)<a\le N/(u_0^j u)}} \frac{1}{a} \ge \frac{1+\lambda/3}{u_0^j u}.$$

This follows since otherwise, summing over $j \ge 0$ we get that

$$\sum_{a\in A}\frac{1}{a} < \frac{1+\lambda/3}{u(1-1/u_0)} < \frac{1+\lambda}{u}$$

which is a contradiction. Hence, Hypothesis A, except for the last claim concerning the case u is fixed (which will be proved in Section 7), follows if we prove the claim when $A \subseteq (\lambda N/u, N/u]$, i.e. if we prove the following hypothesis.

Hypothesis A*. *Fix $\lambda \in (0,1)$. There exists a constant $c = c(\lambda)$ such that the following holds. Let $N \ge u \ge 1$ be such that $N \ge (10u/\lambda)^2$. Let A be a subset of the integers in $(\lambda\frac{N}{u}, \frac{N}{u}]$ such that*

$$\sum_{a\in A}\frac{1}{a} \ge \frac{1+\lambda}{u}.$$

Then there exists an integer $k \in [u, u/\lambda]$ such that

$$|\{(a_1,\dots,a_k)\in A^k\colon N-k < a_1+\cdots+a_k \le N\}| \ge \alpha_{k,u}\frac{|A|^k}{N},$$

where $\alpha_{k,u}$ is a constant with $\alpha_{k,u} = (c/\log u)^k$ as $u\to\infty$.

3. Proving Hypothesis A*: an outline

Our main goal has become to prove Hypothesis A*, a quantitative variant of Proposition 1.2, concerning the number of solutions to $a_1+\cdots+a_k = t$ for some fixed t. In Section 3.1 we state some removal-type results in this spirit. When the number of variables k is bounded, this follows from an arithmetic regularity lemma of Green [7]. However, when k grows, the situation becomes different and we will prove the substitute Theorem 3.4 in Sections 4–6. Hypothesis A* will be deduced from these results in Sections 3.2 and 3.3.

3.1. **An arithmetic regularity lemma for popular sums.** An important tool in graph theory is the triangle removal lemma, which can be proved using Szemerédi's regularity lemma. Green [7] developed an arithmetic version of the regularity lemma, and deduced as a consequence a removal lemma in the arithmetic setting.

Theorem 3.1 (Arithmetic removal lemma)**.** *Let $k \ge 3$ be a positive integer. Let G be a finite abelian group with $|G| = N$, and let $A_1,\cdots,A_k \subseteq G$ be subsets. For any $\eta > 0$ there exists a positive constant $\delta = \delta(k,\eta) > 0$ such that the following statement holds. If the number of solutions to $a_1+\cdots+a_k = 0$ with $a_i \in A_i$ for all i is at most δN^{k-1}, then for*

each i there exists a subset $A_i' \subseteq A_i$ with $|A_i \setminus A_i'| \leq \eta N$, such that there is no solution to $a_1' + \cdots + a_k' = 0$ with $a_i' \in A_i'$ for all i.

By inspecting the proof, one notes that the construction of A_i' is translation-invariant, in the sense that if $A_i = B_i + t$ for some $t \in G$, then one can take $A_i' = B_i' + t$. Using this observation, the following extension of this arithmetic removal lemma quickly follows.

Theorem 3.2 (Removal lemma for popular sums)**.** *Let $k \geq 3$ be a positive integer. Let G be a finite abelian group with $|G| = N$, and let $A_1, \cdots, A_k \subseteq G$ be subsets. For any $\eta > 0$ there exists a positive constant $\delta = \delta(k, \eta) > 0$ such that the following statement holds. For each i there exists a subset $A_i' \subseteq A_i$ with $|A_i \setminus A_i'| \leq \eta N$, such that for each $x \in A_1' + \cdots + A_k'$, the number of solutions to $a_1 + \cdots + a_k = x$ with $a_i \in A_i$ for all i is at least δN^{k-1}.*

While no assumptions on $|A_i|$ are made in the statement, if $|A_i| = o(N)$ for some i then we may take $A_i' = \emptyset$ so that the conclusion holds trivially. Thus the removal lemma is only meaningful when $|A_i| \gg N$ for each i.

Proof. Define $A_1', \cdots, A_k'$ as in Green's proof of the arithmetic removal lemma. Let $x \in A_1' + \cdots + A_k'$, and suppose that there are at most δN^{k-1} solutions to $a_1 + \cdots + a_k = x$ with $a_i \in A_i$ for all i. Theorem 3.1 applied with the sets $A_1, \cdots, A_{k-1}, A_k - x$, along with the observation made above about translation invariance, implies that there is no solution to $a_1' + \cdots + a_k' = x$ with $a_i' \in A_i'$ for all i, which is absurd. □

In other words, Theorem 3.2 asserts that, given a positive density subset $A \subseteq G$, all k-fold sums can be made popular by removing a few elements from A, for any fixed $k \geq 3$. When $k = 2$, Theorem 3.1 is trivially true whereas Theorem 3.2 fails (see [17] for a construction of a counterexample using niveau sets).

It was later observed in [10, 13] that Theorem 3.1 can also be deduced directly from the graph removal lemma, bypassing the arithmetic regularity lemma. In this way Theorem 3.1 can also be generalized to deal with general linear equations using hypergraph removal lemmas; see [12] and references therein. On the other hand, in order to deduce Theorem 3.2 it seems necessary to use the arithmetic regularity lemma due to the translation-invariance property required. Consequently, while [7, Conjecture 9.4] is proved in [12], its extension in the spirit of Theorem 3.2 is still open.

For subsets A in an arbitrary abelian group (not necessarily finite), the following analogue of Theorem 3.2 can be deduced via a Freiman isomorphism.

Corollary 3.3. *Let $k \geq 3$ be a positive integer. Let G be an arbitrary abelian group, and let $A_1, \cdots, A_k \subseteq G$ be finite subsets. Let $A = A_1 \cup \cdots \cup A_k$ and assume that $|A + A| \leq K|A|$ for some $K \geq 1$. For any $\eta > 0$ there exists a positive constant $\delta = \delta(k, \eta, K) > 0$ such that the following statement holds. For each i there exists a subset $A_i' \subseteq A_i$ with $|A_i \setminus A_i'| \leq \eta|A|$, such that for each $x \in A_1' + \cdots + A_k'$, the number of solutions to $a_1 + \cdots + a_k = x$ with $a_i \in A_i$ for all i is at least $\delta|A|^{k-1}$.*

Proof. By Freiman's theorem, there is a Freiman isomorphism $\pi : A \to \widetilde{G}$ of order k from A to a finite abelian group $\widetilde{G}$, with image $\pi(A) = \widetilde{A}$, such that $|\widetilde{A}| = \alpha|\widetilde{G}|$ for some

$\alpha = \alpha(k, K) > 0$. Let $\widetilde{A}_i = \pi(A_i)$ for $1 \le i \le k$. By Theorem 3.2 applied to $\widetilde{A}_1, \cdots, \widetilde{A}_k$ (with η replaced by $\eta\alpha$), there are subsets $\widetilde{A}'_i \subseteq \widetilde{A}_i$ with $|\widetilde{A}_i \setminus \widetilde{A}'_i| \le \eta\alpha|\widetilde{G}| = \eta|A|$, such that for each $\widetilde{x} \in \widetilde{A}'_1 + \cdots + \widetilde{A}'_k$, the number of solutions to $\widetilde{a}_1 + \cdots + \widetilde{a}_k = \widetilde{x}$ with $\widetilde{a}_i \in \widetilde{A}_i$ is at least $\delta|\widetilde{G}|^{k-1}$ for some $\delta = \delta(k, \eta, K) > 0$.

Now let $A'_i = \pi^{-1}(\widetilde{A}'_i)$. Then $|A_i \setminus A'_i| = |\widetilde{A}_i \setminus \widetilde{A}'_i| \le \eta|A|$. For any $x \in A'_1 + \cdots + A'_k$, note that any solution to $\widetilde{a}_1 + \cdots + \widetilde{a}_k = \pi(x)$ with $\widetilde{a}_i \in \widetilde{A}_i$ gives rise to a solution to $a_1 + \cdots + a_k = x$ with $a_i = \pi^{-1}(\widetilde{a}_i)$, and moreover different solutions to the former give different solutions to the latter. The desired conclusion follows immediately. □

We expect some version of Corollary 3.3 to hold as k grows, and in this direction we will prove the following theorem.

Theorem 3.4. *For any $K \ge 1$ and $\eta > 0$, there exist positive integers $m = m(\eta, K)$ and $\ell = \ell(\eta, K)$ and a positive constant $\delta = \delta(\eta, K)$ such that the following statement holds. Let $A \subseteq G$ be a subset in a torsion-free abelian group G with $|A + A| \le K|A|$. Then there exist an element $z \in G$ with $z + \ell A \subset (m + \ell)A$, and a subset $A' \subseteq A$ with $|A'| \ge (1-\eta)|A|$, such that for any positive integer $k > \ell$ and any element $x \in kA'+z$, we have $r_{(k+m)A}(x) \ge (\delta|A|)^{k+m-1}$, where $r_{nA}(x)$ denotes the number of solutions to $a_1 + \cdots + a_n = x$ with $a_1, \cdots, a_n \in A$.*

In the following two subsections we will show how the removal lemmas can be used to prove Hypothesis A*, and the proof of Theorem 3.4 will occupy Sections 4–6. To end this subsection, we give a rough sketch of the main ideas of the proof of Theorem 3.4, motivated by arguments in [11].

We shall first deduce a filling lemma: from the removal lemma for popular sums with a fixed number of summands and work of Tao and Vu [15] we deduce that there is a bounded m and a proper progression P such that $A \subseteq P$ and mA (popularly) contains a translate of P, possibly after removing a small proportion of elements from A.

Now write C for the convex hull of A, so that $C \subseteq P$. After shrinking A a bit, any element $x \in kA$ is then a popular sum in kC. We then use an induction and the Shapley-Folkman theorem (see Lemma 6.1 below) to show that popular sums in kC are also popular in $(k-1)C + A$ (if C is slightly shrunk in an appropriate way). After doing this reduction enough times, we deduce that x is popular in $rC + (k-r)A$, for some bounded r. The final task of finding popular representations of elements in rC can be done through the filling lemma described above since rmA popularly contains a translate of rC. In practice we need to be very careful to always guarantee popularity at each stage.

3.2. **Proof of Hypothesis A* for bounded u.** We divide into two cases depending on whether $u = O_\lambda(1)$ or not. First suppose that $u = O_\lambda(1)$. Since Theorem 3.2 is only applicable for $k \ge 3$, we need to do some initial preparations to handle the case where we would have $k = 2$. Write

$$A' = \{a \in A \colon A \cap (N - 2 - a, N - a] = \emptyset\}.$$

If $|A\backslash A'| \geq \left(\frac{\lambda}{2u}\right)^2 |A|$, the claim follows with $k=2$, so we can assume that $|A\backslash A'| < \left(\frac{\lambda}{2u}\right)^2 |A|$. Then, by assumptions on set A, we have $A' \subseteq (\lambda N/u, N/u]$ and

$$\sum_{a\in A'} \frac{1}{a} \geq \sum_{a\in A} \frac{1}{a} - \frac{u}{\lambda N}|A\setminus A'| \geq \frac{1+3\lambda/4}{u},$$

so that A' has density at least λ/u on the interval $[1, N/u]$. By Corollary 3.3 we may find a subset $A'' \subseteq A'$ with $|A'\setminus A''| \leq (\frac{\lambda}{2u})^2|A'|$, such that for any $3 \leq k \leq u/\lambda$ and any $x \in kA''$, we have

$$\big|\{(a_1,\cdots,a_k)\in A^k : a_1+\cdots+a_k = x\}\big| \geq \delta|A|^{k-1}, \tag{3.1}$$

for some $\delta = \delta(\lambda,u) > 0$. Since

$$\sum_{a\in A''} \frac{1}{a} \geq \sum_{a\in A'} \frac{1}{a} - \frac{u}{\lambda N}|A'\setminus A''| \geq \frac{1+\lambda/2}{u} > \frac{1}{u} + \frac{1}{\sqrt{N}-1}$$

by the lower bound for N, Bleichenbacher's theorem (Proposition 1.2) implies that there exists a positive integer k and $a_1',\cdots,a_k' \in A''$ such that $N-k < a_1'+\cdots+a_k' \leq N$. Note that we necessarily have $k \in [u, u/\lambda]$, and by the choice of A', we must have $k \neq 2$. If $k=1$, the claim follows immediately. If $k \geq 3$, then (3.1) applied to $a_1'+\cdots+a_k'$ gives that

$$\big|\{(a_1,\cdots,a_k)\in A^k : N-k < a_1+\cdots+a_k \leq N\}\big| \geq \delta|A|^{k-1},$$

as desired.

3.3. **Proof of Hypothesis A* for large u, assuming Theorem 3.4.** For the rest of the proof assume that $u \geq U$ for some sufficiently large U depending on λ. Let us now prove by induction on $j \geq 0$ that Hypothesis A* holds when $2^jU \leq u \leq 2^{j+1}U$. Let $A \subseteq (\lambda N/u, N/u]$ be a subset with $\sum_{a\in A}\frac{1}{a} > (1+\lambda)/u$. We wish to find a positive integer $k \in [u, u/\lambda]$ such that

$$\big|\{(a_1,\ldots,a_k)\in A^k : N-k < a_1+\cdots+a_k \leq N\}\big| \geq \left(\frac{c}{\log u}\right)^k \cdot \frac{|A|^k}{N} \tag{3.2}$$

for some $c = c(\lambda) > 0$. In case $j=0$, this follows from the work on case $u = O_\lambda(1)$ once c is small enough.

Assume now that $2^jU \leq u \leq 2^{j+1}U$ for some $j \geq 1$. We shall study popular doubling in A, but first we need to find an appropriate notion of popularity. Write $r_0 = 0$ and

$$r_i = 2^{i-10}\lambda^4\frac{|A|^2}{|2A|}$$

for $i \geq 1$. For $i \geq 0$, let

$$B_i = \{n \in 2A : r_i < r_{2A}(n) \leq r_{i+1}\}.$$

Note that since $\frac{|A|}{|2A|} \geq \frac{\lambda N/u^2}{2N/u} = \frac{\lambda}{2u}$ and $r_{2A}(n) \leq |A|$ for all $n \in 2A$, the set B_i is empty for $i \geq I = 2\log u - 1$. Furthermore

$$\sum_{0<i<I} |B_i| r_i = \frac{1}{2} \sum_{0\leq i<I} |B_i| r_{i+1} - \frac{1}{2}|B_0| r_1 \geq \frac{1}{2} \sum_{n\in 2A} r_{2A}(n) - \frac{1}{2}|2A| \cdot \frac{\lambda^4 |A|^2}{2^9 |2A|} \geq \frac{|A|^2}{4},$$

so that there exists a smallest positive index i_0 such that

$$|B_{i_0}| r_{i_0} \geq \frac{\lambda^4}{512 \log u} |A|^2.$$

We choose

$$E = \{(a_1, a_2) \in A \times A \colon r_{2A}(a_1 + a_2) > r_{i_0}\}.$$

Now

$$|(A \times A) \setminus E| \leq \sum_{i<i_0} |B_i| r_{i+1} \leq |B_0| r_1 + 2\log u \cdot 2 \cdot \frac{\lambda^4}{512 \log u} |A|^2 \leq \frac{\lambda^4}{64} |A|^2. \tag{3.3}$$

Write $D = \cup_{i\geq i_0} B_i = A \overset{E}{+} A \subseteq (2\lambda N/u, 2N/u]$.

Let us first consider the case that $|D| > 8N/u^2$. Then

$$\sum_{\substack{d\in D \\ \lambda \frac{N}{u/2} < d \leq \frac{N}{u/2}}} \frac{1}{d} \geq |D| \frac{u}{2N} > \frac{1+\lambda}{u/2},$$

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

$$|\{(d_1, \dots, d_{k/2}) \in D^{k/2} : N - k/2 < d_1 + \cdots + d_{k/2} \leq N\}| \geq \left(\frac{c}{\log u}\right)^{k/2} \frac{|D|^{k/2}}{N}.$$

Hence, by the definitions of D and i_0, we have

$$\begin{aligned}
&|\{(a_1, \dots, a_k) \in A^k : N - k < a_1 + \cdots + a_k \leq N\}| \\
&\geq |\{(d_1, \dots, d_{k/2}) \in D^{k/2} : N - k/2 < d_1 + \cdots + d_{k/2} \leq N\}| \cdot r_{i_0}^{k/2} \\
&\geq \left(\frac{c}{\log u}\right)^{k/2} \frac{|B_{i_0}|^{k/2}}{N} \cdot r_{i_0}^{k/2} \geq \left(\frac{c}{\log u}\right)^{k/2} \cdot \frac{1}{N} \left(\frac{\lambda^4}{512 \log u} |A|^2\right)^{k/2} \geq \left(\frac{c}{\log u}\right)^{k} \frac{|A|^k}{N},
\end{aligned}$$

provided that $c \leq \frac{\lambda^4}{512}$.

Let us now consider the case that $|D| \leq 8N/u^2$. We need the following lemma.

Lemma 3.5. *Let $(G, +)$ be an abelian group and let $\delta > 0$. If $E \subseteq A \times A$ satisfies*

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

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

$$|A'| \geq (1 - 2\delta)|A| \quad \textit{and} \quad |A' + A'| \leq \frac{K^3}{1 - 6\delta} |A|.$$

Proof. This is a variant of the Balog-Szemerédi-Gowers theorem (see [16, Theorem 2.29]) which can be proved by incorporating the hint for [16, Exercise 2.5.4] to the proof of the Balog-Szemerédi-Gowers theorem in [16, Section 6.4]. See also [6, Lemma 5.1] for a proof of a variant for $A - A$. □

Since $|A| \geq \lambda N/u^2$ and $|D| \leq 8N/u^2$, we have $|A \overset{E}{+} A| \leq (8/\lambda)|A|$. Recall also (3.3). Thus Lemma 3.5 implies that there is a subset $B \subseteq A$ such that

$$|B| \geq (1 - \lambda^2/4)|A| \quad \text{and} \quad |B + B| \leq (20/\lambda)^3|B|.$$

Applying Theorem 3.4 to B with $K = (20/\lambda)^3$ and $\eta = \lambda^2/4$, we obtain an element $z \in \mathbb{Z}$ with $|z| = O_\lambda(N/u)$ and a subset $B' \subseteq B$ with $|B'| \geq (1-\eta)|B|$ satisfying the property that whenever $x \in kB' + z$ for any positive integer k large enough depending on λ, there exists $n \in [k, k + O_\lambda(1)]$ such that $r_{nB}(x) \geq (c|B|)^{n-1}$.

Since $B' \subseteq (\lambda N/u, N/u]$, we have

$$\begin{aligned}\sum_{b\in B'} \frac{1}{b} &\geq \sum_{a\in A} \frac{1}{a} - \frac{u}{\lambda N}\left(|A \setminus B| + |B \setminus B'|\right) \geq \sum_{a\in A} \frac{1}{a} - \frac{u}{\lambda N} \cdot \frac{\lambda^2}{2}|A| \\ &\geq \sum_{a\in A} \frac{1}{a} - \frac{u}{\lambda N} \cdot \frac{\lambda^2}{2} \cdot \frac{N}{u} \sum_{a\in A} \frac{1}{a} \geq \left(1 - \frac{\lambda}{2}\right) \sum_{a\in A} \frac{1}{a} > \frac{1 + \lambda/3}{u}.\end{aligned}$$

Recalling that $|z| = O_\lambda(N/u)$, and writing $N' = N - z$ and $u' = u(N - z)/N$, we have

$$u' = u - \frac{uz}{N} \geq u - O_\lambda(1) \geq \left(1 - \frac{\lambda}{10}\right) u$$

provided that U is large enough. Hence

$$\sum_{\substack{b\in B' \\ b\leq N'/u'}} \frac{1}{b} > \frac{1 + \lambda/3}{u} \geq \frac{1 + \lambda/6}{u'} > \frac{1}{u'} + \frac{1}{\sqrt{N'} - 1}.$$

Hence, by discrete Bleichenbacher's theorem (Proposition 1.2), we find k and $b_1, \cdots, b_k \in B'$ such that

$$N - z - k < b_1 + \cdots + b_k \leq N - z.$$

Write $x := b_1 + \cdots + b_k + z \in (N - k, N]$. Now $x \in kB' + z$ and hence there exists $\ell \in [k, k + O(1)]$ such that $r_{\ell B}(x) \geq (c|B|)^{\ell-1}$. Therefore,

$$|\{(a_1, \ldots, a_\ell) \in A^\ell : N - \ell < a_1 + \cdots + a_\ell \leq N\}| \geq r_{\ell A}(x) \geq r_{\ell B}(x) \geq (c|B|)^{\ell-1} \geq (c|A|/2)^{\ell-1}.$$

This clearly implies (3.2), completing the proof of Hypothesis A*.

4. The filling argument

In this section we carry out the first step in proving Theorem 3.4, that of locating a proper progression P containing A such that mA fills a translate of P for some bounded m.

Lemma 4.1 (Filling lemma)**.** *For any $K \geq 1$, there exists a positive integer $m = m(K)$ such that the following statement holds. Let G be a torsion-free abelian group, and let $A \subseteq G$ be a finite subset with $|A + A| \leq K|A|$. Then there is a proper progression Q of rank $O_K(1)$ with size $|Q| = O_K(|A|)$, such that $A \subseteq Q$ and $g + Q \subseteq mA$ for some $g \in G$.*

Proof. By Freiman's theorem there is a proper progression P of rank $d-1 = O_K(1)$ containing A, such that $|A| = \alpha|P|$ for some $\alpha \gg_K 1$. Thus for any positive integer ℓ, we have

$$|\ell A| \leq |\ell P| \leq \ell^{d-1}|P| \leq \alpha^{-1}\ell^{d-1}|A|.$$

The hypotheses in [15, Theorem 1.21] are then satisfied for ℓ large enough depending on K. Hence there is a proper progression Q' of rank $d' \leq d-1$, such that

$$g + Q' \subseteq \ell A \subseteq g' + kQ',$$

for some constant $k = k(d)$ and some $g, g' \in G$. Hence the iterated sumset $k\ell A$ contains a translate of kQ', which in turn contains a translate of A. Finally, by [15, Corollary 1.11] we may find a proper progression Q containing kQ', such that Q is contained in jkQ' for some $j = j(d)$. Thus for $m' = jk\ell$, the iterated sumset $m'A$ contains a translate of Q, which in turn contains a translate of A. Since $d = O_K(1)$ and $\ell, j, k = O_{d,K}(1)$, we have $m' \leq m$ for some integer m depending only on K. Clearly the claim holds for this m. □

Combining the previous lemma with Corollary 3.3 we obtain the following filling lemma for popular sums.

Lemma 4.2 (Filling lemma, popularity version)**.** *For any $K \geq 1$ and $\eta \in (0, 1/2)$, there exist a large positive integer $m = m(K)$ and a small positive constant $\delta = \delta(K, \eta)$ such that the following statement holds. Let $A \subseteq G$ be a subset in a torsion-free abelian group G with $|A + A| \leq K|A|$. Then there exist a proper progression P of rank $O_K(1)$ with size $|P| = O_K(|A|)$ and a subset $A' \subseteq A$ with $|A'| \geq (1-\eta)|A|$ and $A' \subseteq P$, such that mA' popularly contains some translate of P. That is, for some $g \in G$ we have $r_{mA}(x) \geq \delta|A|^{m-1}$ for any $x \in g + P$.*

Proof. Let $m = m(2K)$ be the constant from Lemma 4.1. By Corollary 3.3, there is a subset $A' \subseteq A$ with $|A'| \geq (1-\eta)|A|$ such that, for each $x \in mA'$ we have $r_{mA}(x) \geq \delta|A|^{m-1}$ for some $\delta = \delta(K, \eta) > 0$. Since $|A' + A'| \leq 2K|A'|$, Lemma 4.1 implies that there is a proper progression P of rank $O_K(1)$ with size $|P| = O_K(|A|)$, with the properties that $A' \subseteq P$ and $g + P \subseteq mA'$ for some $g \in G$. For each $x \in g + P$ we then have $x \in mA'$, and hence $r_{mA}(x) \geq \delta|A|^{m-1}$, as desired. □

Remark 4.3. It is a standard result in additive combinatorics that $3A$ contains a large progression P of small rank. Here we require the extra condition that A is (essentially) contained in a translate of P. A similar result is proved in [14, Lemma 2.5], but it is not enough for us to deduce Theorem 3.4.

When $G = \mathbb{Z}$ and A lies densely inside an interval, Lemma 4.2 can also be proved by a Fourier analytic argument (see [2, Lemma 8.5]).

Via the filling lemma (Lemma 4.2), Theorem 3.4 reduces to the following proposition.

Proposition 4.4. *For any $\alpha, \eta > 0$ and $d \in \mathbb{N}$, there exist a positive integer $\ell = \ell(\alpha, d, \eta)$ and a small positive constant $\delta = \delta(\alpha, d, \eta)$ such that the following statement holds. Let $P = ([-N_1, N_1] \times \cdots \times [-N_d, N_d]) \cap \mathbb{Z}^d$ be a box for some positive integers $N_1, \cdots, N_d$, and let $A \subseteq P$ be a subset with $|A| \geq \alpha |P|$. Then there exists a subset $A' \subseteq A$ with $|A'| \geq (1-\eta)|A|$ with the property that, for any positive integer $k > \ell$ and any element $x \in kA'$, there are at least $(\delta|A|)^{k-\ell}$ ways to write $x = y + a_1 + \cdots + a_{k-\ell}$ with $y \in \ell P$ and $a_1, \cdots, a_{k-\ell} \in A$.*

Proof of Theorem 3.4 assuming Proposition 4.4. First note that if Theorem 3.4 holds for some subset A, then it also holds for any translate of A. From the filling lemma (Lemma 4.2), we obtain a subset $A_1 \subseteq A$ with $|A_1| \geq (1 - \eta/4)|A|$ and a proper progression P of rank $d = O_K(1)$ with size $|P| = O_K(|A|)$, such that $A_1 \subset P$ and nA_1 popularly contains a translate of P for some $n = n(\eta, K)$. By translating A appropriately, we may further assume that P is symmetric.

Let $\pi : \mathbb{Z}^d \to G \supset P$ be the Freiman homomorphism mapping the standard basis vectors in $\mathbb{Z}^d$ to the generators of P. Since P is proper, the map gives a bijection between the box $\pi^{-1}(P)$ and P. Write $\widetilde{P} = \pi^{-1}(P)$ and $\widetilde{A}_1 = \pi^{-1}(A_1)$. Applying Proposition 4.4 to the box $\widetilde{P}$ and the subset $\widetilde{A}_1$, we obtain a positive integer $\ell = \ell(\eta, K)$ and a subset $\widetilde{A}' \subseteq \widetilde{A}_1$ with $|\widetilde{A}'| \geq (1 - \eta/2)|\widetilde{A}_1|$, such that for any $k > \ell$ and $\widetilde{x} \in k\widetilde{A}'$, the number of ways to write

$$\widetilde{x} = \widetilde{y} + \widetilde{a}_1 + \cdots + \widetilde{a}_{k-\ell} \tag{4.1}$$

with $\widetilde{y} \in \ell\widetilde{P}$ and $\widetilde{a}_1, \cdots, \widetilde{a}_{k-\ell} \in \widetilde{A}_1$ is at least $(\delta|A|)^{k-\ell}$, for some positive constant $\delta = \delta(\eta, K) > 0$.

Let $A' = \pi(\widetilde{A}')$. Clearly $|A_1 \setminus A'| = |\widetilde{A}_1 \setminus \widetilde{A}'| \leq (\eta/2)|A|$, and thus $|A \setminus A'| \leq \eta|A|$. Moreover, for any $k > \ell$ and $x \in kA'$, we may find $\widetilde{x} \in k\widetilde{A}'$ such that $\pi(\widetilde{x}) = x$. Via the map π, each representation for $\widetilde{x}$ of the form (4.1) gives rise to a representation x of the form

$$x = y + a_1 + \cdots + a_{k-\ell} \tag{4.2}$$

with $y \in \ell P$ and $a_1, \cdots, a_{k-\ell} \in A_1$. Hence there are at least $(\delta|A|)^{k-\ell}$ such representations for x.

Recall from the output of the filling lemma that nA_1 popularly contains a translate of P. It then easily follows (for example from [6, Lemma 5.3]) that $2\ell n A_1$ popularly contains a translate $z + \ell P$ for some $z \in G$. Thus each representation for $x \in kA'$ of the form (4.2) gives rise to at least $(\delta|A|)^{2\ell n - 1}$ ways to write $z + x$ as a sum of $2\ell n + (k - \ell)$ elements of A, since

$$r_{2\ell n A}(z + y) \geq (\delta|A|)^{2\ell n - 1}$$

if $\delta > 0$ is small enough. We conclude that for any $k > \ell$ and $x \in kA'$, we have

$$r_{(2\ell n + k - \ell)A}(z + x) \geq (\delta|A|)^{k-\ell}(\delta|A|)^{2\ell n - 1} = (\delta|A|)^{2\ell n + k - \ell - 1}.$$

This shows that Theorem 3.4 holds with this choice of ℓ and with $m = (2n-1)\ell$. □

We will prove Proposition 4.4 in Section 6 using geometrical ideas, after establishing some preliminary lemmas in Section 5.

5. Auxiliary results about convex bodies

Notations. In this section and the next, we use normal letters such as A, C, P to denote subsets of $\mathbb{Z}^d$, and boldface letters such as $\mathbf{C}, \mathbf{P}$ to denote convex bodies in $\mathbb{R}^d$. For $t > 0$, we use $t\mathbf{C}, t\mathbf{P}$ to denote dilations of convex bodies in the usual manner.

The aim of this section is to prove two intuitive properties of convex hulls of positive density subsets A of large boxes P in $\mathbb{Z}^d$. The first one, Lemma 5.3 says that, for some constant $\varepsilon > 0$ (depending only on the density of A and on the dimension), the convex hull of A contains a translate of a small dilate of the convex hull of the box. The second one, Lemma 5.4 states that most lattice points in the convex hull of A are away from the boundary of the convex hull.

Before stating and proving these, we state two auxiliary results which will be used in the proofs of the two lemmas.

Lemma 5.1. *For any $\alpha > 0$ and positive integer d, there exist constants $N_0 = N_0(\alpha, d)$ and $c = c(\alpha, d)$ such that the following statement holds. Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d \geq N_0$, and let $P = \mathbf{P} \cap \mathbb{Z}^d$. Let $A \subseteq P$ be a subset with $|A| \geq \alpha|P|$, and let $\mathbf{C} \subseteq \mathbb{R}^d$ be the convex hull of A. Then vol$(\mathbf{C}) \geq c|P|$.*

Proof. We can clearly assume that $\alpha < 1/100$. Write $M_j = \lceil \alpha^3 N_j \rceil$, and let us split the box P into $M_1 \cdots M_d$ fibers

$$P_{i_1,\ldots,i_d} = \{(x_1, \ldots, x_d) \in P \colon x_j \equiv i_j \pmod{M_j}\} \quad \text{with } 0 \leq i_j < M_j \text{ for each } 1 \leq j \leq d,$$

and write $A_{i_1,\ldots,i_d} = A \cap P_{i_1,\ldots i_d}$. For some $i_1, \ldots, i_d$ we must have $|A_{i_1,\ldots,i_d}| \geq \alpha |P_{i_1,\ldots i_d}|$. By the natural bijection

$$\rho \colon P_{i_1,\ldots,i_d} \to \prod_{j=1}^{d} \left[\left\lceil \frac{-N_j - i_j}{M_j} \right\rceil, \left\lfloor \frac{N_j - i_j}{M_j} \right\rfloor \right] \cap \mathbb{Z}^d =: P'$$

$$(x_1, \ldots, x_d) \to \left(\frac{x_1 - i_1}{M_1}, \ldots, \frac{x_d - i_d}{M_d} \right)$$

we can map the corresponding fibre into the box P' which has bounded sidelengths. Write B for the image of $A_{i_1,\ldots,i_d}$, and write $L_1, \cdots, L_d$ for the side lengths of P'. By our choice of M_j, we have $L_j > \alpha^{-1}$ for each j, once N_0 is large enough in terms of α. It follows that

$$|B| \geq \alpha |P'| > \max_j \prod_{i=1, i \neq j}^{d} L_i.$$

Thus B cannot be contained in any $d-1$-dimensional hyperplane, and so B contains $d+1$ points generating a non-trivial simplex Δ, whose volume is at least $1/d!$ since its vertices are lattice points. Hence $\rho^{-1}(\Delta)$ has volume at least

$$\frac{1}{d!} \prod_{i=1}^{d} M_i \gg_{\alpha,d} |P|,$$

and the claim follows since, by convexity, $\rho^{-1}(\Delta) \subseteq \mathbf{C}$. □

Theorem 5.2 (John). *Let $\mathbf{C} \subseteq \mathbb{R}^d$ be a convex body. There exists an invertible linear transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ and a point $x_0 \in \mathbf{C}$ such that*

$$B_d \subseteq T(\mathbf{C} - x_0) \subseteq dB_d,$$

where B_d is the unit ball $\{(x_1, \cdots, x_d) : x_1^2 + \cdots + x_d^2 \leq 1\}$.

Proof. See [16, Theorem 3.13]. □

Lemma 5.3 (Large boxes inside convex sets). *For any $\alpha > 0$ and positive integer d, there exist $N_0 = N_0(\alpha, d) > 0$ and $\beta = \beta(\alpha, d) > 0$ such that the following statement holds. Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d \geq N_0$, and let $P = \mathbf{P} \cap \mathbb{Z}^d$. Let $A \subseteq P$ be a subset with $|A| \geq \alpha|P|$. Let $\mathbf{C} \subseteq \mathbb{R}^d$ be the convex hull of A. Then $x_0 + \beta\mathbf{P} \subseteq \mathbf{C}$ for some $x_0 \in \mathbb{Z}^d$.*

Proof. Since $N_1, \cdots, N_d$ are large enough, Lemma 5.1 implies that

$$\mathrm{vol}(\mathbf{C}) \geq cN_1N_2 \cdots N_d$$

for some positive constant $c = c(\alpha, d)$. Now apply John's theorem to $\mathbf{C}$ to obtain an invertible linear transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ and a point $x_0 \in \mathbf{C}$ such that

$$B_d \subseteq T(\mathbf{C} - x_0) \subseteq dB_d.$$

In particular, we have

$$T^{-1}(B_d) \subseteq \mathbf{C} - x_0 \subseteq 2\mathbf{P},$$

and thus $T^{-1}(e_i) \in 2\mathbf{P}$ for the standard basis vectors $e_1, \cdots, e_d$, so that the (i, j)-entry of T^{-1} is $O(N_i)$ for each $1 \leq i, j \leq d$. Moreover, we have

$$(\det T)(\mathrm{vol}(\mathbf{C})) = \mathrm{vol}(T(\mathbf{C} - x_0)) \leq \mathrm{vol}(dB_d),$$

so that $\det T \ll |P|^{-1}$ and $\det T^{-1} \gg |P|$.

Now consider the (i, j)-entry of T. The bounds on the matrix entries of T^{-1} imply that the determinant of the (j, i)-minor of T^{-1} is $O(|P|/N_j)$. It follows that the (i, j)-entry of T is bounded in absolute by

$$O\left(\frac{1}{\det T^{-1}} \cdot \frac{|P|}{N_j}\right) = O\left(\frac{1}{N_j}\right).$$

It follows that $\|T(x)\|_\infty \ll 1$ for any $x \in \mathbf{P}$. Hence $T(\mathbf{P}) \subseteq \beta^{-1}B_d$ for $\beta > 0$ small enough. This implies that $\beta\mathbf{P} \subseteq T^{-1}(B_d)$ and thus $x_0 + \beta\mathbf{P} \subseteq \mathbf{C}$ as desired.

Finally, to ensure that $x_0 \in \mathbb{Z}^d$, we may replace β by $\beta/2$ and note that $x_0 + (\beta/2)\mathbf{P}$ contains a lattice point for any $x_0 \in \mathbb{R}^d$, once N_0 is large enough. □

Lemma 5.4 (Lattice points near the boundary). *For any $\beta, \eta \in (0, 1)$ and positive integer d, there exist $N_0 = N_0(d, \beta, \eta) > 0$ and $\gamma = \gamma(d, \eta) > 0$ such that the following statement holds. Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d \geq N_0$. Let $\mathbf{C} \subseteq \mathbf{P}$ be a convex body, and assume that $\beta\mathbf{P} \subseteq \mathbf{C}$. Let $\mathbf{C}' = (1-\gamma)\mathbf{C}$. Then $|(\mathbf{C} \setminus \mathbf{C}') \cap \mathbb{Z}^d| \leq \eta \cdot vol(\mathbf{C})$.*

Proof. Choose $\gamma = \gamma(d, \eta)$ so small that $(1+\gamma)^d - (1-2\gamma)^d < \eta$, and let $X = (\mathbf{C} \setminus \mathbf{C}') \cap \mathbb{Z}^d$. Let $\mathbf{B} \subseteq \mathbb{R}^d$ be the unit box $[-1/2, 1/2]^d$. Note that

$$\bigcup_{x \in X} (x + \mathbf{B}) \subseteq (\mathbf{C} \setminus \mathbf{C}') + \mathbf{B}.$$

Since the union above is a disjoint union, we have

$$|X| \leq \mathrm{vol}((\mathbf{C} \setminus \mathbf{C}') + \mathbf{B}).$$

The volume above is at most

$$\mathrm{vol}(\mathbf{C} + \mathbf{B}) - \mathrm{vol}(\{x \in \mathbf{C}' : x + \mathbf{B} \subseteq \mathbf{C}'\}).$$

If N_0 is large enough depending on β and γ, then $\mathbf{B} \subseteq \gamma\beta\mathbf{P}$ and thus $\mathbf{B} \subseteq \gamma\mathbf{C}$. It follows that

$$|X| \leq \mathrm{vol}((1+\gamma)\mathbf{C}) - \mathrm{vol}((1-2\gamma)\mathbf{C}) = \left[(1+\gamma)^d - (1-2\gamma)^d\right] \mathrm{vol}(\mathbf{C}) \leq \eta \cdot \mathrm{vol}(\mathbf{C}),$$

by our choice of γ. □

6. Proof of Proposition 4.4

Recall the notations from the beginning of Section 5. The following result of Shapley and Folkman resembles a simpler and non-popular version of what we wish to prove.

Lemma 6.1 (Shapley-Folkman)**.** *Let d be a positive integer, let $B \subseteq \mathbb{R}^d$, and let $\mathbf{C}$ be the convex hull of B. For any integer $k > d$, one has*

$$k\mathbf{C} = d\mathbf{C} + (k-d)B.$$

Proof. See e.g. [4, Appendix 1] or [3, Corollary on page 435]. □

In order to extend the previous lemma to popular representations, we need to introduce some notation.

Definition 6.2 (ε-regular subsets)**.** Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d$, and let $P = \mathbf{P} \cap \mathbb{Z}^d$ and $P_\varepsilon = \varepsilon\mathbf{P} \cap \mathbb{Z}^d$ for $\varepsilon > 0$. A subset $A \subseteq P$ is called ε-regular, if for each $a \in A$ the small box $a + P_\varepsilon$ centered around a contains at least $\varepsilon|P_\varepsilon|$ elements in A.

Lemma 6.3 (Regularization)**.** *Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d$, and let $P = \mathbf{P} \cap \mathbb{Z}^d$. Let $A \subseteq P$ be a subset with $|A| = \alpha|P|$ for some $\alpha > 0$. Let $\eta > 0$ be real. For any $\varepsilon \in (0, \alpha\eta 100^{-d})$, there is an ε-regular subset $A' \subseteq A$ with $|A'| \geq (1-\eta)|A|$.*

Proof. Let $P_{\varepsilon/2} = (\varepsilon/2)\mathbf{P} \cap \mathbb{Z}^d$, so that the side lengths of $P_{\varepsilon/2}$ are precisely $2\lfloor \varepsilon N_1/2 \rfloor + 1, \cdots, 2\lfloor \varepsilon N_d/2 \rfloor + 1$. Cover P by at most

$$\prod_{i=1}^{d} \left\lceil \frac{2N_i + 1}{2\lfloor \varepsilon N_i/2 \rfloor + 1} \right\rceil \leq 10^d \prod_{i=1}^{d} \min(N_i, \varepsilon^{-1})$$

translates of $P_{\varepsilon/2}$. Define A' by removing from A those translates containing at most $\varepsilon|P_\varepsilon|$ elements of A. Then A' is ε-regular by construction, since any $a \in A'$ lies in a translate of $P_{\varepsilon/2}$ containing at least $\varepsilon|P_\varepsilon|$ elements of A', but this translate of $P_{\varepsilon/2}$ is contained in $a + P_\varepsilon$.

Moreover, since

$$\varepsilon|P_\varepsilon| \le 10^d \varepsilon \prod_{i=1}^{d} \max(\varepsilon N_i, 1),$$

the number of elements in $A \setminus A'$ is at most

$$100^d \varepsilon \prod_{i=1}^{d} \min(N_i, \varepsilon^{-1}) \max(\varepsilon N_i, 1) \le 100^d \varepsilon |P| \le \eta |A|,$$

as desired. □

Proposition 6.4 (popular Shapley-Folkman)**.** *Let $\mathbf{P} = [-N_1, N_1] \times \cdots \times [-N_d, N_d]$ be a box in $\mathbb{R}^d$ for some positive integers $N_1, \cdots, N_d$, and let $P = \mathbf{P} \cap \mathbb{Z}^d$. Let $A \subseteq P$ be a subset. Let $\mathbf{C} \subseteq \mathbb{R}^d$ be the convex hull of A, and assume that $x_0 + \beta\mathbf{P} \subseteq \mathbf{C}$ for some $x_0 \in \mathbb{Z}^d$ and $\beta > 0$. Let $\mathbf{C}' = (1-\gamma)\mathbf{C} + \gamma x_0$ for some $\gamma \in (0, 1/(d+2))$. If A is ε-regular for some $\varepsilon \le \beta\gamma$, then for each positive integer $k > d$ and any element $x \in (k+1)\mathbf{C}'$, there are at least $\delta|P|$ ways to write $x = y + a$ for some $y \in k\mathbf{C}'$ and $a \in A$, where $\delta = \delta(d, \varepsilon) > 0$ is a positive constant.*

Proof. By translation we may assume that $x_0 = 0$, so that $\beta\mathbf{P} \subseteq \mathbf{C}$ and $\mathbf{C}' = (1-\gamma)\mathbf{C}$. Let $x \in (k+1)\mathbf{C}'$ be for some $k > d$. Write $x = (k+1)(1-\gamma)z$ for some $z \in \mathbf{C}$. Note that

$$x = (k - d - \gamma(k+1))z + (d+1)z = (k - d - \gamma(k+1))z + dw + a$$

for some $a \in A$ and $w \in \mathbf{C}$, by the Shapley-Folkman theorem (Lemma 6.1). We thus get a solution $x = y + a$ with

$$y = (k - d - \gamma(k+1))z + dw.$$

We claim that

$$y + P_\varepsilon \subset k\mathbf{C}'. \tag{6.1}$$

If this claim is true, then any $t \in P_\varepsilon$ with $a - t \in A$ gives rise to a representation

$$x = (y + t) + (a - t)$$

with $y + t \in k\mathbf{C}'$. By the ε-regularity of A, there are at least $\varepsilon|P_\varepsilon|$ such elements t, leading to at least

$$\varepsilon|P_\varepsilon| \ge \varepsilon^{d+1} N_1 \cdots N_d \ge \delta|P|$$

representations for some constant $\delta > 0$ depending on d and ε, as desired.

To prove (6.1), take any $t \in \varepsilon\mathbf{P}$. Since $\beta\mathbf{P} \subseteq \mathbf{C}$ and $\varepsilon \le \beta\gamma$, we have $\varepsilon\mathbf{P} \subseteq \gamma\mathbf{C}$, and thus $t \in \gamma\mathbf{C}$. Hence

$$y + t \in (k - d - \gamma(k+1))\mathbf{C} + d\mathbf{C} + \gamma\mathbf{C} \subseteq k(1-\gamma)\mathbf{C} = k\mathbf{C}',$$

as desired. □

Now we are finally ready to prove Proposition 4.4.

Proof of Proposition 4.4. Let $\beta = \min_{d' \le d} \beta(\alpha\eta/8, d') > 0$, where $\beta(\cdot,\cdot)$ is from Lemma 5.3, and let $\gamma = \min\{1/(d+2), \min_{d' \le d} \gamma(d', \eta/4)\} > 0$, where $\gamma(\cdot,\cdot)$ is from Lemma 5.4. Let N_0 be the maximum of the constants $\max_{d' \le d} N_0(\alpha\eta/8, d')$ from Lemma 5.3 and $\max_{d' \le d} N_0(d', \beta, \eta/4)$ from Lemma 5.4.

Without loss of generality we may assume that $N_1, \cdots, N_{d'} \ge N_0$ and $N_{d'+1}, \cdots, N_d < N_0$, for some $0 \le d' \le d$. We may also assume that $d' > 0$ since otherwise $|A|$ is bounded and the conclusion is trivial. Let $\mathbf{P}' = [-N_1, N_1] \times \cdot \times [-N_{d'}, N_{d'}]$, all of whose side lengths are at least $2N_0 + 1$, and let $P' = \mathbf{P}' \cap \mathbb{Z}^{d'}$. We can partition P into $J \le (2N_0+1)^{d-d'}$ smaller boxes $P_1, \cdots, P_J$, with $P_j = P' \times \{t_j\}$ for some $t_j \in \mathbb{Z}^{d-d'}$. For each $1 \le j \le J$, let $A_j \subset P'$ be the set of $a \in P'$ with $(a, t_j) \in A$. Let $\mathcal{J}$ be the set of indices j with $|A_j| \ge (\eta\alpha/4)|P'|$.

Let $\varepsilon \in (0, \beta\gamma)$ be small enough depending on α, d, η. For each $j \in \mathcal{J}$, Lemma 6.3 applied to $A_j \subseteq P'$ implies that there is an ε-regular subset $B_j \subseteq A_j$ with $|A_j \setminus B_j| \le (\eta/4)|A_j|$. Let $\mathbf{C}_j \subseteq \mathbb{R}^{d'}$ be the convex hull of B_j. Lemma 5.3 applied to B_j and $\mathbf{C}_j$ implies that $x_j + \beta\mathbf{P}' \subset \mathbf{C}_j$ for some $x_j \in \mathbb{Z}^{d'}$.

For each $j \in \mathcal{J}$, we may thus apply Proposition 6.4 to B_j and $\mathbf{C}_j$ to define $\mathbf{C}_j'$ and conclude that for any $k > d$ and $x \in (k+1)\mathbf{C}_j'$, there are at least $\delta'|P'|$ ways to write $x = y + b$ for some $y \in k\mathbf{C}_j'$ and $b \in B_j$, where $\delta' = \delta'(d, \varepsilon) > 0$ is a constant. This number of representations is at least $\delta|A|$ for $\delta = \delta'/J$.

Now for $j \in \mathcal{J}$ we write $A_j' = (B_j \cap \mathbf{C}_j') \times \{t_j\}$ and let $A' = \cup_{j \in \mathcal{J}} A_j'$. To finish the proof, we show that the conclusion of Proposition 4.4 holds with this choice of A' and $\ell = d(2N_0+1)^d \ge dJ$. Indeed, let $k > \ell$ and $x \in kA'$ be arbitrary. Assume that

$$x = \sum_{j \in \mathcal{J}} (x_j, k_j t_j),$$

where $x_j \in k_j(B_j \cap \mathbf{C}_j')$ and $\sum k_j = k$. We may choose $\ell_j \in [\min\{d, k_j\}, k_j]$ for $j \in \mathcal{J}$ such that $\sum \ell_j = \ell$. For those $j \in \mathcal{J}$ with $k_j > d$, by iterating the output of Proposition 6.4 we see that the number of ways to write $x_j = y_j + b_{j,1} + \cdots + b_{j,k_j-\ell_j}$ with $y_j \in \ell_j \mathbf{C}_j'$ and $b_{j,1}, \cdots, b_{j,k_j-\ell_j} \in B_j$ is at least $(\delta|A|)^{k_j-\ell_j}$. For those j with $k_j \le d$, we necessarily have $\ell_j = k_j$ and the statement above holds also. Hence we obtained at least $(\delta|A|)^{k-\ell}$ representations

$$x = \sum_{j \in \mathcal{J}} (y_j, \ell_j t_j) + \sum_{j \in \mathcal{J}} \sum_{1 \le i \le k_j - \ell_j} (b_{j,i}, t_j)$$

of the desired form, since $\sum_{j \in \mathcal{J}} (y_j, \ell_j t_j) \in \ell P$ and each $(b_{j,i}, t_j) \in A$.

To show that $A \setminus A'$ is small, observe from our constructions that

$$|A \setminus A'| \le \sum_{j \notin \mathcal{J}} |A_j| + \sum_{j \in \mathcal{J}} |A_j \setminus B_j| + \sum_{j \in \mathcal{J}} |(\mathbf{C}_j \setminus \mathbf{C}_j') \cap \mathbb{Z}^d|.$$

By the definition of $\mathcal{J}$, the first sum above is bounded by $(\eta\alpha/4)|P| \le (\eta/4)|A|$. By the construction of B_j from regularization and Lemma 5.4, both the second and the third sums above are bounded by $\sum_{j \in \mathcal{J}} (\eta/4)|A_j| \le (\eta/4)|A|$. This shows that $|A \setminus A'| \le \eta|A|$, completing the proof. $\square$

As shown in Section 4, Proposition 4.4 implies Theorem 3.4. Hence, as shown in Section 3, this finishes the proof of Hypothesis A*. As shown in Section 2 this implies Hypothesis A and Theorem 1.1 except for the last claims concerning the case u is fixed.

7. Case u is fixed and $v \geq 1000u^2/\lambda^2$ of Hypothesis A

In this section we deduce the last claim of Hypothesis A from the first part of Hypothesis A and the arithmetic removal lemma. Note that by Hypothesis A*, we know that the first part of Hypothesis A actually holds for $N \geq (30v/\lambda)^2$.

We can assume that N is large enough depending on u and λ, since otherwise the claim follows trivially from the discrete Bleichenbacher theorem (Proposition 1.2). Notice first that we can assume that, for every $u' \in [1, e^{-1/u}v]$, one has

$$\sum_{\substack{a\in A\\ N/e^{-1/u}v<a\leq N/u'}} \frac{1}{a} < \frac{1+\lambda/2}{u'}, \tag{7.1}$$

since otherwise the claim follows immediately from the first part of Hypothesis A. Notice also that

$$\sum_{(1-\frac{\lambda}{8})N<a\leq N/u} \frac{1}{a} \leq \frac{\lambda}{4u}. \tag{7.2}$$

Indeed, this is trivially true if $u \geq (1-\lambda/8)^{-1}$, and if $u \leq (1-\lambda/8)^{-1} \leq 8/7$, then each summand is at most $8/(7N)$ and there are at most $\lambda N/8 + 1$ summands. Using these we obtain that

$$\begin{aligned}\sum_{\substack{a\in A\\ \frac{\lambda}{8u}N<a\leq(1-\frac{\lambda}{8})N}} \frac{1}{a} &\geq \sum_{\substack{a\in A\\ N/v<a\leq N/u}} \frac{1}{a} - \sum_{N/v<a\leq N/e^{-1/u}v} \frac{1}{a} - \sum_{\substack{a\in A\\ N/e^{-1/u}v<a\leq\frac{\lambda}{8u}N}} \frac{1}{a} - \sum_{(1-\frac{\lambda}{8})N<a\leq N/u} \frac{1}{a}\\ &\geq \frac{1+\lambda}{u} - \frac{1+\lambda/8}{u} - \frac{\lambda}{4u} - \frac{\lambda}{4u} \geq \frac{\lambda}{4u},\end{aligned}$$

where we used (7.1) to bound the third sum. This implies that

$$|A'| := \left|A \cap \left[\frac{\lambda}{8u}N < a \leq \left(1-\frac{\lambda}{8}\right)N\right]\right| \geq \frac{\lambda}{8u}N \cdot \frac{\lambda}{4u} \gg N \tag{7.3}$$

since λ and u are fixed. Then, by the removal lemma for popular sums (Theorem 3.2), there exist $\delta = \delta(\lambda)$ and $A'' \subseteq A'$ such that $|A''| \geq (1-\lambda^2/(1000u^2))|A'|$ and, for all positive integers $k \leq 8u/\lambda$, $r_{kA'}(n) \geq \delta N^{k-1}$ for every $n \in kA''$.

Assume first that for some $k_0 \in \{1, \dots, \lfloor 8u/\lambda \rfloor\}$,

$$|B| := |k_0 A'' \cap [0.65N, (1-\lambda/40)N]| \geq \frac{\lambda^3}{10000u^3}N. \tag{7.4}$$

Let $b \in B$. Writing $N' = N - b$, $v' = v(N-b)/N$ and $u' = e^{-1/u}v(N-b)/N$, we have, by (7.1),

$$\sum_{\substack{a\in A\\ N'/v'<a\le N'/u'}} \frac{1}{a} = \sum_{\substack{a\in A\\ N/v<a\le N/e^{-1/u}v}} \frac{1}{a} \ge \sum_{\substack{a\in A\\ N/v<a\le N/u}} \frac{1}{a} - \sum_{\substack{a\in A\\ N/e^{-1/u}v<a\le N/u}} \frac{1}{a}$$
$$\ge \frac{1+\lambda}{u} - \frac{1+\lambda/2}{u} = \frac{\lambda}{2u} \ge \frac{2}{u'}$$

since $v \ge 1000u^2/\lambda^2$.

Note also that $N' = Nv'/v \ge (100v/\lambda)^2 v'/v \ge 30v'^2/\lambda^2$ since $v \ge v'$. Hence we can apply the first part of Hypothesis A with N', u', v' obtaining that there exists $k_1 \le v' \le 0.35v$ such that

$$|\{(a_1,\dots,a_{k_1}) \in A^{k_1} \colon N - b - k_1 \le a_1 + \cdots + a_{k_1} \le N - b\}| \ge \alpha'_{v'} \frac{|A|^{k_1}}{N},$$

where $\alpha'_{v'} > 0$ is a constant with $\alpha'_v = v^{-o(v)}$ as $v \to \infty$. Write $k = k_1 + k_0 \le 0.35v + 8u/\lambda \le e^{-1/u}v$. We get that

$$\begin{aligned}
&|\{(a_1,\dots,a_k) \in A^k \colon N - k \le a_1 + \cdots + a_k \le N\}| \\
&\ge \sum_{b\in B} r_{k_0A'}(b) \cdot |\{(a_1,\dots,a_{k_1}) \in A^{k_1} \colon N - b - k_1 \le a_1 + \cdots + a_{k_1} \le N - b\}| \\
&\ge \frac{\lambda^3}{10000u^3} N \cdot \delta N^{k_0-1} \cdot \alpha'_{v'} \frac{|A|^{k_1}}{N} \ge \alpha_v \frac{|A|^k}{N},
\end{aligned}$$

where $\alpha_v = \frac{\lambda^3}{10000u^3} \cdot \delta\alpha'_{v'} = v^{-o_\lambda(v)}$ as $v \to \infty$.

Consider now the case that (7.4) does not hold for any $k_0 \le 8u/\lambda$. Write

$$D = A'' \cap \left[\left(\frac{1}{2} - \frac{\lambda}{80}\right) N, 0.65N\right].$$

Then $|A'' \setminus D| \le \frac{\lambda^2}{1250u^2} N$ since every $a \in A'' \setminus D$ produces some $k_0 a$ counted in (7.4) for some $k_0 \le 8u/\lambda$. Since $|A''| \ge |A'|/2 \ge \frac{\lambda^2}{64u^2} N$ by (7.3), we have

$$|D| \ge \frac{\lambda^2}{64u^2} N - \frac{\lambda^2}{1250u^2} N \ge \frac{\lambda^2}{100u^2} N. \tag{7.5}$$

In particular D is non-empty. Since $A'' \subseteq (N/v, N/u]$, we must have $u \le 2.1$. From the bound

$$|A' \setminus D| \le |A'' \setminus D| + |A' \setminus A''| \le \frac{\lambda^2 N}{500u^2},$$

together with (7.2) we get

$$\sum_{\substack{a\in A\\ \frac{\lambda}{8u}N<a\le N/u}}\frac{1}{a}\le\frac{8u}{\lambda N}|A'\setminus D|+\sum_{(1-\frac{\lambda}{8})N<a\le N/u}\frac{1}{a}+\sum_{(\frac12-\frac{\lambda}{80})N\le a\le 0.65N}\frac{1}{a}$$
$$\le\frac{\lambda}{50u}+\frac{\lambda}{4u}+\log\frac{0.65}{1/2-\lambda/80}+O(1/N)<0.3+0.27\cdot\frac{\lambda}{u}.$$

Let $d\in D$ and write now $N'=N-d, u'=\frac{8u}{\lambda}\cdot\frac{N-d}{N}$ and $v'=1.9ve^{-1/u}\frac{N-d}{N}$. Then

$$\sum_{\substack{a\in A\\ N'/v'<a\le N'/u'}}\frac{1}{a}=\sum_{\substack{a\in A\\ N/(1.9ve^{-1/u})<a\le\lambda N/(8u)}}\frac{1}{a}$$
$$\ge\sum_{\substack{a\in A\\ N/v<a\le N/u}}\frac{1}{a}-\sum_{N/v<a\le\max\{N/v,N/(1.9ve^{-1/u})\}}\frac{1}{a}-\sum_{\substack{a\in A\\ \frac{\lambda}{8u}N\le a\le N/u}}\frac{1}{a}$$
$$\ge\frac{1+\lambda}{u}+\min\{0,\log(1.9e^{-1/u})\}-O(1/N)-0.3-0.27\cdot\frac{\lambda}{u}$$
$$\ge\min\{\log1.9,1/u\}-0.32+0.73\cdot\frac{\lambda}{u}\ge0.73\cdot\frac{\lambda}{u}\ge\frac{1.4}{u'}$$

where we used $u\le2.1$ and $u'\ge2u/\lambda$ in the last two steps. Note also that $N'=Nv'/(1.9ve^{-1/u})\ge(100v/\lambda)^2v'/(2v)\ge(30v'/\lambda)^2$ since $v\ge v'$. Hence we can now apply the first part of Hypothesis A with N',u' and v', to obtain $k_1\le v'$ such that

$$|\{(a_1,\dots,a_{k_1})\in A^{k_1}\colon N-d-k_1\le a_1+\cdots+a_{k_1}\le N-d\}|\ge\alpha'_{v'}\frac{|A|^{k_1}}{N},$$

where $\alpha'_{v'}>0$ is a constant with $\alpha'_v=v^{-o(v)}$ as $v\to\infty$. Write $k=k_1+1\le v'+1\le0.98e^{-1/u}v+1\le e^{-1/u}v$. We then get

$$|\{(a_1,\dots,a_k)\in A^k\colon N-k\le a_1+\cdots+a_k\le N\}|$$
$$\ge\sum_{d\in D}|\{(a_1,\dots,a_{k_1})\in A^{k_1}\colon N-d-k_1\le a_1+\cdots+a_{k_1}\le N-d\}|$$
$$\ge\frac{\lambda^2}{100u^2}N\cdot\alpha'_{v'}\frac{|A|^{k_1}}{N}\ge\alpha_v\frac{|A|^k}{N},$$

where $\alpha_v=\frac{\lambda^2}{100u^2}\cdot\alpha'_{v'}=v^{-o_\lambda(v)}$ as $v\to\infty$.

References

[1] D. Bleichenbacher. The continuous postage stamp problem. *Unpublished manuscript*, 2003.

[2] J. Bourgain. Estimates related to sumfree subsets of sets of integers. *Israel J. Math.*, 97:71–92, 1997.

[3] J. W. S. Cassels. Measures of the non-convexity of sets and the Shapley-Folkman-Starr theorem. *Math. Proc. Cambridge Philos. Soc.*, 78(3):433–436, 1975.

[4] I. Ekeland and R. Temam. *Convex analysis and variational problems*. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1976. Translated from the French, Studies in Mathematics and its Applications, Vol. 1.
[5] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.
[6] A. Granville, D. Koukoulopoulos, and K. Matomäki. When the sieve works. *Duke Math. J.*, 164:1935–1969, 2015.
[7] B. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.
[8] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, Rhode Island, 2004.
[9] H. W. Lenstra jr. and C. Pomerance. Primality testing with gaussian periods. *Preprint, available at* `http://www.math.dartmouth.edu/~carlp/aks06-2015.pdf`, 2011.
[10] D. Král, O. Serra, and L. Vena. A combinatorial proof of the removal lemma for groups. *J. Combin. Theory Ser. A*, 116(4):971–978, 2009.
[11] I. Z. Ruzsa. The Brunn-Minkowski inequality and nonconvex sets. *Geom. Dedicata*, 67(3):337–348, 1997.
[12] A. Shapira. A proof of Green's conjecture regarding the removal properties of sets of linear equations. *J. Lond. Math. Soc. (2)*, 81(2):355–373, 2010.
[13] B. Szegedy. The symmetry preserving removal lemma. *Proc. Amer. Math. Soc.*, 138(2):405–408, 2010.
[14] E. Szemerédi and V. H. Vu. Finite and infinite arithmetic progressions in sumsets. *Ann. of Math. (2)*, 163(1):1–35, 2006.
[15] T. Tao and V. Vu. John-type theorems for generalized arithmetic progressions and iterated sumsets. *Adv. Math.*, 219(2):428–449, 2008.
[16] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, paperback edition, 2010.
[17] J. Wolf. The structure of popular difference sets. *Israel J. Math.*, 179:253–278, 2010.

Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland
E-mail address: `ksmato@utu.fi`

Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom
E-mail address: `Xuancheng.Shao@maths.ox.ac.uk`