INVERSE QUESTIONS FOR THE LARGE SIEVE

BEN GREEN AND ADAM J HARPER

Abstract. Suppose that an infinite set $\mathscr{A}$ occupies at most $\frac{1}{2}(p+1)$ residue classes modulo p, for every sufficiently large prime p. The squares, or more generally the integer values of any quadratic, are an example of such a set. By the large sieve inequality the number of elements of $\mathscr{A}$ that are at most X is $O(X^{1/2})$, and the quadratic examples show that this is sharp. The simplest form of the *inverse large sieve problem* asks whether they are the only examples. We prove a variety of results and formulate various conjectures in connection with this problem, including several improvements of the large sieve bound when the residue classes occupied by $\mathscr{A}$ have some additive structure. Unfortunately we cannot solve the problem itself.

To Roger Heath-Brown on his 60th birthday

Contents

1. Introduction

Notation. Most of our notation is quite standard. When dealing with infinite sets $\mathscr{A}$, we write $\mathscr{A}[X]$ for the intersection of $\mathscr{A}$ with the initial segment $[X] := \{1, \dots, X\}$.

Our primary aim in this paper is to study sets $\mathscr{A}$ of integers with the property that the reduction $\mathscr{A} \pmod p$ occupies at most $\frac{1}{2}(p+1)$ residue classes modulo p for all sufficiently large primes p. It follows from the large sieve that $|\mathscr{A}[X]| \ll X^{1/2}$ for all X (we will recall the details of this argument below). This is clearly sharp up to the value of the implied constant, as shown by taking $\mathscr{A}$ to be the set of squares or more generally the set of integer values taken by any rational quadratic, that is to say quadratic with rational coefficients.

It has been speculated, most particularly by Helfgott and Venkatesh [12, pp 232–233] and by Walsh [20], that quadratics provide the only examples of sets for which the large sieve bound is essentially

sharp. See also [1, Problem 7.4]. One might call problems of this type the "inverse large sieve problem". Unfortunately, we have not been able to prove any statement of this kind, and our aims here are more modest.

Suppose that $\mathscr{A} (\mathrm{mod}\ p) \subset S_p$ for all sufficiently large primes p. Our first set of results consists of improvements to the large sieve bound when S_p looks very much unlike a quadratic set modulo p, for example by having some additive structure.

Theorem 1.1. *Suppose that for each prime $p \leqslant X^{1/2}$ one has a set $S_p \subset \mathbb{Z}/p\mathbb{Z}$ of size at most $(p+1)/2$. Suppose there is some $\delta > 0$ such that, for each p, S_p has at least $(\frac{1}{16} + \delta)p^3$ additive quadruples, that is to say quadruples (s_1, s_2, s_3, s_4) with $s_1 + s_2 = s_3 + s_4$. Suppose that $\mathscr{A} (\mathrm{mod}\, p) \subset S_p$ for all p. Then $|\mathscr{A}[X]| \ll X^{1/2-c\delta^2}$, where $c > 0$ is an absolute constant.*

The condition of having at least $(\frac{1}{16} + \delta)p^3$ additive quadruples will *not* be satisfied by a generic (e.g. randomly selected) set $S_p \subset \mathbb{Z}/p\mathbb{Z}$ of size $(p+1)/2$, which will have $(\frac{1}{16} + o(1))p^3$ such quadruples for large p. But it is a rather general condition corresponding to S_p being additively structured, and certainly we are not aware of any previous improvements to the large sieve bound under comparably general conditions.

An extreme case of the preceding theorem is that in which S_p is in fact an interval. Here a simple calculation, reproduced later, shows that Theorem 1.1 is applicable with the choice $\delta = 1/48$, but we can do rather better.

Theorem 1.2. *Suppose that $\mathscr{A}$ is a set of integers and that, for each prime p, the set $\mathscr{A} (\mathrm{mod}\, p)$ lies in some interval I_p. Let $\varepsilon > 0$ be arbitrary. Then*

(i) *If $|I_p| \leqslant (1-\varepsilon)p$ for at least a proportion ε of the primes in each dyadic interval $[Z, 2Z]$ then $|\mathscr{A}[X]| \ll_\varepsilon (\log\log X)^{C\log(1/\varepsilon)}$, where $C > 0$ is some absolute constant;*

(ii) *If $|I_p| \leqslant \frac{p}{2}$ for all primes then $|\mathscr{A}[X]| \ll (\log\log X)^{\gamma + o(1)}$, where $\gamma = \frac{\log 18}{\log(3/2)} \approx 7.129$;*

(iii) *If $I_p = [\alpha p, \beta p]$ for all primes p and for fixed $0 \leqslant \alpha < \beta < 1$ (not depending on p) then $|\mathscr{A}| = O_{\beta-\alpha}(1)$;*

(iv) *There are examples of infinite sets $\mathscr{A}$ with $|I_p| \leqslant (\frac{1}{2} + \varepsilon)p$ for all p.*

This improves on the results of an unpublished preprint [10] by the first author, in which it was shown that one has $|\mathscr{A}[X]| \ll X^{1/3+o(1)}$ under the condition (ii).

Theorem 1.1 also covers the case in which S_p is an arithmetic progression of length $\frac{1}{2}(p+1)$, where the common difference of this arithmetic progression may depend on p. Here again one could apply Theorem 1.1 with the choice $\delta = 1/48$, but we can also handle this situation with a less restrictive condition on the size of S_p.

Theorem 1.3. *Let $\varepsilon > 0$. Suppose that $\mathscr{A}$ is a set of integers and that, for each prime $p \leqslant X^{1/2}$, the set $\mathscr{A} (\mathrm{mod}\, p)$ lies in some arithmetic progression S_p of length $(1-\varepsilon)p$. Then $|\mathscr{A}[X]| \ll_\varepsilon X^{1/2-\varepsilon'}$, where $\varepsilon' > 0$ depends on ε only.*

We are not aware of any previous results improving the large sieve bound $X^{1/2}$ when the S_p are arbitrary arithmetic progressions, even for $\varepsilon = 1/2$.

After proving the foregoing results, we turn to the "robustness" of the inverse large sieve problem. The aim of these results is to show that if $|\mathscr{A}(\mathrm{mod}\ p)| \leqslant \frac{1}{2}(p+1)$ (or if similar conditions hold), if $|\mathscr{A}[X]| \approx X^{1/2}$, and if $\mathscr{A}$ is even vaguely close to quadratic in structure, then it must in fact approximate a quadratic very closely. Our proof methods here lead to some complicated dependencies between parameters, so we do not state and prove the most general result possible, settling instead for a couple of statements that have relatively clean formulations.

The first and main one concerns finite sets. Here, and henceforth in the paper, we say that a rational quadratic ψ has *height at most* H if it can be written as $\psi(x) = \frac{1}{d}(ax^2+bx+c)$ with $a,b,c,d \in \mathbb{Z}$ and $\max(|a|,|b|,|c|,|d|) \leqslant H$.

Theorem 1.4. *Let $X_0 \in \mathbb{N}$, and let $\varepsilon > 0$. Let $X \in \mathbb{N}$ be sufficiently large in terms of X_0 and ε, and suppose that $H \leqslant X^{1/8}$. Suppose that $A, B \subset [X]$ and that $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p+1$ for all $p \in [X_0, X^{1/4}]$. Then, for some absolute constant $c > 0$, one of the following holds:*

(i) (Better than large sieve) *Either $|A \cap [X^{1/2}]|$ or $|B \cap [X^{1/2}]|$ is $\leqslant X^{1/4-c\varepsilon^3}$;*
(ii) (Behaviour with quadratics) *Given any two rational quadratics ψ_A, ψ_B of height at most H, either $|A \setminus \psi_A(\mathbb{Q})|$ and $|B \setminus \psi_B(\mathbb{Q})| \leqslant HX^{1/2-c}$, or else at least one of $|A \cap \psi_A(\mathbb{Q})|$ and $|B \cap \psi_B(\mathbb{Q})|$ is bounded above by $HX^{1/4+\varepsilon}$.*

We expect that if the large sieve bound is close to sharp for A and B, then there must exist rational quadratics of "small" height containing "many" points of A and B. Together with Theorem 1.4, this provides some motivation for making the following conjecture of the form "almost equality in the large sieve implies quadratic structure".

Conjecture 1.5. *Let $X_0 \in \mathbb{N}$, and let $\rho > 0$. Let $X \in \mathbb{N}$ be sufficiently large in terms of X_0 and ρ. Suppose that $A, B \subset [X]$ and that $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p+1$ for all $p \in [X_0, X^{1/4}]$. Then there exists a constant $c = c(\rho) > 0$ such that one of the following holds:*

(i) (Better than large sieve) *Either $|A \cap [X^{1/2}]|$ or $|B \cap [X^{1/2}]|$ is $\leqslant X^{1/4-c}$;*
(ii) (Quadratic structure) *There are two rational quadratics ψ_A, ψ_B of height at most X^{ρ} such that $|A \setminus \psi_A(\mathbb{Q})|$ and $|B \setminus \psi_B(\mathbb{Q})| \leqslant X^{1/2-c}$.*

The contents of Theorem 1.4 and of Conjecture 1.5 are perhaps a little hard to understand on account of the parameters H, X, ρ and ε. As a corollary we establish the following more elegant statement involving infinite sets.

Theorem 1.6. *Suppose that $\mathscr{A}$ is a set of positive integers and that $|\mathscr{A}(\mathrm{mod}\ p)| \leqslant \frac{1}{2}(p+1)$ for all sufficiently large primes p. Then one of the following options holds:*

(i) (Quadratic structure) *There is a rational quadratic ψ such that all except finitely many elements of $\mathscr{A}$ are contained in $\psi(\mathbb{Q})$;*
(ii) (Better than large sieve) *For each integer k there are arbitrarily large values of X such that $|\mathscr{A}[X]| < \frac{X^{1/2}}{\log^k X}$;*
(iii) (Far from quadratic structure) *Given any rational quadratic ψ, for all X we have $|\mathscr{A}[X] \cap \psi(\mathbb{Q})| \leqslant X^{1/4+o_\psi(1)}$.*

We conjecture that option (iii) is redundant. This is another conjecture of inverse large sieve type, rather cleaner than Conjecture 1.5.

Conjecture 1.7. *Suppose that $\mathscr{A}$ is a set of positive integers and that $|\mathscr{A} (\mathrm{mod}\ p)| \leqslant \frac{1}{2}(p+1)$ for all sufficiently large primes p. Then one of the following options holds:*

(i) (Quadratic structure) *There is a rational quadratic ψ such that all except finitely many elements of $\mathscr{A}$ are contained in $\psi(\mathbb{Q})$;*

(ii) (Better than large sieve) *For each integer k there are arbitrarily large values of X such that $|\mathscr{A}[X]| < \frac{X^{1/2}}{\log^k X}$. In particular, $\liminf_{X\to\infty} X^{-1/2}|\mathscr{A}[X]| = 0$.*

We remark that some very simple properties of rational quadratics are laid down in Appendix A. In particular we draw attention to the fact that given a rational quadratic ψ there are further rational quadratics ψ_1, ψ_2 such that $\psi_1(\mathbb{Z}) \subset \psi(\mathbb{Q}) \cap \mathbb{Z} \subset \psi_2(\mathbb{Z})$. In particular, $|\psi(\mathbb{Q}) \cap [X]| \ll_\psi X^{1/2}$.

Our final task in the paper is to show, elaborating on ideas of Elsholtz [3], that Conjecture 1.5 would resolve the currently unsolved "inverse Goldbach problem" of Ostmann [16, p. 13] (and see also [5, p. 62]). This asks whether the set of primes can be written as a sumset $\mathscr{A} + \mathscr{B}$ with $|\mathscr{A}|, |\mathscr{B}| \geqslant 2$, except for finitely many mistakes. Evidently the answer should be that it cannot be so written.

Theorem 1.8. *Assume Conjecture 1.5. Let $\mathscr{A}, \mathscr{B}$ be two sets of positive integers, with $|\mathscr{A}|, |\mathscr{B}| \geqslant 2$, such that $\mathscr{A} + \mathscr{B}$ contains all sufficiently large primes. Then $\mathscr{A} + \mathscr{B}$ also contains infinitely many composite numbers.*

We remark that much stronger statements that would imply this should be true, but we do not know how to prove them. For example, it is reasonable to make the following conjecture.

Conjecture 1.9. *Let $\delta > 0$. Then if X is sufficiently large in terms of δ, the following is true. Let $A, B \subset [X]$ be any sets with $|A|, |B| \geqslant X^\delta$. Then $A + B$ contains a composite number.*

We do not know how to prove this for any $\delta \leqslant \frac{1}{2}$. If one had it for any $\delta < \frac{1}{2}$, the inverse Goldbach problem would follow.

The proofs of the above theorems are rather diverse and use the large sieve, Gallagher's "larger sieve", and several other tools from harmonic analysis and analytic number theory. The proofs of Theorems 1.1 and 1.3, which involve lifting the additive structure of the S_p to additive structure on $\mathscr{A}[X]$, also involve some ideas of additive combinatorial flavour, although we do not need to import many results from additive combinatorics to prove them. With very few exceptions (for example, Lemma 5.1 depends on standard Fourier arguments given in detail in Lemma 4.1), Sections 3,4,5,6 and 7 may be read independently of one another.

The situation considered in the majority of this paper, in which $\mathscr{A} (\mathrm{mod}\ p)$ is small for *every* prime p (or at least for every prime $p \leqslant X^{1/2}$), may seem rather restrictive. It would be possible to adapt our arguments and prove many of our theorems under weaker conditions, and we leave this to the reader who has need of such results. However, it seems possible that any set $\mathscr{A}$ for which $|\mathscr{A} (\mathrm{mod}\ p)| \leqslant (1-c)p$ for a decent proportion of primes p and for which $|\mathscr{A}[X]| \geqslant X^c$ for infinitely many X has at least some "algebraic structure". Moreover such statements may well be true in finitary settings, in which $\mathscr{A}$ is

restricted to some finite interval $[X]$ and p is only required to range over some (potentially quite small) subinterval of $[X]$. Unfortunately none of our methods come close to establishing such strong results.

Acknowledgements. The authors are very grateful to Jean Bourgain for allowing us to use his ideas in Section 4. We also thank the anonymous referee for his or her very careful reading of the paper. The first-named author is supported by ERC Starting Grant 279438 *Approximate Algebraic Structure and Applications*. He is very grateful to the University of Auckland for providing excellent working conditions on a sabbatical during which this work was completed. The second-named author was supported by a Doctoral Prize from the EPSRC when this work was started; by a postdoctoral fellowship from the Centre de Recherches Mathématiques, Montréal; and by a research fellowship at Jesus College, Cambridge when the work was completed.

2. The large sieve and the larger sieve

The large sieve. Let us begin by briefly recalling a statement of the large sieve bound. The following may be found in Montgomery [15].

Proposition 2.1. *Let $\mathscr{A}$ be a set of positive integers with the property that $\mathscr{A} (\mathrm{mod}\, p) \subset S_p$ for each prime p. Then for any Q, X we have the bound*

$$|\mathscr{A}[X]| \leqslant \frac{X+Q^2}{\sum_{q \leqslant Q} \mu^2(q) \prod_{p|q} \frac{|S_p^c|}{|S_p|}} \leqslant \frac{X+Q^2}{\sum_{p \leqslant Q} \frac{|S_p^c|}{p}}.$$

where $\mu(q)$ denotes the Möbius function.

The second bound is a little crude but has the virtue of being simple: we will use it later on. In the particular case that $|S_p| \leqslant \frac{1}{2}(p+1)$ for all p, discussed in the introduction, the first bound implies upon setting $Q := X^{1/2}$ that

$$|\mathscr{A}[X]| \leqslant \frac{2X}{\sum_{q \leqslant X^{1/2}} \mu^2(q) \prod_{p|q} \frac{p-1}{p+1}} \ll X^{1/2},$$

as we claimed.

The large sieve may also be profitably applied to "small sieve" situations in which $|S_p| = p - O(1)$ (as opposed to "large sieve" situations in which $p - |S_p|$ is large). We will need one such result later on, in §6.

Lemma 2.2. *Suppose that $\mathscr{B} \subset \mathbb{Z}$ is a set with the property that $\mathscr{B} (\mathrm{mod}\, p)$ misses $w(p)$ residue classes, for every prime p. Suppose that the function w has average value k in the (fairly weak) sense that $\frac{1}{Z} \sum_{Z \leqslant p \leqslant 2Z} w(p) \log p = k + O(\frac{1}{\log^2 Z})$ for all Z. Then $|\mathscr{B} \cap [-N, N]| \ll N(\log N)^{-k}$ for all N.*

Proof. In view of Proposition 2.1, it would suffice to know that

$$\sum_{q \leqslant N^{1/2}} \mu^2(q) \prod_{p|q} \frac{w(p)}{p - w(p)} \gg \log^k N$$

for all large N. If we define a multiplicative function $g(n)$, supported on squarefree integers, by $g(p) := w(p)/p$, then it would obviously suffice to know that $\sum_{q \leqslant N^{1/2}} g(q) \gg \log^k N$ for all large N. However there is an extensive theory, dating back to Halász, Wirsing and others, that gives asymptotics

and bounds for sums of multiplicative functions. For example, partial summation and the assumption that $\sum_{Z\leqslant p\leqslant 2Z} w(p)\log p = kZ + O(\frac{Z}{\log^2 Z})$ show that g satisfies the conditions of Theorem A.5 of Friedlander and Iwaniec [6], and therefore

$$\sum_{q\leqslant N} g(q) \sim c_g \log^k N \quad \text{as } N\to\infty,$$

for a certain constant $c_g > 0$. □

The larger sieve. The "larger sieve" was introduced by Gallagher [8]. A pleasant discussion of it may be found in chapter 9.7 of Friedlander and Iwaniec [6]. We will apply it several times in the paper, and we formulate a version suitable for those applications.

Theorem 2.3. *Let $0 < \delta \leqslant 1$, and let $Q > 1$. Let $\mathscr{P}$ be a set of primes. For each prime $p \in \mathscr{P}$, suppose that one is given a set $S_p \subset \mathbb{Z}/p\mathbb{Z}$, and write $\sigma_p := |S_p|/p$. Suppose that there is some set $\mathscr{A}'_p \subset \mathscr{A}$, $|\mathscr{A}'_p[X]| \geqslant \delta|\mathscr{A}[X]|$, such that $\mathscr{A}'_p (\mathrm{mod}\ p) \subset S_p$. Then*

$$|\mathscr{A}[X]| \ll \frac{Q}{\delta^2 \sum_{p\in\mathscr{P}, p\leqslant Q} \frac{\log p}{\sigma_p p} - \log X},$$

provided that the denominator is positive.

Remark. In this paper we will always have δ at least some absolute constant, not depending on X, and very often we will have $\delta \approx 1$.

Proof. We examine the expression

$$I := \sum_{\substack{p\in\mathscr{P}\\ p\leqslant Q}} \sum_{\substack{x,y\in\mathscr{A}[X]\\ x\neq y}} 1_{p|x-y}\log p.$$

On the one hand we have

$$\sum_{p\in\mathscr{P}} 1_{p|n}\log p \leqslant \sum_{p} 1_{p|n}\log p \leqslant \log n,$$

and therefore

$$I \leqslant |\mathscr{A}[X]|^2 \log X.$$

On the other hand, writing $\mathscr{A}(a,p;X)$ for the number of $x\in\mathscr{A}[X]$ with $x\equiv a (\mathrm{mod}\ p)$, we have

$$I = \sum_{\substack{p\in\mathscr{P}\\ p\leqslant Q}} \sum_{a (\mathrm{mod}\, p)} |\mathscr{A}(a,p;X)|^2 \log p - |\mathscr{A}[X]| \sum_{\substack{p\in\mathscr{P}\\ p\leqslant Q}} \log p.$$

Comparing these facts yields

$$\sum_{\substack{p\in\mathscr{P}\\ p\leqslant Q}} \sum_{a (\mathrm{mod}\, p)} |\mathscr{A}(a,p;X)|^2 \log p \leqslant |\mathscr{A}[X]|^2 \log X + O(|\mathscr{A}[X]|Q), \tag{2.1}$$

and so of course, since $\mathscr{A}'_p \subset \mathscr{A}$,

$$\sum_{\substack{p\in\mathscr{P}\\ p\leqslant Q}} \sum_{a (\mathrm{mod}\, p)} |\mathscr{A}'_p(a,p;X)|^2 \log p \leqslant |\mathscr{A}[X]|^2 \log X + O(|\mathscr{A}[X]|Q).$$

However by the Cauchy–Schwarz inequality and the fact that $\mathscr{A}'_p(\mathrm{mod}\ p) \subset S_p$ we have

$$\sum_{a(\mathrm{mod}\, p)} |\mathscr{A}'_p(a,p;X)|^2 \geqslant \frac{|\mathscr{A}'_p[X]|^2}{\sigma_p p} \geqslant \delta^2 \frac{|\mathscr{A}[X]|^2}{\sigma_p p}.$$

Summing over p and rearranging, we obtain the result. □

The larger sieve bound can be a little hard to get a feel for, so we give an example. Suppose that $\mathscr{P}$ consists of all primes and that $\sigma_p = \alpha$ for all p. Take $\delta = 1$. Then, since $\sum_{p\leqslant Q} \frac{\log p}{p} = \log Q + O(1)$, the larger sieve bound is essentially

$$|\mathscr{A}[X]| \ll \frac{Q}{\frac{1}{\alpha}\log Q - \log X}.$$

Taking Q a little larger than X^α, we obtain the bound $|\mathscr{A}[X]| \ll X^{\alpha+o(1)}$. This beats an application of the large sieve when $\alpha < \frac{1}{2}$, that is to say when we are sieving out a majority of residue classes (hence the terminology "larger sieve"). However in the type of problems we are generally considering in this paper, where $\alpha = \frac{1}{2}$, we only recover the bound $|\mathscr{A}[X]| \ll X^{1/2+o(1)}$, as of course we must since nothing has been done to exclude the example where $\mathscr{A}$ is a set of values of a quadratic.

In actual fact one of our three applications of the larger sieve (in the proof of Theorem 1.1) requires an inspection of the above proof, rather than an application of the result itself. This is the observation that when $\sigma_p \approx \frac{1}{2}$ the larger sieve *does* beat the bound $|\mathscr{A}[X]| \ll X^{1/2+o(1)}$ unless $\mathscr{A}$ satisfies a certain "uniform fibres" condition. Recall that if $\mathscr{A}$ is a set of integers then $\mathscr{A}(a,p;X)$ is the number of $x \in \mathscr{A}[X]$ with $x \equiv a (\mathrm{mod}\ p)$.

Lemma 2.4. *Let $\kappa > 0$ and $\eta > 0$ be small parameters. Suppose that $\mathscr{A}$ is a set of integers occupying at most $(p+1)/2$ residue classes modulo p for all p. Say that $\mathscr{A}[X]$ has* η-uniform fibres above p *if*

$$\sum_{a(\mathrm{mod}\, p)} |\mathscr{A}(a,p;X)|^2 \leqslant (2+\eta)|\mathscr{A}[X]|^2/p.$$

Let $\mathscr{P}_{\mathrm{unif}}$ be the set of primes above which $\mathscr{A}[X]$ has η-uniform fibres. Then either $|\mathscr{A}[X]| \leqslant X^{1/2-\kappa}$ or else "most" fibres are η-uniform in the sense that

$$\sum_{\substack{p\leqslant X^{1/2},\\ p\notin\mathscr{P}_{\mathrm{unif}}}} \frac{\log p}{p} \leqslant \frac{3\kappa\log X + O(1)}{\eta}.$$

Proof. Let $\mathscr{P}$ be the set of all primes, and let $Q := X^{1/2-\kappa}$. We proceed as in the proof of the larger sieve until (2.1), which was the inequality

$$\sum_{p\leqslant Q}\sum_{a(\mathrm{mod}\, p)} |\mathscr{A}(a,p;X)|^2 \log p \leqslant |\mathscr{A}[X]|^2 \log X + O(|\mathscr{A}[X]|Q).$$

Now by the Cauchy–Schwarz inequality we have

$$\sum_{a(\mathrm{mod}\, p)} |\mathscr{A}(a,p;X)|^2 \geqslant 2|\mathscr{A}[X]|^2/(p+1)$$

for all p. Using this and the estimate $\sum_{p\leqslant Q}\log p/p = \log Q + O(1)$, we see that the left-hand side of (2.1) is at least

$$2|\mathscr{A}[X]|^2(\log Q + O(1)) + \eta|\mathscr{A}[X]|^2 \sum_{\substack{p\leqslant Q\\ p\notin\mathscr{P}_{\rm unif}}}\frac{\log p}{p}.$$

Therefore

$$\eta\sum_{\substack{p\leqslant Q\\ p\notin\mathscr{P}_{\rm unif}}}\frac{\log p}{p} \leqslant \log X - 2\log Q + O(1) + O(\frac{Q}{|\mathscr{A}[X]|}).$$

If $|\mathscr{A}[X]|\leqslant X^{1/2-\kappa}$ then we are done; otherwise, the term $O(Q/|\mathscr{A}[X]|)$ may be absorbed into the $O(1)$ term and, after a little rearrangement, we obtain

$$\sum_{\substack{p\leqslant Q\\ p\notin\mathscr{P}_{\rm unif}}}\frac{\log p}{p} \leqslant \frac{2\kappa\log X + O(1)}{\eta}.$$

Since

$$\sum_{Q\leqslant p\leqslant X^{1/2}}\frac{\log p}{p} = \kappa\log X + O(1),$$

the claimed bound follows. □

3. Sieving by additively structured sets

Our aim in this section is to establish Theorem 1.1.

Let A be a finite set of integers. As is standard, we write $E(A,A)$ for the *additive energy* of A, that is to say the number of quadruples $(a_1,a_2,a_3,a_4)\in A^4$ with $a_1+a_2=a_3+a_4$. If p is a prime, write $E_p(A,A)$ for the number of quadruples with $a_1+a_2\equiv a_3+a_4 (\mathrm{mod}\ p)$. It is easy to see that $E_p(A,A)\geqslant |A|^4/p$. In situations where this inequality is not tight, we can get a lower bound for the additive energy $E(A,A)$. To do this we will use the *analytic large sieve inequality*, which is something like an approximate version of Bessel's inequality (and which leads, in a non-obvious way, to the large sieve bound that we stated as Proposition 2.1). We cite the following version, which is best possible in various aspects, from Chapter 9.1 of Friedlander and Iwaniec [6].

Proposition 3.1. *Let $0<\delta\leqslant 1/2$, and suppose that $\theta_1,\dots,\theta_R\in\mathbb{R}/\mathbb{Z}$ form a δ-spaced set of points, in the sense that $\|\theta_r-\theta_s\|\geqslant\delta$ for all $r\neq s$ where $\|\cdot\|$ denotes distance to the nearest integer. Suppose that $(a(x))_{M<x\leqslant M+X}$ are any complex numbers, where X is a positive integer. Then*

$$\sum_{r=1}^{R}\Big|\sum_{M<x\leqslant M+X}a(x)e(\theta_r x)\Big|^2 \leqslant (X-1+\delta^{-1})\sum_{M<x\leqslant M+X}|a(x)|^2,$$

where as usual $e(\theta):=\exp\{2\pi i\theta\}$.

Using the analytic large sieve inequality, we shall prove the following lemma.

Lemma 3.2 (Lifting additive energy)**.** *Suppose that $A\subset[X]$. Then we have*

$$\sum_{p\leqslant X^{1/2}}p\big(E_p(A,A)-\frac{|A|^4}{p}\big)\leqslant 3XE(A,A).$$

Proof. Write $r(x)$ for the number of representations of x as a_1+a_2 with $a_1,a_2\in A$. Then

$$E(A,A)=\sum_{x\leqslant 2X} r(x)^2,$$

whilst

$$E_p(A,A)=\sum_{\substack{x,x'\leqslant 2X\\ x\equiv x'(\mathrm{mod}\, p)}} r(x)r(x')=\frac{1}{p}\sum_{a(\mathrm{mod}\, p)}|\sum_{x\leqslant 2X} r(x)e(ax/p)|^2.$$

It follows that

$$\begin{aligned}\sum_{p\leqslant X^{1/2}} pE_p(A,A)&=\sum_{p\leqslant X^{1/2}}\sum_{a(\mathrm{mod}\, p)}|\sum_{x\leqslant 2X} r(x)e(ax/p)|^2\\ &=\sum_{p\leqslant X^{1/2}}\sum_{\substack{a(\mathrm{mod}\, p)\\ a\neq 0}}|\sum_{x\leqslant 2X} r(x)e(ax/p)|^2+\sum_{p\leqslant X^{1/2}} p\frac{|A|^4}{p}.\end{aligned}$$

Now the fractions a/p are $1/X$-spaced, as a,p range over all pairs with $p\leqslant X^{1/2}$ prime and $1\leqslant a\leqslant p-1$. By the analytic form of the large sieve it follows that

$$\sum_{p\leqslant X^{1/2}}\sum_{a(\mathrm{mod}\, p) a\neq 0}|\sum_{x\leqslant 2X} r(x)e(ax/p)|^2\leqslant 3X\sum_{x\leqslant 2X} r(x)^2.$$

Putting all these facts together gives the result. □

Corollary 3.3. *Let $\eta,\delta>0$ be small real numbers with $\eta\leqslant\delta^2$. Suppose that $A\subset[X]$ is a set. Let $\mathscr{P}$ be a set of primes satisfying $36\delta^{-2}\leqslant p\leqslant X^{1/2}$, and suppose that the following are true whenever $p\in\mathscr{P}$:*

(i) *$A(\mathrm{mod}\, p)$ lies in a set S_p of cardinality at most $\frac{1}{2}(p+1)$;*
(ii) *S_p has at least $(\frac{1}{16}+\delta)p^3$ additive quadruples;*
(iii) *A has η-uniform fibres mod p, in the sense that $\sum_{a(\mathrm{mod}\, p)}|A(a;p)|^2\leqslant(2+\eta)|A|^2/p$, where $A(a;p)$ is the number of $x\in A$ with $x\equiv a(\mathrm{mod}\, p)$.*

Then $E(A,A)\geqslant\frac{\delta|A|^4}{3X}|\mathscr{P}|$.

Proof. Suppose that $p\in\mathscr{P}$. We will obtain a lower bound for $E_p(A,A)$ which beats the trivial bound of $E_p(A,A)\geqslant|A|^4/p$. The corollary will then follow quickly from Lemma 3.2. First of all we apply the variance identity

$$\sum_{m=1}^{M}(t_m-\frac{1}{M}\sum_{i=1}^{M}t_i)^2=\sum_{i=1}^{M}t_i^2-\frac{1}{M}(\sum_{i=1}^{M}t_i)^2$$

with $M:=|S_p|$ and the t_i being the $A(a;p)$ with $a\in S_p$. This and the uniform fibres assumption yields

$$\sum_{a\in S_p}\big(|A(a;p)|-\frac{|A|}{|S_p|}\big)^2\leqslant\frac{(2+\eta)|A|^2}{p}-\frac{2|A|^2}{p+1}\leqslant\frac{|A|^2}{p}\big(\eta+\frac{2}{p}\big).$$

Write $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ for the function $f(a) := |A(a;p)|$, and $g : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ for the function which is $|A|/|S_p|$ on S_p and zero elsewhere. We have shown[1] that

$$\|f-g\|_2 \leqslant \frac{|A|}{p}\sqrt{\eta+\frac{2}{p}}.$$

Note also that, by the Cauchy–Schwarz inequality,

$$\|f-g\|_1 \leqslant \|f-g\|_2 \leqslant \frac{|A|}{p}\sqrt{\eta+\frac{2}{p}},$$

and hence

$$\|\hat{f}-\hat{g}\|_4^4 \leqslant \|\hat{f}-\hat{g}\|_\infty^2\|\hat{f}-\hat{g}\|_2^2 \leqslant \|f-g\|_1^2\|f-g\|_2^2 \leqslant \frac{|A|^4}{p^4}(\eta+\frac{2}{p})^2,$$

which of course implies that

$$\|\hat{f}-\hat{g}\|_4 \leqslant \frac{|A|}{p}(\eta^{1/2}+\sqrt{\frac{2}{p}}),$$

where $\hat{f}(r) := \frac{1}{p}\sum_{a\in\mathbb{Z}/p\mathbb{Z}} f(a)e(-ar/p)$, similarly for $\hat{g}$. It follows that

$$\|\hat{f}\|_4 \geqslant \|\hat{g}\|_4 - \frac{\eta^{1/2}|A|}{p} - \frac{\sqrt{2}|A|}{p^{3/2}}.$$

Note, however, that

$$E_p(A,A) = \sum_{a_1+a_2=a_3+a_4} f(a_1)f(a_2)f(a_3)f(a_4) = p^3\|\hat{f}\|_4^4,$$

whilst

$$\|\hat{g}\|_4^4 = \frac{|A|^4}{|S_p|^4}\frac{E_p(S_p,S_p)}{p^3} \geqslant \frac{|A|^4}{16|S_p|^4}(1+16\delta)$$

and so

$$\|\hat{g}\|_4 \geqslant \frac{|A|}{2|S_p|}(1+2\delta) \geqslant \frac{|A|}{p+1}(1+2\delta).$$

Putting these facts together, and remembering that $\eta \leqslant \delta^2$ and $p \geqslant 36\delta^{-2}$, yields

$$\|\hat{f}\|_4 \geqslant \frac{|A|}{p}(1+2\delta-\eta^{1/2}-3/p^{1/2}) \geqslant \frac{|A|}{p}(1+\delta/2)$$

and so $E_p(A,A) \geqslant \frac{|A|^4}{p}(1+\delta)$ whenever $p \in \mathscr{P}$. The result now follows immediately from Lemma 3.2. □

Corollary 3.4. *Let $\kappa > 0$ be a small parameter. Suppose that $A \subset [X]$ and that, for every prime $p \leqslant X^{1/2}$, the set $A \pmod p$ lies in a set S_p of cardinality at most $\frac{1}{2}(p+1)$ and with at least $(\frac{1}{16}+\delta)p^3$ additive quadruples. Then either $|A| \ll X^{1/2-\kappa}$ or else $E(A,A) \geqslant \delta X^{-9\kappa/\delta^2}|A|^3$.*

Proof. Suppose that $|A| \geqslant X^{1/2-\kappa}$ and that $\kappa \log X$ is large enough. (If $\kappa \log X$ is small then $|A| \ll X^{1/2} \ll X^{1/2-\kappa}$ by the usual large sieve bound.) Set $\eta := \delta^2$. By Lemma 2.4 we either have $|A| \leqslant$

[1]The normalisations here are the ones standard in additive combinatorics. Write $\|F\|_2 = (\frac{1}{p}\sum_{x\in\mathbb{Z}/p\mathbb{Z}}|F(x)|^2)^{1/2}$, but $\|\hat{F}\|_2 = (\sum_r |\hat{F}(r)|^2)^{1/2}$ on the Fourier side. Then we have the Parseval identity $\|F\|_2 = \|\hat{F}\|_2$ and Young's inequality $\|\hat{F}\|_\infty \leqslant \|F\|_1$, both of which are used here.

$X^{1/2-\kappa}$ or else A has η-uniform fibres on a set $\mathscr{P} \subset [X^{1/2}]$ of primes satisfying

$$\sum_{\substack{p \leqslant X^{1/2} \\ p \notin \mathscr{P}}} \frac{\log p}{p} \leqslant \frac{4\kappa \log X}{\eta}.$$

This implies that $|\mathscr{P}| \geqslant X^{1/2-8\kappa/\eta}$, and so by Corollary 3.3 and the fact that $|A| \geqslant X^{1/2-\kappa}$ we have $E(A,A) \geqslant \delta X^{-9\kappa/\eta}|A|^3$, which gives the claimed bound. □

The main task for the rest of this section will be to prove the following.

Proposition 3.5 (Differenced larger sieve)**.** *Let X be large, and let $A \subset [X]$ be a set with the property that $A \pmod p$ lies in a set S_p of size at most $\frac{1}{2}(p+1)$ for all primes $p \leqslant X^{1/2}$. Suppose that $E(A,A) \geqslant |A|^3/K$. Then $|A| \leqslant KX^{1/2-c_0}$, where $c_0 > 0$ is an absolute constant.*

Let us pause to see how this and Corollary 3.4 combine to establish Theorem 1.1.

Proof of Theorem 1.1 given Proposition 3.5. Let $\kappa > 0$ be a parameter to be specified shortly. Suppose that $A \subset [X]$, and that $A \pmod p \subset S_p$ for all primes p. Suppose furthermore that $|S_p| = \frac{1}{2}(p+1)$ and that S_p has at least $(\frac{1}{16} + \delta)p^3$ additive quadruples for all p. By Corollary 3.4 we see that either $|A| \ll X^{1/2-\kappa}$, or else $E(A,A) \geqslant \delta X^{-9\kappa/\delta^2}|A|^3$. In this second case it follows from Proposition 3.5 that $|A| \ll_\delta X^{\frac{1}{2}+9\kappa/\delta^2-c_0}$. Choosing $\kappa := c_0\delta^2/10$ gives the result. □

It remains to prove Proposition 3.5. As the reader will soon see, the proof might be thought of as a "differenced larger sieve" argument, in which the larger sieve is not applied to A directly, but rather to intersections of shifted copies of A (as in Lemma 3.7) and to a set H of pairwise differences of elements of A (as in Lemma 3.10). The assumption that A has large additive energy allows one to recover bounds on A from that information (as in Lemma 3.6).

Remark. It is possible to prove Proposition 3.5 with a quite respectable value of the constant c_0. Unfortunately the quality of the final bound in Theorem 1.1 is not really determined by the value of c_0, but by the much poorer bounds that we achieved when trying to force the set A to have uniform fibres mod p. We believe that by reworking Corollary 3.3 a little one could prove Theorem 1.1 with an improved bound $|\mathscr{A}[X]| \ll X^{1/2-c\sqrt{\delta}}$, but this is presumably very far from optimal.

Proof of Proposition 3.5. The argument is a little involved, so we begin with a sketch. Suppose that $E(A,A) \approx |A|^3$. Then it is not hard to show that $|A \cap (A+h)| \approx |A|$ for $h \in H$, where $|H| \approx |A|$. Modulo p, the set $A \cap (A+h)$ is contained in $S_p \cap (S_p + h)$. If, for some $h \in H$, we have $|S_p \cap (S_p+h)| < (\frac{1}{2} - c)p$ then an application of the larger sieve implies that $|A \cap (A+h)| < X^{1/2-c'}$, and hence $|A| \lessapprox X^{1/2-c'}$. The alternative is that $|S_p \cap (S_p + h)| \approx \frac{1}{2}p$ for many p, for all $h \in H$. Using this we can show that there is *some* p for which $|S_p \cap (S_p + h)| \approx \frac{1}{2}p$ for many h. By a result of Pollard, there is no such set S_p.

Let us turn now to the details, formulating a number of lemmas which correspond to the above heuristic discussion. From now on, the assumptions are as in Proposition 3.5.

Lemma 3.6. *Then there is a set $H \subset [-X, X]$, $|H| \geqslant |A|/2K$ such that $|A \cap (A+h)| \geqslant |A|/2K$ for all $h \in H$.*

Proof. This is completely standard additive combinatorics and is a consequence, for example, of the inequalities in [18, §2.6]. It is no trouble to give a self-contained proof: note that $E(A,A) = \sum_x |A \cap (A+x)|^2$ and that we have the trivial bound $|A \cap (A+x)| \leqslant |A|$ for all x. If H is the maximal set with the stated property then

$$E(A,A) = \sum_{x \in H} |A \cap (A+x)|^2 + \sum_{x \notin H} |A \cap (A+x)|^2 \leqslant |H||A|^2 + \frac{|A|}{2K} \sum_x |A \cap (A+x)| = |H||A|^2 + \frac{|A|^3}{2K},$$

from which the statement follows immediately. □

Lemma 3.7. *Let $c > 0$ be a small constant. Set $Q := X^{1/2-c/2}$, and suppose that there is some $h \in H$ such that*

$$\sum_{p \leqslant Q} \frac{\log p}{p} \frac{|S_p \cap (S_p + h)|}{p} < (\frac{1}{2} - c) \log Q.$$

Then $|A| \ll K X^{1/2-c/4}$.

Proof. Note that $A \cap (A+h) \subseteq S_p \cap (S_p + h)$ for all p. Write $\sigma_p := |S_p \cap (S_p + h)|/p$ and apply the larger sieve, Theorem 2.3, with $\delta = 1$ and A replaced by $A \cap (A+h)$. We obtain the bound

$$|A \cap (A+h)| \ll \frac{Q}{\sum_{p \leqslant Q} \frac{\log p}{p \sigma_p} - \log X}, \tag{3.1}$$

provided that the denominator is positive.

Our assumption is that

$$\sum_{p \leqslant Q} \frac{\sigma_p \log p}{p} \leqslant (\frac{1}{2} - c) \log Q.$$

Since $4t + 1/t \geqslant 4$ for all $t > 0$, it follows that

$$\sum_{p \leqslant Q} \frac{\log p}{\sigma_p p} \geqslant 4 \log Q + O(1) - 4(\frac{1}{2} - c) \log Q = (2 + 4c) \log Q + O(1).$$

It is easy to check that the denominator of (3.1) is indeed positive, since $Q = X^{1/2-c/2}$. We obtain the bound

$$|A \cap (A+h)| \ll X^{1/2-c/2+o(1)} \ll X^{1/2-c/4}.$$

Since $|A \cap (A+h)| \geqslant |A|/2K$, the lemma follows. □

Before stating the next lemma, let us isolate a fact which will be needed in the proof. This is basically due to Pollard.

Lemma 3.8 (Pollard)**.** *Let $\varepsilon > 0$ be small, and let $S \subset \mathbb{Z}/p\mathbb{Z}$ be a non-empty set such that $|S| < (1-2\varepsilon)p$. Then there are at most $4\varepsilon|S| + 1$ values of $h \in \mathbb{Z}/p\mathbb{Z}$ such that $|S \cap (S+h)| \geqslant (1-\varepsilon)|S|$.*

Proof. This follows quickly from a well-known result of Pollard [17]. Writing N_i for the number of h such that $|S \cap (S+h)| \geqslant i$, Pollard's result in our setting implies that $N_1 + \cdots + N_r \geqslant r(2|S| - r)$ for all $2|S| - p \leqslant r \leqslant |S|$. Temporarily write H for the set of all $h \in \mathbb{Z}/p\mathbb{Z}$ such that $|S \cap (S+h)| \geqslant (1-\varepsilon)|S|$, and also let $R := |S| - 2\lfloor \varepsilon |S| \rfloor$ and $U := |S| - \lfloor \varepsilon |S| \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function. Then

$$N_{R+1} + \cdots + N_{|S|} \geqslant N_{R+1} + \cdots + N_U \geqslant |H|(U - R) = |H| \lfloor \varepsilon |S| \rfloor.$$

Pollard's result tells us that

$$N_1 + \cdots + N_R \geqslant R(2|S| - R) = |S|^2 - 4\lfloor \varepsilon |S| \rfloor^2.$$

On the other hand we trivially have

$$N_1 + \cdots + N_{|S|} = |S|^2.$$

Combining all these facts leads to the result provided that $|S| \geqslant 1/\epsilon$.

Alternatively, if $|S| < 1/\epsilon$ then $|S \cap (S+h)| \geqslant (1-\varepsilon)|S|$ only if $S \cap (S+h) = S$, in which case $S \cap (S+nh) = S$ for every n. Since S is a proper subset of $\mathbb{Z}/p\mathbb{Z}$, this can only happen when $h = 0$. □

We will also require a simple and standard averaging principle, the proof of which we include here for completeness.

Lemma 3.9. *Let $\varepsilon, \varepsilon'$ be real numbers with $0 < \varepsilon \leqslant \varepsilon'$. Let X be a finite set, let $(\lambda(x))_{x\in X}$ be nonnegative weights, and suppose that $f : X \to [0,1]$ is a function such that $\sum_{x\in X} \lambda(x) f(x) \geqslant (1-\varepsilon)\sum_{x\in X} \lambda(x)$.*

Let $X' \subset X$ be the set of all $x \in X$ such that $f(x) \geqslant 1-\varepsilon'$. Then $\sum_{x\in X'} \lambda(x) \geqslant (1-\frac{\varepsilon}{\varepsilon'})\sum_{x\in X} \lambda(x)$. In particular if $\sum_{x\in X} f(x) \geqslant (1-\varepsilon)|X|$ then there are at least $(1-\frac{\varepsilon}{\varepsilon'})|X|$ values of x such that $f(x) \geqslant 1-\varepsilon'$.

Proof. We have

$$(1-\varepsilon)\sum_{x\in X}\lambda(x) \leqslant \sum_{x\in X}\lambda(x)f(x) = \sum_{x\in X'}\lambda(x)f(x) + \sum_{x\in X\setminus X'}\lambda(x)f(x) \leqslant \sum_{x\in X'}\lambda(x) + (1-\varepsilon')\sum_{x\in X\setminus X'}\lambda(x).$$

Rearranging this inequality gives the first result. The second one follows by taking all the weights $\lambda(x)$ to be 1. □

Lemma 3.10. *Let $c > 0$ be a sufficiently small absolute constant. Let $H \subset [-X, X]$ be a set of size $X^{1/2-c/4}$, and let $Q = X^{1/2-c/2}$. Then there is some $h \in H$ such that*

$$\sum_{p\leqslant Q} \frac{\log p}{p}\frac{|S_p \cap (S_p + h)|}{p} < (\frac{1}{2} - c)\log Q.$$

Proof. Suppose not. Then certainly

$$\sum_{h\in H}\sum_{p\leqslant Q}\frac{\log p}{p}\frac{|S_p\cap(S_p+h)|}{p} \geqslant (\frac{1}{2}-c)|H|\log Q \geqslant (\frac{1}{2}-c)|H|\sum_{p\leqslant Q}\frac{\log p}{p}.$$

Write $\mathscr{P}$ for the set of primes $p \leqslant Q$ such that

$$\sum_{h\in H}\frac{|S_p\cap(S_p+h)|}{p} \geqslant (\frac{1}{2}-c^{1/2})|H|. \tag{3.2}$$

By Lemma 3.9 applied with X the set of primes $p \leqslant Q$, $\lambda(p) = \frac{\log p}{2p}|H|$, $f(p) = \frac{2}{|H|}\frac{1}{p}\sum_{h\in H}|S_p \cap (S_p + h)|$, $\epsilon = 2c$ and $\epsilon' = 2c^{1/2}$ we have[2]

$$\sum_{p\in\mathscr{P}} \frac{\log p}{p} \geqslant (1-c^{1/2}) \sum_{p\leqslant Q} \frac{\log p}{p}.$$

Note that $|S_p \cap (S_p+h)| \leqslant \frac{1}{2}(p+1)$ always. It also follows from Lemma 3.9 applied to the inequality (3.2) that, for all $p \in \mathscr{P}$, there is a set $H'_p \subset H$ with $|H'_p| \geqslant (1-c^{1/4})|H|$ such that $|S_p \cap (S_p+h)| \geqslant (\frac{1}{2}-c^{1/4})p$ for all $h \in H'_p$. On the other hand, by Lemma 3.8 it follows that $|H'_p (\mathrm{mod}\ p)| \leqslant 4c^{1/4}p + 1$, and so all but $c^{1/4}|H|$ elements of H reduce to lie in a set of size $4c^{1/4}p + 1 < \frac{1}{3}p$ modulo p, for all $p \in \mathscr{P}$, a set which satisfies $\sum_{p\in\mathscr{P}} \frac{\log p}{p} \geqslant (1-c^{1/2})(\log Q + O(1))$. We may apply the larger sieve, Theorem 2.3, to this situation, taking $\delta = 1 - c^{1/4}$ and $\sigma_p = 1/3$ for all $p \in \mathscr{P}$. This gives the bound

$$|H| \ll \frac{Q}{3(1-c^{1/4})^2(1-c^{1/2})(\log Q + O(1)) - \log X}$$

provided that the denominator is positive. If c is sufficiently small then the denominator will be positive with our choice of Q, namely $X^{1/2-c/2}$, and we get the bound $|H| \ll X^{1/2-c/2+o(1)}$. This is contrary to assumption. □

We may now conclude the proof of Proposition 3.5. As in the hypothesis of the proposition, let $A \subset [X]$ be a set such that $A(\mathrm{mod}\ p) \subset S_p$ for all p, where $|S_p| \leqslant \frac{1}{2}(p+1)$. Suppose additionally that $E(A,A) \geqslant |A|^3/K$. By Lemma 3.6 there is a set $H \subset [-X, X]$, $|H| \geqslant |A|/2K$, such that $|A \cap (A+h)| \geqslant |A|/2K$ for all $h \in H$. If $|H| < X^{1/2-c/4}$ then the proposition follows, so suppose this is not the case. Then Lemma 3.10 applies and we may conclude that there is an $h \in H$ such that

$$\sum_{p\leqslant Q} \frac{\log p}{p}\frac{|S_p \cap (S_p+h)|}{p} < (\frac{1}{2} - c)\log Q,$$

where $Q = X^{1/2-c/2}$. Finally, by Lemma 3.7, it follows that $|A| \ll KX^{1/2-c/4}$, thereby concluding the proof of the proposition. □

4. Sieving by intervals

Our aim in this section is to establish Theorem 1.2. We begin by recalling the statement of it.

Theorem 1.2. *Suppose that $\mathscr{A}$ is a set of integers and that, for each prime p, the set $\mathscr{A}(\mathrm{mod}\, p)$ lies in some interval I_p. Let $\varepsilon > 0$ be arbitrary. Then*

(i) *If $|I_p| \leqslant (1-\varepsilon)p$ for at least a proportion ε of the primes in each dyadic interval $[Z, 2Z]$ then $|\mathscr{A}[X]| \ll_\varepsilon (\log\log X)^{C\log(1/\varepsilon)}$, where $C > 0$ is some absolute constant;*

(ii) *If $|I_p| \leqslant \frac{p}{2}$ for all primes then $|\mathscr{A}[X]| \ll (\log\log X)^{\gamma+o(1)}$, where $\gamma = \frac{\log 18}{\log(3/2)} \approx 7.129$;*

(iii) *If $I_p = [\alpha p, \beta p]$ for all primes p and for fixed $0 \leqslant \alpha < \beta < 1$ (not depending on p) then $|\mathscr{A}| = O_{\beta-\alpha}(1)$;*

(iv) *There are examples of infinite sets $\mathscr{A}$ with $|I_p| \leqslant (\frac{1}{2}+\varepsilon)p$ for all p.*

[2]The alert reader will observe that our applications of Lemma 3.9 are slightly bogus, since we have $f(p) \leqslant (p+1)/p$ rather than $f(p) \leqslant 1$, as required in the lemma. This can be corrected by instead setting $\lambda(p) = \frac{\log p}{2p}\frac{p+1}{p}|H|$ and $f(p) = \frac{2}{|H|}\frac{1}{p+1}\sum_{h\in H}|S_p \cap (S_p+h)|$, which makes no essential difference to the conclusions about $\mathscr{P}$ and H_p.

The proof of parts (i) and (ii) relies on the following basic lemma.

Lemma 4.1. *Suppose that p is a prime, that $I_p \subset \mathbb{Z}/p\mathbb{Z}$ is an interval of length at most $(1-\varepsilon)p$, and that $A \subset [X]$ is a set with $A(\mathrm{mod}\, p) \subset I_p$. Then there is some integer k, $1 \leqslant k \leqslant \lceil 2/\varepsilon^2 \rceil$, such that*

$$|\sum_{x \leqslant X} 1_A(x) e(kx/p)| \geqslant \varepsilon |A|/32.$$

If $|I_p| \leqslant p/2$ then we have the following more precise conclusion: there is an integer $k \in \{1,2\}$ such that

$$|\sum_{x \leqslant X} 1_A(x) e(kx/p)| \geqslant |A|/3.$$

Proof. We claim that there is a 1-periodic real-valued function

$$f(\theta) = 1 + \sum_{0<|k|\leqslant \lceil 2/\varepsilon^2 \rceil} c_k e(k\theta)$$

such that $f(\theta) \leqslant 0$ when $|\theta| \geqslant \varepsilon/2$.

To construct $f(\theta)$, consider first the convolution $\psi(\theta) := 1_{|\theta|\leqslant \varepsilon/4} * 1_{|\theta|\leqslant \varepsilon/4} = \int_{\mathbb{R}/\mathbb{Z}} 1_{|\theta-\phi|\leqslant\varepsilon/4} 1_{|\phi|\leqslant\varepsilon/4} d\phi$. We have

$$|\hat{\psi}(k)| = |\widehat{1_{|\phi|\leqslant\varepsilon/4}}(k)|^2 = \Big| \int_{\mathbb{R}/\mathbb{Z}} 1_{|\phi|\leqslant\varepsilon/4} e(-k\phi) d\phi \Big|^2 \leqslant \min(\varepsilon, \frac{1}{\pi|k|})^2.$$

From the Fourier inversion formula it follows that

$$\frac{8}{\varepsilon^2}\psi(\theta) = 2 + \sum_{k\neq 0} c_k e(k\theta),$$

where $|c_k| \leqslant \min(8, \frac{1}{\varepsilon^2|k|^2})$. Furthermore, by construction, $\psi(\theta) = 0$ for $|\theta| \geqslant \varepsilon/2$. Define

$$f(\theta) := 1 + \sum_{0<|k|\leqslant K} c_k e(k\theta),$$

where $K := \lceil 2/\varepsilon^2 \rceil$. Since

$$\sum_{|k|>K} |c_k| \leqslant \sum_{|k|\geqslant K+1} \frac{1}{\varepsilon^2|k|^2} \leqslant \frac{2}{\varepsilon^2 K} \leqslant 1,$$

it follows that f has the required properties. Now there is some $\beta \in [0,1]$ (depending on I_p) such that $\|\frac{x}{p} + \beta\| \geqslant \varepsilon/2$ whenever $x \in A$, where $\|\cdot\|$ denotes distance to the nearest integer. This means that $f(\frac{x}{p} + \beta) \leqslant 0$, and so

$$1 + \sum_{0<|k|\leqslant \lceil 2/\varepsilon^2 \rceil} c_k e(k(\frac{x}{p} + \beta)) \leqslant 0.$$

It follows that

$$|A| \leqslant - \sum_{0<|k|\leqslant \lceil 2/\varepsilon^2 \rceil} c_k \sum_{x\leqslant X} 1_A(x) e(k(\frac{x}{p} + \beta)).$$

Using the triangle inequality, one obtains

$$|A| \leqslant \sum_{0<|k|\leqslant \lceil 2/\varepsilon^2 \rceil} |c_k| |\sum_{x\leqslant X} 1_A(x) e(kx/p)|.$$

To conclude the proof of the lemma, we observe that

$$\sum_{0<|k|\leqslant K} |c_k| \leqslant \frac{32}{\varepsilon},$$

an estimate that follows upon splitting into the ranges $0 < |k| \leqslant 1/\varepsilon$ and $|k| > 1/\varepsilon$.

For the second statement, simply note that the function $f(\theta) = 1 - 2\cos\theta + \cos 2\theta$ satisfies $f(\theta) \leqslant 0$ when $|\theta| \leqslant \pi/2$; rewriting the left-hand side as $2\cos\theta(\cos\theta - 1)$, this becomes clear. The rest of the argument proceeds as before. □

We turn now to the proof of Theorem 1.2 (i). The general scheme of the argument, and in particular the use of Vinogradov's estimate (Proposition 4.3 below) was suggested to us by Jean Bourgain. We are very grateful to him for allowing us to include it here. The heart of the matter is the proof of the following lemma, from which Theorem 1.2 (i) follows rather easily by an iteration argument (or equivalently induction on X).

Lemma 4.2. *Suppose that $A \subset [X]$ and that $A(\mathrm{mod}\ p)$ lies in an interval I_p of length at most $(1-\varepsilon)p$, for at least ε of all primes in each dyadic interval. Suppose that $X > X_0(\varepsilon)$. Then there is a subinterval of $[X]$ of length $\exp(\log^{7/10} X)$ containing at least $c\varepsilon^5|A|$ points of A, where $c > 0$ is a small absolute constant.*

Indeed, before proving this lemma let us explain how it implies Theorem 1.2 (i). We set $X_0 = X$ and $A_0 = \mathscr{A}[X]$, and by repeated application of the lemma we construct numbers X_i and sets A_i such that $A_i \subset [X_i]$, $A_i(\mathrm{mod}\ p)$ lies in an interval I_p of length at most $(1-\varepsilon)p$, for at least ε of all primes in each dyadic interval, $\log X_{i+1} = \log^{7/10} X_i$ and $|A_{i+1}| \geqslant c\varepsilon^5|A_i|$. This procedure terminates when we first have $X_{i+1} \leqslant X_0(\varepsilon)$, which will happen after $\ll \log\log\log X$ iterations. Consequently we have $|\mathscr{A}[X]| \leqslant (c^{-1}\varepsilon^{-5})^{O(\log\log\log X)} X_0(\varepsilon) \ll_\varepsilon (\log\log X)^{C\log(1/\varepsilon)}$, as claimed[3] in Theorem 1.2.

Proof of Lemma 4.2. Suppose that p is a prime such that $A(\mathrm{mod}\ p) \subset I_p$. By Lemma 4.1, there is some k, $1 \leqslant k \leqslant \lceil 2/\varepsilon^2 \rceil$, such that

$$|\sum_{a\in A} e(ka/p)| \geqslant \varepsilon|A|/32. \tag{4.1}$$

Let Y, $1 \ll Y \ll X$, be a parameter to be selected later (we will in fact take $Y = \exp(c\log^{7/10} X)$). We may choose a single k so that the preceding estimate holds for $\gg \varepsilon^2$ of the primes in $[Y, 2Y]$ for which we know that $A(\mathrm{mod}\ p) \subset I_p$, that is for $\gg \varepsilon^3$ of all the primes in $[Y, 2Y]$. Now we use the following fact: there is a weight function $w : [Y, 2Y] \to \mathbb{R}_{\geqslant 0}$ such that

(i) $w(p) \geqslant 1$ for all primes $p \in [Y, 2Y]$;
(ii) $\sum_{Y\leqslant n\leqslant 2Y} w(n) \leqslant 10\pi(Y)$;
(iii) $w(n) = \sum_{d|n:d\leqslant Y^{1/2}} \lambda_d$, where $\sum_{d\leqslant Y} \frac{|\lambda_d|}{d} \ll \log^3 Y$.

Such a function can be constructed in the form $w(n) = (\sum_{d|n} \mu(d)\psi(\frac{\log d}{\log Y}))^2$, where $\psi \in C^\infty(\mathbb{R})$ is supported on $|x| \leqslant \frac{1}{4}$, is bounded in absolute value by 1, and $\psi(0) = 1$. Property (i) is then clear,

[3]As we have written things, we need to have $X_0(\varepsilon) = \exp(C\log^{100}(1/\epsilon))$ (say) in order for the parameter Y in the proof of Lemma 4.2 to be large enough. But we remark that by taking more care of the final iterations in the proof of Theorem 1.2 (i), one could obtain a bound $|\mathscr{A}[X]| \ll (\log\log X)^{C\log(1/\varepsilon)}$ for all X, with an absolute implied constant (not depending on ε).

whilst bound (ii) can be verified by expanding out and interchanging the order of summation. To check (iii), we note that it is clear that $|\lambda_d| \leqslant \sum_{[d_1,d_2]=d} 1 \leqslant \tau_3(d)$, the number of ways of writing d as a product of three nonnegative integers. The claimed bound is then an easy exercise. It follows from (4.1) and the above properties that

$$\sum_{Y\leqslant n\leqslant 2Y} w(n)|\sum_{a\in A} e(ka/n)|^2 \geqslant c\varepsilon^5\pi(Y)|A|^2. \tag{4.2}$$

Expanding out and applying the triangle inequality yields

$$\sum_{a,a'\in A} |\sum_{Y\leqslant n\leqslant 2Y} w(n)e(\frac{k(a-a')}{n})| \geqslant c\varepsilon^5\pi(Y)|A|^2. \tag{4.3}$$

We now claim that, if Y is chosen judiciously, the contribution to this from those pairs a, a' with $|a-a'| \geqslant Y^{10}$ (say) can be ignored. Indeed suppose, on the contrary, that

$$|\sum_{Y\leqslant n\leqslant 2Y} w(n)e(\frac{x}{n})| \geqslant \frac{c}{100}\varepsilon^5\pi(Y), \tag{4.4}$$

for some $x := k(a-a')$ satisfying $Y^{10} \leqslant x \ll \varepsilon^{-2}X$. By property (iii) of $w(n)$ and the triangle inequality, this implies that

$$\sum_{d\leqslant Y^{1/2}} |\lambda_d| |\sum_{Y/d\leqslant n'\leqslant 2Y/d} e(\frac{x}{dn'})| \geqslant \frac{c}{100}\varepsilon^5\pi(Y).$$

By the upper bound (iii) for $\sum |\lambda_d|/d$ it follows that there is some $d \leqslant Y^{1/2}$ such that

$$|\sum_{Y/d\leqslant n'\leqslant 2Y/d} e(\frac{x}{dn'})| \gg \varepsilon^5 \log^{-4} Y \frac{Y}{d}. \tag{4.5}$$

At this point we invoke the following powerful estimate of Vinogradov.

Proposition 4.3. *Let $\delta > 0$ be small and Y be large, and suppose that $x \geqslant Y^5$. Suppose that $|\sum_{Y\leqslant n\leqslant 2Y} e(x/n)| \geqslant \delta Y$. Then $x \geqslant \exp(c\log^{3/2} Y/\log^{1/2}(1/\delta))$.*

Proof. Using e.g. Theorem 8.25 of Iwaniec and Kowalski [13], one obtains that

$$|\sum_{Y\leqslant n\leqslant 2Y} e(x/n)| \ll Ye^{-c\frac{\log^3 Y}{\log^2 x}}.$$

Thus we must have $1/\delta \gg \exp(c\log^3 Y/\log^2 x)$, from which the conclusion of the proposition quickly follows. □

Applying this Proposition to (4.5) leads to a contradiction unless

$$\frac{x}{d} > \exp\left(c\frac{\log^{3/2} Y}{(\log\log Y)^{1/2} + (\log(1/\epsilon))^{1/2}}\right),$$

which would imply that $X \gg \epsilon^2 x > \exp(c(\log^{3/2} Y)/((\log\log Y)^{1/2} + (\log(1/\epsilon))^{1/2}))$. With $Y = \exp(c\log^{7/10} X)$ and $X \geqslant \exp(C\log^{100}(1/\epsilon))$, say, this will not be so. It follows that we were wrong to assume (4.4), and so indeed the contribution to (4.3) of those pairs a, a' with $|a - a'| \geqslant Y^{10}$ may be ignored. Thus we have

$$\sum_{\substack{a,a' \in A \\ |a-a'| \leqslant Y^{10}}} \Big| \sum_{Y \leqslant n \leqslant 2Y} w(n) e(\frac{k(a-a')}{n}) \Big| \geqslant \frac{c}{2}\varepsilon^5 \pi(Y)|A|^2. \tag{4.6}$$

Finally, we may apply the trivial bound to the inner sum, recalling from (ii) above that $\sum_{Y \leqslant n \leqslant 2Y} w(n) \leqslant 10\pi(Y)$. We obtain

$$\sum_{\substack{a,a' \in A \\ |a-a'| \leqslant Y^{10}}} 1 \gg \varepsilon^5 |A|^2,$$

which implies that there is a subinterval of $[X]$ of length Y^{10} containing $\gg \varepsilon^5|A|$ elements of A. This concludes the proof of Lemma 4.2, and hence of Theorem 1.2 (i). □

Proof of Theorem 1.2 (ii) (sketch). We proceed as above, with the following changes.

- Use the second conclusion of Lemma 4.1 to conclude that there is some $k \in \{1, 2\}$ such that
$$\sum_{Y \leqslant p \leqslant 2Y} |\sum_{a \in A} e(ka/p)|^2 \geqslant \frac{1}{18}(\pi(2Y) - \pi(Y))|A|^2.$$
This takes the place of (4.2).
- Expand out as in (4.3) to get
$$\sum_{a,a' \in A} |\sum_{Y \leqslant p \leqslant 2Y} e(\frac{k(a-a')}{p})| \geqslant \frac{1}{18}(\pi(2Y) - \pi(Y))|A|^2.$$
- Choose $Y = \exp(\log^{2/3+o(1)} X)$, and use Jutila [14, Theorem 2] (which is a Vinogradov-type estimate for $\sum_{p \leqslant P} e(x/p)$) to show that the contribution from those pairs with $|a - a'| \geqslant Y^{10}$ can be ignored, so we have
$$\sum_{\substack{a,a' \in A \\ |a-a'| \leqslant Y^{10}}} |\sum_{Y \leqslant p \leqslant 2Y} e(\frac{k(a-a')}{p})| \geqslant (\frac{1}{18} - o(1))\pi(Y)|A|^2.$$
- Conclude that there is some interval of length $\sim Y^{10}$ containing at least $(\frac{1}{18} - o(1))|A|$ points of A, and proceed iteratively as before.

Remark. We could have used Jutila's bound in the proof of Theorem 1.2 (i) as well, instead of using the weight function w . We chose not to do this in the interests of self-containment and of variety. Note that Jutila's paper predates Vaughan's identity [19] for prime number sums, and his argument would be a little more accessible if this device were used. A model for such an argument may be found in the paper of Granville and Ramaré [9].

Proof of Theorem 1.2 (iii). This is essentially a consequence of Jutila [14, Corollary, p126]. A slight variant of that Corollary shows that the number of $p \in [x^{1/2}, 2x^{1/2}]$ with $\alpha \leqslant \{x/p\} \leqslant \beta$ is $\sim (\beta-\alpha)(\pi(2x^{1/2}) - \pi(x^{1/2}))$, and so all elements of $\mathscr{A}$ are bounded by $O_{\beta-\alpha}(1)$.

Proof of Theorem 1.2 (iv). We take $\mathscr{A}$ to consist of the numbers $a_i = \prod_{p \leqslant X_i} p$, for some extremely rapidly-growing sequence $X_1 < X_2 < \dots$. Given a prime p, suppose that $X_i \leqslant p \leqslant X_{i+1}$. Then $a_{i+1}, a_{i+2}, \dots$ all reduce to zero (mod p), and so $\mathscr{A}(\mathrm{mod}\ p) = \{0, a_1, \dots, a_i\}$. By choosing the X_i sufficiently rapidly growing we may ensure that $0 < a_1 < \dots < a_{i-1} < \varepsilon p$. Regardless of the value of $a_i (\mathrm{mod}\ p)$ (which we cannot usefully control) the set $\mathscr{A}(\mathrm{mod}\ p)$ will be contained in some interval of length at most $(\frac{1}{2} + \varepsilon)p$.

Remark. With $\mathscr{A}$ as constructed above, the shape of $|\mathscr{A}[X]|$ is $\log_* X$. Thus there is still a considerable gap between the bound of (i) and the construction given here. We expect, however, that the correct bound in (i) is of $\log_*$ type, which would follow assuming vaguely sensible conjectures on exponential sums $\sum_{n \leqslant Y} e(x/n)$. If, for example, the conclusion of Proposition 4.3 were instead that $x \geqslant \exp(Y^{1/10})$ then we would get a $\log_*$-type bound on $\mathscr{A}[X]$ in this case.

5. Sieving by arithmetic progressions

In this section we shall prove Theorem 1.3, whose statement was as follows.

Theorem 1.3. *Let $\varepsilon > 0$. Suppose that $\mathscr{A}$ is a set of integers and that, for each prime $p \leqslant X^{1/2}$, the set $\mathscr{A}(\mathrm{mod}\, p)$ lies in some arithmetic progression S_p of length $(1-\varepsilon)p$. Then $|\mathscr{A}[X]| \ll_\varepsilon X^{1/2-\varepsilon'}$, where $\varepsilon' > 0$ depends on ε only.*

Suppose to begin with that $|S_p| = \frac{1}{2}(p+1)$ for each odd prime p. Then (dilating S_p to the interval $\{1, \dots, \frac{1}{2}(p+1)\}$, which preserves the additive energy)

$$E_p(S_p, S_p) = (\frac{p+1}{2})^2 + 2\sum_{h=1}^{\frac{1}{2}(p-1)} (\frac{p+1}{2} - h)^2 \geqslant \frac{p^3}{12}.$$

Thus Theorem 1.1 is applicable with the choice $\delta = 1/48$, and the result follows in this case.

Now we turn to the proof of Theorem 1.3 for arbitrary $\varepsilon > 0$. We begin with a result which should be compared to Corollary 3.3, but which is simpler to state and prove than that result.

Lemma 5.1. *Let $\varepsilon > 0$ be small, and suppose that $A \subset [X]$ is a set and $\mathscr{P} \subset [X^{1/2}]$ is a set of primes such that $|\mathscr{P}| \geqslant 2/\varepsilon^2$. If $S_p \subset \mathbb{Z}/p\mathbb{Z}$ is an arithmetic progression of length at most $(1-\varepsilon)p$, and if $A(\mathrm{mod}\, p) \subset S_p$ for each prime $p \in \mathscr{P}$, then $E(A,A) \gg \frac{\varepsilon^4 |A|^4}{X} |\mathscr{P}|$, where the constant implicit in the $\gg$ notation is absolute.*

Proof. In view of Lemma 3.2, and our assumption that $|\mathscr{P}| \geqslant 2/\varepsilon^2$, it will suffice to show that

$$E_p(A,A) - \frac{|A|^4}{p} \gg \frac{\varepsilon^4 |A|^4}{p}$$

for each prime $p \in \mathscr{P}$ that is greater than $2/\varepsilon^2$ (that being a positive proportion of all the primes in $\mathscr{P}$). As in the proof of Corollary 3.3 we have

$$E_p(A,A) = p^3 \sum_{r \in \mathbb{Z}/p\mathbb{Z}} \big|\frac{1}{p}\sum_{x \leqslant X} 1_A(x)e(-xr/p)\big|^4.$$

The contribution from the $r = 0$ term is evidently equal to $|A|^4/p$. By Lemma 4.1 (which was stated in the case S_p is an interval, but may easily be adapted to the case S_p a progression by dilation), if $p > \lceil 2/\varepsilon^2 \rceil$ then there is some nonzero r satisfying $|\sum_{x \leqslant X} 1_A(x)e(rx/p)| \geqslant \varepsilon|A|/32$. The result follows immediately. □

The other major ingredient that we shall need is an analogue of Lemma 3.10 that applies when the sets S_p have size at most $(1-\varepsilon)p$, rather than size at most $\frac{1}{2}(p+1)$ as in that lemma. The following result provides this.

Lemma 5.2. *Let $c > 0$ be a sufficiently small absolute constant. Let $H \subset [-X, X]$ be a set of size $X^{1/2-c/4}$, and let $Q = X^{1/2-c/2}$. Suppose that for each prime $p \in \mathscr{P}'$ we have a subset $S_p \subset \mathbb{Z}/p\mathbb{Z}$ such that $\frac{1}{10}p \leqslant |S_p| \leqslant (1-\varepsilon)p$, where $\mathscr{P}'$ is a subset of the primes $p \leqslant Q$ satisfying $\sum_{p \in \mathscr{P}'} \frac{\log p}{p} \geqslant \frac{1}{3}\log Q$. Then there is some $h \in H$ such that*

$$\sum_{p \in \mathscr{P}'} \frac{\log p}{p}\frac{|S_p \cap (S_p + h)|}{p} < (1 - c\varepsilon)\sum_{p \in \mathscr{P}'} \frac{\log p}{p}\frac{|S_p|}{p}.$$

Proof. The proof is quite close to the proof of Lemma 3.10, so we shall give a fairly brief account. If the conclusion were false then we would have

$$\sum_{h \in H}\sum_{p \in \mathscr{P}'} \frac{\log p}{p}\frac{|S_p \cap (S_p + h)|}{p} \geqslant (1 - c\varepsilon)|H|\sum_{p \in \mathscr{P}'} \frac{\log p}{p}\frac{|S_p|}{p}.$$

In this case, if we let $\mathscr{P}$ denote the subset of primes $p \in \mathscr{P}'$ for which

$$\sum_{h \in H} \frac{|S_p \cap (S_p + h)|}{p} \geqslant (1 - c^{1/2}\varepsilon)|H|\frac{|S_p|}{p},$$

then applying Lemma 3.9 with $X = \mathscr{P}'$, $\lambda_p = \frac{\log p}{p^2}|H||S_p|$ and $f(p) = |H|^{-1}\sum_{h \in H} \frac{|S_p \cap (S_p+h)|}{|S_p|}$ yields

$$\sum_{p \in \mathscr{P}} \frac{\log p}{p}\frac{|S_p|}{p} \geqslant (1 - c^{1/2})\sum_{p \in \mathscr{P}'} \frac{\log p}{p}\frac{|S_p|}{p} \geqslant \frac{1}{40}\log Q.$$

As in the proof of Lemma 3.10, it also follows from Lemma 3.9 that, for all $p \in \mathscr{P}$, there is a set $H'_p \subset H$ with $|H'_p| \geqslant (1 - c^{1/4})|H|$ such that $|S_p \cap (S_p + h)| \geqslant (1 - c^{1/4}\varepsilon)|S_p|$ for all $h \in H'_p$. Finally, using Lemma 3.8 applied to S_p (and with ϵ replaced by $c^{1/4}\epsilon$) it follows that $|H'_p (\mathrm{mod}\ p)| \leqslant 4c^{1/4}\varepsilon p + 1$, and so all but $c^{1/4}|H|$ elements of H reduce to lie in a set of size $4c^{1/4}\varepsilon p + 1 < \frac{1}{100}p$ modulo p, for all $p \in \mathscr{P}$, a set which satisfies $\sum_{p \in \mathscr{P}} \log p/p \geqslant \frac{1}{40}\log Q$. We may apply the larger sieve, Theorem 2.3, to this situation, taking $\delta = 1 - c^{1/4}$ and $\sigma_p = 1/100$ for all $p \in \mathscr{P}$. This gives the bound

$$|H| \ll \frac{Q}{100(1 - c^{1/4})^2(1/40)\log Q - \log X}$$

provided that the denominator is positive. If c is sufficiently small then this will be so with our choice of Q, namely $X^{1/2-c/2}$, and we get the bound $|H| \ll X^{1/2-c/2+o(1)}$. This is contrary to assumption. □

Now we can prove Theorem 1.3, by applying the foregoing lemmas repeatedly to the intersection of A (and of the sets S_p) with shifted copies of itself. The key point here is that, since any subset of A will lie in the arithmetic progression S_p when reduced modulo p, we can use Lemma 5.1 throughout this "intersecting process" to obtain a lower bound on additive energy. In particular, we don't need to keep track of any "uniformity of fibres" throughout the process. Eventually we will obtain a subset of A that has cardinality quite close to $|A|$, but lies modulo p in a multiply intersected copy of S_p having size $< (1/2-c)p$ (for most primes p). The theorem will then follow immediately from the larger sieve.

Proof of Theorem 1.3. We assume that ε is small, and that X is sufficiently large in terms of ε. This is certainly permissible for proving the theorem. We will prove that $|A| < X^{1/2-c(\varepsilon)}$, where $c(\varepsilon) = K^{-1/\varepsilon}$ for a large absolute constant $K > 0$. Suppose, for a contradiction, that $|A| \geqslant X^{1/2-c(\varepsilon)}$. We proceed iteratively, setting $A_0 = A$ and constructing a sequence of sets $A_i \subset A$, $i = 1, 2, 3, \dots$ such that

$$|A_i| \geqslant X^{1/2-3^i K^{-1/\varepsilon}}. \tag{5.1}$$

The sets A_i will satisfy $A_i (\mathrm{mod}\ p) \subset S_p^i$, where $S_p^i \subset S_p$, and where

$$\sum_{p \leqslant Q} \frac{\log p}{p} \frac{|S_p^i|}{p} < (1 - c\varepsilon/2)^i (\log Q + O(1)). \tag{5.2}$$

Here c is the absolute constant from Lemma 5.2, and $Q = X^{1/2-c/2}$. After $O(1/\varepsilon)$ steps we will, in particular, have

$$\sum_{p \leqslant Q} \frac{\log p}{p} \frac{|S_p^i|}{p} < (1/2 - c) \log Q,$$

and therefore, writing $\sigma_p^i := |S_p^i|/p$, we will have

$$\sum_{p \leqslant Q} \frac{\log p}{\sigma_p^i p} \geqslant 4 \log Q + O(1) - 4(\frac{1}{2} - c) \log Q = (2 + 4c) \log Q + O(1),$$

as argued in the proof of Lemma 3.7. Using the larger sieve, this implies that $|A_i| \ll X^{1/2-c/2}$, which contradicts the lower bound (5.1) provided that K was chosen large enough.

We will show how the set A_{i+1} is obtained from A_i, and verify that it satisfies the size bound (5.1) and that its reductions modulo primes p satisfy the bound (5.2). Firstly, if $A_i \subset A$ satisfies (5.1) then Lemma 5.1 implies that

$$E(A_i, A_i) \gg \frac{\varepsilon^4 |A_i|}{X^{1/2} \log X} |A_i|^3 \gg \frac{\varepsilon^4 X^{-3^i K^{-1/\varepsilon}}}{\log X} |A_i|^3 \geqslant 2X^{-2 \cdot 3^i K^{-1/\varepsilon}} |A_i|^3,$$

provided that X is large enough in terms of ε. Using Lemma 3.6, it follows that there is a set $H \subset [-X, X]$ such that $|H| \geqslant |A_i| X^{-2 \cdot 3^i K^{-1/\varepsilon}} \geqslant X^{1/2-3^{i+1} K^{-1/\varepsilon}}$, and such that

$$|A_i \cap (A_i + h)| \geqslant X^{1/2-3^{i+1} K^{-1/\varepsilon}}$$

for all $h \in H$. The set A_{i+1} will be of the form $A_i \cap (A_i + h)$, for suitably chosen $h \in H$. Note that any such choice will indeed satisfy the size bound (5.1). In view of (5.2), we may assume that the sets

S_p^i satisfy

$$(1/2-c)\log Q \leqslant (1-c\varepsilon/2)^{i+1}(\log Q + O(1)) \leqslant \sum_{p\leqslant Q}\frac{\log p}{p}\frac{|S_p^i|}{p} < (1-c\varepsilon/2)^i(\log Q + O(1)),$$

the lower bound holding because if it failed we could simply set $A_{i+1}=A_i$. Now let $\mathscr{P}'$ denote the set of primes $p\leqslant Q$ for which $|S_p^i|\geqslant \frac{1}{10}p$. We must have

$$\sum_{p\in\mathscr{P}'}\frac{\log p}{p}\geqslant\frac{1}{3}\log Q \qquad \text{and} \qquad \sum_{p\in\mathscr{P}'}\frac{\log p}{p}\frac{|S_p^i|}{p}\geqslant\frac{1}{2}\sum_{p\leqslant Q}\frac{\log p}{p}\frac{|S_p^i|}{p},$$

say, because otherwise the lower bound we just assumed would be violated. Thus we can apply Lemma 5.2, deducing that for some $h\in H$ we have

$$\sum_{p\in\mathscr{P}'}\frac{\log p}{p}\frac{|S_p^i\cap(S_p^i+h)|}{p} < (1-c\varepsilon)\sum_{p\in\mathscr{P}'}\frac{\log p}{p}\frac{|S_p^i|}{p} \leqslant \sum_{p\in\mathscr{P}'}\frac{\log p}{p}\frac{|S_p^i|}{p} - \frac{c\varepsilon}{2}\sum_{p\leqslant Q}\frac{\log p}{p}\frac{|S_p^i|}{p}.$$

Finally, if we set $A_{i+1}=A_i\cap(A_i+h)$ for this choice of h, and correspondingly set $S_p^{i+1}=S_p^i\cap(S_p^i+h)$, then the upper bound (5.2) will indeed be satisfied. □

6. Robustness of the inverse large sieve problem

In this section we prove Theorem 1.4. Let us begin by recalling the statement.

Theorem 1.4. *Let $X_0\in\mathbb{N}$, and let $\varepsilon>0$. Let $X\in\mathbb{N}$ be sufficiently large in terms of X_0 and ε, and suppose that $H\leqslant X^{1/8}$. Suppose that $A,B\subset[X]$ and that $|A(\mathrm{mod}\ p)|+|B(\mathrm{mod}\ p)|\leqslant p+1$ for all $p\in[X_0,X^{1/4}]$. Then, for some absolute constant $c>0$, one of the following holds:*

(i) (Better than large sieve) *Either $|A\cap[X^{1/2}]|$ or $|B\cap[X^{1/2}]|$ is $\leqslant X^{1/4-c\varepsilon^3}$;*

(ii) (Behaviour with quadratics) *Given any two rational quadratics ψ_A,ψ_B of height at most H, either $|A\setminus\psi_A(\mathbb{Q})|$ and $|B\setminus\psi_B(\mathbb{Q})|\leqslant HX^{1/2-c}$, or else at least one of $|A\cap\psi_A(\mathbb{Q})|$ and $|B\cap\psi_B(\mathbb{Q})|$ is bounded above by $HX^{1/4+\varepsilon}$.*

The proof of Theorem 1.4 requires a number of preliminary ingredients, which we assemble now. We start with some results concerning *quasisquares*, that is to say squarefree integers that are quadratic residues modulo "many" primes (see, for example, [2, Section 12.14]). The first result is not the one actually required later on (which is Lemma 6.2), but it may be of independent interest and its proof motivates the proof of the result needed later.

Lemma 6.1. *Let $\eta>0$, and suppose that $Y\geqslant Z>2$. Suppose that $\mathscr{P}\subset[Z,2Z]$ is a set consisting of at least $\eta Z/\log Z$ of the primes in $[Z,2Z]$. Then the number of squarefree $q\in[1,Y]$ such that $\left(\frac{q}{p}\right)=1$ for all $p\in\mathscr{P}$ is at most $8(6\log Y/\eta)^{3\log Y/\log Z}$.*

Proof. First of all, let $\mathscr{Q}$ denote the set of all q that are squarefree, lie in $[1,Y]$ and are a quadratic residue modulo all $p\in\mathscr{P}$. Each $q\in\mathscr{Q}$ is either congruent to 1(mod 4), or it is congruent to 2 or 3(mod 4). Let us assume that at least half of the elements of $\mathscr{Q}$ are congruent to 2 or 3(mod 4), and henceforth redefine $\mathscr{Q}$ to consist of those elements only, and aim to show that $|\mathscr{Q}|\leqslant 4(6\log Y/\eta)^{3\log Y/\log Z}$. (The proof when at least half of the elements are congruent to 1(mod 4) is essentially the same.)

Let $k \geqslant 3$ be the smallest integer for which $Z^k > Y^2$. Then if $n \in [Z^k, 2^k Z^k]$ is any product of k distinct primes from $\mathscr{P}$, and if $q \in \mathscr{Q}$, we have that the Jacobi symbol $\left(\frac{q}{n}\right) = \prod_{p|n} \left(\frac{q}{p}\right) = 1$. Let $\mathscr{S}$ be the set of all such n; then clearly

$$|\mathscr{S}| = \binom{|\mathscr{P}|}{k} \geqslant \frac{|\mathscr{P}|^k}{k^k} \geqslant \left(\frac{\eta}{k \log Z}\right)^k Z^k. \tag{6.1}$$

(Of course this is only true if $|\mathscr{P}| \geqslant k$, but otherwise the bound we shall derive is trivial anyway.)

Finally, note that if q is squarefree and congruent to 2 or 3(mod 4), and if $n \in \mathscr{S}$ (so, in particular, n is odd), then $\left(\frac{q}{n}\right) = \left(\frac{4q}{n}\right) = \chi_{4q}(n)$, where $\chi_{4q}(n)$ is a primitive character modulo $4q$. (It is the primitive quadratic character corresponding to the fundamental discriminant $4q$.) The multiplicative form of the large sieve [13, Theorem 7.13] implies that

$$\sum_{\substack{q \leqslant Q, \\ q \equiv 2 \text{ or } 3 (\mathrm{mod}\ 4), \\ q \text{ squarefree}}} \left| \sum_{n \in \mathscr{S}} a_n \chi_{4q}(n) \right|^2 \leqslant (16Q^2 + N) \sum_{n \in \mathscr{S}} |a_n|^2,$$

for any set $\mathscr{S} \subset [N]$ and for any Q and any coefficients a_n. Applying this with our particular set $\mathscr{S}$, and with $a_n = 1$, yields

$$|\mathscr{Q}||\mathscr{S}|^2 \leqslant (16Y^2 + 2^k Z^k)|\mathscr{S}| < 2^{k+2} Z^k |\mathscr{S}|,$$

and therefore by (6.1)

$$|\mathscr{Q}| \leqslant \frac{2^{k+2} Z^k}{|\mathscr{S}|} \leqslant 4(\frac{2k \log Z}{\eta})^k.$$

Noting that $k \leqslant \frac{2 \log Y}{\log Z} + 1 \leqslant \frac{3 \log Y}{\log Z}$, the result follows. □

Remarks. The conclusion of Lemma 6.1 is nontrivial when Y is any fixed power of Z, and even for somewhat larger values of Y. It seems to us that the bound obtained here is stronger than could (straightforwardly) be obtained using the real character sum estimate of Heath-Brown [11], which comes with an unspecified factor of $(QN)^\varepsilon$.

Now we present the result we need later on. Of course more general statements are possible, but we leave their formulation as an exercise to the interested reader.

Lemma 6.2. *Suppose that $Z \geqslant Y^{1/5}$ are large. The number of squarefree $q \in [1, Y]$ which are squares modulo 95% of the primes in $[Z, 2Z]$ is $O(\log^{10} Z)$. Furthermore, all of these q apart from $q = 1$ are at least Y^ϱ, for a certain absolute constant $\varrho > 0$.*

Proof. The argument for the first part closely follows the preceding. Write $\mathscr{Q}$ for the set of all squarefree $q \in [1, Y]$ which are squares modulo at least 95% of the primes $p \in [Z, 2Z]$. Write $\mathscr{P}_q$ for the set of these primes; note carefully that $\mathscr{P}_q$ may depend on q. Write $\mathscr{S}$ for the set of products of 10 distinct primes from $[Z, 2Z]$, and write $\mathscr{S}_q$ for the set of products of 10 distinct primes from $\mathscr{P}_q$. Note that $\left(\frac{q}{n}\right) = 1$ whenever $n \in \mathscr{S}_q$. Furthermore,

$$|\mathscr{S}_q| = \binom{|\mathscr{P}_q|}{10} \geqslant (1 - o(1))(1 - \frac{1}{20})^{10} \binom{|\mathscr{P}|}{10} > 0.59 \binom{|\mathscr{P}|}{10} = 0.59 |\mathscr{S}|$$

for every $q \in \mathscr{Q}$ (the key point here is that $0.59 > 0.5$). It follows that for every $q \in \mathscr{Q}$ we have $\sum_{n \in \mathscr{S}} \left(\frac{q}{n}\right) \geqslant 0.18 |\mathscr{S}|$. The proof now concludes as before.

To prove the second part, we use a form of the prime number theorem for the real character $\chi(m) = \left(\frac{q}{m}\right)$ (which, provided $q > 1$ is squarefree, is always a non-principal character of conductor at most $4q$). This tells us (see e.g. Theorem 7 in Gallagher's paper [7]) that

$$\sum_{Z \leqslant m \leqslant 2Z} \Lambda(m)\chi(m) = O(Z \max(e^{-c\sqrt{\log Z}}, e^{-c \log Z/\log q}))$$

if χ has no exceptional zero, and

$$\sum_{Z \leqslant m \leqslant 2Z} \Lambda(m)\chi(m) = \frac{Z^\beta - (2Z)^\beta}{\beta} + O(Z \max(e^{-c\sqrt{\log Z}}, e^{-c \log Z/\log q})),$$

if χ does have an exceptional zero $\beta \in (\frac{1}{2}, 1)$. Either way, we have the 1-sided inequality

$$\sum_{Z \leqslant m \leqslant 2Z} \Lambda(m)\chi(m) \leqslant O(Z \max(e^{-c\sqrt{\log Z}}, e^{-c \log Z/\log q})).$$

Since

$$\sum_{Z \leqslant m \leqslant 2Z} \Lambda(m) = Z(1 + o(1))$$

by the prime number theorem, it follows that if $q \leqslant Y^\varrho$ with ϱ small enough then at most 95% of the primes in $[Z, 2Z]$ are such that $(q|p) = \chi(p) = 1$. □

Proof of Theorem 1.4. A key role will be played by primes p that are close to $X^{1/4}$. It is convenient to introduce some terminology concerning them. Let C be a large absolute constant to be specified later. We say that $p \sim X^{1/4}$ if $X^{1/4-C\varepsilon} < p < X^{1/4}$. Furthermore, we will say that a certain property holds "for at least 1% of primes $p \sim X^{1/4}$" if the set $\mathscr{P}$ of primes for which this property holds satisfies

$$\sum_{p \sim X^{1/4} : p \in \mathscr{P}} \frac{\log p}{p} \geqslant 0.01 \sum_{p \sim X^{1/4}} \frac{\log p}{p}.$$

The weighting of $\log p/p$ is included with some later applications of the larger sieve in mind. We begin with some preliminary analysis using the larger sieve, strongly based on the work of Elsholtz [3].

Lemma 6.3. *Let A, B be sets such that $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p+1$ for all primes $p \in [X_0, X^{1/4}]$. Let $\varepsilon > 0$, and suppose that X is sufficiently large in terms of X_0 and ε. Then either $A \cap [X^{1/2}]$ or $B \cap [X^{1/2}]$ has size at most $X^{1/4-c\varepsilon^3}$, or else for at least 99% of primes $p \sim X^{1/4}$ we have both $|A(\mathrm{mod}\ p)| \leqslant (\frac{1}{2} + \varepsilon)p$ and $|B(\mathrm{mod}\ p)| \leqslant (\frac{1}{2} + \varepsilon)p$. We may take $c = 2^{-10}$.*

Proof. Write $\alpha_p := |A(\mathrm{mod}\ p)|/p$, $\beta_p := |B(\mathrm{mod}\ p)|/p$. Thus we are assuming that $\alpha_p + \beta_p \leqslant 1 + \frac{1}{p}$. Suppose the final statement of the lemma is false. Then

$$\sum_{p \sim X^{1/4}} \frac{\log p}{p}\big((1-2\alpha_p)^2 + (1-2\beta_p)^2\big) \geqslant 2^{-5}\varepsilon^2 \sum_{p \sim X^{1/4}} \frac{\log p}{p} \geqslant 2^{-5}\varepsilon^3 \log X.$$

If $c < 2^{-8}$, we can remove the contribution from $X^{1/4-2c\varepsilon^3} \leqslant p \leqslant X^{1/4}$ trivially to get

$$\sum_{p < X^{1/4-2c\varepsilon^3}} \frac{\log p}{p}\big((1-2\alpha_p)^2 + (1-2\beta_p)^2\big) \geqslant 2^{-6}\varepsilon^3 \log X.$$

We claim that if a, b are positive real numbers with $a + b \leqslant 1$ then

$$\frac{1}{a} + \frac{1}{b} \geqslant 4 + 2(1-2a)^2 + 2(1-2b)^2.$$

To see this, apply the inequality

$$\frac{1}{x} + \frac{1}{1-x} = 4 + \frac{(1-2x)^2}{x(1-x)} \geqslant 4 + 4(1-2x)^2$$

with $x = a$ and $x = b$ in turn, and add the results.

Applying this together with the above, we obtain

$$\sum_{X_0 \leqslant p \leqslant X^{1/4-2c\varepsilon^3}} \frac{\log p}{p}(\frac{1}{\alpha_p} + \frac{1}{\beta_p}) \geqslant (1 - 8c\varepsilon^3)\log X + 2^{-5}\varepsilon^3 \log X - O(1) \geqslant (1 - 8c\varepsilon^3 + 2^{-6}\varepsilon^3)\log X.$$

Here we used the fact (an estimate of Mertens) that $\sum_{X_0 \leqslant p \leqslant Z} \frac{\log p}{p} = \log Z + O_{X_0}(1)$. Note also that we only have $\alpha_p + \beta_p \leqslant 1 + \frac{1}{p}$, and not $\alpha_p + \beta_p \leqslant 1$; the introduction of the $O(1)$ term takes care of this as well, the full justification of which we leave to the reader[4]. Without loss of generality the contribution from the α_p is at least that from the β_p, so

$$\sum_{X_0 \leqslant p \leqslant X^{1/4-2c\varepsilon^3}} \frac{\log p}{p}\frac{1}{\alpha_p} \geqslant (\frac{1}{2} - 4c\varepsilon^3 + 2^{-7}\varepsilon^3)\log X > (\frac{1}{2} + 2^{-8}\varepsilon^3)\log X$$

if $c \leqslant 2^{-10}$. Then, however, the larger sieve implies that

$$|A \cap [X^{1/2}]| \ll \frac{X^{1/4-2c\varepsilon^3}}{\varepsilon^3 \log X} < X^{1/4-c\varepsilon^3},$$

and the result follows. □

Suppose now that the hypotheses are as in Theorem 1.4. Replace ε by $\varepsilon/2C$ (the statement of the theorem does not change). Let ψ_A, ψ_B be rational quadratics of height at most H. If option (ii) of the theorem does not hold then at least $HX^{1/4+2C\varepsilon}$ elements of A lie in $\psi_A(\mathbb{Q})$ and at least $HX^{1/4+2C\varepsilon}$ elements of B lie in $\psi_B(\mathbb{Q})$. Suppose also that option (i) of Theorem 1.4 does not hold. Then we may apply Lemma 6.3 to conclude that both $|A(\mathrm{mod}\ p)|$ and $|B(\mathrm{mod}\ p)|$ are $\leqslant (\frac{1}{2} + \varepsilon)p$ for at least 99% of all primes $p \sim X^{1/4}$. (We urge the reader to recall the special meaning of this notation.) Using this information together with the fact that $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p + 1$ for all $p \sim X^{1/4}$, we will deduce that in fact

$$|A \setminus \psi_A(\mathbb{Q})|, |B \setminus \psi_B(\mathbb{Q})| \ll HX^{1/2-c}, \tag{6.2}$$

this being the other conclusion of Theorem 1.4 (ii). It suffices to prove this for A, the proof for B being identical. Write $\psi = \psi_A$, and suppose that $p \sim X^{1/4}$. Set

$$T_p := A(\mathrm{mod}\ p) \cap (\psi(\mathbb{Q}) \cap \mathbb{Z})(\mathrm{mod}\ p),$$

$$U_p := A(\mathrm{mod}\ p) \setminus (\psi(\mathbb{Q}) \cap \mathbb{Z})(\mathrm{mod}\ p).$$

[4]We are working with the condition $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p + 1$, rather than the cleaner condition $|A(\mathrm{mod}\ p)| + |B(\mathrm{mod}\ p)| \leqslant p$, so that we can formulate Theorem 1.6 to include the case in which $\mathscr{A}$ is the set of values of a quadratic. Note, however, that in this case Lemma 6.3 is vacuous anyway. Therefore this small point really can be ignored.

We know that $|A(\mathrm{mod}\ p)| \leqslant (\frac{1}{2}+\varepsilon)p$ for at least 99% of all primes $p \sim X^{1/4}$. For these primes, then,

$$|T_p| + |U_p| \leqslant (\frac{1}{2}+\varepsilon)p.$$

We claim that $|U_p| \leqslant 2\varepsilon p$ for at least 98% of all primes $p \sim X^{1/4}$. If this failed, we would have

$$|T_p| \leqslant (\frac{1}{2}-\varepsilon)p \tag{6.3}$$

for at least 1% of the primes $p \sim X^{1/4}$. Write $\psi(x) = \frac{1}{d}(ax^2+bx+c)$ with a,b,c,d having no common factor and $|a|,|b|,|c|,|d| \leqslant H$, and note that $\psi^{-1}(\mathbb{Z}) \subset \frac{1}{a}\mathbb{Z}$. Note furthermore that

$$\{x \in \mathbb{Z} : \psi(\frac{x}{a}) \in A \cap \psi(\mathbb{Q})\} \subset \bigcap_{p \sim X^{1/4}} \{x \in [-C_1HX^{1/2}, C_1HX^{1/2}] : \psi(\frac{x}{a})(\mathrm{mod}\ p) \in T_p\}, \tag{6.4}$$

where C_1 is some absolute constant. If $p \sim X^{1/4}$, the condition that $\psi(\frac{x}{a})(\mathrm{mod}\ p) \in T_p$ forces $x(\mathrm{mod}\ p)$ to lie in some set $S_p \subset \mathbb{Z}/p\mathbb{Z}$ of residue classes with $|S_p| \leqslant 2|T_p|$. We now have a large sieve problem to which Proposition 2.1 may be applied. If p is such that (6.3) holds then we have $|S_p^c|/|S_p| \geqslant \varepsilon$, and so

$$\sum_{q \leqslant X^{1/4}} \mu^2(q) \prod_{p|q} \frac{|S_p^c|}{|S_p|} \geqslant \sum_{p \sim X^{1/4}} \frac{|S_p^c|}{|S_p|} \gg \varepsilon \frac{X^{1/4-C\varepsilon}}{\log X}.$$

Here, $X^{1/4-C\varepsilon}/\log X$ is a crude lower bound for the size of a set of primes constituting 1% of all $p \sim X^{1/4}$, this lower bound being attained when all the primes congregate at the bottom of the interval $X^{1/4-C\varepsilon} \leqslant p \leqslant X^{1/4}$. It follows from (6.4) and Proposition 2.1 that $|A \cap \psi(\mathbb{Q})| < HX^{1/4+2C\varepsilon}$ if X is large enough, contrary to assumption.

Now let us write $A = A_\psi \cup E$, where A_ψ consists of those $x \in A$ for which $x(\mathrm{mod}\ p) \in T_p$ for at least 97% of $p \sim X^{1/4}$, and E consists of those $x \in A$ such that $x(\mathrm{mod}\ p) \in U_p$ for at least 3% of $p \sim X^{1/4}$. The idea here is that A_ψ satisfies a large number of local conditions suggesting that its elements lie in $\psi(\mathbb{Q})$. We would like to relate A_ψ to $A \cap \psi(\mathbb{Q})$, and show that E is small. With this idea in hand, we can divide the task of proving (6.2) into two subclaims, namely

$$|A_\psi \setminus \psi(\mathbb{Q})| \ll HX^{1/2-c} \qquad \text{and} \qquad |E| \ll X^{1/4}. \tag{6.5}$$

Of course, we could tolerate a weaker bound for $|E|$, but as it turns out we need not settle for one.

We start with the first claim, which is quite straightforward given results we established earlier. Let Δ be the discriminant of ψ. If x is an integer then so is $4adx + \Delta d^2$, and furthermore if $x = \psi(n)$ then $4adx + \Delta d^2 = (2an+b)^2$. Therefore if $x \in A_\psi$ then $4adx + \Delta d^2$ is an integer which is a square modulo p for at least 97% of all $p \sim X^{1/4}$. By a simple averaging argument, $4adx + \Delta d^2$ is a square modulo at least 95% of all $p \in [Z, 2Z]$ for some $Z \sim X^{1/4}$. It follows from Lemma 6.2 that

$$4adx + \Delta d^2 = n_i s^2, \tag{6.6}$$

where $s \in \mathbb{Z}$ and n_i is one of at most $O(\log^{10} X)$ squarefree integers, with $n_1 = 1$ and $n_i \geqslant X^\varrho$ if $i \geqslant 2$, where $\varrho > 0$ is an absolute constant. The number of $x \in [X]$ for which (6.6) holds for a given i is $\ll H\sqrt{X/n_i}$, and so the number of x for which this holds for *some* $i \geqslant 2$ is $\ll H \log^{10} X \cdot X^{(1-\varrho)/2} \ll HX^{1/2-\varrho/4}$. If $i = 1$, so that $n_1 = 1$, then $4adx + \Delta d^2$ is a square and so $x \in \psi(\mathbb{Q})$. This concludes the proof of the first bound in (6.5).

We turn now to the proof of the second bound in (6.5). Recall first of all that $|U_p| \leqslant 2\varepsilon p$ for at least 98% of all $p \sim X^{1/4}$, and also that for every $x \in E$ we have $x \pmod p \in U_p$ for at least 3% of all $p \sim X^{1/4}$. For every $x \in E$, both of these events occur for at least 1% of all $p \sim X^{1/4}$.

Write E_p for the subset of E whose elements belong to U_p modulo p. By the preceding facts we have

$$\sum_{\substack{p\sim X^{1/4}\\ |U_p|\leqslant 2\varepsilon p}} \frac{\log p}{p}|E_p| = \sum_{x\in E}\sum_{\substack{p\sim X^{1/4}\\ |U_p|\leqslant 2\varepsilon p}} \frac{\log p}{p} 1_{x \pmod p \in U_p} \geqslant \frac{1}{100}|E| \sum_{p\sim X^{1/4}} \frac{\log p}{p}.$$

Writing $\mathscr{P} := \{p \sim X^{1/4} : |U_p| \leqslant 2\varepsilon p \text{ and } |E_p| \geqslant \frac{1}{200}|E|\}$, it follows that

$$\sum_{p\in\mathscr{P}} \frac{\log p}{p} \geqslant \frac{1}{200} \sum_{p\sim X^{1/4}} \frac{\log p}{p}.$$

Applying the larger sieve (that is Theorem 2.3) with the choices $\delta = \frac{1}{200}$, $Q = X^{1/4}$ and $\sigma_p = 2\varepsilon$, we obtain $|E| \ll X^{1/4}(\frac{1}{2^{20}\varepsilon}\sum_{p\sim X^{1/4}} \frac{\log p}{p} - \log X)^{-1}$, provided that the term in parentheses is positive. That term is $> (2^{-21}C - 1)\log X$, which is positive if $C > 2^{22}$ (say). Thus we get the bound $|E| \ll X^{1/4}$. This completes the proof of (6.5), and hence (6.2) and Theorem 1.4. □

We turn now to the proof of Theorem 1.6, the stability theorem for a single infinite set $\mathscr{A}$. Again, we begin by recalling the statement.

Theorem 1.6. *Suppose that $\mathscr{A}$ is a set of positive integers and that $|\mathscr{A} \pmod p| \leqslant \frac{1}{2}(p+1)$ for all sufficiently large primes p. Then one of the following options holds:*

(i) (Quadratic structure) *There is a rational quadratic ψ such that all except finitely many elements of $\mathscr{A}$ are contained in $\psi(\mathbb{Q})$;*
(ii) (Better than large sieve) *For each integer k there are arbitrarily large values of X such that $|\mathscr{A}[X]| < \frac{X^{1/2}}{\log^k X}$;*
(iii) (Far from quadratic structure) *Given any rational quadratic ψ, for all X we have $|\mathscr{A}[X] \cap \psi(\mathbb{Q})| \leqslant X^{1/4+o_\psi(1)}$.*

Proof. Suppose that neither item (ii) nor item (iii) holds. Then there is an $\varepsilon > 0$ and a rational quadratic ψ such that $|\mathscr{A}[X] \cap \psi(\mathbb{Q})| > X^{1/4+\varepsilon}$ for arbitrarily large values of X. For any such X we may apply Theorem 1.4 with $A = \mathscr{A}[X]$ to conclude that either option (ii) of our present theorem holds, or else for infinitely many X we have $|\mathscr{A}[X] \setminus \psi(\mathbb{Q})| \ll X^{1/2-c}$. (We note in passing that Lemma 6.3 is redundant inside the proof of Theorem 1.4 in this setting, being trivially true.) We will deduce from this and further applications of the fact that $|\mathscr{A} \pmod p| \leqslant \frac{1}{2}(p+1)$ for p sufficiently large that either (i) or (ii) of Theorem 1.6 holds. Let $\tilde{\psi}$ be a rational quadratic satisfying the conclusions of Lemma A.1, thus $\psi(\mathbb{Q}) \cap \mathbb{Z} \subset \tilde{\psi}(\mathbb{Z})$ and for all sufficiently large primes p the reductions (mod p) of $\psi(\mathbb{Q}) \cap \mathbb{Z}$ and of $\tilde{\psi}(\mathbb{Z})$ are the same. Suppose that option (ii) of Theorem 1.6 does not hold for the infinitely many values of X for which we have $|\mathscr{A}[X] \setminus \psi(\mathbb{Q})| \ll X^{1/2-c}$. Then there is some integer k such that (letting X range through these values)

$$\limsup_{X\to\infty} X^{-1/2}\log^k X |\mathscr{A}[X] \cap \psi(\mathbb{Q})| = \infty. \tag{6.7}$$

We claim that this implies statement (i) of Theorem 1.4, and in fact the stronger conclusion $|\mathscr{A}\setminus\psi(\mathbb{Q})| \leqslant k+1$. Suppose this statement is false. Then there are elements $x_1,\dots,x_{k+1}$ in $\mathscr{A}$ but not in $\psi(\mathbb{Q})$. Since x lies in $\psi(\mathbb{Q})$ if and only if $4adx+\Delta d^2$ is the square of a rational number, it follows that none of $4adx_i+\Delta d^2$ is a square. Set $m_i := 4adx_i+\Delta d^2$, and suppose that p is a prime such that $(m_i|p)=-1$. If p is sufficiently large then $x_i \notin (\psi(\mathbb{Q})\cap\mathbb{Z})(\mathrm{mod}\ p)$ and hence $x_i \notin \tilde{\psi}(\mathbb{Z})(\mathrm{mod}\ p)$.

For each prime p, let $k(p)$ be the number of indices $i\in\{1,\dots,k+1\}$ such that $(m_i|p)=-1$. From the above reasoning and the assumption that $x_i\in\mathscr{A}$ it follows that $\mathscr{A}\cap\tilde{\psi}(\mathbb{Z})(\mathrm{mod}\ p)$ must occupy a set of size at most $\frac{1}{2}(p+1)-k(p)$ for all sufficiently large primes p. Define a set $\mathscr{B}\subset\mathbb{Z}$ by $\mathscr{A}\cap\tilde{\psi}(\mathbb{Z})=\tilde{\psi}(\mathscr{B})$. Thus $|\tilde{\psi}(\mathscr{B})(\mathrm{mod}\ p)| \leqslant \frac{1}{2}(p+1)-k(p)$ for all sufficiently large primes p, which implies that $|\mathscr{B}(\mathrm{mod}\ p)| \leqslant p-2k(p)+1$. Note also that $\mathscr{A}[X]\cap\psi(\mathbb{Q}) \subset \tilde{\psi}(\mathscr{B}\cap[c_1\sqrt{X},c_2\sqrt{X}])$ for some constants c_1,c_2 depending only on $\tilde{\psi}$. We may now apply Lemma 2.2 to the set $\mathscr{B}$. In that lemma we may take $w(p)=2k(p)-1$, where $k(p)$ is the number of i for which $(m_i|p)=-1$, or equivalently (if $p>2$) for which $\chi_i(p)=-1$ where $\chi_i(n)=(4m_i|n)$ is a real Dirichlet character and $(\,|\,)$ denotes the Kronecker symbol. Thus $k(p)=\frac{1}{2}(k+1)-\frac{1}{2}\sum_{i=1}^{k+1}\chi_i(p)$, and so $w(p)=k-\sum_{i=1}^{k+1}\chi_i(p)$. The conditions of Lemma 2.2 are easily satisfied by the prime number theorem for characters with a fairly crude error term. It follows from Lemma 2.2 and the above discussion that

$$|\mathscr{A}[X]\cap\psi(\mathbb{Q})| \leqslant |\mathscr{B}\cap[c_1\sqrt{X},c_2\sqrt{X}]| \ll X^{1/2}(\log X)^{-k},$$

contrary to (6.7). □

7. Composite numbers in $\mathscr{A}+\mathscr{B}$

Recall from the introduction the following conjecture of "inverse large sieve" type.

Conjecture 1.5. *Let $X_0\in\mathbb{N}$, and let $\rho>0$. Let $X\in\mathbb{N}$ be sufficiently large in terms of X_0 and ρ. Suppose that $A,B\subset[X]$ and that $|A(\mathrm{mod}\ p)|+|B(\mathrm{mod}\ p)|\leqslant p+1$ for all $p\in[X_0,X^{1/4}]$. Then there exists a constant $c=c(\rho)>0$ such that one of the following holds:*

(i) (Better than large sieve) *Either $|A\cap[X^{1/2}]|$ or $|B\cap[X^{1/2}]|$ is $\leqslant X^{1/4-c}$;*
(ii) (Quadratic structure) *There are two rational quadratics ψ_A,ψ_B of height at most X^ρ such that $|A\setminus\psi_A(\mathbb{Q})|$ and $|B\setminus\psi_B(\mathbb{Q})|\leqslant X^{1/2-c}$.*

Our aim in this section is to prove Theorem 1.8, which is the following statement.

Theorem 1.8. *Assume Conjecture 1.5. Let $\mathscr{A},\mathscr{B}$ be two sets of positive integers, with $|\mathscr{A}|,|\mathscr{B}|\geqslant 2$, such that $\mathscr{A}+\mathscr{B}$ contains all sufficiently large primes. Then $\mathscr{A}+\mathscr{B}$ also contains infinitely many composite numbers.*

This follows quite straightforwardly from the following fact.

Lemma 7.1. *There is $\rho>0$ with the following property. Suppose that ψ_A,ψ_B are two rational quadratics of height at most X^ρ and that $A\subset\psi_A(\mathbb{Q})\cap[X]$ and $B\subset\psi_B(\mathbb{Q})\cap[X]$ are sets of positive integers. Suppose that $A+B$ contains no composite number. Then at least one of A and B has cardinality $\ll X^{1/3}$.*

Proof of Theorem 1.8 given Lemma 7.1. Let $\mathscr{A}, \mathscr{B}$ be two sets of positive integers with $|\mathscr{A}|, |\mathscr{B}| \geqslant 2$ such that $\mathscr{A} + \mathscr{B}$ coincides with the set of primes on $[X_0, \infty)$.

We claim that there are infinitely many X such that either $|\mathscr{A}[X]|$ or $|\mathscr{B}[X]|$ has cardinality at most $X^{1/2-c}$. This, however, is contrary to a theorem[5] of Elsholtz [3], which implies that $|\mathscr{A}[X]|, |\mathscr{B}[X]| \gg X^{1/2} \log^{-5} X$ for all sufficiently large X.

It remains to prove the claim. Let ρ be as in Lemma 7.1. For each X, write $A := \mathscr{A} \cap (X^{1/4}, X]$ and $B := \mathscr{B} \cap (X^{1/4}, X]$. If $p \in [X_0, X^{1/4}]$ then $A+B$ contains no multiple of p, since any such number would be a nontrivial multiple of p (and hence composite) and lies in $\mathscr{A} + \mathscr{B}$. For these primes p, then, we have $|A (\mathrm{mod}\ p)| + |B (\mathrm{mod}\ p)| \leqslant p$, since $B (\mathrm{mod}\ p)$ cannot intersect $(-A) (\mathrm{mod}\ p)$. Assuming Conjecture 1.5, for each X one of the two options (i) or (ii) of that conjecture holds.

If (i) holds for infinitely many X then without loss of generality we have $|A \cap [X^{1/2}]| \ll X^{1/4-c}$ for infinitely many X. By Elsholtz [3] we have $|\mathscr{A}[X^{1/4}]| \ll X^{1/8+o(1)}$, and therefore $|\mathscr{A}[X^{1/2}]| \ll X^{1/4-c}$ for infinitely many X, thereby establishing the claim.

Suppose, then, that (ii) holds for all sufficiently large X. That is, there are rational quadratics ψ_A, ψ_B of height X^ρ such that $|A \setminus \psi_A(\mathbb{Q})|, |B \setminus \psi_B(\mathbb{Q})| \leqslant X^{1/2-c}$. Write $A' := A \cap \psi_A(\mathbb{Q})$ and $B' := B \cap \psi_B(\mathbb{Q})$. Certainly $A' + B'$ contains no composite numbers. By Lemma 7.1, for all X at least one of A', B' has cardinality $\ll X^{1/3}$, and this means that indeed either $\mathscr{A}[X]$ or $\mathscr{B}[X]$ has size at most $X^{1/2-c}$ for infinitely many X. □

Proof of Lemma 7.1. Write $\psi_A(x) = \frac{1}{d_A}(a_A x^2 + b_A x + c_A)$, $\psi_B(x) = \frac{1}{d_B}(a_B x^2 + b_B x + c_B)$. Here $a_A, a_B, b_A, b_B, c_A, c_B, d_A, d_B$ are integers, all of magnitude at most $H = X^\rho$. Set $Y := (H^2 X)^{1/4}$. If $A + B$ contains no composite number then the set $(A \cap (Y, X]) + (B \cap (Y, X])$ contains no multiple of any prime $p \leqslant Y$. Note also that $\psi_A^{-1}(\mathbb{Z}) \subset \frac{1}{a_A}\mathbb{Z}$ and $\psi_B^{-1}(\mathbb{Z}) \subset \frac{1}{a_B}\mathbb{Z}$. Set

$$S_A := \{x \in \mathbb{Z} : \psi_A(\frac{x}{a_A}) \in A \cap (Y, X]\}, \quad S_B := \{x \in \mathbb{Z} : \psi_B(\frac{x}{a_B}) \in B \cap (Y, X]\},$$

and note that $\psi_A(\frac{x_A}{a_A}) + \psi_B(\frac{x_B}{a_B}) \neq 0 (\mathrm{mod}\ p)$ whenever $x_A \in S_A$, $x_B \in S_B$, and $p \leqslant Y$ is a prime. To prove Lemma 7.1, it suffices to show that either $|S_A|$ or $|S_B|$ has size $\ll X^{1/3}$. Note furthermore (by completing the square) that $S_A, S_B \subset [-4(H^2X)^{1/2}, 4(H^2X)^{1/2}]$.

We will focus attention only on those primes $p \leqslant Y$ for which $(-a_A a_B d_A d_B | p) = 1$, that is to say for which $-a_A a_B d_A d_B$ is a square modulo p. We look for such primes amongst the $p \leqslant Y$ with $p \equiv 1 (\mathrm{mod}\ 8)$. Since both -1 and 2 are squares modulo such a prime, it certainly suffices to additionally ensure that $(q_1|p) = 1, \dots, (q_k|p) = 1$, where $q_1, \dots, q_k$ are the distinct odd primes appearing in $a_A a_B d_A d_B$. This is equivalent to the union of $(q_1 - 1) \dots (q_k - 1)/2^k$ congruence conditions modulo $q_1 \dots q_k$. Together with the condition $p \equiv 1 (\mathrm{mod}\ 8)$, we get the union of at least $2^{-k-2}\phi(q)$ congruence conditions modulo $q := 8q_1 \dots q_k$. Since $|a_A|, |a_B|, |d_A|, |d_B| \leqslant H$ we have $q \leqslant 8H^4$. Now we invoke [13, Corollary 18.8], a quantitative version of Linnik's theorem on the least prime in an arithmetic progression, which implies that there are at least $\frac{Y}{\phi(q)\sqrt{q}\log Y}$ primes $p \leqslant Y$ satisfying each of these congruence conditions, and hence $\gg \frac{2^{-k}Y}{\sqrt{q}\log Y}$ such primes in total. (Here we used the fact that $H = X^\rho$ with ρ sufficiently small.) Write $\mathscr{P}$ for the set of such primes, thus $|\mathscr{P}| \gg X^{1/4-o(1)}H^{-2}$.

[5]The state-of-the-art here is $|\mathscr{A}[X]| \gg X^{1/2}/\log X \log\log X$: see [4].

Now suppose that p is such a prime and that $-a_Ad_A/a_Bd_B \equiv m^2 (\mathrm{mod}\ p)$. Then we have

$$\psi_A(\frac{x_A}{a_A}) + \psi_B(\frac{x_B}{a_B}) = \frac{1}{a_Ad_A}(x_A^2 + b_Ax_A + a_Ac_A) + \frac{1}{a_Bd_B}(x_B^2 + b_Bx_B + a_Bc_B)$$
$$\equiv \frac{1}{a_Ad_A}\big((x_A + c_1)^2 - (mx_B + c_2)^2 + c_3\big),$$

where c_1, c_2, c_3 do not depend on x_A, x_B. Therefore for each prime $p \in \mathscr{P}$ we have one of the following alternatives.

(i) $c_3 \equiv 0$ modulo p. Then whenever $x_A \in S_A (\mathrm{mod}\ p)$ we must have $\frac{1}{m}(x_A + c_1 - c_2) \notin S_B (\mathrm{mod}\ p)$, whence $|S_A (\mathrm{mod}\ p)| + |S_B (\mathrm{mod}\ p)| \leqslant p$.

(ii) $c_3 \not\equiv 0$ modulo p. Then we have $\psi_A(\frac{x_A}{a_A}) + \psi_B(\frac{x_B}{a_B}) \equiv 0 (\mathrm{mod}\ p)$ whenever

$$(x_A + mx_B + c_1 + c_2)(x_A - mx_B + c_1 - c_2) \equiv -c_3,$$

an equation which has $p-1$ solutions (x_A, x_B). This solution set must be disjoint from $S_A \times S_B$. Since to each x_A there are at most two x_B, and to each x_B there are at most two x_A, this forces at least one of $S_A (\mathrm{mod}\ p), S_B (\mathrm{mod}\ p)$ to have size $\leqslant 7p/8$, say.

In both cases at least one of $S_A (\mathrm{mod}\ p), S_B (\mathrm{mod}\ p)$ has size $\leqslant 7p/8$. Without loss of generality the first holds for at least half the elements of $\mathscr{P}$. Finally the large sieve, as in Proposition 2.1, tells us that $|S_A| \ll \frac{(H^2X)^{1/2}}{|\mathscr{P}|} \ll H^3X^{1/4+o(1)}$. If ρ is small enough then this is $\ll X^{1/3}$, as required. □

Appendix A. Basic facts about rational quadratics

Lemma A.1. *Suppose that ψ is a rational quadratic such that $\psi(\mathbb{Q}) \cap \mathbb{Z}$ is nonempty. Then there is another rational quadratic $\tilde{\psi}$ with $\psi(\mathbb{Q}) = \tilde{\psi}(\mathbb{Q})$ such that $\psi(\mathbb{Q}) \cap \mathbb{Z} \subset \tilde{\psi}(\mathbb{Z})$. Furthermore, for all sufficiently large primes p the reductions $(\mathrm{mod}\, p)$ of $\psi(\mathbb{Q}) \cap \mathbb{Z}$ and of $\tilde{\psi}(\mathbb{Z})$ are the same.*

Proof. Write $\psi(x) = \frac{1}{d}(ax^2 + bx + c)$ with $a, b, c, d \in \mathbb{Z}$, and simply define $\tilde{\psi}(x) := \psi(\frac{1}{a}x)$. The first property, that $\psi(\mathbb{Q}) = \tilde{\psi}(\mathbb{Q})$, is immediate. Moreover if $\psi(u/v)$ is an integer, with u/v a rational in lowest terms, then $v|a$. It follows that $\psi(\mathbb{Q}) \cap \mathbb{Z} \subset \psi(\frac{1}{a}\mathbb{Z}) = \tilde{\psi}(\mathbb{Z})$, the second required property.

To get the last statement (about reductions mod p), let x_0 be a rational such that $\psi(x_0) \in \mathbb{Z}$ and write $x_0 := r/s$ in lowest terms. Then $\psi(x_0 + d\mathbb{Z}) \subset \mathbb{Z}$ (since $s|a$, as noted above), and so $\tilde{\psi}(ax_0 + ad\mathbb{Z}) \subset \mathbb{Z}$. Thus, writing $P \subset \mathbb{Z}$ for the infinite arithmetic progression $ax_0 + ad\mathbb{Z}$, we see that $\tilde{\psi}(P) \subset \psi(\mathbb{Q}) \cap \mathbb{Z}$. However for p a sufficiently large prime, $P (\mathrm{mod}\ p)$ is all of $\mathbb{Z}/p\mathbb{Z}$, thereby concluding the proof. □

References

[1] E. Croot and V. Lev, *Open problems in additive combinatorics,* in Additive Combinatorics, CRM Proc. Lecture Notes **43**, 207–233, Amer. Math. Soc., Providence, RI, 2007.

[2] J.-M. De Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers,* Graduate Studies in Math. **134**, AMS 2012.

[3] C. Elsholtz, *The inverse Goldbach problem,* Mathematika **48** (2001), 151–158.

[4] C. Elsholtz and A. J. Harper, *Additive decompositions of sets with restricted primes factors,* to appear in Trans. Amer. Math. Soc., preprint available at arXiv:1309.0593.

[5] P. Erdős, *Problems and Results in Combinatorial Number Theoery, III,* in Number theory day (Proc. Conf., Rockefeller Univ., New York, 1976), pp. 4372. Lecture Notes in Math., Vol. **626**, Springer, Berlin, 1977.

[6] J. Friedlander and H. Iwaniec. *Opera de Cribro*. American Mathematical Society Colloquium Publications **57**, 2010.
[7] P. X. Gallagher, *A Large Sieve Density Estimate near $\sigma = 1$,* Inventiones math. **11** (1970), 329–339.
[8] P. X. Gallagher, *A larger sieve,* Acta Arith. **18** (1971), 77–81.
[9] A. Granville and O. Ramaré, *Explicit bounds on exponential sums and the scarcity of squarefree binomial coefficients,* Mathematika **43** (1996), 73–107.
[10] B. Green, *On a variant of the large sieve,* preprint available at arXiv:0807.5037.
[11] D. R. Heath-Brown, *A mean value estimate for real character sums,* Acta Arith. **72** (1995), 235–275.
[12] H. A. Helfgott and A. Venkatesh, *How small must ill-distributed sets be?,* in Analytic Number Theory: Essays in honour of Klaus Roth, 224–234, Cambridge Univ. Press, Cambridge, 2009.
[13] H. Iwaniec and E. Kowalski. *Analytic Number Theory.* American Mathematical Society Colloquium Publications **53**, 2004.
[14] M. Jutila, *On numbers with a large prime factor. II.,* J. Indian Math. Soc. (N.S.) **38** (1974), 125–130.
[15] H. L. Montgomery, *A note on the large sieve,* J. London Math. Soc. **43** (1968) 93–98.
[16] H.-H. Ostmann, *Additive Zahlentheorie. Teil I: Allgemeine Untersuchungen,* Springer- Verlag, Berlin-Heidelberg-New York, 1968.
[17] J. M. Pollard, *A generalisation of the theorem of Cauchy and Davenport*, J. London Math. Soc. **8** (1974), 460–462.
[18] T. Tao and V. H. Vu. *Additive Combinatorics*. Reprint with corrections. Cambridge Studies in Advanced Mathematics **105**, 2007.
[19] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, Comptes Rendus Acad. Sci. Paris, Serie A **285** (1977), 981–983.
[20] M. N. Walsh, *The inverse sieve problem in high dimensions*, Duke Math. J. **161** (2012), no. 10, 2001–2022.

Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, England
E-mail address: ben.green@maths.ox.ac.uk

Jesus College, Cambridge CB5 8BL, England
E-mail address: A.J.Harper@dpmms.cam.ac.uk