ACTA ARITHMETICA
159.3 (2013)

Sums of positive density subsets of the primes

by

Kaisa Matomäki (Turku)

1. Introduction. Let us first quickly state our main result.

Theorem 1.1. *Let A and B be subsets of the primes with positive relative lower densities α and β. Then the lower density of $A+B$ in the natural numbers is at least*

$$(1 - o_{\alpha+\beta\to 0}(1))\frac{\alpha}{e^{\gamma}\log\log(1/\beta)}, \tag{1.1}$$

where γ is the Euler–Mascheroni constant.

It might look surprising that the lower bound (1.1) is not symmetric in α and β (though of course one gets the same bound with α and β interchanged, and could replace (1.1) by the maximum of these two bounds, which is symmetric). However the two sets indeed have non-symmetric roles in the situation and the lower bound is asymptotically sharp as Examples 1.2 (case $\alpha = \beta$) and 3.4 (general case) below demonstrate.

Studying additive properties of positive density subsets of the primes has become much more accessible during the last decade thanks to works of Green and Tao; first Green [4] showed that any subset of the primes with positive relative upper density contains 3-term arithmetic progressions, and later Green and Tao [7] generalised this to arbitrarily long arithmetic progressions—before their celebrated theorem it was not even known that the primes themselves contain infinitely many 4-term arithmetic progressions!

Methods used for studying primes in arithmetic progressions are typically applicable also to studying Goldbach type problems and so is the case with Green–Tao methods (see [9] for very general quantitative results). However, for Goldbach type problems, that is, for sums of primes, introducing positive density subsets cannot be straightforward as the following example shows.

2010 *Mathematics Subject Classification*: Primary 11B30, 11P32.
Key words and phrases: sum sets, primes, positive density subsets, Goldbach type problems.

DOI: 10.4064/aa159-3-1

Example 1.2. Let $m \in \mathbb{N}$ and let $\mathcal{P}$ be the set of primes that are 1 (mod m). Then no integer $\not\equiv l$ (mod m) is the sum of l primes from $\mathcal{P}$. In this example $\mathcal{P}$ has relative density $1/\varphi(m)$, while the sum set $l\mathcal{P}$ has density at most $1/m$ in the natural numbers.

Consider the case where m is the product of the first k primes. Then $m/\varphi(m) = (1 + o_{m\to\infty}(1))e^{\gamma} \log\log \varphi(m)$ by Lemma 4.1 below, and hence, writing $\delta = 1/\varphi(m)$ for the relative density of $\mathcal{P}$, the density of $l\mathcal{P}$ in the natural numbers is only

$$(1 + o_{\delta\to 0}(1))\frac{\delta}{e^{\gamma} \log\log(1/\delta)}.$$

However, Li and Pan [12] have managed to show that when the subsets of the primes are dense enough, this kind of phenomenon cannot occur. Plugging this information into an adaption of Green's approach, they proved the following variant of the ternary Goldbach conjecture.

Theorem. *Let $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ be subsets of the primes with positive relative lower densities α_1, α_2 and α_3. If $\alpha_1 + \alpha_2 + \alpha_3 > 2$, then for every sufficiently large odd integer n there exist primes $p_i \in \mathcal{P}_i$, $i = 1, 2, 3$, such that $n = p_1 + p_2 + p_3$.*

This is sharp in the sense that there are examples in which $\alpha_1 + \alpha_2 + \alpha_3 = 2$ and the conclusion is not true. However X. Shao [15] has very recently shown that in case $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{P}_3$ it is enough that $\alpha_i > 5/8$ (which is again sharp).

A natural question is then: What happens for smaller densities? Chipeniuk and Hamel [3] proved essentially the following theorem, again using an adaption of methods of Green and Green–Tao together with arguments which show that there cannot be too bad obstructions like that in Example 1.2.

Theorem. *There exist absolute positive constants C_1 and C_2 such that if A is a subset of the primes with positive relative lower (or upper) density $\delta < 1/e$, then the lower (or upper) density of $A + A$ in the natural numbers is at least*

$$C_1\delta \exp\bigl(-C_2(\log(1/\delta))^{2/3}(\log\log(1/\delta))^{1/3}\bigr).$$

Notice that the bound here is in particular $\delta^{1+o(1)}$. On the other hand Chipeniuk and Hamel [3] pointed out Example 1.2 and anticipated that the bound there is the right answer in the spirit of Freiman's theorem. Theorem 1.1 shows that this is indeed the case (it is clear from Theorem 2.1 below that in case $A = B$ lower densities can be replaced by upper densities also in Theorem 1.1).

Sum sets of positive density subsets of the primes did actually receive some attention already before the breakthroughs of Green and Green–Tao. Indeed, Ramaré and Ruzsa [14] proved Theorem 1.1 with $1 - o(1)$ replaced by $c - o(1)$ for some constant c. Until recently the authors of [3] as well as the current author have been unaware of that work. In light of [14], the achievement of Theorem 1.1 is getting the right constant.

Ramaré and Ruzsa actually showed more general results for subsets of "sifted sequences", which roughly means sequences to which the Selberg sieve can be applied. The enveloping sieve they developed to achieve this can also be incorporated into the Green–Tao method (see [6]).

To handle congruence problems similar to the one in Example 1.2, one needs to show a corresponding result in the cyclic group $\mathbb{Z}_m$. Our version of this is the following.

THEOREM 1.3. *Let m be a square-free natural number and let A and B be subsets of $\mathbb{Z}_m^*$ with positive relative densities α and β. Then*

$$|A+B| \geq (1 - o_{\beta\to 0}(1)) \frac{\alpha}{e^{\gamma} \log\log(1/\beta)} m.$$

This is asymptotically best possible when m is the product of the first l primes, as Example 3.4 below shows. By best possible we mean that, for any $\varepsilon > 0$, one can find many pairs (α, β) (and thus A, B) such that $1 - o(1)$ cannot be replaced by $1 + \varepsilon$. This of course immediately implies that also Theorem 1.1 is best possible in the same sense.

In [3] and [14] similar results to Theorem 1.3 appear with essentially the same densities as in respective results for the primes. Proofs of those results in [3] and [14] proceed along different lines from ours. On the other hand our deduction of the result for the primes from a result in $\mathbb{Z}_m^*$ follows [3] though we need to work more carefully as we cannot afford to lose constant multiples because we aim for an asymptotically sharp result.

Instead of Theorem 1.3 we will prove a more precise statement in which the dependence on β is explicit rather than asymptotic (see Theorem 3.3 below). Using it we will prove the following explicit version of the main theorem in Section 7.

THEOREM 1.4. *Let A and B be subsets of the primes with positive relative lower densities α and β. Let q_n denote the nth prime and let $l \in \mathbb{N}$ be such that*

$$\beta > \prod_{j=2}^{l} \frac{2}{\varphi(q_j)}.$$

Then the lower density of $A+B$ in the natural numbers is at least

$$\alpha \frac{\varphi(q_1 \cdots q_l)}{q_1 \cdots q_l}. \tag{1.2}$$

One would expect this to hold with 2 replaced by 1 in the lower bound for β (see the discussion after Theorem 3.3). Notice that the lower bound (1.2) depends very mildly on β; for instance taking $l = 54$ shows that if $\beta > 2\cdot 10^{-84}$, then $A+B$ has lower density at least $\alpha/10$.

2. Finite case and inverse questions. Instead of Theorem 1.1, we will show the following finite version which is also best possible as we will show in Section 9.

THEOREM 2.1. *Let $\varepsilon > 0$. There exists $\gamma_1 = \gamma_1(\varepsilon)$ such that if $\gamma_0 \in (0,\gamma_1)$, there exists $n_0 = n_0(\gamma_0,\varepsilon)$ such that the following holds for every $n \ge n_0$ and $\alpha,\beta \in (\gamma_0,\gamma_1)$.*

Let $A, B \subseteq \mathbb{P}\cap[1,n]$ with relative densities α and β. Then

$$|A+B| \ge (1-\varepsilon)\frac{\alpha+\beta}{e^\gamma \log\log(1/(\alpha\beta))}n.$$

Notice that here $A+B$ is a subset of $[1,2n]$, so the factor $(\alpha+\beta)/2$ appears in its density. However a simple trick recovers α for Theorem 1.1:

Deduction of Theorem 1.1 from Theorem 2.1. Let $\varepsilon > 0$ be small and let A and B be as in Theorem 1.1. We can assume that $\beta \le \alpha \le \gamma_1(\varepsilon^2)$. Let n be large and take

$$A' = A\cap[1,(1-\varepsilon^2)n] \quad\text{and}\quad B' = B\cap[1,\varepsilon^2 n],$$

so that A' and B' have relative densities $\alpha' \ge (1-\varepsilon^2)^2\alpha$ and $\beta' \ge \varepsilon^2(1-\varepsilon^2)\beta$ in $\mathbb{P}\cap[1,n]$. Furthermore $A'+B' \subseteq (A+B)\cap[1,n]$, so, applying Theorem 2.1 to A' and B', we get

$$|(A+B)\cap[1,n]| \ge (1-\varepsilon^2)\frac{\alpha'+\beta'}{e^\gamma\log\log(1/(\alpha'\beta'))}n \ge (1-\varepsilon)\frac{\alpha}{e^\gamma\log\log(1/\beta)}n$$

when ε, α and β are small enough, and the claim follows. ■

Before going to the proof of Theorem 2.1 in Sections 3–6, we turn to discussing an inverse question: As shown, Theorem 1.1 is best possible in general and so is Theorem 2.1, but can we classify the worst case examples? To simplify the discussion, let us consider only the finite case.

As already discussed, a major hindrance for us is bad distribution in arithmetic progressions. To quantify this, we introduce a bit of notation which will be used throughout the paper.

Let $W \ll \log\log n$ be a large parameter and set $m = \prod_{p\le W} p$. We split A and B into residue classes modulo m. For any set $C \subseteq \mathbb{Z}$, integer $q \ge 1$ and $r \in \mathbb{Z}_q$, write

$$C_q[r] := \{c \in C\colon c \equiv r \pmod q\} \tag{2.1}$$

and let

$$\begin{aligned}\alpha_m[r] &:= \frac{|A_m[r]|}{|\mathbb{P}_m[r]\cap[1,n]|}, \quad \alpha_m := \min_{r\in\mathbb{Z}_m^*}\alpha_m[r],\\ \beta_m[r] &:= \frac{|B_m[r]|}{|\mathbb{P}_m[r]\cap[1,n]|}, \quad \beta_m := \min_{r\in\mathbb{Z}_m^*}\beta_m[r].\end{aligned}$$

Proposition 3.1 below immediately implies the following.

THEOREM 2.2. *For every $\varepsilon > 0$ there exists $W = W(\varepsilon)$ such that the following holds when*

$$\log\log n \gg W \quad \textit{and} \quad m = \prod_{p\le W} p.$$

Let $A, B \subseteq \mathbb{P}\cap[1,n]$ be such that $\alpha_m, \beta_m \ge \varepsilon$. Then

$$|A+B| \ge (1-\varepsilon)\frac{\alpha_m+\beta_m}{2}n.$$

Hence we see that if A and B are not too badly distributed in residue classes modulo a certain fixed m, we immediately get a greatly improved lower bound for $A+B$. Theorem 2.2 is best possible (except for ε) as the example $A = \mathbb{P}\cap[1,\alpha n]$ and $B = \mathbb{P}\cap[1,\beta n]$ demonstrates. To find more examples, for $N \ge 1$, $d \in \mathbb{Z}_N^*$, $\theta\in\mathbb{R}$ and $\delta > 0$, define

$$U_N(d,\theta,\delta) = \left\{k\in\mathbb{Z}_N\colon \left\|\frac{d}{N}k-\theta\right\| \le \delta\right\}, \tag{2.2}$$

where we write $\|x\|$ for the distance from x to the nearest integer(s). We will abuse notation by writing $U_N(d,\theta,\delta)$ also for the set of integers whose reduction modulo N is in the set.

If now

$$A = \mathbb{P}\cap[3,n]\cap U_N(d,\theta,\alpha/2) \quad\text{and}\quad B = \mathbb{P}\cap[3,n]\cap U_N(d,\theta',\beta/2),$$

then

$$A+B \subseteq \{2k\colon k\le n\}\cap U_N(d,\theta+\theta',(\alpha+\beta)/2),$$

so such choice looks like a good candidate for an example with almost equality in Theorem 2.2.

For many subsets of the primes one can actually guarantee equidistribution in residue classes with small modulus, and so it is natural to ask when the lower bound in Theorem 2.2 is sharp. We show that something similar to the above example must happen.

THEOREM 2.3. *Let $\varepsilon > 0$ and $K \ge 1$. There exists $\gamma_1 = \gamma_1(K,\varepsilon)$ such that for each $\gamma_0 \in (0,\gamma_1)$, there exists $W = W(\gamma_0,\varepsilon)$ such that the following holds when*

$$\log\log n \gg W, \quad m = \prod_{p\le W} p, \quad \textit{and} \quad N \in \mathbb{P}\cap[2n/m, 4n/m].$$

Let $A, B \subseteq \mathbb{P} \cap [1, n]$ be such that $\alpha_m, \beta_m \in (\gamma_0, \gamma_1)$. If

$$|A + B| \leq K \cdot (\alpha_m \beta_m)^{1/2} n,$$

then there exist sequences $d_r, d'_r \in \mathbb{Z}_N$ and $\theta_r, \theta'_r \in \mathbb{R}$ for $r \in \mathbb{Z}_m^*$ such that

$$\Big| \bigcup_{r \in \mathbb{Z}_m^*} \{r + km \in A_m[r] \colon k \in U_N(d_r, \theta_r, \varepsilon)\} \Big| \geq (1 - \varepsilon)|A|,$$

$$\Big| \bigcup_{r \in \mathbb{Z}_m^*} \{r + km \in B_m[r] \colon k \in U_N(d'_r, \theta'_r, \varepsilon)\} \Big| \geq (1 - \varepsilon)|B|.$$

This is the first result in this direction and is quite unsatisfying for several reasons:

(i) It is inelegant.
(ii) It only works for small α and β (though one could probably prove a weaker result for larger α and β).
(iii) It tells nothing about d_r and d'_r (the proof tells something).
(iv) The (hidden) dependencies are quite bad.

3. An outline of the arguments. Let us first describe the elements of the proof of Theorem 2.1. Recalling the notation $C_q[r]$ from (2.1), we will prove that an analogue of Theorem 2.1 holds inside single residue classes modulo the product of sufficiently many smallest primes.

PROPOSITION 3.1. *Let $\varepsilon > 0$. There exists $W_0 = W_0(\varepsilon)$ such that the following holds when*

$$W_0 \leq W \ll \log \log n, \quad m = \prod_{p \leq W} p, \quad \textit{and} \quad r, s \in \mathbb{Z}_m^*.$$

Let $A \subseteq \mathbb{P}_m[r] \cap [1, n]$ and $B \subseteq \mathbb{P}_m[s] \cap [1, n]$ with relative densities $\alpha, \beta > \varepsilon$. Then

$$|A + B| \geq (1 - \varepsilon)(\alpha + \beta)\frac{n}{m}.$$

This is proved in Section 6 refining the work of Chipeniuk and Hamel [3], which itself, similarly to work of Hamel and Łaba [10] on sum sets of subsets of random sets, adapts the method of Green [4] and Green–Tao [7]. The process is roughly:

1. Embed A and B into $\mathbb{Z}_N$ with $N \asymp n/m$.
2. Convolve weighted characteristic functions f and g of these sets.
3. Split f and g into bounded and "uniform" parts.
4. Uniform parts contribute much to the convolution only rarely while bounded parts contribute a positive main term.

Proposition 3.1 clearly implies that, for any $\varepsilon_0 > 0$, $W_0(\varepsilon_0) \leq W \ll \log\log n$ and $A, B \subseteq \mathbb{P} \cap [1, n]$,

$$|A+B| = \sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ a,b\in\mathbb{Z}_m^*}} |A_m[a] + B_m[b]| \tag{3.1}$$
$$\geq (1-\varepsilon_0) \sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ \alpha_m[a],\beta_m[b]>\varepsilon_0}} \{\alpha_m[a] + \beta_m[b]\} \frac{n}{m}.$$

The right hand side can be evaluated using the following weighted version of Theorem 1.3.

Corollary 3.2. *Let m be square-free and let $u_r, v_r \in [0,1]$ for each $r \in \mathbb{Z}_m^*$. Assume that $u = \mathbb{E}_{r\in\mathbb{Z}_m^*} u_r$ and $v = \mathbb{E}_{r\in\mathbb{Z}_m^*} v_r$ are positive. Then*

$$\sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ u_a v_b \neq 0}} \{u_a + v_b\} \geq (1 - o_{u+v\to 0}(1)) \frac{u+v}{e^\gamma \log\log(1/(uv))} m.$$

This corollary is derived from Theorem 1.3 at the end of Section 5.

Deduction of Theorem 2.1 from (3.1) *and Corollary 3.2.* We can assume that ε is small and that γ_1 is so small that $o_{u+v\to 0}(1)$-term in Corollary 3.2 is $< \varepsilon^2$ whenever $u, v \in (0, \gamma_1)$. Let $\varepsilon_0 = \varepsilon^2\gamma_0^2$ and let W be $W_0(\varepsilon_0)$ in Proposition 3.1 and $m = \prod_{p\leq W} p$ and $n_0 = \exp(\exp(W))$. For every $r \in \mathbb{Z}_m^*$, define

$$u_r = \max\{0, \alpha_m[r] - \varepsilon_0\} \quad \text{and} \quad v_r = \max\{0, \beta_m[r] - \varepsilon_0\}.$$

Notice that $\mathbb{E}_{r\in\mathbb{Z}_m^*} u_r \geq \alpha - \varepsilon_0$ and $\mathbb{E}_{r\in\mathbb{Z}_m^*} v_r \geq \beta - \varepsilon_0$. Applying first (3.1) and then Corollary 3.2, we see that

$$|A+B| \geq (1-\varepsilon_0) \sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ u_a,v_b>0}} \{u_a + v_b\} \frac{n}{m} \geq (1-\varepsilon^2)^2 \frac{\alpha+\beta-2\varepsilon_0}{e^\gamma \log\log(1/(\alpha\beta))} n$$
$$\geq (1-\varepsilon) \frac{\alpha+\beta}{e^\gamma \log\log(1/(\alpha\beta))} n$$

when ε is small enough. ∎

In the first part of Section 5 we will prove the following more precise statement instead of Theorem 1.3. The proof proceeds by induction on the number of prime factors of m.

Theorem 3.3. *Let $m = p_1 \cdots p_k$ where $p_1 < \cdots < p_k$ and let $A, B \subseteq \mathbb{Z}_m^*$. Assume that*

$$|B| \geq \varphi(m) \prod_{i=1}^{l} \frac{2}{\varphi(p_i)} \tag{3.2}$$

for some $l \in \{0, \dots, k\}$. *Then*

$$|A+B| \geq |A| \frac{m}{\varphi(m)} \cdot \frac{\varphi(p_1 \cdots p_l)}{p_1 \cdots p_l}. \tag{3.3}$$

A natural guess is that this holds with the factor 2 replaced by 1, which would be best possible when the counterpart of (3.2) holds with equality:

Example 3.4. Let m be as in Theorem 3.3 and let $l \leq k$ and $A_0 \subseteq \mathbb{Z}^*_{p_1 \cdots p_l}$ be non-empty. Choose

$$\begin{aligned} B &= \{b \in \mathbb{Z}_m^* \colon b \equiv 1 \pmod{p_1 \cdots p_l}\}, \\ A &= \{a \in \mathbb{Z}_m^* \colon a \ (\mathrm{mod}\ p_1 \cdots p_l) \in A_0\}. \end{aligned}$$

Then $|B| = \varphi(m) \prod_{i=1}^{l} \frac{1}{\varphi(p_i)}$ and

$$|A+B| = |\{c \in \mathbb{Z}_m \colon c - 1 \ (\mathrm{mod}\ p_1 \cdots p_l) \in A_0\}| = |A| \frac{m}{\varphi(m)} \cdot \frac{\varphi(p_1 \cdots p_l)}{p_1 \cdots p_l},$$

so equality holds in (3.3).

When p_i are the first l primes, the right hand side is by Lemma 4.1 below

$$(1 + o_{l \to \infty}(1)) \frac{|A|/\varphi(m)}{e^\gamma \log\log(\varphi(m)/|B|)} m,$$

which shows that the bound in Theorem 1.3 is asymptotically best possible. This implies that also Theorem 1.1 is asymptotically best possible.

Deduction of Theorem 1.3 from Theorem 3.3. Let $m = p_1 \cdots p_k$ with $p_1 < \cdots < p_k$. Given β, choose l for which

$$\prod_{i=1}^{l} \frac{2}{\varphi(p_i)} \leq \beta < \prod_{i=1}^{l-1} \frac{2}{\varphi(p_i)},$$

so (3.3) holds by Theorem 3.3. By Lemma 4.1 below,

$$\frac{\varphi(p_1 \cdots p_l)}{p_1 \cdots p_l} \geq \frac{1 - o_{\beta \to 0}(1)}{e^\gamma \log\log(1/\beta)},$$

and the claim follows. ∎

For the proof of Theorem 1.1, we still need to

1. prove Proposition 3.1 (in Section 6);
2. prove Theorem 3.3 (in Section 5);
3. show that Theorem 1.3 implies Corollary 3.2 (in Section 5).

Theorem 1.4 will be derived from Proposition 3.1 and Theorem 3.3 in Section 7. Theorem 2.3 will be proved in Section 8. In this case we have automatically satisfactory distribution modulo m, so we only need to prove the following counterpart of Proposition 3.1 from which Theorem 2.3 follows easily.

Proposition 3.5. *Let $\varepsilon > 0$ and $K \geq 1$. There exists $\gamma_1 = \gamma_1(K, \varepsilon)$ such that for each $\gamma_0 \in (0, \gamma_1)$, there exists $W = W(\gamma_0, \varepsilon)$ such that the following holds when*

$$\log\log n \gg W, \quad m = \prod_{p \leq W} p, \quad r, s \in \mathbb{Z}_m^*, \quad \text{and} \quad N \in \mathbb{P} \cap [2n/m, 4n/m].$$

Let $A \subseteq \mathbb{P}_m[r] \cap [1, n]$ and $B \subseteq \mathbb{P}_m[s] \cap [1, n]$ with relative densities $\alpha, \beta > \varepsilon$. If

$$|A + B| \leq K(\alpha\beta)^{1/2} \frac{n}{m},$$

then there exist $d \in \mathbb{Z}_N^*$ and $\theta, \theta' \in \mathbb{R}$ such that

$$|\{r + km \in A \colon k \in U(d, \theta, \varepsilon)\}| \geq (1 - \varepsilon)|A|,$$
$$|\{s + km \in B \colon k \in U(d, \theta', \varepsilon)\}| \geq (1 - \varepsilon)|B|.$$

The proof of this is similar to the proof of Proposition 3.1. The main additional tool is a theorem due to Green and Ruzsa [5], which says that any set in $\mathbb{Z}_N$ with a low but positive density and small doubling is contained in some set $U_N(d, \theta, \delta)$ (see Lemma 4.7 below).

4. Auxiliary results

4.1. Lemmas needed in the proof of Theorem 1.1. Let us first quickly prove the well-known result on $m/\varphi(m)$ which we have already used many times.

Lemma 4.1. *For $m \in \mathbb{N}$,*

$$m/\varphi(m) \leq (1 + o_{m\to\infty}(1))e^\gamma \log\log \varphi(m).$$

If m is the product of all primes up to some $M \geq 1$, "$\leq$" can be replaced by "$=$".

Proof. When m is the product of all primes $\leq M$, the claim follows from Mertens' formula since, by the prime number theorem, $M = (1 + o_{m\to\infty}(1)) \log m$. Otherwise let q_i be the ith prime and choose l such that

$$q_1 \cdots q_l \geq m > q_1 \cdots q_{l-1}.$$

Then it is easy to see that

$$\frac{m}{\varphi(m)} \leq \frac{q_1 \cdots q_l}{\varphi(q_1 \cdots q_l)} = (1 + o_{l\to\infty}(1))e^\gamma \log\log \varphi(q_1 \cdots q_{l-1})$$
$$\leq (1 + o_{m\to\infty}(1))e^\gamma \log\log \varphi(m). \ \blacksquare$$

The second lemma is a very simple yet useful inequality.

LEMMA 4.2. *Let $\alpha_1 \geq \cdots \geq \alpha_k \geq 0$ and let δ and $\delta_1, \ldots, \delta_k$ be real numbers such that*

$$\frac{1}{j}\sum_{i=1}^{j}\delta_i \geq \delta \quad \textit{for every } j = 1, \ldots, k.$$

Then

$$\sum_{i=1}^{k}\alpha_i\delta_i \geq \delta\sum_{i=1}^{k}\alpha_i. \tag{4.1}$$

Proof. Writing $\alpha_{k+1} = 0$, the left hand side of (4.1) is, by partial summation,

$$\sum_{j=1}^{k}(\alpha_j - \alpha_{j+1})\sum_{i=1}^{j}\delta_i \geq \delta\sum_{j=1}^{k}j(\alpha_j - \alpha_{j+1}) = \delta\sum_{i=1}^{k}\alpha_i. \blacksquare$$

In the following lemma we write $r_{A+B}(n)$ for the number of representations of n as a sum $a + b$ with $a \in A$ and $b \in B$.

LEMMA 4.3. *Let $p \in \mathbb{P}$ and $A, B \subseteq \mathbb{Z}_p$. Then, for any $m \leq r \leq \min\{|A|, |B|\}$,*

$$|\{n \in \mathbb{Z}_p \colon r_{A+B}(n) \geq m\}| \geq \min\left\{p, |A| + |B| - r - \frac{m-1}{r}p\right\}. \tag{4.2}$$

Proof. This follows as in [1, Proposition 4]: Pollard's generalisation [13] of the Cauchy–Davenport inequality gives

$$\sum_{n\in\mathbb{Z}_p}\min\{r, r_{A+B}(n)\} \geq r\min\{p, |A| + |B| - r\}. \tag{4.3}$$

Writing N_m for the left hand side of (4.2), the left hand side of (4.3) is $\leq rN_m + (m-1)(p - N_m)$. Rearranging, one gets

$$N_m \geq \min\left\{p, \frac{(|A| + |B| - r)r - (m-1)p}{r - m + 1}\right\}$$

and the claim follows since $r - m + 1 \leq r$. $\blacksquare$

Next we show that we can restrict A and B to so-called downsets. For this notion and further compression operations in related problems, see for instance [2, 8]. In the following definition we order the elements of $\mathbb{Z}_p$ in the natural way $0 < 1 < 2 < \cdots < p - 1$.

DEFINITION 4.4. A set $C \subseteq \prod_{i=1}^{k}\mathbb{Z}_{p_i}^{*}$ is a *downset* if $(i_1, \ldots, i_k) \in C$ whenever there exists $(j_1, \ldots, j_k) \in C$ such that $1 \leq i_l \leq j_l$ for every $l = 1, \ldots, k$.

Lemma 4.5. *Let $p_1 < \cdots < p_k$, $G = \prod_{i=1}^k \mathbb{Z}_{p_i}$ and $A, B \subseteq G^*$. Then there exist downsets $A', B' \subseteq G^*$ such that*

$$|A'| = |A|, \quad |B'| = |B|, \quad \text{and} \quad |A' + B'| \le |A + B|.$$

Proof. Stronger statements in $\mathbb{Z}_n^k$ and $\mathbb{Z}_2^k$ are proved in [2, 8]. For completeness we provide a proof in our setting.

Write $m = p_1 \cdots p_k$ and think of G as $\mathbb{Z}_{m/p_k} \times \mathbb{Z}_{p_k}$. For $q \,|\, m$, write $\pi_q \colon \mathbb{Z}_m \to \mathbb{Z}_q$ for the reduction map $\pi_q(x) = x \pmod q$ and recall the notation $C_q[r]$ from (2.1). Notice that, for any $C \subseteq G$,

$$C = \bigcup_{c_1 \in \pi_{m/p_k}(C)} \{c_1\} \times C_{m/p_k}[c_1].$$

Set

$$A^{(k)} = \bigcup_{a_1 \in \pi_{m/p_k}(A)} \{a_1\} \times \{1, \ldots, |A_{m/p_k}[a_1]|\},$$
$$B^{(k)} = \bigcup_{b_1 \in \pi_{m/p_k}(B)} \{b_1\} \times \{1, \ldots, |B_{m/p_k}[b_1]|\}.$$

Clearly $A^{(k)}, B^{(k)} \subseteq G^*$, $|A^{(k)}| = |A|$ and $|B^{(k)}| = |B|$. Furthermore

$$\begin{aligned}
|A + B| &= \sum_{n \in \pi_{m/p_k}(A+B)} |(A+B)_{m/p_k}[n]| \\
&\ge \sum_{n \in \pi_{m/p_k}(A+B)} \max_{\substack{a \in \pi_{m/p_k}(A) \\ b \in \pi_{m/p_k}(B) \\ a+b=n}} \{|A_{m/p_k}[a] + B_{m/p_k}[b]|\} \\
&\ge \sum_{n \in \pi_{m/p_k}(A+B)} \max_{\substack{a \in \pi_{m/p_k}(A) \\ b \in \pi_{m/p_k}(B) \\ a+b=n}} \{\min\{p_k, |A_{m/p_k}[a]| + |B_{m/p_k}[b]| - 1\}\} \\
&= \sum_{n \in \pi_{m/p_k}(A^{(k)}+B^{(k)})} |(A^{(k)} + B^{(k)})_{m/p_k}[n]| = |A^{(k)} + B^{(k)}|,
\end{aligned}$$

where the second inequality follows from the Cauchy–Davenport inequality (the case $m = r = 1$ of Lemma 4.3).

Now the sets $A^{(k)}$ and $B^{(k)}$ have a downset type property with respect to the last coordinate (i.e. $(a_1, a_2) \in A^{(k)}$ whenever $(a_1, a_2') \in A^{(k)}$ for some $a_2' \ge a_2 \ge 1$). Applying the same process to each of the remaining $k-1$ coordinates in turn and noticing that the process does not spoil the downsetness of coordinates handled before, we finally end up with a downset with the desired properties. ■

The following lemma shows how a certain structure in B forces $A + B$ to be large.

Lemma 4.6. *Let $p_k > \cdots > p_1 > 2$ and $G = \prod_{i=1}^k \mathbb{Z}_{p_i}$. Let $A \subseteq G^*$, $\mathcal{I} \subseteq \{1, \ldots, k\}$ and*

$$B_{\mathcal{I}} = \{(j_1, \ldots, j_k) \in G^* \colon j_i = 1 \text{ for } i \notin \mathcal{I}, \text{ and } j_i \in \{1,2\} \text{ for } i \in \mathcal{I}\}.$$

Then

$$|A + B_{\mathcal{I}}| \geq |A| \prod_{i \in \mathcal{I}} \frac{p_i}{\varphi(p_i)}.$$

Proof. Write $\mathbf{0} := (0, \ldots, 0) \in G$, $\mathbf{1} := (1, \ldots, 1) \in G$, and $\boldsymbol{e}_i$ for the element of G with the ith component 1 and others 0. Notice that for any set $C \subseteq G^*$ one has $|C + \{\mathbf{0}, \boldsymbol{e}_i\}| \geq |C| p_i/\varphi(p_i)$. Hence

$$|A + B_{\mathcal{I}}| = \Big|A + \{\mathbf{1}\} + \sum_{i \in \mathcal{I}} \{\mathbf{0}, \boldsymbol{e}_i\}\Big| \geq |A| \prod_{i \in \mathcal{I}} \frac{p_i}{\varphi(p_i)}. \ \blacksquare$$

We are not going to use this lemma very much, but it somewhat reveals what is happening and also suggests an alternative approach to Theorem 1.3: One could try to show that any large enough set B must contain $B_{\mathcal{I}}$ for some large $\mathcal{I}$. By Lemma 4.5 we can assume that A and B are downsets and thus B is surely more likely to contain large $B_{\mathcal{I}}$ than a typical set but there is still no guarantee that B contains a large $B_{\mathcal{I}}$. However, using [8, Proposition 3.2] (expansion in Hamming balls) one can show that $3B$ contains $B_{\mathcal{I}}$ so large that the lower bound in Theorem 1.3 follows for $|A + 3B|$.

4.2. Lemmas needed in the proof of Theorem 2.3. Recall the definition of $U_N(d, \theta, \delta)$ from (2.2). Noticing that, for any $N \geq 1$, $d \in \mathbb{Z}_N^*$, $a \in \mathbb{Z}_N$ and $\delta > 0$,

$$U_N(d, a/N, \delta) = \{a\overline{d} + i\overline{d} \in \mathbb{Z}_N \colon -\delta N \leq i \leq \delta N\},$$

the following lemma is an immediate consequence of [5, Theorem 1.2]:

Lemma 4.7. *Let $\delta > 0$ and $K \geq 1$. There exists $\gamma_0 = \gamma_0(K, \delta)$ such that the following holds when $N \in \mathbb{P}$.*

Let $A \subseteq \mathbb{Z}_N$ be such that $|A| \leq \gamma_0 N$ and $|A - A| \leq K|A|$. Then there exist $d \in \mathbb{Z}_N^*$ and $\theta \in \mathbb{R}$ such that

$$A \subseteq U_N(d, \theta, \delta).$$

For a function $h \colon \mathbb{Z}_N \to \mathbb{C}$ define the normalized Fourier transform

$$\widehat{h}(\xi) := \frac{1}{N} \sum_{x \in \mathbb{Z}_N} h(x) e(-x\xi/N).$$

The following lemma is a rather simple form of the well-known principle that large Fourier coefficients mean regularity (for a sharp result for (i), see [5, Lemma 3.2] which is a rewording of [11, Theorem 1]).

LEMMA 4.8. *Let $N \geq 1$, $f\colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$, $\xi \in \mathbb{Z}_N^*$ and $\delta, \varepsilon > 0$.*

(i) *If $|\widehat{f}(\xi)| \geq (1 - 8\varepsilon\delta^2)\widehat{f}(0)$, then there exists $\theta \in \mathbb{R}$ such that*

$$\sum_{n \in U_N(\xi,\theta,\delta)} f(n) \geq (1-\varepsilon) \sum_{n \in \mathbb{Z}_N} f(n).$$

(ii) *If there exists $\theta \in \mathbb{R}$ such that*

$$\sum_{n \in U_N(\xi,\theta,\delta)} f(n) \geq (1-\varepsilon) \sum_{n \in \mathbb{Z}_N} f(n),$$

then $|\widehat{f}(\xi)| \geq (1 - 2\varepsilon - 20\delta^2)\widehat{f}(0)$.

Proof. For any $\theta \in \mathbb{R}$,

$$\begin{aligned}\mathrm{Re}(\widehat{f}(\xi)e(\theta)) = \frac{1}{N}\biggl(& \sum_{n \in U_N(\xi,\theta,\delta)} f(n) \cos\Bigl(2\pi\Bigl(\frac{\xi}{N}n - \theta\Bigr)\Bigr) \\ & + \sum_{n \notin U_N(\xi,\theta,\delta)} f(n) \cos\Bigl(2\pi\Bigl(\frac{\xi}{N}n - \theta\Bigr)\Bigr)\biggr).\end{aligned}$$

To prove (i), choose θ such that $|\widehat{f}(\xi)| = e(\theta)\widehat{f}(\xi)$. Since $\cos(2\pi x) \leq 1 - 8x^2$ for $x \in [-1/2, 1/2]$ we get

$$\begin{aligned}(1 - 8\varepsilon\delta^2)\widehat{f}(0) \leq |\widehat{f}(\xi)| &= \mathrm{Re}(\widehat{f}(\xi)e(\theta)) \\ &\leq \frac{1}{N}\Bigl(\sum_{n \in U_N(\xi,\theta,\delta)} f(n) + (1 - 8\delta^2) \sum_{n \notin U_N(\xi,\theta,\delta)} f(n)\Bigr) \\ &= \frac{1}{N}\Bigl((1 - 8\delta^2) \sum_{n \in \mathbb{Z}_N} f(n) + 8\delta^2 \sum_{n \in U_N(\xi,\theta,\delta)} f(n)\Bigr),\end{aligned}$$

from which the claim follows by rearranging.

To prove (ii), choose θ as in the claim. Since $\cos(2\pi x) \geq \max\{-1, 1 - 20x^2\}$,

$$\begin{aligned}|\widehat{f}(\xi)| \geq \mathrm{Re}(\widehat{f}(\xi)e(\theta)) &\geq \frac{1}{N}\Bigl((1 - 20\delta^2) \sum_{n \in U_N(\xi,\theta,\delta)} f(n) - \sum_{n \notin U_N(\xi,\theta,\delta)} f(n)\Bigr) \\ &\geq (1 - 20\delta^2 - 2\varepsilon)\hat{f}(0). \ \blacksquare\end{aligned}$$

For a set $E \subseteq A \times B$, we write

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

for the restricted sum set. The Balog–Szemerédi–Gowers theorem (see [16, Theorem 2.29]) lets us pass from a restricted sum set to a normal sum set. We need the following version which is a mixture of [16, Theorem 2.29

and Exercise 2.5.4] and can be proved by incorporating the hint in [16, Exercise 2.5.4] into the proof of the Balog–Szemerédi–Gowers theorem in [16, Section 6.4].

Lemma 4.9. *Let A and B be additive sets in a group Z, and let $E \subset A \times B$ be such that*

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

for some $\delta > 0$ and $K \leq 1$. Then there exist subsets $A' \subseteq A$ and $B' \subseteq B$ such that

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

5. Sum sets in $\mathbb{Z}_m^*$. In this section we execute steps 2 and 3 of the agenda at the end of Section 3.

Proof of Theorem 3.3. We can clearly assume that $p_1 > 2$ and by Lemma 4.5 we can assume that A and B are downsets. We start by handling a couple of extremal values of l:

- If $l = 0$, then $B = \mathbb{Z}_m^*$ and the claim follows from Lemma 4.6 with $\mathcal{I} = \{1, \ldots, k\}$.
- If $l = k$, then the claim becomes $|A+B| \geq |A|$, which is trivial.
- If $l = k-1$, then $|B| \geq p_k - 1 \geq 2$, and so, in the notation of Lemma 4.6, B contains $B_{\mathcal{I}}$ for some $|\mathcal{I}| \geq 1$ and the claim follows from Lemma 4.6.

Hence we can assume that $1 \leq l \leq k-2$. We prove the claim by induction on k, that is, the number of prime factors of m. The above takes care of $k \leq 2$, so we can proceed to the induction step. For that assume that the claim holds with m/p_1 in place of m.

Think of $\mathbb{Z}_m^*$ as $\mathbb{Z}_{p_1}^* \times \mathbb{Z}_{m/p_1}^*$ and write

$$A = \bigcup_{i=1}^{p_1-1} \{i\} \times A_i \quad \text{and} \quad B = \bigcup_{i=1}^{p_1-1} \{i\} \times B_i.$$

Since A and B are downsets, we have $|A_1| \geq \cdots \geq |A_{p_1-1}|$ and $|B_1| \geq \cdots \geq |B_{p_1-1}|$. We will split into three cases which we first discuss informally:

Case 1: If B_1 is very large, then $A + (\{1\} \times B_1)$ is large.

Case 2: If B_2 is not too small, then $(A + (\{2\} \times B_2)) \cup (\{2\} \times (A_1 + B_1))$ is large.

Case 3: If neither of these holds, then even $B_{(p_1+1)/2}$ must be rather large and we can again show that $A+B$ is large.

Let us now turn to the rigorous treatment of these three cases.

Case 1: $|B_1| \geq \varphi(m/p_1)\prod_{i=2}^{l} \frac{2}{\varphi(p_i)}$. Since $A+B$ contains

$$A + (\{1\} \times B_1) = \bigcup_{i=2}^{p_1} \{i\} \times (A_{i-1} + B_1),$$

by the induction hypothesis

$$\begin{aligned} |A+B| &\geq \sum_{i=1}^{p_1-1} |A_i + B_1| \geq \sum_{i=1}^{p_1-1} |A_i| \frac{m/p_1}{\varphi(m/p_1)} \cdot \frac{\varphi(p_2\cdots p_l)}{p_2\cdots p_l} \\ &= |A| \frac{m}{\varphi(m)} \cdot \frac{\varphi(p_1\cdots p_l)}{p_1\cdots p_l}. \end{aligned}$$

Case 2: $|B_2| \geq \varphi(m/p_1)\prod_{i=2}^{l+1} \frac{2}{\varphi(p_i)}$. Since $A+B$ contains

$$\{2\} \times (A_1 + B_1) \quad \text{and} \quad A + (\{2\} \times B_2) = \bigcup_{i=3}^{p_1+1} \{i\} \times (A_{i-2} + B_2),$$

by the induction hypothesis

$$\begin{aligned} |A+B| &\geq |A_1 + B_1| + \sum_{i=1}^{p_1-1} |A_i + B_2| \\ &\geq \Big(|A_1| + \sum_{i=1}^{p_1-1} |A_i|\Big) \frac{m/p_1}{\varphi(m/p_1)} \cdot \frac{\varphi(p_2\cdots p_{l+1})}{p_2\cdots p_{l+1}} \\ &\geq \frac{\varphi(p_{l+1})}{p_{l+1}} \Big(\frac{|A|}{p_1-1} + |A|\Big) \frac{m}{\varphi(m)} \cdot \frac{\varphi(p_1\cdots p_l)}{p_1\cdots p_l} \\ &\geq |A| \frac{m}{\varphi(m)} \cdot \frac{\varphi(p_1\cdots p_l)}{p_1\cdots p_l}. \end{aligned}$$

Case 3: $|B_1| < \varphi(m/p_1)\prod_{i=2}^{l} \frac{2}{\varphi(p_i)}$ and $|B_2| < \varphi(m/p_1)\prod_{i=2}^{l+1} \frac{2}{\varphi(p_i)}$. Notice that automatically

$$|B_1| \geq \frac{|B|}{p_1-1} \geq \frac{\varphi(m)}{p_1-1} \prod_{i=1}^{l} \frac{2}{\varphi(p_i)} \geq \varphi(m/p_1) \prod_{i=2}^{l+1} \frac{2}{\varphi(p_i)}.$$

Furthermore

$$|B| = |B_1| + \cdots + |B_{p_1-1}| \leq |B_1| + \frac{p_1-3}{2}|B_2| + \frac{p_1-1}{2}|B_{(p_1+1)/2}|,$$

so that

$$\begin{aligned} |B_{(p_1+1)/2}| &\geq \frac{2}{p_1-1}\Big(|B| - |B_1| - \frac{p_1-3}{2}|B_2|\Big) \\ &\geq \Big(\varphi(m/p_1)\prod_{i=2}^{l} \frac{2}{\varphi(p_i)}\Big) \cdot \frac{2}{p_1-1}\Big(2 - 1 - \frac{p_1-3}{2} \cdot \frac{2}{p_{l+1}-1}\Big) \end{aligned}$$

Now the product after $\cdot$ is

$$\geq \frac{2}{p_1-1} - \frac{2}{p_{l+1}-1} = \frac{2(p_{l+1}-p_1)}{(p_1-1)(p_{l+1}-1)} \geq \frac{4}{(p_{l+1}-1)(p_{l+2}-1)},$$

and hence

$$|B_{(p_1+1)/2}| \geq \varphi(m/p_1) \prod_{i=2}^{l+2} \frac{2}{\varphi(p_i)}.$$

Now $A+B$ contains the disjoint sets

$$\{1\} \times (A_{(p_1+1)/2} + B_{(p_1+1)/2}), \quad \left\{\frac{p_1+3}{2}\right\} \times (A_1 + B_{(p_1+1)/2}),$$
$$\{i\} \times (A_{i-1} + B_1) \quad \text{for } 2 \leq i \leq p_1,\ i \neq (p_1+3)/2,$$

so by the induction hypothesis

$$\begin{aligned} |A+B| &\geq (|A_1| + \cdots + |A_{p_1-1}| - |A_{(p_1+1)/2}|) \cdot \frac{m/p_1}{\varphi(m/p_1)} \cdot \frac{\varphi(p_2\cdots p_{l+1})}{p_2\cdots p_{l+1}} \\ &\quad + (|A_{(p_1+1)/2}| + |A_1|) \cdot \frac{m/p_1}{\varphi(m/p_1)} \cdot \frac{\varphi(p_2\cdots p_{l+2})}{p_2\cdots p_{l+2}} \\ &= \frac{m}{\varphi(m)} \prod_{i=1}^{l} \frac{\varphi(p_i)}{p_i} \\ &\quad \cdot \frac{\varphi(p_{l+1})}{p_{l+1}} \biggl(\sum_{i=1}^{p_1-1} |A_i| + \biggl(1 - \frac{1}{p_{l+2}}\biggr)|A_1| - \frac{1}{p_{l+2}}|A_{(p_1+1)/2}| \biggr). \end{aligned}$$

By Lemma 4.2 with $\alpha_j = |A_j|$ and $\delta = (p_1 - 2/p_{l+2})/(p_1-1)$, the product after $\cdot$ is

$$\geq \frac{p_{l+1}-1}{p_{l+1}} \cdot |A| \frac{p_1 - 2/p_{l+2}}{p_1-1} = |A| \biggl(1 + \frac{p_{l+1} - 2\frac{p_{l+1}}{p_{l+2}} - p_1 + \frac{2}{p_{l+2}}}{p_{l+1}(p_1-1)}\biggr) > |A|$$

since $p_{l+1} - p_1 \geq 2$. ∎

Proof of Corollary 3.2. For $x \in [0,2]$, let

$$A_x = \{a \in \mathbb{Z}_m^* \colon u_a > x\},\ B_x = \{b \in \mathbb{Z}_m^* \colon v_b > x\}, \quad C_x = \bigcup_{t\in[0,x]} (A_t + B_{x-t}).$$

Then

$$\sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c \\ u_a v_b \neq 0}} \{u_a + v_b\} = \sum_{c\in\mathbb{Z}_m} \int_{\substack{x \\ c\in C_x}} 1\, dx = \int_0^2 |C_x|\, dx. \tag{5.1}$$

Choose $v' \in [0,1]$ so that

$$\sum_{b\in\mathbb{Z}_m^*} \min\{v', v_b\} = (1-v)v\varphi(m). \tag{5.2}$$

Such a v' exists since the left hand side is continuous in v' and equals 0 for $v'=0$ and $v\varphi(m)$ for $v'=1$. By (5.2),

$$(1-v)v\varphi(m) = v\varphi(m) + \sum_{b\in B_{v'}} (v'-v_b) \geq v\varphi(m) - |B_{v'}| \ \Rightarrow\ \frac{|B_{v'}|}{\varphi(m)} \geq v^2.$$

By (5.1),

$$\sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ u_a v_b\neq 0}} \{u_a+v_b\} \geq \int_0^{v'} |A_0+B_x|\,dx + \int_{v'}^{v'+1} |A_{x-v'}+B_{v'}|\,dx.$$

Applying Theorem 1.3 to each sum set shows that this is

$$\begin{aligned}
&\geq (1-o_{u+v\to 0}(1))\frac{m}{\varphi(m)}\Big(\int_0^{v'} \frac{|B_x|}{e^\gamma \log\log(1/u)}\,dx + \int_0^1 \frac{|A_x|}{e^\gamma\log\log(1/v^2)}\,dx\Big)\\
&\geq \frac{1-o_{u+v\to 0}(1)}{e^\gamma \log\log(1/(uv))}\cdot\frac{m}{\varphi(m)}\Big(\sum_{b\in\mathbb{Z}_m^*}\min\{v_b,v'\} + \sum_{a\in\mathbb{Z}_m^*} u_a\Big)\\
&\geq (1-o_{u+v\to 0}(1))\frac{u+v}{e^\gamma\log\log(1/(uv))}m. \ \blacksquare
\end{aligned}$$

6. Proof of Proposition 3.1. In this section we prove Proposition 3.1 following arguments in [3]. As a first step we reduce to $\mathbb{Z}_N$ for a prime $N \asymp n/m$ and introduce weights coming from Green's modified von Mangoldt function $\lambda_{d,m,N}\colon \mathbb{Z}^+ \to \mathbb{R}$ defined by

$$\lambda_{d,m,N}(x) := \begin{cases} \dfrac{\varphi(m)}{mN}\log(mx+d) & \text{if } x\leq N \text{ and } mx+d \text{ is prime},\\ 0 & \text{otherwise.}\end{cases}$$

For any $N > n/m$ and set $C\subseteq \mathbb{P}_m[d]\cap[1,n]$ we define a set $\widetilde{C}$ and a function $h_C\colon \mathbb{Z}_N\to\mathbb{R}_{\geq 0}$ by

$$\widetilde{C} := (m^{-1}(C-d))\cap\{1,\dots,N\} \quad\text{and}\quad h_C(x) := N\,1_{\widetilde{C}}(x)\lambda_{d,m,N}(x).$$

Choice of N, d and m will always be clear from the context. Notice that $|C| = |\widetilde{C}|$. The prime number theorem in arithmetic progressions implies the following.

Lemma 6.1. *Let $\gamma_0>0$, $N>n/m$ with $m \ll (\log n)^A$ for some absolute constant A. Let $d\in\mathbb{Z}_m^*$ and assume that $C\subseteq\mathbb{P}_m[d]\cap[1,n]$ has positive relative density $\gamma'>\gamma_0$. Then*

$$\sum_{x\in\mathbb{Z}_N} h_C(x) = (1+o_{n\to\infty}(1))\gamma'\frac{n}{m}.$$

Proof. This follows by an obvious modification of [3, proof of Lemma 8]. ■

Let m, r, s, A and B be as in Proposition 3.1 and think of $\widetilde{A}$ and $\widetilde{B}$ as subsets of $\mathbb{Z}_N$. Then taking $N > 2n/m$ we have

$$|A+B| = |\widetilde{A}+\widetilde{B}|,$$

and by the above lemma,

$$\mathbb{E}h_A = (1+o_{n\to\infty}(1))\alpha\frac{n}{mN} \quad \text{and} \quad \mathbb{E}h_B = (1+o_{n\to\infty}(1))\beta\frac{n}{mN}.$$

For two functions $f, g\colon \mathbb{Z}_N \to \mathbb{C}$, we define the convolution

$$f * g(x) := \sum_{y\in\mathbb{Z}_N} f(y)g(x-y).$$

Since $\operatorname{supp}(f*g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g)$, Proposition 3.1 follows from the following lemma.

Lemma 6.2. *Let $\varepsilon > 0$. There exists $W_0 = W_0(\varepsilon)$ such that the following holds when*

$$W_0 \le W \ll \log\log n, \quad m = \prod_{p\le W} p.$$

If A and B are as in Proposition 3.1, *then*

$$|\{x \in \mathbb{Z}_N\colon (h_A * h_B)(x) > 0\}| \ge (1-\varepsilon)\min\{\mathbb{E}h_A + \mathbb{E}h_B, 1\}\cdot N.$$

Before going to the proof we need some notation. For $f\colon \mathbb{Z}_N \to \mathbb{C}$, we use the standard norm notations

$$\|f\|_p := \Big(\sum_{x\in\mathbb{Z}_N} |f(x)|^p\Big)^{1/p} \quad \text{for } 0 < p < \infty, \qquad \|f\|_\infty := \max_{x\in\mathbb{Z}_N} |f(x)|.$$

Furthermore, we say that a function $\nu\colon \mathbb{Z}_N \to \mathbb{R}_{\ge 0}$ is *η-pseudorandom* if

$$\|\widehat{\nu} - 1_{\xi=0}\|_\infty \le \eta.$$

We will be particularly interested in functions satisfying the following definition.

Definition 6.3. Let $\mathcal{F}(\eta, C)$ be the set of functions $f\colon \mathbb{Z}_N \to \mathbb{R}_{\ge 0}$ such that

(i) f is majorized by some η-pseudorandom function ν (i.e. $f(x) \le \nu(x)$ for every $x \in \mathbb{Z}_N$),

(ii) $\|\widehat{f}\|_3 \le C$.

Our interest is understandable due to the following lemma.

Lemma 6.4. *Let $\varepsilon > 0$. There exist $C = C(\varepsilon)$ and $W_0 = W_0(\varepsilon)$ such that the following holds when*

$$W_0 \le W \le \log\log N, \quad m = \prod_{p\le W} p \quad \textit{and} \quad r \in \mathbb{Z}_m^*.$$

Let $C \subseteq \mathbb{P}_m[r]$ with relative density at least ε. Then

$$h_C \in \mathcal{F}(2(\log\log W)/W, C).$$

Proof. This follows with $\nu = N\lambda_{r,m,N}$ from work of Green [4, Lemma 6.2, its proof and Lemma 6.6]. ∎

Next we quote a result which says that any $f \in \mathcal{F}(\eta, C)$ can be divided into bounded and uniform components.

LEMMA 6.5. *Let ε_0, σ and C be positive parameters. Then there exists $\eta = \eta(\varepsilon_0, \sigma, C)$ such that, for any $f \in \mathcal{F}(\eta, C)$, there exist functions $f_1, f_2 \colon \mathbb{Z}_N \to \mathbb{R}_{\geq 0}$ such that $f = f_1 + f_2$ and*

(i) $0 \leq f_1(x) \leq 1 + \sigma$ *for all* $x \in \mathbb{Z}_N$*;*
(ii) $\mathbb{E}f_1 = \mathbb{E}f$*;*
(iii) $\|\widehat{f_1}\|_\infty \leq 1 + \sigma$ *and* $\|\widehat{f_2}\|_\infty \leq \varepsilon_0$*;*
(iv) $\|\widehat{f_i}\|_3 \leq C$ *for* $i = 1, 2$.

Proof. This is originally due to Green [4] and is contained in the proof of [6, Proposition 5.1] (see also [3, Lemma 11]). ∎

We will show the following strengthening of [3, Lemma 13]. This together with Lemma 6.4 obviously implies Lemma 6.2 and hence Proposition 3.1.

PROPOSITION 6.6. *For every $\varepsilon > 0$ and $C > 0$, there exists $\eta = \eta(\varepsilon, C)$ such that if $N \in \mathbb{P}$ and $f, g \in \mathcal{F}(\eta, C)$ are such that $\mathbb{E}f = \alpha \in (\varepsilon, 1]$ and $\mathbb{E}g = \beta \in (\varepsilon, 1]$, then*

$$|\{x \in \mathbb{Z}_N \colon (f * g)(x) > 0\}| \geq (1 - \varepsilon)\min\{\alpha + \beta, 1\} \cdot N.$$

The main difference to [3, Lemma 13] is that there was $\frac{1}{2}(\alpha + \beta)$ in the place of $\alpha + \beta$. To get this sharper result we utilize Pollard's theorem (Lemma 4.3) instead of looking at L_1-norm estimates as in [3]. Note that Pollard's theorem was invoked in a similar context also in the work of Li and Pan [12] on the ternary problem.

Proof of Proposition 6.6. Let $\sigma = \varepsilon^3/10 < \alpha\beta\varepsilon/10$ and let ε_0 be small and η be $\eta(\varepsilon_0, \sigma, C)$ from Lemma 6.5. Write $f = f_1 + f_2$ and $g = g_1 + g_2$ as in Lemma 6.5. The claim follows once we have shown that

$$|\{x \in \mathbb{Z}_N \colon (f_1 * g_1)(x) \geq \sigma^4\alpha\beta N\}| \geq \min\{\alpha + \beta - 6\sigma, 1\}N$$

and

$$\big|\big\{x \in \mathbb{Z}_N \colon |(f_i * g_j)(x)| \geq \tfrac{1}{10}\sigma^4\alpha\beta N\big\}\big| \leq \sigma N \quad \text{for } (i, j) \neq (1, 1).$$

The latter follows for small enough ε_0 (depending on ε and C) by estimating the L_2-norm of $f_i * g_j$ exactly as in [3, proof of Lemma 13]. For the former, write

$$A = \{a \in \mathbb{Z}_N \colon f_1(a) \geq \alpha\sigma\} \quad \text{and} \quad B = \{b \in \mathbb{Z}_N \colon g_1(b) \geq \beta\sigma\}.$$

Now

$$\alpha = \mathbb{E} f_1 \leq \frac{1}{N}(\alpha\sigma N + (1+\sigma)|A|) \ \Rightarrow\ |A| \geq \frac{1-\sigma}{1+\sigma}\alpha N \geq (1-2\sigma)\alpha N$$

and similarly $|B| \geq (1-2\sigma)\beta N$. Hence, by Lemma 4.3 with $m = \sigma^2 N$ and $r = \sigma N$, $r_{A+B}(n) \geq \sigma^2 N$ for at least

$$\min\{1, (\alpha+\beta)(1-2\sigma) - 2\sigma\}N \geq \min\{1, \alpha+\beta-6\sigma\}N$$

values $n \in \mathbb{Z}_N$. But clearly $f_1 * g_1(n) \geq \sigma^4 \alpha\beta N$ for these n, finishing the proof. ∎

As pointed out before the proof, this implies Proposition 3.1, which in turn, together with Corollary 3.2 proved in the previous section, implies Theorem 2.1 as shown in Section 3. Hence also Theorem 1.1 follows. ∎

7. Proof of Theorem 1.4. The proof of Theorem 1.4 is a modification of the proof of Theorem 1.1 and does not need any new ideas but we provide the proof for completeness.

Let $\varepsilon > 0$ be so small that $(1-3\varepsilon)\beta > \prod_{i=2}^{l} \frac{2}{\varphi(q_i)}$, let $\varepsilon_0 = \alpha\beta\varepsilon^2/2$, $W = W_0(\varepsilon_0)$ in Proposition 3.1, and let $n \gg \exp(\exp(W))/\varepsilon$ be so large that

$$\frac{|A \cap [3,(1-\varepsilon)n]|}{|\mathbb{P} \cap [1,(1-\varepsilon)n]|} \geq (1-\varepsilon)\alpha \quad \text{and} \quad \frac{|B \cap [3,\varepsilon n]|}{|\mathbb{P} \cap [1,\varepsilon n]|} \geq (1-\varepsilon)\beta.$$

Analogously to the deduction of Theorem 1.1 from Theorem 2.1 we take $A' = A \cap [3,(1-\varepsilon)n]$ and $B' = B \cap [3,\varepsilon n]$. With $m = \prod_{2<p\leq W} p$, let

$$D = \left\{ r \in \mathbb{Z}_m^* \colon \frac{|B'_m[r]|}{|\mathbb{P}_m[r] \cap [1,\varepsilon n]|} \geq \varepsilon\beta \right\},$$

and notice that when n is large enough,

$$\begin{aligned}(1-\varepsilon)\beta \leq \frac{|B'|}{|\mathbb{P} \cap [1,\varepsilon n]|} &\leq \sum_{r \in \mathbb{Z}_m^*} \frac{|B'_m[r]|}{(1-\varepsilon)\varphi(m)|\mathbb{P}_m[r] \cap [1,\varepsilon n]|} \\ &\leq \frac{1}{1-\varepsilon}\left(\frac{|D|}{\varphi(m)} + \varepsilon\beta\right),\end{aligned}$$

so that $|D| \geq (1-3\varepsilon)\beta\varphi(m)$.

For any $q \geq 1$ and $s \in \mathbb{Z}_q$, let

$$\alpha'_q[s] = \frac{|A'_q[s]|}{|\mathbb{P}_q[s] \cap [1,n]|} \quad \text{and} \quad \beta'_q[s] = \frac{|B'_q[s]|}{|\mathbb{P}_q[s] \cap [1,n]|}.$$

By Proposition 3.1,

$$\begin{aligned}|(A+B)\cap[1,n]| \geq |A'+B'| &= \sum_{c\in\mathbb{Z}_{2m}} \max_{\substack{a+b=c\\ a,b\in\mathbb{Z}_{2m}^*}} |A'_{2m}[a]+B'_{2m}[b]| \\ &\geq (1-\varepsilon_0)\sum_{c\in\mathbb{Z}_{2m}} \max_{\substack{a+b=c\\ a,b\in\mathbb{Z}_{2m}^*\\ \alpha'_{2m}[a],\beta'_{2m}[b]>\varepsilon_0}} \{\alpha'_{2m}[a]+\beta'_{2m}[b]\}\frac{n}{2m}.\end{aligned}$$

Notice that $\mathbb{Z}_{2m}^* \cong \mathbb{Z}_m^*$ and when $r\in\mathbb{Z}_m^*$ and $r'\in\mathbb{Z}_{2m}^*$ is such that $r'\equiv r \pmod{m}$, then $\alpha'_{2m}[r'] = \alpha'_m[r]$ and $\beta'_{2m}[r'] = \beta'_m[r]$. Hence

$$|(A+B)\cap[1,n]| \geq (1-\varepsilon_0)\sum_{c\in\mathbb{Z}_m} \max_{\substack{a+b=c\\ a,b\in\mathbb{Z}_m^*\\ \alpha'_m[a]-\varepsilon_0>0,\, b\in D}} \{\alpha'_m[a]-\varepsilon_0\}\frac{n}{2m}.$$

Writing $A'_x = \{a\in\mathbb{Z}_m^* \colon \alpha'_m[a] > x+\varepsilon_0\}$, the right hand side equals

$$(1-\varepsilon_0)\frac{n}{2m}\int_0^1 |A'_x + D|\,dx \geq (1-\varepsilon_0)\frac{n}{2m}\int_0^1 |A'_x|\,dx \cdot \frac{m}{\varphi(m)}\cdot\frac{\varphi(q_2\cdots q_l)}{q_2\cdots q_l}$$

by Theorem 3.3 since $|D| \geq (1-3\varepsilon)\beta\varphi(m) \geq \varphi(m)\prod_{i=2}^{l}\frac{2}{\varphi(q_i)}$. By definition of A'_x, the integral here is

$$\sum_{a\in\mathbb{Z}_m^*,\, \alpha'_m[a]>\varepsilon_0} (\alpha'_m[a]-\varepsilon_0) \geq (1-\varepsilon)^3\alpha\varphi(m) - \varepsilon_0\varphi(m) \geq (1-4\varepsilon)\alpha\varphi(m).$$

Hence

$$|(A+B)\cap[1,n]| \geq (1-4\varepsilon)\alpha\frac{\varphi(q_1\cdots q_l)}{q_1\cdots q_l}\cdot n$$

for every $\varepsilon>0$ and every large enough n and the claim follows.

8. Proof of Proposition 3.5.

8. Proof of Proposition 3.5. Recall the notation $U_N(d,\theta,\delta)$ from Section 4.2 and the notation $\mathcal{F}(\eta, C)$ from Definition 6.3. Proposition 3.5 follows from the following proposition as Proposition 3.1 follows from Proposition 6.6.

PROPOSITION 8.1. *Let $\varepsilon>0$, $K\geq 1$, $C>0$ and $N\in\mathbb{P}$. There exists $\gamma_1 = \gamma_1(K,\varepsilon)$ such that for each $\gamma_0\in(0,\gamma_1)$ there exists $\eta = \eta(\gamma_0,\varepsilon,C)$ such that the following holds.*

Let $f,g\in\mathcal{F}(\eta,C)$ be such that $\mathbb{E}f = \alpha\in(\gamma_0,\gamma_1)$ and $\mathbb{E}g = \beta\in(\gamma_0,\gamma_1)$. If

$$|\{x\in\mathbb{Z}_N \colon (f*g)(x)>0\}| \leq K(\alpha\beta)^{1/2}N,$$

then there exist $d\in\mathbb{Z}_N^*$ and $\theta_1,\theta_2\in\mathbb{R}$ such that

$$\frac{1}{N}\sum_{x\in U_N(d,\theta_1,\varepsilon)} f(x) \geq \alpha(1-\varepsilon) \quad \textit{and} \quad \frac{1}{N}\sum_{x\in U_N(d,\theta_2,\varepsilon)} g(x) \geq \beta(1-\varepsilon).$$

This proposition will quickly follow from the following lemma.

Lemma 8.2. *Let $\varepsilon > 0$, $K \ge 1$ and $N \in \mathbb{P}$. There exists $\gamma_1 = \gamma_1(K, \varepsilon)$ such that for each $\gamma_0 \in (0, \gamma_1)$ there exists $\sigma' = \sigma'(\gamma_0, \varepsilon)$ such that the following holds.*

Let $f, g\colon \mathbb{Z}_N \to [0,1]$ be such that $\mathbb{E}f = \alpha \in (\gamma_0, \gamma_1)$ and $\mathbb{E}g = \beta \in (\gamma_0, \gamma_1)$. If

$$|\{x \in \mathbb{Z}_N\colon (f * g)(x) \ge \sigma' N\}| \le K(\alpha\beta)^{1/2} N, \tag{8.1}$$

then there exists $\xi \in \mathbb{Z}_N$ such that $|\widehat{f}(\xi)| \ge (1-\varepsilon)\alpha$ and $|\widehat{g}(\xi)| \ge (1-\varepsilon)\beta$.

Proof. We can assume that ε and γ_1 are small. Let $\delta = \gamma_0^2\varepsilon^2$, $\sigma' = \delta^4\gamma_0^4/2$ and let

$$A = \{x \in \mathbb{Z}_N\colon f(x) \ge \delta\alpha\} \quad \text{and} \quad B = \{x \in \mathbb{Z}_N\colon g(x) \ge \delta\beta\}.$$

Then $|A| \ge (1-\delta)\alpha N$ and $|B| \ge (1-\delta)\beta N$. Writing

$$E = \left\{(a,b) \in A \times B\colon r_{A+B}(a+b) \ge \delta^2 \frac{|A|\,|B|}{|A+B|}\right\},$$

one has $|E| \ge (1-\delta^2)|A|\,|B|$ and

$$x \in A \overset{E}{+} B \ \Rightarrow\ (f*g)(x) \ge \delta\alpha \cdot \delta\beta \cdot \delta^2 \frac{|A|\,|B|}{|A+B|} \ge (1-\delta)^2\delta^4(\alpha\beta)^2 N \ge \sigma' N,$$

so by assumption (8.1),

$$|A \overset{E}{+} B| \le K(\alpha\beta)^{1/2}N = K\left(\frac{\alpha N}{|A|} \cdot \frac{\beta N}{|B|}\right)^{1/2} |A|^{1/2}|B|^{1/2}.$$

By Lemma 4.9 we can find sets $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \ge (1-\delta)|A|$, $|B'| \ge (1-\delta)|B|$ and

$$\begin{aligned}|A'+B'| &\le \frac{K^3}{1-6\delta}\left(\frac{\alpha N}{|A|} \cdot \frac{\beta N}{|B|}\right)^{3/2} |A|^{1/2}|B|^{1/2} \\ &\le \min\{(3K)^3|A'|^{1/2}|B'|^{1/2}, (3K)^3(\alpha\beta)^{1/2}N\}.\end{aligned}$$

By [16, Corollary 2.24],

$$\frac{|A'+B'-(A'+B')|}{|A'+B'|} \le (3K)^{O(1)}.$$

Hence, by Lemma 4.7 when γ_1 is small enough (depending on K and ε), there exist $d \in \mathbb{Z}_N^*$ and $\theta' \in \mathbb{R}$ such that

$$A' + B' \subseteq U_N(d, \theta', \varepsilon).$$

In particular there is $\theta \in \mathbb{R}$ such that

$$A' \subseteq U_N(d, \theta, \varepsilon).$$

Hence by Lemma 4.8, there is $\xi \in \mathbb{Z}_N$ such that

$$\widehat{f1_{A'}}(\xi) \geq (1 - 20\varepsilon^2)\widehat{f1_{A'}}(0) = (1 - 20\varepsilon^2)(\widehat{f}(0) - \widehat{f1_{A\setminus A'}}(0) - \widehat{f1_{\mathbb{Z}_N\setminus A}}(0)).$$

Thus

$$\begin{aligned}|\widehat{f}(\xi)| &= |\widehat{f1_{A'}}(\xi) + \widehat{f1_{A\setminus A'}}(\xi) + \widehat{f1_{\mathbb{Z}_N\setminus A}}(\xi)| \\ &\geq (1 - 20\varepsilon^2)\widehat{f}(0) - 2\widehat{f1_{A\setminus A'}}(0) - 2\widehat{f1_{\mathbb{Z}_N\setminus A}}(0) \\ &\geq \frac{1}{N}\Big((1 - 20\varepsilon^2)\sum_{x\in\mathbb{Z}_N} f(x) - 2\delta N - 2\cdot\delta\alpha\cdot N\Big) \geq \alpha(1 - \varepsilon).\end{aligned}$$

Similarly $|\widehat{g}(\xi)| \geq \beta(1 - \varepsilon)$. ∎

Proof of Proposition 8.1. We can clearly assume that ε is small. Let γ_1 and σ' be $\gamma_1(3K, \varepsilon^3/2)$ and $\sigma'(\gamma_0/2, \varepsilon^3/2, 3K)$ in Lemma 8.2. Write $f = f_1 + f_2$ and $g = g_1 + g_2$ as in Lemma 6.5 with $\sigma < 1/2$ and $\varepsilon_0 < \varepsilon^3\gamma_0/2$ so small that, for $(i, j) \neq (1, 1)$,

$$|\{x \in \mathbb{Z}_N : (f_i * g_j)(x) \geq \sigma'/10\}| \leq \frac{1}{10}(\alpha\beta)^{1/2}N$$

(possible by estimating the L_2-norm of $f_i * g_j$ exactly as in [3, proof of Lemma 13]). Hence by assumption

$$|\{x \in \mathbb{Z}_N : (f_1 * g_1)(x) \geq \sigma'\}| \leq 2K(\alpha\beta)^{1/2}N.$$

Applying Lemma 8.2 to $f_1/(1 + \sigma)$ and $f_2/(1 + \sigma)$, we see that $|\widehat{f_1}(\xi)| \geq (1 - \varepsilon^3/2)\alpha$ and $|\widehat{g_1}(\xi)| \geq (1 - \varepsilon^3/2)\beta$, which immediately implies that $|\widehat{f}(\xi)| \geq (1 - \varepsilon^3)\alpha$ and $|\widehat{g}(\xi)| \geq (1 - \varepsilon^3)\beta$, and the claim follows from Lemma 4.8. ∎

9. An additional example.

The following example demonstrates that $1 - \varepsilon$ in Theorem 2.1 cannot be replaced by $1 + \varepsilon$ in general.

EXAMPLE 9.1. We can assume that $\beta \leq \alpha$. Let m be the product of the first l primes and let $\varepsilon > 0$. Then we have the following two examples.

1. For $\beta' \in (0, 1]$, choose

$$\begin{aligned}A &= \{p \in [2, n] \cap \mathbb{P} : p \equiv 1 \pmod{m}\}, \\ B &= \{p \in [2, \beta' n] \cap \mathbb{P} : p \equiv 1 \pmod{m}\}.\end{aligned}$$

Now $\alpha = (1 + o_{n\to\infty}(1))/\varphi(m)$, $\beta = (1 + o_{n\to\infty}(1))\beta'/\varphi(m)$ and

$$\begin{aligned}|A + B| &\leq |\{k \in [4, (1 + \beta')n] : k \equiv 2 \pmod{m}\}| \leq (1 + \beta')\frac{n}{m} \\ &= (1 + o_{n\to\infty}(1))(\alpha + \beta)\frac{\varphi(m)}{m}n.\end{aligned}$$

When α and $\beta \geq \alpha^4$ are small enough (depending on ε) and n is large enough (depending on α), we deduce by Lemma 4.1 that

$$|A+B| \leq (1+\varepsilon^2)\frac{\alpha+\beta}{e^\gamma \log\log(1/\alpha)}n < (1+\varepsilon)\frac{\alpha+\beta}{e^\gamma \log\log(1/(\alpha\beta))}n.$$

2. For $\alpha' \in (0,1]$, choose

$$A = [2, \alpha' n] \cap \mathbb{P} \quad \text{and} \quad B = \{p \in [2, n/\varphi(m)] \cap \mathbb{P} \colon p \equiv 1 \ (\mathrm{mod}\ m)\}.$$

Now $\alpha = (1+o_{n\to\infty}(1))\alpha'$, $\beta = (1+o_{n\to\infty}(1))/\varphi(m)^2$, and

$$\begin{aligned}|A+B| &\leq |\{k \in [4, (\alpha' + 1/\varphi(m))n] \colon \gcd(k-1,m) = 1\}| + l \\ &= (1+o_{n\to\infty}(1))(\alpha+\beta^{1/2})\frac{\varphi(m)}{m}n.\end{aligned}$$

When α and $\beta \leq \alpha^4$ are small enough (depending on ε) and n is large enough (depending on β), we infer by Lemma 4.1 that

$$|A+B| \leq (1+\varepsilon^2)\frac{\alpha+\beta^{1/2}}{e^\gamma \log\log(1/\beta^{1/2})}n < (1+\varepsilon)\frac{\alpha+\beta}{e^\gamma \log\log(1/(\alpha\beta))}n.$$

Acknowledgments. The author would like to thank Mariah Hamel for pointing out the work of Ramaré and Ruzsa [14], and Andrew Granville for advice and for unintentionally inspiring Theorem 2.3. She is also grateful to the referee for his or her comments which led to adding Theorem 1.4 and helped to improve the exposition.

The author was supported by the Academy of Finland grant no. 137883.

References

[1] N. Alon, A. Granville, and A. Ubis, *The number of sumsets in a finite field*, Bull. London Math. Soc. 42 (2010), 784–794.

[2] B. Bollobás and I. Leader, *Sums in the grid*, Discrete Math. 162 (1996), 31–48.

[3] K. Chipeniuk and M. Hamel, *On sums of sets of primes with positive relative density*, J. London Math. Soc. (2) 83 (2011), 673–690.

[4] B. Green, *Roth's theorem in the primes*, Ann. of Math. (2) 161 (2005), 1609–1636.

[5] B. Green and I. Z. Ruzsa, *Sets with small sumset and rectification*, Bull. London Math. Soc. 38 (2006), 43–52.

[6] B. Green and T. Tao, *Restriction theory of the Selberg sieve, with applications*, J. Théor. Nombres Bordeaux 18 (2006), 147–182.

[7] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) 167 (2008), 481–547.

[8] B. Green and T. Tao, *Freiman's theorem in finite fields via extremal set theory*, Combin. Probab. Comput. 18 (2009), 335–355.

[9] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) 171 (2010), 1753–1850.

[10] M. Hamel and I. Łaba, *Arithmetic structures in random sets*, Integers 8 (2008), A04, 21 pp.

[11] V. F. Lev, *Distribution of points on arcs*, Integers 5 (2005), A11, 6 pp.

[12] H. Li and H. Pan, *A density version of Vinogradov's three primes theorem*, Forum Math. 22 (2010), 699–714.

[13] J. M. Pollard, *A generalisation of the theorem of Cauchy and Davenport*, J. London Math. Soc. (2) 8 (1974), 460–462.

[14] O. Ramaré and I. Z. Ruzsa, *Additive properties of dense subsets of sifted sequences*, J. Théor. Nombres Bordeaux 13 (2001), 559–581.

[15] X. Shao, *A density version of the Vinogradov three prime theorem*, arXiv:1206.6139v1[math.NT], 2012.

[16] T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge Stud. Adv. Math. 105, Cambridge Univ. Press, Cambridge, 2010.

Kaisa Matomäki
Department of Mathematics
University of Turku
20014 Turku, Finland
E-mail: ksmato@utu.fi

Received on 2.4.2012
and in revised form on 13.11.2012 (7020)