Proc. Indian Acad. Sci. (Math. Sci.) (2022) 132:62
https://doi.org/10.1007/s12044-022-00698-z

The number of rationals determined by large sets of sifted integers

OLIVIER RAMARÉ

Institut de Mathématiques de Marseille, CNRS/Aix Marseille Université,
U.M.R. 7373, Site Sud, Campus de Luminy, Case 907,
13288 Marseille Cedex 9, France
E-mail: olivier.ramare@univ-amu.fr

MS received 4 January 2022; revised 28 May 2022; accepted 30 May 2022

Abstract. We prove that the number of fractions h_1/h_2 of integers h_1, h_2 a subset $\mathcal{A} \subset \mathcal{H} \cap [1, X]$ is at least $\alpha X/(\log X)^{3/2}$, where $\mathcal{H}$ is the set $p-1$, p being a prime such that $p+1$ is a sum of two coprime squares. So, this number of fractions is $\gg_\varepsilon \alpha^{1+\varepsilon}|\mathcal{A}|^2$, where ε is any positive real number. We take this opportunity to describe a geometrical view of the sieve and its usage to study integer sequences.

Keywords. Quotient sets; Selberg sieve; Brun–Titchmarsh theorem; multiplication table problem.

2010 Mathematics Subject Classification. 11B30; 11N36.

1. Introduction

In [1], we proved that a subset $\mathcal{A}$ of $[1, X]$, of cardinality at least αX, produces more than constant times $\alpha^{2+\varepsilon}X^2$ distinct fractions a/a', with a and a' from $\mathcal{A}$. Here $\varepsilon > 0$ is a given real number, the constant may depend on ε (and we prove that this dependance is indeed heavy) and X is large enough. The question appeared as Problem 22 in [17], though in a different form, and a first result was obtained in [16]. From a generic viewpoint, we considered thinner sets in [2] (see also its addendum in [18]) and noted that multiplicativity plays a large role in the behaviour of $|\mathcal{A}/\mathcal{A}|$. Our two fields of experiments were the set of primes minus 1 and the set of sums of two squares minus 1, as the random case shows a change of regime around density $1/\sqrt{\log X}$.

We consider here a similar problem, but with dense subsets of sufficiently sifted sequences, in the sense of [15]. The angle we take is different and may be resumed by saying that such sets are close to subsets of intersections of arithmetic progressions, hence may be expected to behave like subsets of density of the integers.

We will be more precise in subsection 1.2 but in this Introduction, it is enough to say that a sufficiently sifted sequence is a sequence $\mathcal{H}$ which can be upper-sifted by a sieve of dimension $\kappa \geq 0$ and such that there exists three positive constants c_1, c_2 and $X_0(\mathcal{H})$ such that, when $X \geq X_0(\mathcal{H})$, we have

Published online: 26 October 2022

$$sc_1 \frac{X}{(\log X)^\kappa} \leq \#\{h \leq X, h \in \mathcal{H}\} \leq c_2 \frac{X}{(\log X)^\kappa}. \tag{1}$$

The upper bound comes from an upper sieve, say from Selberg's sieve, while the lower bound has to be obtained in some manner. Here are some examples of sufficiently sifted sequences:

(i) The set of primes of dimension $\kappa = 1$,
(ii) The set of sums of two coprime squares, of dimension $\kappa = 1/2$,
(iii) The set of sums n of two coprime squares such that $n + 1$ is also such a sum, see [7],
(iv) The set of primes p such that $p + 1$ is a sum of two coprime squares, see [10], [8] and [19],
(v) The set of sums of two coprime squares that are also of the form $x^2 + xy + y^2$, with x and y coprime, see [5],
(vi) The set of natural integers, of dimension $\kappa = 0$.

Given any such set $\mathcal{H}$ and some integer c, the set $\mathcal{H} - c$ is also sufficiently sifted and of the same dimension. It is expected that the intersection of any two such sequences, if infinite, is also sufficiently sifted, but we are still light-years away from proving that.

We consider in what follows the sufficiently sifted sequence $\mathcal{H}$ to be fixed, the bound X to be larger than $X_0(\mathcal{H})$ and subsets, say $\mathcal{A}$ of $\mathcal{H} \cap [1, X]$ such that

$$|\mathcal{A}| \gg \alpha \frac{c_1 X}{(\log X)^\kappa} \tag{2}$$

for some positive α. We are concerned with the dependence in α in subsequent estimates, so we shorten the above in $|\mathcal{A}| \gg_{\mathcal{H}} \alpha X/(\log X)^\kappa$ and say that $\mathcal{A}$ is relatively dense with respect to $\mathcal{H}$. Here is our main result.

Theorem 1. *Let $\varepsilon > 0$. Let $\mathcal{H}$ be a given sufficiently sifted set of strict dimension κ as described above. Let α be a real number in $(0, 1]$ and X be a real number $\geq X_0(\mathcal{H})$. When $\mathcal{A}$ is a subset of $\mathcal{H} \cap [1, X]$ with $|\mathcal{A}| \gg_{\mathcal{H}} \alpha X/(\log X)^\kappa$, we have $|\mathcal{A}/\mathcal{A}| \gg_{\varepsilon,\mathcal{H}} \alpha^{1+\varepsilon}|\mathcal{A}|^2$.*

An optimal result would have α^ε rather than $\alpha^{1+\varepsilon}$. To explain what we mean by *strict* dimension, see (4) and the paragraph that follows, the concept of sieve dimension being rather well-known. When $\mathcal{H} = \mathbb{N}$, the paper [1, Theorem 1.1] proves that the lower bound $\alpha^{2+\varepsilon} X^2$ is available, but we have not been able to adapt the proof to this case. Getting the lower bound $\alpha^{4+\varepsilon} X^2/(\log X)^{2\kappa}$ is rather straightforward, the main work here is to get α^3 rather than α^4.

We have restricted our attention to the quotient $\mathcal{A}/\mathcal{A}$, as it is believed to be more regular than the product set $\mathcal{A} \cdot \mathcal{A}$ when we encounter the *multiplication table problem*. The readers interested in such questions can refer [11], [4] and [2].

A part of our work is to handle such general sequences and then to prove the above theorem. Let us mention a corollary of Theorem 1 that may be used to measure future progress on this question.

COROLLARY 2

Let $\varepsilon > 0$. There exists a positive constant $C(\varepsilon)$ with the following property: Let $\eta \in (0, 1)$ be a parameter. Let $\mathcal{P}(\eta, X)$ be the sequence of primes $p \leq X$ such that $\|p\pi\| \leq \eta$. We

have more than $C(\varepsilon)\eta^{3+\varepsilon}X^2/(\log X)^2$ *fractions of the form* $(p_1-1)/(p_2-1)$, *where* p_1 *and* p_2 *belong to* $\mathcal{P}(\eta, X)$, *provided* $X \geq X_0(\eta)$.

Let us now turn our attention to a precise definition of sufficiently sifted sequences.

1.1 *Geometric sieve context*

We start by selecting, for every prime p, a large subset $\mathcal{K}_p$ of $\mathbb{Z}/p\mathbb{Z}$. Let us review this choice in our examples:

(i) For the set of primes, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$.
(ii) For the set of sums of two coprime squares, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$ when $p \equiv 3[4]$ and $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z}$ otherwise.
(iii) For the set of sums n of two coprime squares such that $n+1$ is also such a sum, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0, -1\}$ when $p \equiv 3[4]$ and $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z}$ otherwise.
(iv) For the set of primes p such that $p+1$ is a sum of to coprime squares, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0, -1\}$ when $p \equiv 3[4]$ and $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$ otherwise.
(v) For the set of sums of two coprime squares that are also of the form x^2+xy+y^2, with x and y coprime, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$ when $p \in \{5, 7, 11\} \bmod 12$ and $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z}$, see [5, Proposition 6.2].
(vi) For the set of natural integers, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z}$.
(vii) And finally, for the set of primes minus 1, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{1\}$.

For every square-free integer d, we also consider the subset $\mathcal{K}_d \subset \mathbb{Z}/d\mathbb{Z}$ that corresponds by the Chinese Remainder theorem to $\prod_{p|d} \mathcal{K}_p$. We define $\mathcal{K} = (\mathcal{K}_d)$. In the vocabulary of [13, Chapter 2] or of [15], $\mathcal{K}$ is a *square-free multiplicatively split compact set*. This comes from the fact that $\mathcal{K}$ may be viewed as a subset of $\hat{\mathbb{Z}} = \prod_p \hat{\mathbb{Z}}_p$, the product of the p-adic integers, which is also the inverse limit of $(\mathbb{Z}/q\mathbb{Z})$, with the obvious choice of morphisms. We will not use this viewpoint here. A particular property of $\mathcal{K}$ is to be underlined

$$\forall d|q, \forall a \in \mathcal{K}_d, \quad \sum_{\substack{b \in \mathcal{K}_q \\ b \equiv a[d]}} 1 = \frac{|\mathcal{K}_q|}{|\mathcal{K}_d|}. \tag{3}$$

Let us pick a subcompact set $\mathcal{K}' \subset \mathcal{K}$. This one does not need to be multiplicatively split. This is equivalent to the choice of a coherent sequence of subsets $\mathcal{K}'_d \subset \mathbb{Z}/d\mathbb{Z}$, where coherent means that the canonical surjection from $\mathbb{Z}/q\mathbb{Z}$ to $\mathbb{Z}/d\mathbb{Z}$ indeed sends $\mathcal{K}'_q$ onto $\mathcal{K}'_d$, whenever d and q are chosen so that $d|q$.

It is time to turn to quantification. We assumed that our sieve is of *strict* dimension κ and this corresponds to two facts: first $\mathcal{K}$ is non-empty, i.e. none of the subsets $\mathcal{K}_p$ is empty, and second, that we have

$$\sum_{p \leq Q} \frac{(p - \kappa - |\mathcal{K}_p|)^2}{p} \log p = \mathcal{O}(1) \tag{4}$$

when Q goes to infinity. The readers will swiftly check that the sieve dimensions we have announced earlier for our examples are correct. The dimension is usually defined by

$$\sum_{p \leq Q} \frac{p - |\mathcal{K}_p|}{p} \log p = \kappa \log Q + \mathcal{O}(1)$$

which is a clear consequence of (4). Technical assumptions are then added on $|\mathcal{K}_p|$. As it turns out, in our case, hypothesis (4) is easy to state and avoids all the technical assumptions. So we added the qualification *strict* in front of 'dimension' to handle this situation. Concerning $\mathcal{K}'$, we only assume it is small in the sense that

$$|\mathcal{K}'_r| \leq d(r)^m \tag{5}$$

for some m, and where $d(r)$ is the number of divisors of r.

1.2 *Sufficiently sifted sequences*

Now that we have the notion of compact sets, we may turn to sequences of integers, and see how both are related. Let us first consider our examples. We know that a prime in $[\sqrt{X}, X]$ has no prime factors below $\sqrt{X}$. We know that sums of two coprime squares do not have any prime factors congruent to 3 modulo 4. Two points are to be inspected with attention:

(a) We detect primes in this manner, but only the ones in given intervals. This is enough for us since we only consider fixed X. Furthermore, the primes below $\sqrt{X}$ are negligible in numbers.
(b) A point that is clear with sums of two squares: when we consider integers without any prime factors $\leq \sqrt{X}$, that is, $\equiv 3[4]$, this includes also primes in $(\sqrt{X}, X]$ that are congruent to 3 modulo 4! Again, such numbers are less numerous than sums of two squares.

These two points underline the fact that the fit between the easily-defined sequences and the sieve setting is imperfect, and that some tweaking is required. A convenient tool to connect these two universes is the next notion.

DEFINITION 3

A sequence $(u_n)_{n\leq N}$ of complex numbers is said to be *carried by* $\mathcal{K}$ whenever

$$\forall n \leq N, \left[u_n \neq 0 \implies \forall q \leq \sqrt{X}, n \in \mathcal{K}_q\right]. \tag{6}$$

Here is how we define sufficiently sifted sequences.

DEFINITION 4

Let $\mathcal{K}$ be a square-free multiplicative compact set. An infinite sufficiently sifted sequence $\mathcal{H}$ carried by $\mathcal{K}$ is a sequence such that there exists c_1 and X_0 such that, for every $X \geq X_0$ large enough,

(a) the characteristic function of $\mathcal{H} \cap [\sqrt{X}, X]$ is carried by $\mathcal{K}$,
(b) $|\mathcal{H} \cap [\sqrt{X}, X]| \geq c_1 X/(\log X)^\kappa$.

We deduce from our hypotheses that we also have $|\mathcal{H} \cap [\sqrt{X}, X]| \geq c_2 X/(\log X)^\kappa$ as stated earlier. Note that this defines the dimension κ uniquely, whether one starts from the sequence $\mathcal{H}$ or from the compact set $\mathcal{K}$. In a more general setting, we would not specify $q \leq \sqrt{X}$ either in Definition 3 or in Definition 4, but keep $q \leq Q$ for some parameter Q; such a precision is not required here and our wish of simplicity asks for the simple choice $Q = \sqrt{X}$.

1.3 *Some distribution results on sufficiently sifted sequences*

We finally reach the point where we can prove results on sequences through sieve tools. Here is our general theorem.

Theorem 5. *Let $\varepsilon > 0$ and $m \geq 0$. Let $\mathcal{K}$ be a multiplicatively split compact set, with a small subset $\mathcal{K}'$ that verifies (5). Let $(u_n)_{n\leq N}$ be a a sequence carried by $\mathcal{K}$. We have, when $R < \sqrt{N}$,*

$$\sum_{r\leq R} |\mathcal{K}_r| \left| \sum_{n\in\mathcal{K}'_r} u_n \right|^2 \ll_{\varepsilon,\mathcal{K}} R^{\varepsilon} \sum_n |u_n|^2 \frac{N}{(\log N)^{\kappa}}.$$

The corollary we use for the proof of Theorem 1 is the following.

COROLLARY 6

Let $\varepsilon > 0$. Let $\mathcal{K}$ be a multiplicatively split compact set. Let $(u_n)_{n\leq N}$ be a sequence carried by $\mathcal{K}$. We have, when $R < \sqrt{N}$,

$$\sum_{r\leq R} |\mathcal{K}_r| \left| \sum_{n\equiv 0[r]} u_n \right|^2 \ll_{\varepsilon,\mathcal{K}} R^{\varepsilon} \sum_n |u_n|^2 \frac{N}{(\log N)^{\kappa}}.$$

Elliott [3] has proven a similar inequality, regarding it as dual to the Turan–Kubilius inequality, but restricting the moduli to the prime powers. He was, however, able to extend in this case, the summation to all moduli $\leq N$ and to dispense with the term R^{ε}. See for instance, [3, Theorem 3.1], where the letter q is used throughout the book to denote a prime power (related to the prime q_0).

2. Aritheoremetical auxiliaries on multiplicative compact sets

Given a square-free multiplicative compact set $\mathcal{K}$, we would like to be able to handle the quantity $|\mathcal{K}_d|$. Hypothesis (4) tells us that $|\mathcal{K}_p|$ is on average equal to $p - \kappa$.

Lemma 7. *There exists a positive constant c_3 such that, for every integer r, we have*

$$c_3 \frac{r}{(\log\log 3r)^{\kappa}} \leq |\mathcal{K}_r| \leq r.$$

Proof. We assume r to be square-free. By (4), we have

$$|\mathcal{K}_p| - p + \kappa \ll \sqrt{p/\log p}.$$

In particular, when $p \geq \kappa + A$ for some A, we have $|\mathcal{K}_p| - p + \kappa \leq |p - \kappa|/2$. We then consider

$$\log \prod_{\substack{p|d,\\ p>\kappa+A}} \frac{|\mathcal{K}_p|}{p-\kappa} = \sum_{\substack{p|r,\\ p>\kappa+A}} \log\Big(1+\frac{|\mathcal{K}_p|-p+\kappa}{p-\kappa}\Big)$$
$$= \sum_{\substack{p|r,\\ p>\kappa+A}} \frac{|\mathcal{K}_p|-p+\kappa}{p-\kappa} + \mathcal{O}\Big(\sum_{\substack{p|r,\\ p>\kappa+A}} \Big(\frac{|\mathcal{K}_p|-p+\kappa}{p-\kappa}\Big)^2\Big)$$
$$= \mathcal{O}\Big(\Big(\sum_{\substack{p|r,\\ p>\kappa+A}} \frac{(|\mathcal{K}_p|-p+\kappa)^2}{p}\Big)^{1/2}\Big) + \mathcal{O}(1) = \mathcal{O}(1).$$

The lemma then follows in a classical manner. Rapidly, we check that, with $D = \log r$, for $r \geq 2(\kappa + A)$,

$$-\log \prod_{\substack{p|d,\\ p>\kappa+A}} (1-\kappa/p) = -\sum_{\substack{p|r,\\ p>\kappa+A}} \log\Big(1-\frac{\kappa}{p}\Big)$$
$$= \sum_{\substack{p|r,\\ \kappa+A<p\leq D}} \frac{\kappa}{p} + \sum_{\substack{p|r,\\ p\geq D}} \frac{\kappa}{p} + \mathcal{O}(1)$$
$$\geq \kappa \log\log D + \mathcal{O}\Big(\frac{\log r}{D\log D}\Big) + \mathcal{O}(1),$$

hence the result. □

We end this section by the classical "sub-multiplicativity" property of the divisor functions, and a proof is given, for instance, in [14, Lemma 12].

Lemma 8. *We have* $d_k(rs) \leq d_k(r)d_k(s)$.

3. On sequences supported by compact sets

We consider the vector space $\mathscr{F}_q$ of functions from $\mathbb{Z}/q\mathbb{Z}$ to $\mathbb{Z}$ that vanish out of $\mathcal{K}_q$, which we endow with the hermitian product

$$[f|g]_{\mathcal{K}_q} = \frac{1}{|\mathcal{K}_q|} \sum_{a\in\mathcal{K}_q} f(a)\overline{g(a)}. \tag{7}$$

A definition is required here to clarify our subsequent steps.

DEFINITION 9

A sequence $(\mathscr{K}_q)_{q\leq Q}$ is said to be an orthonormal system on $\mathcal{K}$ if

(a) For all $q \in \mathcal{Q}$, $\mathscr{K}_q \subset \mathscr{F}_q$.
(b) Let ℓ and q be both in $\mathcal{Q}$ with $\ell|q$ and let χ be an element of $\mathscr{K}_\ell$. Then $\tilde{\chi}$ defined by $\tilde{\chi}(x) = \chi(x + \ell\mathbb{Z})$ if $x \in \mathcal{K}_q$ and $\tilde{\chi}(x) = 0$ otherwise, is in $\mathscr{K}_q$.

(c) For all $(\chi_1, \chi_2) \in \mathscr{K}_q^2$, we have

$$[\chi_1|\chi_2]_{\mathcal{K}_q} = \begin{cases} 0 & \text{if } \chi_1 \neq \chi_2, \\ 1 & \text{if } \chi_1 = \chi_2 \end{cases} \tag{8}$$

(d) $|\mathscr{K}_q| = |\mathcal{K}_q|$.

(e) If χ comes (according to (c)) from $\mathscr{K}_{\ell_1}$ and from $\mathscr{K}_{\ell_2}$, then χ comes from $\mathscr{K}_{(\ell_1,\ell_2)}$, where (ℓ_1, ℓ_2) is the gcd of ℓ_1 and ℓ_2.

By [12, Theorem 2], there exists an orthonormal system for $\mathcal{K}$, since this set is supposed to be multiplicatively split.

We shall call *characters* the elements of $\mathscr{K}_q$, even though they are usually not linked with any group structure. The notion of *induced* character is natural from (3), while the one of *conductor* is simply established from (e). Let $\mathscr{K}_q^*$ be the set of characters of conductor q.

We consider the non-negative multiplicative function h defined by (see [13, (2.5)])

$$h(d) = \mu^2(d) \prod_{p \| d} \Big(\frac{p}{|\mathcal{K}_p|} - 1 \Big). \tag{9}$$

This definition is valid because we have assumed our compact set $\mathcal{K}$ to be square-free. We further define (see [13, (2.7)])

$$G_d(Q) = \sum_{\substack{\delta \leq Q, \\ [d,\delta] \leq Q}} h(\delta), \tag{10}$$

where $[d, \delta] = \operatorname{lcm}(d, \delta)$ is the least common multiple of d and δ. The readers who are used to the Selberg sieve may be surprised by this definition (see Section 4). This sequence of summatory functions is controlled in a two step process: First, and since $\delta \leq [d, \delta] \leq d\delta$, we obviously have

$$G_d(Q/d) \leq G_d(Q) \leq G_1(Q). \tag{11}$$

And secondly, as a consequence of (4), we have

$$G_1(Q) = C(\mathcal{K})(\log Q)^\kappa (1 + o(1)) \tag{12}$$

for some positive constant $C(\mathcal{K})$. This is a consequence of any decent theorem on averages of multiplicative functions, as can be found in [21] or in [9].

We define

$$S(\alpha) = \sum_{n \leq N} u_n e(n\alpha), \ (\alpha \in \mathbb{R}/\mathbb{Z}) \tag{13}$$

and

$$S(\chi) = \sum_{n \leq N} u_n \chi(n), \ (\chi \in \mathscr{K}_q, q \in \mathcal{Q}), \tag{14}$$

distinction between (13) and (14) being clear from the context. Let us note that (6) ensures the fundamental equality $S(\chi) = S(\chi')$ whenever χ and χ' are induced by a same character.

We have, by [12, (14)], case $K = 1$,

$$\sum_{a \mod {}^*q} |S(a/q)|^2 = \sum_{f|q} \Big(\sum_{d|q/f} \mu\Big(\frac{q}{df}\Big) \frac{df}{|\mathcal{K}_{df}|} \Big) \sum_{\chi \in \mathscr{K}_f^*} |S(\chi)|^2. \tag{15}$$

This is a local version of [13, Theorem 2.1], which we recall in this context.

Lemma 10. *Let (u_n) being a sequence of complex numbers carried by $\mathcal{K}$. We have*

$$\sum_{d\le Q} G_d(Q) \sum_{\chi\in\mathscr{K}_f^*} |S(\chi)|^2 = \sum_{q\le Q}\sum_{a \bmod {}^* q} |S(a/q)|^2.$$

This leads us to a theorem analogous to [13, Theorem 5.2].

Theorem 11. *Let (u_n) being a sequence of complex numbers carried by $\mathcal{K}$. For any $R \le \sqrt{N}$, we have*

$$\sum_{r\le R}\sum_{\chi\in\mathscr{K}_r^*} |S(\chi)|^2 \le \frac{2N}{G_1(\sqrt{N}/R)}\sum_n |u_n|^2.$$

Proof. We simply find that

$$\begin{aligned}\sum_{r\le R}\sum_{\chi\in\mathscr{K}_r^*} |S(\chi)|^2 &\le \min_{r\le R}\frac{1}{G_r(\sqrt{N})}\sum_{d\le\sqrt{N}} G_d(\sqrt{N})\sum_{\chi\in\mathscr{K}_f^*}|S(\chi)|^2\\ &\le \frac{1}{G_1(\sqrt{N}/R)}\sum_{d\le Q} G_d(\sqrt{N})\sum_{\chi\in\mathscr{K}_f^*}|S(\chi)|^2\\ &= \frac{1}{G_1(\sqrt{N}/R)}\sum_{q\le Q}\sum_{a \bmod {}^* q}|S(a/q)|^2\end{aligned}$$

and the classical large sieve inequality concludes. □

We also derive from Lemma 10 a kind of Brun–Titchmarsh theorem.

Lemma 12. *Let (u_n) being a sequence of complex numbers carried by $\mathcal{K}$ with $u_n \in [0,1]$. We have*

$$\Big|\sum_{n\le N} \mathbf{1}_{n\in\mathcal{K}_r'} u_n\Big| \le |\mathcal{K}_r'|\frac{2N}{|\mathcal{K}_r| G_1(\sqrt{N}/r)}. \tag{16}$$

Proof. By orthogonality, we write

$$\sum_{n\le N}\mathbf{1}_{n\in\mathcal{K}_r'}u_n = \sum_{\chi\in\mathscr{K}_r}[\mathbf{1}_{\mathcal{K}_r'}|\chi]_{\mathcal{K}_r}S(\chi)$$

and thus

$$\begin{aligned}\Big|\sum_{n\le N}\mathbf{1}_{n\in\mathcal{K}_r'}u_n\Big|^2 &\le \sum_{\chi\in\mathscr{K}_r}\big|[\mathbf{1}_{\mathcal{K}_r'}|\chi]_{\mathcal{K}_r}\big|^2\sum_{\chi\in\mathscr{K}_r}|S(\chi)|^2 = \big\|\mathbf{1}_{\mathcal{K}_r'}\big\|_{\mathcal{K}_r}^2\sum_{\chi\in\mathscr{K}_r}|S(\chi)|^2\\ &\le |\mathcal{K}_r'|\sum_{\chi\in\mathscr{K}_r}|S(\chi)|^2/|\mathcal{K}_r|.\end{aligned}$$

By Lemma 10, the sum over χ is not more than

$$\sum_{q\le\sqrt{N}}\sum_{a \bmod {}^*q}|S(a/q)|^2/G_r(\sqrt{N}),$$

which the large sieve inequality bounds above by

$$\sum_n |u_n|^2\frac{2N}{|\mathcal{K}_r|G_r(\sqrt{N})}.$$

We gather our estimates and get

$$\Big|\sum_{n\le N}\mathbf{1}_{n\in\mathcal{K}_r'}u_n\Big|^2\le|\mathcal{K}_r'|\sum_n|u_n|^2\frac{2N}{|\mathcal{K}_r|G_r(\sqrt{N})}\le|\mathcal{K}_r'|\frac{2N\sum_n|u_n|^2}{|\mathcal{K}_r|G_1(\sqrt{N}/r)},$$

on using (11). Since $u_n\in[0,1]$, we have

$$\sum_n|u_n|^2\le\sum_n u_n$$

and the lemma follows readily. □

4. A divertimento: Comparing two definitions

This section is not required for the final proof. We defined the family of functions (G_d) in (10) while the usual definition is, for instance, given in [6, Chapter 3, (1.3)], the function g therein being our function h, when one sets $\omega(p)=p-|\mathcal{K}_p|$. Let us bridge the gap between these two sets of definitions. In the summation over δ such that $[d,\delta]\le Q$, we may write $\delta=md'$, where m is prime to d, and d' in fact divides d, as we may assume δ to be square-free. The condition $[d,\delta]\le Q$ reduces to $dm\le Q$, and we may sum freely on d'. We readily check that

$$\sum_{d'|d}h(d')=\frac{d}{|\mathcal{K}_d|}$$

which implies that

$$G_d(Q)=\frac{d}{|\mathcal{K}_d|}\sum_{\substack{m\le Q/d,\\(m,d)=1}}h(m). \tag{17}$$

This expression links clearly the two sets of functions: $dG_d(Q)/|\mathcal{K}_d|$ is the function that is denoted by $G_d(Q/d)$ in [6]. The inequalities (11) which are obvious in our setting were discovered and proven are in [20].

When the sieve is not square-free, our definition still holds while the one used more classically does not.

5. Proof of Theorem 5

5.1 *A general version of the Ramanujan sums*

We conclude this part with a generalization of the Ramanujan sums. We define

$$c_q(\mathcal{K}, \mathcal{K}', n) = |\mathcal{K}_q| \sum_{\chi \in \mathscr{K}_q^*} [\mathbf{1}_{\mathcal{K}_q'} | \chi]_{\mathcal{K}_q} \chi(n) \tag{18}$$

so that $c_q(\hat{\mathbb{Z}}, \{0\}, n)$ is the usual Ramanujan sum[1]. Note that (21) below may serve as a definition as well. We find that, by definition,

$$|\mathcal{K}_q| \, [\mathbf{1}_{\mathcal{K}_q'} | \chi]_{\mathcal{K}_q} = \sum_{\mathfrak{k} \in \mathcal{K}_q'} \overline{\chi}(\mathfrak{k}).$$

We reach at this level our first main inequality (on using Parseval for $\mathbf{1}_{\mathcal{K}_q'}$):

$$\Big| \sum_n u_n c_q(\mathcal{K}, \mathcal{K}', n) \Big|^2 \leq |\mathscr{K}_q^*| |\mathcal{K}_q'| \sum_{\chi \in \mathscr{K}_r^*} |S(\chi)|^2. \tag{19}$$

These generalized Ramanujan sums give us a decomposition of $\mathbf{1}_{\mathcal{K}_q'}$ that will be of the first step in the proof of Theorem 5:

$$\mathbf{1}_{\mathcal{K}_q'} = \sum_{\chi \in \mathscr{K}_q} [\mathbf{1}_{\mathcal{K}_q'} | \chi]_{\mathcal{K}_q} \chi = \frac{1}{|\mathcal{K}_q|} \sum_{f | q} c_f(\mathcal{K}, \mathcal{K}', \cdot) \tag{20}$$

which leads to

$$c_q(\mathcal{K}, \mathcal{K}', \cdot) = \sum_{f | q} \mu(q/f) |\mathcal{K}_f| \mathbf{1}_{\mathcal{K}_f'}. \tag{21}$$

Proof of Theorem 5. We use the decomposition (20) to write

$$\sum_n u_n \mathbf{1}_{n \in \mathcal{K}_r'} = \frac{1}{|\mathcal{K}_r|} \sum_{f | r} \sum_n u_n c_f(\mathcal{K}, \mathcal{K}', n).$$

Let us denote our sum by S, i.e.,

$$S = \sum_{r \leq R} d(r)^m |\mathcal{K}_r| \Big| \sum_n u_n \mathbf{1}_{n \in \mathcal{K}_r'} \Big|^2. \tag{22}$$

By employing the above decomposition, we find that

$$\begin{aligned} S &\leq \sum_{r \leq R} \frac{d(r)^m d(r)}{|\mathcal{K}_r|} \sum_{f | r} \Big| \sum_n u_n c_f(\mathcal{K}, \mathcal{K}', n) \Big|^2 \\ &\leq \sum_{f \leq R} \sum_{f | r \leq R} \frac{d(r)^{m+1}}{|\mathcal{K}_r|} \Big| \sum_n u_n c_f(\mathcal{K}, \mathcal{K}', n) \Big|^2. \end{aligned}$$

[1] *A priori*, associated to the orthonomal system $(e(\cdot a/q))_{a,q}$. We show in (21) below that this definition does not depend on the chosen orthonormal system since this is indeed the orthonormal projection of $\mathbf{1}_{\mathcal{K}_q'}$ on the space generated by $\mathscr{K}_q^*$.

We continue with

$$S \le \sum_{f\le R} |\mathcal{K}'_f||\mathcal{K}_f| \sum_{f|r\le R} \frac{d(r)^{m+1}}{|\mathcal{K}_r|}\Big|\sum_n u_n c_f(\mathcal{K},\mathcal{K}',n)\Big|^2/(|\mathcal{K}_f||\mathcal{K}'_f|)$$
$$\le \sum_{f\le R} |\mathcal{K}'_f||\mathcal{K}_f| \sum_{f|r\le R} \frac{d(r)^{m+1}}{|\mathcal{K}_r|} \sum_{\chi\in\mathscr{K}_f^*} |S(\chi)|^2$$

by using equation (19). We are in a position to use Theorem 11. We appeal to Lemma 8 and then to Lemma 7 to replace $|\mathcal{K}_r|$ by $c_3 r/(\log\log 3r)^\kappa$ and majorize $|\mathcal{K}_f|$ by f. The theorem follows swiftly.

6. Proof of Theorem 1

Let us specify that a sufficiently sifted sequence $\mathcal{H}$ is fixed throughout this part. We start with a subset $\mathcal{A} \in [1, X] \cap \mathcal{H}$ such that $|\mathcal{A}| \ge \alpha c_1 X/(\log X)^\kappa$. By maybe removing elements in $\mathcal{A}$, we further assume that

$$\alpha c_1 \le |\mathcal{A}|\frac{(\log X)^\kappa}{X} \le 2c_1\alpha. \tag{23}$$

We proceed as in the beginning of the main proof in [1] and define

$$\mathcal{M}(\mathcal{A}, r) = \#\{(a, b) \in \mathcal{A}^2, \gcd(a, b) = r\}. \tag{24}$$

We shall find one r for which $|\mathcal{M}(\mathcal{A}, r)|$ and use the inequality

$$m(\mathcal{A}) = \max_{r\ge1} |\mathcal{M}(\mathcal{A}, r)| \ge |\mathcal{A}/\mathcal{A}|. \tag{25}$$

The proof starts by the following inequality, valid when $X \ge X_0$:

$$|\mathcal{A}|^2 \le \#\{(a, b) \in \mathcal{A}^2\} \le \sum_{r\ge1} \mathcal{M}(\mathcal{A}, r). \tag{26}$$

Let us shorten the sum on the RHS.

6.1 *Using a rough upper bound*

For r larger than $R_1 = (\log X)^{2\kappa+1}$, we bound above $\mathcal{M}(\mathcal{A}, r)$ by X^2/r^2 and thus

$$\sum_{r\ge R_1} \mathcal{M}(\mathcal{A}, r) \le X^2 \sum_{r\ge R_1} r^{-2} \le 2X^2/R_1.$$

We take X large enough that this upper bound be $\le \frac{1}{4}(2\alpha c_1 X/\log^\kappa X)^2$, which is also $\le \frac{1}{4}|\mathcal{A}|^2$.

6.2 *Using a Brun–Titchmarsh like upper bound*

For r between $R_2 = C_2\alpha^{-2}$ for a large enough constant C_2, we use (16) with (u_n) the characteristic function of $\mathcal{A}$, $\mathcal{K}' = \{0\}$ and $Q = \sqrt{X}$:

$$\sum_{R_2<r\le R_1} \mathcal{M}(\mathcal{A}, r) \le \frac{X + Q^2}{G_1(Q/R_1)^2} \sum_{R_2\le r\le R_1} \frac{1}{|\mathcal{K}_r|^2} \ll \frac{X^2}{(\log X)^{2\kappa} R_2}.$$

We select C_2 so that this upper bound be $\leq \frac{1}{4}(2\alpha c_1 X/\log^\kappa X)^2$. We have thus reached

$$\tfrac{1}{2}|\mathcal{A}|^2 \leq \sum_{r\leq R_2} \mathcal{M}(\mathcal{A}, r).$$

Thus at least one $\mathcal{M}(\mathcal{A}, r)$ is larger than $\frac{1}{4}(\alpha X/\log^\kappa X)^2/R_2$ and this already proves that $|\mathcal{A}/\mathcal{A}| \geq C_2^{-1}\alpha^2|\mathcal{A}|^2$.

6.3 *Using a large sieve extension of a Brun–Titchmarsh like upper bound*

We further lower the exponent of α by shortening the sum over r some more. On using Theorem 5 with $R = R_2$ and $m = 0$, we reach, for $R \in [R_3, R_2]$,

$$\sum_{R<r\leq 2R} \mathcal{M}(\mathcal{A}, r) \leq \frac{1}{R} \max_{R<r\leq 2R} \frac{r}{|\mathcal{K}_r|} \sum_{R<r\leq 2R} |\mathcal{K}_r| \Big| \sum_{\substack{a\in\mathcal{A},\\ a\equiv 0[r]}} 1\Big|^2$$
$$\ll_{\varepsilon,\mathcal{K}} \log(1/\alpha)^\kappa \frac{\alpha^{1+\varepsilon}X^2}{R(\log X)^{2\kappa}} \ll_{\varepsilon,\mathcal{K}} \frac{\alpha^{1-\varepsilon}X^2}{R(\log X)^{2\kappa}}.$$

We sum over $R = R_2, R_2/2^1, R_2/2^2, \ldots$ until we reach below

$$R_3 = C_3\alpha^{-1-\varepsilon}, \tag{27}$$

where C_3 is large enough that the above bound is $\leq \frac{1}{4}(\alpha c_1 X/\log^\kappa X)^2$. We have reached

$$|\mathcal{A}/\mathcal{A}| \gg_{\varepsilon,\mathcal{K}} \alpha^{3+\varepsilon}(X/\log^\kappa X)^2 \tag{28}$$

and this concludes the proof of Theorem 1.

References

[1] Cilleruelo J, Ramana D S and Ramaré O. The number of rational numbers determined by large sets of integers, *Bull. London Math. Soc.* **42(3)** (2010) 517–526

[2] Cilleruelo J, Ramana D S and Ramaré O, Quotients and products of zero-density subsets of the set of positive integers, *Tr. Mat. Inst. Steklova* **296** (Analiticheskaya i Kombinatornaya Teoriya Chisel) (2017) 58–71; English version published in *Proc. Steklov Inst. Math.* **296(1)** (2017) 52–64

[3] Elliott P D T A, Duality in analytic number theory, volume 122 of Cambridge Tracts in Mathematics (1997) (Cambridge: Cambridge University Press)

[4] Ford K, Extremal properties of product sets, *Proc. Steklov Inst. Math.* **303(1)** (2018) 220–226; published in Russian in *Tr. Mat. Inst. Steklova* **303** (2018) 239–245

[5] Fouvry É, Levesque C and Waldschmidt M, Representation of integers by cyclotomic binary forms, *Acta Arith.* **184(1)** (2018) 67–86

[6] Halberstam H and Richert H-E, Sieve methods (1974) (London–New York: Academic Press (a subsidiary of Harcourt Brace Jovanovich, Publishers)) London Mathematical Society Monographs, No. 4

[7] Indlekofer K H, Scharfe Abschätchung für die Anzahlfunction der b-Zwillinge, *Acta Arith.* **26** (1974/75) 207–212

[8] Iwaniec H, Primes of the type $\varphi(x, y) + a$, where φ is a quadratic form, *Acta Arith.* **21** (1972) 203–234

[9] Levin B V and Fainleib A S, Application of some integral equations to problems of number theory, *Russian Math. Surveys* **22** (1967) 119–204
[10] Linnik Ju V, An asymptotic formula in an additive problem of Hardy–Littlewood, *Izv. Akad. Nauk SSSR Ser. Mat.* **24** (1960) 629–706
[11] Mangerel A P, The Multiplication Table and its Generalizations, PhD thesis (2014) (University of Toronto)
[12] Ramaré O, An explicit result of the sum of seven cubes, *Manuscripta Math.* **124(1)** (2007) 59–75
[13] Ramaré O, Arithmetical aspects of the large sieve inequality, volume 1 of Harish-Chandra Research Institute Lecture Notes (2009) (New Delhi: Hindustan Book Agency), with the collaboration of D S Ramana
[14] Ramaré O, On Bombieri's asymptotic sieve, *J. Number Theory* **130(5)** (2010) 1155–1189
[15] Ramaré O and Ruzsa I M, Additive properties of dense subsets of sifted sequences. *J. Théorie N. Bordeaux* **13** (2001) 559–581
[16] Sándor C, On the minimal gaps between products of members of a sequence of positive density, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **48** (2005) 3–7
[17] Sárközy A, Unsolved problems in number theory, *Period. Math. Hungar.* **42(1-2)** (2001) 17–35
[18] Shteĭnikov Yu N, Addendum to: Cilleruelo J, Ramana D S and Ramaré O, Quotients and products of zero-density subsets of the set of positive integers [MR3640773] *Tr. Mat. Inst. Steklova* **296** (Analiticheskaya i Kombinatornaya Teoriya Chisel) (2017) 260–264; English version published in *Proc. Steklov Inst. Math.* **296(1)** (2017) 251–255
[19] Teräväinen J. The Goldbach problem for primes that are sums of two squares plus one, *Mathematika* **64(1)** (2018) 20–70, 2018
[20] van Lint J E and Richert H E, On primes in arithmetic progressions, *Acta Arith.* **11** (1965) 209–216
[21] Wirsing E, Das asymptotische Verhalten von Summen über multiplikative Funktionen, *Math. Ann.* **143** (1961) 75–102

COMMUNICATING EDITOR: Sanoli Gun