Submitted exclusively to the London Mathematical Society
doi:10.1112/0000/000000

A note on additive properties of dense subsets of sifted sequences

O. Ramaré

Abstract

In this paper we show that if A is a subset of the primes with positive lower relative density δ then $A+A$ must have positive lower density at least $C_1\delta/\log\log(1/\delta)$ in the natural numbers. Our argument uses the techniques developed by the author and I.Ruzsa in their work on additive properties of dense subsequences of sufficiently sifted sequences. The result is optimal and improves on recent work of K.Chipeniuk & M.Hamel. We continue by proving several similar results, by successively replacing the sequence of primes by the sequence of sums of two squares, by the sequence of those integers n that are such that n and $n+1$ are both a sum of two squares and finally by the sequence of primes p that are such that $p+1$ is a sum of two squares. The second part of this paper contains a heuristical argument that leads to several conjectures concerning the existence of k-term arithmetic progressions within these sequences. We conclude with some conjectures belonging to the Ramsey part of additive number theory.

1. *Introduction and some results*

In recent years there has been much progress made towards understanding additive properties of the primes. One of the first important structural result on the primes is due to Šnirel'man [**21**], who showed that the primes is an asymptotic additive basis. In 1937 Vinogradov [**23**] proved his celebrated Theorem stating that the order of this basis is not more than 4. More recently, Sárkőzy [**20**] proved that every dense subsequence of the primes is also an asymptotic additive basis. In 2001 and expanding on a previous work of the author [**16**], the author & Ruzsa [**18**] obtained a wide generalization of this result as a well as optimal bounds concerning the order of such sequences as asymptotic additive basis.

The strategy we have developed is to embed the primes in a 'weighted sifted' sequence where they have positive relative density and to adapt Šnirel'man's approach in this setting. The heuristic which we explain more fully in the next section is that 'weighted sifted' sequences behave like arithmetic progressions and should thus share their properties. This is the starting point of [**7**] and [**6**]. Since the paper [**18**] has been recently overlooked, we complement it by this note; no result is proved here that is not already contained in this previous paper, but this different presentation as well as the heuristical approach proposed section 2 may be helpful.

We prove here that:

Theorem 1.1. *Let A be a subset of the primes with positive relative lower density $1/k$. Then there exists an absolute constant C_1 such that $A+A$ has positive lower density at least $C_1/[k\log\log(3k)]$ in the natural numbers.*

We can replace in the above statement lower densities by upper densities. Indeed, the proof goes by selecting a bound X up to which $A\cap[1,X]$, has enough elements and then showing that the cardinality of $(A+A)\cap[1,X]$ is large enough. We can either assume that the initial statement is valid for every X, or only for X belonging to some sequence. This

---

2000 *Mathematics Subject Classification* 11B05, 11P32, 11N36 (primary), 11B13 (secondary).

improves on [**20**] who has the lower bound C_1/k^4 and on [**2**] who has the lower bound $C_1/[k\exp(-C_2(\log k)^{2/3}(\log\log k)^{1/3})]$. It also improves on this latter work in that we only need a *lower* density, and not a density, and the conclusion gives a lower density and not only an upper one. Moreover our bound is optimal up to the constant C_1, as shown by Sárkőzy [**20**].

Our method is extremely flexible and we prove also:

THEOREM 1.2. *Let A be a subset of the sums of two squares with positive relative lower density $1/k$. Then there exists an absolute constant C_1 such that $A+A$ has positive lower density at least $C_1/[k\sqrt{\log\log(3k)}]$ in the natural numbers.*

The example developed in [**18**] shows that this lower bound is optimal, again up to the constant C_1. The next two theorems deal with more difficult sequences.

THEOREM 1.3. *Let $\mathcal{C}$ be the sequence of those integers n that such that n and $n+1$ are simultaneously a sum of two coprime squares. Let A be a subset of $\mathcal{C}$ with positive relative lower density $1/k$. Then there exists an absolute constant C_1 such that $A+A$ has positive lower density at least $C_1/[k\log\log(3k)]$ in the natural numbers .*

THEOREM 1.4. *Let $\mathcal{D}$ be the sequence of those primes p such that $p+1$ is a sum of two coprime squares. Let A be a subset of $\mathcal{D}$ with positive relative lower density $1/k$. Then there exists an absolute constant C_1 such that $A+A$ has positive lower density at least $C_1/[k(\log\log(3k))^{3/2}]$ in the natural numbers.*

The sequences $\mathcal{C}$ and $\mathcal{D}$ above are not known to have an asymptotic density, though they are believed to possess one. We know however that (see [**13**] for the lower bound)

$$\left|\left\{c\le X, c\in\mathcal{C}\right\}\right| \asymp X/\log X$$

and that (see [**14**] for the lower bound)

$$\left|\left\{d\le X, d\in\mathcal{D}\right\}\right| \asymp X/(\log X)^{3/2}.$$

Both lower bounds are consequences of the sieve reversal process of Iwaniec, while both upper bounds are consequences of the Selberg sieve, see [**8**] for instance.

2. *A heuristical argument and conjectures*

We first exhibit a majorant for the characteristic function of the primes betwen X and $2X$. Let us recall that the Ramanujan sum can be defined by

$$c_r(n) = \sum_{\substack{d|r,\\ d|n}} d\mu(r/d). \tag{2.1}$$

Our construction starts by choosing a parameter R strictly less than X. Then the function (used also in [**10**], [**4**], [**5**], [**3**] and [**22**])

$$\Lambda_R(n) = \sum_{r\le R} \frac{\mu(r)c_r(n)}{\phi(r)} \tag{2.2}$$

takes a constant value on every integer n that is coprime with every r that is less than R. This value is, say

$$Y = \sum_{r \le R} \frac{\mu^2(r)}{\phi(r)}. \tag{2.3}$$

We can show that this value is asymptotic to $\log R$ when R grows, but we will not use this here. The function $\Lambda_R(n)$ has a main drawback: we do not control its values on integers that are not coprime with some r satifying $r \le R$. So we consider instead

$$\beta_R(n) = (\Lambda_R(n)/Y)^2 = \left| \sum_{r \le R} \frac{\mu(r) c_r(n)}{\phi(r)} / Y \right|^2. \tag{2.4}$$

This is the majorising function we propose. It is indeed the majorising function that the Selberg sieve proposes, by following another path which shows that it is optimal in some sense. The path followed here is fully documented in [**17**, chapter 11].

The following Lemma specifies how far off we are from the prime numbers.

Lemma 2.1. *When $R^2 = o(X)$, we have*

$$\sum_{X < n \le 2X} \beta_R(n) \sim \frac{\log X}{\log R} \cdot \sum_{X < p \le 2X} 1.$$

So if we take for instance $R = X^{1/3}$, we get a *weighted sequence* which is about three times larger than the sequence of primes. On taking $R = \sqrt{X}/\log X$, we get an even better sequence in terms of density.

On opening the square, employing (2.1) and shuffling terms, we reach:

$$\beta_R(n) = \sum_{\substack{r_1, r_2 \le R, \\ d_1 | r_1, d_2 | r_2}} \frac{d_1 \mu(r_1/d_1) d_2 \mu(r_2/d_2)}{\phi(r_1)\phi(r_2)} \mathbb{1}_{n \equiv 0[\mathrm{lcm}(d_1,d_2)]} / Y^2 \tag{2.5}$$

the summation being on the four integer variables d_1, d_2, r_1 and r_2 verifying the stated conditions. Here is the main point exhibited by this expression:

$\beta_R(n)$ is defined in terms of congruences.
The moduli of these congruences are all no more than R^2.

Note that R^2 is a parameter at our disposal. When $R^2 = o(X)$ you can expect to be able to handle such congruences; if need be, one can even choose R^2 to be a some small power of X, as in [**7**].

Once these properties have been established (density and expression in terms of arithmetic progressions), it is straighforward to see that the weighted sequence should share many properties of arithmetic progressions, since, in essence, it is *a linear combination of a small number of arithmetic progressions*. This analogy has some limitations. There are some local obstructions, and for instance we loose a factor $\log\log k$ to some power in our estimates. I do not know in this context how to state a conjecture that would be as precise as Mann's Theorem (see [**9**]).

The similarity of treatment with the case of primes however leads to the following conjecture:

Conjecture 1. *There are arbitrarily long arithmetic progressions in the sequence of sums of two squares, and in each of the sequences $\mathcal{C}$ and $\mathcal{D}$.*

It is plausible that [**7**] and [**6**] may be generalized to these cases, and this would settle the above conjecture. The general line would be to establish a Szemeredi kind of theorem for the weighted majorant sequence ; the property for the sequence $\mathcal{C}$ (resp. $\mathcal{D}$) would follow by considering it as a subsequence of density of the enveloping (majorant) one. This is not more than a sketch, but it seems to me to be a very plausible one. Let us mention [**11**] where the author studies in details the 4-terms arithmetic progressions among sums of two squares.

On investigating monochromatic sums as done in [**12**] we propose the following conjectures:

CONJECTURE 2. *If one partitions the sequence of sums of two squares into k subsequences, where k is a positive natural number, there exists a constant C_1 such that every large enough integer is a sum of at most $C_1 k\sqrt{\log\log k}$ sums of two squares, each belonging to the same part.*

CONJECTURE 3. *If one partitions the sequence $\mathcal{C}$ into k subsequences, where k is a positive natural number, there exists a constant C_1 such that every large enough integer is a sum of at most $C_1 k\log\log k$ elements of $\mathcal{C}$, each belonging to the same part.*

CONJECTURE 4. *If one partitions the sequence $\mathcal{D}$ into k subsequences, where k is a positive natural number, there exists a constant C_1 such that every large enough integer is a sum of at most $C_1 k(\log\log k)^{3/2}$ elements of $\mathcal{D}$, each belonging to the same part.*

In the case of primes, a very precise result on this question is proved in [**15**].

In the four cases considered, the host sequence has density some (negative) power of $\log X$ when counting them up to X. One way to state that is to say that the *dimension* κ is finite. Since the Selberg sieve provides an upper bound for the characteristic function of thinner sets, one may wonder how thin one can take this host sequence. We do not have any example, save an extreme one: the thinner sequence for which the Selberg sieve works is surely the sequence of squares, and we know that there are no 4-terms arithmetic progressions made only of squares (one may read [**1**] on connected subjects). However, many properties still hold true as shown for instance in [**12**] (see also [**19**]).

3. *Proof of theorems 1.1, 1.2, 1.3 and 1.4*

Each of the sequences we consider is *sufficiently sifted* in the terminology of [**18**], and this is the fact that we need to establish. Let us start by recalling the definition of a sufficiently sifted sequence:

DEFINITION 3.1. *A sequence $\mathcal{A}$ of integers is said to be* sufficiently sifted *if there exist parameters $X_0 \ge 1$, $c_1, c_2 > 0$, $\kappa \ge 0$, $s_0 \ge 2$, $\xi \in [0, \frac{1}{2}[$, $\alpha > 0$, a sequence $(\mathcal{K}_p)_p$ such that $\mathcal{K}_p \subset \mathbb{Z}/p\mathbb{Z}$ and finally a sequence $(\mathcal{A}_X)_{X\ge 1}$ of subsets of $\mathcal{A}$ such that*

(H_1) *When $X \ge X_0$, we have $A(X) \ge c_1 X/\log^\kappa X$.*

(H_2) *$A(X) - A_X(X) = o\big(X(\log X)^{-\kappa}\big)$.*

(H_3) *For every prime p not more than X^{1/s_0}, we have $\mathcal{A}_X + p\mathbb{Z} \subset \mathcal{K}_p$.*

(H_4) *$\sum_{p\le X}(1 - |\mathcal{K}_p|/p)\log p = \kappa\log X + \mathcal{O}(1)$.*

(H_5) *We have $|\mathcal{K}_p| \ge p - c_2 p^\xi$.*

Here, as usual, $A(X)$ (resp. $\mathcal{A}_X(X)$) denotes the number of elements of $\mathcal{A}$ (resp. $\mathcal{A}_X$) that are below X.

It is best to motivate this definition by looking at the case of prime numbers. We would like to use a sieve approach and say that the primes are the ones that, for every prime p, are not congruent to 0 modulo p. Alas, such a definition would rule out *every* integer, save 1. We overcome this difficulty by using the sieve property locally. Practically, we define $\mathcal{A}_X$, for every positive real number X, to be the set of primes between $\sqrt{X}$ (excluded) and X. This time, for every prime p less than $X^{1/2}$, every element of $\mathcal{A}_X$ falls modulo p on a non-zero class. The above definition shows that the primes indeed form a well-sifted sequence with the choices $s_0 = 2$, $\kappa = 1$, $\mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$, $c_2 = 1$ and finally $\xi = 0$.

Examined under this light, the above definition looks natural enough, but there are a number of difficulties that we now point out. The *density* κ is in fact the only biunivocally defined parameter above. Indeed, the sieve shows that with the conditions above, we have $A(X) \asymp X/\log^\kappa X$. This being so, hypothesis (H_4) constrains somehow the choice of $(\mathcal{K}_p)$ but not strongly: a smaller or larger set $\mathcal{K}_p$ for some prime p may also satisfy (H_3). Finally, in the case of primes, we could have taken for $\mathcal{A}_X$ the sequence of those primes between $X/\log X$ and X. This shows that any s_0 strictly less than 1 would then be possible, i.e. that even the level at which we sieve is unclear when we start from the end-product.

Once this definition has been set and properly commented, it is not difficult to show that the sequences we consider are indeed sufficiently sifted. Let us give some details to establish this fact for the sequence $\mathcal{D}$. We set

$$\begin{cases} & \mathcal{K}_2 = \mathbb{Z}/2\mathbb{Z} \setminus \{0\}, \\ \text{when } p \equiv 3[4] & \mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0, 1\}, \\ \text{when } p \equiv 1[4] & \mathcal{K}_p = \mathbb{Z}/p\mathbb{Z} \setminus \{0\}. \end{cases} \tag{3.1}$$

For $X \geq 1$, we select $\mathcal{D}_X = \mathcal{D} \cap (\sqrt{X}, X]$, so that hypotheses (H_1), (H_2), (H_3), (H_4) and (H_5) are easily seen to be satisfied with $\kappa = 3/2$, $s_0 = 2$, $\xi = 0$, $c_2 = 2$ and $A(X) - A_X(X) = \mathcal{O}(\sqrt{X})$.

Let us recall [**18**, Theorem 1].

LEMMA 3.1. *For $i \in \{1, 2\}$, let $\mathcal{A}_i$ be a sufficiently sifted sequence of dimension κ_i, let $k_i \geq 1$ be a real number and let $\mathcal{A}_i^* \subset \mathcal{A}_i$ be such that $A_i^*(X) \geq A_i(X)/k_i$ for $X \geq X_1$. Then we have*

$$(\mathcal{A}_1^* + \mathcal{A}_2^*)(X) \gg_{\mathcal{A}_1,\mathcal{A}_2} X/\big(k_1(\log\log 3(k_1 + k_2))^{\kappa_1}\big).$$

By $\gg_{\mathcal{A}_1,\mathcal{A}_2}$, we mean that the implied constant may depend on all the parameters required to define the sufficiently sifted sequences $\mathcal{A}_1$ and $\mathcal{A}_2$.

Theorem 1.4 follows by choosing $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{D}$.

Concerning the three other sequences, the proof is similar, and we simply mention that the dimension κ of the sequence of integers that sums of two coprime squares is $1/2$, while the dimension of $\mathcal{C}$ is 1.

References

1. T.C. Brown, A.R. Freedman, and Peter Jau-Shyong Shiue. Progressions of squares. *Australas. J. Combin.*, 27:187–192, 2003.

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

3. J.B. Friedlander and D.A. Goldston. Some singular series averages and the distribution of Goldbach numbers in short intervals. *Illinois J. Math.*, 39(1), 1995.

4. D.A. Goldston. On Bombieri and Davenport's theorem concerning small gaps between primes. *Mathematika*, 1, 1992.

5. D.A. Goldston. A lower bound for the second moment of primes in short intervals. *Expo. Math.*, 13(4):366–376, 1995.
6. B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, (167):481–547, 2008. Available at `http://fr.arxiv.org/pdf/math.NT/0404188`.
7. B.J. Green. Roth's theorem in the primes. *Ann. of Math.*, 3(161):1609–1636, 2005.
8. H. Halberstam and H.E. Richert. Mean value theorems for a class of arithmetic functions. *Acta Arith.*, 43:243–256, 1971.
9. H. Halberstam and K.F. Roth. Sequences. *2nd edition, Springer-Verlag*, page 292pp, 1983.
10. D.R. Heath-Brown. The ternary Goldbach problem. *Rev. Mat. Iberoamericana*, 1(1):45–59, 1985.
11. D.R. Heath-Brown. Linear relations amongst sums of two squares. In *Number theory and algebraic geometry*, volume 303 of *London Math. Soc. Lecture Note Ser.*, pages 133–176. Cambridge Univ. Press, Cambridge, 2003.
12. N. Hegyvári and F. Hennecart. On monochromatic sums of squares and primes. *J. Number Theory*, 124(2):314–324, 2007.
13. K.H. Indlekofer. Scharfe Abschätchung für die Anzahlfunction der b-Zwillinge. *Acta Arith.*, 26:207–212, 1974/75.
14. H. Iwaniec. Primes of the type $\varphi(x,y)+a$ where φ is a quadratic form. *Acta Arith.*, 21:203–234, 1972.
15. D.S. Ramana and O. Ramaré. Additive energy of dense sets of primes and monochromatic sums. *Submitted*, 2012.
16. O. Ramaré. On Snirel'man's constant. *Ann. Scu. Norm. Pisa*, 21:645–706, 1995. `http://math.univ-lille1.fr/~ramare/Maths/Article.pdf`.
17. O. Ramaré. *Arithmetical aspects of the large sieve inequality*, volume 1 of *Harish-Chandra Research Institute Lecture Notes*. Hindustan Book Agency, New Delhi, 2009. With the collaboration of D. S. Ramana.
18. O. Ramaré and I.M. Ruzsa. Additive properties of dense subsets of sifted sequences. *J. Théorie N. Bordeaux*, 13:559–581, 2001.
19. I.Z. Ruzsa. On an additive property of squares and primes. *Acta Arith.*, 49:281–289, 1988.
20. A. Sárkőzy. On finite addition theorems. *Astérisque*, (258):xi–xii, 109–127, 1999. Structure theory of set addition.
21. L.G. Šnirel'man. Über additive Eigenschaften von Zahlen. *Math. Ann.*, 107:649–690, 1933.
22. R.C. Vaughan. Moments for primes in arithmetic progressions. *Duke Math. J.*, (2):371–383, 2003.
23. I.M. Vinogradov. Representation of an odd number as a sum of three primes. *Dokl. Akad. Nauk SSSR*, 15:291–294, 1937.

Olivier Ramaré
CNRS / Laboratoire Paul Painlevé,
Université Lille 1,
59 655 Villeneuve d'Ascq, cedex,
France.

ramare@math.univ-lille1.fr