follow that

(7)
$$\sum_{n \leqslant x} f(n)^2 \sim x \quad \text{as} \quad x \to \infty.$$

On the other hand, it is not hard to show that if there is an $h = h(x)$ such that $\log h = o\big((\log x)^{1/2}\big)$ and

(8)
$$\int_2^x \left(A(u+h) - A(u) - \frac{h}{\log u} \right)^2 du = o\big(h^2 x (\log x)^{-2}\big),$$

then both (1) and (7) hold. From (1) and (7) we see that

(9)
$$\sum_{n \leqslant x} \big(f(n) - 1\big)^2 = o(x),$$

from which it follows that $f(n)$ is near 1 for almost all n.

If we take $\mathscr{A}$ to be the set of prime numbers then we have (1), since this is the prime number theorem. If the Riemann Hypothesis is assumed, then (8) holds for prime numbers with $h = \exp\big((\log x)^{1/3}\big)$, for example. (See [1].)

I am happy to thank Professor Pál Erdős for his comments, and also Professor Carl Pomerance, who pointed out an error in and a simplification of my original argument.

Reference

[1]　B. Saffari and R. C. Vaughan, *On the fractional parts of x/n and related sequences II*, Ann. Inst. Fourier (Grenoble) 27 (2) (1977), pp. 1–30.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
Ann Arbor, MI 48109

On an additive property of squares and primes

by

Imre Z. Ruzsa (Budapest)

To Uncle Paul, with epsilonial love

1. Introduction. The additive property in the title is that of being an essential component. Essential components are traditionally defined via the Schnirelmann density. The *Schnirelmann density* $\sigma(A)$ of a set A of integers is defined by the formula

$$\sigma(A) = \inf A(n)/n,$$

where n runs over the natural numbers and we use $A(n)$ to denote the *counting function* of our set A, that is, the number of its elements between 1 and n (the nonpositive elements are not taken into account).

This concept of density was introduced by and named after L. G. Schnirelmann [10], who proved the inequality

(1.1)
$$\sigma(A+B) \geqslant \sigma(A) + \sigma(B) - \sigma(A)\sigma(B)$$

and used it to show that every set of positive density is a basis, and that the set P of primes is an asymptotic basis (that is, the sumset $P + \ldots + P$ with a sufficiently large number of summands contains all large integers), which was the first unconditional result concerning the Goldbach conjecture.

A set H is called a (Schnirelmann) *essential component* if $\sigma(A+H) > \sigma(A)$ whenever $0 < \sigma(A) < 1$. By (1.1), sets of positive density always have this property. The first essential component of density 0 was discovered by Khintchine [4]; it was the set Q of squares. A few years later Erdős [1] found that every basis is an essential component; he proved this in the effective form

(1.2)
$$\sigma(A+H) \geqslant \sigma(A) + \sigma(A)\big(1 - \sigma(A)\big)/(2h),$$

if H is a basis of order h. A much stronger version of (1.2) was found by Plünnecke [8]; he proved

(1.3)
$$\sigma(A+H) \geqslant \sigma(A)^{1 - 1/h},$$

which is, in this generality, the best possible order of magnitude.

Write $\sigma(A) = 1/K$. Since the squares "almost" form a basis of order 3, and the primes "almost" form a basis of order 2, in the sense that $Q+Q+Q$, resp. $P'+P'$ ($P' = P \cup \{0, 1\}$) has a positive density, by Plünnecke's method one can obtain

$$(1.4) \qquad\qquad \sigma(A+Q) \gg K^{-2/3},$$

$$(1.5) \qquad\qquad \sigma(A+P') \gg K^{-1/2}.$$

On the other hand, let A consist of the numbers that are $\equiv 1 \pmod{m}$, with, say, $m = [K]$. Then up to K the sums $A+P'$, resp. $A+Q$, are identical with $1+P'$ and $1+Q$, and this shows that

$$(1.6) \qquad\qquad \sigma(A+Q) \ll K^{-1/2},$$

$$(1.7) \qquad\qquad \sigma(A+P') \ll 1/\log K.$$

For the squares, Plünnecke [6] proved already in 1959

$$(1.8) \qquad\qquad \sigma(A+Q) \gg K^{-3/4}$$

and later [7] improved it to

$$(1.9) \qquad\qquad \sigma(A+Q) \gg K^{-1/2}$$

by analytical methods; this means that (1.6) gives the correct order of magnitude. I do not know of any similar investigation concerning the primes. Our first result fills this gap.

THEOREM 1. *Let A be any set of integers of Schnirelmann density $\sigma(A)$ $= 1/K$, $K > 2$, and let P' denote the set of primes, 0 and 1. We have*

$$\sigma(A+P') \geqslant c/\log K$$

with some absolute constant $c > 0$.

Clearly no inequality like (1.1) can hold for the asymptotical lower density

$$\underline{d}(A) = \liminf A(n)/n;$$

just think of the set of multiples of a fixed number m, and it was much later that Kneser [5] found an analogous but more complicated inequality for the lower density of sums (see also Halberstam–Roth [2]). This seemed to imply that the Schnirelmann density is better suited for the study of set addition than the asymptotic one. On the other hand, the asymptotical density is invariant under translation and the inclusion of a finite number of elements, while the Schnirelmann density is very sensitive to the first elements. Our main objective is to investigate the impact of addition of P and Q on the asymptotic lower density of sets; it will turn out that the increment is much greater than in the Schnirelmann case.

2. Results on the asymptotic density. As before, we put

$$P = \{\text{primes}\} = \{2, 3, 5, \ldots\}, \qquad Q = \{\text{squares}\} = \{0^2, 1^2, 2^2, \ldots\}.$$

THEOREM 2. *Let A be a set of integers with lower asymptotic density*

$$\underline{d}(A) = 1/K.$$

For every $\varepsilon > 0$ we have

$$(2.1) \qquad\qquad \underline{d}(A+Q) \geqslant K^{-(\log 2 + \varepsilon)/\log\log K}$$

if $K > K_0(\varepsilon)$. On the other hand, for every ε and $K > K_0(\varepsilon)$ there is a set A with $\underline{d}(A) = 1/K$ and

$$(2.2) \qquad\qquad \underline{d}(A+Q) \leqslant K^{-(\log 2 - \varepsilon)/\log\log K}.$$

A lower estimate of the type $CK^{-1/2}$ was given by Plünnecke [7]. However, while this was optimal for the Schnirelmann density, we have just seen that the case of the asymptotic density is rather different.

THEOREM 3. *Let A be a set with $\underline{d}(A) = 1/K$, $K > 3$. We have*

$$(2.3) \qquad\qquad \underline{d}(A+P) \geqslant c/\log\log K$$

with some absolute constant $c > 0$. On the other hand, there is a constant C such that for every $K > 3$ one can find a set A with $\underline{d}(A) = 1/K$ and

$$(2.4) \qquad\qquad \underline{d}(A+P) \leqslant C/\log\log K.$$

The construction of the examples to show (2.2) and (2.4) is very simple. Choose a positive integer $m < K$ and let A be the set of multiples of m. Then obviously

$$\cdot\, \underline{d}(A) = 1/m > 1/K$$

(to make the density exactly $1/K$, one can select a subset).

In $A+Q$, a residue class $a \pmod{m}$ is represented if a is a quadratic residue modulo m. Now if $m = p_1 \ldots p_k$ with different odd primes p_i, then the number of quadratic residues is

$$\frac{p_1+1}{2} \cdots \frac{p_1+1}{2},$$

hence

$$(2.5) \qquad\qquad \underline{d}(A+Q) \leqslant 2^{-k}(1+p_1^{-1}) \ldots (1+p_k^{-1}).$$

In $A+P$, a residue class $a \pmod{m}$ is represented if there are primes $p \equiv a \pmod{m}$. This is possible if either a is a prime divisor of m or it is coprime to m, hence

$$\underline{d}(A+P) = \frac{k+\varphi(m)}{m} = k/m + (1-p_1^{-1}) \ldots (1-p_k^{-1})$$

where φ denotes Euler's function.

Now let $m = p_1 p_2 \ldots p_k$, the product of the first k odd primes where k is selected so that

$$p_1 \ldots p_k \leqslant K < p_1 \ldots p_k p_{k+1}.$$

(2.2) and (2.4) now follow from (2.5), (2.6) and the well-known estimates for p_k and the related products (see Hardy–Wright [3]).

3. Preliminaries. Our proof follows Schnirelmann's footsteps. We take all the sets $a+R$ ($R = P$ or Q), $a \in A$, we show that the intersection of any two is comparatively thin and we infer that the union must be big. This is expressed in the following lemma.

LEMMA 3.1. *Let* $X_1, \ldots, X_n$ *be finite sets,*

$$(3.2) \qquad Y = \bigcup X_i, \quad S = \sum |X_i|, \quad S_2 = \sum_{i<j} |X_i \cap X_j|.$$

We have

$$(3.3) \qquad |Y| \geqslant S^2/(S + 2S_2).$$

This is not the best possible estimate for $|Y|$ in terms of S and S_2, but for our purposes even the weaker

$$(3.4) \qquad |X| \geqslant \tfrac{1}{3} \min(S, S^2/S_2)$$

suffices.

The proof of Lemma 3.1 is standard. Let f_i be the indicator function of X_i, g that of Y and apply the Cauchy–Schwarz inequality for g and $\sum f_i$.

4. The primes. Put $T = A + P$. To prove Theorems 1 and 2 we need a lower estimate for $T(x)$ for large x.

LEMMA 4.1. *The number of positive integers* $n \leqslant x$ *for which both* $n+a$ *and* $n+b$ *are primes* ($a \neq b$) *is*

$$\leqslant c_1 \frac{x}{\log^2 x} g(a-b),$$

where c_1 *is an absolute constant and*

$$g(n) = \prod_{p \mid a-b} (1 + 1/p).$$

This is a standard application of Brun's or Selberg's sieve and it can be explicitly found in Prachar [9].

Let $A^* = A \cap [1, x/2]$ and $P^* = P \cap [1, x/2]$. Clearly $|A^* + P^*|$ is a lower estimate for $T(x)$. We apply Lemma 3.1, in the form (3.4) for the sets $X_a = a + P^*$, $a \in A^*$. Obviously

$$(4.2) \qquad S = \sum |X_a| = |A^*| \, |P^*| = A(x/2)\,\pi(x/2).$$

For different a and b we have by Lemma 4.1

$$(4.3) \qquad |X_a \cap X_b| \leqslant c_1 \frac{x}{\log^2 x} g(a-b).$$

LEMMA 4.4. *Let* $B \subset [1, y]$ *a set of integers,* $|B| \geqslant y/L$, $L \geqslant 3$. *We have*

$$(4.5) \qquad \sum_{b \in B} g(b) \leqslant c_2 |B| \log \log L.$$

Let us assume this lemma now and postpone its proof to the end of this section.

By (4.2) and (4.3) we have for any fixed a

$$(4.6) \qquad \sum_{b \neq a} |X_a \cap X_b| \leqslant c_3 \frac{x}{\log^2 x} \log \log M$$

where $M = \max(3, x/A(x/2))$.

Now we apply Lemma 3.1 in the form (3.4). This yields

$$(4.7) \qquad T(x) \geqslant c_5 \min\left(\frac{lx}{\log x}, \, x/\log\log \frac{3x}{l} \right),$$

where $l = A(x/2)$.

To obtain Theorem 1, we apply $l \geqslant x/(2K)$ by the Schnirelmann density condition and then (4.7) yields the required estimate for $x \geqslant 2K$; for $x < 2K$ we use the obvious $T(x) \geqslant 1 + \pi(x)$.

To obtain Theorem 3, we observe that the condition $\underline{d}(A) = 1/K$ implies $l \geqslant x/(3K)$ for large x and then (4.7) yields

$$\underline{d}(T) \geqslant c_6/\log\log K. \quad \blacksquare$$

Proof of Lemma 4.4. By the power-mean inequality we have with any $k \geqslant 1$

$$(4.8) \qquad \beta \stackrel{\text{def}}{=} |B|^{-1} \sum_{b \in B} g(b) \leqslant \left(|B|^{-1} \sum_{b \in B} g(b)^k \right)^{1/k} \leqslant |B|^{-1/k} \left(\sum_{n \leqslant y} g(n)^k \right)^{1/k}.$$

We represent the function $g(n)^k$ in the form

$$g(n)^k = \sum_{d \mid n} h(d).$$

Since g is multiplicative, so is h, and clearly

$$h(p^j) = g(p^j)^k - g(p^{j-1})^k = \begin{cases} (1+1/p)^{k-1} & \text{for } j = 1, \\ 0 & \text{for } j \geqslant 2. \end{cases}$$

Hence

$$\sum_{n \leqslant y} g(n)^k = \sum_{d \leqslant y} h(d)\,[x/d] \leqslant y \sum_{d \leqslant y} h(d)/d \leqslant y \prod_p (1 + h(p)/p).$$

To estimate this product we distinguish the cases $p \leqslant k$ and $p > k$. For $p \leqslant k$ we use

$$1 + h(p)/p \leqslant 1 + h(p) = (1 + 1/p)^k,$$

whence

$$\prod_{p \leqslant k} (1 + h(p)/p) \leqslant \prod_{p \leqslant k} (1 + 1/p)^k \leqslant (c_3 \log k)^k \qquad (k \geqslant 2)$$

with an absolute constant c_3.

For $p > k$ we have

$$(1 + 1/p)^k \leqslant e^{k/p} \leqslant 1 + 2k/p,$$

thus

$$\prod_{k > p} (1 + h(p)/p) \leqslant \prod_{p > k} (1 + 2k/p^2) \leqslant \exp 2k \sum_{p > k} p^{-2} \leqslant c_4.$$

Thus we obtained

$$\sum_{n \leqslant y} g(n)^k \leqslant y(c \log k)^k.$$

Combined with (4.8) this means

$$\beta \leqslant (y/|B|)^{1/k} \log k.$$

Lemma 4.4 follows by taking $k = \max(2, \log(x/|B|))$.

5. The squares. Now we put $T = A + Q$. The proof will be parallel to that in the previous section, with much difference in the details. To estimate $T(x)$ like there, we need a bound for

$$|(Q + a) \cap (Q + b)|.$$

This is the same as the number of solutions of

$$a - b = m^2 - n^2 = (m - n)(m + n).$$

This is, in general, too large; we need a quantity that is bounded (on the average). This can be achieved by restricting the choice of a, b to small numbers and of m, n to big ones. To be more exact, let

$$A^* = A \cap [1, L\sqrt{x}], \qquad Q^* = Q \cap [x/4, x - L\sqrt{x}].$$

Clearly

$$T(x) \geqslant |A^* + Q^*|.$$

Let $\tau_L(n)$ denote the number of divisors of n up to L.

Lemma 5.1. *For $a, b \in A^*$, $a \neq b$, we have*

$$|(Q^* + a) \cap (Q^* + b)| \leqslant \tau_L(|a - b|).$$

Proof. Assume $a > b$. A solution (m, n) of the equation

$$a - b = m^2 - n^2 = (m - n)(m + n)$$

is uniquely determined by the value of $d = m - n$, which is a positive divisor of $a - b$ and which is at most L, since $m, n \in Q^*$ means $m, n \geqslant \sqrt{x}/2$ and hence

$$d = m - n = \frac{a - b}{m + n} \leqslant \frac{L\sqrt{x}}{\sqrt{x}} = L.$$

We need also an analogue of Lemma 4.4.

Lemma 5.2. *Let $B \subset [1, y]$ be a set of integers, $|B| \geqslant y/L$, $L \geqslant 3$. For every $\varepsilon > 0$ we have with a suitable constant c_ε*

$$(5.3) \qquad \sum_{b \in B} \tau_L(b) \leqslant c_\varepsilon |B| L^{(\log 2 + \varepsilon)/\log\log L}.$$

We assume this lemma now and postpone its proof to the end of this section.

Put $L = 3K$; for large x we have

$$S = |Q^*| |A^*| \geqslant A(L\sqrt{x})(\sqrt{x}/2 - L) > x.$$

By Lemma 5.2, for a fixed a we have

$$\sum_{b \in A^*, b \neq a} |(Q^* + a) \cap (Q^* + b)| \leqslant c_\varepsilon |A^*| L^{(\log 2 + \varepsilon)/\log\log L}$$

hence

$$S_2 \leqslant c_\varepsilon |A^*|^2 L^{(\log 2 + \varepsilon)/\log\log L}$$

and now an application of Lemma 3.1 yields Theorem 2.

Proof of Lemma 5.2. Like in (4.8), we have

$$(5.4) \qquad \beta \overset{\text{def}}{=} |B|^{-1} \sum_{b \in B} \tau_L(b) \leqslant |B|^{-1/k} \left(\sum_{n \leqslant y} \tau_L(n)^k \right)^{1/k}.$$

We assume k to be an integer. $\tau_L(b)^k$ is the number of k-tuples $(d_1, \ldots, d_k)$ where every $d_i | b$, thus

$$\sum_{n \leqslant x} \tau_L(n)^k \leqslant x \sum_{d_1, \ldots, d_k \leqslant L} [d_1, \ldots, d_k]^{-1}.$$

Lemma 5.5. *Let $d_1, \ldots, d_k$ be natural numbers. For a set*

$$J \subset I = \{1, \ldots, k\}$$

write

$$\Delta_J = \gcd\{d_i : i \in J\},$$

$$\delta_J = \Delta_J/(\Delta_J, \text{lcm}\{d_j : j \notin I\}).$$

We have $\delta_\emptyset = 1$,

(5.6)
$$\prod_{J \subset I} \delta_J = [d_1, \ldots, d_k]$$

and

(5.7)
$$d_j = \prod_{j \in J} \delta_J$$

for all j.

For a proof, just compare the exponent of an arbitrary prime in both sides of (5.6) and (5.7).

(5.7) shows that (i) if all the d_j are $\leqslant L$, then so are all the δ_J, (ii) for two different k-tuples, the corresponding systems (δ_J) will also be different.

By the Lemma we have

$$\sum_{d_1, \ldots, d_k \leqslant L} [d_1, \ldots, d_k]^{-1} \leqslant \sum_{\delta_J \leqslant L} \Big(\prod_{J \subset I, J \neq \emptyset} \delta_J \Big)^{-1}$$

$$\leqslant \Big(\sum_{a \leqslant L} 1/a \Big)^{2^k - 1} \leqslant (1 + \log L)^{2^k - 1}.$$

Combining our estimates we obtain

$$\beta \leqslant (x/|B|)^{1/k} (1 + \log L)^{(2^k - 1)/k}.$$

To obtain the Lemma, we put

$$k = \left[\frac{\log \log L}{\log 2 + \varepsilon} \right].$$

This concludes the proof of Lemma 5.2 and hence that of Theorem 2.

6. Concluding remarks. Since we are interested in small densities, according to the terminology of Plünnecke, we are speaking about *extraordinary components* (außenordentliche Komponente) rather than essential components. He calls a set H an extraordinary component if the *impact function* (Wirkungsfunktion)

$$f(x) = \inf \{ \sigma(A + H): \sigma(A) \geqslant x \}$$

satisfies $f(x)/x \to \infty$ as $x \to 0$. It is known that there are extraordinary components that are not essential components and it seems likely but is not yet proved that every essential component is an extraordinary component.

With my method it would be possible to determine this impact function for the sets P and Q and small values of x exactly, but it does not work for any thinner set. For the set of cubes rather than the squares I cannot tell anything nontrivial. On the other hand, Plünnecke's method of [7] works equally well also in this case and yields $f(x) \gg x^{4/5}$. Obviously $f(x) \ll x^{2/3}$ and one would think this is the correct order of magnitude. Also I feel that

for the cubes and the asymptotical density the set of multiples of a "highly composite" m is about the worst possible, thus

$$f(1/K) \gg K^{-c/\log \log K}$$

with some $c > 0$, but presently I have no idea of how to prove this.

References

[1] P. Erdős, *On the arithmetical density of the sum of two sequences, one of which forms a basis for the integers*, Acta Arith. 1 (1936), pp. 197–200.

[2] H. Halberstam and K. F. Roth, *Sequences*, Clarendon Press, Oxford 1966.

[3] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 4th ed., Clarendon Press, Oxford 1968.

[4] A. Khintchine, *Über ein metrisches Problem der additiven Zahlentheorie*, Mat. Sbornik 40 (1933), pp. 180–189.

[5] M. Kneser, *Abschätzung der asymptotischen Dichte von Summenmengen*, Math. Zeitschr. 58 (1953), pp. 459–484.

[6] H. Plünnecke, *Über ein metrisches Problem der additiven Zahlentheorie*, J. Reine Angew. Math. 197 (1957), pp. 97–103.

[7] — *Über die Dichte der Summe zweier Mengen, deren eine die Dichte null hat*, ibid. 205 (1960/61), pp. 1–20.

[8] — *Eine zahlentheoretische Anwendung der Graphentheorie*, ibid. 243 (1970), pp. 171–183.

[9] K. Prachar, *Primzahlverteilung*, Berlin–Göttingen–Heidelberg 1957.

[10] L. G. Schnirelmann, *Über additive Eigenschaften von Zahlen*, Math. Annalen 107 (1933), pp. 649–690.