ON THE EASIER WARING PROBLEM FOR POWERS OF PRIMES. I

By PAUL ERDÖS

[Communicated by Mr H. Davenport]

[Dedicated to Prof. E. Landau on his 60th birthday]

[*Received* 16 November, *read* 7 December 1936]

The famous theorem of Schnirelmann states that a constant c exists such that every integer greater than one may be expressed as the sum of at most c primes. Recently, Heilbronn, Landau and Scherk proved that this holds with $c=71$. Probably the true value of c is 3. Another well-known theorem (Hilbert's solution of Waring's problem) is that every integer is the sum of a bounded number of positive kth powers. If we omit the restriction that all the kth powers are positive, the problem is referred to as the easier Waring problem and the proof of the result is then much simpler.

It is an interesting but very difficult question, which I cannot answer, whether every integer is the sum of a bounded number of kth powers of primes. It would suffice to prove that for some fixed l the integers of the form $p_1^k+p_2^k+\ldots+p_l^k$ (where the p's and the q's throughout this paper denote primes), have a positive density. I can, however, prove

Theorem I. *The positive integers of the form* $p_1^2+p_2^2-q_1^2-q_2^2$ *have positive density.*

I then deduce

Theorem II. *A constant* c *exists such that every integer is the sum of at most* c *positive and negative squares of primes.*

This follows from Theorem I by Schnirelmann's well-known

Lemma. *If* $n_1=1, n_2, n_3, \ldots$ *is a sequence of positive integers such that, for every* $n \geqslant 1$, *there are at least* αn *of them not exceeding* n, *where* α *is positive and independent of* n, *then every positive integer can be expressed as the sum of a bounded number of the* $n_1, n_2, \ldots$.

I shall prove by more complicated arguments in another paper that the density of each of the sets of integers

$$p_1^2+p_2^2-p_3^2, \quad \sum_{r=1}^{4} p_r^3-\sum_{r=1}^{4} q_r^3, \quad \sum_{r=1}^{2^k} \epsilon_r p_r^k, \quad (\epsilon_r=\pm 1),$$

is positive. I conjecture that the density of integers of the form $p_1^2+p_2^2+p_3^2$ is also positive, but this seems to be very deep.

Throughout this paper n denotes a sufficiently large positive integer, the c's and γ's are positive numbers independent of n, not necessarily the same at each time of occurrence. Also $0<\gamma<1$, and γ will be used only as an exponent of n in applications of Lemma 2.

LEMMA 1. *Let $k>0$, $x, x_1, x_2, \ldots, x_r$ be $r+2$ integers. Write*

$$f(x)=\prod_{s=1}^{r}(x-x_s), \tag{1}$$

$$g(x)=\begin{cases}\prod\limits_{p\mid f(x)}\left(1+\dfrac{k}{p}\right) & \text{if } f(x)\neq 0,\\ 0 & \text{if } f(x)=0.\end{cases}$$

Then if $\max|x_s|\leqslant n$, and ϵ is any positive number, there exists $g=g(r,k,\epsilon)$ such that, for $n>n(\epsilon)$,

$$\sum_{\substack{x=1\\ g(x)>g}}^{n} g(x)<\epsilon n. \tag{2}$$

It suffices to prove that

$$E=\sum_{x=1}^{n} g(x)^2\leqslant c(k,r)\,n, \tag{3}$$

for then we can take $g=c/\epsilon$, and

$$\sum_{\substack{x=1\\ g(x)>g}}^{n} g(x)<\frac{1}{g}\sum_{x=1}^{n} g(x)^2\leqslant\frac{\epsilon}{c}cn$$

by (3).

Now
$$E=\sum_{\substack{x=1\\ f(x)\neq 0}}^{n}\prod_{p\mid f(x)}\left(1+\frac{k}{p}\right)^2.$$

For the primes $p\geqslant k$, we note that

$$\left(1+\frac{k}{p}\right)^2\leqslant 1+\frac{3k}{p}.$$

Hence
$$E\leqslant\sum_{\substack{x=1\\ f(x)\neq 0}}^{n}\prod_{p<k}\left(1+\frac{k}{p}\right)^2\prod_{p\mid f(x)}\left(1+\frac{3k}{p}\right)$$

$$\leqslant c(k)\sum_{\substack{x=1\\ f(x)\neq 0}}^{n}\sum_{d\mid f(x)}\frac{\mu^2(d)\,(3k)^{\lambda(d)}}{d},$$

where $\lambda(d)$ denotes the number of different prime factors of d. We invert the order of summation. The number of solutions of $f(x)\equiv 0 \pmod d$, $1\leqslant x\leqslant n$, does not exceed $r^{\lambda(d)}\left(\left[\frac{n}{d}\right]+1\right)$, and $|f(n)|\leqslant(2n)^r$.

Hence
$$E \leqslant c(k) \sum_{d=1}^{(2n)^r} \frac{\mu^2(d)(3kr)^{\lambda(d)}}{d}\left(\frac{n}{d}+1\right)$$
$$\leqslant c(k)\, n \prod_{p=2}^{\infty}\left(1+\frac{3kr}{p^2}\right)+c(k) \prod_{p<(2n)^r}\left(1+\frac{3kr}{p}\right)$$
$$\leqslant c(k,r)\, n,$$

since
$$\sum_{p \leqslant N} \frac{1}{p} = o(\log N).$$

LEMMA 2 (Brun*). *Let there be given an arithmetical progression*
$$x \equiv \Delta \pmod{D}, \quad 0 < \Delta \leqslant D,$$
and r consecutive primes $p_1 < p_2 \ldots < p_r$, none of which divides D. With each prime p_t let there be associated k residues $R_{s,t}$ $(s = 1, 2, \ldots, k)$, where k is an assigned integer less than each of the p's. If $\nu(x, n)$ is the number of the $x \leqslant n$ such that
$$x \not\equiv R_{s,t} \pmod{p_t}, \quad (s = 1, 2, \ldots, k,\ t = 1, 2, \ldots, r)$$
then
$$\nu(x, n) < \frac{cn}{D} \prod_{p}\left(1-\frac{k}{p}\right),$$
provided that all the p's do not exceed n^γ where γ is a certain positive numerical constant less than 1.

This is proved by Brun for $k = 1$, 2 and the proof is the same for general k. It is also clear on examining Brun's proof that the result still holds† when D depends upon n if $D < n^\epsilon$, where ϵ is sufficiently small.

LEMMA 3 (Schnirelmann‡). *If a is a positive integer, the number of solutions of the equation*
$$p_1 - p_2 = a,$$
in primes $p_2 < p_1 \leqslant n$, , is less than
$$c \frac{n}{(\log n)^2} \prod_{p \mid a}\left(1+\frac{1}{p}\right).$$

LEMMA 4. *With suitable c, there exist two positive integers a, b with highest common factor (a, b) such that*

(1)
$$(a, b) < c,$$
$$c \log n < a < b < 10 \log n,$$
$$\prod_{p \mid a}\left(1+\frac{1}{p}\right) < c, \quad \prod_{q \mid b}\left(1+\frac{1}{q}\right) < c;$$

* "Le crible d'Eratosthène et le théorème de Goldbach", *Videnskapsselskapets Skrifter*, I, Mat. Naturv. Klasse 1920, No. 3, Kristiania.

† *Loc. cit.* p. 22, eqn. (20).

‡ "Über additive Eigenschaften von Zahlen", *Math. Annalen*, 107 (1933), 649–690, p. 670. Lemma 3 can be deduced from Lemma 2.

(2) *each of the equations*

$$p_1-p_2=a, \quad q_1-q_2=b$$

has more than $cn/(\log n)^3$ *solutions in primes* p, q *satisfying*

$$p_1<\frac{n}{60\log n}, \quad q_1<\frac{n}{60\log n}.$$

Let $\nu(x)$ be the number of solutions of

$$p_3-p_4=x$$

with $$p_3<n/60\log n.$$

We first find a lower bound for

$$\nu=\sum_{x=1}^{10\log n}\nu(x).$$

Split the interval $(0, n/60\log n)$ into $1+[n/600(\log n)^2]$ subintervals each of length $10\log n$ except the first, which may be ignored, and containing $\nu_1, \nu_2, \ldots$ primes respectively. Then evidently on taking any two primes in each subinterval, we have

$$2\nu\geqslant\nu_1^2+\nu_2^2+\ldots-\nu_1-\nu_2-\ldots.$$

Now

$$\frac{n}{30(\log n)^2}>\nu_1+\nu_2+\ldots>\frac{n}{120(\log n)^2},$$

since the number of primes not exceeding N lies between $N/\frac{3}{2}\log N$ and $\frac{3}{2}N/\log N$. Further, if the sum of the ν's is given, the sum of their squares will be a minimum when all the ν's are equal. Hence, by taking

$$\nu_i=\frac{n}{120(\log n)^2}\Big/\left[\frac{n}{600(\log n)^2}\right]>5$$

in $\sum\nu_i^2$, we obtain

$$2\nu>25\left[\frac{n}{600(\log n)^2}\right]-\frac{n}{30(\log n)^2}>\frac{n}{120(\log n)^2}-25,$$

whence

$$\nu>\frac{n}{241(\log n)^2}. \tag{4}$$

Now an upper bound for $\nu(x)$ is given by Lemma 3 as

$$\nu(x)<\frac{cn}{(\log n)^3}\prod_{p\mid x}\left(1+\frac{1}{p}\right). \tag{5}$$

Consider first the values of x for which $P=\prod_{p\mid x}\left(1+\frac{1}{p}\right)>c$. Then from Lemma 1 on taking $r=1$, $k=1$, $x_1=0$, $g=c$, we have

$$\sum_{\substack{x<10\log n\\ P>c}}\nu(x)<c\,\frac{\epsilon n\log n}{(\log n)^3}<\frac{n}{482(\log n)^2}.$$

Hence, from (4),

$$\sum_{\substack{x<10\log n\\ P\leqslant c}}\nu(x)>\frac{n}{482(\log n)^2}. \tag{6}$$

Hence, noting from (5) that, for these x, $\nu(x) < cn/(\log n)^3$, there must clearly be at least $c \log n$ terms in the sum in (6) for which

$$x < 10 \log n, \quad \prod_{p \mid x} \left(1 + \frac{1}{p}\right) \leqslant c, \quad \nu(x) > cn/(\log n)^3.$$

Hence there must be $c \log n$ integers x satisfying these conditions and also the further condition

$$c \log n < x < 10 \log n.$$

Hence among these integers there exist two, say a and b, such that

$$c \log n < a, \quad b < 10 \log n$$

and

$$a - b < c, \quad \text{whence} \quad (a, b) = (a, a-b) < c.$$

THEOREM 1. *The integers of the form $p_1^2 + p_2^2 - q_1^2 - q_2^2$ have a positive density.*

Let a, b be the integers of Lemma 4. Consider the two sets of integers defined by

$$p_1^2 - p_2^2, \quad q_1^2 - q_2^2,$$

where the p's and q's are given by

$$p_1 - p_2 = a, \quad q_1 - q_2 = b, \quad p_1, q_1 < n/60 \log n. \tag{7}$$

Then

$$p_1^2 - p_2^2 = a(2p_2 + a) < \tfrac{1}{3}n + a^2,$$

and similarly

$$q_1^2 - q_2^2 = b(2q_2 + b) < \tfrac{1}{3}n + b^2.$$

Since there are at least $cn/(\log n)^3$ values of p_2 and also of q_2, there are at least $cn^2/(\log n)^6$ even integers m, not necessarily all different, given by

$$m = 2ap_2 + 2bq_2. \tag{8}$$

Let $\nu_1(m)$ denote the number of times m occurs. Clearly $m < \dfrac{2n}{3} < n$, and so

$$\sum_{m=1}^{n} \nu_1(m) \geqslant cn^2/(\log n)^6. \tag{9}$$

We estimate $\nu_1(m)$ by Brun's method. If $p_2 \leqslant m^\gamma$ there is at most one value of q_2 for each p_2, and similarly if $q_2 \leqslant m^\gamma$. These p_2 and q_2 give a contribution of at most $2m^\gamma$ to $\nu_1(m)$. Suppose then that $p_2 > m^\gamma$, $q_2 > m^\gamma$. Let p be an arbitrary prime not exceeding m^γ with $p \nmid 2ab$. Then

$$p_2 \not\equiv 0, \quad -a, \quad \frac{m}{2a}, \quad \frac{m}{2a} + \frac{b^2}{a} \pmod{p},$$

by (7) and (8). These four residues are all different if we assume $p \nmid M$, where

$$M = m(m + 2a^2)(m + 2b^2)(m + 2a^2 + 2b^2).$$

Denote by $\nu_2(m)$ the number of integers $x < m/2a$ for which

$$x \not\equiv 0, \quad -a, \quad \frac{m}{2a}, \quad \frac{m}{2a} + \frac{b^2}{a} \pmod{p},$$

for any of the primes $p < m^\gamma$ not dividing $2ab$; and also for which

$$m - 2ax \equiv 0 \pmod{2b},$$

so that x may be one of $[a, b]$ different residues $(\mathrm{mod}\, b)$. Hence, by Lemma 2,

$$\nu_2(m) < \frac{cm}{2ab}[a,b] \prod_{\substack{p<m^\gamma \\ p \nmid 2abM}} \left(1 - \frac{4}{p}\right).$$

On using the inequality

$$1 < \left(1 + \frac{5}{p}\right)\left(1 - \frac{4}{p}\right), \quad p > 20,$$

for the primes $p > 20$ dividing $2abM$, we have

$$\begin{aligned} \nu_2(m) &< \frac{cm}{ab} \prod_{p \mid 2ab}\left(1 + \frac{5}{p}\right) \prod_{p \mid M}\left(1 + \frac{5}{p}\right) \prod_{20<p<m^\gamma}\left(1 - \frac{4}{p}\right) \\ &< \frac{cn}{(\log n)^6} \prod_{p \mid M}\left(1 + \frac{5}{p}\right), \end{aligned} \tag{10}$$

since $$\prod_{20<p<m^\gamma}\left(1 - \frac{4}{p}\right) < \frac{c}{(\log m)^4}, \quad \prod_{p \mid 2ab}\left(1 + \frac{5}{p}\right) < c$$

from (1) of Lemma 4.

Now apply Lemma 1 with

$$k = 5, \quad r = 4, \quad x_1 = 0, \quad x_2 = -2a^2, \quad x_3 = -2b^2, \quad x_4 = -2a^2 - 2b^2.$$

Then $$\sum_{\substack{m=1 \\ \prod_{p \mid M}\left(1+\frac{5}{p}\right) > c}}^{n} \nu_2(m) < \frac{\epsilon n^2}{(\log n)^6}.$$

But $$\nu_1(m) \leqslant \nu_2(m) + 2m^\gamma, \tag{11}$$

and so $$\sum_{\substack{m=1 \\ \prod_{p \mid M}\left(1+\frac{5}{p}\right) \geqslant c}}^{n} \nu_1(m) < \frac{\epsilon n^2}{(\log n)^6} + 2n^{1+\gamma} < \frac{2\epsilon n^2}{(\log n)^6}. \tag{12}$$

Hence, from (9), (11), by choice of a suitable ϵ,

$$\sum_{\substack{m=1 \\ \prod_{p \mid M}\left(1+\frac{5}{p}\right) < c}}^{n} \nu_1(m) \geqslant cn^2/(\log n)^6. \tag{13}$$

But for the m in (13) we have, from (10), (11),

$$\nu_2(m) < cn/(\log n)^6, \quad \nu_1(m) < cn/(\log n)^6.$$

Hence in (13) at least

$$\frac{cn^2}{(\log n)^6} \Big/ \frac{cn}{(\log n)^6} = cn$$

of the $\nu_1(m)$ are not zero. Hence also there must be at least cn integers of the form

$$2ap_2 + 2bq_2 + a^2 + b^2.$$

These integers are obviously less than n and are also of the form

$$p_1^2 - p_2^2 + p_3^2 - p_4^2.$$

This proves Theorem **I.**

I should like to express my deep gratitude to Mr Davenport and Prof. Mordell for their help with my manuscript.