Annales Univ. Sci. Budapest., Sect. Comp. 32 (2010) 13-21

ON SUMS OF POWERS OF PRIME FACTORS OF AN INTEGER

J.-M. DeKoninck (Québec, Canada)
F. Luca (Morelia, Mexico)

Abstract. In this paper, we look at the positive integers n such that the equality

$$\sum_{p^{\alpha_p}||n} p^{\alpha_p} = \left(\sum_{p|n} p\right)^2$$

holds.

1. Introduction

For every positive integer

$$n = \prod_{p^{\alpha_p}||n} p^{\alpha_p}$$

let

$$B(n) = \sum_{p^{\alpha_p}||n} p^{\alpha_p} \qquad \text{and} \qquad \beta(n) = \sum_{p|n} p.$$

Plainly, $B(n) = \beta(n)$ if and only if n is squarefree. In this paper, we look at the positive integers n such that $B(n) = \beta(n)^2$. Let $\mathcal{A}$ be the set of such n. For a positive real number x we write $\mathcal{A}(x) = \mathcal{A} \cap [1, x]$. Since $\mathcal{A}$ contains all squares of primes, we get that $\#\mathcal{A}(x) \geq \pi(x^{1/2}) \gg x^{1/2}/\log x$. In this note, we show that $\mathcal{A}$ contains a lot more numbers, and in fact our main result is the following.

Theorem 1. *The estimates*

$$\frac{x}{\exp\left((2\sqrt{34/3}+o(1))\sqrt{\log x\log\log x}\right)}\le$$

$$\le \#\mathcal{A}(x)\le \frac{x}{\exp\left(\left(\frac{1}{\sqrt{2}}+o(1)\right)\sqrt{\log x\log\log x}\right)}$$

hold as $x\to\infty$.

Throughout, we use the Vinogradov symbols $\gg$ and $\ll$ and the Landau symbols O and o with their regular meanings. We use log for the natural logarithm and p, q and r with or without subscripts for prime numbers.

Acknowledgments. This work was done in April of 2006, while the second author was in residence at the Centre de Recherche Mathématique of the Université de Montréal for the thematic year *Analysis and Number Theory*. This author thanks the organizers for the opportunity of participating in this program. He was also supported in part by Grants SEP-CONACyT 46755, PAPIIT IN104505 and a Guggenheim Fellowship. The first author was supported in part by a grant from NSERC.

2. The upper bound

For a positive integer n we write $P(n)$ for the largest prime factor of n. Let us consider that following sets:

$$\mathcal{A}_1(x)=\{n\le x\mid n=p^2\},$$

$$\mathcal{A}_2(x)=\{n\le x\mid P(n)<y\}$$

and

$$\mathcal{A}_3(x)=\left\{n\le x\mid n\notin\mathcal{A}_2(x),\ P(n)^2\mid n\right\},$$

where y is a parameter which depends on x to be chosen later and which satisfies $\exp((\log\log x)^2)\le y\le x$, and $P(n)$ denotes the largest prime factor of n.

Plainly,

$$\#\mathcal{A}_1(x)=\pi(x^{1/2})\ll\frac{x^{1/2}}{\log x}. \tag{1}$$

From standard estimates on smooth numbers [1], we know that if we set $u = \log x/\log y$, then

$$\#\mathcal{A}_2(x) \ll \frac{x}{\exp((1+o(1))u\log u)} \qquad (x\to\infty) \tag{2}$$

in our range for y versus x, while

$$\#\mathcal{A}_3(x) \le \sum_{\substack{p \text{ prime}\\ p\ge y}} \left\lfloor \frac{x}{p^2} \right\rfloor \le x\sum_{n\ge y}\frac{1}{n^2} \ll \frac{x}{y}. \tag{3}$$

Let $\mathcal{A}_4(x) = \mathcal{A}(x)\setminus(\mathcal{A}_1(x)\cup\mathcal{A}_2(x)\cup\mathcal{A}_3(x))$. If $n\in\mathcal{A}_4(x)$, then we can write $n = P(n)m$, where $m>1$ (note that $\omega(n)>1$ since n belongs to $\mathcal{A}(x)$ but not to $\mathcal{A}_1(x)$). Furthermore, since $n\notin\mathcal{A}_3(x)$, we have $P(n)\nmid m$. Since $n\in\mathcal{A}(x)$, we can write

$$P(n)+B(m) = B(n) = \beta(n)^2 = (\beta(m)+P(n))^2,$$

so that

$$P(n)^2 + (2\beta(m)-1)P(n) + (\beta(m)^2 - B(m)) = 0.$$

Hence, $P(n)$ is determined in at most two ways by m. Furthermore, note that for the positive integers n under consideration, we have that $P(n)\ge y$, implying that $m\le x/y$, so that

$$\#\mathcal{A}_4(x) \le 2\sum_{m\le x/y} \ll \frac{x}{y}. \tag{4}$$

From estimates (1), (2), (3) and (4), we immediately deduce that

$$\#\mathcal{A}(x) \le \#\mathcal{A}_1(x)+\#\mathcal{A}_2(x)+\#\mathcal{A}_3(x)+\#\mathcal{A}_4(x) \ll$$
$$\ll \frac{x^{1/2}}{\log x} + \frac{x}{y} + \frac{x}{\exp((1+o(1))u\log u)}.$$

To minimize the right hand side above we choose $y = \exp(u\log u)$, which amounts to

$$\log^2 y = \log x \log\left(\frac{\log x}{\log y}\right).$$

Thus, we get that $\log y = (1+o(1))\sqrt{\log x \log\log x}$ as $x \to \infty$, and with this choice of y versus x we obtain

$$\#\mathcal{A}(x) \ll \frac{x}{\exp\left(\left(\frac{1}{\sqrt{2}}+o(1)\right)\sqrt{\log x \log\log x}\right)}$$

as $x \to \infty$, thus establishing the upper bound in Theorem 1.

3. The lower bound

Let N be a large odd integer which is not a multiple of 3. In particular, $N^2 \equiv 1 \pmod{24}$. Let M and k be some positive functions of N to be specified more precisely later on, where for now $M = N^{1+o(1)}$ and $k = o(\log N)$ as $N \to \infty$. We assume that k is an integer congruent to 17 modulo 24. Let $c_1 = (1-1/1.1)^{1/3}$, $c_2 = (1-1/6.9)^{1/3}$, and let $p \in \mathcal{I} = [c_1 N^{2/3}, c_2 N^{2/3}]$ and $q_1 < \cdots < q_k \in \mathcal{J} = [M/2, M]$ be all primes congruent to 1 modulo 24. Let

$$N_1 = N - p - \sum_{i=1}^{k} q_i \qquad \text{and} \qquad N_2 = N^2 - p^3 - \sum_{i=1}^{k} q_i. \tag{5}$$

Note that N_1 is odd and $N_2 \equiv 7 \pmod{24}$. Furthermore, assuming that

$$kM = o(N), \tag{6}$$

then

$$N_1 = N(1 - p/N + O(kM/N)) = N(1+o(1))$$

and

$$N_2 = N^2(1 - p^3/N^2 + O(kM/N^2)),$$

so that

$$\frac{N_1^2}{N_2} = \frac{N^2(1+o(1))}{N^2(1-p^3/N^2+o(1))} \in [1.1+o(1), 6.9+o(1)]$$

as $N \to \infty$, because $p \in \mathcal{I}$. Thus, from the application of Theorem 16 on page 139 in Hua's book [2] mentioned on page 156 of the same book, we have that the number of solutions $(p_1, \ldots, p_7)$ of the system of equations

$$p_1 + \cdots + p_7 = N_1 \qquad \text{and} \qquad p_1^2 + \cdots + p_7^2 = N_2 \tag{7}$$

is

$$\geq c_3 \frac{N^4}{(\log N)^7}, \tag{8}$$

where c_3 is some positive absolute constant. We now show that for large N, most of such solutions have $p_i \neq p_j$ for $i \neq j$ in $\{1, \ldots, 7\}$ and also that $p_i \notin \{p, q_1, \ldots, q_k\}$ for any $i \in \{1, \ldots, 7\}$.

Let us count the solutions to the system of equations (7) having $p_i = p_j$ for some $i \neq j$. We assume that $p_6 = p_7$. Since $p_i \leq N$ for all $i = 1, \ldots, 7$, it follows that the triplet (p_4, p_5, p_6) can be chosen in at most $O(N^3/(\log N)^3)$ ways. Assume now that p_4, p_5 and p_6 have been chosen and that $p_6 = p_7$. Then

$$p_1 + p_2 + p_3 = A \qquad \text{and} \qquad p_1^2 + p_2^2 + p_3^2 = B, \tag{9}$$

where $A = N_1 - (p_4 + p_5 + 2p_6)$ and $B = N_2 - (p_4^2 + p_5^2 + 2p_6^2)$. Expressing p_3 versus p_1 and p_2 from the left equation above and inserting the answer into the right equation above we get

$$p_1^2 + p_2^2 + (A - p_1 - p_2)^2 = B,$$

or

$$p_1^2 + p_2^2 + p_1 p_2 - Ap_1 - Ap_2 + A^2/2 = B/2.$$

This last equation above can be rewritten as

$$\left(p_1 + \frac{p_2}{2} - \frac{A}{2}\right)^2 + \frac{3}{4}\left(p_2 - \frac{A}{3}\right)^2 = \frac{3B - A^2}{6},$$

or

$$3U^2 + V^2 = 2(3B - A^2), \quad \text{where} \quad U = 2p_1 + p_2 - A \text{ and } V = 3p_2 - A.$$

Note that p_1 and p_2 determine U and V uniquely and vice-versa. It is well known that if m is any fixed positive integer then the number of integer solutions (x, y) of the equation $3x^2 + y^2 = m$ is $O(2^{\omega_1(m)})$, where $\omega_1(m)$ is the number of prime divisors of m which are congruent to 1 modulo 3. It is well known that $\omega_1(m) \leq \omega(m) \leq c_4 \log m / \log\log m$, where $\omega(m)$ is the total number of distinct prime factors of m and c_4 is some absolute constant. Since $2(B - A^2) < 2N^2$, it follows that the number of possibilities for (p_1, p_2, p_3) once p_4, p_5 and p_6 are fixed is $\leq \exp(3c_4 \log N / \log\log N)$ provided N is sufficiently large. Hence, the total number of solutions of the system of equations (7) with $p_i = p_j$ for some $i \neq j$ is

$$\ll \frac{N^3}{(\log N)^3} \exp\left(3c_4 \frac{\log N}{\log\log N}\right) = N^{3+o(1)}.$$

Comparing this with estimate (8), it follows that for large N, at least

$$\frac{c_3}{2} \frac{N^4}{(\log N)^7} \tag{10}$$

solutions $(p_1, \ldots, p_7)$ exist with all the components distinct.

We now find an upper bound for the number of solutions for which $p_i \in \in \{p, q_1, \ldots, q_k\}$. Since $p \ll N^{2/3}$, $q_i \leq M$ and $M = N^{1+o(1)} > N^{2/3}$ for large N, it follows that $p_i \leq M$ for some $i = 1, \ldots, 7$. Assume that $p_7 \leq M$. Then the quadruplet (p_4, p_5, p_6, p_7) can be chosen in at most $O(N^3 M/(\log N)^4)$ ways. Now for each one of these choices we are led again to a system of equations (9) with suitable A and B depending on p_4, p_5, p_6 and p_7, which, by the previous argument, admits at most $\exp(3c_4 \log N/\log\log N)$ solutions provided N is sufficiently large. Thus, the number of such possibilities does not exceed

$$c_5 \frac{N^3 M}{(\log N)^4} \exp\left(3c_4 \frac{\log N}{\log\log N}\right)$$

for some positive constant c_5. We require that the above upper bound is smaller than a half of the expression shown at (10). This will be the case if

$$M \leq \frac{c_3}{4c_5} \frac{N}{(\log N)^3} \exp\left(-3c_4 \frac{\log N}{\log\log N}\right),$$

and at least

$$\frac{c_3}{4} \frac{N^4}{(\log N)^7}$$

solutions $(p_1, \ldots, p_7)$ exist where all the primes p_i are distinct and do not belong to $\{p, q_1, \ldots, q_k\}$. Hence, we see that it suffices to choose M smaller than

$$N \exp\left(-4c_4 \frac{\log N}{\log\log N}\right),$$

and then in light of the fact that we have assumed that $k = o(\log N)$, it follows that inequality (6) also holds.

With such choices $(p_1, \dots, p_7, p, q_1, \dots, q_k)$, we note that the number

$$n = p^3 \prod_{i=1}^{k} q_i \prod_{j=1}^{7} p_j^2$$

satisfies, recalling relations (5) and (7),

$$B(n) = p^3 + \sum_{i=1}^{k} q_i + \sum_{j=1}^{7} p_j^2 = N^2 = \left(p + \sum_{i=1}^{k} q_i + \sum_{j=1}^{7} p_j \right)^2 = \beta(n)^2.$$

Further, note that

$$n \le (c_2 N^{2/3})^3 M^k N^{14} := x. \tag{11}$$

The number of such n is, by the above argument and unique factorization,

$$\ge \frac{c_3}{4} \frac{N^4}{(\log N)^7} \pi'(\mathcal{I}) \binom{\pi'(\mathcal{J})}{k},$$

where $\pi'(\mathcal{I})$ and $\pi'(\mathcal{J})$ denote the number of primes congruent to 1 modulo 24 in the intervals $\mathcal{I}$ and $\mathcal{J}$, respectively. Since certainly

$$\pi'(\mathcal{I}) \ge \frac{(c_2 - c_1 + o(1))}{\phi(24)} \frac{N^{2/3}}{\log(N^{2/3})}$$

as $N \to \infty$, where ϕ stands for Euler's function, we get that $\pi'(\mathcal{I}) \ge$ $\ge c_5 N^{2/3}/\log N$ for large N, where c_5 is some appropriate positive constant. Furthermore, by a similar argument, we get

$$\pi'(\mathcal{J}) \ge c_6 \frac{M}{\log N},$$

where c_6 is also some appropriate constant. Hence, the number of such n is at least

$$\ge c_7 \frac{N^{4+2/3}}{(\log N)^8} \binom{\lfloor c_6 M/\log N \rfloor}{k} \ge c_7 \frac{N^{14/3}}{(\log N)^8} \frac{(c_8 M)^k}{(\log N)^k k!},$$

where $c_8 = c_6/2$, provided that N is large, because $M = N^{1+o(1)}$ and $k = = o(\log N)$. We thus get, using estimate (11), that

$$\#\mathcal{A}(x) \geq \frac{c_7 c_8^k N^{14/3} M^k}{(\log N)^{k+8} k^k} = \tag{12}$$
$$= \frac{x}{\exp\left(34/3 \log N + (k+8)\log\log N + k\log k + O(k)\right)}.$$

In light of (11) we get

$$\log x = k \log M + 16 \log N + O(1),$$

and since $\log M = (1 + o(1)) \log N$, we arrive at $\log x = (1 + o(1))k \log N$. In order to optimize the lower bound of $\#\mathcal{A}(x)$ shown at (12), we choose k and N such that $\log N$, $k \log\log N$ and $k \log k$ all have the same order of magnitude. This suggests choosing

$$k = \left\lfloor c_9 \sqrt{\frac{\log x}{\log\log x}} + O(1) \right\rfloor,$$

where c_9 is a constant to be optimally chosen later on, and $O(1)$ accounts for the fact that $k \equiv 17 \pmod{24}$. Thus, since

$$k = c_9(1 + o(1)) \left(\frac{\log x}{\log\log x}\right)^{1/2},$$

and $k \log N = (1 + o(1)) \log x$, we get

$$\log N = c_9^{-1}(1 + o(1))(\log x \log\log x)^{1/2},$$

$$k \log k = \frac{c_9}{2}(1 + o(1))(\log x \log\log x)^{1/2},$$

and

$$k \log\log N = \frac{c_9}{2}(1 + o(1))(\log x \log\log x)^{1/2},$$

showing that

$$\frac{34}{3}\log N + (k+8)\log\log N + k\log k + O(k) =$$
$$= \left(\frac{34}{3c_9} + c_9 + o(1)\right)(\log x \log\log x)^{1/2}.$$

Thus, the optimal constant c_9 is the one which minimizes the function $h(t) = 34/(3t) + t$. Since $h'(t) = -34/(3t^2) + 1$, we get that the minimum of this function is achieved at $t_0 = \sqrt{34/3}$, for which $h(t_0) = 2\sqrt{34/3}$. Thus, choosing $c_9 = t_0$, we get that

$$\#\mathcal{A}(x) \geq \frac{x}{\exp\left((2\sqrt{34/3} + o(1))\sqrt{\log x \log\log x}\right)},$$

thus completing the proof of the theorem.

References

[1] **Hildebrand A.,** On the number of positive integers $\leq x$ and free of prime factors $> y$, *J. Number Theory,* **22** (1986), 289-307.

[2] **Hua L.K.,** *Additive theory of prime numbers*, Translations of Mathematical Monographs **13**, AMS, 1965.

(Received June 7, 2007)

J.-M. De Koninck
Département de mathématiques
Université Laval
Québeck G1K 7P4
Canada
jmdk@mat.ulaval.ca

F. Luca
Instituto de Matemáticas
Universidad Autónoma de México
C.P. 58089 Morelia
Michoacán, México
fluca@matmor.unam.mx