

ON INTEGERS OF THE FORM $p + 2^k$

LAURENT HABSIEGER AND XAVIER-FRANÇOIS ROBLLOT

ABSTRACT. We investigate the density of integers that may be written as $p + 2^k$, where p is a prime and k a nonnegative integer.

1. INTRODUCTION

Troughout this paper, the symbol p will denote a prime and k will be a nonnegative integer. Romanov [R] proved that the integers of the form $p + 2^k$ have positive density. He also raised the following question : does there exists an arithmetic progression consisting only of odd numbers, no term of which is of the form $p + 2^k$? Erdős [E] found such an arithmetic progression by considering integers which are congruent to 172677 modulo 5592405 = $(2^{24} - 1)/3$. Thus the density of numbers of the form $p + 2^k$ is less than $1/2$, the trivial bound obtained from the odd integers. For convenience we introduce

$$\underline{d} = \liminf_{x \rightarrow \infty} \frac{\#\{p + 2^k \leq x\}}{x/2} \quad \text{and} \quad \bar{d} = \limsup_{x \rightarrow \infty} \frac{\#\{p + 2^k \leq x\}}{x/2}.$$

The aim of this paper is to give an explicit version of the estimates $0 < \underline{d} \leq \bar{d} < 1$.

Theorem 1. *We have*

$$0.1866 < \underline{d} \leq \bar{d} < 0.9819.$$

This range is pretty large and Bombieri conjectured the more precise value 0.86 [P].

In section 2, we obtain the lower bound $0.1866 < \underline{d}$, by slightly refining a straightforward application of a recent result of Pintz and Ruzsa [PR], in their study of Linnik's approximation of Goldbach problem (see also [HB-P]). In section 3, we get the upper bound, using computations on residue classes.

2. THE LOWER BOUND

Let N be a large integer and put $L = \lfloor \log N / \log 2 \rfloor$. Define the functions

$$r(n) = \#\{(p, k) : n = p + 2^k, p \leq N, 1 \leq k \leq L\}$$

and

$$s(N) = \#\{(p_1, p_2, k_1, k_2) : p_1 - p_2 = 2^{k_2} - 2^{k_1}, p_j \leq N, 1 \leq k_j \leq L, j = 1, 2\},$$

so that

$$s(N) = \sum_{n=1}^N r^2(n).$$

Pintz and Ruzsa [PR] proved the following lemma.

Partially supported by the European Community IHRP Program, within the Research Training Network "Algebraic Combinatorics in Europe", grant HPRN-CT-2001-00272.

Lemma 1. *For N large enough, we have*

$$s(N) \leq \frac{2}{\log^2 2} CN,$$

where $C < 5.3636$.

Let $d(N)$ denote the number of positive integers $n \leq N$ which may be written in the form $n = p + 2^k$. The Cauchy-Schwarz inequality implies easily that

$$(\pi(N)L)^2 \leq d(N)s(N),$$

where $\pi(N)$ denotes the number of primes $p \leq N$. we deduce from Lemma 1 and from the prime number theorem that $2Cd(N) \geq (1 + o(1))N$, and the lower bound $\underline{d} \geq 1/C > 0.1864$ follows from the definitions.

To get the bound from the theorem, we need further notations. Put

$$\epsilon_N = \frac{\sum_{1 \leq n \leq N, r(n) > 0} r(n)}{\sum_{1 \leq n \leq N, r(n) > 0} 1} \quad \text{and} \quad \epsilon = \frac{2}{\underline{d} \log 2}.$$

By the definitions, there exists a subsequence of $(\epsilon_N)_{N \in \mathbb{N}}$ which converges to ϵ . Let us now refine the Cauchy-Schwarz inequality by studying

$$\Delta_N = \sum_{1 \leq n \leq N, r(n) > 0} (r(n) - \epsilon_N)^2,$$

so that

$$\begin{aligned} \Delta_N &= \sum_{1 \leq n \leq N} r^2(n) - \frac{\left(\sum_{1 \leq n \leq N} r(n)\right)^2}{\sum_{1 \leq n \leq N, r(n) > 0} 1} = s(N) - \frac{(\pi(N)L)^2}{d(N)} \\ &\leq \left(5.3636 - \frac{1}{\underline{d}} + o(1)\right) \frac{2N}{\log^2 2}, \end{aligned}$$

for infinitely many N . Without loss of generality we may assume that $\epsilon \in]15, 15.5[$: otherwise we would get either $\underline{d} \geq 0.19$ which would be better, or $\underline{d} \leq 0.1862$ which is false. For infinitely many N we thus have

$$\begin{aligned} \Delta_N &\geq \sum_{1 \leq n \leq N, r(n) > 0} (15 - \epsilon_N)^2 \geq \left(\sum_{1 \leq n \leq N, r(n) > 0} (15 - \epsilon)^2 + o(1)\right) N \\ &= \left(\frac{\underline{d}}{2} \left(15 - \frac{2}{\underline{d} \log 2}\right)^2 + o(1)\right) N. \end{aligned}$$

We deduce from these estimates the inequality

$$\frac{\underline{d}}{2} \left(15 - \frac{2}{\underline{d} \log 2}\right)^2 \leq \frac{2}{\log^2 2} \left(5.3636 - \frac{1}{\underline{d}}\right),$$

which may be written as

$$56.25 \log^2 2 \underline{d}^2 - (15 \log 2 + 5.3636) \underline{d} + 1 \leq 0.$$

The lower bound $\underline{d} \geq 0.1866$ then follows.

3. THE UPPER BOUND

A. Basic ideas.

Let us introduce further notations. Let M be a positive odd integer and let ω denote the order of 2 in $(\mathbb{Z}/M\mathbb{Z})^*$. For \overline{m} a residue class modulo M , put

$$f_M(\overline{m}) = \{\overline{k} \in \mathbb{Z}/\omega\mathbb{Z} : \overline{m} - 2^{\overline{k}} \in (\mathbb{Z}/M\mathbb{Z})^*\},$$

and

$$\delta_M(\nu) = |\{\overline{m} \in \mathbb{Z}/M\mathbb{Z} : |f_M(\overline{m})| = \nu\}|.$$

The basic tool to get an upper bound for \overline{d} is the following lemma.

Lemma 2. *With the previous notations, we have*

$$\overline{d} \leq \sum_{\nu=0}^{\omega} \delta_M(\nu) \min\left(\frac{1}{M}, \frac{2\nu}{\omega\varphi(M)\log 2}\right),$$

where φ denotes Euler's function.

Proof. Let \overline{m} be a congruence class modulo M , with $|f_M(\overline{m})| = \nu$. Let us study the proportion of odd integers congruent to \overline{m} that may be written in the form $p + 2^k$. This proportion is clearly at most $1/M$, and we only need to prove the alternative upper bound.

Since all the primes but a finite number are invertible modulo M , there almost surely exist ν congruence equations $\overline{m} = \overline{p}_i + 2^{\overline{k}_i}$, $i \in \{1, \dots, \nu\}$, such that any representation $p + 2^k$ comes from one of these congruence equations. The number of primes up to N which are congruent to p_i modulo M is equivalent to $N/(\varphi(M)\log N)$, while the number of powers of 2 which are congruent to $2^{\overline{k}_i}$ modulo M is equivalent to $\log N/(\omega \log 2)$. Thus the number of integers congruent to \overline{m} that may be written in the form $p + 2^k$ is at most $(\nu/(\varphi(M)\omega \log 2) + o(1))N$. This implies that the proportion of odd integers enjoying these properties is at most $2\nu/(\varphi(M)\omega \log 2)$ and the lemma follows.

□

This lemma provides a non trivial upper bound for \overline{d} as soon as there exist residue classes \overline{m} modulo M such that

$$f_M(\overline{m}) < \frac{\omega\varphi(M)\log 2}{2M}, \quad (1)$$

a condition that occur for a small number of classes. The main problem is to compute the distribution of the $f_M(\overline{m})$'s in an efficient way. The direct computation of all the $f_M(\overline{m})$'s is quickly limited by memory problems. However one can obtain significant results this way.

Take $M = 23205 = (2^{24} - 1)/723$, so that $\omega = 24$ and $\varphi(M) = 9216$. The condition (1) is equivalent to $f_M(\overline{m}) \leq 3$. We find

$$(\delta_M(0), \delta_M(1), \delta_M(2), \delta_M(3)) = (0, 48, 720, 320),$$

and we get this way $\overline{d} \leq 0.985049$.

B. Refined algorithms and results.

It appears that the function f_M takes a very few possible values, when compared to the subset set of $\mathbb{Z}/\omega\mathbb{Z}$. So let us introduce

$$g_M(I) = \{\bar{m} \in \mathbb{Z}/M\mathbb{Z} : f_M(\bar{m}) = I\} \quad \text{and} \quad G_M(I) = |g_M(I)|,$$

for $I \subset \mathbb{Z}/\omega\mathbb{Z}$. Note that

$$\delta_M(\nu) = \sum_{|I|=\nu} G_M(I).$$

So it is sufficient to know the distribution of the $G_M(I)$'s to compute an upper bound for \bar{d} .

The main advantage of the function g_M is that it is easily computable by induction on the number of prime factors of M . The initial case is given by $g_0(\{0\}) = \{0\}$.

Let M_1, M_2 be two positive odd squarefree integers, with $M_2 = pM_1$ for some prime p not dividing M_1 . Let ω_1, ω_2 and ω_p denote the order of 2 in $(\mathbb{Z}/M_1\mathbb{Z})^*$, $(\mathbb{Z}/M_2\mathbb{Z})^*$ and $(\mathbb{Z}/p\mathbb{Z})^*$, respectively. The image of f_p is easy to compute. There is the subset $I_{p,0} = \{\bar{2}^{\bar{k}} \in (\mathbb{Z}/p\mathbb{Z})^* : \bar{k} \in \mathbb{Z}/\omega_p\mathbb{Z}\}$ with $G_p(I_{p,0}) = p - \omega_p$, for each $\bar{j} \in \mathbb{Z}/\omega_p\mathbb{Z}$ the subset $I_{p,\bar{j}} = \{\bar{2}^{\bar{k}} \in (\mathbb{Z}/p\mathbb{Z})^* : \bar{k} \in \mathbb{Z}/\omega_p\mathbb{Z}, \bar{k} \neq \bar{j}\}$ with $G_p(I_{p,\bar{j}}) = 1$. Now, let I_2 and I_p be in the image of f_{M_2} and f_p respectively. Denote by \tilde{I}_2 and \tilde{I}_p the subsets of $\mathbb{Z}/M_1\mathbb{Z}$ which are inverse image of I_2 and I_p by the map on subsets induced by the natural surjections $\mathbb{Z}/M_1\mathbb{Z} \rightarrow \mathbb{Z}/M_2\mathbb{Z}$ and $\mathbb{Z}/M_1\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z}$ respectively. Then it is easy to see that $\tilde{I}_2 \cap \tilde{I}_p$ is in the image of f_{M_1} with $G_{M_1}(\tilde{I}_2 \cap \tilde{I}_p) = G_{M_2}(I_2)G_p(I_p)$, and that all subsets in the image of f_{M_1} are obtained in this way.

This constructions allows to build recursively the image of f_M . It also enables us to know how many classes have the same image. Therefore, one can compute $G_M(I)$ without knowing $g_M(I)$.

Let us give an example. For $M = 5592405 = 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 241 = (2^{24} - 1)/3$, we have $\omega = 24$. There are 16401 subsets in the image of f_M , which is much fewer than 2^{24} . Each of these subsets is obtained in r ways, with $1 \leq r \leq 250068$. Only subsets of cardinality at most 3 lead to an improved upper bound. The empty set appears 48 times. Each of the singletons from $\mathbb{Z}/24\mathbb{Z}$ appears 540 times. For 2-subsets, the situation is slightly more complicated to describe. The subsets of the form $\{a, a \pm 8\}$ appear 3625 times (there are 24 of them) while those of the form $\{a, a + 12\}$ appear 7170 times (there are 12 of them). There are 224 interesting 3-subsets, appearing 3, 6, 225 or 9520 times.

This method requires much less memory than the algorithm from the previous subsection. It is still possible to save a bit more memory. Indeed the representation problem (by an invertible plus a power of 2) is invariant when multiplied by a power of 2. So we can use a representative of a collection of subsets, each of them being obtained by translation from the representative, instead of subsets of $\mathbb{Z}/\omega\mathbb{Z}$.

The best result found so far is given by $M = 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 31 \cdot 41 \cdot 73 \cdot 241 \cdot 257$. It leads to the improvement

$$\bar{d} < 0.9818818607968211912960156368,$$

and the upper bound from Theorem 1 follows. This computation took 35 minutes on an Intel Xeon 2.4GHz with a memory stack of 2.1Go. Indeed, the real limitation

is the memory. Note that during the computations, subsets for which $G_M(I)$ was quite large and thus unlikely to contribute in the density were dropped (still there were a total of 4469837 different subsets at the end). Hence the density obtained may be a little bigger than the actual density for this value of M .

REFERENCES

1. P. Erdős, *On integers of the form $2^k + p$ and some related problems*, Summa Brasil. Math. **2** (1950), 113–123.
2. D. R. Heath-Brown and J.-C. Puchta, *Integers represented as a sum of primes and powers of two*, Asian J. Math. **6** (2002), 535–565.
3. J. Pintz, personal communication (2002).
4. J. Pintz and I. Z. Ruzsa, *On Linnik's approximation to Goldbach's problem, I*, Acta Arith. **109.2** (2003), 169–194.
5. N. P. Romanov, *Über einige Sätze der additiven Zahlentheorie*, Math. Ann. **109** (1934), 668–678.

IGD UMR 5028 MATHÉMATIQUES, UNIVERSITÉ CLAUDE BERNARD LYON 1, 43 BOULEVARD
DU 11 NOVEMBRE 1918, 69622 VILLEURBANNE CEDEX, FRANCE
E-mail address: Laurent.Habsieger@igd.univ-lyon1.fr, roblot@igd.univ-lyon1.fr