THE DENSITY OF MINIMAL DIVIDING ODD SUBSETS FOR THE EVEN NUMBERS IS ASYMPTOTICALLY NORMAL

MATHIS ANTONETTI

Abstract. In this notice, we introduce the problem of minimal dividing odd subsets for the even numbers and we show that the density of such subsets of n elements is asymptotically normal (that is at least decreasing as $\frac{1}{n}$). We argue that understanding the problem of minimal dividing odd subset might lead to new approaches to solving NP-hard problems.

1. Introduction

The object of study of this paper are the minimal dividing odd subsets for the even numbers, i.e. the subsets E of $2\mathbb{N}+1 = \{1, 3, 5, ...\}$ such that the binary composition $E + E = \{a + b \mid a, b \in E\}$ contains $2[\![1, m]\!] = \{2, 4, ..., 2m\}$ with m as large as possible. For example, under the Goldbach conjecture [Feliksiak(2021)] , it is clear that $\{1, p_2, ..., p_n\}$ is an odd dividing subset for the even numbers but of course, it is not minimal.

More precisely, we define

$$m(E) = \max\{m \mid 2[\![1, m]\!] \subset E + E\}, \tag{1}$$

and for any $n \in \mathbb{N}+1$,

$$E_n = \underset{E \subset 2\mathbb{N}+1, \mathrm{Card}(E)=n}{\operatorname{argmax}} m(E). \tag{2}$$

Then by definition, E_n contains all the subsets E of at most n elements such that $E + E$ contains $[\![1, m]\!]$ with m as large as possible. In the sequel, we are interested in $m(E_n) = \max_{E \in E_n} m(E) = \min_{E \in E_n} m(E)$ and more precisely in $d(n) = \frac{n}{m(E_n)}$. In fact, $d(n)$ is the density of odd numbers necessary to retrieve the even numbers up to $2m(E_n)$. That is why d is an interesting function to study.

2. Main result

Theorem 2.1. *Let $n \in \mathbb{N}+1$, we have*

$$d(n) \leq \frac{n}{2n(p(n)+1) - 2p(n)(2p(n)+1) - 1}, \tag{3}$$

where $p(n) = \begin{cases} \frac{n}{4} \text{ if } 4 \mid n \\ \lfloor \frac{n-1}{4} \rfloor \text{ otherwise} \end{cases}$.

From this result, we deduce immediately the following corollary.

Date: April 2024.

Corollary 2.2. *We have* $d(n) = O(\frac{1}{n})$ *when* $n \to +\infty$.

In other words, the density of minimal dividing odd subsets for the even numbers is asymptotically normal. To prove this result, we need the following lemmas :

Lemma 2.3. *Let* $p \in \mathbb{N}$ *and* $n \in \mathbb{N} + 2p + 1$. *Define* $(u_k(p,n))_{1 \leq k \leq n}$ *by induction as follows*

(4)
$$u_k(p,n) = \begin{cases} u_{k-1}(p,n) + 2 \text{ if } k \in [\![2, p+1]\!] \cup [\![n-p+1, n]\!] \\ u_{k-1}(p,n) + 2(p+1) \text{ otherwise} \end{cases}, \quad u_1(p,n) = 1.$$

We have $2[\![1, u_n(p,n)]\!] = E_{n,p} + E_{n,p}$ *where* $E_{n,p} = \{u_k(p,n) \mid k \in [\![1,n]\!]\}$.

Lemma 2.4. *We have* $E_{n,p} + E_{n,p} \subset E_{n+1,p} + E_{n+1,p}$.

Proof. Lemma 2.4. Clearly, we have $u_{n-p}(p, n+1) = u_{n-p}(p,n)$ (since n only matters for the terms $u_k(p,n)$ with $k \geq n-p+1$). More generally, the following relation holds

(5) $$\forall i \leq n-p, \quad u_i(p, n+1) = u_i(p,n).$$

Thus we have

(6) $$\forall r, q \in [\![1, n-p]\!], \quad u_q(p,n) + u_r(p,n) = u_q(p,n+1) + u_r(p,n+1),$$

and

(7) $$\forall k \in [\![1, p+1]\!], u_{n-p+k}(p, n+1) = u_n(p,n) + 2k.$$

Let $k \in [\![1,p]\!]$ and $h \in [\![1,n]\!]$. According to (6), it suffices to prove the following result

$$\exists a, b \in [\![1, n+1]\!], \quad u_{n-p+k}(p,n) + u_h(p,n) = u_a(p,n+1) + u_b(p,n+1).$$

<u>If $p+2 \leq h \leq n-p$:</u> We obtain with (5),

$$u_{n-p+k}(p,n) + u_h(p,n) = (u_n(p,n) + 2(k+1)) + (u_h(p,n) - 2(p+1)) = u_{n-p+k}(p,n+1) + u_{h-1}(p,n+1).$$

<u>If $n-p+1 \leq h \leq n$:</u> If $n-p+1 \leq h+k \leq n+1$, we have

$$\begin{aligned} u_{n-p+k}(p,n) + u_h(p,n) &= (u_n(p,n) + 2(h+k-(n-p))) + (u_h(p,n) - 2(h-(n-p+1)) - 2(p+1)) \\ &= u_{h+k}(p,n+1) + u_{n-m-1}(p,n+1). \end{aligned}$$

Otherwise, if $n+2 \leq h+k \leq n+p$, we obtain

$$\begin{aligned} u_{n-p+k}(p,n) + u_h(p,n) &= (u_n(p,n) + 2(h+k-(n+1))) + (u_h(p,n) - 2(h-(n-p+1))) \\ &= u_{h+k-(p+1)}(p,n+1) + u_{n-m}(p,n+1). \end{aligned}$$

<u>If $1 \leq h \leq p+1$:</u> If $1 \leq h+k \leq p+1$, we have

$$\begin{aligned} u_{n-p+k}(p,n) + u_h(p,n) &= (u_n(p,n) - 2p) + (u_h(p,n) + 2k) \\ &= u_{n-p}(p,n+1) + u_{k+h}(p,n+1). \end{aligned}$$

Otherwise, if $p+2 \leq h+k \leq 2p+1$, we obtain

$$\begin{aligned} u_{n-p+k}(p,n) + u_h(p,n) &= (u_n(p,n) + 2(k+h-(p+1))) + (u_h(p,n) - 2(h-1)) \\ &= u_{n-2p+k+h-1}(p,n+1) + u_1(p,n+1). \end{aligned}$$

□

Proof. Lemma 2.3. We proceed by induction over n. The initial case $n = 2p+1$ is obvious since we have $u_k(p, 2p+1) = u_{k-1}(p, 2p+1) + 2$ for all $k \leq 2p+1$ so that for all $q \leq 2p$, we have

$$\begin{aligned} 4q &= (2q-1) + (2q+1) = u_q(p, 2p+1) + u_{q+1}(p, 2p+1), \\ 4q+2 &= 2(2q+1) = 2u_{q+1}(p, 2p+1). \end{aligned}$$

Thus $2[\![1, u_{2p+1}(p, 2p+1)]\!] = 2[\![1, 4p+1]\!] = E_{2p+1,p} + E_{2p+1,p}$.

Now, assume that $2[\![1, u_n(p, n)]\!] = E_{n,p} + E_{n,p}$. Then using Lemma 2.4, we obtain $2[\![1, u_n(p, n)]\!] \subset E_{n+1,p} + E_{n+1,p}$. Moreover, using (7), one obtains the following property

$$\begin{aligned} \forall j, k \in [\![1, p+1]\!], \quad u_{n-p+j}(p, n+1) + u_{n-p+k}(p, n+1) &= (u_n(p, n) + 2j) + (u_n(p, n) + 2k) \\ &= 2(u_n(p, n) + j + k), \end{aligned}$$

and $u_{n+1}(p, n+1) = u_n(p, n) + 2(p+1)$.

Hence $2[\![1, u_{n+1}(p, n+1)]\!] \setminus \{2u_n(p, n) + 2\} \subset E_{n+1,p} + E_{n+1,p}$. Finally, since

$$2u_n(p,n)+2 = (u_n(p,n)-2p)+(u_n(p,n)+2(p+1)) = u_{n-p}(p,n+1)+u_{n+1}(p,n+1),$$

we actually have $2[\![1, u_{n+1}(p, n+1)]\!] \subset E_{n+1,p}+E_{n+1,p}$. Since the elements of $E_{n+1,p}$ are odd, the elements of $E_{n+1,p} + E_{n+1,p}$ are even and $\max(E_{n+1,p} + E_{n+1,p}) = 2\max(E_{n+1,p}) = 2u_{n+1}(p, n+1)$. The result follows. □

Proof. Theorem 2.1. We clearly have $n \geq 2p(n)+1$ so using Lemma 2.3, we obtain $m(E_{n,p(n)}) \leq m(E_n)$ according to the definition (2) of E_n. Thus

$$d(n) \leq \frac{n}{m(E_{n,p(n)})} = \frac{n}{2n(p(n)+1) - 2p(n)(2p(n)+1) - 1}.$$

□

Proof. Corollary 2.2. We have

$$2n(p(n)+1) - 2p(n)(2p(n)+1) - 1 = 2(p(n)+1)(n - 2p(n)) + 2p(n) - 1 \underset{n\to+\infty}{\sim} \frac{n^2}{4}.$$

Thus using (3), we obtain

$$\limsup_{n\to+\infty} nd(n) \leq 4.$$

In particular, we have $d(n) = O(\frac{1}{n})$. □

3. An exhaustive search algorithm

To find $d(n)$, we can compute E_n by using $F_{n+1} \backslash F_n = \{2(m(F_n)+1)-k \mid k \in F_n\}$ for all candidate F_n. The resulting algorithm is a time-efficient exhaustive search.

Algorithm 1 Exhaustive search of E_n

```
E_n ← {{1}}
C ← {{1}}
S ← {1}
N ← {∅}
for t = 1 to n do
    C̃, S̃, Ñ ← ∅, ∅, ∅
    for i = 1 to Card(C) do
        for j = 1 to Card(C_i) do
            if 2S_i + 2 − (C_i)_j > (C_i)_{Card(C_i)} then
                C̃ ← C̃ ∪ {C_i ∪ {2S_i + 2 − (C_i)_j}}
                N̂ ← {(C_i)_k + 2S_i + 2 − (C_i)_j | k ∈ ⟦j + 1, Card(C_i)⟧} ∪ {2(2S_i + 2 − (C_i)_j)}
                for k = 1 to 1 + S_i − (C_i)_j do
                    J ← 2(S_i + 1 + k)
                    if J ∉ N_i then
                        if J ∈ N̂ then
                            N̂ ← N̂ \ J
                        else
                            S̃ ← S̃ ∪ {−1 + ⌊J/2⌋}
                            Ñ ← N_i ∪ N̂
                            break
                        end if
                    end if
                    if k = 1 + S_i − (C_i)_j then
                        S̃ ← S̃ ∪ {⌊J/2⌋}
                        Ñ ← N_i ∪ N̂
                    end if
                end for
            end if
        end for
    end for
    C, S, N ← C̃, S̃, Ñ
    E_n ← {C_i | i ∈ argmax(S)}
end for
```

The complexity of such algorithm is $O(n!)$ because $\mathrm{Card}(C) = O(n!)$ at the last step of the first for loop. This suggest that the minimal dividing subset problem is actually NP-hard.

4. Experiments

Using $F_{n+1} \setminus F_n = \{2(m(F_n)+1) - k \mid k \in F_n\}$ for all $F_{n+1} \in E_{n+1}, F_n \in E_n$, it is easy to implement an efficient exhaustive search to get $m(E_n)$ and $d(n)$. With this implementation in Python, we obtained the following figure.

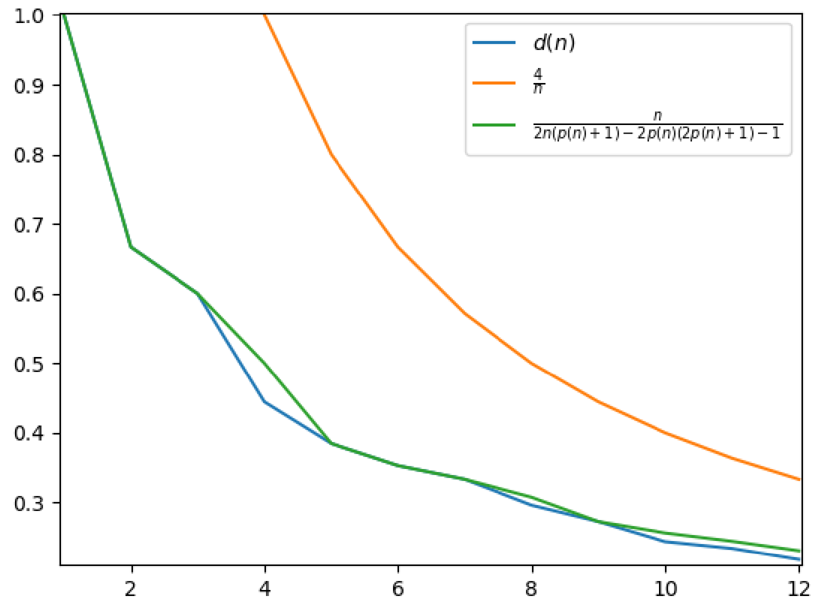

Figure 1. Comparison of $d(n)$, $\frac{n}{2np(n)-2p(n)(2p(n)+1)-1}$ and $\frac{4}{n}$ for $n = 1, ..., 12$.

We can observe that our inequality dictates almost perfectly the behavior of $d(n)$ for small n. Since the complexity of searching such E_n is at least exponential, we cannot go much further than $n = 12$ in practice.

5. Conclusion

We have introduced the concept of dividing odd subset for the even numbers and we studied its properties. In particular, we have shown that the density $d(n)$ of minimal such is asymptotically normal by deriving an inequality that seems to accurately describe the behavior of $d(n)$. This problem seems to be NP-hard depending on n since the complexity of the natural exhaustive search algorithm derived in section 3 is worse than exponential. This could be an interesting avenue toward solving more efficiently NP-hard problems [Michael Garey(1979)].

References

[Feliksiak(2021)] Jan Feliksiak. 2021. The Binary Goldbach Conjecture. *ScienceOpen Preprints* (2021). https://doi.org/10.14293/S2199-1006.1.SOR-.PPIU46T.v1

[Michael Garey(1979)] David Johnson Michael Garey. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness.*

Email address: `mathis.antonetti@gmail.com`