ON THE MINIMUM OVERLAP PROBLEM

T. AGAMA

Abstract. In this note we study the minimum overlap problem. We obtain the following crude inequality for the problem

$$M(n) < \mathcal{D}(k)(1-o(1))\frac{n}{4}$$

where $\mathcal{D}(k) > 1$.

1. **Introduction and problem statement**

The minimum overlap problem was first posed by then then Hungarian mathematician Paul Erdős. The problem is often stated in the following way:
Let $A = \{a_i\}$ and $B = \{b_j\}$ be any two complementary subsets, a splitting of the set $\{1, 2 \ldots, n\}$ such that $|A| = |B| = \frac{n}{2}$. Let M_k denotes the number of solutions to the equation $a_i - b_j = k$, where $-n \leq k \leq n$. Let us denote by $M(n) := \min_{A,B} \max_k M_k$. Then the problem asks for an estimate for $M(n)$ for sufficiently large values of n. There has been significant progress in estimating from below and above the quantity $M(n)$. Erdős [1] managed to obtain the following upper and lower bounds

$$M(n) < (1+o(1))\frac{n}{2} \quad \textbf{and} \quad M(n) > \frac{n}{4}.$$

The lower bound was improved to (see [2])

$$M(n) > (1-2^{-\frac{1}{2}})n$$

and latter to (see [2])

$$M(n) > \sqrt{(4-\sqrt{15})}(n-1)$$

the most recent of which is [2]

$$M(n) > \sqrt{(4-\sqrt{15})}n.$$

The upper bound, to the contrary, developed quite steadily overtime in the aftermath of Erdős's result (see [1])

$$M(n) < (1+o(1))\frac{2n}{5},$$

a result due to Motzkin, Ralston and Selfridge. The best known upper bound concerning this problem is due to Haugland [3], given by

$$M(n) < (1+o(1))0.38093n.$$

Date: August 15, 2020.
2000 *Mathematics Subject Classification.* Primary 54C40, 14E20; Secondary 46E25, 20C20.
Key words and phrases. minimum; overlap.

In this note we obtain the following crude upper bound to the problem

Theorem 1.1. *Let $A = \{a_i\}$ and $B = \{b_j\}$ be any two complementary subsets, a splitting of the set $\{1, 2 \dots, n\}$ such that $|A| = |B| = \frac{n}{2}$. Let M_k denotes the number of solutions to the equation $a_i - b_j = k$, where $-n \leq k \leq n$. Let us denote by $M(n) := \min_{A,B} \max_k M_k$, then for a fixed k we have the inequality*

$$M(n) < \mathcal{D}(k)(1 - o(1))\frac{n}{4}$$

where $\mathcal{D}(k) > 1$.

2. Preliminary result

Theorem 2.1. *Let $\{r_j\}_{j=1}^{n}$ and $\{h_j\}_{j=1}^{n}$ be any sequence of real numbers, and let r and h be any real numbers satisfying $\sum\limits_{j=1}^{n} r_j = r$ and $\sum\limits_{j=1}^{n} h_j = h$, and*

$$(r^2 + h^2)^{1/2} = \sum_{j=1}^{n} (r_j^2 + h_j^2)^{1/2},$$

then

$$\sum_{j=2}^{n} r_j h_j = \sum_{j=2}^{n} h_j \left(\sum_{i=1}^{j} r_i + \sum_{i=1}^{j-1} r_i \right) - 2 \sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}.$$

Proof. Consider a right angled triangle, say ΔABC in a plane, with height h and base r. Next, let us partition the height of the triangle into n parts, not necessarily equal. Now, we link those partitions along the height to the hypotenuse, with the aid of a parallel line. At the point of contact of each line to the hypotenuse, we drop down a vertical line to the next line connecting the last point of the previous partition, thereby forming another right-angled triangle, say $\Delta A_1 B_1 C_1$ with base and height r_1 and h_1 respectively. We remark that this triangle is covered by the triangle ΔABC, with hypotenuse constituting a proportion of the hypotenuse of triangle ΔABC. We continue this process until we obtain n right-angled triangles $\Delta A_j B_j C_j$, each with base and height r_j and h_j for $j = 1, 2, \dots n$. This construction satisfies

$$h = \sum_{j=1}^{n} h_j \text{ and } r = \sum_{j=1}^{n} r_j$$

and

$$(r^2 + h^2)^{1/2} = \sum_{j=1}^{n} (r_j^2 + h_j^2)^{1/2}.$$

Now, let us deform the original triangle ΔABC by removing the smaller triangles $\Delta A_j B_j C_j$ for $j = 1, 2, \dots n$. Essentially we are left with rectangles and squares piled on each other with each end poking out a bit further than the one just above, and we observe that the total area of this portrait is given by the relation

$$\begin{aligned}\mathcal{A}_1 &= r_1h_2 + (r_1+r_2)h_3 + \cdots (r_1+r_2+\cdots+r_{n-2})h_{n-1} + (r_1+r_2+\cdots+r_{n-1})h_n\\ &= r_1(h_2+h_3+\cdots h_n) + r_2(h_3+h_4+\cdots+h_n) + \cdots + r_{n-2}(h_{n-1}+h_n) + r_{n-1}h_n\\ &= \sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}.\end{aligned}$$

On the other hand, we observe that the area of this portrait is the same as the difference of the area of triangle ΔABC and the sum of the areas of triangles $\Delta A_jB_jC_j$ for $j=1,2,\ldots,n$. That is

$$\mathcal{A}_1 = \frac{1}{2}rh - \frac{1}{2}\sum_{j=1}^{n} r_jh_j.$$

This completes the first part of the argument. For the second part, along the hypotenuse, let us construct small pieces of triangle, each of base and height (r_i,h_i) $(i=1,2\ldots,n)$ so that the trapezoid and the one triangle formed by partitioning becomes rectangles and squares. We observe also that this construction satisfies the relation

$$(r^2+h^2)^{1/2} = \sum_{i=1}^{n}(r_i^2+h_i^2)^{1/2},$$

Now, we compute the area of the triangle in two different ways. By direct strategy, we have that the area of the triangle, denoted $\mathcal{A}$, is given by

$$\mathcal{A} = 1/2\bigg(\sum_{i=1}^{n} r_i\bigg)\bigg(\sum_{i=1}^{n} h_i\bigg).$$

On the other hand, we compute the area of the triangle by computing the area of each trapezium and the one remaining triangle and sum them together. That is,

$$\mathcal{A} = h_n/2\bigg(\sum_{i=1}^{n} r_i + \sum_{i=1}^{n-1} r_i\bigg) + h_{n-1}/2\bigg(\sum_{i=1}^{n-1} r_i + \sum_{i=1}^{n-2} r_i\bigg) + \cdots + 1/2r_1h_1.$$

By comparing the area of the second argument, and linking this to the first argument, the result follows immediately. □

Corollary 2.2. Let $f:\mathbb{N}\longrightarrow\mathbb{C}$, then we have the decomposition

$$\sum_{n\leq x-1}\sum_{j\leq x-n} f(n)f(n+j) = \sum_{2\leq n\leq x} f(n)\sum_{m\leq n-1} f(m).$$

Proof. Let us take $f(j)=r_j=h_j$ in Theorem 2.1, then we denote by $\mathcal{G}$ the partial sums

$$\mathcal{G} = \sum_{j=1}^{n} f(j)$$

and we notice that

$$\sum_{j=1}^{n}\sqrt{(h_j^2+r_j^2)}=\sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}$$
$$=\sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}$$
$$=\sqrt{2}\sum_{j=1}^{n}f(j).$$

Since $\sqrt{(\mathcal{G}^2+\mathcal{G}^2)}=\mathcal{G}\sqrt{2}=\sqrt{2}\sum\limits_{j=1}^{n}f(j)$ our choice of sequence is valid and, therefore the decomposition is valid for any arithmetic function. □

Lemma 2.3. *(Area method) Let* $f:\mathbb{N}\longrightarrow\mathbb{C}$*. If*

$$\sum_{n\leq x}f(n)f(n+l_0)>0$$

then there exist some constant $1>\mathcal{C}(l_0)>0$ *such that*

$$\sum_{n\leq x}f(n)f(n+l_0)<\frac{1}{\mathcal{C}(l_0)x}\sum_{2\leq n\leq x}f(n)\sum_{m\leq n-1}f(m).$$

Proof. By Corollary 2.2, we obtain the identity by taking $f(j)=r_j=h_j$

$$\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j)=\sum_{2\leq n\leq x}f(n)\sum_{m\leq n-1}f(m).$$

Next we observe that

$$\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j)\gg\sum_{n\leq x}\sum_{j\leq x}f(n)f(n+j)$$
$$=\sum_{n\leq x}f(n)f(n+1)+\sum_{n\leq x}f(n)f(n+2)$$
$$+\cdots\sum_{n\leq x}f(n)f(n+l_0)+\cdots\sum_{n\leq x}f(n)f(n+x)$$
$$\geq|\mathcal{M}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)$$
$$+|\mathcal{N}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)$$
$$+\cdots+\sum_{n\leq x}f(n)f(n+l_0)+\cdots+|\mathcal{R}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)$$
$$=\bigg(|\mathcal{M}(l_0)|+|\mathcal{N}(l_0)|+\cdots+1$$
$$+\cdots+|\mathcal{R}(l_0)|\bigg)\sum_{n\leq x}f(n)f(n+l_0)$$
$$\geq\mathcal{C}(l_0)x\sum_{n<x}f(n)f(n+l_0).$$

where $\min\{\mathcal{M}(l_0)|, |\mathcal{N}(l_0)| \ldots, |\mathcal{R}(l_0)|\} = \mathcal{C}(l_0)$. By inverting this inequality, the result follows immediately. $\square$

3. Main result

We begin this section by introducing an arithmetic function on particular sets of integers.

Definition 3.1. Let $A = \{a_i\}$ and $B = \{b_j\}$ be any two complementary subsets, a splitting of the set $\{1, 2 \ldots, n\}$ such that $|A| = |B| = \frac{n}{2}$. Then we consider the following arithmetic function

$$\vee(c_i) = \begin{cases} 1 & \textbf{if} \quad c \in A \cup B \\ 0 & \textbf{otherwise}. \end{cases}$$

Lemma 3.2. *Let $A = \{a_i\}$ and $B = \{b_j\}$ be any two complementary subsets, a splitting of the set $\{1, 2 \ldots, n\}$ such that $|A| = |B| = \frac{n}{2}$, then we have*

$$\sum_{1 \leq i \leq n} \vee(a_i) = \frac{n}{2}$$

and

$$\sum_{1 \leq j \leq n} \vee(b_j) = \frac{n}{2}.$$

Proof. This is an easy consequence of the size of $|A \cup B| = n$ and the size of each complementary subset. $\square$

Theorem 3.3. *Let $A = \{a_i\}$ and $B = \{b_j\}$ be any two complementary subsets, a splitting of the set $\{1, 2 \ldots, n\}$ such that $|A| = |B| = \frac{n}{2}$. Let M_k denotes the number of solutions to the equation $a_i - b_j = k$, where $-n \leq k \leq n$. Let us denote by $M(n) := \min_{A,B} \max_k M_k$, then for a fixed k we have the inequality*

$$M(n) < \mathcal{D}(k)(1 - o(1))\frac{n}{4}$$

where $\mathcal{D}(k) > 1$.

Proof. Let k be fixed with $-n \leq k \leq n$, then the underlying problem is to estimate the correlation

$$\sum_{1 \leq i \leq n} \vee(a_i) \vee (a_i + k).$$

Applying the area method, there exist some constant $1 > \mathcal{R}(k) > 0$ such that

$$\sum_{1 \leq i \leq n} \vee(a_i) \vee (a_i + k) < \frac{1}{\mathcal{R}(k)2n} \sum_{2 \leq i \leq n} \vee(a_i) \sum_{s \leq i-1} \vee(a_s).$$

Applying partial summations on the right-hand side of the inequality, we have the following

$$\begin{aligned}\sum_{2\leq i\leq n}\vee(a_i)\sum_{s\leq i-1}\vee(a_s) &\leq \sum_{1\leq i\leq n}(i-1)\vee(a_i)\\ &= \sum_{1\leq i\leq n} i\vee(a_i)-\sum_{1\leq i\leq n}\vee(a_i)\\ &= n\sum_{1\leq i\leq n}\vee(a_i)-\int\limits_{i=1}^{n}\sum_{1\leq i\leq k}\vee(a_i)dk-\frac{n}{2}\\ &\leq \frac{n^2}{2}-\frac{n}{2}.\end{aligned}$$

It follows that

$$\sum_{1\leq i\leq n}\vee(a_i)\vee(a_i+k)<\frac{1}{\mathcal{R}(k)2n}\left(\frac{n^2}{2}-\frac{n}{2}\right)$$

and the claim upper bound follows, where $0<\mathcal{R}(k)<1$. □

[1].

References

1. Finch, Steven, *Erdös' minimum overlap problem*, Citeseer, 2004.
2. Guy, Richard, *Unsolved problems in number theory*, Springer Science & Business Media, vol.1, Taylor & Francis, 2004.
3. Haugland, Jan Kristian, *The minimum overlap problem revisited*, arXiv preprint arXiv:1609.08000, 2016.

Department of Mathematics, African Institute for Mathematical sciences
E-mail address: theophilus@aims.edu.gh/ emperordagama@yahoo.com

[1] .