A COMBINATORIAL "INTERMEDIATE VALUE" LEMMA

E. KOWALSKI

We consider the following problem, which I raised around 2001 to L. Habsieger (with a vague motivation coming from combinatorial tricks that could be played with large-sieve inequalities for automorphic forms, if certain conditions could be met):

Problem 1. *Show that, for any integer $m \geqslant 1$, there exists an integer $N \geqslant 1$ with the following property: for any map*

$$\{1, \ldots, N\} \rightarrow \{1, \ldots, m\}$$

there exists a subset $I \subset \{1, \ldots, N\}$ such that

$$\sum_{i \in I} f(i) = N.$$

From the consideration of constant functions $f(i) = k$ for all i, it is seen that any integer N with this property must be divisible by all integers $1 \leqslant k \leqslant m$, and therefore

$$N \geqslant N_0 = \operatorname{lcm}(1, \ldots, m)$$

(which by the Prime Number Theorem, grows like $\exp(m + o(m))$). It turns out that, conversely, this integer N_0 has the stated property. This was proved by L. Habsieger very quickly after I had asked the question.

Lemma 2 (Habsieger)**.** *For any $m \geqslant 1$, the integer $N = \operatorname{lcm}(1, \ldots, m)$ has the property described in the statement of Problem 1, except possibly for $m = 5$ and $m = 6$, where $N = 5!$ and $N = 2 \cdot 5!$, respectively, have this property.*

Proof. Let N be an integral multiple of the lcm of $1, \ldots, m$. First, we fix some notation: for a map f as above, we let

$$\int f = \sum_{1 \leqslant i \leqslant N} f(i), \quad \int_I f = \sum_{i \in I} f(i), \quad f_i = |f^{-1}(\{i\})|,$$

so that, in particular, we have

$$f_1 + f_2 + \cdots + f_m = N \tag{1}$$

$$\int f = \sum_j j f_j \geqslant N. \tag{2}$$

The proof is, essentially, separated in two cases, corresponding either to situations where f assumes a lot of "large" values, or to cases when f assumes often the value 1.

<u>First case.</u> Unless f is the constant function 1 (for which no work is required), we have

$$\int f = N + M > m(m-1) \tag{3}$$

where

$$M = f_2 + 2f_3 + \cdots + (m-1)f_m.$$

(note that $N \geqslant m(m-1)$ since $m \mid N$, $m-1 \mid N$). Assume first that

$$2M < N, \tag{4}$$

(which means that f takes mostly small values).

Comparing with (1), we find that

$$\sum_{j \geqslant 1} (j-1) f_j < M$$

and hence

$$\sum_{j \geqslant 2} j f_j < 2M, \text{ since } j \leqslant 2(j-1) \text{ for } j \geqslant 2,$$

and consequently, there are at most M values of $k \leqslant N$ such that $f(k) \geqslant 2$. Since we have assumed $2M < N$, letting

$$J = \{i \mid f(i) \geqslant 2\}$$

we get by the above that

$$2f_2 + \cdots + mf_m = \int_J f < 2M < N.$$

The number f_1 of values of k where $f(k) = 1$ is

$$N - (f_2 + f_3 + \cdots + f_m) \geqslant N - \int_J f$$

(by (1) again), hence adding to J a subset of size $N - \int_J f$ of $f^{-1}(1)$, we obtain a set I where the values of f sum to N.

Second case. We now consider what happens if (4) fails (so $M \geqslant N/2$); in fact, we assume the inequality

$$M > \frac{m^2}{4} + \frac{m(m-1)^2}{2}. \tag{5}$$

Let

$$i = \inf\{j \mid f_j \geqslant m-1\}, \tag{6}$$

(which exists because otherwise we would have

$$N \leqslant \int f < m(m-1),$$

which is impossible).

Now the idea is to write $N = i \times N/i$, i.e., to achieve the value N as i times N/i – this is where the assumption $i \mid N$ is used. To do this, we can produce the necessary number of i's either from the f_i instances of integers where $f(k) = i$, or by *grouping together* i points where the same value $j > i$ is attained. In other words, we exploit the fact that $i \cdot j = j \cdot i$.

For $j > i$, we say that a *j-block* (of f, which is fixed throughout) is a subset of $f^{-1}(\{j\})$ of order i. So, there exist $\lfloor f_j/i \rfloor$ disjoint j-blocks. Moreover, a subset $I \subset \{1, \ldots, N\}$ is called *admissible* if it is a disjoint union of j-blocks with $j > i$ (where j may not be the same for all blocks). For such a set, the values of $\int_I f$ satisfies

$$\int_I f = \lambda i, \text{ with } \lambda = \sum_{j > i} j \lambda_j$$

where, for each $j > i$, the number

$$\lambda_j \leqslant \left\lfloor \frac{f_j}{i} \right\rfloor$$

counts the disjoint j-blocks contained in I. Let

$$\mu = \sum_{j>i} j \left\lfloor \frac{f_j}{i} \right\rfloor,$$

so that $\int_I f \leqslant \mu i$ if I is admissible.

Now, we claim that (because of (5)), we have

$$i(f_i + \mu) \geqslant N. \tag{7}$$

If this fact is taken for granted, we now select an admissible set I such that $\int_I f = \lambda i$ with $\lambda \leqslant N/i$ as close as possible to N/i, say

$$N/i = \lambda + \delta, \quad \delta \geqslant 0.$$

Because of (7), we have

$$0 \leqslant \delta \leqslant f_i \ ; \tag{8}$$

indeed, we split in two cases: (1) if there is no j-block with $j > i$ disjoint from I, it must be the case that $\lambda = \mu$, and then $f_i \geqslant N/i - \mu$ because of (7); (2) if, on the other hand, there exists a j-block J with $j > i$ disjoint from I, we notice that $I' = I \cup J$ is admissible, and the choice of λ means that

$$\int_I f = (\lambda + j)i > N,$$

hence $\lambda + m \geqslant \lambda + j > N/i$ and so $\delta \leqslant m - 1 \leqslant f_i$ by construction.

The point is that, because of (8) we can add to I any subset of $f^{-1}(i)$ of order $N/i - \lambda \leqslant f_i$, obtaining a set I' such that

$$\int_{I'} f = i(N/i - \lambda) + i\lambda = N.$$

Now we check that (7) holds if (5) does, thereby completing this first case. We have

$$i(\mu + f_i) = if_i + \sum_{j>i} ij \left\lfloor \frac{f_j}{i} \right\rfloor \geqslant if_i + \sum_{j>i} (jf_j - i) \geqslant \sum_{j\geqslant i} jf_j - i(m-i).$$

Then, since we assumed that $\int f \geqslant N + M$, and since the definition (6) of i leads to

$$\sum_{j<i} jf_j < (m-1) \sum_{j\leqslant i} j = (m-1)\frac{i(i-1)}{2},$$

we have

$$\begin{aligned} i(f_i + \mu) &\geqslant N + M - i(m-i) - (m-1)i(i-1)/2 \\ &\geqslant N + (M - m^2/4 - m(m-1)^2/2) \geqslant N \end{aligned}$$

because of (5).

There remains to be seen when the two conditions (5) and (4) can be reconciled, or in other words when

$$\frac{N}{2} > \frac{m^2}{4} + \frac{m(m-1)^2}{2} \ ;$$

an easy computation shows that this is true for all $m \geqslant 7$ if $N = \text{lcm}(1, \ldots, m)$, and for $m = 5$, $m = 6$ if $N = 120$, 240, respectively. So the following last step completes the proof.

Third case. We show directly that the desired property holds with $N = \text{lcm}(1, \ldots, m)$ when $m \in \{2, 3, 4, 5, 6\}$ in turn, working with a putative counterexample f:

– For $m = 2$, $N = 2$, the result is clear (if $f_2 = 0$, then $f_1 = 2$).

– For $m = 3$, $N = 6$, we are done unless $f_3 = 0$ or $f_3 = 1$; if $f_3 = 0$, then we can use "induction", so we can assume $f_3 = 1$. Then we are done if $f_1 \geqslant 3$, or if $f_2 \geqslant 3$, and one of these holds.

– For $m = 4$, $N = 12$, we use induction if $f_4 = 0$, and count by 4's if $f_4 \geqslant 3$, so we can restrict to $f_4 = 1$ or 2. In the second case, we are done unless $f_2 < 2$ and $f_1 < 4$, but then $f_3 \geqslant 4$ clearly which again finishes the proof. In the first case $f_4 = 1$, say $f(k_1) = 4$; for any subset (not containing k_1) with 6 elements, we can get a subset I where f sums to 6, getting $I_1 = I \cup \{k_1\}$ where $\int_{I_1} f = 10$. The complement to I_1 still has 5 elements, where f takes values $\{1, 2, 3\}$, and no subset sums to 2: this is clearly impossible. $\square$

Remark 3. It is quite likely that $N = \text{lcm}(1, \ldots, m)$ is also suitable when $m = 5$ and $m = 6$ (note $N = 60$ for both values), but a nice, direct argument, is still missing...