THE CARDINALITY OF SOME SYMMETRIC DIFFERENCES

PO-YI HUANG, WEN-FONG KE, AND GÜNTER F. PILZ

(Communicated by Jim Haglund)

Abstract. In this paper, we prove that for positive integers $k$ and $n$, the cardinality of the symmetric differences of \{$1, 2, \ldots, k$\}, \{$2, 4, \ldots, 2k$\}, \{$3, 6, \ldots, 3k$\}, \ldots, \{$n, 2n, \ldots, kn$\} is at least $k$ or $n$, whichever is larger. This solved a problem raised by Pilz in which binary composition codes were studied.

1. Introduction

The symmetric difference of two sets $A$ and $B$, denoted by $A \Delta B$, is $(A \setminus B) \cup (B \setminus A)$. For any positive integers $k$ and $n$, the cardinality of the symmetric difference \{$1, 2, \ldots, k$\} $\Delta$ \{$2, 4, \ldots, 2k$\} $\Delta$ \cdots $\Delta$ \{$n, 2n, \ldots, kn$\} is of interest in several different situations. Here we mention three of them:

- “To love or not to love”. Let us take $k = 3$. Suppose three people with numbers 1, 2, and 3 on their back enter an empty room. Then three more people with numbers 2, 4, and 6 go into this room. Now two people have the same number, namely 2; they fall in love and leave the room. So only numbers 1, 3, 4 and 6 remain. Next, people with numbers 3, 6, and 9 come in. Numbers 3 and 6 find partners, and only the three people with numbers 1, 4, and 9 remain, and so on. The conjecture is: “There will always be at least 3 people in the room.” This is easy to show, but it seems considerably harder for a general $k$ greater than 3.

- “Summands in binary polynomials”. Over $\mathbb{Z}_2$, consider the sum of polynomials

\[
(1 + x + \cdots + x^k) \circ x + (1 + x + \cdots + x^k) \circ x^2 + \cdots + (1 + x + \cdots + x^k) \circ x^n,
\]

\[
= (1 + x + \cdots + x^k) + (1 + x^2 + \cdots + x^{2k}) + \cdots + (x^n + x^{2n} + \cdots + x^{kn}).
\]
Will there always be at least \( k \) summands present? Equivalently, has the symmetric difference
\[
\{1,2,\ldots,k\} \Delta \{2,4,\ldots,2k\} \Delta \cdots \Delta \{n,2n,\ldots,kn\}
\]
always at least \( k \) elements?

- “Codes by composition”. Encode a binary message \((a_1,a_2,\ldots,a_n)\) of length \( n \) as a polynomial by composition as
\[
a_1(1+x+\cdots+x^k) \circ x + a_2(1+x+\cdots+x^k) \circ x^2 + \cdots + a_n(1+x+\cdots+x^k) \circ x^n.
\]

There is a good reason to do this kind of coding; see [3]. A positive answer in item 2 would give a positive indication that the minimal weight of these codewords is \( k \).

With some extensive experimental data, the following conjecture, with a convenient name, was raised in [3].

**1-2-3 Conjecture.** The cardinality of the symmetric difference
\[
\{1,2,\ldots,k\} \Delta \{2,4,\ldots,2k\} \Delta \cdots \Delta \{n,2n,\ldots,kn\}
\]
is always at least \( k \).

It was shown in [3] that the conjecture holds true for \( k \leq 6 \). In private communications, E. Fried (Budapest) proved it for \( k = 7 \) and \( k = 8 \), and P. Fuchs (Linz) for \( k \geq 10^{12} \).

In this paper, we prove a slightly general version of the conjecture. Some notation can be useful for our discussion.

For \( k,s \in \mathbb{N} \), let \( I_k = \{1,2,\ldots,k\} \) and \( sI_k = \{s,2s,\ldots,ks\} \). For \( 1 \leq u < v \), put \( D_{k\times[u,v]} = uI_k \Delta (u+1)I_k \Delta \cdots \Delta vI_k \). When \( u = 1 \), we use \( D_{k\times[1,v]} \) instead of \( D_{k\times[1,u]} \) and denote by \( d_k(n) \) the cardinality of \( D_{k\times n} \). It is obvious that if \( 1 < s < n \), then \( D_{k\times n} = D_{k\times s} \Delta D_{k\times [s+1,n]} \). Also,
\[
(1:1) \quad D_{k\times n} = D_{n\times k} \quad \text{for all} \quad k \text{ and } n.
\]

Now, we modify the conjecture (but still keep the same name) and shorten it using the prepared notation.

**1-2-3 Conjecture.** For all \( k,n \in \mathbb{N} \), \( d_k(n) \geq \max\{n,k\} \).

It is easy to see that \( D_{k\times k} = \{1^2,2^2,\ldots,k^2\} \), which has \( k \) elements. The fact \([1:1]\) tells us that it suffices to show

**Restricted 1-2-3 Conjecture.** For all \( n > k \), it follows that \( d_k(n) \geq n \).

In this paper, we will show that this conjecture has a positive answer.

2. **The case when \( k < n \leq 2k \)**

Let \( k \) and \( w \) be fixed such that \( 1 \leq w \leq k \), and let \( n = k + w \). First, we make two general observations:

1. For any positive integer \( a \), the set \( aI_k \) contains at most \( \lfloor \sqrt{k} \rfloor \) many squares.

   To see this, we notice that the greatest common divisor \( u \), say, of the squares in \( aI_k \) is itself a square and is a multiple of \( a \). Hence \( u \in aI_k \). Therefore, the squares contained in \( aI_k \) are contained in \( \{u,4u,9u,\ldots,\lfloor \sqrt{k} \rfloor^2 u\} \), and so there are at most \( \lfloor \sqrt{k} \rfloor \) of them.
(2) Let $s$ and $t$ be distinct integers with $k < s \leq w$ and $k < t \leq w$. Then $st \not\in sI_k$ and $st \not\in tI_k$ since $st > sk$ and $st > tk$. Suppose that $\ell$ is the greatest common divisor of $s$ and $t$. Then $\ell \leq |t - s| \leq w - 1$, and

$$sI_k \cap tI_k \subseteq \{st/\ell, 2st/\ell, \ldots, (\ell - 1)st/\ell\}.$$ 

Therefore, $|sI_k \cap tI_k| \leq \ell - 1$, and the number of cancellations taking place in $sI_k$ and $tI_k$ is at most $2(\ell - 1)$.

Since $D_{k \times k} = \{1^2, 2^2, \ldots, k^2\}$, and the number of squares in $(k+1)I_k$ is less than or equal to $\sqrt{k}$, we have $d_k(k+1) \geq 2k - 2\sqrt{k}$. Also, since $k+1$ and $k+2$ are coprime to each other, $(k+1)I_k$ and $(k+2)I_k$ do not have anything in common. Thus $d_k(k+2) \geq 3k - 4\sqrt{k}$. Now, $k+2$ and $k+3$ are also coprime, and $(k+1)I_k$ and $(k+3)I_k$ have at most one element in common. We have $d_k(k+3) \geq 4k - 6\sqrt{k} - 2$. Finally, counting in the possible cancellations among $(k+1)I_k$, $(k+2)I_k$, $(k+3)I_k$, and $(k+4)I_k$, we have $d_k(k+4) \geq 5k - 8\sqrt{k} - 6$. Thus, for $w = 1, 2, 3, 4$, we want

$$\begin{align*}
2k - 2\sqrt{k} &\geq k + 1, \\
3k - 4\sqrt{k} &\geq k + 2, \\
4k - 6\sqrt{k} - 2 &\geq k + 3, \\
5k - 8\sqrt{k} - 6 &\geq k + 4.
\end{align*}$$

As long as $k \geq 9$, the above inequalities hold.

In the following, we assume that $k \geq 9$ and $w \geq 5$. For $2 \leq \ell \leq w - 1$, put

$$C_\ell = \{a \mid \ell \text{ divides } a, \text{ and } k + 1 \leq a \leq k + w\}.$$ 

Then $|C_\ell| \leq \lceil \frac{w}{\ell} \rceil$, and so the number of cancellations among the $aI_k$'s, $a \in C_\ell$, can be no more than

$$\left(\frac{\lceil \frac{w}{\ell} \rceil}{2}\right) \cdot 2(\ell - 1) = \left(\frac{w}{\ell}\right)(\lceil \frac{w}{\ell} \rceil - 1)(\ell - 1) \leq (\frac{w}{\ell} + 1)\frac{w}{\ell}(\ell - 1) < (\frac{w}{\ell} + 1)w.$$ 

Therefore, the total number of cancellations occurring in $(k+1)I_k$, $\ldots$, $(k+w)I_k$ is at most

$$\sum_{\ell=2}^{w-1} \left(\frac{\lceil \frac{w}{\ell} \rceil}{2}\right) \cdot 2(\ell - 1) < \sum_{\ell=2}^{w-1} (\frac{w}{\ell} + 1)w < \sum_{\ell=2}^{w-1} (\frac{w}{\ell} + \frac{w}{\ell})w = \sum_{\ell=2}^{w-1} \frac{2w^2}{\ell} = 2w^2 \sum_{\ell=2}^{w-1} \frac{1}{\ell} < 2w^2 \cdot \ln(w - 1).$$

After cancellations there are at least $kw - 2w^2\ln(w - 1)$ many elements left in $D_{k \times [k+1, k+w]}$.

Now, $D_{k \times (k+w)}$ has at least $2k$ elements as long as $kw - 2w^2 \cdot \ln(w - 1) \geq 3k$, or equivalently,

$$k \geq \frac{2w^2 \cdot \ln(w - 1)}{w - 3}.$$ 

Note that the function $f(x) = \frac{2x^2 \cdot \ln(x-1)}{x-3}$ is increasing for $x \geq 5$. For each given $k$, let $w_k = \max\{w \geq 1 \mid w \text{ satisfies } (2.1))$. Table 1 gives various $k$ and $w_k$. 

Table 1

<table>
<thead>
<tr>
<th>$k$</th>
<th>$w_k$</th>
<th>$k$</th>
<th>$w_k$</th>
<th>$k$</th>
<th>$w_k$</th>
<th>$k$</th>
<th>$w_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>5</td>
<td>100</td>
<td>15</td>
<td>1000</td>
<td>105</td>
<td>10000</td>
<td>753</td>
</tr>
<tr>
<td>50</td>
<td>8</td>
<td>200</td>
<td>27</td>
<td>2000</td>
<td>189</td>
<td>20000</td>
<td>1381</td>
</tr>
<tr>
<td>70</td>
<td>11</td>
<td>300</td>
<td>38</td>
<td>4000</td>
<td>341</td>
<td>40000</td>
<td>2548</td>
</tr>
<tr>
<td>90</td>
<td>13</td>
<td>600</td>
<td>68</td>
<td>8000</td>
<td>620</td>
<td>80000</td>
<td>4725</td>
</tr>
</tbody>
</table>

From Table 1 we know that, for example, if $k = 70$, then each of the sets $D_{k\times(k+s)}$, $5 \leq s \leq 11$, has at least $2k = 140$ elements in there. As another example, let $k = 40000$. Then each of the sets $D_{k\times(k+s)}$, $5 \leq s \leq 2548$, has at least 80000 elements in there.

We note that if a prime $p$ occurs in $\{k+1, \ldots, k+s\}$, $1 \leq s \leq k$, then $D_{k\times(k+s)}$ has at least $k$ elements in it as the elements of $p\mathbb{I}_k$ cannot be canceled. To see this, we assume that $ps = qt \in p\mathbb{I}_k \cap q\mathbb{I}_t$ for some integer $q > k$ with $q \not\equiv p$, and $s, t \in I_k$. Then $p$ divides $q$ since $t \leq k < p$. From $p > k$, we infer that $q > 2k$.

**Lemma 2.1.** Suppose that $w_k \geq 5$ and that there are two distinct primes among $k+1, \ldots, k+w_k$. Then $D_{k\times(k+s)}$ has at least $2k$ elements for all $s$ with $5 \leq s \leq k$.

**Proof.** If $5 \leq s \leq w_k$, then we have seen from $k \geq \frac{2^{s/2} \ln(s-1)}{3}$ that $D_{k\times(k+s)}$ has at least $2k$ elements. On the other hand, if $w_k < s \leq k$, then the two primes between $k+1$ and $k+w_k$ give us what we want. $\square$

Therefore, we assume that $w_k \geq 5$, and we would like to have two distinct primes among $k+1, k+2, \ldots, k+w_k$. This brings us to the prime gaps consideration.

A prime gap is the difference between two successive prime numbers. The $n$-th prime gap is the difference between the $(n+1)$-th and the $n$-th prime number. One writes $g(p)$ for the the gap $q-p$, where $q$ is the next prime to $p$. A prime gap is said to be maximal if it is larger than all gaps between smaller primes. The notation for the $n$-th maximal prime gap is $g_n$. Table 2 shows $g_n$ for $1 \leq n \leq 15$. For example, for any prime $p$ less than 9551, the prime gap $g(p)$ is less than 36. That is to say that for any prime $p$ with $p < 9551$, there must be a prime in the set $\{p, p+1, \ldots, p+35\}$.

Table 2

<table>
<thead>
<tr>
<th>$n$</th>
<th>$g_n$</th>
<th>$p_n$</th>
<th>$n$</th>
<th>$g_n$</th>
<th>$p_n$</th>
<th>$n$</th>
<th>$g_n$</th>
<th>$p_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>113</td>
<td>11</td>
<td>36</td>
<td>9551</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>18</td>
<td>523</td>
<td>12</td>
<td>44</td>
<td>15683</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>20</td>
<td>887</td>
<td>13</td>
<td>52</td>
<td>19609</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>23</td>
<td>9</td>
<td>22</td>
<td>1129</td>
<td>14</td>
<td>72</td>
<td>31397</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>89</td>
<td>10</td>
<td>34</td>
<td>1327</td>
<td>15</td>
<td>86</td>
<td>155921</td>
</tr>
</tbody>
</table>

If $p$ is a prime with $p > 2k$ and the maximal prime gap $g_m = g(q) < \frac{2k}{q}$, where $q$ is the first prime to have $g(q) = g_m$, then there must exist at least two primes among $k+1, k+2, \ldots, k+w_k$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Take $k = 300$; then $w_k = 38$. Now, 601 is the prime just bigger then $2k = 600$, and the maximal prime gap for primes less than 887 is at most $g_7 = 18$. Thus, $g(601) \leq 18 < \frac{600}{2}$. By Lemma 2.1, $D_{k \times (k+s)}$ contains at least $2k = 600$ elements for any $s$ with $5 \leq s \leq 300$.

For those $k$ which are less than 300, we can check easily using GAP [11] that $d_k(n) \geq n$ for $k < n \leq 2k$. Actually, a small program in GAP running on a modern PC takes about 90 seconds to verify it.

For those $k$ that are greater than 300, we will argue in the following that $d_k(n) \geq n$ for $k < n \leq 2k$ by using the monotone increasing property of $w_k$ and maximal prime gaps.

In [4, p. 368], one finds

$$\text{(2:2) for } n \geq 2, \quad p_n \geq n(\ln n + \ln \ln n - 1.0072629)$$

and

$$\text{(2:3) for } n \geq 7022, \quad p_n \leq n(\ln n + \ln \ln n - 0.9385).$$

Therefore, for $n \geq 7022$, we have

$$(n+1) \ln(n+1) - n \ln n = n(\ln(n+1) + \ln(n+1) \leq n \cdot \frac{1}{n} + \ln(n+1) = 1 + \ln(n+1)$$

and

$$(n+1) \ln n - n \ln n = n \ln\left(\frac{\ln(n+1)}{\ln n}\right) + \ln(n+1)
\leq n \cdot \left(\frac{\ln(n+1)}{\ln n} - 1\right) + \ln(n+1)
\leq n \cdot \left(\frac{\ln 7023}{\ln 7022} - 1\right) + \ln(n+1)
< 0.00002 n + \ln(n+1).$$

The last two inequalities hold since the function $f(x) = \frac{\ln(x+1)}{\ln x}$ is decreasing and $\frac{\ln 7023}{\ln 7022} < 0.00002$. Hence

$$g(p_n) = p_{n+1} - p_n
\leq (n+1)(\ln(n+1) + \ln \ln(n+1) - 0.9385) - n(\ln n + \ln \ln n - 1.0072629)
\leq (1 + \ln(n+1)) + (0.00002n + \ln(n+1)) + 0.0687629n - 0.9385.$$

As $\ln x + \ln \ln x$ is increasing and concave downward, we have

$$\ln(n+1) + \ln \ln(n+1)
\leq (\ln 7023 + \ln \ln 7023) + (\ln x + \ln \ln x)|_{x=7023} \cdot (n - 7022)
< \ln 7023 + \ln 7023 + 0.0001585n - 1.1
< 11 + 0.0001585n,$$

and so

$$g(p_n) = p_{n+1} - p_n < 12 + 0.069n < 0.071n \quad \text{for } n \geq 7022.$$

Now we consider $D_{k \times n}$ with $k > p_{7022} = 70919$ and $k + 5 \leq n = k + w \leq 2k$. If $k \geq \frac{2w^2 \ln(w-1)}{w-3}$, then $d_k(n) \geq 2k \geq n$ as we have seen in Lemma 2.1. Thus, suppose that $k < \frac{2w^2 \ln(w-1)}{w-3}$. Since $k > 70919$, we certainly have $w > 9$, and
so \( \ln \ln(3w) = \ln(\ln 3 + \ln w) > \ln(3 \ln 3) > 0 \). Also, from \( w^2 - 9w > 0 \), we get \( 2w^2 < 3w(w - 3) \). Combining these with \( \ln 3 > 1.098 \) (hence \( \ln 3 - 1.0072629 > 0 \)), we get

\[
2w^2 \ln(w - 1) < 3w(w - 3) \ln w
\]

\[
< 3w(w - 3)(\ln w + \ln 3 - 1.0072629 + \ln \ln(3w))
\]

\[
< 3w(w - 3)(\ln(3w) + \ln(3w) - 1.0072629)
\]

\[
\leq (w - 3)p_{3w}
\]

by (2.2). Therefore,

\[
p_{7022} < k < \frac{2w^2 \ln(w - 1)}{w - 3} < p_{3w}.
\]

As \( g(p_{3w}) < 0.213w \) by (2.3), the prime gaps for the primes between \( p_{7022} \) and \( p_{3w} \) are all smaller than \( 0.213w \). This means that there is a prime between \( k \) and \( k + 0.213w \), and another one between \( k + 0.213w \) and \( k + 2 \cdot 0.213w \). In particular, there are at least two distinct primes in \([k, k + w]\), and so \( D_k(n) \geq 2k \geq n \) again by Lemma 2.1.

Next, suppose that \( k \leq p_{7022} \). From the prime gap table above, we see that for any prime \( p \) with \( k \leq p \leq p_{7022} = 70919 < 155928 \), the prime gap \( g(p) \) is less than or equal to \( g_{14} = 72 \). Also, for \( 2000 < k \leq 70919 \), we have \( w_k \geq 189 > 144 \). Thus, there are at least two primes between \( k \) and \( k + w_k \).

For any \( k \) with \( 600 < k \leq 2000 \), \( w_k \geq 68 \). For any prime \( p \) with \( k \leq p \leq 2000 \), there are at least two primes between \( k \) and \( k + w_k \).

For any \( k \) with \( 300 < k \leq 600 \), \( w_k \geq 38 \), and for any prime \( p \) with \( k \leq p \leq 600 \), the prime gap \( g(p) \) is at most 34. Thus, for any \( k \) with \( 600 < k \leq 2000 \), there are at least two primes between \( k \) and \( k + w_k \).

For any \( k \) with \( 300 < k \leq 600 \), \( w_k \geq 38 \), and for any prime \( p \) with \( k \leq p \leq 600 \), the prime gap \( g(p) \) is at most 18. Again, for \( 300 < k \leq 600 \), there are at least two primes between \( k \) and \( k + w_k \).

Therefore, for all \( k \) with \( 300 < k \leq p_{7022} \), there are always two primes between \( k \) and \( k + w_k \). By Lemma 2.1, \( d_k(n) \geq 2k > n \) for any such \( k \) and any \( n \) with \( k + 5 \leq n \leq 2k \).

We have now shown that \( d_k(n) > n \) for all \( k \) and \( n \) with \( k > 300 \) and \( k < n \leq 2k \). Hence we can announce that the Restricted 1-2-3 Conjecture is true for any \( n, k \in \mathbb{N} \) with \( k < n \leq 2k \).

Induction kicks in from here. The starting ground is that \( d_1(n) = n \) for all \( n > 1 \). Let \( k > 1 \), and assume that we have the Restricted 1-2-3 Conjecture verified up to \( k - 1 \). That is, assume that \( d_k(m) \geq m \) when \( s \leq k - 1 < m \). We want to show that \( d_k(n) \geq n \) for all \( n > k \). From the above, we also know that this is true for \( n \) up to \( 2k \). So, we assume that \( n > 2k \) and also that \( d_k(m) \geq m \) when \( 2k \leq m < n \). And we continue. . .

3. A REDUCTION

The first step is to make certain that we do not need to care too much for \( n \) large enough. Namely, we will show that it is sufficient to restrict \( n \) to be no larger than \( \text{LCM}(I_k) \), the least common multiple of \( \{1, 2, \ldots, k\} \).


Let \( T \) be a nonempty subset of \( I_k \) with \(|T| = \ell\). Then we have
\[
\sum_{\emptyset \neq S \subseteq T} (-2)^{|S|-1} = \sum_{i=1}^{\ell} \binom{\ell}{i} (-2)^{i-1} \\
= (-2)^{-1} \cdot \sum_{i=1}^{\ell} \binom{\ell}{i} (-2)^i \\
= (-2)^{-1} \cdot (((-2) + 1)^\ell - 1) \\
= (-2)^{-1} \cdot ((-1)^\ell - 1) \\
= \begin{cases} 0, & \text{if } \ell \text{ is even;} \\
1, & \text{if } \ell \text{ is odd.} \\
\end{cases}
\]

For convenience and by abusing notation, if \( T = \emptyset \), we put \( \sum_{\emptyset \neq S \subseteq T} (-2)^{|S|-1} = 0 \). Using this identity, we have

**Theorem 3.1.** Let \( n \geq k \). Then
\[
d_k(n) = \sum_{\emptyset \neq S \subseteq I_k} \left\lfloor \frac{n \min S}{\text{LCM}(S)} \right\rfloor \cdot (-2)^{|S|-1}.
\]

**Proof.** Denote by \( 2^{I_k} \) the power set of \( I_k \). Define \( \theta : \mathbb{N} \to 2^{I_k} \) by \( \theta(m) = \{ s \in I_k \mid m \in sI_n \} \). For \( \emptyset \neq S \subseteq I_k \), set
\[
\tilde{S} = \theta^{-1}(S) = \{ m \in \mathbb{N} \mid \theta(m) = S \}
\]
and
\[
\overline{S} = \{ m \in \mathbb{N} \mid \theta(m) \supseteq S \}.
\]

Note that if \( \emptyset \neq S \subseteq T \), then \( m \in \overline{S} \) whenever \( m \in \overline{T} \), and \( \overline{S} \) is the disjoint union of \( \overline{T} \) for subsets \( T \) of \( I_k \) containing \( S \). Therefore, \( |\overline{S}| = \sum_{S \subseteq T \subseteq I_k} |\overline{T}| \).

An integer \( m \geq 1 \) will appear in \( D_{k \times n} \) if and only if \( \theta(m) \) is a nonempty set with an odd number of elements in it. Therefore, we have
\[
d_k(n) = \sum_{m \geq 1} \left( \sum_{\emptyset \neq S \subseteq \theta(m)} (-2)^{|S|-1} \right)
\]
\[
= \sum_{\emptyset \neq T \subseteq I_k} |\overline{T}| \cdot \left( \sum_{\emptyset \neq S \subseteq T} (-2)^{|S|-1} \right)
\]
\[
= \sum_{\emptyset \neq T \subseteq I_k} \sum_{\emptyset \neq S \subseteq T} |\overline{T}| \cdot (-2)^{|S|-1}
\]
\[
= \sum_{\emptyset \neq S \subseteq I_k} \left( \sum_{S \subseteq T \subseteq I_k} |\overline{T}| \right) \cdot (-2)^{|S|-1}
\]
\[
= \sum_{\emptyset \neq S \subseteq I_k} |\overline{S}| \cdot (-2)^{|S|-1}.
\]

To finish the proof, we notice that for any nonempty subset \( S \) of \( I_k \),
\[
\overline{S} = \{ m \geq 1 \mid m \in sI_n \text{ for all } s \in S \}
\]
\[
= \{ m \geq 1 \mid \text{LCM}(S) \text{ divides } m \text{ and } m \leq sn \text{ for all } s \in S \}.
\]

Therefore, \( |\overline{S}| = \left\lfloor \frac{n \min S}{\text{LCM}(S)} \right\rfloor \). \qed
Suppose that \( n = a + b \cdot \text{LCM}(I_k) \), where \( a, b \in \mathbb{N} \) with \( a \leq \text{LCM}(I_k) \). For any nonempty subset \( S \) of \( I_k \), since \( \text{LCM}(S) \) divides \( \text{LCM}(I_k) \), \( \frac{b \cdot \text{LCM}(I_k) \cdot \min S}{\text{LCM}(S)} \) is an integer, and we have
\[
\left\lfloor \frac{a - \min S}{\text{LCM}(S)} \right\rfloor = \left\lfloor \frac{a + b \cdot \text{LCM}(I_k) \cdot \min S}{\text{LCM}(S)} \right\rfloor = \left\lfloor \frac{a - \min S}{\text{LCM}(S)} \right\rfloor + \frac{b \cdot \text{LCM}(I_k) \cdot \min S}{\text{LCM}(S)} = \left\lfloor \frac{a - \min S}{\text{LCM}(S)} \right\rfloor + b \cdot \frac{\text{LCM}(I_k) \cdot \min S}{\text{LCM}(S)}.
\]
This makes \( d_k(a + b \cdot \text{LCM}(I_k)) = d_k(a) + b \cdot d_k(\text{LCM}(I_k)) \). If we can show that \( d_k(n) \geq n \) for all \( n \leq \text{LCM}(I_k) \), then we are done.

With the above preparation, we make the assumption that
\[ 2k < n \leq \text{LCM}(I_k) \]
for the rest of the paper, and move on.

4. The case when \( n > 2k \)

We start with an easy observation.

**Lemma 4.1.** Let \( p \) and \( q \) be distinct primes which are greater than \( \max\{k, \sqrt{n}\} \) and less than or equal to \( n \). Then for any \( s, t \in \mathbb{N} \) with \( s \leq \left\lfloor \frac{n}{p} \right\rfloor \) and \( t \leq \left\lfloor \frac{n}{q} \right\rfloor \), we have \((sp)I_k \cap (tq)I_k = \emptyset\).

**Proof.** Suppose that \( spa = tqb \) for some \( a, b \in I_k \). Then \( p \mid tb \). Since \( p \) is a prime larger than \( k \), we must have \( p \nmid t \). From \( t \leq \left\lfloor \frac{n}{q} \right\rfloor \), we reach a contradiction that \( pq \leq tq \leq n \). \( \square \)

**Remark 4.2.** Suppose that \( p \) is a prime such that \( \max\{k, \sqrt{n}\} < p \leq n \). Then
\[ pI_k \Delta 2pI_k \Delta \cdots \Delta \left\lfloor \frac{n}{p} \right\rfloor pI_k = p(I_k \Delta 2I_k \Delta \cdots \Delta \left\lfloor \frac{n}{p} \right\rfloor I_k), \]
which has at least \( \max\{k, \left\lfloor \frac{n}{p} \right\rfloor\} \) many elements by the induction hypothesis. Combining this with Lemma 4.1, our goal is then to show that
\[ \sum_{p \in \mathcal{P}} \max\{k, \left\lfloor \frac{n}{p} \right\rfloor\} \geq n, \]
where \( \mathcal{P} \) is defined to be
\[ \mathcal{P} = \{ p \mid p \text{ is a prime and } \max\{k, \sqrt{n}\} < p \leq n \}. \]

We will use the following results from number theory, where \( \pi(x) \) denotes the number of primes less than or equal to \( x \).

**Proposition 4.3** ([2, Theorem 2 and Corollary to Theorem 3] and [5, (3.5), (3.6), (3.8)]).

1. \( 2^k \leq \text{LCM}(I_k) \leq 4^k \).
2. \( \pi(x) > x/\ln x \) for \( x \geq 17 \).
3. \( \pi(x) < 1.25506x/\ln x \) for \( x > 1 \).
4. \( \pi(2x) - \pi(x) > 3x/(5\ln x) \) for \( x > 20.5 \).
4.1. **The case when $2k < n \leq k^2$.** Let $\left\lceil \frac{n}{k} \right\rceil = m$. Thus, $m \geq 3$ and $km \geq n$. Since each prime in $P$ contributes at least $k$ elements to the set $D_{k \times n}$ (see Remark 1.2), we aim to show that $|P| \geq m$.

Therefore, we want $\pi(n) - \pi(k) \geq m$. It is easy to verify that $d_k(n) \geq n$ for $k$ and $n$ with $k \leq 20$ and $2k < n \leq k^2$. So we assume that $k \geq 21$.

Let $v = m - 1$. Then $2 \leq v < k - 1$ and $n \geq vk + 1$, and the goal is to show that $\pi(vk + 1) - \pi(k) \geq v + 1$.

For $v = 2, 3$ or $4$, we have

$$\pi(vk + 1) - \pi(k) \geq \pi(2k) - \pi(k) \geq 3k/(5 \ln k) \geq 3 \cdot 21/(5 \ln 21) > 4.1.$$  

Since $\pi(vk + 1) - \pi(k)$ is an integer, it is at least $5$, which is greater than $v - 1$.

On the other hand, for $v \geq 5$ we have

$$\pi(vk + 1) - \pi(k) \geq \pi(vk) - \pi(k) \geq \frac{vk}{\ln(vk)} - \frac{1.25506k}{\ln k}$$  

$$> \frac{vk}{\ln(k^2)} - \frac{1.3k}{\ln k} \geq \frac{vk}{2 \ln k} - \frac{1.3k}{\ln k}$$  

$$= \frac{k}{\ln k} \cdot \left( \frac{v}{2} - 1.3 \right) \geq \frac{21}{21} \cdot \left( \frac{v}{2} - 1.3 \right)$$  

$$> 6 \cdot \left( \frac{v}{2} - 1.3 \right) > v + 1.$$  

Thus the case when $2k < n \leq k^2$ is done.

4.2. **The case when $n > k^2$.** In this case, $\max\{k, \sqrt{n}\} = \sqrt{n}$. Let $a_0 = n$, $a_i = \frac{n}{\sqrt{k}}$ for $i = 1, \ldots, \ell$, and $a_{\ell + 1} = \sqrt{n}$, where $\ell$ is the largest integer such that $\frac{n}{a_{\ell + 1} a_{\ell + 1}} \leq \sqrt{n} < \frac{n}{a_0 a_0}$. Here we have $\left\lceil \frac{n}{p} \right\rceil \leq k$ if $a_i < p \leq a_0$, and $\left\lceil \frac{n}{p} \right\rceil = k + i$ if $a_{i + 1} < p \leq a_i$ ($i = 1, 2, \ldots, \ell$); therefore,

$$\sum_{p \in P} \max\{k, \left\lceil \frac{n}{p} \right\rceil\} = \sum_{\pi < p < n} \max\{k, \left\lceil \frac{n}{p} \right\rceil\}$$  

$$= \sum_{a_1 < p \leq a_0} \max\{k, \left\lceil \frac{n}{p} \right\rceil\} + \sum_{a_2 < p \leq a_1} \max\{k, \left\lceil \frac{n}{p} \right\rceil\} + \cdots + \sum_{a_{\ell + 1} < p \leq a_\ell} \max\{k, \left\lceil \frac{n}{p} \right\rceil\}$$  

$$= \sum_{a_1 < p \leq a_0} k + \sum_{a_2 < p \leq a_1} \left\lceil \frac{n}{p} \right\rceil + \cdots + \sum_{a_{\ell + 1} < p \leq a_\ell} \left\lceil \frac{n}{p} \right\rceil$$  

$$= \sum_{a_1 < p \leq a_0} k + \sum_{a_2 < p \leq a_1} (k + 1) + \cdots + \sum_{a_{\ell + 1} < p \leq a_\ell} (k + \ell)$$  

$$= \sum_{i=0}^{\ell} (k + i)(\pi(a_i) - \pi(a_{i + 1}))$$  

$$= \sum_{i=0}^{\ell} k(\pi(a_i) - \pi(a_{i + 1})) + \sum_{i=1}^{\ell} i(\pi(a_i) - \pi(a_{i + 1}))$$  

$$= \left( k\pi(a_0) - k\pi(a_{\ell + 1}) \right) + \left( \sum_{i=1}^{\ell} \pi(a_i) - \ell\pi(a_{\ell + 1}) \right)$$  

$$= k\pi(n) + \left( \sum_{i=1}^{\ell} \pi(a_i) \right) - (k + \ell)\pi(\sqrt{n}).$$
Using Proposition 1.3 we have
\[
\pi(\sqrt{n}) < 1.25506 \cdot \frac{\sqrt{n}}{\ln \sqrt{n}} < 2.52 \cdot \frac{\sqrt{n}}{\ln n},
\]
and if \(n \geq 17\),
\[
\sum_{i=1}^{\ell} \frac{n}{\ln(n + i)} > \sum_{i=1}^{\ell} \frac{n}{\ln(n + i + 1)} = \frac{n}{\ln n} \sum_{i=1}^{\ell} \frac{1}{\ln n + i} > \frac{n}{\ln n} (\ln(k + \ell + 1) - \ln(k + 1)).
\]
The last inequality came from \(\sum_{i=1}^{\ell} \frac{n}{\ln n + i} \geq \int_{k+1}^{k+\ell+1} \frac{1}{x} dx\). Also, from \(\frac{n}{\sqrt{n}} \leq \sqrt{n} < \frac{n}{k+1}\) we have \(k + \ell < \sqrt{n} \leq k + \ell + 1\), and so \(\frac{1}{\sqrt{n}} \leq \ln(n) \leq \ln(k + \ell + 1)\). This yields
\[
\sum_{p \in P} \max\{k, \lfloor \frac{n}{p} \rfloor\} > k \frac{n}{\ln n} + \frac{n}{\ln n} (\ln(k + \ell + 1) - \ln(k + 1)) - (k + \ell) \sqrt{n} \frac{1}{\ln n} \cdot 2.52 
\]
\[
\geq \frac{n}{2} + \frac{n}{\ln n} \cdot (k - \ln(k + 1) - 2.52).
\]

To finish the task, we need only to have \(\frac{k - \ln(k+1) - 2.52}{\ln n} \geq \frac{1}{2}\). To this end, we use the fact that \(n \leq \text{LCM}(I_k) < 4^k\). So
\[
\frac{k - \ln(k+1) - 2.52}{\ln n} \geq \frac{k - \ln(k+1) - 2.52}{\ln(\text{LCM}(I_k))} > \frac{k - \ln(k+1) - 2.52}{k \cdot \ln 4}.
\]

Now,
\[
\frac{k - \ln(k+1) - 2.52}{k \cdot \ln 4}
\]
is increasing for all \(k\) and is more than \(\frac{1}{2}\) when \(k = 18\).
\[
\frac{k - \ln(k+1) - 2.52}{\ln(\text{LCM}(I_k))} \geq \frac{1}{2} \quad \text{for } k = 8, 9, \ldots, 17.
\]
So, we see that, indeed,
\[
\sum_{p \in P} \max\{k, \lfloor \frac{n}{p} \rfloor\} \geq n \text{ when } n \geq k^2 \geq 64.
\]
For \(k < 8\), we have to check \(d_k(n)\) for \(k^2 < n \leq \text{LCM}(I_k)\). Since \(k^2 > \text{LCM}(I_k)\) for \(k = 2, 3, 4\), this amounts to checking the cases \(k = 5\) with \(25 < n \leq \text{LCM}(5) = 60\), \(k = 6\) with \(36 < n \leq 60\), and \(k = 7\) with \(49 < n \leq 420\). Again, a simple computer routine verifies that these are all fine. Therefore, we conclude that the Restricted 1-2-3 Conjecture is true for all \(k\) and \(n\) with \(n \geq k^3\), and as well conclude our proof for the Restricted 1-2-3 Conjecture.

Finally, for the “Codes by composition” (see the introduction), one needs even more. Here, we mention as an open problem:

**Extended 1-2-3 Conjecture.** For every finite, nonempty subset \(I\) of the natural numbers, the symmetric difference of the sets \(iI_k, i \in I\), has at least \(k\) elements.

Note that it is not true if one changes “at least \(k\) elements” to “at least \(k\) or \(n\) elements, whichever is larger”!
REFERENCES


Department of Mathematics and National Center for Theoretical Sciences (South), National Cheng Kung University, 1 University Road, Tainan 701, Taiwan
E-mail address: pyhuang@mail.ncku.edu.tw

Department of Mathematics and National Center for Theoretical Sciences (South), National Cheng Kung University, 1 University Road, Tainan 701, Taiwan
E-mail address: wfke@mail.ncku.edu.tw

Department of Algebra, Johannes Kepler Universität Linz, Altenberger Strasse 69, 4040 Linz, Austria
E-mail address: guenter.pilz@jku.at