ACTA ARITHMETICA
132.2 (2008)

Minimal Niven numbers

by

H. Fredricksen (Monterey, CA), E. J. Ionascu (Columbus, GA),
F. Luca (Morelia) and P. Stănică (Monterey, CA)

1. Motivation. A positive integer n is a *Niven number* (or a *Harshad number*) if it is divisible by the sum of its (decimal) digits. For instance, 2007 is a Niven number since 9 divides 2007. A *q-Niven number* is an integer k which is divisible by the sum of its base q digits, call it $s_q(k)$ (if $q = 2$, we shall use $s(k)$ for $s_2(k)$). Niven numbers have been extensively studied by various authors (see Cai [3], Cooper and Kennedy [4], De Koninck and Doyon [5], De Koninck, Doyon and Kátai [6], Grundman [7], Mauduit, Pomerance and Sárközy [11], Mauduit and Sárközy [12], Vardi [16], to cite just a few of the most recent works).

In this paper, we define a natural sequence in relation to q-Niven numbers. For a fixed but arbitrary $k \in \mathbb{N}$ and a base $q \geq 2$, one may ask whether or not there exists a q-Niven number whose sum of digits is precisely k. We will show that the answer is affirmative. Therefore, it makes sense to define a_k to be the smallest positive multiple of k such that $s_q(a_k) = k$. In other words, a_k is the smallest Niven number whose sum of digits is k. We denote by c_k the companion sequence $c_k = a_k/k$, $k \in \mathbb{N}$. Obviously, a_k and c_k depend on q, but we will not make this explicit to avoid cluttering the notation.

In this paper, we give two different constructive techniques, for the binary and nonbinary cases, yielding sharp upper bounds for a_k. We find elementary upper bounds true for all k, and then better nonelementary ones true for most odd k.

Throughout the paper, we use the Vinogradov symbols $\gg$ and $\ll$ and the Landau symbols O and o with their usual meanings. The constants

2000 *Mathematics Subject Classification*: 11L20, 11N25, 11N37.

Key words and phrases: sum of digits, Niven numbers.

Work by F. L. was started in the Spring of 2007 while he visited the Naval Postgraduate School. He would like to thank this institution for its hospitality. H. F. acknowledges support from the National Security Agency under contract RMA54. Research of P. S. was supported in part by a RIP grant from Naval Postgraduate School.

implied by such symbols are absolute. We write x for a large positive real number, and p and q for prime numbers. If $\mathcal{A}$ is a set of positive integers, we write $\mathcal{A}(x) = \mathcal{A} \cap [1, x]$. We write $\ln x$ for the natural logarithm of x and $\log x = \max\{\ln x, 1\}$.

Acknowledgements. We thank Professor Igor Shparlinski for useful suggestions.

2. Easy proof for the existence of a_k. In this section we present a simple argument that shows that the sequence a_k is well defined. First we assume that k satisfies $\gcd(k, q) = 1$. By Euler's theorem, we can find an integer t such that $q^t \equiv 1 \pmod{k}$, and then define $K = 1 + q^t + q^{2t} + \cdots + q^{(k-1)t}$. Obviously, $K \equiv 0 \pmod{k}$, and also $s_q(K) = k$. Hence, in this case, K is a Niven number whose digits in base q are only 0's and 1's and whose sum is k.

If k is not coprime to q, we write $k = ab$ where $\gcd(b, q) = 1$ and a divides q^n for some $n \in \mathbb{N}$. As before, we can find $K \equiv 0 \pmod{b}$ with $s_q(K) = b$. Let $u = \max\{n, \lceil \log_q K \rceil\} + 1$, and define $K' = (q^u + q^{2u} + \cdots + q^{au})K$. Certainly $k = ab$ is a divisor of K' and $s_q(K') = ab = k$. Therefore, a_k is well defined for every $k \in \mathbb{N}$.

This argument gives a large upper bound, namely of size $\exp(O(k^2))$ for a_k.

We remark that if m is the minimal q-Niven number corresponding to k, then $q-1$ must divide $m - s_q(m) = kc_k - k = (c_k - 1)k$. This observation turns out to be useful in the calculation of c_k for small values of k. For instance, in base ten, the following table of values of a_k and c_k can be established easily by using the previous simple observation. As an example, if $k = 17$ then 9 has to divide $c_{17} - 1$ and so we need only check $10, 19, 28$.

k	10	11	12	13	14	15	16	17	18	19	20	21	22	23
c_k	19	19	4	19	19	13	28	28	11	46	199	19	109	73
a_k	190	209	48	247	266	195	448	476	198	874	3980	399	2398	1679

3. Elementary bounds for a_k in the binary case. For each positive integer k we set $n_k = \lceil \log_2 k \rceil$. Thus, n_k is the smallest positive integer with $k \le 2^{n_k}$. Assuming that $k \in \mathbb{N}$ $(k > 1)$ is odd, we let t_k be the multiplicative order of 2 modulo k, and so, $2^{t_k} \equiv 1 \pmod{k}$. Obviously, $t_k \ge n_k$ and $t_k \mid \phi(k)$, where ϕ is Euler's totient function. Thus,

$$n_k \le t_k \le k - 1. \tag{1}$$

LEMMA 1. *For every odd integer $k > 1$, every integer $x \in \{0, 1, \ldots, k-1\}$ can be represented as a sum modulo k of exactly n_k distinct elements of*

$$D = \{2^i : i = 0, \ldots, n_k + k - 2\}.$$

Proof. We find the required representation in a constructive way. Let us start with an example. If $x = 0$ and $k = 2^{n_k} - 1$, then since $x \equiv k \pmod{k}$, we have a representation as required by writing $k = 1 + 2 + \cdots + 2^{n_k - 1}$ (note that $n_k - 1 \le n_k + k - 2$ is equivalent to $k \ge 1$).

Any $x \in \{0, 1, \ldots, k-1\}$ has at most n_k bits of which at most $n_k - 1$ are ones. Next, let us illustrate the construction when this binary representation of x contains exactly $n_k - 1$ ones, say

$$x = 2^{n_k - 1} + 2^{n_k - 2} + \cdots + 2 + 1 - 2^j \quad \text{for some } j \in \{0, 1, \ldots, n_k - 1\}.$$

First, we assume $j \le n_k - 2$. Using $2^{j+1} = 2^j + 2^j \equiv 2^j + 2^{j+t_k} \pmod{k}$, we write

$$x \equiv 2^{j+t_k} + 2^{n_k - 1} + \cdots + 2^{j+2} + 2^j + 2^{j-1} + \cdots + 1 \pmod{k},$$

where both $j + t_k \le n_k - 2 + k - 1 = n_k + k - 3$ and $j + t_k > n_k - 1$ are true according to (1). Therefore all exponents are distinct and they lie in the required range, which gives us a representation of x as a sum of exactly n_k different elements of D modulo k.

If $j = n_k - 1$, then $x = 2^{n_k - 1} - 1$. We consider $x + k$ instead of x. By the definition of n_k, we must have $k \ge 2^{n_k - 1} + 1$. Hence, $x + k \ge 2^{n_k}$, which implies that the binary representation of $x + k$ starts with 2^{n_k} and it has at most n_k ones. Indeed, if $s(x + k) \ge n_k + 1$, then $x + k \ge 2^{n_k} + 2^{n_k - 1} + \cdots + 2 + 1 = 2^{n_k + 1} - 1$, which in turn contradicts the inequality $x + k \le k - 1 + k = 2k - 1 \le 2^{n_k + 1} - 3$ since k is odd. If $s(x + k) = n_k$, then we are done ($k \ge 3$). If $s(x + k) = n_k - 1$, then we proceed as before and observe that this time $j + t_k \le n_k + k - 2$ for every $j \in \{0, 1, 2, \ldots, n_k - 1\}$, and $j + t_k > n_k$ if $j > 0$, which is an assumption that we can make because in order to obtain $n_k - 1$ ones, two of the powers of 2, out of $1, 2, 2^2, \ldots, 2^{n_k - 1}$, must be missing.

If $s(x + k) < n_k - 1$, then for every zero in the representation of $x + k$ which is preceded by a one and followed by l ($l \ge 0$) other zeros, we can fill out the zero gap in the following way. If such a zero is the coefficient of 2^j, then we replace 2^{j+1} by $2^j + 2^{j-1} + \cdots + 2^{j-l} + 2^{j-l+t_k}$. This will give $l + 2$ ones instead of a one and $l + 1$ zeros. We fill out all gaps this way except the gap corresponding to the smallest power of 2 and $l \ge 1$, where in order to ensure the inequality $j' + t_k > n_k$ ($j' = j - l + 1 > 0$) one will replace 2^{j+1} by $2^j + 2^{j-1} + \cdots + 2^{j-l+1} + 2^{j-l+1+t_k}$. The result will be a representation in which all the additional powers $2^{j'+t_k}$ will be distinct and the total number of powers of two is n_k. The maximum exponent of these powers is at most $j' + t_k \le n_k + k - 2$.

If the representation of x starts with $2^{n_k - 1}$, then the technique described above can be applied directly to x making sure that all zero gaps are completely filled. Otherwise, we apply the previous technique to $x + k$. ∎

EXAMPLE 2. Let $k = 11$. Then $n_{11} = 4$ and $t_{11} = 10$. Suppose that we want to represent 9 as a sum of 4 distinct terms modulo 11 from the set $D = \{1, 2, \ldots, 2^{13}\}$. Since $9 = 2^3 + 1$, we have $9 = 2^2 + 2 + 2 + 1$, so $9 \equiv 2^2 + 2 + 2^{11} + 1 \pmod{11}$. If we want to represent $7 = 2^2 + 2^1 + 2^0$ then, since this representation does not contain 2^3, we look at $7 + 11 = 18 = 2^4 + 2 = 2^3 + 2^3 + 2 = 2^3 + 2^2 + 2^2 + 2$. Thus, $7 \equiv 2^3 + 2^2 + 2^{12} + 2 \pmod{11}$.

We note that the representation given by Lemma 1 is not unique. If this construction is applied in such a way that the zero left when appropriate is always the one corresponding to the largest power of 2, we will obtain the largest such representation. In the previous example, we can fill out the smallest gap first and leave a zero from the gap corresponding to 2^3, so $7 \equiv 18 \equiv 2^4 + 2 = 2^3 + 2^3 + 1 + 1 \equiv 2^3 + 2^{13} + 1 + 2^{10} \pmod{11}$.

Recall that $2^\alpha \parallel m$ means that $2^\alpha \mid m$ but $2^{\alpha+1} \nmid m$. We write $\mu_2(m)$ for the exponent α.

THEOREM 3. *For all positive integers k and l, there exists a positive integer n having the following properties*:

(a) $s(nk) = lk$,
(b) $n \le (2^{lk+n_k} - 2^{\mu_2(k)})/k$.

Proof. It is clear that if k is a power of 2, say $k = 2^s$, then we can take $n = 2^{lk} - 1$, and so, $s(kn) = s(2^s + 2^{s+1} + \cdots + 2^{s+lk-1}) = lk$. In this case, the upper bound in part (b) is sharp since $n_k = s = \mu_2(k)$.

Furthermore, if k is of the form $k = 2^m d$ for some positive integers m, d with odd $d \ge 3$, then assuming we can find an integer $n \le (2^{2^m ld + n'_k} - 1)/d$, where $n'_k = \lceil \log_2 d \rceil$, such that $s(nd) = 2^m ld$, then nk satisfies (a) since $s(nk) = s(2^m nd) = s(nd) = 2^m ld = lk$. We observe that (b) is also satisfied in this case, because $(2^{2^m ld + n'_k} - 1)/d = (2^{lk+n_k} - 2^m)/k$.

Thus, without loss of generality, we may assume that $k \ge 3$ is odd. Consider the integer $M = 2^{lk+n_k} - 1 = 1 + 2^1 + \cdots + 2^{lk+n_k-1}$, and so $s(M) = lk + n_k$. By Lemma 1, we can write

$$M \equiv 2^{j_1} + 2^{j_2} + \cdots + 2^{j_{n_k}} \pmod{k}, \tag{2}$$

where $0 \le j_1 < \cdots < j_n \le k + n_k - 2 < lk + n_k - 1$. Therefore, we may take

$$n = \frac{M - (2^{j_1} + 2^{j_2} + \cdots + 2^{j_{n_k}})}{k},$$

which is an integer by (2) and satisfies $s(nk) = s(M - (2^{j_1} + 2^{j_2} + \cdots + 2^{j_{n_k}})) = lk$. ∎

COROLLARY 4. *The sequence $(a_k)_{k \ge 1}$ satisfies*

$$2^k - 1 \le a_k \le 2^{k+n_k} - 2^{\mu_2(k)}. \tag{3}$$

Proof. The first inequality in (3) follows from the fact that if $s(a_k) = k$, then $a_k \geq 1 + 2 + \cdots + 2^{k-1} = 2^k - 1$. The second inequality follows from Theorem 3 by taking $l = 1$, and from the minimality condition in the definition of a_k. ∎

We have computed a_k and c_k for all $k = 1, \ldots, 128$,

$$c_1 = 1, \quad c_2 = 3, \quad c_3 = 7, \ \ldots, \ c_{20} = 209715, \ \ldots$$

and the graph of $k \mapsto \ln(c_k)$ against the functions $k \mapsto \ln(2^k)$ and $k \mapsto \ln(2^k - 1) - \ln(k)$ is shown in Figure 1.

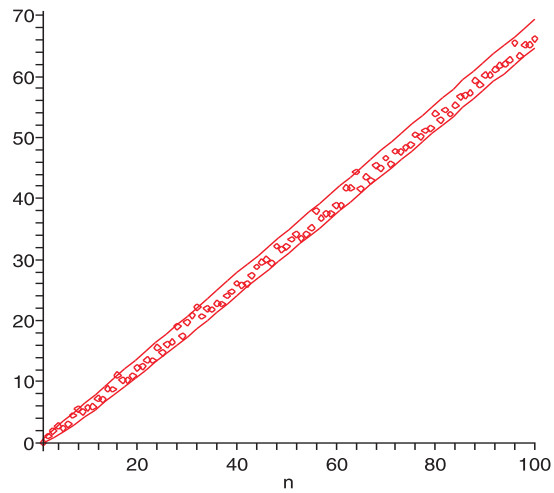

Fig. 1. The graphs of $k \mapsto \ln(c_k)$ and $k \mapsto \ln(2^k)$, $k \mapsto \ln(2^k - 1) - \ln(k)$

The right hand side of inequality (3) is sharp when $k = 2^s$, as we have already seen. For $k = 2^s - 1$, we get values of c_k very close to $2^k - 1$ but, in general, numerical evidence shows that $c_k/2^k$ is closer to zero more often than it is to 1. In fact, we show in Section 6 that this is indeed the case at least for odd indices (see Corollary 11, Corollary 12 and relation (23)).

4. Improving binary estimates and some closed formulae. In order to obtain better bounds for a_k, we introduce the following classes of odd integers. For a positive integer m we define

$$\mathcal{C}_m = \Big\{ k \equiv 1 \pmod 2 : 2^{k+m} - 1 \equiv \sum_{i=1}^{m} 2^{j_i} \pmod k \\ \text{for } 0 \leq j_1 < \cdots < j_m \leq m + k - 2 \Big\}.$$

Observe that $\mathcal{C}_m \subseteq \mathcal{C}_{m+1}$. Indeed, if k is in $\mathcal{C}_m$, then $2^{k+m} - 1 \equiv 2^{j_1} + \cdots + 2^{j_m}$ for some $0 \leq j_1 < \cdots < j_m \leq m + k - 2$. Multiplying the above congruence

by 2 and adding 1 to both sides, we get $2^{k+m+1} - 1 \equiv 1 + 2^{j_1+1} + \cdots + 2^{j_m+1}$, a representation which implies that k belongs to $\mathcal{C}_{m+1}$. Note also that Lemma 1 shows that every odd integer $k \geq 3$ belongs to $\mathcal{C}_u$, where $u = \lceil \log k / \log 2 \rceil$. Hence, $2\mathbb{N} + 1 = \bigcup_{m \in \mathbb{N}} \mathcal{C}_m$.

THEOREM 5. *For every $k \in \mathcal{C}_1$, we have*

$$2^k - 1 < a_k < 2^{k+1} - 1.$$

In particular, $c_k/2^k \to 0$ as $k \to \infty$ through $\mathcal{C}_1$. Furthermore, $a_k = 2^{k+1} - 1 - 2^{j_1}$, where $j_1 = j_0 + st_k$ with $s = \lfloor (k - 1 - j_0)/t_k \rfloor$ and $0 \leq j_0 \leq t_k - 1$ is such that $2^{k+1} - 1 \equiv 2^{j_0} \pmod{k}$.

Proof. We know that $2^k - 1 \not\equiv 0 \pmod{k}$ (see [13, Problem 37, p. 109]). Hence, an integer of binary length k whose sum of digits is k is not divisible by k. Therefore, $a_k > 2^k - 1$.

Next, we assume that a_k is an integer of binary length $k+1$ and sum of digits k; that is, $a_k = 2^{k+1} - 1 - 2^j$ for some $j = 0, \ldots, k-1$. But $2^{k+1} - 1 \equiv x \pmod{k}$, and by hypothesis there exists j_0 such that $x = 2^{j_0}$ for some $j_0 \in \{0, \ldots, t_k - 1\}$. In order to obtain a_k, we need to subtract the highest power of 2 possible because of the minimality of a_k. So, we need to take the greatest exponent $j_1 = j_0 + st_k \leq k - 1$, leading to $s = \lfloor (k - 1 - j_0)/t_k \rfloor$. Hence, $a_k = 2^{k+1} - 1 - 2^{j_1}$. ∎

Based on the above argument, we can compute, for instance, $a_5 = 55 = 2^6 - 1 - 2^3$, since $2^3 - 1 \equiv 2^3 \pmod{5}$. Similarly, $a_{29} = 2^{30} - 1 - 2^5 = 1073741791$, since $2^{30} - 1 \equiv 2^5 \pmod{29}$, and $a_{25} = 2^{26} - 1 - 2^{19} = 66584575$, since $2^{26} - 1 \equiv 2^{19} \pmod{25}$, or perhaps the more interesting example $a_{253} = 2^{254} - 1 - 2^{242}$.

THEOREM 6. *If $m \in \mathbb{N}$ and $k \in \mathcal{C}_{m+1} \setminus \mathcal{C}_m$, then*

$$2^{k+m-1} - 1 < a_k < 2^{k+m} - 1.$$

Thus, $c_k/2^k \to 0$ as $k \to \infty$ in $\mathcal{C}_m$ for any fixed m.

Proof. Similar to the proof of Theorem 5. ∎

THEOREM 7. *For all integers $k = 2^i - 1 \geq 3$, we have*

$$a_k \leq 2^{k+k^-} + 2^k - 2^{k-i} - 1, \tag{4}$$

where k^- is the least positive residue of $-k$ modulo i. Furthermore, the bound (4) is tight when $k = 2^i - 1$ is a Mersenne prime. In this case, we have $c_k/2^k \to 1/2$ as $k \to \infty$ through Mersenne primes, assuming that this set is infinite.

Proof. For the first claim, we show that the sum of binary digits of the bound of the upper bound on (4) is exactly k, and also that this number

is a multiple of k. From the definition of k^-, we find that $k + k^- = i\alpha$ for some positive integer α. Since

$$\begin{aligned} 2^{k+k^-} + 2^k - 2^{k-i} - 1 &= 2^{k-i}(2^i - 1) + 2^{i\alpha} - 1 \\ &= (2^i - 1)(2^{k-i} + 2^{i(\alpha-1)} + 2^{i(\alpha-2)} + \cdots + 1), \end{aligned}$$

we find that $2^{k+k^-} + 2^k - 2^{k-i} - 1$ is divisible by k. Further, $k^- \geq 1$ since k is not divisible by i (see the proof of Theorem 5), and

$$\begin{aligned} s(2^{k+k^-} + 2^k - 2^{k-i} - 1) &= s(2^{k+k^- -1} + \cdots + 2 + 1 + 2^k - 2^{k-i}) \\ &= s(2^{k+k^- -1} + \cdots + 2^k + \cdots + \widehat{2^{k-i}} + \cdots + 2 + 1 + 2^k) \\ &= s(2^{k+k^-} + 2^{k-1} + \cdots + \widehat{2^{k-i}} + \cdots + 2 + 1) = k, \end{aligned}$$

where $\hat{t}$ means that t is omitted. The first claim is proved.

We now consider a Mersenne prime $k = 2^i - 1$. First, we show that $k \in \mathcal{C}_i \setminus \mathcal{C}_{i-1}$. Since $u = \lceil \log k / \log 2 \rceil = i$, by Lemma 1 we know that $k \in \mathcal{C}_i$. Suppose by way of contradiction that $k \in \mathcal{C}_{i-1}$. Then

$$2^{k+i-1} - 1 \equiv 2^{j_1} + \cdots + 2^{j_{i-1}} \pmod{k} \tag{5}$$

with some $0 \leq j_1 < j_2 < \cdots < j_{i-1} \leq k + i - 3$. Since k is prime, we have $2^{k-1} \equiv 1 \pmod{k}$, and so $2^{k+i-1} - 1 \equiv 2^i - 1 \equiv 0 \pmod{k}$.

Because $2^i \equiv 1 \pmod{k}$, we can reduce all powers 2^j of 2 modulo k to powers with exponents less than or equal to $i-1$. We get at most $i-1$ such terms. But in this case, the sum of at least one and at most $i-1$ distinct members of the set $\{1, 2, \ldots, 2^{i-1}\}$ is positive and less than the sum of all of them, which is k. So, the equality (5) is impossible.

To finish the proof, we need to choose the largest representation $x = 2^{j_1} + \cdots + 2^{j_i}$, with $0 \leq j_1 < \cdots < j_i \leq k + i - 2$, such that $2^{k+i} - 1 \equiv x \pmod{k}$. But $2^{k+i} - 1 \equiv 2^{i+1} - 1 \equiv 1 \pmod{k}$. Since the exponents j are all distinct, the way to accomplish this is to take $j_i = k+i-2$, $j_{i-1} = k+i-3, \ldots, j_2 = k$, and finally j_1 to be the greatest integer with the property that the resulting x satisfies $x \equiv 1 \pmod{k}$. Since $x = 2^{j_1} + 2^k(1 + 2 + \cdots + 2^{i-2}) = 2^{j_1} + 2^k(2^{i-1} - 1) \equiv 2^{j_1} + 2^i - 2 \equiv 2^{j_1} - 1 \pmod{k}$, we need to have $2^{j_1} \equiv 2 \pmod{k}$. Since the multiplicative order of 2 modulo k is clearly i, we have to take the largest $j_1 = 1 + si$ such that $1 + si < k$. But i must be prime too, and so $2^{i-1} \equiv 1 \pmod{i}$. This implies $k = 2^i - 1 \equiv 1 \pmod{i}$. Therefore, $j_1 = k - i$. So, $a_k = 2^{k+i} - 1 - x = 2^{k+i} - 1 - 2^{k-i} - 2^{k+i-1} + 2^k = 2^{k+i-1} + 2^k - 2^{k-i} - 1$ and the inequality given in our statement becomes an equality since $k^- = i - 1$ in this case.

Regarding the limit claim, we observe that

$$\frac{c_k}{2^k} = \frac{k+1}{2k} + \frac{1}{k} - \frac{1}{k2^i} - \frac{1}{k2^k} \to \frac{1}{2}$$

as i (and hence k) goes to infinity. ∎

Between the two extremes, Theorems 6 and 7, we find out that the first situation is more predominant (see Corollary 12). Next, we give quantitative results on the sets $\mathcal{C}_m$. We start with a result which shows that $\mathcal{C}_1$ is of asymptotic density zero as one would less expect.

5. $\mathcal{C}_1$ is of density zero. Here, we show that $\mathcal{C}_1$ is of asymptotic density zero. For the purpose of this section only, we omit the index and simply write

$$\mathcal{C} = \{1 \le n : 2^{n+1} - 1 \equiv 2^j \pmod{n} \text{ for some } j = 1, 2, \ldots\}.$$

It is clear that $\mathcal{C}$ contains only odd numbers. Recall that for a positive real number x and a set $\mathcal{A}$ we put $\mathcal{A}(x) = \mathcal{A} \cap [1, x]$. We prove the following estimate.

THEOREM 8. *The estimate*

$$\#\mathcal{C}(x) \ll \frac{x}{(\log\log x)^{1/3}}$$

holds for all $x > e^e$.

Proof. We let x be large, and put q for the smallest prime exceeding

$$y = \frac{1}{2}\left(\frac{\log\log x}{\log\log\log x}\right)^{1/2}.$$

Clearly, for large x the prime q is odd and its size is $q = (1 + o(1))y$ as $x \to \infty$. For an odd prime p we write t_p for the order of 2 modulo p, defined at the beginning of Section 3. Recall that this is the smallest positive integer k such that $2^k \equiv 1 \pmod{p}$. Clearly, $t_p \mid p - 1$. We put

$$\mathcal{P} = \{p \text{ prime} : p \equiv 1 \pmod{q} \text{ and } t_p \mid (p-1)/q\}. \tag{6}$$

The effective version of Lagarias and Odlyzko of Chebotarev's density theorem (see [10], or p. 376 in [14]) shows that there exist absolute constants A and B such that

$$\#\mathcal{P}(t) = \frac{\pi(t)}{q(q-1)} + O\left(\frac{t}{\exp(A\sqrt{\log t}/q)}\right) \tag{7}$$

for all real numbers t as long as $q \le B(\log t)^{1/8}$. In particular, (7) holds when $x > x_0$ is sufficiently large and uniformly in $t \in [z, x]$, where we take $z = \exp((\log\log x)^{100})$.

We use the above estimate to compute the sum of the reciprocals of the primes $p \in \mathcal{P}(u)$, where we put $u = x^{1/100}$. We have

$$S = \sum_{p \in \mathcal{P}(u)} \frac{1}{p} = \sum_{\substack{p \in \mathcal{P} \\ p \le z}} \frac{1}{p} + \sum_{\substack{p \in \mathcal{P} \\ z < p \le u}} \frac{1}{p} = S_1 + S_2.$$

For S_1, we only use the fact that every prime $p \in \mathcal{P}$ is congruent to 1 modulo q. By the Brun–Titchmarsh inequality we have

$$S_1 \le \sum_{\substack{p \le z \\ p \equiv 1 \,(\mathrm{mod}\, q)}} \frac{1}{p} \ll \frac{\log\log z}{\phi(q)} \ll \frac{\log\log\log x}{q} = O(1).$$

For S_2, we are in the range where estimate (7) applies so by Abel's summation formula,

$$\begin{aligned} S_2 = \sum_{\substack{p \in \mathcal{P} \\ z \le p \le u}} \frac{1}{p} &\ll \int_z^u \frac{d\#\mathcal{P}(t)}{t} = \frac{\#\mathcal{P}(t)}{t}\Big|_{t=z}^{t=u} \\ &\quad + \int_z^u \left(\frac{\pi(t)}{q(q-1)t^2} + O\left(\frac{t^{-1}}{\exp(A\sqrt{\log t}/q)} \right) \right) dt \\ &= \int_z^u \frac{dt}{q(q-1)t\log t} + O\left(\frac{1}{q^2}\right) + O\left(\int_z^u \frac{dt}{q(q-1)t(\log t)^2} \right) \\ &= \frac{\log\log u - \log\log z}{q(q-1)} + O\left(\frac{1}{q^2}\right) = \frac{\log\log x}{q(q-1)} + O(1). \end{aligned}$$

In the above estimates, we used the fact that

$$\pi(t) = \frac{t}{\log t} + O\left(\frac{t}{(\log t)^2} \right),$$

as well as the fact that

$$\frac{t}{\exp(A\sqrt{\log t}/q)} = O\left(\frac{t}{q^2(\log t)^2} \right)$$

uniformly for $t \ge z$. To summarize, we have

$$S = \frac{\log\log x}{q(q-1)} + O(1) = \frac{\log\log x}{q^2} + O\left(\frac{\log\log x}{q^3} + 1 \right) = \frac{\log\log x}{q^2} + O(1). \tag{8}$$

We next eliminate a few primes from $\mathcal{P}$ defined in (6). Namely, we let

$$\mathcal{P}_1 = \{p : t_p < p^{1/2}/(\log p)^{10}\},$$

and

$$\mathcal{P}_2 = \{p : p-1 \text{ has a divisor } d \text{ in } [p^{1/2}/(\log p)^{10}, p^{1/2}(\log p)^{10}]\}.$$

A well-known elementary argument (see, for example, Lemma 4 in [1]) shows that

$$\#\mathcal{P}_1(t) \ll \frac{t}{(\log t)^2}, \tag{9}$$

therefore by the Abel summation formula one easily gets

$$\sum_{p\in\mathcal{P}_1}\frac{1}{p}=O(1).$$

As for $\mathcal{P}_2$, results of Indlekofer and Timofeev from [9] show that

$$\#\mathcal{P}_2(t)\ll\frac{t\log\log t}{(\log t)^{1+\delta}},$$

where $\delta=2-(1+\ln\ln 2)/\ln 2=0.08\ldots$, so again by Abel's summation formula one gets

$$\sum_{p\in\mathcal{P}_2}\frac{1}{p}=O(1).$$

We thus arrive at the conclusion that letting $\mathcal{Q}=\mathcal{P}\setminus(\mathcal{P}_1\cup\mathcal{P}_2)$, we have

$$S'=\sum_{p\in\mathcal{Q}(u)}\frac{1}{p}=S-\sum_{p\in\mathcal{P}_1(u)\cup\mathcal{P}_2(u)}\frac{1}{p}=\frac{\log\log x}{q^2}+O(1). \tag{10}$$

Now let us go back to the numbers $n\in\mathcal{C}$. Let $\mathcal{D}_1$ be the subset of $\mathcal{C}(x)$ consisting of the numbers free of primes in $\mathcal{Q}(u)$. By the Brun sieve,

$$\begin{aligned}\#\mathcal{D}_1&\ll x\prod_{p\in\mathcal{Q}(u)}\Big(1-\frac{1}{p}\Big)=x\exp\Big(-\sum_{p\in\mathcal{Q}(u)}\frac{1}{p}+O\Big(\sum_{p\in\mathcal{Q}(u)}\frac{1}{p^2}\Big)\Big)\\&\ll x\exp(-S'+O(1))\ll x\exp\Big(-\frac{\log\log x}{q^2}\Big)\\&=\frac{x}{(\log\log x)^{4+o(1)}}\ll\frac{x}{(\log\log x)^3}.\end{aligned} \tag{11}$$

Assume from now on that $n\in\mathcal{C}(x)\setminus\mathcal{D}_1$. Thus, $p\,|\,n$ for some prime $p\in\mathcal{Q}(u)$. Assume that $p^2\,|\,n$ for some $p\in\mathcal{Q}(u)$. Denote by $\mathcal{D}_2$ the subset of such $n\in\mathcal{C}(x)\setminus\mathcal{D}_1$. Keeping $p\in\mathcal{Q}(u)$ fixed, the number of $n\le x$ with the property that $p^2\,|\,n$ is $\le x/p^2$. Summing up now over all primes $p\equiv 1 \pmod q$ not exceeding $x^{1/2}$, we see that the number of such $n\le x$ is at most

$$\#\mathcal{D}_2\le\sum_{\substack{p\le x^{1/2}\\ p\equiv 1\,(\mathrm{mod}\,q)}}\frac{x}{p^2}\ll\frac{x}{q^2\log q}\ll\frac{x}{\log\log x}. \tag{12}$$

Let $\mathcal{D}_3=\mathcal{C}(x)\setminus(\mathcal{D}_1\cup\mathcal{D}_2)$. Write $n=pm$, where p does not divide m. We may also assume that $n\ge x/\log x$ since there are only at most $x/\log x$ positive integers n failing this condition. Put $t=t_p$. The definition of $\mathcal{C}$ implies that

$$2^{mp+1}\equiv 2^j+1 \pmod p$$

for some $j=1,\ldots,t$, and since $2^p\equiv 2 \pmod p$, we get $2^{mp+1}\equiv 2^{m+1} \pmod p$. We note that $2^{m+1} \pmod p$ determines $m\le x/p$ uniquely modulo t.

We estimate the number of values that m can take modulo t. Writing $X = \{2^j \pmod p\}$, we see that $\#\{m \pmod p\} \le I/t$, where I is the number of solutions (x_1, x_2, x_3) to the equation

$$x_1 - x_2 - x_3 = 0, \quad x_1, x_2, x_3 \in X. \tag{13}$$

Indeed, note that if m and j are such that $2^{m+1} \equiv 1 + 2^j \pmod p$, then $(x_1, x_2, x_3) = (2^{m+1+y}, 2^y, 2^{j+y})$ for $y = 0, \dots, t-1$ is also a solution of (13), and conversely, every solution $(x_1, x_2, x_3) = (2^{y_1}, 2^{y_2}, 2^{y_3})$ of (13) arises from $2^{m+1} \equiv 1+2^j \pmod p$, where $m+1 = y_1 - y_2$ and $j = y_3 - y_2$, by multiplying it with 2^{y_2}.

To estimate I, we use exponential sums. For a complex number z put $\mathbf{e}(z) = \exp(2\pi i z)$. Using the fact that for $z \in \{0, 1, \dots, p-1\}$ the sum

$$\frac{1}{p} \sum_{a=0}^{p-1} \mathbf{e}(az/p)$$

is 1 if and only if $z = 0$ and is 0 otherwise, we get

$$I = \frac{1}{p} \sum_{x_1,x_2,x_3 \in X} \sum_{a=0}^{p-1} \mathbf{e}(a(x_1 - x_2 - x_3)/p).$$

Separating the term for $a = 0$, we get

$$I = \frac{(\#X)^3}{p} + \frac{1}{p} \sum_{a=1}^{p-1} \sum_{x_1,x_2,x_3 \in X} \mathbf{e}(a(x_1 - x_2 - x_3)/p) = \frac{t^3}{p} + \frac{1}{p} \sum_{a=1}^{p-1} T_a T_{-a}^2,$$

where we put $T_a = \sum_{x_1 \in X} \mathbf{e}(ax_1/p)$. A result of Heath-Brown and Konyagin [8] says that if $a \neq 0$, then

$$|T_a| \ll t^{3/8} p^{1/4}.$$

Thus,

$$I = \frac{t^3}{p} + O\Big(t^{3/8} p^{-3/4} \sum_{a=1}^{p-1} |T_a|^2\Big) = \frac{t^3}{p} + O(t^{11/8} p^{1/4}),$$

where the last estimate follows by completing the inner sum (with $a = 0$) and appealing to Parseval's formula. This leads to

$$\#\{m \pmod t\} \le \frac{I}{t} \le \frac{t^2}{p} + O(t^{3/8} p^{1/4}).$$

Since also $m \le x/p$, it follows that the number of acceptable values for m is

$$\ll \frac{x}{pt}\left(\frac{t^2}{p} + t^{3/8} p^{1/4}\right) \ll \frac{xt}{p^2} + \frac{x}{t^{5/8} p^{3/4}}$$

(note that $x/pt \geq 1$ because $pt < p^2 < u^2 < x$). Hence,

$$\#\mathcal{D}_3 \leq \sum_{p\in\mathcal{Q}(u)} \frac{xt}{p^2} + \sum_{p\in\mathcal{Q}(u)} \frac{x}{t^{5/8}p^{3/4}} = T_1 + T_2.$$

For the first sum T_1 above, we observe that $t \leq p/q$, therefore $t/p^2 \leq 1/(pq)$. Thus, the first sum above is

$$T_1 \ll \sum_{p\in\mathcal{Q}(u)} \frac{x}{pq} \ll \frac{xS'}{q} \ll \frac{x\log\log x}{q^3} \ll x\,\frac{(\log\log\log x)^{3/2}}{(\log\log x)^{1/2}}, \tag{14}$$

where we used again estimate (10). Finally, for T_2, we change the order of summation to get

$$T_2 \leq x \sum_{t\geq t_0} \frac{1}{t^{5/8}} \sum_{\substack{p\in\mathcal{Q}(u)\\ t(p)=t}} \frac{1}{p^{3/4}}, \tag{15}$$

where $t_0 = t_0(q)$ can be taken to be any lower bound on the smallest $t = t_p$ that can show up. We will consider it later. For the moment, note that for a fixed t, p is a prime factor of $2^t - 1$. Every such p is of the form $p = qt\lambda + 1$ for some positive integer λ. Since $t > p^{1/2}(\log p)^{10} > (qt\lambda)^{1/2}(\log t)^{10}$, we see that $\lambda < t/q$ uniformly in t and q once x is large. Thus,

$$\sum_{\substack{p\in\mathcal{Q}(u)\\ t(p)=t}} \frac{1}{p^{3/4}} < \frac{1}{t^{3/4}q^{3/4}} \sum_{1\leq\lambda\leq t/q} \frac{1}{\lambda^{3/4}} \ll \frac{1}{t^{3/4}q^{3/4}} \left(\frac{t}{q}\right)^{1/4} = \frac{1}{t^{1/2}q}.$$

Hence,

$$T_2 \ll \frac{x}{q} \sum_{t\geq t_0} \frac{1}{t^{9/8}}.$$

Since $p \notin \mathcal{P}_1 \cup \mathcal{P}_2$, we get $t_p > p^{1/2}(\log p)^{10}$. Since $p \geq 2q+1$, it follows that $t \gg q^{1/2}(\log q)^{10}$. Thus, for large x we may take $t_0 = q^{1/2}(\log q)^9$ and get an upper bound for T_2. Hence,

$$\begin{aligned} T_2 &\ll \frac{x}{q} \sum_{t>q^{1/2}(\log q)^9} \frac{1}{t^{9/8}} \ll \frac{x}{q} \int_{q^{1/2}(\log q)^9}^{\infty} \frac{1}{s^{9/8}}\,ds \\ &\ll \frac{x}{q}\left(-\frac{1}{s^{1/8}}\Big|_{q^{1/2}(\log q)^9}^{\infty}\right) \ll \frac{x}{q^{1+1/16}(\log q)^{1/8}} \ll \frac{x}{q^{17/16}(\log q)^{1/8}} \\ &\ll \frac{x(\log\log\log x)^{13/32}}{(\log\log x)^{17/32}}. \end{aligned} \tag{16}$$

Combining the bounds (14) and (16), we get

$$\#\mathcal{D}_3 \ll \frac{x}{(\log\log x)^{1/3}},$$

which together with the bounds (11) and (12) completes the proof of the theorem. ∎

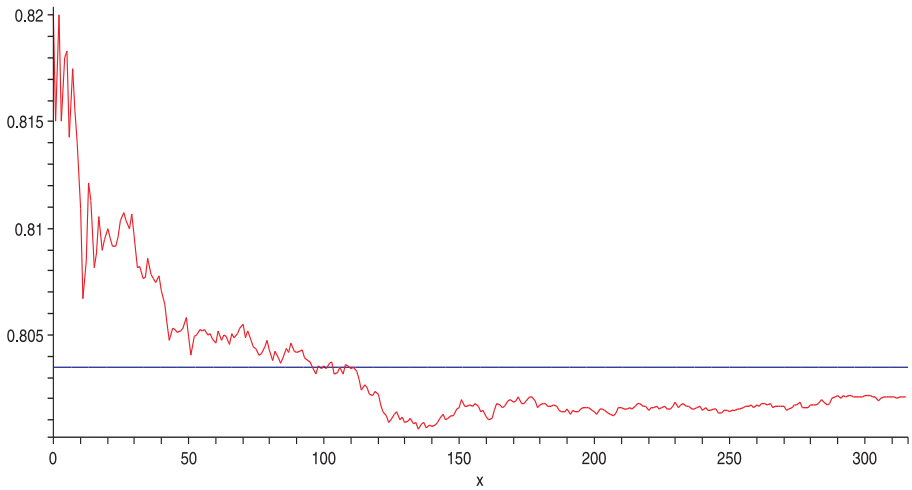

Fig. 2. The graph of $2\frac{\#\mathcal{C}_2(x)}{x}$, $1 \le x \le 63201$, x odd

Although the density of $\mathcal{C}_1$ is zero, one may try to calculate the densities of $\mathcal{C}_m$ $(m>1)$ hoping that they are positive and approach 1 as $m \to \infty$. In Figure 2 we have numerically calculated the density of $\mathcal{C}_2$ within the odd integers up to 63201. Nevertheless, we abandoned this idea having conjectured that the density of each $\mathcal{C}_m$ is still zero. However, the next section gives a way of proving that $c_k/2^k$ goes to zero in arithmetic average over odd integers k.

6. The sets $\mathcal{C}_m$ for large m. In this section, we prove the following result.

THEOREM 9. *Put* $m(k) = \lfloor \exp(4000(\log\log\log k)^3) \rfloor$. *The set of odd positive integers* k *such that* $k \in \mathcal{C}_{m(k)}$ *is of asymptotic density* $1/2$.

In particular, most odd positive integers k belong to $\mathcal{C}_{m(k)}$.

Proof. Let x be large. We put

$$y = (\log\log x)^3.$$

We start by discarding some of the odd positive integers $k \le x$. We start with

$$\mathcal{A}_1 = \{k \le x : q^2 \mid k, \text{ or } q(q-1) \mid k, \text{ or } q^2 \mid \phi(k) \text{ for some prime } q \ge y\}.$$

Clearly, if $n \in \mathcal{A}_1$, then there exists some prime $q \ge y$ such that either $q^2 \mid n$, or $q(q-1) \mid n$, or $q^2 \mid p-1$ for some prime factor p of n, or n is a multiple of

two primes $p_1 < p_2$ such that $q \mid p_i - 1$ for both $i = 1$ and 2. The number of integers in the first category is

$$\le \sum_{y<q\le x^{1/2}} \left\lfloor \frac{x}{q^2} \right\rfloor \le x \sum_{y<q\le x^{1/2}} \frac{1}{q^2} \ll x \int_y^{x^{1/2}} \frac{dt}{t^2} \ll \frac{x}{y} = \frac{x}{(\log\log x)^3} = o(x)$$

as $x \to \infty$. Similarly, the number of integers in the second category is

$$\le \sum_{y<q<x^{1/2}+1} \left\lfloor \frac{x}{q(q-1)} \right\rfloor \ll x \sum_{y\le q\le x^{1/2}+1} \frac{1}{q^2} \ll \frac{x}{y} = \frac{x}{(\log\log x)^3} = o(x)$$

as $x \to \infty$. The number of integers in the third category is

$$\begin{aligned} &\le \sum_{y<q\le x^{1/2}} \sum_{\substack{p\le x \\ p\equiv 1 \,(\mathrm{mod}\, q^2)}} \left\lfloor \frac{x}{p} \right\rfloor \le x \sum_{y<q\le x^{1/2}} \sum_{\substack{p\le x \\ p\equiv 1 \,(\mathrm{mod}\, q^2)}} \frac{1}{p} \\ &\ll x \sum_{y<q\le x^{1/2}} \frac{\log\log x}{\phi(q^2)} \ll x \log\log x \sum_{y<q\le x^{1/2}} \frac{1}{q^2} \\ &\ll \frac{x\log\log x}{y} = \frac{x}{(\log\log x)^2} = o(x) \end{aligned}$$

as $x \to \infty$, while the number of integers in the fourth and most numerous category is

$$\begin{aligned} &\le \sum_{y<q\le x^{1/2}} \sum_{\substack{p_1<p_2<x \\ p_i\equiv 1 \,(\mathrm{mod}\, q),\, i=1,2}} \left\lfloor \frac{x}{p_1p_2} \right\rfloor \le x \sum_{y<q\le x^{1/2}} \sum_{\substack{p_1<p_2<x \\ p_i\equiv 1 \,(\mathrm{mod}\, q),\, i=1,2}} \frac{1}{p_1p_2} \\ &\le x \sum_{y<q\le x^{1/2}} \frac{1}{2}\Bigg(\sum_{\substack{p\le x \\ p\equiv 1 \,(\mathrm{mod}\, q)}} \frac{1}{p} \Bigg)^2 \ll x \sum_{y<q\le x^{1/2}} \left(\frac{\log\log x}{\phi(q)} \right)^2 \\ &\ll x(\log\log x)^2 \sum_{y<q\le x^{1/2}} \frac{1}{q^2} \ll \frac{x(\log\log x)^2}{y} = \frac{x}{\log\log x} = o(x) \end{aligned}$$

as $x \to \infty$. We now let

$$\mathcal{Q} = \{p : t_p \le p^{1/3}\},$$

and let $\mathcal{A}_2$ be the set of $k \le x$ divisible by some $q \in \mathcal{Q}$ with $q > y$. To estimate $\#\mathcal{A}_2$, we begin by estimating the counting function $\#\mathcal{Q}(t)$ of $\mathcal{Q}$ for positive real numbers t. Clearly,

$$2^{\#\mathcal{Q}(t)} \le \prod_{q\in\mathcal{Q}(t)} q \le \prod_{s\le t^{1/3}} (2^s - 1) < 2^{\sum_{s\le t^{1/3}} s} \le 2^{t^{2/3}},$$

so

$$\#\mathcal{Q}(t) \le t^{2/3}. \tag{17}$$

By Abel's summation formula, we now get

$$\#\mathcal{A}_2 \le \sum_{\substack{y\le q\le x\\ q\in\mathcal{Q}}} \left\lfloor \frac{x}{q}\right\rfloor \le x \sum_{\substack{y\le q\le x\\ q\in\mathcal{Q}}} \frac{1}{q} \ll x \int_y^x \frac{d\#\mathcal{Q}(t)}{t}$$
$$\ll \frac{x}{y^{1/3}} = \frac{x}{\log\log x} = o(x)$$

as $x \to \infty$.

Recall now that $P(m)$ stands for the largest prime factor of the positive integer m. Known results from the theory of distribution of smooth numbers show that uniformly for $3 \le s \le t$, we have

$$\Psi(t,s) = \#\{m \le t : P(m) \le s\} \ll t\exp(-u/2), \tag{18}$$

where $u = \log t/\log s$ (see [15, Section III.4]). Thus, putting

$$z = \exp(32(\log\log\log x)^2),$$

we conclude that

$$\Psi(t,y) \ll \frac{t}{(\log\log x)^5} \tag{19}$$

uniformly for large x once $t > z$, because in this case $u = \log t/\log y \ge \frac{32}{3}\log\log\log x$, therefore

$$\frac{u}{2} \ge \frac{16}{3}\log\log\log x,$$

so, in particular, $u/2 > 5\log\log\log x$ for all large x. Furthermore, if $t > Z = \exp((\log\log x)^2)$, then

$$u = \frac{\log t}{\log y} = \frac{(\log\log x)^2}{3\log\log\log x},$$

so $u/2 > 2\log\log x$ once x is sufficiently large. Thus, in this range, inequality (19) can be improved to

$$\Psi(t,y) \ll \frac{x}{\exp(2\log\log x)} \ll \frac{x}{(\log x)^2}. \tag{20}$$

Now for a positive integer m, we write $d(m,y)$ for the largest divisor d of m which is y-smooth, that is, $P(d) \le y$. Let $\mathcal{A}_3$ be the set of $k \le x$ having a prime factor p exceeding z^{10} such that $d(p-1,y) > p^{1/10}$. To estimate $\#\mathcal{A}_3$, we fix a y-smooth number d and a prime p with $z^{10} < p < d^{10}$ such that $p \equiv 1 \pmod d$, and observe that the number of $n \le x$ which are multiples of this prime p is $\le \lfloor x/p\rfloor$. Note also that $d > p^{1/10} > z$. Summing up over

all the possibilities for d and p, we conclude that $\#\mathcal{A}_3$ does not exceed

$$\sum_{\substack{z<d\\ P(d)\le y}}\sum_{\substack{p\le x\\ p\equiv 1 \pmod d}}\left\lfloor\frac{x}{p}\right\rfloor \le x\sum_{\substack{z<d\\ P(d)\le y}}\sum_{\substack{p\le x\\ p\equiv 1 \pmod d}}\frac{1}{p} \ll x\sum_{\substack{z<d\\ P(d)\le y}}\frac{\log\log x}{\phi(d)}$$

$$\ll x(\log\log x)^2\sum_{\substack{z<d\\ P(d)\le y}}\frac{1}{d} \ll x(\log\log x)^2\int_z^x \frac{d\Psi(t,y)}{t}$$

$$\ll x(\log\log x)^2\left(\frac{\Psi(t,y)}{t}\bigg|_{t=z}^{t=x} + \int_z^x \frac{\Psi(t,y)\,dt}{t^2}\right)$$

$$\ll \frac{x}{(\log\log x)^3} + x(\log\log x)^2\int_z^x \frac{\Psi(t,y)\,dt}{t^2}.$$

In the above estimates, we used, aside from the Abel summation formula and inequality (19), also the minimal order of the Euler function $\phi(d)/d \gg 1/\log\log x$ valid for all $d \in [1,x]$. It remains to bound the above integral. For this, we split it at Z and use estimates (19) and (20). In the smaller range, we have

$$\int_z^Z \frac{\Psi(t,y)\,dt}{t^2} \ll \frac{1}{(\log\log x)^5}\int_z^Z \frac{dt}{t} \ll \frac{\log Z}{(\log\log x)^5} \ll \frac{1}{(\log\log x)^3}.$$

In the larger range, we use estimate (20) and get

$$\int_Z^x \frac{\Psi(t,y)\,dt}{t^2} \ll \frac{1}{(\log x)^2}\int_Z^x \frac{dt}{t} \ll \frac{1}{\log x}.$$

Putting these together we get

$$\#\mathcal{A}_3 \ll \frac{x}{(\log\log x)^3} + x(\log\log x)^2\left(\frac{1}{(\log\log x)^3} + \frac{1}{\log x}\right) = o(x)$$

as $x \to \infty$.

Now let $l = d(k, z^{10})$. Put

$$w = \exp(1920(\log\log\log x)^3),$$

and put $\mathcal{A}_4$ for the set of $k \le x$ such that $l > w$. Note that each such k has a divisor $d > w$ such that $P(d) \le z^{10}$. Since for such d we have

$$\frac{\log d}{\log(z^{10})} = 6\log\log\log x,$$

we find that in the range $t \ge w$, putting now $u = \log t/\log(z^{10})$, we have $u/2 > 3\log\log\log x$ for large x, so

$$\Psi(t, z^{10}) < \frac{t}{(\log\log x)^3} \tag{21}$$

uniformly for such t once x is large. Furthermore, if

$$t > Z_1 = \exp(1280 \log\log x (\log\log\log x)^2),$$

then $u = \log t/\log(z^{10}) > 4\log\log x$ and therefore $u/2 > 2\log\log x$. In particular,

$$\Psi(t, z^{10}) \ll \frac{x}{(\log x)^2} \tag{22}$$

in this range. By an argument already used previously, $\#\mathcal{A}_4$ is at most

$$\begin{aligned}
&\leq \sum_{\substack{w<d<x\\ P(d)\le z^{10}}} \left\lfloor \frac{x}{d} \right\rfloor \leq x \sum_{\substack{w<d<x\\ P(d)\le z^{10}}} \frac{1}{d} \ll x \int_w^x \frac{d\Psi(t, z^{10})}{t} \\
&\ll x\left(\frac{\Psi(t, z^{10})}{t}\bigg|_{t=w}^{t=x} + \int_w^x \frac{\Psi(t, z^{10})\, dt}{t^2} \right) \\
&\ll x\left(\frac{1}{(\log\log x)^3} + \int_w^{Z_1} \frac{\Psi(t, z^{10})\, dt}{t^2} + \int_{Z_1}^x \frac{\Psi(t, z^{10})\, dt}{t^2} \right) \\
&\ll x\left(\frac{1}{(\log\log x)^3} + \frac{\log Z_1}{(\log\log x)^3} + \frac{\log x}{(\log x)^2} \right) = o(x)
\end{aligned}$$

as $x \to \infty$, where the above integral was estimated by splitting it at Z_1 and using estimates (21) and (22) for the lower and upper ranges respectively.

Let $\mathcal{A}_5$ be the set of $k \le x$ which are coprime to all primes $p \in [y, z^{10}]$. By the Brun method,

$$\#\mathcal{A}_5 \ll x \prod_{y\le q\le z} \left(1 - \frac{1}{q}\right) \ll \frac{x \log y}{\log z} \ll \frac{x}{\log\log\log x} = o(x)$$

as $x \to \infty$.

We next let $\mathcal{A}_6$ be the set of $k \le x$ such that $P(k) < w^{100}$. Clearly,

$$\#\mathcal{A}_6 = \Psi(x, w^{100}) \ll x \exp\left(-c_1 \frac{\log x}{(\log\log\log x)^3} \right) = o(x)$$

as $x \to \infty$, where $c_1 = 1/384000$.

Finally, we let

$$\mathcal{A}_7 = \{k \le x : dp \,|\, k \text{ for some } p \equiv 1 \pmod d \text{ and } p < d^3\}.$$

Assume that $k \in \mathcal{A}_7$. Then there is a prime factor p of k and a divisor d of $p-1$ of size $d > p^{1/3}$ such that $dp \,|\, k$. If we fix d and p, the number of such

$n \le x$ is $\le \lfloor x/(dp) \rfloor$. Thus,

$$\begin{aligned}\#\mathcal{A}_7 &\le \sum_{y\le p\le x} \sum_{\substack{d\mid p-1\\ d>p^{1/3}}} \left\lfloor \frac{x}{dp} \right\rfloor \le x \sum_{y\le p\le x} \frac{1}{p} \sum_{\substack{d\mid p-1\\ d>p^{1/3}}} \frac{1}{d} \\ &\ll \sum_{y\le p\le x} \frac{1}{p}\left(\frac{\tau(p-1)}{p^{1/3}}\right) \ll x \sum_{y\le p\le x} \frac{\tau(p-1)}{p^{1+1/3}} \ll x \sum_{y\le p\le x} \frac{1}{p^{5/4}} \\ &\ll x \int_y^x \frac{dt}{t^{5/4}} \ll \frac{x}{y^{1/4}} = \frac{x}{(\log\log x)^{3/4}} = o(x)\end{aligned}$$

as $x \to \infty$. Here, we used $\tau(m)$ for the number of divisors of the positive integer m and the fact that $\tau(m) \ll_\varepsilon m^\varepsilon$ for all $\varepsilon > 0$ (with the choice of $\varepsilon = 1/12$).

From now on, $k \le x$ is odd and not in $\bigcup_{1\le i\le 7} \mathcal{A}_i$. From what we have seen above, most odd integers k below x have this property. Then $l \le w$ because $k \notin \mathcal{A}_4$. Further, k/l is square-free because $k \notin \mathcal{A}_1$. Moreover, if $p \,|\, k/l$, then $p > z^{10} > y$, therefore $t_p > p^{1/3}$ because $k \notin \mathcal{A}_2$. Since $k \notin \mathcal{A}_3$, we get $d(p-1, y) < p^{1/10}$, so $t'_p = t_p/\gcd(t_p, d(p-1,y)) > p^{1/3-1/10} > p^{1/5}$ for all such p. Moreover, t'_p is divisible only by primes $> z > y$, so if p_1 and p_2 are distinct primes dividing k/l, then t'_{p_1} and t'_{p_2} are coprime because $k \notin \mathcal{A}_1$. Finally, $l > y$ because $k \notin \mathcal{A}_5$. Furthermore, for large x we have $w > y$, so $k > l$ and in fact k/l is divisible by a prime $> w^{100}$ because $k \notin \mathcal{A}_6$.

We next put $n = \operatorname{lcm}[d(\phi(k), y), \phi(l)]$. We let n_0 stand for the minimal positive integer such that $n_0 \equiv -k+1 \pmod{\phi(l)}$ and let $m = n_0 + l\phi(l)$. Note that

$$m \le 2l\phi(l) \le 2w^2 = 2\exp(3840(\log\log\log x)^3).$$

We may also assume that $k > x/\log x$ since there are only at most $x/\log x = o(x)$ positive integers k failing this property. Since $k > x/\log x$, we get

$$m < 2\exp(3840(\log\log\log x)^3) < \lfloor \exp(4000(\log\log\log k)^3) \rfloor = m(k)$$

for large x. We will now show that this value for m works. First of all $m + k = n_0 + l\phi(l) + k \equiv 1 \pmod{\phi(l)}$ so

$$\begin{aligned}2^{m+k} - 1 \equiv 1 \equiv 2^{\phi(l)-1} + 2^{\phi(l)-2} + \cdots + 2^{\phi(l)-(n_0-1)} + 2^{\phi(l)-(n_0-1)} + \\ + 2^{x_1 n} + \cdots + 2^{x_t n} \pmod{l},\end{aligned}$$

where $t = l\phi(l)$ and $x_1, \ldots, x_t$ are any nonnegative integers. Let

$$U = 2^{m+k} - 1 - 2^{\phi(l)-1} - \cdots - 2^{\phi(l)-(n_0-1)} - 2^{\phi(l)-(n_0-1)}.$$

Then

$$U \equiv \sum_{i=1}^{t} 2^{x_i \phi(l)} \pmod{l}$$

for any integers $x_1, \dots, x_t$. Let p be any prime divisor of k/l. Clearly, $\gcd(t_p, n) = d(t_p, y)$, because $t_p \mid \phi(k)$ and $n \notin \mathcal{A}_1$. In particular,

$$t'_p = \frac{t_p}{\gcd(t_p, n)} \geq \frac{t_p}{\gcd(d(\phi(k), y), p-1)} \geq p^{1/3-1/10} > p^{1/5}.$$

Let $X = \{2^{jn} \pmod{p}\}$. Clearly, the order of 2^n modulo p is precisely t'_p. So, $\#X = t'_p > p^{1/5}$. A recent result of Bourgain, Glibichuk and Konyagin (see Theorem 5 in [2]) shows that there exists absolute constant T such that for all integers λ, the equation

$$\lambda \equiv 2^{x_1 n} + \cdots + 2^{x_t n} \pmod{p}$$

has integer solutions $0 \leq x_1, \dots, x_t < t'_p$ once $t > T$. In fact, for large p the number of such solutions

$$N(t, p, \lambda) = \#\{(x_1, \dots, x_t) : 0 \leq x_1, \dots, x_t \leq t_p\}$$

satisfies

$$N(t, p, \lambda) \in \left[\frac{\#X^t}{2p}, \frac{2\#X^t}{p}\right]$$

independently of the parameter λ and uniformly in the number t. In particular, if we let $N_1(t, p, \lambda)$ be the number of such solutions with $x_i = x_j$ for some $i \neq j$, then $N_1(t, p, \lambda) \ll t^2 \#X^{t-1}/p$. Indeed, the pair (i, j) with $i \neq j$ can be chosen in $O(t^2)$ ways, and the common value of $x_i = x_j$ can be chosen in $\#X$ ways. Once these two data are chosen, the number of ways of choosing $x_s \in \{0, 1, \dots, t'_p - 1\}$ with $s \in \{1, \dots, t\} \setminus \{i, j\}$ such that

$$\lambda - 2^{x_i n} - 2^{x_j n} \equiv \sum_{\substack{1 \leq s \leq t \\ s \neq i,j}} 2^{x_s n} \pmod{p}$$

is $N(t-2, p, \lambda - 2^{x_i n} - 2^{x_j n}) \ll \#X^{t-2}/p$ for $t > T + 2$. In conclusion, if all solutions $x_1, \dots, x_t$ have two components equal, then $p^{1/5} \ll \#X \ll t^2$, so $p \ll t^{10}$. For us, $t \leq 2w^2$, so $p \ll w^{20}$. Since $P(k) = P(k/l) > w^{100}$, it follows that at least for the largest prime factor $p = P(k)$ of k, we may assume that $x_1, \dots, x_t$ are all distinct modulo p for a suitable value of λ.

We apply the above result with $\lambda = U$, $t = l\phi(l)$ (note that since $t > y$, it follows that $t > T + 2$ does indeed hold for large values of x), and write $\mathbf{x}(p) = (x_1(p), \dots, x_t(p))$ for a solution of

$$U \equiv 2^{x_1(p) n} + \cdots + 2^{x_t(p) n} \pmod{p}, \qquad 0 \leq x_1(p) \leq \cdots \leq x_t(p) < t'_p.$$

We also assume that for at least one prime (namely the largest one) the $x_i(p)$'s are distinct. Now choose integers $x_1, \ldots, x_t$ such that

$$x_i \equiv x_i(p) \pmod{t'_p}$$

for all $p \mid k/l$. This is possible by the Chinese Remainder Lemma since the numbers t'_p are coprime as p varies over the distinct prime factors of k/l. We assume that for each i, x_i is the minimal nonnegative integer in the corresponding arithmetic progression modulo $\prod_{p|k/l} t'_p$. Further, since $nx_i(p)$ are distinct modulo t'_p when $p = P(k)$, it follows that nx_i are also distinct for $i = 1, \ldots, t$. Hence, for such x_i's, $U - \sum_{i=1}^{t} 2^{x_i n}$ is a multiple of all $p \mid k/l$, and since k/l is square-free, we get $U \equiv \sum_{i=1}^{t} 2^{x_i n} \pmod{k/l}$. But this congruence is also valid modulo l, so it is valid modulo $\operatorname{lcm}[l, k/l] = k$, since l and k/l are coprime. Thus,

$$U \equiv \sum_{i=1}^{t} 2^{x_i n} \pmod{k},$$

or

$$2^{k+m-1} - 1 \equiv 2^{\phi(l)-1} + \cdots + 2^{\phi(l)-(n_0-1)} + 2^{\phi(l)-(n_0-1)} + \sum_{i=1}^{t} 2^{x_i n} \pmod{k}.$$

As we have said, the numbers $x_i n$ are distinct and they can be chosen of sizes at most $n \operatorname{lcm}[t'_p : p \mid k/l] \le \phi(k) \le k$. Finally, nx_i are divisible by $\phi(l)$ whereas none of the numbers $\phi(l) - j$ for $j = 1, \ldots, n_0 - 1$ is unless $n_0 = 1$. Thus, assuming that $n_0 \neq 1$, we infer that all the $m = t + n_0$ exponents are distinct except $\phi(l) - (n_0 - 1)$ that appears twice.

Let us first justify that $n_0 \neq 1$. Recalling the definition of n_0, if this were not so then $\phi(l) \mid k$. However, we have just said that l has a prime factor $p > y$. If $\phi(l) \mid k$, then k is divisible by both p and $p - 1$ for some $p > y$, and this is impossible since $n \notin \mathcal{A}_1$.

Finally, to deal with the repetition of the exponent $\phi(l) - (n_0 - 1)$, we replace it by $\phi(l) - (n_0 - 1) + t_k$, where as usual t_k is the order of 2 modulo k. We show that with this replacement, all the exponents are distinct. Indeed, the replacement will not change the value of $2^{\phi(l)-(n_0-1)+t_k} \pmod{k}$. Assume that after this replacement, $\phi(l) - (n_0 - 1) + t_k$ is still one of the remaining exponents. If it has become a multiple of n, it is in particular divisible by t_p for all primes $p \mid l$. Since $t_p \mid t_k$ and $t_p \mid \phi(l)$ for all primes $p \mid l$, we find that $t_p \mid n_0 - 1$, so $t_p \mid k$. Since l is divisible by some prime $p > y$ (because $k \notin \mathcal{A}_5$), we see that $t_p \mid k$. Since $k \notin \mathcal{A}_2$, we get $t_p > p^{1/3}$. Thus, k is divisible by a prime $p > y$ and a divisor d of $p - 1$ with $d > p^{1/3}$, and this is false since $n \notin \mathcal{A}_7$. Hence, this is impossible, so it must be the case that

$$\phi(l) - (n_0 - 1) + t_k \in \{\phi(l) - 1, \ldots, \phi(l) - (n_0 - 1)\}.$$

This shows that $t_k \le n_0 \le l\phi(l) \le 2^{10}w^{20}$. However, t_k is a multiple of $t_{P(k)} \ge P(k)^{1/3}$, showing that $P(k) \le 2^{30}w^{60}$, which is false for large x since $k \notin \mathcal{A}_6$. Thus, the new exponents are all distinct for our values of k. As far as their sizes go, note that since k has at least two odd prime factors, it follows that $t_k \mid \phi(k)/2$, therefore

$$\phi(l) - (n_0 - 1) + t_k \le w + \phi(k)/2 < w + k/2 < k$$

since $k > 2w$ for large x. Thus, we have obtained a representation of $2^{k+m}-1$ modulo m of the form

$$2^{j_1} + \cdots + 2^{j_m} \pmod{k}$$

where $0 \le j_1 < \cdots < j_m \le k$, which shows that $k \in \mathcal{C}_m$. Since $m \le m(k)$ and $\mathcal{C}_m \subseteq \mathcal{C}_{m(k)}$, the conclusion follows. ∎

Remark 10. The above proof shows that in fact the number of odd $k < x$ such that $k \notin \mathcal{C}_{m(k)}$ is $O(x/\log\log\log x)$.

Corollary 11. *For large x, the inequality $c_k/2^k < 2^{m(k)-\log k/\log 2}$ holds for all odd $k < x$ with at most $O(x/\log\log\log x)$ exceptions.*

Proof. This follows from the fact that $c_k = a_k/k \le 2^{k+m}/k$, where $k \in \mathcal{C}_m$ (see Theorem 6), together with Theorem 9 and Remark 10. ∎

Corollary 12. *The estimate*

$$\frac{1}{x} \sum_{\substack{1 \le k \le x \\ k \text{ odd}}} \frac{c_k}{2^k} = O\left(\frac{1}{\log\log\log x}\right)$$

holds for all x.

Proof. If $k \le x/\log x$ is odd, then $c_k/2^k \le 1$, so

$$\sum_{\substack{k \le x/\log x \\ k \text{ odd}}} \frac{c_k}{2^k} \le \frac{x}{\log x}.$$

If $k \in [x/\log x, x]$ but $k \notin \mathcal{C}_{m(k)}$, then still $c_k/2^k \le 1$ and, by Corollary 11, the number of such k's is $O(x/\log\log\log x)$. Thus,

$$\sum_{\substack{k \in [x/\log x, x] \\ k \notin \mathcal{C}_{m(k)} \\ k \text{ odd}}} \frac{c_k}{2^k} \ll \frac{x}{\log\log\log x}.$$

For the remaining odd values of $k \le x$, we have

$$\frac{c_k}{2^k} \le 2^{m(k)-\log k/\log 2},$$

so it suffices to show that

$$2^{m(k)-\log k/\log 2} < \frac{1}{\log\log\log x},$$

which is equivalent to

$$\log k/\log 2 - m(k) > \log\log\log\log x/\log 2,$$

which in turn is implied by

$$\log(x/\log x) - \log\log\log\log x > (\log 2)\exp(4000(\log\log\log x)^3),$$

and this is certainly true for large x. Thus, indeed,

$$\sum_{\substack{1\le k\le x\\ k\,\text{odd}}} \frac{c_k}{2^k} = O\bigg(\frac{x}{\log\log\log x}\bigg),$$

which is what we wanted to prove. ∎

In particular,

$$\frac{1}{x}\sum_{\substack{k\le x\\ k\,\text{odd}}} \frac{c_k}{2^k} = o(1) \tag{23}$$

as $x \to \infty$. One can adapt these techniques to deduce that the whole sequence $c_k/2^k$ is convergent to 0 in arithmetic average. In order to do so, the sets $\mathcal{C}_m$ should be suitably modified and an analog of Theorem 9 for these new sets should be proved. We leave this for a subsequent work.

7. Existence and bounds for a_k in base $q > 2$. Let $q \ge 2$ be a fixed integer and let x be a positive real number. Put

$$V_k(x) = \{0 \le n < x : s_q(n) = k\},$$
$$V_k(x; h, m) = \{0 \le n < x : s_q(n) = k,\ n \equiv h \pmod m\}.$$

Mauduit and Sárközy proved in [12] that if $\gcd(m, q(q-1)) = 1$, then there exists some constant c_0 depending on q such that if we put

$$l = \min\{k, (q-1)\lfloor \log x/\log q\rfloor - k\},$$

then $V_k(x)$ is well distributed in residue classes modulo m provided that $m < \exp(c_0 l^{1/2})$.

Taking $m = k$ and $h = 0$, we deduce that if $k < \exp(c_0 l^{1/2})$, then

$$\#V_k(x; 0, k) = (1 + o(1))\#V_k(x)/k$$

as $x \to \infty$ uniformly in our range for k. The condition on k is equivalent to $\log k \ll l^{1/2}$, which is implied by $k + O((\log k)^2) \ll \log x$. Thus, we have the following result.

LEMMA 13. *Let $q \ge 2$ be fixed. There exists a constant c_1 such that if k is any positive integer with $\gcd(k, q(q-1)) = 1$, then $V_k(x)$ is well distributed in arithmetic progressions of modulus k whenever $x > \exp(c_1 k)$.*

Corollary 2 of [12] implies that if

$$\Delta = \left| \frac{q-1}{2\log q} \log x - k \right| = o(\log x) \quad \text{as } x \to \infty, \tag{24}$$

then

$$\#V_k(x) = \frac{x}{(\log x)^{1/2}} \exp\left(-c_3 \frac{\Delta^2}{\log x} + O\left(\frac{\Delta^3}{(\log x)^2} + \frac{1}{(\log x)^{1/2}} \right) \right)$$

holds with some explicit constant c_3 depending on q. As a corollary, we deduce the following result.

LEMMA 14. *If condition* (24) *is satisfied, then* $V_k(x) \neq \emptyset$.

In case k and q are coprime but k and $q-1$ are not, we may apply instead Theorem B of [11] with $m = k$ and $h = 0$ to arrive at a similar result.

LEMMA 15. *Assume that* $q \geq 2$ *is fixed. There exists a constant* c_4 *depending only on* q *such that if* k *is a positive integer with* $\gcd(k, q) = 1$, *and* $x \geq \exp(c_4 k)$, *then* $V_k(x; 0, k) \neq \emptyset$.

One can even remove the coprimality condition on q and k. Assume that x is sufficiently large such that

$$\Delta \leq c_5 (\log x)^{5/8}, \tag{25}$$

where c_5 is some suitable constant depending on q. Using Theorem C and Lemma 5 of [11] with $m = k$ and $h = 0$, we obtain the following result.

LEMMA 16. *Assume that both estimates* (25) *and* $k < 2^{(\log x)^{1/4}}$ *hold. Then* $V_k(x; 0, k) \neq \emptyset$.

A sufficient condition on x for Lemma 16 above to hold is that $x > \exp(c_6 k)$, where c_6 is a constant that depends on q. Putting Lemmas 15 and 16 together we obtain the next theorem.

THEOREM 17. *For all* $q \geq 2$ *there exists a constant* c_6 *depending on* q *such that for all* $k \geq 1$ *there exists* $n \leq \exp(c_6 k)$ *with* $s_q(kn) = k$.

Consequently, $a_k = \exp(O(k))$ for all k, and in particular it is nonzero. The example of Lemma 18 below shows that $a_k = \exp(o(k))$ does not always hold as $k \to \infty$.

LEMMA 18. *If* $q > 2$, *then*

$$a_{q^m} = q^m (2q^{(q^m-1)/(q-1)} - 1).$$

If $q = 2$, *then* $a_{2^m} = 2^m (2^{2^m} - 1)$.

Proof. The fact that $s_q(a_{q^m}) = q^m$ for all $q \geq 2$ is immediate. We now show the minimality of the given a_{q^m} with this property. Define

$$\alpha_m = \frac{q^m - 1}{q - 1}.$$

Note that every digit of $q^{\alpha_m} - 1$ in base q is maximal, so $q^{\alpha_m} - 1$ is minimal such that $s_q(q^{\alpha_m} - 1) = q^m - 1$. Since

$$q^{\alpha_m} - 1 = (q-1)q^{\alpha_m - 1} + (q-1)q^{\alpha_m - 2} + \cdots + (q-1),$$

a_{q^m} must contain the least term q^t, where $t > \alpha_m - 1$ is such that its sum of digits is q^m and $q^m \mid a_{q^m}$. The least term is obviously q^{α_m}, and it just happens that a_{q^m} such defined satisfies the above mentioned conditions. ∎

References

[1] W. D. Banks, M. Z. Garaev, F. Luca and I. E. Shparlinski, *Uniform distribution of fractional parts related to pseudoprimes*, Canad. J. Math., to appear.

[2] J. Bourgain, A. A. Glibichuk and S. V. Konyagin, *Estimates for the number of sums and products and for exponential sums in finite fields of prime order*, J. London Math. Soc. 73 (2006), 380–398.

[3] T. Cai, *On 2-Niven numbers and 3-Niven numbers*, Fibonacci Quart. 34 (1996), 118–120.

[4] C. N. Cooper and R. E. Kennedy, *On consecutive Niven numbers*, ibid. 21 (1993), 146–151.

[5] J.-M. De Koninck and N. Doyon, *On the number of Niven numbers up to x*, ibid. 41 (2003), 431–440.

[6] J.-M. De Koninck, N. Doyon and I. Kátai, *On the counting function for the Niven numbers*, Acta Arith. 106 (2003), 265–275.

[7] H. G. Grundman, *Sequences of consecutive Niven numbers*, Fibonacci Quart. 32 (1994), 174–175.

[8] D. R. Heath-Brown and S. Konyagin, *New bounds for Gauss sums derived from kth powers*, Quart. J. Math. 51 (2000), 221–235.

[9] H.-K. Indlekofer and N. M. Timofeev, *Divisors of shifted primes*, Publ. Math. Debrecen 60 (2002), 307–345.

[10] L. C. Lagarias and A. M. Odlyzko, *Effective versions of Chebotarev's density theorem*, in: Algebraic Number Fields, A. Fröhlich (ed.), Academic Press, New York, 1977, 409–464.

[11] C. Mauduit, C. Pomerance and A. Sárközy, *On the distribution in residue classes of integers with a fixed digit sum*, Ramanujan J. 9 (2005), 45–62.

[12] C. Mauduit and A. Sárközy, *On the arithmetic structure of the integers whose sum of digits is fixed*, Acta Arith. 81 (1997), 145–173.

[13] I. Niven, H. S. Zuckerman and H. L. Montgomery, *An Introduction to the Theory of Numbers*, 5th ed., Wiley, 1991.

[14] F. Pappalardi, *On Hooley's theorem with weights*, Rend. Sem. Mat. Univ. Politec. Torino 53 (1995), 375–388.

[15] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Univ. Press, 1995.

[16] I. Vardi, *Niven numbers*, §2.3 in: Computational Recreations in Mathematics, Addison-Wesley, 1991, 19 and 28–31.

Department of Applied Mathematics
Naval Postgraduate School
Monterey, CA 93943, U.S.A.
E-mail: half@nps.edu
pstanica@nps.edu

Department of Mathematics
Columbus State University
Columbus, GA 31907, U.S.A.
E-mail: ionascu_eugen@colstate.edu

Instituto de Matemáticas
Universidad Nacional Autónoma de México
C.P. 58089
Morelia, Michoacán, México
E-mail: fluca@matmor.unam.mx

Received on 15.8.2007 (5493)