PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 85, Number 1, May 1982

ON THE DISTRIBUTION OF SELF-NUMBERS

U. ZANNIER

ABSTRACT. Self-numbers are those integers which cannot be expressed as $a + f(a)$, where $f(a)$ denotes the sum of the digits of a in a given scale. Here I prove that the number of self-numbers less than or equal to a large number x equals $Lx + O(\log^2 x)$, where L is a strictly positive constant.

1. Introduction. Many papers have been devoted to "digitaddition sequences", i.e. sequences $\{a_n\}$ where $a_{n+1} = a_n + f(a_n)$, and $f(a)$ is the sum of the digits of a in a fixed scale (see for example [**1, 2, 3**]). In particular self-numbers, numbers that cannot be written in the form $n + f(n)$, have been studied, and it has been proved, for example, that there are infinitely many of them in any scale (see [**2**]).

However only a few papers deal with asymptotic results; here the following theorem is proved.

THEOREM. *Let $A(x)$ denote the number of self-numbers (in the scale of 2) less than or equal to x. Then we have the formula*

$$A(x) = Lx + O((\log x)f([x]))$$

where $L > 0$, $f(a)$ is as above, and the constant implicit in the O is effectively computable.

An analogous theorem holds in any scale, the proof being a little more complicated but completely similar.

2. Notation. If S is any finite set we denote the number of its elements by $|S|$. We denote by $[x]$ the largest integer less than or equal to x.

If $a = \sum_j a_j 2^j$ is the expression of a in the binary scale we put $f(a) = \sum_j a_j$.

Finally we set $A'_j(x) = \{n \leq x\colon n = r + f(r)$ has exactly j solutions$\}$ and $A_j(x) = |A'_j(x)|$. It is clear that $A_0(x)$ is the number of self-numbers less than or equal to x.

3. Proof of the theorem. We separate the proof in two parts: in the first one, we give only a lower bound for $A_0(x)$ of the type $A_0(x) > Dx$, with $D > 0$, and in the second one we prove $A_0(x) = Lx + O((\log x) f([x])) = Lx + O(\log^2 x)$. By the first part, L must be > 0 and so the demonstration will be complete.

Let $B(n) = \{a\colon a + f(a) = n\}$. We have $B(n) \cap B(m) = \varnothing$ when $n \neq m$, and $\bigcup_{n \leq x} B(n) \subset [1, x]$ and so $\sum_{n \leq x} |B(n)| \leq [x]$.

Received by the editors July 28, 1980 and, in revised form, December 8, 1980.
1980 *Mathematics Subject Classification.* Primary 10A30, 10A99.

©1982 American Mathematical Society
0002-9939/82/0000-1046/$02.25

But $\Sigma_{n \leq x} | B(n) | = A_1(x) + 2A_2(x) + 3A_3(x) + \ldots$, and comparing these results with the obvious equation

$$[x] = A_0(x) + A_1(x) + A_2(x) + \ldots$$

we obtain

$$A_0(x) \geq A_2(x) + 2A_3(x) + 3A_4(x) + \ldots . \tag{1}$$

We shall now construct many numbers n such that $| B(n) |$ is at least 2, hence, using inequality (1), we will obtain the desired lower bound.

We note that $\cup_{j \geq 2} A'_j(x)$ is not empty if $x \geq 5$ (in fact we have for instance $5 = 3 + f(3) = 4 + f(4)$). Pick then $n_0 \in \cup_{j \geq 2} A'_j(x)$ (for example, $n_0 = 5$). Then there exist a, a' with $a \neq a'$ and

$$n_0 = a + f(a) = a' + f(a').$$

Let k be so large that $2^k > \max(a, a')$ and set

$$a_n = a + 2^k n \qquad a'_n = a' + 2^k n.$$

Then it is clear that $f(a_n) = f(a) + f(2^k n)$ and similarly for a'.

Let us consider the sets

$$C(y) = \{n_0 + 2^k n + f(2^k n), n \leq y\} = \{m_1, m_2, \ldots, m_c\}$$

with $c = c(y) = | C(y) |$.

Suppose that $m_j = n_0 + 2^k n_{jv} + f(2^k n_{jv})$, $v = 1, 2, \ldots, r_j$. We have $[y] = r_1 + r_2 + \cdots + r_c$. We show that $| B(m_j) |$ is at least $2r_j$ (in particular $| B(m_j) | \geq 2$).

In fact $m_j = a_{jv} + f(a_{jv}) = a'_{jv} + f(a'_{jv})$, $v = 1, 2, \ldots, r$, and the a_s, a'_s are all distinct because a, $a' < 2^k$.

It follows that $| B(m_1) | + | B(m_2) | + \cdots + | B(m_c) | \geq 2[y]$ and setting $H_b = | \{i \leq c : | B(m_i) | = b\} |$, we have $\Sigma_b bH_b \geq 2[y]$.

Set $M = \max(m_j)$. Then $m_j \in A'_{|B(m_j)|}(M)$ and we obtain

$$\sum_{b \geq 2} (b-1) A_b(M) \geq \sum_b (b-1) H_b \geq \frac{1}{2}\Big(\sum_b bH_b\Big) \geq [y].$$

Now inequality (1) implies

$$A_0(M) \geq [y]. \tag{2}$$

From the trivial inequality $f(a) \leq \log 2a/\log 2$ we obtain

$$m_j \leq n_o + 2^k y + \log 2^{k+1} y/\log 2 \leq Ky$$

if K is sufficiently large, so that by (2) we have $A_0(Ky) \geq [y]$ and we conclude that there exists a constant $D > 0$ such that $A_0(y) \geq Dy$. This completes the first part.

We shall now obtain a formula which says that $A_0(x)$ is "almost additive" in a sense that will be clear in a moment.

Fix an integer k and choose n such that

$$2^k + k + 2 \leq n < 2^{k+1}. \tag{3}$$

We want to prove that n is a self-number if and only if $n' = n - 2^k - 1$ is (cf. [**2**]).

In fact let $n' = a + f(a)$. Then $a \leq 2^k - 1$ and it follows that $f(a + 2^k) = f(a) + 1$ and $n = n' + 2^k + 1 = a + 2^k + f(a + 2^k)$.

Conversely suppose that $n = a + f(a)$. Then

$$f(a) \leqslant \log 2a/\log 2 < \log 2^{k+2}/\log 2 = k + 2$$

whence $2^k < a < 2^{k+1}$.

We obtain $f(a - 2^k) = f(a) - 1$ and $n' = a - 2^k + f(a - 2^k)$, whence n is a self-number if n' is.

What we have proved means that if m satisfies (3) then

$$A_0(m) - A_0(2^k + k + 2) = A_0(m - 2^k - 1) - A_0(k + 1)$$

and so

$$A_0(m) = A_0(m - 2^k) + A_0(2^k) + O(k) \tag{4}$$

and the same estimate obviously holds also when $2^k \leqslant m \leqslant 2^k + k + 2$.

From (4) it follows in particular that

$$A_0(2^{k+1}) = 2A_0(2^k) + O(k) \tag{5}$$

and iteration of this formula gives

$$A_0(2^{k+s}) = 2^s A_0(2^k) + O\left(\sum_{r=1}^{s} 2^r(k + s - r)\right). \tag{6}$$

Now we have

$$\sum_{r=1}^{s} 2^r k + \sum_{r=1}^{s} 2^r(s - r) = (2^{s+1} - 2)k + \sum_{r=0}^{s-1} 2^{s-t} t$$

$$= O\left(2^s k + 2^s \sum_{r=0}^{\infty} t 2^{-t}\right) = O(2^s k + 2^s) = O(2^s k)$$

whence after division by 2^{k+s} (6) becomes

$$A_0(2^{k+s})/2^{k+s} = A_0(2^k)/2^k + O(k/2^k). \tag{7}$$

This implies in particular that the sequence $A_0(2^r)/2^r$ is Cauchy, so that $A_0(2^r)/2^r \to L$. Now, passing to the limit for $s \to \infty$ in (7) we deduce that $A_0(2^k) = L2^k + O(k)$.

Now, making use of (4) we may write, when $m = \sum_{j=0}^{k} a_j 2^j$,

$$A_0(m) = \sum_{j=0}^{k} A_0(a_j 2^j) + O(kf(m)) = \sum_{j=0}^{k} a_j A_0(2^j) + O((\log m) f(m))$$

$$= L\sum_0^k a_j 2^j + O\left(\sum_0^k a_j j\right) + O((\log m) f(m)) = Lm + O((\log m) f(m))$$

and the theorem is proved.

4. Remarks. When a particular scale has been chosen one can avoid the use of the first part; in fact if $L = 0$ we should have $A_0(m) \leqslant K(\log m) f(m) < K' \log m$ for some explicit K'. In case this is false, its falsity can be verified with a finite number of operations, so that the result in the first part becomes really essential only if one deals with the general case.

We note also that the constant L has not been determined, but, using the theorem with an explicit constant in the O, one may obviously calculate it with any degree of accuracy. I will give here some examples.

From the proof of our theorems it follows easily that, when the scale is g, one has the estimate

$$\left|\frac{A_0(m)}{m} - L\right| \leqslant 2\,\frac{\log mg^3}{\log g}\,f(m)\,\frac{1}{m} \tag{8}$$

where of course L depends on g.

Here are some results when $g = 2, 4, 10$. (I have not made computations for odd g, since in this case it is not difficult to show that self-numbers are precisely the odd numbers. I omit for simplicity the easy proof.)

$g = 2$

m	$A_0(m)$	$A_0(m)/m$
262,144	66,237	0.252674
524,288	132,470	0.252666
786,432	198,704	0.252665
$1{,}048{,}576 = 2^{20}$	264,938	0.252665

$g = 4$

m	$A_0(m)$	$A_0(m)/m$
262,144	54,917	0.209491
524,288	109,827	0.209478
786,432	164,737	0.209473
1,048,576	219,647	0.209471

$g = 10$

m	$A_0(m)$	$A_0(m)/m$
250,000	24,451	0.097804
500,000	48,896	0.097792
750,000	73,340	0.097787
1,000,000	97,785	0.097785

Now let us set in formula (8) $m = 1{,}048{,}576$ when $g = 2$ or 4 and $m = 1{,}000{,}000$ when $g = 10$. In these cases $f(m) = 1$ and we easily obtain

$$\left|\frac{A_0(m)}{m} - L\right| < 10^{-4}$$

whence $L = 0.252, 0.209, 0.097$ when $g = 2, 4, 10$ respectively, and where the values are correct to three decimals.

I want to express my thanks to Professor Stolarsky for his kind assistance and also to my friends S. Bussino, M. Dell'Orso and R. Martinolli for the above and other computations relative to the problem of self-numbers.

References

1. M. Gardner, *Mathematical games*, Sci. Amer. **232** (1975), 113–114.
2. B. S. Recaman, *Solution to problem* E 2408, Amer. Math. Monthly **81** (1974), 407.
3. K. B. Stolarsky, *The sum of a digitaddition series*, Proc. Amer. Math. Soc. **59** (1976), 1–5.

Department of Mathematics, Scuola Normale Superiore, Pisa, Italy