THE FIRST FUNCTION AND ITS ITERATES

CARL POMERANCE

For Ron Graham on his 80th birthday.

Abstract. Let $s(n)$ denote the sum of the positive divisors of n except for n iteself. Discussed since Pythagoras, s may be the first function of mathematics. Pythagoras also suggested iterating s, so perhaps considering the first dynamical system. The historical legacy has left us with some colorful and attractive problems, mostly still unsolved. Yet the efforts have been productive in the development of elementary, computational, and probabilistic number theory. In the context of the Catalan–Dickson conjecture and the Guy–Selfridge counter conjecture, we discuss the geometric mean of the numbers $s(s(2n))/s(2n)$, thus extending recent work of Bosma and Kane. We also show that almost all odd numbers are in the image of every iterate of s, and we establish an asymptotic formula for the size of the set $s^{-1}(n)$ when n is odd.

1. Introduction

Let $\sigma(n)$ denote the sum of the natural divisors of the positive integer n. Let $s(n)$ be the sum of only the proper divisors of n, so that $s(n) = \sigma(n) - n$. "Perfect" and "amicable" numbers are attributed to Pythagoras. A perfect number is one, like 6, where $s(n) = n$, and an amicable number, like 220, is not perfect, but satisfies $s(s(n)) = n$. (The name "amicable" stems from the pair of numbers n and $s(n) = m$, where s of one is the other.) Euclid found the formula $2^{p-1}(2^p - 1)$ that gives perfect numbers whenever the second factor is prime, and Euler proved that all even perfect numbers are given by this formula. No odd perfect numbers are known. The search for even perfect numbers spurred theoretical developments, such as the Lucas–Lehmer primality test, a forerunner of all of modern primality testing (see [25]).

The ancient Greeks also distinguished two types of non-perfect numbers, the "deficient" ones, where $s(n) < n$, and the "abundant" ones, where $s(n) > n$. This concept spurred the development of probabilistic number theory, when Davenport showed that the sets of deficient and abundant numbers each have an asymptotic density, and more generally showed that $s(n)/n$ has a distribution function.

Prominent among the many unsolved problems about s is the century-old Catalan–Dickson conjecture ([3], [4]). This asserts that starting from any positive integer n and iterating s, one arrives eventually at 1, then stopping at 0, or one enters a cycle, such as a 1-cycle (perfect numbers), a 2-cycle (amicable pairs), or a higher order cycle. That is, the conjecture asserts that every orbit is bounded. This conjecture has helped to spur on modern factorization algorithms (since one needs the prime factorization of n to compute $s(n)$). The first number in doubt is 276, where thousands of iterates have been computed, reaching beyond 175 decimal digits. Although we know of no unbounded sequences of this type, Guy and Selfridge [11] came up with a "counter" conjecture, namely that for almost all[1] even seeds, the sequence is unbounded, while for almost all odd seeds it is bounded.

We extend the definition of s to include $s(0) = 0$, and for $n \ge 0$, we let $s_k(n)$ denote the k-th iterate of s at n. The sequence $n, s(n), s_2(n), \dots$ is known as the "aliquot" sequence with seed n. Numbers in a cycle under the s-iteration are called "sociable". It's known (see [14]) that the set of even sociable numbers has asymptotic density 0 and the set

2010 *Mathematics Subject Classification.* 11N37, 11N64, 11A25.

[1]Whenever we say "almost all" we mean except for a set of asymptotic density 0.

of odd sociable numbers has upper asymptotic density at most the density of the set of odd abundant numbers (which is ≈ 0.002). It is conjectured that the set of odd sociable numbers has density 0. It is not known if there are infinitely many sociable numbers, though it is conjectured that this is the case.

The evidence either for the Catalan–Dickson conjecture or the Guy–Selfridge counter conjecture is mixed. Perhaps pointing towards Guy–Selfridge is the theorem of Lenstra [15] that there exist arbitrarily long strictly increasing aliquot sequences, and the strengthening of Erdős [9] that for each k and almost all abundant numbers n, the aliquot sequence with seed n strictly increases for k steps. (The asymptotic density of the abundant numbers is known to be ≈ 0.2476 [13].) These theoretical results are borne out in practice, where it is found not uncommonly that certain divisors (such as 24), known as "drivers", persist for long stretches in the sequence, and when these divisors are abundant, the sequence grows geometrically at least as long as this persistence.

However, the persistence of an abundant driver is not abolute, and its dominance can be broken, sometimes by a "down driver", such as 2 (where the number is not divisible by 4 nor 3), which can also persist and tends to drive the sequence lower geometrically. It is rare for a sequence to switch parity, this occurs if and only if one hits a square or its double. Since the even numbers are where the principal disagreement in the two conjectures lie, Bosma and Kane [2] considered the geometric mean, on average for $s(2n)/2n$. Namely, they showed that there is a real number β such that

$$\frac{1}{x}\sum_{n\le x}\log(s(2n)/2n)\to\beta \quad \text{as } x\to\infty,$$

and that $\beta \approx -0.03$ is *negative*. This may be interpreted as evidence in favor of Catalan–Dickson and against Guy–Selfridge.

However, very little is known about the set of numbers $s(n)$, with even less known about $s_k(n)$ when $k \ge 2$, so statistical results about seeds n already seem less relevant when one proceeds a single step into the aliquot sequence. We do know that almost all odd numbers are of the form $s(n)$, a result whose proof depends on approximations to Goldbach's conjecture that even numbers are the sum of two primes. In this paper we delve more deeply into this idea showing that almost all odd numbers are of the form $s_k(n)$ for every k. More mysterious are the even numbers in the range of s. Erdős [7] showed that a positive proportion of even numbers are not of the form $s(n)$, while it was only very recently shown in [17] that a positive proportion of even numbers are of the form $s(n)$. In [23] we give a heuristic argument with some numerical evidence that the asymptotic density of the even numbers of the form $s(n)$ is about $1/3$, also see [27].

Our principal result is an extension of the Bosma–Kane theorem to the next iterate.

Theorem 1.1. *The average value of* $\log(s_2(2n)/s(2n))$ *is asymptotically equal to the average value of* $\log(s(2n)/2n)$. *That is,*

$$\frac{1}{x}\sum_{2\le n\le x}\log(s_2(2n)/s(2n)) \sim \frac{1}{x}\sum_{1\le n\le x}\log(s(2n)/2n)\sim\beta, \quad \text{as } x\to\infty.$$

.

(The reason the first sum excludes $n = 1$ is that $s_2(2) = 0$.) The proof of Theorem 1.1 uses some ideas from [17] and [18].

We also consider the sets $s^{-1}(n)$, obtaining an asymptotic formula for the size of $s^{-1}(n)$ when $n > 1$ is odd. The set $s^{-1}(1)$ is the set of primes, but for $n > 1$, $s^{-1}(n)$ is finite (in fact every preimage of n under s is smaller than n^2).

We mention some other recent work on s. In Bosma [1] the aliquot sequence is computed for each seed to 10^6 until it terminates, cycles, or surpasses 10^{99}. He found that about $1/3$ of the even seeds are in this last category, perhaps lending some support to Guy–Selfridge. In Troupe [28] it is shown that the normal number of prime factors of $s(n)$ is $\log\log n$, thus extending the theorem of Hardy and Ramanujan about the normal number of prime factors of n. In [26] it is shown that for all large x the number of amicable numbers $n \le x$ is smaller than $x/\exp(\sqrt{\log x})$.

By way of notation, we have $\omega(n)$ as the number of distinct prime divisors of the natural number n and $\Omega(n)$ as the total number of prime factors of n with multiplicity. We let $P^+(n)$ denote the largest prime factor of $n > 1$, with $P^+(1) = 1$. We reserve the letters p, q for primes, and we write $p^a \parallel n$ if $p^a \mid n$ and $p^{a+1} \nmid n$. We let $\log_k x$ denote the k-fold iteration of the natural logarithm at x, and when we use this notation we assume that the argument is large enough for the expression to be defined. We write $f(x) \ll g(x)$ if $f(x) = O(g(x))$, and we write $f(x) \asymp g(x)$ if $f(x) \ll g(x) \ll f(x)$.

2. The double iterate

In this section we prove Theorem 1.1.

Proof. It follows from [9] and [10] (also see [22] and [18]) that $s_2(n)/s(n) \sim s(n)/n$ as $n \to \infty$ on a set of asymptotic density 1. Thus,

$$\log(s_2(n)/s(n)) = \log(s(n)/n) + o(1)$$

on this same set. However, some terms here are unbounded, both on the positive side and the negative side, and it is conceivable that a set of asymptotic density 0 could be of consequence when averaging. This event commonly occurs. For example, when computing the average of $\tau(n)$, the number of divisors of n, one finds the normal size of $\tau(n)$ is considerably smaller than the average size.

By the theorem of Bosma and Kane,

$$\sum_{n \le x} \log(s(2n)/2n) \sim \beta x, \quad \text{as } x \to \infty,$$

where $\beta \approx -0.03$. Thus, we need to show that for each fixed $\epsilon > 0$, there is some number B such that the contribution to the two sums in our theorem from terms which have absolute value $> B$ has absolute value at most ϵx.

Note that no term in the Bosma–Kane sum is smaller than $-\log 2$. However, the terms $\log(s_2(2n)/s(2n))$ can be smaller than this; for example, when $n = 2$ we have $-\log 4$. However, for $n > 1$ we have

$$\log(s_2(2n)/s(2n)) \ge \log(1/s(2n)) \ge -\log(3n).$$

Now if $s_2(2n)/s(2n) < 1/2$, we have $s(2n)$ odd, which implies that n or $2n$ is a square. Thus,

$$\frac{1}{x} \sum_{\substack{1<n\le x \\ s_2(2n)/s(2n)<1/2}} \log(s_2(2n)/s(2n)) \ge -\frac{2\log(3x)}{x^{1/2}}.$$

So it remains to show that large positive values are of small consequence. A useful result is the following.

Theorem E. *Uniformly for every positive number x, the number of positive integers $n \le x$ with $s(n)/n > y$ is at most*

$$x/\exp(\exp((e^{-\gamma} + o(1))y)), \quad \textit{as } y \to \infty.$$

This result is essentially due to Erdős, see [14, Theorem B].

We can use Theorem E as follows. Let y_0 be so large that the count in Theorem E is smaller than $x/\exp(\exp(y/2))$ for all $y \ge y_0$. Let $B \ge y_0$ be a large fixed number. The contribution to the Bosma–Kane sum from those $n \le x$ with $B^{2^{j-1}} < s(2n)/2n \le B^{2^j}$ is at most

$$\frac{2x\log(B^{2^j})}{\exp(\exp(B^{2^{j-1}}/2))}.$$

Summing this for $j \ge 1$ we get a quantity that is $\ll x \log B/\exp(\exp(B/2)))$. Since B may be fixed as arbitrarily large, the numbers $n \le x$ with $s(2n)/2n > B$ give a contribution of vanishing importance.

We need to show the same result for $s_2(2n)/s(2n)$, and this is the heart of the argument. Let $y = y(x) = (\log_2 x)/(\log_3 x)^2$.

Proposition 2.1. *But for $O(x/y^{4/3})$ integers $n \le x$ we have*

$$\left|\frac{s_2(n)}{s(n)} - \frac{s(n)}{n}\right| \ll \frac{\log_4 x}{\log_3 x} \cdot \frac{\sigma(n)}{n}.$$

Proof. We first show that we may assume that for each integer $m \le y$ there are 5 distinct primes $p_1, \dots, p_5$ with each $p_i \,\|\, n$ and $p_i \equiv -1 \pmod{m}$. To see this, let P_m denote the set of primes $p \in (y^2, x^{1/5}]$ with $p \equiv -1 \pmod{m}$. Using [24] we have

$$S_m := \sum_{p \in P_m} \frac{1}{p} = \frac{\log(\log(x^{1/5})/\log y)}{\varphi(m)} + O(1),$$

uniformly for each $m \le y$. Thus, for x sufficiently large, we have each

$$S_m \ge \frac{\log_2 x}{2m}.$$

The number of $n \le x$ divisible by some p^2 with $p > y^2$ is $O(x/y^2)$, so we may assume that the numbers n we are considering are never divisible by p^2 for p in some P_m. The number of $n \le x$ not divisible by 5 different members of P_m is, via the sieve,

$$\ll \left(1 + S_m + \frac{1}{2}S_m^2 + \frac{1}{6}S_m^3 + \frac{1}{24}S_m^4\right)\frac{x}{\exp(S_m)} \ll \frac{x(\log_2 x)^4}{\exp\left(\frac{\log_2 x}{2m}\right)} \le \frac{x(\log_2 x)^4}{\exp\left(\frac{1}{2}(\log_3 x)^2\right)}.$$

Summing for $m \le y$, we get an estimate that is $\ll x(\log_2 x)^5/\exp(\frac{1}{2}(\log_3 x)^2)$. Thus, but for a negligible set of $n \le x$, $m^5 \mid \sigma(n)$ for every $m \le y$.

For $n \le x$, let $m = m(n)$ be the largest divisor of n supported on the primes $\le y$. If $p^a \mid m(n)$ with $p^a > y^2$, then $a \ge 3$. The number of such n is at most

$$\sum_{\substack{p \le y \\ p^a > y^2}} \frac{x}{p^a} \ll \frac{x}{y^{4/3}},$$

so we may assume that each $p^a \mid m(n)$ has $p^a \le y^2$. We claim that if n has not so far been excluded, then $m(n) = m(s(n))$. If $p^a \le y$, then by the above paragraph, $p^{5a} \mid \sigma(n)$, so $p^a \,\|\, m(s(n))$. Thus, if $p \nmid m(n)$, then $p \nmid s(m(n))$ and if $p^a \,\|\, m(n)$, then $p^a \,\|\, s(n)$. It remains to consider the case of $p^a \,\|\, m(n)$ with $y < p^a \le y^2$. Let b be the largest integer with $p^b \le y$. Then $p^b \ge y^{1/2}$, so that $5b > a$. But, as we have seen, $p^{5b} \mid \sigma(n)$. Thus $p^a \,\|\, m(s(n))$.

To complete the proof of the proposition, we must show that primes $p > y$ do not overly influence the values $s(n)/n$ and $s_2(n)/s(n)$. The number of integers $n \le x$ with $\omega(n) > 3\log_2 x$ is $\ll x/\log x$ by a well-known result of Hardy and Ramanujan. So, we may assume that $\omega(n) \le 3\log_2 x$. The sum of reciprocals of the first $\lfloor 3\log_2 x \rfloor$ primes larger than y is $\ll \log_4 x/\log_3 x$. Thus,

$$\frac{\sigma(n/(m(n))}{n/m(n)} \ll \frac{\log_4 x}{\log_3 x},$$

and so

$$\left|\frac{s(n)}{n} - \frac{s(m(n))}{m(n)}\right| \ll \frac{\log_4 x}{\log_3 x} \cdot \frac{\sigma(m(n))}{m(n)} \le \frac{\log_4 x}{\log_3 x} \cdot \frac{\sigma(n)}{n}.$$

It thus suffices to prove a similar result for $s_2(n)/s(n)$. By a result of [5], the number of $n \le x$ with $P^+(n) \le x^{1/\log_3 x}$ is at most $x/(\log_2 x)^{(1+o(1))\log_4 x}$, as $x \to \infty$. Thus, we may assume that $P^+(n) > x^{1/\log_3 x}$. Suppose that $\omega(s(n)) > 7\log_2 x \log_3 x$. Write $n = pk$ where $p = P^+(n)$. By a result above we may assume that $p \nmid k$. Thus,

$$s(n) = ps(k) + \sigma(k). \tag{1}$$

Since $\omega(s(n)) > 7\log_2 x \log_3 x$, there is a divisor u of $s(n)$ with $u < x^{1/\log_3 x}$ and $\omega(u) > 7\log_2 x$. Let u_1 be the largest divisor of u that is coprime to n. From (1) we have u_1 coprime to $s(k)$. Since we may assume that $\omega(n) \le 3\log_2 x$, we have $\omega(u_1) > 4\log_2 x$. Reading (1) as a congruence mod u_1, we see that for a given choice of k and u_1, p is determined mod u_1. Since $k < x^{1-1/\log_3 x}$ and $u_1 < x^{1/\log_3 x}$, it follows that the number of choices for p, and thus for n is

$$\le \sum_{k<x^{1-1/\log_3 x}} \sum_{\substack{u_1<x^{1/\log_3 x}\\ \omega(u_1)>4\log_2 x}} \frac{x}{ku_1} \ll \frac{x}{\log x},$$

again using the result of Hardy and Ramanujan mentioned above. Thus, we may assume that $\omega(s(n)) \le 7\log_2 x \log_3 x$. Since the reciprocal sum of the first $7\log_2 x\log_3 x$ primes $> y$ is $\ll \log_4 x/\log_3 x$ and $m(s(n)) = m(n)$, we have

$$\left|\frac{s_2(n)}{s(n)} - \frac{s(m(n))}{m(n)}\right| \ll \frac{\log_4 x}{\log_3 x} \cdot \frac{\sigma(m(n))}{m(n)} \le \frac{\log_4 x}{\log_3 x} \cdot \frac{\sigma(n)}{n}.$$

This completes the proof of the proposition. □

We now complete the proof of the theorem. We have seen that the negative terms in the sums with large absolute values are negligible, and that large positive values of $\log(s(n)/n)$ are also negligible. Since $s(n)/n \ll \log_2 n$ for $n \ge 3$, it follows that for $n \le x$ with $s_2(2n)/s(2n) \ge 1$, we have $\log(s_2(n)/s(n)) \ll \log_3 x$. Thus, the contribution to the sum from those terms not satisfying the inequality in Proposition 2.1 is $\ll x(\log_3 x)/y^{4/3} = o(x)$. Since $\sum_{n\le x} \sigma(n)/n \ll x$, it follows that the difference of the two sums for those $n \le x$ which satisfy the inequality of Proposition 2.1 is $\ll x\log_4 x/\log_3 x = o(x)$. This completes the proof of the theorem. □

Corollary 2.2. *We have*

$$\sum_{n\le x} \frac{s_2(n)}{s(n)} \sim \sum_{n\le x} \frac{s(n)}{n} \quad \text{as } x \to \infty$$

and

$$\sum_{n\le x} \frac{s_2(n)}{n} \sim \sum_{n\le x} \left(\frac{s(n)}{n}\right)^2 \quad \text{as } x \to \infty.$$

These results follow from the tools used to prove Theorem 1.1. Note that, where ζ is the Riemann zeta function,

$$\sum_{n\le x} \frac{s(n)}{n} \sim (\zeta(2)-1)x, \quad \sum_{n\le x} \left(\frac{s(n)}{n}\right)^2 \sim \left(\frac{\zeta(2)^2\zeta(3)}{\zeta(4)} - 2\zeta(2) + 1\right) x \ \text{ as } x \to \infty.$$

In [10] the following conjecture is proposed.

Conjecture 2.3. *If A is a set of natural numbers of asymptotic density* 0*, then $s^{-1}(A)$ has asymptotic density* 0.

Theorem 2.4. *Assuming Conjecture 2.3, then for each integer $k \ge 2$ there is a set A_k of asymptotic density* 1 *such that*

$$\frac{1}{x} \sum_{\substack{n\le x\\ n\in A_k}} \log(s_k(n)/s_{k-1}(n)) \to \beta, \quad \text{as } x \to \infty.$$

Proof. We first note that a consequence of the conjecture is that if A has density 0, then $s_k^{-1}(A) = \{n : s_k(n) \in A\}$ has density 0. This is clear for $k = 1$. For $k \geq 2$, by the case $k = 1$ and induction, $s_k^{-1}(A) = s_{k-1}^{-1}(s^{-1}(A))$ has density 0.

Let A be the set of integers n such that for each $m \leq \log_2 n/(\log_3 n)^2$ there are at least 5 distinct primes $p \,\|\, n$ with $p \equiv -1 \pmod n$. We have seen in the proof of Proposition 2.1 that A has asymptotic density 1. If we also insist that members n of A satisfy $\omega(n) \leq 3\log_2 n$, then A still has asymptotic density 1. Letting B denote the complement of A, the conjecture implies that $s_j^{-1}(B)$ has density 0 for each $j < k$. Let A_k be the part of A lieing outside of each of these sets $s_j^{-1}(B)$ so that A_k has density 1. We have seen in the proof of Proposition 2.1 that if $n, s(n) \in A$, then $s_2(n)/s(n) \sim s(n)/n$. However, if $n \in A_k$ then all of $n, s(n), \dots, s_{k-1}(n)$ are in A, so all of the ratios $s_{j+1}(n)/s_j(n)$ are asymptotic to each other. This proves the theorem. □

In [2] the authors also study the full sum $\sum_{1<n\leq x} \log(s(n)/n)$, showing that it is asymptotically $-e^{-\gamma} x \log_2 x$. We can prove this for $\log(s_2(n)/s(n))$, with the proviso that n runs over composite numbers to avoid undefined summands. The sum of $\log(s(n)/n)$ is analyzed by singling out those n with the same smallest prime factor q. The terms when $q < \exp((\log x)^\epsilon)$ account for a vanishingly small portion of the sum when ϵ is small (and so this is another case where the asymptotics are dominated by those terms corresponding to a set of density 0). To prove the result for $\log(s_2(n)/s(n))$, one reduces to the case when $p = P^+(n) > x^{1/\log_2 x}$, writes $n = pm$, and assumes that m has no prime factors up to $(\log_2 x)^2$. Fixing a prime q in the range $\exp((\log x)^\epsilon) < q < \exp((\log x)^{1-\epsilon})$ and a value for m, one counts primes $p \leq x/m$ such that $s(pm)$ has least prime factor q. This can be done asymptotically correctly using the fundamental lemma of the sieve, see [12].

We do not have an analogue of Theorem 2.4 for $\log(s_k(n)/s_{k-1}(n))$ since presumably the sum of these terms is principally supported on a set of n of density 0.

3. The inverse image

For a positive integer n, let $G(n)$ denote the set of pairs of primes p, q with $n = p + q$.

Theorem 3.1. *For each odd integer $n > 1$,*

$$|s^{-1}(n)| = \frac{1}{2}\#G(n-1) + O\left(\frac{n(\log_2 n)^3}{(\log n)^3}\right).$$

For each even integer $n > 1$,

$$|s^{-1}(n)| \leq \frac{n}{\exp((1/\sqrt{2} + o(1))\sqrt{\log n \log_2 n})} \quad \text{as } n \to \infty.$$

Proof. Suppose that $s(m) = n$, and write $m = pk$, where $p = P^+(m)$. If $p \nmid k$, then

$$ps(k) + \sigma(k) = s(m) = n, \tag{2}$$

so that n and k determine p. Thus, given n, k there are at most 2 choices for p: the one determined by (2) and $P^+(k)$. Also note that since k is a proper divisor of m, we have $k < s(m) = n$.

First fix an even number n. If m is odd, then it is a square, say it is p^2l^2, where p is still $P^+(m)$. Then $pl^2 < n$ and if $p \nmid l$,

$$n = s(m) = (p+1)\sigma(l^2) + p^2 s(l^2).$$

Given l, the right side is increasing with p, so there is at most 1 choice for p. So in all, there are at most 2 choices for p: the 1 determined from the equation above and $P^+(l)$. Hence l is an integer smaller than $\sqrt{n}$ and it determines at most 2 choices for p. Hence there are at most $O(\sqrt{n})$ choices for m when m is odd.

Now assume that m is even. Then $n = s(m) > m/2$. But the number of integers $m < 2n$ determined by $m/P^+(m)$ in at most 2 ways is at most $n/L^{1+o(1)}$ as $n \to \infty$, where $L = \exp(\sqrt{\frac{1}{2}\log n \log_2 n})$. Indeed, if $p \le L$, then by a result of de Bruijn [5], the number of choices for m is at most $n/L^{1+o(1)}$ as $n \to \infty$. But if $p > L$, then $k < 2n/L$ and the number of choices for m is $< 4n/L$. This completes the argument when n is even.

Now assume that n is odd and let

$$\mathcal{I}_j(n) = \{m : s(m) = n,\ \Omega(m) = j\}.$$

Note that since $n > 1$, we have $\mathcal{I}_j(n) = \emptyset$ when $j = 0, 1$. If $m \in \mathcal{I}_2(n)$, then either $m = pq$ with $p > q$ or $m = p^2$. In the latter case we have $n = p + 1$ and in the former case, $n - 1 = p + q$. Thus,

$$\#\mathcal{I}_2(n) = \frac{1}{2}\#G(n-1) + O(1).$$

The remaining cases for j are split into $j = 3$, $4 \le j \le 6\log_2 n$, and $j > 6\log_2 n$. Suppose that $m \in \mathcal{I}_3(n)$, and write $m = pqr$, where $p \ge q \ge r$ are primes. We may assume that $k = qr > n/(\log n)^3$, since otherwise there are at most $2n/(\log n)^3$ choices for m. But, pq, as a proper divisor of m, is smaller than n. This implies that $p/r < (\log n)^3$. If $m = p^3$, then $n = p^2 + p + 1$, so this gives at most 1 choice. If $m = p^2 r$ with $p > r$, then a variant of (2) shows that r is determined by p, n, and since $p^2 < n$, there are at most $\sqrt{n}$ choices in this case. If $m = pr^2$ with $p > r$, then (2) shows that p is determined by r, n and so again there are at most $\sqrt{n}$ choices for m. Finally assume that $p > q > r$ and let $l = q + r + 1$. Then, with x a polynomial variable,

$$(x - q)(x - r) \equiv x^2 + x + n \pmod{l}.$$

Indeed, by (2) and using $\sigma(k) = k + l$, we have

$$n = pl + k + l \equiv k = qr \pmod{l}.$$

For a given choice of n, l, the polynomial $x^2 + x + n$ has at most $l^{o(1)}$ roots modulo l. Further, $r < q < r(\log n)^3$ and $r < \sqrt{n}$, so $l \ll \sqrt{n}(\log n)^3$. We conclude that there are at most $n^{1/2+o(1)}$ choices for m in all cases. That is,

$$\#\mathcal{I}_3(n) \le n^{1/2+o(1)} \quad \text{as } n \to \infty.$$

Now suppose that $j \ge 4$. Write m as $r_1 r_2 \dots r_j$, where $r_1 \le r_2 \le \cdots \le r_j = p$. As in the case $j = 3$, but for $O(n/(\log n)^3)$ choices for m, we may assume that $k = m/p = r_1 r_2 \dots r_{j-1} > n/(\log n)^3$. And since m/r_1 is a proper divisor of m, we have $m/r_1 < n$, so that $r_j/r_1 < (\log n)^3$. Since $n/(\log n)^3 < k < n$, we have

$$n^{1/(j-1)}/(\log n)^3 < r_1 < n^{1/(j-1)},\ n^{1/(j-1)}/(\log n)^{3/(j-1)} < r_{j-1}.$$

The number of choices for k is at most

$$\begin{aligned}
\sum_{r_1 \le r_2 \le \cdots \le r_{j-2} < r_1(\log n)^3} \ \sum_{r_{j-1} < n/r_1 r_2 \dots r_{j-2}} 1 &\le \sum_{r_1, r_2, \dots, r_{j-2}} \pi(n/r_1 r_2 \dots r_{j-2}) \\
&\ll \sum_{r_1, r_2, \dots, r_{j-2}} \frac{n}{r_1 r_2 \dots r_{j-2} \log(n^{1/(j-1)}/(\log n)^{3/(j-1)})} \\
&\ll \frac{jn}{\log n} \sum_{r_1, r_2, \dots, r_{j-2}} \frac{1}{r_1 r_2 \dots r_{j-2}} \\
&\ll \frac{jn}{\log n}\left(\frac{4\log_2 n}{\log(n^{1/(j-1)}/(\log n)^3)}\right)^{j-2} \\
&\ll \frac{j^{j-1} n (4\log_2 n)^{j-1}}{(\log n)^{j-1}}.
\end{aligned}$$

We thus have

$$\sum_{4 \le j \le 6\log_2 n} \#\mathcal{I}_j(n) \ll n(\log_2 n)^3/(\log n)^3.$$

Finally, when $j > 6\log_2 n$, we have $\Omega(k) > 6\log_2 n - 1$, so it follows from Lemmas 12 and 13 in [16] that the number of choices for k is at most $o(n/(\log n)^3)$ as $n \to \infty$. Since the number of choices for m is at most twice the number of choices for k, this concludes our proof. □

Remark. An averaging argument shows that our estimate for $\#\mathcal{I}_3(n)$ is close to best possible. To do better in the case of n odd one would have to improve the above argument for $j = 4$. In the case of n even we have not been able to prove that there are infinitely many n with $|s^{-1}(n)| \geq 3$.

4. The infinite intersection

Let

$$S_\infty = \bigcap_{k\geq 1} s_k(\mathbf{N}).$$

Clearly S_∞ contains all of the sociable numbers, and it is likely to contain all of the odd numbers except for 5. This last assertion follows from the slightly stronger form of Goldbach's conjecture that every even number starting with 8 is the sum of two different primes, for then, every odd number starting with 9 is of the form $s(pq) = p + q + 1$ where $p > q > 2$. (One also should note that $s(11) = 1$, $s_2(9) = 3$, $s_2(49) = 7$, and $5 \notin s(\mathbf{N})$.)

We note that an even stronger form of Goldbach's conjecture is *nearly* true. Recall that $G(n)$ denotes the set of pairs p, q of primes with $p + q = n$. Let

$$\mathcal{G}(n) = \frac{n^2}{\varphi(n)(\log n)^2} \prod_{p\nmid n} \left(1 - \frac{1}{(p-1)^2}\right).$$

Theorem G. *There is a positive constant δ, such that for each large number x, there are two sets $E_C(x), E_S(x)$ such that if even $n \in (\frac{1}{2}x, x]$ and $n \notin E_C(x) \cup E_S(x)$, then*

$$\#G(n) \sim \mathcal{G}(n)$$

uniformly as $x \to \infty$. Moreover, $\#E_C(x) \leq x^{1-\delta}$ and either $E_S(x) = \emptyset$ or there is an integer $r(x) > \log x$ with $E_S(x)$ being the set of even $n \in (\frac{1}{2}x, x]$ with $(r(x), n) > r(x)/\log_2 x$.

This result follows from the proof in [19], see especially the argument in Section 8. It is likely that later results on Goldbach's conjecture can be used as well, giving larger numerical values for δ, for example, see Pintz [20]. We shall use Theorem G to show the following.

Theorem 4.1. *All odd numbers belong to S_∞ except for a set of asymptotic density* 0.

Proof. We first prove the theorem under the assumption that $E_C(x)$ is always empty. Let x be large. If $E_S(x) \neq \emptyset$, then

$$\#E_S(x) \leq \sum_{\substack{d|r(x)\\ d\leq \log_2 x}} \sum_{\substack{n\leq x\\ r(x)/d|n}} 1 \leq \frac{x}{r(x)} \sum_{\substack{d|r(x)\\ d\leq\log_2 x}} d \leq \frac{x(\log_2 x)^2}{r(x)} < \frac{x(\log_2 x)^2}{\log x}.$$

So, all but $o(x)$ odd integers $n \in (\frac{1}{2}x, x]$ have $n - 1 \notin E_S(x)$. We claim that all of these n are in S_∞. For consider the $\#G(n-1) \sim \mathcal{G}(n-1)$ pairs p, q with $p + q = n - 1$. By a sieve argument, for each odd $n \in (\frac{1}{2}x, x]$, the number of these pairs with $pq \leq \epsilon x^2$ is $O(\epsilon\mathcal{G}(n-1))$. Let $\epsilon > 0$ be a fixed small constant so that at least half of the pairs p, q for n have $pq > \epsilon x^2$. Let k be a fixed integer with $2^k\epsilon > 1$. Thus, there is a fixed constant $c > 0$ such that for some choice of $x_1 \in \{2^{-i}x^2 : i = 0, 1, \dots, k\}$ with at least $c\mathcal{G}(n-1)$

pairs p, q with $p+q=n-1$, $pq \in (\frac{1}{2}x_1, x_1]$. The number of such $pq \in E_S(x_1)$ is at most

$$2 \sum_{\substack{d \mid r(x_1) \\ d \le \log_2 x_1}} \sum_{\substack{q \le \frac{1}{2}n \\ n-q \text{ prime} \\ r(x_1)/d \mid q(n-q)-1}} 1.$$

The inner sum here puts q in at most $r(x_1)^{o(1)}$ residue classes mod $r(x_1)/d$. So, if $r(x_1)$ is large, say $r(x_1) > (\log x_1)^3$, then ignoring the primality conditions on $q, n-q$, the number of pairs we are counting is $\ll n/(\log n)^{2.9} = o(\mathcal{G}(n-1))$. So, assume that $r(x_1) \le (\log x_1)^3$. Then the inner sum above is $\le \mathcal{G}(n-1) r(x_1)^{o(1)}/\varphi(r(x_1)/d)$. This too is $o(\mathcal{G}(n-1))$. We conclude that there is some fixed constant $c' > 0$ such that if x is sufficiently large, each odd $n \in (\frac{1}{2}x, x]$ not in $E_S(x)$ has some x_1 as above and at least $c'n/(\log n)^2$ pairs p, q with $p+q=n-1$, $pq \in (\frac{1}{2}x_1, x_1]$, and $pq-1 \notin E_S(x_1)$.

We claim that even the existence of a single pair p, q in $G(n-1)$ with $pq \in (\frac{1}{2}x_1, x_1]$ and $pq-1 \notin E_S(x_1)$ is enough to allow the process to continue indefinitely. That is, let $n_1 = pq$. Then there is some $n_2 > \epsilon n_1^2$ with $s(n_2) = n_1$, then some n_3, etc. This proves the claim and establishes the theorem in the case that the exceptional sets $E_C(x)$ are always empty.

To establish the theorem with the exceptional sets $E_C(x)$ taken into account, we proceed similarly. First, we eliminate those $n \in (\frac{1}{2}x, x]$ with $n-1$ in $E_C(x)$, which is clearly a negligible set. The exceptional set $E_C(x_1)$ also can eliminate some values of $n \in (\frac{1}{2}x, x]$. Since each odd $n \in (\frac{1}{2}x, x]$ with $n-1$ not in $E_C(x) \cup E_S(x)$ gives rise to $\ge c'n/(\log n)^2$ numbers $n_1 \in (\frac{1}{2}x_1, x_1]$ with $n_1 - 1$ not in $E_S(x_1)$, the number of these n which give rise to $> (c'/2)n/(\log n)^2$ numbers $n_1 \in (\frac{1}{2}x_1, x_1]$ with $n_1 - 1$ in $E_C(x_1)$ is at most

$$\frac{x_1^{1-\delta}}{(c'/4)x/(\log x)^2} \ll x^{1-2\delta}(\log x)^2.$$

Each remaining n_1 gives rise to at least $c'n_1/(\log n_1)^2$ numbers $n_2 \in (\frac{1}{2}x_2, x_2]$ with $n_2 - 1$ not in $E_S(x_2)$, where $x_2 \asymp x_1^2 \asymp x^4$. The number of n where more than $(c'/2)n_1/(\log n_1)^2$ values of n_2 are in $E_C(x_2)$ is $\ll x^{1-4\delta}(\log x)^4$. Let j be an integer with $2^j\delta > 2$. So, after iterating this procedure j times, we have that the exceptional set $E_C(x_j)$ does not rule out any $n \in (\frac{1}{2}x, x]$ that has not already been ruled out. So, replacing c' with $c'/2$ (since an exceptional set can cut down the growth of the number of n_i's at level i), we can continue the process to infinity. □

Remark. Though S_∞ contains almost all odd numbers, it is difficult to say for sure if a specific odd number that is not sociable is in S_∞. One may assume in Theorem G that the exceptional modulus $r(x)$, when it exists, is a bit larger, say $r(x) > (\log x)^{3/2}$. Then one has from the proof of Theorem 4.1 that almost all primes are in S_∞. Even the existence of one prime in S_∞ is enough to show that $1 \in S_\infty$. So, we can identify one non-sociable positive integer in S_∞. Is there another that can be identified? Are there any even non-sociables in S_∞? It is not hard to show, using Lenstra [15] or Erdős [9], that for every k there are infinitely many even numbers in $s_k(\mathbf{N})$.

References

[1] W. Bosma, *Aliquot sequences with small starting values*, preprint available from `http://www.math.ru.nl/~bosma/`.
[2] W. Bosma and B. Kane, *The aliquot constant*, Quart J. Math. **63** (2012), 309–323.
[3] E. Catalan, *Propositions et questions diverses*, Bull. Soc. Math. France **16** (1888), 128–129.
[4] L. E. Dickson, *Theorems and tables on the sum of the divisors of a number*, Quart J. Math. **44** (1913), 264–296.
[5] N. G. de Bruijn, *On the number of positive integers $\le x$ and free of prime factors $> y$*, Nederl. Acad. Wetensch. Proc. Ser. A **54** (1951), 50–60.

[6] ———, *On the number of positive integers $\leq x$ and free of prime factors $> y$. II*, Nederl. Acad. Wetensch. Proc. Ser. A **69** (1966), 239–247.

[7] P. Erdős, *Über die Zahlen der Form $\sigma(n) - n$ und $n - \varphi(n)$*, Elem. Math. **28** (1973), 83–86.

[8] ———, *On the distribution of numbers of the form $\sigma(n)/n$ and on some related questions*, Pacific J. Math. **52** (1974), 59–65.

[9] ———, *On asymptotic properties of aliquot sequences*, Math. Comp. **30** (1976), 641–645.

[10] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math., vol. 85, Birkhäuser Boston, Boston, MA, 1990, pp. 165–204.

[11] R. K. Guy and J. L. Selfridge, *What drives an aliquot sequence?*, Math. Comp. **29** (1975), 101–107.

[12] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press, London, 1974.

[13] M. Kobayashi, *On the density of the abundant numbers*, Ph.D. Dissertation, Dartmouth College, 2010.

[14] M. Kobayashi, P. Pollack, and C. Pomerance, *On the distribution of sociable numbers*, J. Number Theory **129** (2009), 1990–2009.

[15] H. W. Lenstra, Jr., *Problem 6064*, Amer. Math. Monthly **82** (1975), 1016. *Solution by the proposer*, op. cit. **84** (1977), 580.

[16] F. Luca and C. Pomerance, *Irreducible radical extensions and Euler function chains*, Combinatorial number theory, de Gruyter, Berlin, 2007, pp. 351–361; Integers **7** (2007), no. 2, #A25.

[17] ———, *The range of the sum-of-proper-divisors function*, Acta Arith. **168** (2015), 187–199.

[18] ———, *Local behavior of the composition of the aliquot and co-totient functions*, to appear.

[19] H. L. Montgomery and R. C. Vaugan, *The exceptional set in Goldbach's problem*, Acta Arith. **27** (1975), 353–370.

[20] J. Pintz, *Recent results on the Goldbach conjecture*, in: Elementare und analytische Zahlentheorie, 220–254, Schr. Wiss. Ges. Johann Wolfgang Goethe Univ. Frankfurt am Main, 20, Franz Steiner Verlag Stuttgart, Stuttgart, 2006.

[21] P. Pollack, *On the greatest common divisor of a number and its sum of divisors*, Michigan Math. J. **60** (2011), 199–214.

[22] ———, *Some arithmetic properties of the sum of proper divisors and the sum of prime divisors*, Illinois J. Math. **58** (2014), 125–147.

[23] P. Pollack and C. Pomerance, *Some problems of Erdős on the sum-of-divisors function*, Trans. Amer. Math. Soc. (Ser. B), to appear.

[24] C. Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. **293/294** (1977), 217–222.

[25] ———, *Primality testing: variations on a theme of Lucas*, Proceedings of the 13th Meeting of the Fibonacci Association, Congressus Numerantium **201** (2010), 301–312.

[26] ———, *On amicable numbers*, in Analytic number theory (in honor of Helmut Maier's 60th birthday), M. Rassias and C. Pomerance, eds., Springer, Cham, Switzerland, 2015, pp. 321–327.

[27] C. Pomerance and H.-S. Yang, *Variant of a theorem of Erdős on the sum-of-proper-divisors function*, Math. Comp. **83** (2014), no. 288, 1903–1913.

[28] L. Troupe, *On the number of prime factors of values of the sum-of-proper-divisors function*, J. Number Theory, to appear.

Mathematics Department, Dartmouth College, Hanover, NH 03755, USA
E-mail address: carl.pomerance@dartmouth.edu