The range of the sum-of-proper-divisors function

Florian Luca
School of Mathematics, University of the Witwatersrand
P. O. Box Wits 2050, South Africa
E-mail: florian-luca@wits.ac.za

Carl Pomerance
Department of Mathematics
Dartmouth College
Hanover, NH 03755-3551, USA
E-mail: carl.pomerance@dartmouth.edu

Abstract

Answering a question of Erdős, we show that a positive proportion of even numbers are in the form $s(n)$, where $s(n) = \sigma(n) - n$, the sum of proper divisors of n.

1 Introduction

For a positive integer n, let $s(n) = \sigma(n) - n$, the sum of the proper divisors of n. The function s has been studied since antiquity; it may be the first function ever defined by mathematicians. Beginning with Pythagoras, we have looked for cycles in the dynamical system formed when iterating s. There are still a number of unsolved problems connected with this dynamical system: Are there infinitely many cycles? Examples of cycles are $6 \to 6$ and $220 \to 284 \to 220$; about 12 million are known. Does the set of numbers involved in some cycle have asymptotic density 0? We know the upper density is bounded above by about 0.002. Is there an unbounded

2010 *Mathematics Subject Classification*: Primary 11A25; Secondary 11N37.
Key words and phrases: sum of proper divisors, aliquot sum, applications of sieve methods.

orbit? The least starting value in question is $n = 276$. (For references on these questions, see [KPP].)

Perhaps a more basic question with the function s is to identify its image: What numbers are of the form $s(n)$? Note that if p, q are different primes then $s(pq) = p + q + 1$. Not many even numbers are of this form, but a slightly stronger version of Goldbach's conjecture (every even number starting with 8 is the sum of two different primes) implies that every odd number starting with 9 is in the range of s. Since $s(2) = 1$, $s(4) = 3$, and $s(8) = 7$, while $s(n) = 5$ has no solutions, it then follows from this slightly stronger Goldbach conjecture that every odd number except 5 is in the range of s. Moreover, this slightly stronger form of Goldbach's conjecture is known to be usually true. There are many papers in this line, a recent survey is [P].

So, almost all odd numbers (in the sense of asymptotic density) are of the form $s(n)$. In a short, beautiful paper, Erdős [E73] looked at the even values of s, showing that a positive proportion of even numbers are missed. He raised the issue of whether the asymptotic density of even values exists, saying that it is not even known if the lower density is positive. Similar questions are asked for the function $s_\varphi(n) := n - \varphi(n)$, where φ is Euler's function. Again, almost all odd numbers are attained by s_φ, but even less is known about even values, compared with $s(n)$. In fact, the Erdős argument (that s misses a positive proportion of even values) fails for s_φ.

These thoughts were put in a more general context in [EGPS]. There the following conjecture is formulated.

Conjecture 1.1. *If $\mathcal{A}$ is a set of natural numbers of asymptotic density 0, then $s^{-1}(\mathcal{A})$ also has asymptotic density 0.*

If this is true, one consequence would be that the set of even values of s does not have density 0. Indeed, if $\mathcal{A}$ is the set of even numbers in the range of s, then

$$s^{-1}(\mathcal{A}) = \{n \text{ even} : n, n/2 \text{ not squares}\} \cup \{n^2 : n \text{ odd}\},$$

so $s^{-1}(\mathcal{A})$ has asymptotic density $\frac{1}{2}$. Thus, if Conjecture 1.1 is true, then $\mathcal{A}$ does not have asymptotic density 0.

In this paper we prove the following theorem.

Theorem 1.2. *The set of even numbers of the form $s(n)$ for some integer n has positive lower density.*

With a few superficial changes the proof of Theorem 1.2 can be adapted to show the following more general result: *For any two fixed positive integers a, b, a positive proportion of numbers in the residue class $a \pmod b$ are of the form $s(n)$.* Since asymptotically all odd numbers are of the form $s(n)$, this result has new content only in the case that a, b are both even.

Essentially the same proof will show that numbers of the form $s_\varphi(n)$ contain a positive proportion of all even numbers (or any residue class).

It is hoped that the methods in this paper can be of help in proving Conjecture 1.1.

It seems likely that the asymptotic density of even numbers in the range of s exists. In some numerical work in [PY] it appears that the even numbers in the range have density about $\frac{1}{3}$ and the density of even numbers missing is about $\frac{1}{6}$. In [CZ] it is shown that the lower density of the set of even numbers missing from the range is more than 0.06. The proof of Theorem 1.2 that we present is effective, but we have made no effort towards finding some explicit lower bound for the lower density of even values of s.

2 Notation and lemmas

We have the letters p, q, r, π, with or without dashes or subscripts representing prime numbers. We let $\tau(n)$ denote the number of positive divisors of n. We say a positive integer n is deficient if $s(n) < n$. We let $P(n)$ denote the largest prime factor of n when $n > 1$, and we let $P(1) = 1$. We say a positive integer n is z-smooth if $P(n) \le z$. For each prime p and natural number n, we let $v_p(n)$ denote the exponent of p in the prime factorization of n. For each large number n, let

$$y = y(n) = \log\log n / \log\log\log n.$$

Lemma 2.1. *On a set of asymptotic density 1 we have*

(1) $p^{2a} \mid \sigma(n)$ *for every prime power* $p^a \le y$,

(2) $P(\gcd(n, \sigma(n))) \le y$,

(3) $\sigma(n)/\gcd(n, \sigma(n))$ *is divisible by every prime* $p \le y$,

(4) *and every prime factor of* $s(n)/\gcd(n, \sigma(n))$ *exceeds* y.

Proof. (1) Let x be large, let $y = y(x)$, and let d be an integer with $1 < d \le y$. The integers $n \le x$ with $d^2 \mid \sigma(n)$ include all $n \le x$ which are precisely divisible (i.e., divisible to just the first power) by two different primes p_1, p_2 in the residue class $-1 \pmod d$. The complementary set where $d^2 \nmid \sigma(n)$ is contained in the union of the set of those $n \le x$ divisible by the square of a prime $p > y$ and the set of those $n \le x$ which are not divisible by two different primes $p \equiv -1 \pmod d$ with $p \in (y, \sqrt{x})$. The number of $n \le x$ divisible by the square of a prime $p > y$ is at most $x\sum_{p>y} 1/p^2 \ll x/(y\log y)$, so these numbers are negligible. Let $\mathcal{P}_d(y,\sqrt{x})$ denote the set of primes $p \equiv -1 \pmod d$ with $p \in (y, \sqrt{x})$. Note that the prime number theorem for residue classes implies that

$$\sum_{p\in\mathcal{P}_d(y,\sqrt{x})} \frac{1}{p} = \frac{\log(\log x/\log y)}{\varphi(d)} + O(1),$$

uniformly for $d \le y$. The number of $n \le x$ which are not divisible by 2 different primes in $\mathcal{P}_d(y,\sqrt{x})$ is, by the sieve (see [HR, Theorem 2.2]),

$$\begin{aligned}
&\ll x\left(1 + \sum_{p\in\mathcal{P}_d(y,\sqrt{x})}\frac{1}{p}\right)\prod_{p\in\mathcal{P}_d(y,\sqrt{x})}\left(1-\frac{1}{p}\right)\\
&\ll \frac{x\log\log x}{\varphi(d)}\exp\left(-\frac{\log(\log x/\log y)}{\varphi(d)}\right)\\
&\le \frac{x\log\log x}{\varphi(d)}\exp\left(-\frac{\log(\log x/\log y)}{d}\right)\\
&\ll \begin{cases} \dfrac{x}{\varphi(d)}, & \text{if } \frac{1}{2}y < d \le y,\\[2mm] \dfrac{x}{\varphi(d)\log\log x}, & \text{if } d \le \frac{1}{2}y. \end{cases}
\end{aligned}$$

Letting d run over primes and powers of primes, we see that the number of integers $n \le x$ which do not have the property in (1) is $\ll x/\log y = o(x)$ as $x \to \infty$.

(2) In [ELP, Theorem 8], it is shown that on a set of asymptotic density 1, $\gcd(n, \varphi(n))$ is the largest divisor of n supported on the primes at most $\log\log n$. Virtually the same proof establishes the analogous result for $\gcd(n, \sigma(n))$, so that for almost all n, $\gcd(n,\varphi(n)) = \gcd(n,\sigma(n))$. (Also see [E56, EGPS, KS, Pol].) That the assertion (2) usually holds, it suffices to note that the number of $n \le x$ divisible by a prime in $(y, \log\log x]$ is $o(x)$ as $x \to \infty$.

(3) Let $y = y(x)$, where x is large. This assertion will follow from (1) for $n \le x$ if for each prime power p^a with $p^a \le y < p^{a+1}$, we have $p^{2a} \nmid n$. But, the number of $n \le x$ which fail to have this condition is at most

$$x\sum_{p\le y}\frac{1}{y} \ll \frac{x}{\log y}.$$

(4) For this part, we have seen that we may assume that for each prime $p \le y$, we have $v_p(\sigma(n)) > v_p(n)$. Thus, $v_p(s(n)) = v_p(n) = v_p(\gcd(n, \sigma(n))$ for such primes p. □

Lemma 2.2. *The set of numbers n with $P(n) > n^{1/2}$ and $\pi^2 \mid s(n)$ for some prime $\pi > y(n)$ has asymptotic density* 0.

Proof. Assume that $n \in (x, 2x]$ and that $n = pm$ where $p = P(n) > x^{1/2}$. Let $y = y(x)$ and say $\pi^2 \mid s(n)$ where $\pi > y$. We have

$$s(n) = ps(m) + \sigma(m) \equiv 0 \pmod{\pi^2}.$$

Thus, if $\pi \mid s(m)$, then $\pi \mid \sigma(m)$, so that $\pi \mid m$. By part (2) of Lemma 2.1 this occurs only for $o(x)$ choices for n, so assume that $\pi \nmid s(m)$. The above congruence thus places p in a residue class $R_{\pi,m} \pmod{\pi^2}$ determined by π and m. Since $s(n) \ll x \log\log x$, there is some constant c such that $\pi \le c(x \log\log x)^{1/2}$. First assume that $\pi > \log x$, so that $\pi \in I := (\log x, c(x \log\log x)^{1/2}]$. Using only that $p \le 2x/m$ is an integer in a residue class mod π^2, we have that the number of choices for n is at most

$$\sum_{\pi \in I} \sum_{m < 2x^{1/2}} \left(1 + \frac{2x}{m\pi^2}\right) \ll \sum_{\pi \in I} \left(x^{1/2} + \frac{x \log x}{\pi^2}\right) \ll \frac{x}{\log\log x}.$$

So it remains to consider values of $\pi \in (y, \log x]$. For this we use the Brun–Titchmarsh inequality to count the number of triples π, m, p, getting

$$\sum_{y < \pi \le \log x} \sum_{m < 2x^{1/2}} \sum_{\substack{p \le 2x/m \\ p \equiv R_{\pi,m} \pmod{\pi^2}}} 1 \ll \sum_{\pi} \sum_{m} \frac{x}{m\pi^2 \log(x/m\pi^2)}$$
$$\ll \sum_{\pi} \frac{x \log x}{\pi^2 \log x} \ll \frac{x}{y \log y}.$$

This completes the proof. □

We remark that it would be nice to remove the condition $P(n) > n^{1/2}$ in Lemma 2.2, but we don't know how to do this. Note that Lemmas 2.1, 2.2 imply that for a positive proportion of squarefree integers n we have $s(n)$ squarefree.

Lemma 2.3. *The set of deficient numbers n for which $s(n)$ is non-deficient has asymptotic density* 0.

This result follows from [EGPS, Theorem 5.1] and the continuity of the distribution function for $\sigma(n)/n$.

Lemma 2.4. *As n tends to infinity through a set of asymptotic density* 1 *we have* $\tau(s(n)) = (\log n)^{\log 2 + o(1)}$.

This result follows from the estimates in [T]. We remark that our proof does not depend on this lemma, we could have used the weaker inequality $\tau(s(n)) \le n^{o(1)}$ which holds for all n as $n \to \infty$, but we thought it good to highlight some other recent research concerning the statistical study of $s(n)$.

Lemma 2.5. *On a set of integers n of asymptotic density* 1 *we have*

$$\sum_{\substack{r \mid \sigma(n) \\ r > (\log\log n)^2}} \frac{1}{r} \le 1.$$

This follows by the method of proof of [DL, Lemma 5].

3 Proof of the theorem

In this section we prove Theorem 1.2.

Proof. We identify a set of integers $\mathcal{A}$ such that every member of $s(\mathcal{A})$ is even and $s(\mathcal{A})$ has positive lower density. We shall pile on a number of conditions for $\mathcal{A}$ to satisfy. For our initial choice for $\mathcal{A}$, we take the set of even deficient numbers. This set has a positive density, see [K]. Let x be large; we study $\mathcal{A}(x) := \mathcal{A} \cap [1, x]$. We assume that each member n of $\mathcal{A}(x)$ is of the form

$$\begin{aligned} &n = pm, \quad p \in \left(\frac{x}{2m}, \frac{x}{m}\right], \quad m = q\ell = qrk, \\ &k \le x^{1/60}, \quad r \in (x^{1/15}, x^{1/12}], \quad q \in (x^{7/20}, x^{11/30}]. \end{aligned}$$

So $n = pm = pq\ell = pqrk$. Note that n, m, ℓ, k are all even deficient numbers, each running through a positive proportion of numbers to their respective bounds: $n \le x$, $m \le x^{7/15}$, $\ell \le x^{1/10}$, and $k \le x^{1/60}$. We assume that each of these 4 variables satisfy the properties in the lemmas. We also assume that k has no prime factors in $(y(k), y(x)]$.

Let $y = y(x)$. Say $\delta > 0$ is such that $\#\mathcal{A}(x) \ge \delta x$ for all large x. For each y-smooth integer d, let $\mathcal{A}_d(x)$ denote the subset of $\mathcal{A}(x)$ consisting of those members n with largest y-smooth divisor equal to d. By standard results on smooth numbers (see [dB]), there is some constant c such that the reciprocal sum of those y-smooth numbers $d > y^c$ is less than $\frac{1}{3}\delta \log y$.

Note that if $d \le y^c$ is y-smooth, then the number of integers $n \le x$ with greatest y-smooth divisor equal to d is

$$(3.1) \qquad (1+o(1))\frac{x}{d}\prod_{p\le y}\left(1-\frac{1}{p}\right) = (1+o(1))\frac{x}{e^\gamma d\log y}$$

uniformly as $x \to \infty$. Let $\mathcal{D}$ denote the set of y-smooth numbers $d \le y^c$ with

$$\#\mathcal{A}_d(x) \ge \frac{\delta}{6}\frac{x}{d\log y}.$$

We have from (3.1) that for large x,

$$(3.2) \qquad \sum_{d\in\mathcal{D}} \#\mathcal{A}_d(x) \le \frac{x}{\log y}\sum_{d\in\mathcal{D}}\frac{1}{d},$$

and, by definition,

$$\sum_{\substack{P(d)\le y\\ d\le y^c\\ d\notin\mathcal{D}}} \#\mathcal{A}_d(x) < \frac{\delta}{6}\frac{x}{\log y}\sum_{P(d)\le y}\frac{1}{d} = (1+o(1))\frac{\delta}{6}e^\gamma x.$$

Using $\sum_{d\le y^c,\, P(d)\le y}\#\mathcal{A}_d(x) > \frac{2}{3}\delta x$, we thus have, for x large,

$$\sum_{\substack{P(d)\le y\\ d\le y^c\\ d\notin\mathcal{D}}} \#\mathcal{A}_d(x) < \frac{1}{3}\delta x, \quad \sum_{d\in\mathcal{D}}\#\mathcal{A}_d(x) > \frac{1}{3}\delta x,$$

which, with the upper bound (3.2) just seen for this last sum, gives

$$(3.3) \qquad \sum_{d\in\mathcal{D}}\frac{1}{d} > \frac{1}{3}\delta\log y.$$

For $d \in \mathcal{D}$ and a positive integer u, let $R_d(u)$ denote the number of representations of u in the form $s(n)$ for $n \in \mathcal{A}_d(x)$. By the definition of $\mathcal{D}$,

$$\sum_u R_d(u) = \#\mathcal{A}_d(x) \gg \frac{x}{d\log y}$$

uniformly for all $d \in \mathcal{D}$. Note too that if $d \ne d'$, then we cannot have both $R_d(u), R_{d'}(u) > 0$. Indeed, by Lemma 2.1, if $R_d(u) > 0$, then d is the largest y-smooth divisor of u.

We will show that

$$(3.4) \qquad \sum_u R_d(u)^2 \ll \frac{x}{d\log y}$$

uniformly for each $d \in \mathcal{D}$, so that from Cauchy's inequality, it will follow, using (3.3), that

$$\#s(\mathcal{A}(x)) \geq \sum_{d\in\mathcal{D}} \#s(\mathcal{A}_d(x)) \geq \sum_{d\in\mathcal{D}} \frac{\left(\sum_u R_d(u)\right)^2}{\sum_u R_d(u)^2} \gg \sum_{d\in\mathcal{D}} \frac{x}{d\log y} \gg x.$$

The sum $\sum_u R_d(u)^2$ counts solutions to $s(n) = s(n')$ for $n, n' \in \mathcal{A}_d(x)$, with $n = pm, n' = p'm'$. We have

$$ps(m) + \sigma(m) = p's(m') + \sigma(m'). \tag{3.5}$$

Suppose that $m = m'$. Since $m > 1$ (which implies that $s(m) > 0$), we deduce that $p = p'$. This situation contributes $\sum_u R_d(u)$ to $\sum_u R_d(u)^2$, which is easily seen to be $\ll x/(d\log y)$. Thus, we may assume that $m \neq m'$.

By Lemma 2.1, we have $\gcd(m, \sigma(m)) = \gcd(m', \sigma(m')) = d$, so that $d \mid (s(m), s(m'))$. Write $\gcd(s(m), s(m')) = dh$. By Lemma 2.1, every prime factor of h exceeds y. Moreover, since $P(m) = q > m^{7/9}$, it follows from Lemma 2.2 that we may assume that $s(m)$ is not divisible by the square of any prime $\pi > y$. Hence, h is squarefree.

We have from (3.5),

$$p\frac{s(m)}{dh} - p'\frac{s(m')}{dh} = \frac{\sigma(m') - \sigma(m)}{dh}. \tag{3.6}$$

For fixed m, m', we count the number of pairs of primes $p,\ p'$ that satisfy this equation. Note that $\sigma(m) \neq \sigma(m')$, since if they would be equal, we would then get from (3.5) that $ps(m) = p's(m')$, and since

$$\min\{p, p'\} > \max\{m, m'\} > \max\{s(m), s(m')\},$$

we would get that $s(m) = s(m')$, so $m = m'$, which is false. Let u, u' be the integral solution of the linear equation (3.6) in $p,\ p'$ with $u > 0$ and minimal. Then

$$p = u + \frac{s(m')}{dh}t \qquad \text{and} \qquad p' = u' + \frac{s(m)}{dh}t$$

are both primes and $0 \leq t \leq (x/m)/(s(m')/dh) = xdh/(ms(m'))$. Let

$$A = \frac{s(m)}{dh} \times \frac{s(m')}{dh} \times \frac{|\sigma(m) - \sigma(m')|}{dh} =: A_1A_2A_3, \text{ say.}$$

By the sieve ([HR, Theorem 2.2]), the number of such $p \leq x/m$ is

$$\ll \frac{xdh}{ms(m')(\log(xdh/ms(m')))^2}\frac{A}{\varphi(A)} \ll \frac{xdh}{mm'(\log x)^2}\frac{A_1}{\varphi(A_1)}\frac{A_2}{\varphi(A_2)}\frac{A_3}{\varphi(A_3)}, \tag{3.7}$$

where the second inequality follows because $ms(m') \le mm' \le x^{14/15}$ and $s(m') \gg m'$. Since $s(m)/(dh)$ and $s(m')/(dh)$ are deficient, it follows that

$$\frac{A_1}{\varphi(A_1)} \ll 1, \qquad \frac{A_2}{\varphi(A_2)} \ll 1$$

However, $A_3/\varphi(A_3)$ is not small. In fact, by Lemma 2.1, we may assume that A_3 is divisible by all primes $\le y = y(x)$, so $\log y \ll A_3/\varphi(A_3) \ll \log\log x$. Write $A_3 = A_{3,1}A_{3,2}A_{3,3}$, where $A_{3,1}$ is the largest divisor with $P(A_{3,1}) \le (\log\log x)^2$ and $A_{3,2}$ is the largest divisor of what remains with $P(A_{3,2}) \le \log x$. Since A_3 has $O(\log x/\log\log x)$ distinct prime factors, it follows that $A_{3,3}/\varphi(A_{3,3}) \sim 1$ as $x \to \infty$ and so

$$\frac{A_1A_2A_3}{\varphi(A_1)\varphi(A_2)\varphi(A_3)} \ll \frac{A_3}{\varphi(A_3)} \ll \frac{A_{3,2}}{\varphi(A_{3,2})}\log y. \tag{3.8}$$

Let $A'_{3,2}$ be the largest divisor of $A_{3,2}$ which is coprime to $\sigma(m)$. By Lemma 2.5, we may assume that $A_{3,2}/\varphi(A_{3,2}) \ll A'_{3,2}/\varphi(A'_{3,2})$. From (3.7), we now have the problem of showing that for $d \in \mathcal{D}$,

$$\frac{x\log y}{(\log x)^2}\sum_{m,m'}\frac{dhA'_{3,2}}{mm'\varphi(A'_{3,2})} \ll \frac{x}{d\log y}, \tag{3.9}$$

where $dh = \gcd(s(m), s(m'))$.

3.1 $h > x^{1/3}$

We first sum over m, m' with $h > x^{1/3}$, showing that the contribution to (3.9) is small. With $m = q\ell$ and $h \mid s(m)$, we have

$$s(m) = qs(\ell) + \sigma(\ell) \equiv 0 \pmod{h}. \tag{3.10}$$

In addition, h and $\sigma(\ell)$ are coprime. Indeed, if some prime $\pi \mid \gcd(h, \sigma(\ell))$, then $\pi = q$ or $\pi \mid s(\ell)$. In the latter case, $\pi \mid \ell$, so $\pi \mid n$. But $\pi \mid \sigma(\ell)$ implies that $\pi \mid \sigma(n)$, so we have a contradiction to our assumption that the properties in Lemma 2.1 hold. If $\pi = q$, since $\pi \mid \sigma(\ell)$, we again get $\pi \mid \gcd(n, \sigma(n))$, a contradiction. So, given h, ℓ we have from (3.10) that q is in a fixed coprime residue class modulo h; say

$$q \equiv a_{h,\ell} \pmod{h}.$$

Similarly, we have $m' = q'\ell'$ and $q' \equiv a_{h,\ell'} \pmod{h}$.

Since $h \mid \gcd(s(m), s(m'))$, (3.5) implies that $h \mid \sigma(m) - \sigma(m')$, so that $m \equiv m' \pmod{h}$. With (3.10) we get that

$$\frac{\ell\sigma(\ell)}{s(\ell)} \equiv -q\ell = -m \equiv -m' = -q'\ell' \equiv \frac{\ell'\sigma(\ell')}{s(\ell')} \pmod{h},$$

which implies

$$s(\ell')\ell\sigma(\ell) - s(\ell)\ell'\sigma(\ell') \equiv 0 \pmod{h}. \tag{3.11}$$

The absolute value of the left-hand side is $< 2\max\{\ell^3, \ell'^3\} < 2x^{3/10}$. Thus, for $h > x^{1/3}$, then it must be the case that the integer in the left-hand side of the above congruence must be the zero integer. We thus get that

$$\frac{\ell\sigma(\ell)}{s(\ell)} = \frac{\ell'\sigma(\ell')}{s(\ell')}, \quad \text{or equivalently,} \quad \frac{\ell^2}{s(\ell)} + \ell = \frac{\ell'^2}{s(\ell')} + \ell'. \tag{3.12}$$

We have $\gcd(\ell, s(\ell)) = \gcd(\ell', s(\ell')) = d$. Further, by property (3) in Lemma 2.1, $d\,\mathrm{rad}(d) \mid \gcd(\sigma(\ell), \sigma(\ell'))$, where $\mathrm{rad}(d)$ is the largest squarefree divisor of d. Hence, $\gcd(\ell^2, s(\ell)) = d$ and the same is true for $\gcd(\ell'^2, s(\ell'))$. Putting $\ell = d\lambda,\ \ell' = d\lambda'$, we get that

$$\frac{d\lambda^2}{s(\ell)/d} - \frac{d\lambda'^2}{s(\ell')/d} = \ell - \ell',$$

and the two fractions appearing in the left-hand side above are reduced. So, their denominators must be equal, that is, $s(\ell)/d = s(\ell')/d$, therefore $s(\ell) = s(\ell')$. Now equation (3.12) gives

$$\ell^2 + \ell s(\ell) = \ell'^2 + \ell' s(\ell),$$

and since the function $t^2 + ts(\ell)$ is increasing in t, this gives $\ell = \ell'$. Thus, in the case $h > x^{1/3}$, we must have $\ell = \ell'$ and the congruence classes $a_{h,\ell}, a_{h,\ell'}$ of q and q' modulo h are the same.

Summing the expression in (3.9) over m, m' where $h \mid \gcd(s(m), s(m'))$, $h > x^{1/3}$, and using the maximal order of $A'_{3,2}/\varphi(A'_{3,2})$, we have

$$\frac{dx\log\log x}{(\log x)^2} \sum_{m,m',h} \frac{h}{mm'} = \frac{dx\log\log x}{(\log x)^2} \sum_{q,q',\ell,h} \frac{h}{qq'\ell^2}.$$

Since $\ell = \ell'$ and $m \neq m'$, we have $q \neq q'$; assume that $q > q'$. Since $q \equiv q' \equiv a_{h,\ell} \pmod{h}$, the sum of $1/q$ above is $O((\log x)/h)$, even forgetting that q is prime. Thus, the above sum reduces to

$$\frac{dx\log\log x}{\log x} \sum_{q',\ell,h} \frac{1}{q'\ell^2} \le \frac{dx\log\log x}{\log x} \sum_{q',\ell} \frac{\tau(s(q'\ell))}{q'\ell^2} \le x(\log x)^{O(1)} \sum_{q',\ell} \frac{1}{q'\ell^2},$$

by Lemma 2.4. Now $\sum 1/q' \ll 1$ and $\sum 1/\ell^2 \ll x^{-1/15}$, so we have the estimate

$$x^{14/15}(\log x)^{O(1)} = O\left(\frac{x}{d\log y}\right),$$

which is consistent with (3.9).

3.2 $h \le x^{1/3}$

We now consider values of h with $h \le x^{1/3}$. Since $s(m')$ is deficient, we have $s(m')/\varphi(s(m')) \ll 1$, so that $A'_{3,2}/\varphi(A'_{3,2}) \ll A''_{3,2}/\varphi(A''_{3,2})$, where $A''_{3,2}$ is the largest divisor of $A'_{3,2}$ coprime to $s(m')$. Fix m', h with $h \mid s(m')$ and consider numbers m that can arise. As noted before,

$$m \equiv \sigma(m) \equiv \sigma(m') \equiv m' \pmod{h}.$$

Since $h \mid s(m)$, $\gcd(m, \sigma(m)) = d$, we have $\gcd(m, h) = \gcd(\sigma(m), h) = 1$. Recall that $m = qrk$. Thus, the above congruences, rewritten as

$$qrk \equiv (q+1)(r+1)\sigma(k) \equiv m' \pmod{h},$$

determine $u := qr \pmod{h}$ and $v := q + r \pmod{h}$, when k, m' are given. As we have seen, Lemma 2.2 allows us to assume that h is squarefree. This implies that the number of solutions to the congruence $t^2 - vt + u \equiv 0 \pmod{h}$ is at most $\tau(h)$. That is, there are at most $\tau(h)$ pairs $a, b \pmod{h}$ such that we have $q \equiv a \pmod{h}$ and $r \equiv b \pmod{h}$. Let $\mathcal{S}_{h,k}$ denote the set of pairs a, b that arise for m', h, k.

For m', h, k, and the pair a, b in $\mathcal{S}_{h,k}$ all given, define

$$f_{m',h,k,a,b}(qr) = f(qr) = \sum_{\substack{\pi \mid \sigma(kqr) - \sigma(m') \\ (\log\log x)^2 < \pi \le \log x \\ \pi \nmid h\sigma(kqr)}} \frac{1}{\pi},$$

where π runs over primes. Note that if $f(qr) \le 1$, then $A''_{3,2}/\varphi(A''_{3,2}) \ll 1$. Say k, r are given and $\pi \mid \sigma(kqr) - \sigma(m')$ and $\pi \nmid \sigma(kqr)$. Since

$$q\sigma(kr) = -\sigma(kr) + \sigma(kqr) \equiv -\sigma(kr) + \sigma(m') \pmod{\pi},$$

if kr, π are fixed, then q is in a residue class modulo π, say $c_{\pi,kr} \pmod{\pi}$. To summarize, with m', h, kr, a, b fixed, if $m = kqr$ has $\pi \mid A''_{3,2}$, we have $q \equiv c_{\pi,kr} \pmod{\pi}$, $q \equiv a \pmod{h}$, $r \equiv b \pmod{h}$. Since $\pi \nmid h$, the two congruences for q may be combined to put q in a single residue class modulo πh. Thus, using $q > x^{7/20}$, $h \le x^{1/3}$, $\pi \le \log x$, and the Brun–Titchmarsh inequality,

$$\sum_{qr} \frac{f(qr)}{qr} \ll \sum_{\pi} \frac{1}{\pi} \sum_{r} \frac{1}{r} \sum_{q} \frac{1}{q} \ll \sum_{\pi} \frac{1}{\pi\varphi(\pi h)} \sum_{r} \frac{1}{r} \ll \sum_{\pi} \frac{1}{\pi^2 h} \sum_{r} \frac{1}{r}.$$

To estimate $\sum_r \frac{1}{r}$ we consider two ranges for h. Since $r \equiv b \pmod{h}$, we have

$$\sum_{r} \frac{1}{r} \ll \begin{cases} \dfrac{\log x}{x^{1/20}}, & \text{if } h > x^{1/20}, \\[2ex] \dfrac{1}{h}, & \text{if } h \le x^{1/20}. \end{cases}$$

Here, we are using that $r \in (x^{1/15}, x^{1/12}]$, a trivial estimate when $h > x^{1/20}$, and the Brun–Titchmarsh inequality with partial summation (as well as $\varphi(h) \gg h$) in the second case. Thus,

$$\sum_{qr} \frac{f(qr)}{qr} \ll \begin{cases} \dfrac{\log x}{hx^{1/20}}, & \text{if } h > x^{1/20}, \\[2mm] \dfrac{1}{h^2(\log\log x)^2}, & \text{if } h \le x^{1/20}. \end{cases} \tag{3.13}$$

The expression in (3.9) for $h \le x^{1/3}$ can be dealt with as follows. Fix m', h. Since $A'_{3,2}/\varphi(A'_{3,2}) \ll 1$ or $\log\log x/\log y$ depending on whether $f(qr) \le 1$ or $f(qr) > 1$,

$$\begin{aligned} \frac{x\log y}{(\log x)^2}\sum_m \frac{dhA'_{3,2}}{mm'\varphi(A'_{3,2})} &\ll \frac{x\log y}{(\log x)^2}\frac{dh}{m'}\sum_k \frac{1}{k}\sum_{\mathcal{S}_{h,k}}\left(\sum_{f(qr)\le 1}\frac{1}{qr} + \sum_{f(qr)>1}\frac{\log\log x}{qr\log y}\right) \\ &\le \frac{x\log y}{(\log x)^2}\frac{dh}{m'}\sum_k \frac{1}{k}\sum_{\mathcal{S}_{h,k}}\sum_{qr}\left(\frac{1}{qr} + \frac{f(qr)\log\log x}{qr\log y}\right). \end{aligned} \tag{3.14}$$

First assume that $x^{1/20} < h \le x^{1/3}$. By (3.13) we have

$$\sum_{qr}\left(\frac{1}{qr} + \frac{f(qr)\log\log x}{qr\log y}\right) \ll \frac{\log x\log\log x}{hx^{1/20}\log y}.$$

Thus, (3.14) and $\sum_k 1/k \ll (\log x)/d\log y$ imply that

$$\begin{aligned} \frac{x\log y}{(\log x)^2}\sum_m \frac{dhA'_{3,2}}{mm'\varphi(A'_{3,2})} &\ll \frac{x\log y}{(\log x)^2}\frac{dh}{m'}\sum_k \frac{1}{k}\frac{\tau(h)\log x\log\log x}{hx^{1/20}\log y} \\ &\ll \frac{x\log y}{(\log x)^2}\frac{dh}{m'}\frac{\log x}{d\log y}\frac{\tau(h)\log x\log\log x}{hx^{1/20}\log y} \\ &= \frac{x^{19/20}\log\log x}{\log y}\frac{\tau(h)}{m'}. \end{aligned}$$

Now we sum over choices for m', h. We have

$$\sum_{h|s(m')} \tau(h) \le \tau(s(m'))^2 \le (\log x)^{1.4},$$

using Lemma 2.4. Further, $\sum_{m'} 1/m' \ll (\log x)/d\log y$. Thus, the sum in (3.9) is at most $(x^{19/20}/d)(\log x)^{O(1)}$ when $x^{1/20} < h \le x^{1/3}$, which is certainly consistent with the inequality in (3.9).

It remains to consider the case $h \le x^{1/20}$. By (3.13), we have

$$\sum_{qr}\left(\frac{1}{qr}+\frac{f(qr)\log\log x}{qr\log y}\right) \ll \frac{1}{h^2}+\frac{\log\log x}{h^2\log y(\log\log x)^2} \ll \frac{1}{h^2}.$$

Thus, from (3.14) and $\sum_k 1/k \ll (\log x)/d\log y$,

$$\frac{x\log y}{(\log x)^2}\sum_m \frac{dhA'_{3,2}}{mm'\varphi(A'_{3,2})} \ll \frac{x\log y}{(\log x)^2}\frac{dh}{m'}\sum_k \frac{1}{k}\frac{\tau(h)}{h^2} \ll \frac{x}{\log x}\frac{\tau(h)}{hm'}.$$

Now $h \mid s(m')$ and we are assuming that $s(m')$ is deficient. Thus,

$$\sum_h \frac{\tau(h)}{h} \le \left(\sum_h \frac{1}{h}\right)^2 < 4.$$

So, summing the previous expression over h, m' we get an estimate which is $\ll x/d\log y$. This completes the proof of (3.9) and the theorem. □

Acknowledgements

Research of F. L. on this project was carried on while he visited the Mathematics Department of Dartmouth College in Spring 2014. F. L. thanks the Mathematics Department of Dartmouth College for their hospitality. We thank Paul Pollack for his interest in the paper, and we thank the referee for a careful reading and helpful queries.

References

[CZ] Y.-G. Chen and Q.-Q. Zhao, *Nonaliquot numbers*, Publ. Math. Debrecen 78 (2011), no. 2, 439–442.

[dB] 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.

[DL] J.-M. De Koninck and F. Luca, *On the composition of the Euler function and the sum of divisors function*, Colloq. Math. 108 (2007), no. 1, 31–51.

[E56] P. Erdős, *On perfect and multiply perfect numbers*, Ann. Mat. Pura Appl. 42 (1956), 253–258.

[E73] P. Erdős, *Über die Zahlen der Form* $\sigma(n) - n$ *und* $n - \varphi(n)$, Elem. Math. 11 (1973), 83–86.

[EGPS] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions* Analytic Number Theory, Proc. Conf. in honor of Paul T. Bateman, B. C. Berndt, et al. eds., Birkhauser, Boston, 1990, pp. 165–204.

[ELP] P. Erdős, F. Luca, and C. Pomerance, *On the proportion of numbers coprime to a given integer*, Anatomy of integers, 47–64, CRM Proc. Lecture Notes, 46, Amer. Math. Soc., Providence, RI, 2008.

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

[KS] I. Kátai and M. V. Subbarao, *Some further remarks on the iterates of the φ and the σ-functions*, Annales Univ. Sci. Budapest, Sect. Comp. 26 (2006), 51–63.

[K] M. Kobayashi, *On the density of abundant numbers*, PhD thesis, Dartmouth College, 2010.

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

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

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

[PY] C. Pomerance and H.-S. Yang, *Variant of a theorem of Erdős on the sum-of-proper-divisors function*, Math. Comp. 83 (2014), 1903–1913.

[T] L. Troupe, *On the number of prime factors of values of the sum-of-proper-divisors function*, arXiv 1405.3587, J. Number Theory (to appear).