

# LEHMER'S TOTIENT PROBLEM WITH FEWER THAN SIXTEEN PRIME FACTORS

DEEP BHATTACHARJEE

ABSTRACT. A composite number  $n$  whose totient divides  $n - 1$  was known to have at least fourteen prime factors. We prove that it has at least sixteen, at least 16001 if  $(n - 1)/\varphi(n) \geq 3$ , and more than  $10^8$  if  $3 \mid n$ , and that  $n < 2^{2^{k-7}}$  if  $n$  has  $k$  prime factors. For the companion equation  $\varphi(n) \mid n + 1$  we show that the nine known solutions are the only ones with at most eight prime factors, and that a further solution prime to 3 has at least sixteen. Each statement is reduced by hand to a finite computation, which independent programs carry out. Lehmer's conjecture remains open, and we state the assertion that would settle it.

## 1. INTRODUCTION

**1.1. Two questions of Lehmer.** Every prime  $p$  satisfies  $\varphi(p) = p - 1$ , and so  $\varphi(p)$  divides  $p - 1$ . In 1932 Lehmer [19] asked whether any composite number  $n$  has the same property,

$$\varphi(n) \mid n - 1. \quad (1.1)$$

Nobody has found one, and nobody has shown that none exists. Lehmer proved that a composite solution of (1.1) must be odd, squarefree and divisible by at least seven distinct primes. The question is Problem B37 in Guy's collection [16]. Lehmer's paper also lists solutions of the companion equation<sup>1</sup>

$$\varphi(n) \mid n + 1. \quad (1.2)$$

The two equations differ only in a sign. Equation (1.1) has no known composite solution, whereas (1.2) has the solutions

$$1, 2, 3, 15, 255, 65535, 83623935, 4294967295, 6992962672132095, \quad (1.3)$$

and no others are known. Except for 2, each of them satisfies  $n + 1 = 2\varphi(n)$ . Five of them are products of the first Fermat primes<sup>2</sup>, namely  $3, 15 = 3 \cdot 5, 255 = 3 \cdot 5 \cdot 17, 65535 = 3 \cdot 5 \cdot 17 \cdot 257$  and  $4294967295 = 3 \cdot 5 \cdot 17 \cdot 257 \cdot 65537$ . Apart from 1 and 2, the other two are  $83623935 = 3 \cdot 5 \cdot 17 \cdot 353 \cdot 929$  and  $6992962672132095 = 83623935 \cdot 83623937$ . These numbers form sequence A203966 of the OEIS [28].

Much of the literature on (1.1) concerns the number  $\omega(n)$  of distinct prime factors of a composite solution. Pomerance [31] proved that a composite solution with  $k$  prime factors satisfies  $n < k^{2^k}$ , so that for each  $k$  the question is a finite one, and Burek and Žmija [7] sharpened this to  $n \leq 2^{2^k} - 2^{2^{k-1}}$ , using a lemma that goes back to Heath-Brown's work on

2020 *Mathematics Subject Classification*. Primary 11A25; Secondary 11A41, 11A51, 11D68, 11N36, 11Y11, 11Y16, 11Y70.

*Key words and phrases*. Lehmer's totient problem, Euler's totient function, large sieve, Fermat primes.

<sup>1</sup>The comments to sequence A050474 of the OEIS [28] record this, and a misprint in Lehmer's summary where  $3 \cdot 5 \cdot 353 \cdot 929$  stands for  $83623935 = 3 \cdot 5 \cdot 17 \cdot 353 \cdot 929$ .

<sup>2</sup>The Fermat numbers  $F_i = 2^{2^i} + 1$  are prime for  $0 \leq i \leq 4$ , and  $F_0 F_1 \cdots F_{r-1} = 2^{2^r} - 1$ . For  $1 \leq r \leq 5$  the number  $n = 2^{2^r} - 1$  is therefore a product of primes  $F_i$ , and  $\varphi(n) = 2^{1+2+\cdots+2^{r-1}} = 2^{2^r-1} = (n+1)/2$ . The chain breaks at  $r = 6$  because  $F_5 = 641 \cdot 6700417$ , as Euler found.

odd perfect numbers [18] and was sharpened by Cook [10] and Nielsen [26].<sup>3</sup> In binary, their bound allows  $n$  one digit for each subset of the set of its prime factors. Cohen and Hagis [8] settled all  $k \leq 13$  and obtained  $\omega(n) \geq 14$ . When  $3 \mid n$  much more is true: Hagis [17] proved  $\omega(n) \geq 298848$ , and Burcsi, Czirbusz and Farkas [6] raised this to  $\omega(n) \geq 40000000$ . In the same paper they report that a computation of Renze [33], made available as a Mathematica notebook, gives  $\omega(n) \geq 15$  in general, and Pinch [30] outlined a proof of the same bound in a poster presented at ANTS VII. Neither account has been refereed. Since  $p - 1 \mid n - 1$  for every prime  $p$  dividing a solution, a composite solution of (1.1) would be a Carmichael number;<sup>4</sup> Banks, Nevans and Pomerance [3] discuss this and the connection with Giuga's conjecture.<sup>5</sup>

More is known when the quotient  $M = (n - 1)/\varphi(n)$  exceeds 2. Hagis [17] proved  $\omega(n) \geq 1991$  for  $M = 3$ , and Grytczuk and Wójtowicz [15] gave lower bounds for  $\omega(n)$  that grow exponentially with  $M \geq 4$  (see [37] for both). The prime factors of a solution form what Golomb [14] called a  $G$ -sequence: none of them divides another minus one. Meijer [22], who points out the connection with (1.1), proved with the Brun–Selberg sieve that the product of  $p/(p - 1)$  over the elements of a  $G$ -sequence up to  $x$  is at most a constant times  $\log \log x$ . Yamada [37] made this explicit with the large sieve and deduced that  $M < 16.03235 \log \log \omega(n)$  for every composite solution of  $\varphi(n) \mid n \pm 1$  with  $\omega(n) \geq 4$ , so that  $\omega(n)$  is at least doubly exponential in  $M$ .

Norman [27] claimed that  $M \geq 3$  for every composite solution of (1.1), and hence  $\omega(n) \geq 1991$ . The proof rests on [27, Lemma 2.1], which asserts that  $\psi(G)/|G|^2 > \varphi(m)/(2m)$  for every noncyclic nilpotent group  $G$  of order  $4m$  with  $m$  odd and squarefree, where  $\psi(G)$  is the sum of the orders of the elements of  $G$ . The lemma is false. For  $G = C_2 \times C_2 \times C_m$  with  $m$  prime to 3 we have  $\psi(G) = 7 \prod_{p \mid m} (p^2 - p + 1)$ , since  $\psi$  is multiplicative over factors of coprime order, and the ratio of  $\psi(G)/|G|^2$  to  $\varphi(m)/(2m)$  is  $\frac{7}{8} \prod_{p \mid m} (1 + \frac{1}{p(p-1)})$ . The product over all primes  $p \geq 5$  is  $\frac{4}{7} \zeta(2) \zeta(3) / \zeta(6) = 1.1106 \dots < \frac{8}{7}$ , so the ratio is less than 1; for  $m = 5$  the two sides are  $147/400$  and  $2/5$ . The proof of the lemma uses  $\psi(V)/16 \geq \frac{1}{2}$  for the Sylow 2-subgroup  $V$ , whereas  $\psi(C_2 \times C_2) = 7$ . The claim  $M \geq 3$  is therefore unproved; if it holds, Theorem 1.1 below gives  $\omega(n) \geq 16001$  for every composite solution.

Less is known about (1.2). The comments to sequences A050474 and A203966 of the OEIS [28] report that the M.Sc. thesis of Wong [36] shows that a further solution of  $n + 1 = 2\varphi(n)$  would need at least eight prime factors, that there are no further solutions of  $n + 1 = 2\varphi(n)$  below  $10^{31}$  and none of (1.2) below  $1.6 \cdot 10^{30}$  (Aleksyev), and that  $(n + 1)/\varphi(n) = 3$  would force at least 33 prime factors. None of this has appeared in a refereed paper, and we know of no refereed bound on  $\omega(n)$  for (1.2).

**1.2. Results.** We write  $\omega(n)$  for the number of distinct prime factors of  $n$ . Our first theorem goes one prime further than the bound announced by Renze and by Pinch, and its proof includes a complete proof of that bound.

**Theorem 1.1.** *Let  $n$  be composite with  $\varphi(n) \mid n - 1$ . Then  $\omega(n) \geq 16$ . If  $(n - 1)/\varphi(n) \geq 3$ , then  $\omega(n) \geq 16001$ , and if  $3 \mid n$ , then  $\omega(n) > 10^8$ .*

<sup>3</sup>Heath-Brown proved that an odd perfect number with  $k$  distinct prime factors is less than  $4^{4^k}$ , and Nielsen lowered this to  $2^{4^k}$ . The lemma of Cook and Nielsen concerns integers  $x_1 \leq \dots \leq x_m$  for which  $\prod (1 - 1/x_i)$  lies just below a given fraction; we state and sharpen it in Section 6.1.

<sup>4</sup>By Korselt's criterion a composite  $n$  satisfies  $a^{n-1} \equiv 1 \pmod{n}$  for every  $a$  prime to  $n$  exactly when  $n$  is squarefree and  $p - 1 \mid n - 1$  for every prime  $p \mid n$ . A solution of (1.1) is squarefree by Lemma 2.1, and  $p - 1 \mid \varphi(n) \mid n - 1$ .

<sup>5</sup>Giuga conjectured in 1950 that an integer  $n > 1$  with  $1^{n-1} + 2^{n-1} + \dots + (n - 1)^{n-1} \equiv -1 \pmod{n}$  is prime. He showed that a composite  $n$  satisfies this congruence exactly when  $p \mid n/p - 1$  and  $p - 1 \mid n/p - 1$  for every prime  $p \mid n$ ; the second condition makes  $n$  a Carmichael number.

Apart from Norman's claim, we know of no earlier claim beyond fifteen prime factors in general. The second assertion raises the bound 1991 of Hags for the quotient 3 to 16001, and for  $3 \nmid n$  it also improves the bounds of Grytczuk and Wójtowicz for the quotients from 4 to 7. For  $3 \mid n$  the third assertion raises the bound of Buresi, Czirbusz and Farkas from 40000000 to more than  $10^8$ . Our second theorem bounds the size of a solution in terms of the number of its prime factors.

**Theorem 1.2.** *Let  $n$  be composite with  $\varphi(n) \mid n - 1$ , let  $k = \omega(n)$ , and put  $M = (n - 1)/\varphi(n)$ .*

- (i)  $n < 2^{2^{k-7}}$ .
- (ii) If  $M \geq 3$ , then  $n < 2^{2^{k-15981}}$ .
- (iii) If  $3 \mid n$ , then  $n < 2^{2^{k-10^8}}$ .

Part (i) divides the number of binary digits allowed by the bound of Burek and Żmija by  $2^7 = 128$ . Together with Theorem 1.1 it shows that a solution with sixteen prime factors, the first case not yet excluded, lies below  $2^{512} < 1.35 \cdot 10^{154}$ .

The other results concern (1.2).

**Theorem 1.3.** *The solutions of  $\varphi(n) \mid n + 1$  with  $\omega(n) \leq 8$  are exactly the nine numbers (1.3).*

**Theorem 1.4.** *Every solution of  $\varphi(n) \mid n + 1$  with  $n > 3$  and  $3 \nmid n$  has  $\omega(n) \geq 16$ .*

**Theorem 1.5.** *Let  $n > 3$  satisfy  $\varphi(n) \mid n + 1$  with  $(n + 1)/\varphi(n) \geq 3$ . Then  $\omega(n) \geq 16001$ , and  $\omega(n) > 10^8$  if  $3 \mid n$ .*

The next theorem explains the list (1.3). Say that  $n$  is a *Fermat-type* solution if  $n + 1 = 2\varphi(n)$ ; by convention  $n = 1$  is one.

**Theorem 1.6.** *Let  $n_0$  be a Fermat-type solution.*

- (i) *For a prime  $q \nmid n_0$ , the number  $n_0q$  is a Fermat-type solution if and only if  $q = n_0 + 2$ .*
- (ii) *For primes  $p < q$  not dividing  $n_0$ , the number  $n_0pq$  is a Fermat-type solution if and only if*

$$(p - n_0 - 1)(q - n_0 - 1) = n_0^2 + n_0 + 1.$$

- (iii) *The Fermat-type solutions in (1.3), which are all of them except 2, are exactly the numbers obtained from 1 by repeated use of (i) and (ii): every such extension of one of them is again in (1.3).*

Theorem 1.3 contains the bound reported from Wong's thesis and goes one prime factor further: a solution of (1.2) outside (1.3) has at least nine prime factors, whatever the quotient  $(n + 1)/\varphi(n)$ . Theorem 1.5 raises the 33 prime factors of the observation on the quotient 3 recorded in the OEIS to 16001. Part (iii) of Theorem 1.6 places 83623935 in the same family as the products of Fermat primes: it is the two-prime extension of 255 coming from the factorisation  $255^2 + 255 + 1 = 65281 = 97 \cdot 673$ , just as 4294967295 comes from the trivial factorisation  $65281 = 1 \cdot 65281$  (Figure 15).

The last result of this section is not new. It follows from the bound of Meijer [22], and Yamada [37] proved it with  $c = 1/16.03235$  when  $\omega(n) \geq 4$ . We give a short proof by the large sieve in Section 2.2, because the estimate behind it explains the size of the bounds in Theorems 1.1 and 1.5.

**Theorem 1.7** (Meijer, Yamada). *There is an absolute and effective constant  $c > 0$  with the following property. Let  $n$  be composite with  $\varphi(n) \mid n - 1$ , or let  $n > 3$  with  $\varphi(n) \mid n + 1$ , and let  $M$  be the quotient  $(n - 1)/\varphi(n)$  or  $(n + 1)/\varphi(n)$ . Then*

$$\omega(n) \geq \exp(\exp(cM)) - 2.$$

The first primes alone give only  $\omega(n) \geq \exp(c'M)$ , since the product of  $p/(p-1)$  over the first  $K$  primes grows like  $\log K$ . Theorem 1.7 uses the fact that no prime factor of a solution divides another prime factor minus one, and it is proved by the large sieve (Section 2.2). It explains why the bounds 16001 and  $10^8$  in Theorems 1.1 and 1.5 lie so far above the bounds 33 and 1540 that the first primes give.

**1.3. Theory and computation.** Every search in this paper rests on a theorem, proved by hand, that reduces the statement to finitely many cases and shows that the search covers all of them. The results divide as follows.

*Proved entirely by hand.* Theorem 1.7 and Meijer's bound (Theorem 2.7), by the large sieve; the sharp product lemma (Theorem 6.2) and the size bound  $n < 2^{2^{k-2}}$  (Proposition 6.11); the reductions of Section 2, up to the evaluation of the explicit finite products in Lemmas 2.2 and 2.3; parts (i) and (ii) of Theorem 1.6; and the equivalence of Proposition 9.1.

*Proved by hand up to a finite computation.*

- $\omega(n) \geq 16$  in Theorem 1.1, and Theorem 1.4: the interval bounds, the divisibility condition for the last two primes and the completeness of the search (Propositions 3.2 and 3.5, Theorem 3.6, Corollary 3.3), and the reduction of the last three primes to divisors in a residue class (Proposition 4.1, Theorem 4.4, Proposition 4.5), leave a search over at most fifteen primes, which ends in 33865004 two-prime problems for fifteen (Section 5).
- The bounds 16001 and  $10^8$  in Theorems 1.1 and 1.5: Corollary 2.6 reduces them to the branch-and-bound of Proposition 2.5.
- Theorem 1.2: Corollary 6.10 reduces part (i) to the tree  $\mathcal{T}_7$  of 33678 nodes (Theorem 6.15), and parts (ii) and (iii) to the evaluation of two explicit rational numbers.
- Theorem 1.3: the same reductions leave 228260561 two-prime problems for eight primes (Section 7.4).
- Theorem 1.6(iii): the factorisation of eight integers.

Every search was carried out by at least two programs written independently, and every count reported here is the same for all of them (Section 8). A search that finds nothing proves something only if it would have found what it looks for, and the programs were therefore also run on equations that have solutions.

**1.4. Scope.** Theorems 1.1–1.7 bound the number of prime factors and the size of a new solution and describe the known ones; they do not decide whether (1.1) has a composite solution or whether (1.2) has solutions beyond (1.3). Section 9 estimates the cost of the next cases and states the assertion, equivalent to Lehmer's conjecture, that this paper does not prove.

**1.5. Method.** The bound  $\omega(n) \geq 16$  of Theorem 1.1 and Theorems 1.3 and 1.4 rest on three reductions proved in Sections 2 to 4. The first (Section 2) shows that one may assume  $n + \varepsilon = 2\varphi(n)$ , with  $\varepsilon = -1$  for Lehmer's equation and  $\varepsilon = +1$  for the companion, and, except in Theorem 1.3, that  $3 \nmid n$ . The second (Section 3) shows that once the first  $j$  prime factors are fixed, the next one lies in an interval whose ends are explicit functions of them (Proposition 3.2), that for each  $k$  only finitely many sequences of primes satisfy these conditions (Corollary 3.3), and that the last two prime factors are determined by a divisibility condition (Proposition 3.5). Each value of  $k$  is therefore a finite problem, and the traversal of an explicit finite tree finds every solution (Theorem 3.6). The bounds for the two signs of  $\varepsilon$  differ only by terms of size  $1/\varphi(n)$ ; for  $p_1 \geq 5$  they give identical trees for every  $k \leq 14$ , and for  $k = 15$  the same nodes down to depth 13 (Figure 8). The third reduction (Section 4) treats the last three primes together: once  $p_{k-2}$  is fixed, the last two correspond to a divisor of an explicit integer  $N$  in a known residue class modulo an integer  $c$  (Proposition 4.1). Such divisors are the lattice points of a coset near a

hyperbola (Lemmas 4.2 and 4.3). Since the two complementary divisors lie in the same class, their sum lies in a single class modulo  $c$ , which reduces the problem to one square test for each value of the sum (Section 4.4), and when  $c^3 \geq N + 2c^2$  there are at most four such divisors (Proposition 4.5). With these reductions the case  $k \leq 14$  becomes a tree of at most 31962 nodes and the case  $k = 15$  a set of 33865004 problems of this kind, which we solved by computer. One of the routes never factors an integer, so that for every  $k \leq 15$  the conclusion depends on no primality proof.

The companion equation with eight prime factors and  $3 \mid n$  behaves differently. After the Fermat primes 3, 5, 17, 257, 65537 the product  $\prod p/(p-1)$  falls short of 2 by only  $2^{-31}$ , the interval for the sixth prime has length  $2^{33}$ , and each of the  $1.3 \cdot 10^8$  admissible choices of that prime leaves a two-prime problem whose modulus is far below the cube root of the integer to be split. We treat these problems by the sum route of Section 4.4, combined with trial division for the small divisors and sieved by congruences that primes satisfy, and we settle the 1967265 cases in which this would be slow by factoring the integer completely, with every prime factor proved prime (Section 7.4). Theorem 1.6 rests on a short identity together with the factorisation of eight integers.

The bounds for a quotient of at least 3 and for  $3 \mid n$ , in Theorems 1.1 and 1.5, come from independence. The prime factors of a solution form a set in which no prime divides another minus one, and over such a set the product of  $p/(p-1)$  grows only like the double logarithm of its size (Theorem 2.7). A branch-and-bound over these sets (Proposition 2.5) shows that such a set must be very large before the product reaches 3, as it must when the quotient is at least 3 and  $3 \nmid n$ , or reaches  $\frac{8}{3}$  with all its primes congruent to 2 modulo 3, as it must for the prime factors other than 3 when  $3 \mid n$ .

Theorem 1.2 is proved in Section 6. Its first ingredient is a sharp form of the lemma of Cook and Nielsen (Theorem 6.2), in which the bound for the product of the remaining prime factors depends on how far the product  $\prod p/(p-1)$  over the prime factors already chosen falls short of  $M$ . This lemma alone gives  $n < 2^{2^{k-2}}$  (Proposition 6.11), and more generally  $n < 2^{2^{k-s}}$  for every solution whose first prime factors are small in a precise sense. In every other solution the prime factors grow rapidly, and for  $s = 7$  an exhaustive search over their possible initial segments, which visits 33678 of them, shows that there are none. Parts (ii) and (iii) combine the lemma of Cook and Nielsen with Proposition 2.5.

The paper is organised as follows. Section 2 proves the reductions, the bounds for large quotients and for  $3 \mid n$ , Theorem 1.5, and, by the large sieve, Theorem 1.7. Section 3 proves the interval bounds and the completeness of the search, and Section 4 the reduction of the last three primes to divisors in a residue class. Section 5 completes the proof of Theorem 1.1, Section 6 proves Theorem 1.2, and Section 7 proves Theorems 1.3, 1.4 and 1.6. Section 8 summarises the checks on the computations, and Section 9 states the assertion that remains open.

**Notation.** Throughout,  $n = p_1 p_2 \cdots p_k$  with primes  $p_1 < \cdots < p_k$ , and  $\varepsilon \in \{-1, +1\}$  selects the equation  $\varphi(n) \mid n + \varepsilon$ ; we put  $M = (n + \varepsilon)/\varphi(n)$ . For  $0 \leq j \leq k$  let

$$A_j = \prod_{i \leq j} p_i, \quad B_j = \prod_{i \leq j} (p_i - 1), \quad P_j = \frac{A_j}{B_j} = \prod_{i \leq j} \frac{p_i}{p_i - 1},$$

with  $A_0 = B_0 = P_0 = 1$  and  $p_0 = 1$ . We write  $m = k - j$  for the number of primes still to be chosen at depth  $j$ , and  $r_1 < r_2 < \cdots$  for the primes from 5 on.

## 2. REDUCTIONS

**Lemma 2.1.** *Let  $n > 1$  satisfy  $\varphi(n) \mid n + \varepsilon$  and put  $M = (n + \varepsilon)/\varphi(n)$ . Suppose  $n$  is composite if  $\varepsilon = -1$ , and  $n > 3$  if  $\varepsilon = +1$ .*

- (i)  $n$  is odd, squarefree and composite, and  $M \geq 2$ .
- (ii) For primes  $p, q$  dividing  $n$ , possibly equal,  $p \nmid q - 1$  and  $p \nmid M$ .
- (iii)  $\prod_{p|n} \frac{p}{p-1} = M - \frac{\varepsilon}{\varphi(n)}$ .

*Proof.* If  $n > 2$  is even then  $\varphi(n)$  is even and  $n + \varepsilon$  is odd. If  $p^2 \mid n$  then  $p$  divides  $\varphi(n)$ , hence  $n + \varepsilon$ , which is impossible since  $p \mid n$ . For  $\varepsilon = +1$  a prime solution satisfies  $n - 1 \mid 2$ , so  $n \leq 3$ ; thus  $n$  is composite in both cases. For odd squarefree composite  $n$  we have  $\varphi(n) < n - 1$ , so  $M \geq 2$ . If  $p, q \mid n$  then  $q - 1 \mid \varphi(n) \mid n + \varepsilon$  while  $p \mid n$ , so  $p \nmid q - 1$ ; similarly  $p \nmid M\varphi(n) = n + \varepsilon$ . Finally  $n/\varphi(n) = (n + \varepsilon)/\varphi(n) - \varepsilon/\varphi(n)$ .  $\square$

The next lemma shows how the prime 3 enters, and that it does so differently for the two equations.

**Lemma 2.2.** *In the situation of Lemma 2.1, suppose  $3 \mid n$ . Then every other prime factor of  $n$  is congruent to 2 modulo 3, and*

$$M \equiv \begin{cases} 1 & (\text{mod } 3), \quad \varepsilon = -1, \\ 2 & (\text{mod } 3), \quad \varepsilon = +1. \end{cases}$$

Consequently  $M \geq 4$  if  $\varepsilon = -1$ , and  $M = 2$  or  $M \geq 5$  if  $\varepsilon = +1$ . If  $M \geq 4$  then  $\omega(n) \geq 1540$ .

*Proof.* For a prime  $p \mid n$  with  $p \neq 3$ , Lemma 2.1(ii) gives  $3 \nmid p - 1$ , so  $p \equiv 2$  and  $p - 1 \equiv 1 \pmod{3}$ . Hence  $\varphi(n) = 2 \prod_{p \neq 3} (p - 1) \equiv 2 \pmod{3}$ , while  $n + \varepsilon \equiv \varepsilon \pmod{3}$ . The relation  $n + \varepsilon = M\varphi(n)$  becomes  $2M \equiv \varepsilon$ , which is the stated congruence. Suppose  $M \geq 4$ . Then  $n/\varphi(n) > 4$ : for  $\varepsilon = -1$  because  $n/\varphi(n) = M + 1/\varphi(n)$  by Lemma 2.1(iii), and for  $\varepsilon = +1$  because then  $M \geq 5$  and  $n/\varphi(n) = M - 1/\varphi(n) \geq 5 - \frac{1}{2}$ . On the other hand, if  $q_1 < q_2 < \dots$  are the primes congruent to 2 modulo 3 from 5 on, then

$$\frac{n}{\varphi(n)} = \frac{3}{2} \prod_{p|n, p \neq 3} \frac{p}{p-1} \leq \frac{3}{2} \prod_{i=1}^{k-1} \frac{q_i}{q_i - 1},$$

since  $p/(p-1)$  decreases in  $p$ . An exact computation in rational arithmetic shows that the right-hand side is  $3.99985\dots$  for  $k = 1539$  and first exceeds 4 at  $k = 1540$ , where the last prime in the product is  $q_{1539} = 27947$ .  $\square$

For  $\varepsilon = -1$  this is a weak form of the bounds of Hagis and of Burcsi, Czirbusz and Farkas mentioned in the introduction, and Corollary 2.6 below goes much further, but the lemma needs no search and suffices wherever we use it. For  $\varepsilon = +1$  it shows why 3 cannot be removed in the same way:  $M = 2$  is compatible with  $3 \mid n$ , and each Fermat prime 5, 17, 257, 65537 is congruent to 2 modulo 3, as are 353 and 929.

**Lemma 2.3.** *In the situation of Lemma 2.1, suppose  $3 \nmid n$  and  $\omega(n) \leq 32$ . Then  $M = 2$ .*

*Proof.* All prime factors are at least 5. By Lemma 2.1(iii) and monotonicity,

$$M = \frac{n}{\varphi(n)} + \frac{\varepsilon}{\varphi(n)} \leq \prod_{i \leq k} \frac{r_i}{r_i - 1} + \frac{1}{\prod_{i \leq k} (r_i - 1)},$$

where  $k = \omega(n)$ , since  $n/\varphi(n) = \prod_{p|n} p/(p-1)$  and  $\varphi(n) = \prod_{p|n} (p-1)$ , and the  $i$ -th smallest prime factor of  $n$  is at least  $r_i$ . An exact computation shows that the right-hand side is less than 3 for every  $k \leq 32$ . (For  $k = 32$  the product is  $2.99315\dots$  and the second term is below  $10^{-53}$ ; at  $k = 33$  the product exceeds 3.) Since  $M$  is an integer and  $M \geq 2$  by Lemma 2.1(i),  $M = 2$ .  $\square$

**Corollary 2.4.** *Let  $n > 3$  satisfy  $\varphi(n) \mid n + 1$  with  $\omega(n) \leq 7$ . Then  $n + 1 = 2\varphi(n)$ .*

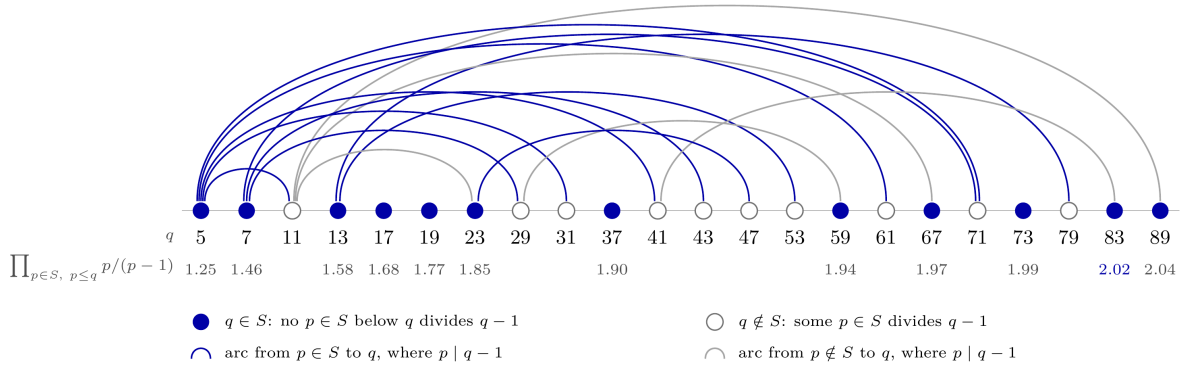


FIGURE 1. The primes from 5 to 89, with an arc from  $p$  to  $q$  whenever  $p$  divides  $q-1$ ; an independent set contains no two primes joined by an arc. The set  $S$  of filled primes is chosen one prime at a time, each the least prime joined to none of those chosen before it. Below each prime of  $S$  is the product of  $p/(p-1)$  over the primes of  $S$  up to it, which first exceeds 2 at 83, the eleventh.

*Proof.* Let  $k = \omega(n) \geq 2$ . The prime factors of  $n$ , in increasing order, are at least 3,  $r_1, \dots, r_{k-1}$ , and Lemma 2.1(iii) gives, as in the proof of Lemma 2.3,

$$M \leq \frac{3}{2} \prod_{i < k} \frac{r_i}{r_i - 1} + \frac{1}{2 \prod_{i < k} (r_i - 1)}.$$

The maximum of the right-hand side over  $2 \leq k \leq 7$  is  $2.92357\dots < 3$ , attained at  $k = 7$ , and so  $M = 2$ .  $\square$

**2.1. Independent sets of primes.** Call a finite set of primes *independent* if none of its elements divides another minus one. Golomb [14] and Meijer [22] call an increasing sequence of odd primes with this property a  $G$ -sequence. By Lemma 2.1(ii) the prime factors of a solution form an independent set, whereas the proofs above used only that the  $i$ -th smallest of them is at least the  $i$ -th prime of a given list. Independence is a much stronger condition. Each prime  $q$  of the set excludes the primes congruent to 1 modulo  $q$ , which make up the proportion  $1/(q-1)$  of all primes by the prime number theorem for arithmetic progressions, and so the product of  $p/(p-1)$  over an independent set grows far more slowly with its size than the product over the first primes. Choosing at each step the least prime from 5 on that the primes already chosen allow gives the independent set

$$5, 7, 13, 17, 19, 23, 37, 59, 67, 73, 83, \dots$$

(Figure 1). Its product passes 2 at the eleventh prime, where the first primes from 5 on need only seven, and it passes 3 only at the 100470-th prime, 5160959, where they need 33 (Figure 2).

**Proposition 2.5.** (i) No independent set of at most 16000 primes  $p \geq 5$  has  $\prod p/(p-1) \geq 2.999999$ .

(ii) No independent set of at most  $10^8$  primes  $p \equiv 2 \pmod{3}$  with  $p \geq 5$  has  $\prod p/(p-1) \geq \frac{8}{3}(1 - 10^{-6})$ .

(iii) No independent set of at most  $10^8$  primes  $p \equiv 2 \pmod{3}$  with  $p \geq 5$  has  $\prod p/(p-1) \geq \frac{10}{3}(1 - 10^{-6})$ .

*Proof.* Let  $Q$  be the set of primes allowed,  $K$  the size and  $V$  the target. For an independent set  $S \subseteq Q$  whose largest element is  $p$  (with  $p = 3$  if  $S$  is empty), call the primes  $q \in Q$  with  $q > p$  and  $q \not\equiv 1 \pmod{\text{every element of } S}$  the *candidates* for  $S$ . If  $|S| < K$ , let  $B(S)$  be the product

of  $q/(q-1)$  over  $S$  and over the  $K - |S|$  smallest candidates for  $S$ . An independent set  $T$  of at most  $K$  primes of  $Q$  whose  $|S|$  smallest elements form  $S$  has its other elements among the candidates for  $S$ , and there are at most  $K - |S|$  of them; as  $q/(q-1)$  decreases in  $q$ , the product over  $T$  is at most  $B(S)$ .

The search starts from the empty set and treats each set  $S$  that it creates as follows. If the product over  $S$  is at least  $V$ , it stops and reports  $S$ . If  $|S| = K$  or  $B(S) < V$ , it discards  $S$ . Otherwise it creates the sets  $S \cup \{q\}$  for the candidates  $q$  with

$$\prod_{p \in S} \frac{p}{p-1} \cdot \frac{q}{q-1} \cdot \prod_{q'} \frac{q'}{q'-1} \geq V, \quad (2.1)$$

where  $q'$  runs over the  $K - |S| - 1$  candidates for  $S$  that follow  $q$ . The left-hand side of (2.1) is at least  $B(S \cup \{q\})$ , because every candidate for  $S \cup \{q\}$  is a candidate for  $S$ , and it decreases as  $q$  increases, so the sets created from  $S$  come from its first few candidates. Let  $T$  be as above with product at least  $V$ . Every initial segment  $S$  of  $T$  has  $B(S) \geq V$ , and if  $S \cup \{q\}$  is also an initial segment, then  $q$  satisfies (2.1); so the search creates the initial segments of  $T$  one after another until it stops.

In each of the three cases the search ends without reporting a set. It creates 90320 sets in case (i), the largest of them with 46 elements, and 9 sets in case (ii) and 2 in case (iii). Two programs written independently, one of them in PARI/GP [29], carry out the three searches and create the same sets. For (ii) and (iii), which involve primes up to  $2 \cdot 10^{10}$ , they run through the primes once for each set created. The products are handled through their logarithms, and every bound is a sum of a few terms  $\log(q/(q-1))$  and a difference of two running sums of such terms. Every running sum here stays below 4; it has fewer than  $2 \cdot 10^6$  terms in case (i), where the primes below  $3 \cdot 10^7$  suffice, and fewer than  $1.1 \cdot 10^8$  in cases (ii) and (iii). In double precision each term is computed with an absolute error below  $3 \cdot 10^{-16}$  and each addition with one below  $2^{-52}$ , so a bound carries an error below  $10^{-8}$  in case (i) and below  $2 \cdot 10^{-7}$  in cases (ii) and (iii). The PARI/GP programs work with 57 significant digits in case (i) and 38 in cases (ii) and (iii), and their errors are below  $10^{-45}$  and  $10^{-25}$ . A set is discarded, or a candidate passed over, only when exact rational arithmetic confirms it or when the logarithm of its bound lies below  $\log V$  by more than a margin that exceeds these errors:  $10^{-7}$  in case (i) and  $10^{-6}$  in cases (ii) and (iii) in double precision, and  $10^{-40}$  and  $10^{-20}$  in PARI/GP. Rounding can therefore add sets to the search, but never remove one.  $\square$

The search for (i) grows quickly with  $K$ : it creates 2567 sets for  $K = 10^4$  and 19711 for  $K = 14000$ . The least size of an independent set of primes from 5 on with product at least 3 therefore lies between 16001 and 100470, the length of the set chosen one prime at a time above. For the primes  $p \equiv 2 \pmod{3}$  the products grow much more slowly: the independent set chosen from them in the same way has 2307386 elements below  $2 \cdot 10^8$ , and its product is only  $2.0703 \dots$ , far below  $\frac{8}{3}$ .

**Corollary 2.6.** *Let  $n$  be as in Lemma 2.1, with  $M = (n + \varepsilon)/\varphi(n)$ .*

- (i) *If  $M \geq 3$  and  $3 \nmid n$ , then  $\omega(n) \geq 16001$ .*
- (ii) *If  $3 \mid n$  and  $\varepsilon = -1$ , then  $\omega(n) \geq 10^8 + 2$ .*
- (iii) *If  $3 \mid n$ ,  $\varepsilon = +1$  and  $M \geq 3$ , then  $\omega(n) \geq 10^8 + 2$ .*

*Proof.* Let  $k = \omega(n)$ . By Lemma 2.1 the prime factors of  $n$  form an independent set, and  $n/\varphi(n) = M - \varepsilon/\varphi(n)$ .

(i) The prime factors are at least 5, and  $n/\varphi(n) \geq 3 - 1/\varphi(n) > 2$ . The product of  $p/(p-1)$  over the six smallest primes from 5 on is  $1.9490 \dots$ , so  $k \geq 7$  and  $\varphi(n) \geq 4 \cdot 6 \cdot 10 \cdot 12 \cdot 16 \cdot 18 \cdot 22 > 10^6$ . Hence  $n/\varphi(n) > 2.999999$ , and Proposition 2.5(i) gives  $k > 16000$ .

(ii) By Lemma 2.2,  $M \geq 4$ , and the  $k-1$  prime factors other than 3 are congruent to 2 modulo 3. The product of  $p/(p-1)$  over them is  $\frac{2}{3}n/\varphi(n) = \frac{2}{3}(M+1/\varphi(n)) > \frac{8}{3}$ , so  $k-1 > 10^8$  by Proposition 2.5(ii).

(iii) By Lemma 2.2,  $M \geq 5$ , and the prime factors other than 3 are as in (ii). The product over them is  $\frac{2}{3}(M-1/\varphi(n)) \geq \frac{10}{3} - \frac{2}{3\varphi(n)} \geq 3$ , while over the five smallest primes  $p \equiv 2 \pmod{3}$  from 5 on it is  $1.5818\dots$ . So there are at least six of them,  $\varphi(n) \geq 2 \cdot 4 \cdot 10 \cdot 16 \cdot 22 \cdot 28 \cdot 40 > 10^6$ , and the product exceeds  $\frac{10}{3}(1-10^{-6})$ . Proposition 2.5(iii) gives  $k-1 > 10^8$ .  $\square$

For  $\varepsilon = -1$ , part (ii) improves the bound  $\omega(n) \geq 40000000$  of Burcsi, Czirbusz and Farkas [6], who completed sequences of primes  $p \equiv 2 \pmod{3}$  in a similar way by the least primes allowed. Part (i) replaces the bound 33 that the products over the first primes give, and for  $M = 3$  the bound 1991 of Hagis [17]; part (iii) replaces the bound 1540 that the products give.

*Proof of Theorem 1.5 and of the last two assertions of Theorem 1.1.* Let  $n$  satisfy (1.1) or (1.2), with  $n$  composite in the first case and  $n > 3$  in the second. If  $3 \nmid n$  and  $M \geq 3$ , then  $\omega(n) \geq 16001$  by Corollary 2.6(i). If  $3 \mid n$ , then  $\omega(n) > 10^8$  by Corollary 2.6(ii) for (1.1), where  $M \geq 4$  holds automatically, and by Corollary 2.6(iii) for (1.2) when  $M \geq 3$ .  $\square$

By Corollary 2.6, a solution with at most 16000 prime factors has  $M = 2$ , and if it solves (1.1) it is prime to 3. The searches below concern the case  $n + \varepsilon = 2\varphi(n)$ .

**2.2. The growth of independent sets.** Proposition 2.5 needs a search because the product of  $p/(p-1)$  over an independent set grows very slowly with the size of the set. Over the first  $K$  primes the product is asymptotic to  $e^\gamma \log K$  by Mertens' theorem, so that the products alone give only  $\omega(n) \geq \exp(c'M)$  for some  $c' > 0$ . Over an independent set of  $K$  primes the product is at most a constant multiple of  $\log \log K$ , and this forces  $\omega(n)$  to be at least doubly exponential in  $M$  (Theorem 1.7).

For a set  $S$  of primes and  $x \geq 2$  put

$$P_S(x) = \prod_{p \in S, p \leq x} \frac{p}{p-1}, \quad \delta_S(x) = \prod_{p \in S, p \leq x} \left(1 - \frac{1}{p-1}\right).$$

An independent set with at least two elements does not contain 2, since 2 divides  $q-1$  for every odd prime  $q$ . For a set  $S$  of odd primes every factor of  $\delta_S(x)$  is at least  $\frac{1}{2}$ , and since  $\frac{p}{p-1}(1 - \frac{1}{p-1}) = 1 - \frac{1}{(p-1)^2}$ ,

$$\delta_S(x) P_S(x) \leq 1. \tag{2.2}$$

**Theorem 2.7** (Meijer [22]). *There is an absolute constant  $c_0$  such that every independent set  $S$  of  $K \geq 3$  primes satisfies*

$$\prod_{p \in S} \frac{p}{p-1} \leq c_0 \log \log K.$$

Meijer proved the bound  $P_S(x) \leq c \log \log x$  with the Brun–Selberg sieve, and Yamada [37] gave explicit constants with the large sieve. We include a short proof by the large sieve, with constants left unspecified. Each element  $p$  of  $S$  excludes from  $S$  the primes congruent to 1 modulo  $p$ , so that a prime above  $x$  can belong to  $S$  only if it avoids one residue class modulo every element of  $S$  up to  $x$ ; the large sieve bounds the number of such primes by  $\delta_S(x)$  times the number of all primes in the range.

**Lemma 2.8.** *Let  $S$  be an independent set of odd primes and  $N \geq 2^{88}$ . Then*

$$\#\{q \in S : N^{1/44} < q \leq N\} \leq \frac{132 N \delta_S(N^{1/44})}{\log N}.$$

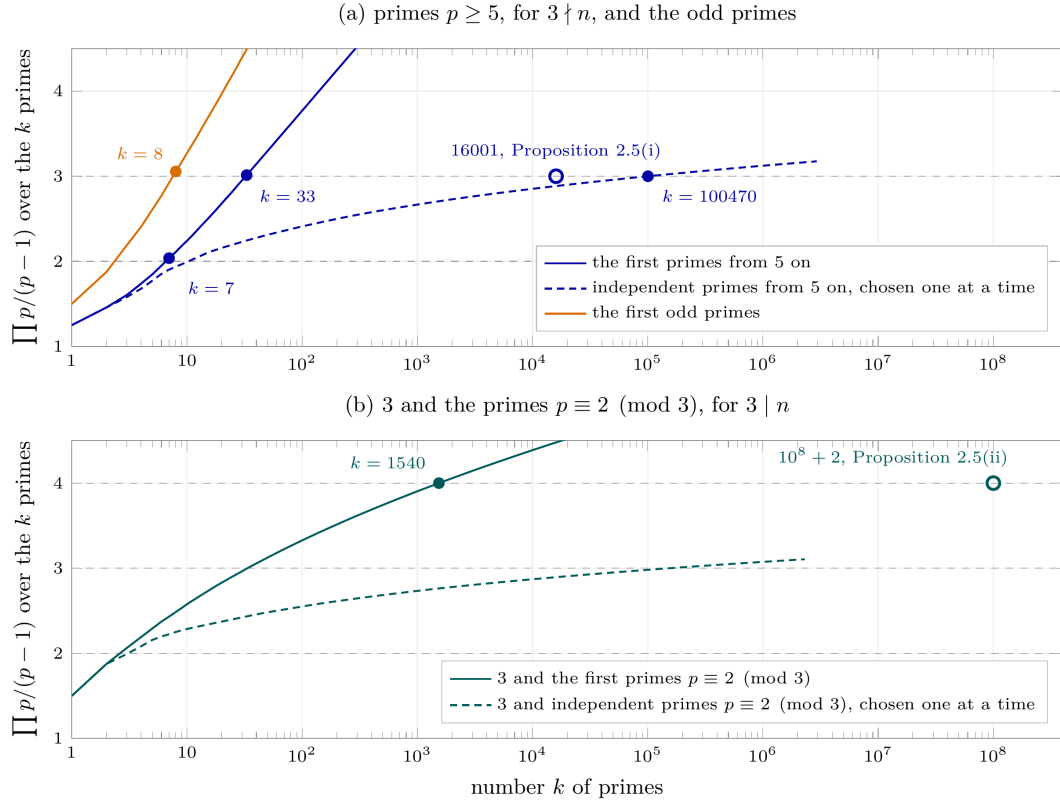


FIGURE 2. The product of  $p/(p-1)$  over the first  $k$  primes of a list (solid) and over the independent set chosen from the same list one prime at a time (dashed), for the primes below  $2 \cdot 10^8$ , with  $k$  on a logarithmic scale. Where a list has both curves, the largest product over an independent set of  $k$  of its primes lies between them. (a) For the primes from 5 on the product passes 2 at  $k = 7$ , so a solution of (3.1) prime to 3 has at least seven prime factors (Section 5), and 3 at  $k = 33$  (Lemma 2.3); for the odd primes it passes 3 at  $k = 8$ , where the argument of Corollary 2.4 stops. The independent set passes 3 only at  $k = 100470$ , and no independent set of primes from 5 on reaches 3 before  $k = 16001$  (open circle). (b) For 3 and the primes  $p \equiv 2 \pmod{3}$  the product passes 4 at  $k = 1540$  (Lemma 2.2), and no independent set formed by 3 and such primes reaches 4 before  $k = 10^8 + 2$  (open circle).

*Proof.* Put  $Q = N^{1/2}$  and  $z = N^{1/44} = Q^{1/22}$ , so that  $z \geq 4$ . For a prime  $p \leq z$  let  $\Omega_p$  consist of the residue class 0 modulo  $p$  and, if  $p \in S$ , also of the class 1; for  $z < p \leq Q$  let  $\Omega_p$  be empty. Let  $\omega(p) < p$  be the number of classes in  $\Omega_p$ . A prime  $q \in S$  with  $z < q \leq N$  lies in none of these classes: it is not divisible by a prime  $p \leq z < q$ , and for  $p \in S$  it is not congruent to 1 modulo  $p$ , because  $S$  is independent. The large sieve [23, 24] states that a set of integers in  $[1, N]$  that avoids the classes of  $\Omega_p$  for every prime  $p \leq Q$  has at most  $(N + Q^2)/L$  elements, where

$$L = \sum_{q \leq Q} \mu^2(q) G(q), \quad G(q) = \prod_{p|q} g(p), \quad g(p) = \frac{\omega(p)}{p - \omega(p)}.$$

Here  $g(p) = 1/(p-1)$  for  $p \leq z$  with  $p \notin S$ ,  $g(p) = 2/(p-2)$  for  $p \leq z$  with  $p \in S$ , and  $g(p) = 0$  for  $p > z$ ; in every case  $g(p) \leq 6/p$ .

Let  $\Pi$  be the product of the primes up to  $z$  and put  $\sigma = 1/\log z$ , so that  $Q^\sigma = e^{22}$ . Then  $L$  is the sum of  $G(q)$  over the divisors  $q \leq Q$  of  $\Pi$ . A divisor  $q > Q$  has  $(q/Q)^\sigma > 1$ , and so

$$\sum_{q|\Pi, q>Q} G(q) \leq Q^{-\sigma} \sum_{q|\Pi} G(q) q^\sigma = e^{-22} \prod_{p \leq z} (1 + g(p)p^\sigma).$$

For  $p \leq z$  we have  $0 < \sigma \log p \leq 1$ , hence  $p^\sigma - 1 \leq (e-1)\sigma \log p$  by convexity, and  $1 + g(p)p^\sigma \leq (1 + g(p)) \exp(g(p)(p^\sigma - 1))$ . With  $g(p) \leq 6/p$  and the elementary bound<sup>6</sup>  $\sum_{p \leq z} (\log p)/p < \log z + \log 4$  this gives

$$\prod_{p \leq z} (1 + g(p)p^\sigma) \leq \exp\left(6(e-1)\left(1 + \frac{\log 4}{\log z}\right)\right) \prod_{p \leq z} (1 + g(p)) < e^{20.7} \prod_{p \leq z} (1 + g(p)),$$

because  $z \geq 4$  and  $12(e-1) < 20.7$ . Since  $\prod_{p \leq z} (1 + g(p)) = \sum_{q|\Pi} G(q)$  and  $e^{20.7-22} < \frac{1}{3}$ , the divisors  $q > Q$  contribute less than a third of this sum, and  $L > \frac{2}{3} \prod_{p \leq z} (1 + g(p))$ . Finally  $1 + g(p) = p/(p-1)$  for  $p \notin S$  and  $1 + g(p) = \frac{p}{p-1} \left(1 - \frac{1}{p-1}\right)^{-1}$  for  $p \in S$ , so that

$$\prod_{p \leq z} (1 + g(p)) = \frac{1}{\delta_S(z)} \prod_{p \leq z} \left(1 - \frac{1}{p}\right)^{-1} > \frac{\log z}{\delta_S(z)},$$

since the product  $\prod_{p \leq z} (1 - 1/p)^{-1}$  is at least  $\sum_{n \leq z} 1/n > \log z$ . With  $\log z = (\log N)/44$  we obtain  $L > \log N/(66 \delta_S(z))$ , and the large sieve, with  $N + Q^2 = 2N$ , gives the lemma.  $\square$

**Lemma 2.9.** *Let  $S$  be an independent set of odd primes and  $y \geq 2^{88}$ . Then*

$$\sum_{q \in S, y < q \leq y^2} \frac{1}{q} \leq 269 \delta_S(y^{1/44}).$$

*Proof.* Let  $a = \lfloor \log_2 y \rfloor \geq 88$  and  $b = \lceil 2 \log_2 y \rceil$ , so that  $b \leq 2a + 2$ . The interval  $(y, y^2]$  is covered by the intervals  $(2^{i-1}, 2^i]$  with  $a < i \leq b$ . For each such  $i$ , Lemma 2.8 with  $N = 2^i$  applies to the elements of  $S$  in  $(2^{i-1}, 2^i]$ , because  $2^{i/44} \leq 2^{i-1}$ , and  $\delta_S(2^{i/44}) \leq \delta_S(y^{1/44})$ , because  $2^i > y$  and  $\delta_S$  is decreasing. Each of these elements exceeds  $2^{i-1}$ , so their reciprocals add up to at most

$$\frac{132 \cdot 2^i \delta_S(y^{1/44})}{2^{i-1} i \log 2} = \frac{264 \delta_S(y^{1/44})}{i \log 2}.$$

Since  $\sum_{a < i \leq b} 1/i \leq \log(b/a) \leq \log(2 + 2/88) < 0.705$  and  $264 \cdot 0.705/\log 2 < 269$ , the lemma follows.  $\square$

*Proof of Theorem 2.7.* By Mertens' theorem  $\prod_{p \leq x} (1 - 1/p)^{-1} = e^\gamma \log x + O(1)$ , so there is an absolute constant  $c_1$  with  $\prod_{u < p \leq u^{44}} p/(p-1) \leq c_1$  for all  $u \geq 2$ . Let  $S$  be an independent set of odd primes. Then  $P_S(y) \leq c_1 P_S(y^{1/44})$  for  $y \geq 2^{44}$ , and Lemma 2.9 and (2.2) give, for  $y \geq 2^{88}$ ,

$$\sum_{q \in S, y < q \leq y^2} \frac{1}{q} \leq \frac{269}{P_S(y^{1/44})} \leq \frac{269 c_1}{P_S(y)}. \quad (2.3)$$

Put  $y_0 = 2^{88}$ ,  $y_{j+1} = y_j^2$  and  $a_j = P_S(y_j)$ . Since  $-\log(1 - 1/q) \leq 2/q$ , (2.3) gives  $a_{j+1} \leq a_j e^{c_2/a_j}$  with  $c_2 = 538 c_1$ , and also  $a_{j+1} \leq c_1 a_j$ . If  $a_j \geq c_2$ , then  $a_{j+1} \leq a_j(1 + 2c_2/a_j) = a_j + 2c_2$ , because  $e^t \leq 1 + 2t$  for  $0 \leq t \leq 1$ ; if  $a_j < c_2$ , then  $a_{j+1} < c_1 c_2$ . By induction  $a_j \leq c_3 + 2c_2 j$  for

<sup>6</sup>Let  $m = \lfloor z \rfloor$ . Legendre's formula gives  $\log m! \geq \sum_{p \leq m} \lfloor m/p \rfloor \log p > m \sum_{p \leq m} (\log p)/p - \sum_{p \leq m} \log p$ , while  $\log m! \leq m \log m$  and  $\sum_{p \leq m} \log p < m \log 4$ , because the product of the primes up to  $m$  is less than  $4^m$ . Dividing by  $m$  gives the bound.

all  $j \geq 0$ , where  $c_3 = \max\{c_1 c_2, \prod_{p \leq 2^{88}} p/(p-1)\} \geq a_0$ . For  $x > y_0$  let  $j$  be the least index with  $y_j \geq x$ . Then  $2^{j-1} \log y_0 < \log x$ , and  $P_S(x) \leq a_j \leq c_3 + 2c_2(1 + \log_2(\log x / \log y_0))$ . Together with  $P_S(x) \leq a_0 \leq c_3$  for  $x \leq y_0$ , this gives  $P_S(x) \leq c_4 \log \log x$  for all  $x \geq 3$ , with an absolute constant  $c_4$ .

Now let  $S$  have  $K \geq 3$  elements, so that its elements are odd. At most  $K$  of them exceed  $K^2$ , and each of those contributes a factor at most  $1 + 1/K^2$ , so that together they contribute less than  $e^{1/K} < 2$ . Hence

$$\prod_{p \in S} \frac{p}{p-1} < 2P_S(K^2) \leq 2c_4 \log \log K^2 = 2c_4(\log 2 + \log \log K) \leq 20 c_4 \log \log K,$$

since  $\log 2 < 9 \log \log 3$ .  $\square$

*Proof of Theorem 1.7.* Let  $n$  be as in Lemma 2.1 and  $k = \omega(n) \geq 2$ . The prime factors of  $n$  form an independent set, and  $n/\varphi(n)$  is the product of  $p/(p-1)$  over them. If  $k \geq 3$ , Theorem 2.7 and Lemma 2.1(iii) give  $M \leq c_0 \log \log k + 1/\varphi(n) < (c_0 + 3) \log \log(k+2)$ , since  $\log \log 5 > \frac{1}{3}$ . If  $k = 2$ , then  $n/\varphi(n) \leq \frac{3}{2} \cdot \frac{5}{4}$  and  $\varphi(n) \geq 8$ , so that  $M \leq 2 < 7 \log \log 4$ . In both cases  $\log \log(k+2) \geq cM$  with  $c = 1/\max\{c_0 + 3, 7\}$ , which is the assertion.  $\square$

The constants in Theorems 2.7 and 1.7 are effective but large, and the numerical bounds of Corollary 2.6 come from Proposition 2.5, not from the sieve. What the sieve explains is the shape of those bounds. Heuristically, the elements of  $S$  up to  $x$  leave a proportion  $\delta_S(x)$  of the primes near  $x$  available, and a set that takes every available prime gains  $\log P_S$  at the rate  $\delta_S(x) d \log \log x$ , since  $\sum_{q \leq x} 1/q$  grows like  $\log \log x$ . The derivative of  $P_S$  with respect to  $\log \log x$  is then  $\delta_S(x) P_S(x)$ , which is at most 1 by (2.2), so that  $P_S(x)$  grows at most like  $\log \log x$ , the rate of Theorem 2.7. The independent sets chosen one prime at a time in Section 2.1 grow at this rate (Table 1). Between  $x = 10^3$  and  $x = 10^8$  the product of  $p/(p-1)$  over the elements up to  $x$  differs from a linear function of  $\log \log x$  by less than 0.003, with slope 0.82 for the primes from 5 on and 0.41 for the primes  $p \equiv 2 \pmod{3}$ , which are half of all primes. The first set needs 100470 primes to reach the product 3, where the first primes from 5 on need 33, and the second has not reached  $\frac{8}{3}$  below  $2 \cdot 10^8$  (Section 2.1). This is why the bounds of Corollary 2.6 exceed 33 and 1540 by so much.

TABLE 1. The independent sets chosen one prime at a time from the primes from 5 on and from the primes  $p \equiv 2 \pmod{3}$  from 5 on: the number of elements up to  $x$  and the product of  $p/(p-1)$  over them.

| $x$    | $\log \log x$ | primes from 5 on |         | $p \equiv 2 \pmod{3}$ |         |
|--------|---------------|------------------|---------|-----------------------|---------|
|        |               | elements         | product | elements              | product |
| $10^3$ | 1.933         | 62               | 2.3430  | 45                    | 1.6489  |
| $10^4$ | 2.220         | 407              | 2.5792  | 293                   | 1.7626  |
| $10^5$ | 2.443         | 2 908            | 2.7593  | 2 185                 | 1.8564  |
| $10^6$ | 2.626         | 22 602           | 2.9078  | 17 396                | 1.9328  |
| $10^7$ | 2.780         | 183 730          | 3.0344  | 142 723               | 1.9979  |
| $10^8$ | 2.913         | 1 538 281        | 3.1447  | 1 207 989             | 2.0546  |

### 3. INTERVALS FOR THE PRIME FACTORS

Fix  $k \geq 2$ , a sign  $\varepsilon$ , and a squarefree odd  $n = p_1 \cdots p_k$  with

$$n + \varepsilon = 2\varphi(n). \tag{3.1}$$

By Lemma 2.1(iii),

$$\prod_{i=1}^k \frac{p_i}{p_i - 1} = 2 - \frac{\varepsilon}{\varphi(n)}. \quad (3.2)$$

For  $\varepsilon = -1$  the product lies just above 2, and for  $\varepsilon = +1$  just below it. Everything in this section is a consequence of (3.2), and the two signs are treated together.

**Proposition 3.1.**  $P_j < 2$  for every  $j < k$ .

*Proof.* For  $\varepsilon = +1$  this is clear, since  $P_j \leq P_k = 2 - 1/\varphi(n)$ . For  $\varepsilon = -1$ , suppose  $P_j \geq 2$  with  $j < k$ . Since  $p_{j+1} - 1 < n - 1$  we have  $p_{j+1}/(p_{j+1} - 1) > 1 + 1/(n - 1)$ , and the remaining factors exceed 1, so  $P_k > 2(1 + 1/(n - 1)) = 2 + 1/\varphi(n)$  by (3.1), contradicting (3.2).  $\square$

At depth  $j$  put  $T = 2/P_j$ , which exceeds 1 by the proposition, and

$$C_j = 2B_j - A_j, \quad T - 1 = \frac{C_j}{A_j}.$$

Thus  $C_j$  is a positive integer, and  $C_j/A_j$  measures how far the primes chosen so far fall short of the target 2.

**Proposition 3.2.** Let  $0 \leq j \leq k - 2$ ,  $m = k - j$ ,  $W = (p_j + 1)^m$  and  $x = p_{j+1}/(p_{j+1} - 1)$ . Then

$$\begin{aligned} x^m > T \quad \text{and} \quad x < T + \frac{1}{A_j W} \quad \text{if } \varepsilon = -1, \\ x^m \geq T - \frac{1}{A_j W} \quad \text{and} \quad x < T \quad \text{if } \varepsilon = +1. \end{aligned} \quad (3.3)$$

In particular  $p_{j+1}$  lies in an interval  $[L_j, U_j]$  whose endpoints are determined by  $p_1, \dots, p_j$ ,  $k$  and  $\varepsilon$ .

*Proof.* Let  $\rho = \prod_{i>j} p_i/(p_i - 1)$ . By (3.2),

$$\rho = \frac{2 - \varepsilon/\varphi(n)}{P_j} = T - \frac{\varepsilon}{P_j \varphi(n)}.$$

There are  $m \geq 2$  factors in  $\rho$ ; each exceeds 1 and none exceeds  $x$ , because  $p/(p - 1)$  decreases in  $p$ . Hence  $x < \rho \leq x^m$ . The remaining primes are odd and larger than  $p_j$  (larger than 2 when  $j = 0$ ), so  $p_i - 1 \geq p_j + 1$  for  $i > j$ , and  $\varphi(n) = B_j \prod_{i>j} (p_i - 1) \geq B_j W$ . Therefore

$$0 < \frac{1}{P_j \varphi(n)} \leq \frac{1}{P_j B_j W} = \frac{1}{A_j W},$$

and the four inequalities follow from  $x < \rho \leq x^m$ . Finally, for  $\varepsilon = +1$  the quantity  $T - 1/(A_j W)$  exceeds 1, since  $T - 1 = C_j/A_j \geq 1/A_j$ .  $\square$

Apart from the correction terms, (3.3) says  $x < T \leq x^m$ , that is  $1/(T - 1) < p_{j+1} - 1 \leq 1/(T^{1/m} - 1)$ . Since  $T^{1/m} - 1$  is close to  $(T - 1)/m$ , the interval has length about  $(m - 1)/(T - 1) = (m - 1)A_j/C_j$  (Figure 3). The upper end obeys a bound that makes the search finite.

**Corollary 3.3.** Let  $0 \leq j \leq k - 2$ , let  $p_1 < \dots < p_j$  be odd primes with  $P_j < 2$ , and let  $p_{j+1} > p_j$  be a prime that satisfies the inequality for  $x^m$  in (3.3). Then  $p_{j+1} < 1 + 3(k - j)B_j/C_j < 4kA_j$ . If  $n = p_1 \cdots p_k$  satisfies (3.1), then moreover  $p_k \leq 2B_{k-1} + 1$ , and

$$A_j < (4k)^{2^j} \quad (0 \leq j \leq k), \quad n < (4k)^{2^k}.$$

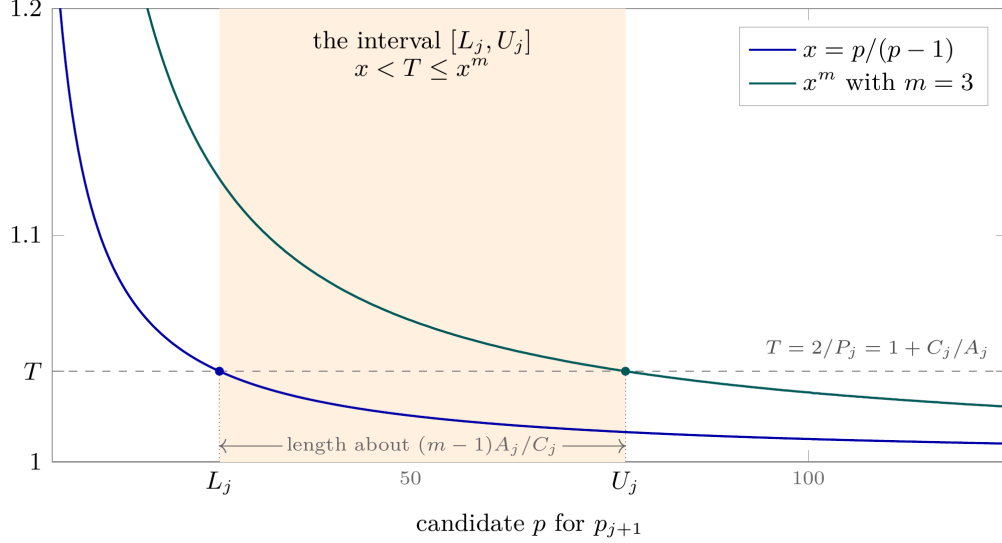


FIGURE 3. The interval of Proposition 3.2 for  $T = 1.04$  and  $m = 3$ , with the correction terms  $1/(A_j W)$  left out. The candidate  $p$  must satisfy  $x < T$ , because the other  $m - 1$  factors of  $\rho$  exceed 1, and  $x^m \geq T$ , because none of them exceeds  $x$ . Here the two conditions give  $26 < p < 77$ .

*Proof.* Let  $m = k - j$  and  $x = p_{j+1}/(p_{j+1} - 1)$ , so that  $p_{j+1} - 1 = 1/(x - 1)$ . By (3.3),  $x^m \geq T'$ , where  $T' = T$  if  $\varepsilon = -1$  and  $T' = T - 1/(A_j W)$  if  $\varepsilon = +1$ . As  $T - 1 = C_j/A_j$  with  $C_j \geq 1$ , and  $W \geq 4$ , in both cases  $T' - 1 \geq (C_j - \frac{1}{4})/A_j \geq 3C_j/(4A_j) > 0$  and  $T' \leq T$ . For  $t > 1$  we have  $t^{1/m} - 1 = (t - 1)/\sum_{i < m} t^{i/m} \geq (t - 1)/(mt)$ , and so

$$p_{j+1} - 1 = \frac{1}{x - 1} \leq \frac{1}{T'^{1/m} - 1} \leq \frac{mT'}{T' - 1} \leq \frac{2mB_j}{A_j} \cdot \frac{4A_j}{3C_j} < \frac{3mB_j}{C_j}.$$

Since  $C_j \geq 1$  and  $B_j \leq A_j - 1$  for  $j \geq 1$ , while  $A_0 = B_0 = C_0 = 1$ , this gives  $p_{j+1} < 4kA_j$ . For a solution, the equation  $A_{k-1}p_k + \varepsilon = 2B_{k-1}(p_k - 1)$  gives  $p_k C_{k-1} = 2B_{k-1} + \varepsilon$ , and  $C_{k-1} \geq 1$  by Proposition 3.1, so that  $p_k \leq 2B_{k-1} + 1 < 4kA_{k-1}$ . By Propositions 3.1 and 3.2 the first part applies to  $p_{j+1}$  for every  $j \leq k - 2$ . Hence  $A_{j+1} < 4kA_j^2$  for  $0 \leq j < k$ , and induction gives  $A_j \leq (4k)^{2^j - 1}$ .  $\square$

Every node of the search below satisfies the hypotheses of Corollary 3.3, so the tree has finitely many nodes and the search terminates. The bound  $n < (4k)^{2^k}$  has the shape of the bound  $n < k^{2^k}$  of Pomerance [31]; Theorem 1.2 replaces it by  $n < 2^{2^{k-7}}$ .

The inequalities (3.3) are tested in integers: for instance  $x^m > T$  reads  $p^m A_j > 2B_j(p - 1)^m$ . A floating-point estimate is used only to find a starting point, which is then corrected by exact comparisons.

**Lemma 3.4** (congruence prune). *No prime factor of  $n$  divides  $p - 1$  for another prime factor  $p$  of  $n$ .*

This is Lemma 2.1(ii). A candidate  $p$  for  $p_{j+1}$  is discarded if  $p \mid p_i - 1$  or  $p_i \mid p - 1$  for some  $i \leq j$ . The primes of the interval  $[L_j, U_j]$  that are not discarded are called *admissible*.

The last two primes are determined together by a divisibility condition.

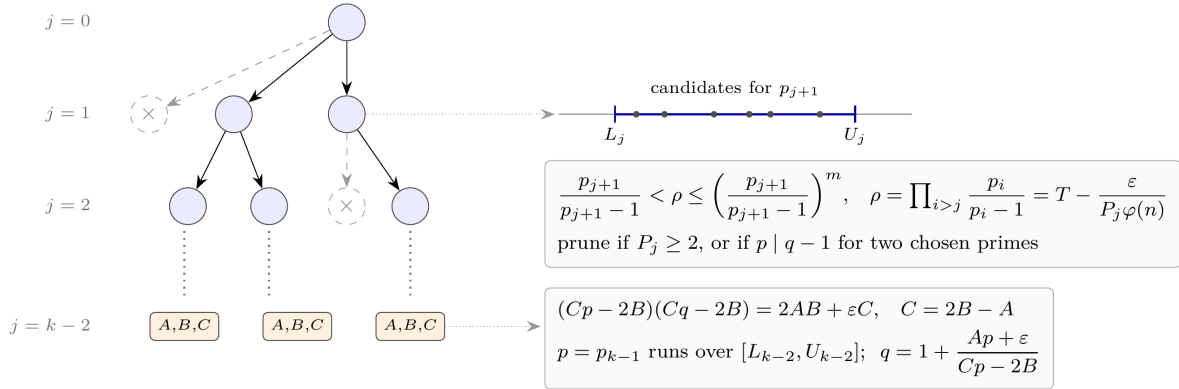


FIGURE 4. The search for  $n = p_1 \cdots p_k$  with  $n + \varepsilon = 2\varphi(n)$ . A node at depth  $j$  is an increasing sequence of primes  $p_1 < \cdots < p_j$ . Its children are the primes in the interval  $[L_j, U_j]$  of Proposition 3.2 that pass the congruence prune; dashed branches are cut. At depth  $k-2$  the pair  $(p_{k-1}, p_k)$  is a lattice point on the hyperbola (3.5), found by running  $p$  over  $[L_{k-2}, U_{k-2}]$  or by factoring  $2AB + \varepsilon C$ .

**Proposition 3.5.** *Let  $A = A_{k-2}$ ,  $B = B_{k-2}$ ,  $C = 2B - A$ ,  $p = p_{k-1}$  and  $q = p_k$ . Then*

$$t := Cp - 2B > 0, \quad t \mid Ap + \varepsilon, \quad q = 1 + \frac{Ap + \varepsilon}{t}, \quad (3.4)$$

and equivalently

$$(Cp - 2B)(Cq - 2B) = 2AB + \varepsilon C. \quad (3.5)$$

*Proof.* Equation (3.1) reads  $Apq + \varepsilon = 2B(p-1)(q-1)$ . Collecting the terms in  $q$  gives  $q(Cp - 2B) = 2B(p-1) + \varepsilon$ . The right-hand side is positive, so  $t > 0$ , and  $q-1 = (2Bp - 2B + \varepsilon - Cp + 2B)/t = (Ap + \varepsilon)/t$ . Multiplying  $Cpq - 2B(p+q) + 2B - \varepsilon = 0$  by  $C$  and completing the product gives  $(Cp - 2B)(Cq - 2B) = 4B^2 - 2BC + \varepsilon C = 2AB + \varepsilon C$ .  $\square$

Once  $p_1, \dots, p_{k-2}$  are fixed, the prime  $p_{k-1}$  runs over the interval of Proposition 3.2 with  $m = 2$ , and for each candidate  $p$  the last prime is determined by (3.4). Alternatively, by (3.5),  $t$  is a divisor of the integer

$$D = 2AB + \varepsilon C$$

lying in a known residue class modulo  $C$ , so the candidates can be read off from the divisors of  $D$  when  $D$  can be factored. The first route costs time proportional to the length of the interval, the second the cost of factoring  $D$ . Figure 4 summarises the procedure.

The search is a depth-first traversal of the tree whose nodes at depth  $j \leq k-2$  are the increasing sequences  $(p_1, \dots, p_j)$  of primes with  $p_1 \geq p_{\min}$ , where each  $p_{i+1}$  lies in the interval determined by  $(p_1, \dots, p_i)$  and passes the congruence prune, and where  $P_i < 2$  for all  $i \leq j$ . At depth  $k-2$  each candidate  $p_{k-1}$  is tested with (3.4): one checks that  $t > 0$ , that  $t \mid Ap + \varepsilon$ , that  $q = 1 + (Ap + \varepsilon)/t$  is a prime larger than  $p$ , that  $p$  and  $q$  pass the congruence prune, and finally that (3.1) holds exactly.

**Theorem 3.6** (completeness). *If  $n = p_1 \cdots p_k$  with primes  $p_{\min} \leq p_1 < \cdots < p_k$  satisfies (3.1), then the traversal visits  $(p_1, \dots, p_{k-2})$ , examines the candidate  $p_{k-1}$ , and records  $n$ .*

*Proof.* By induction on  $j$  the node  $(p_1, \dots, p_j)$  is visited for every  $j \leq k-2$ : the prime  $p_{j+1}$  lies in the interval of Proposition 3.2, it passes the prune by Lemma 3.4, and  $P_{j+1} < 2$  by Proposition 3.1. At depth  $k-2$  the candidate  $p = p_{k-1}$  satisfies (3.4) with  $q = p_k$ . A primality

test never declares a prime to be composite, so  $q$  is accepted, and the final check of (3.1) succeeds.  $\square$

The search is rigorous for two reasons. First, all arithmetic is exact:  $A_j, B_j, C_j$ , the interval endpoints and the quantities in (3.4) are integers or ratios of integers compared by cross multiplication. Second, the probabilistic primality tests used by the programs<sup>7</sup> can err only by declaring a composite number prime, never by declaring a prime composite. Such an error could add a node or accept a composite  $q$ , but it can never remove an actual solution. A search that finds nothing is therefore a proof that there is nothing, and the completeness of the enumeration route depends on no primality proof. The exact check of (3.1) would not catch such an error, since it uses  $\prod(p_i - 1)$  in place of  $\varphi(n)$ . This matters only for solutions that are reported, and every solution reported in this paper, including those used in the checks of Section 8, has all its prime factors below  $10^{14}$ , where the tests are proofs. The divisor route depends on a complete factorisation of  $D$ , which is certified whenever  $D < \psi_{13}$ , since then every prime factor of  $D$  lies below that bound.

#### 4. THE LAST THREE PRIMES

For  $k \leq 14$  the search of Section 3 is cheap because the interval for  $p_{k-1}$  is never longer than about  $5 \cdot 10^7$ . For  $k = 15$  this is no longer true. After the prime  $p_{13}$  has been chosen the interval for  $p_{14}$  can be longer than  $10^{14}$ , and there are tens of millions of choices of  $p_{13}$ . Neither route of Section 3 copes with this: running  $p_{14}$  over its interval is far too slow, and factoring tens of millions of integers of about fifty digits is a computation of a different order. In this section we describe two further ways of finding the divisors of  $2AB + \varepsilon C$  that can occur in Proposition 3.5. Both use the residue class in which these divisors lie: the first lists lattice points near a hyperbola, and the second scans the possible values of the sum of two complementary divisors.

**4.1. Reduction to divisors in a residue class.** Fix a node  $(p_1, \dots, p_{k-3})$  at depth  $k-3$  and a candidate  $s = p_{k-2}$  in its interval that passes the congruence prune. Put

$$A' = A_{k-3}s, \quad B' = B_{k-3}(s-1), \quad c = 2B' - A', \quad N = 2A'B' + \varepsilon c.$$

If  $c \leq 0$  then  $P_{k-2} = A'/B' \geq 2$ , and by Proposition 3.1 no solution has  $p_1, \dots, p_{k-3}, s$  as its first  $k-2$  primes; such  $s$  are discarded. So let  $c \geq 1$ . By Proposition 3.5, applied with  $A = A'$  and  $B = B'$ , the remaining primes  $p < q$  of a solution satisfy  $(cp - 2B')(cq - 2B') = N$  with both factors positive. Let  $r$  be the least positive residue of  $-2B'$  modulo  $c$ .

By (3.2) the last three primes satisfy

$$\frac{s}{s-1} \cdot \frac{p}{p-1} \cdot \frac{q}{q-1} = T - \frac{\varepsilon}{P_{k-3}\varphi(n)}, \quad T = \frac{2}{P_{k-3}},$$

so the point  $(s, p, q)$  lies next to the surface on which this product equals  $T$ . Fixing  $s$  cuts the surface in a curve, and the pair  $(p, q)$  lies exactly on the hyperbola  $(cp - 2B')(cq - 2B') = N$  beside that curve (Figure 5).

**Proposition 4.1.** *The integer  $t = cp - 2B'$  is a divisor of  $N$  with*

$$t \equiv r \pmod{c}, \quad 1 \leq t \leq \sqrt{N}.$$

<sup>7</sup>Below the number  $\psi_{13} = 3317044064679887385961981$  the programs use a strong probable-prime test to the thirteen bases  $2, 3, \dots, 41$ , and above it a Baillie-PSW test. Sorenson and Webster [34] showed that every composite number below  $\psi_{13}$  fails the first test for at least one of these bases, so there it is a proof. The Baillie-PSW test [2, 32] combines a strong probable-prime test to base 2 with a strong Lucas test; no composite number is known to pass it.

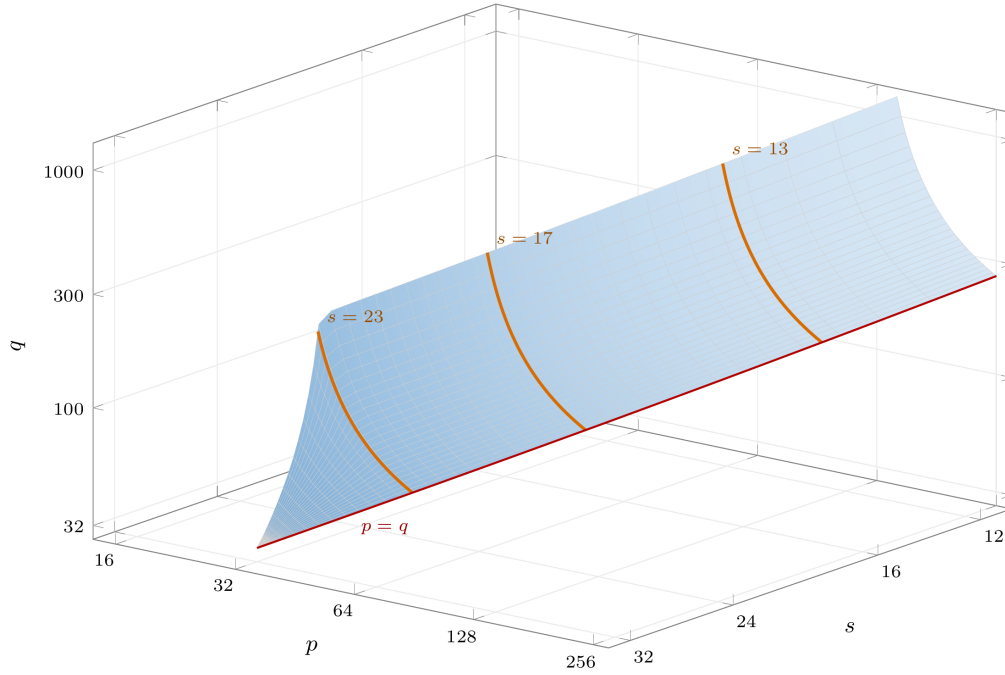


FIGURE 5. The surface  $\frac{s}{s-1} \cdot \frac{p}{p-1} \cdot \frac{q}{q-1} = T$  with  $s < p < q$ , for  $T = 1.1$  on logarithmic axes. Each slice with fixed  $s$ , three of which are marked, falls steeply from large values of  $q$  to the edge  $p = q$ . The slices shrink as  $s$  grows, and the surface ends at the point  $s = p = q$  where  $(s/(s-1))^3 = T$ .

Moreover  $\gcd(r, c) = 1$  and  $N \equiv r^2 \pmod{c}$ . Conversely, every divisor  $t$  of  $N$  with these properties gives integers  $p = (t + 2B')/c$  and  $q = (N/t + 2B')/c$  satisfying  $A'pq + \varepsilon = 2B'(p-1)(q-1)$ .

*Proof.* The congruence holds because  $t \equiv -2B' \pmod{c}$ , and  $t \geq 1$  because both factors are positive, and  $t \leq \sqrt{N}$  because  $p < q$  makes  $t$  the smaller of the two factors. Since  $c = 2B' - A'$  we have  $\gcd(2B', c) = \gcd(2B', A')$ , which is 1 because  $A'$  is odd and, by the congruence prune, no prime factor of  $A'$  divides  $B'$ ; hence  $\gcd(r, c) = 1$ . Further  $N \equiv 2A'B' \equiv (2B')^2 \equiv r^2 \pmod{c}$ , using  $A' \equiv 2B'$ . Conversely, let  $t$  be such a divisor. Then  $N/t \equiv r^2 r^{-1} = r \pmod{c}$ , so both  $t$  and  $N/t$  are congruent to  $-2B'$  and  $p$  and  $q$  are integers. Expanding  $(cp - 2B')(cq - 2B') = N$  and using  $4B'^2 - 2A'B' = 2B'c$  gives  $c(cpq - 2B'(p+q) + 2B' - \varepsilon) = 0$ , which, as  $c \geq 1$  and  $c = 2B' - A'$ , is the equation  $A'pq + \varepsilon = 2B'(p-1)(q-1)$ .  $\square$

So the two-prime problem at depth  $k-2$  becomes the following question: given  $N$ , a modulus  $c$  and a class  $r$  prime to  $c$ , list the divisors of  $N$  up to  $\sqrt{N}$  that are congruent to  $r$  modulo  $c$ . Lenstra [20] proved that there are at most 11 such divisors when  $c > N^{1/3}$ ,<sup>8</sup> and gave a polynomial-time algorithm to find them; Coppersmith, Howgrave-Graham and Nagaraj [11] extended the algorithmic side down to  $c > N^{1/4+\delta}$ . We use a more elementary method suited to our range, and in Section 4.4 a still simpler one based on the fact that  $t$  and  $N/t$  lie in the same class.

<sup>8</sup>In the form used here: if  $0 \leq r < c < N$ ,  $\gcd(r, c) = 1$  and  $c > N^{1/3}$ , then at most 11 positive divisors of  $N$  are congruent to  $r$  modulo  $c$ . The bound counts all divisors in the class, not only those up to  $\sqrt{N}$ .

**4.2. Lattice points near a hyperbola.** Let  $r'$  be the residue of  $Nr^{-1}$  modulo  $c$ , so that  $r' = r$  in the situation of Proposition 4.1, and put

$$w = \frac{N - rr'}{c}, \quad \alpha \equiv wr^{-1}, \quad \beta \equiv r'r^{-1} \pmod{c}.$$

A divisor  $t \equiv r$  is written  $t = r + cu$  with  $u \geq 0$ , and then  $N/t = r' + cv$  for an integer  $v$ . Expanding  $(r + cu)(r' + cv) = N$  gives

$$cuv + r'u + rv = w, \quad (4.1)$$

so the pairs  $(u, v)$  are the lattice points on a hyperbola. Reducing (4.1) modulo  $c$  shows that each of them lies in the coset

$$v \equiv \alpha - \beta u \pmod{c}. \quad (4.2)$$

The coset has one point in every  $c$  consecutive values of  $v$  above a given  $u$ , so by itself it excludes little; combined with the shape of the hyperbola it leaves only a few points to test.

**Lemma 4.2.** *Let  $0 \leq u_0 \leq u_1$  and put*

$$v^+ = \left\lfloor \frac{\lfloor N/(r + cu_0) \rfloor - r'}{c} \right\rfloor, \quad v^- = \left\lceil \frac{\lceil N/(r + cu_1) \rceil - r'}{c} \right\rceil.$$

*If  $t = r + cu$  divides  $N$  and  $u_0 \leq u \leq u_1$ , then  $N/t = r' + cv$  with  $v^- \leq v \leq v^+$  and  $v$  satisfying (4.2). Consequently, with  $H = v^+ - v^- + 1$ , the integer  $u$  lies in the set*

$$\{u \in [u_0, u_1] : (\alpha - \beta u - v^-) \bmod c < H\}, \quad (4.3)$$

*and there is no such  $t$  at all if  $H \leq 0$ .*

*Proof.* Since  $t \equiv r \pmod{c}$  and  $\gcd(r, c) = 1$ , we have  $N/t \equiv Nr^{-1} \equiv r' \pmod{c}$ , so  $v = (N/t - r')/c$  is an integer. The cofactor  $N/t$  is an integer between  $N/(r + cu_1)$  and  $N/(r + cu_0)$ , hence between the ceiling of the first and the floor of the second, and this gives  $v^- \leq v \leq v^+$ ; in particular  $H \geq 1$ . The congruence (4.2) was shown above. Put  $x = (\alpha - \beta u - v^-) \bmod c$ , so that  $0 \leq x < c$  and  $v - v^- \equiv x \pmod{c}$ . As  $v - v^- \geq 0$ , it is at least the least non-negative member  $x$  of its class, and so  $x \leq v - v^- \leq H - 1$ .  $\square$

If  $H \geq c$  the set (4.3) is the whole of  $[u_0, u_1]$ , and every  $u$  in the box is tested. If  $H < c$  it consists of the  $u$  for which an arithmetic progression modulo  $c$  falls into a window of length  $H$ , and its elements can be listed in increasing order without looking at the others, at a cost of  $O(\log c)$  arithmetic operations each, by the following elementary recursion.

**Lemma 4.3.** *Let  $0 \leq a < g$  and  $0 < L \leq R < g$ . The least  $y \geq 0$  with  $L \leq ay \bmod g \leq R$  is found as follows. If  $a = 0$  there is no such  $y$ . If  $a \geq 1$  and  $a \lceil L/a \rceil \leq R$ , it is  $\lceil L/a \rceil$ . Otherwise it is  $\lceil (L + gz)/a \rceil$ , where  $z$  is the least integer  $\geq 0$  with*

$$(-R) \bmod a \leq (g \bmod a) z \bmod a \leq (-L) \bmod a,$$

*and if no such  $z$  exists there is no  $y$ . The last problem is of the same kind, with  $(g, a, L, R)$  replaced by  $(a, g \bmod a, (-R) \bmod a, (-L) \bmod a)$ .*

*Proof.* If  $a = 0$  then  $ay \bmod g = 0 < L$  for every  $y$ . In the second case  $ay \leq R < g$  for  $y = \lceil L/a \rceil$ , so  $ay \bmod g = ay \in [L, R]$ , while  $0 \leq ay < L$  for smaller  $y \geq 0$ . Otherwise no multiple of  $a$  lies in  $[L, R]$ . Since any  $a$  consecutive integers contain a multiple of  $a$ , this gives  $R - L \leq a - 2$ ; moreover  $a \nmid R$ , and no multiple of  $a$  lies in  $[-R, -L]$ , so the residues  $(-R) \bmod a, \dots, (-L) \bmod a$  of  $-R, \dots, -L$  increase without wrapping around and  $0 < (-R) \bmod a \leq (-L) \bmod a < a$ . A solution  $y$  satisfies  $ay = gz + x$  with  $z = \lfloor ay/g \rfloor$  and  $x = ay \bmod g \in [L, R]$ , and  $z \geq 1$  because  $z = 0$  would make  $x = ay$  a multiple of  $a$  in  $[L, R]$ . Conversely, for a given  $z \geq 0$  a solution with this  $z$  exists exactly when  $[L + gz, R + gz]$  contains a multiple of  $a$ ; that multiple is unique,

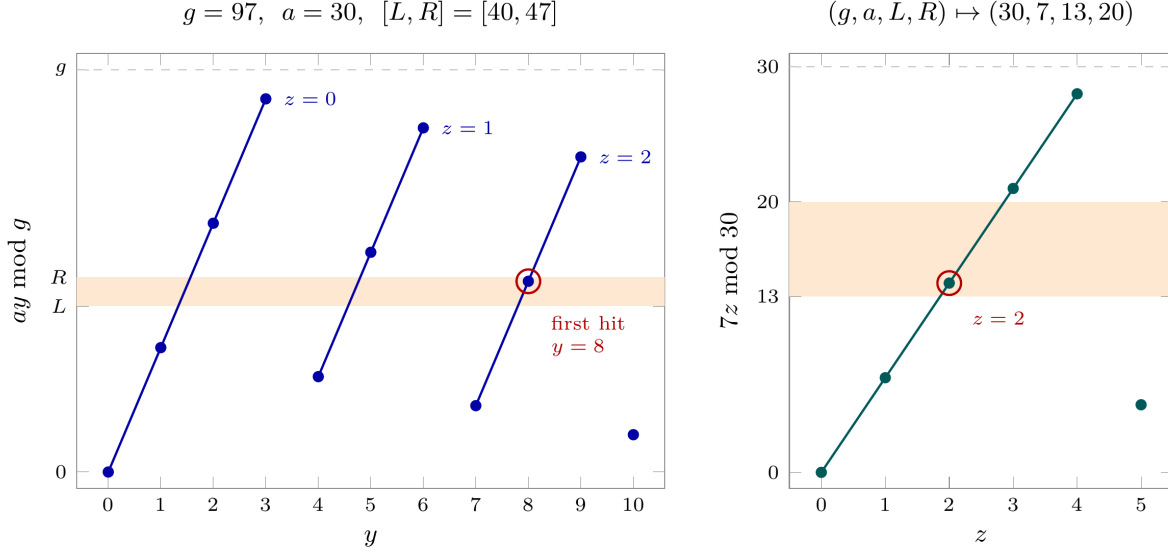


FIGURE 6. One step of Lemma 4.3 with  $g = 97$ ,  $a = 30$  and  $[L, R] = [40, 47]$ . Left: the residues  $ay \bmod g$  rise in runs, one for each value of  $z = \lfloor ay/g \rfloor$ , and the first run misses the window because  $a\lceil L/a \rceil = 60 > R$ . Right: the reduced problem  $(30, 7, 13, 20)$  asks for the least  $z$  with  $7z \bmod 30$  in  $[13, 20]$ . It is  $z = 2$ , and the lemma returns  $y = \lceil (40 + 2 \cdot 97)/30 \rceil = 8$ , for which  $ay \bmod g = 46$ .

because  $R - L < a$ , and the solution is  $y = \lceil (L + gz)/a \rceil$ . The interval contains a multiple of  $a$  exactly when  $gz \equiv -x \pmod{a}$  for some  $x \in [L, R]$ , that is, when  $gz \bmod a = (g \bmod a)z \bmod a$  lies in the stated range. The intervals  $[L + gz, R + gz]$  are disjoint and increase with  $z$ , since  $R - L < g$ , so the least  $z$  gives the least  $y$ .  $\square$

The recursion replaces  $(g, a)$  by  $(a, g \bmod a)$ , as in Euclid's algorithm, so it stops after  $O(\log g)$  steps; Figure 6 follows one step. It is applied with  $g = c$  and  $a = -\beta \bmod c$  as follows. Starting from  $u = u_0$ , let  $b = (\alpha - \beta u - v^-) \bmod c$ . If  $b < H$ , then  $u$  is in (4.3). Otherwise the next element of (4.3), if it does not exceed  $u_1$ , is  $u + y$ , where  $y$  is the least integer  $\geq 0$  with  $c - b \leq ay \bmod c \leq c - b + H - 1$ ; this window lies inside  $[1, c - 1]$ , so Lemma 4.3 applies. Every  $u$  found in this way is tested by one exact division of  $N$  by  $r + cu$ , and the search resumes at  $u + 1$ .

**4.3. Choosing the boxes.** The range  $0 \leq u \leq U = \lfloor (\lfloor \sqrt{N} \rfloor - r)/c \rfloor$  is cut into consecutive intervals  $[u_0, u_1]$ , each giving a box  $[u_0, u_1] \times [v^-, v^+]$  in Lemma 4.2 (Figure 7). Correctness does not depend on how this is done; the choice only affects the running time. Since  $v \approx K/u$  with  $K = N/c^2$ , a box with  $u_1 - u_0 = \lambda u_0$  has area close to  $K\lambda^2/(1 + \lambda)$ , and as the coset has one point per area  $c$  we choose  $\lambda$  with  $\lambda^2/(1 + \lambda) = c^3/N$ , so that each box holds about one point of (4.2). There are two regimes. If  $c^3 > N$ , then  $\lambda$  exceeds 1.6 and grows like  $c^3/N$ , so that  $O(\log N / \log(2 + c^3/N))$  boxes cover the whole range: this is Lenstra's range, and the work is a handful of Euclidean recursions. If  $c^3 < N$ , the boxes shrink to width about  $u_0(c^3/N)^{1/2}$ , the values of  $u$  below  $(N/c^3)^{1/2}$  are tested one at a time, and the number of boxes grows like  $(N/c^3)^{1/2} \log N$ . When this estimate exceeds a fixed threshold we factor  $N$  completely instead and read off its divisors in the class.

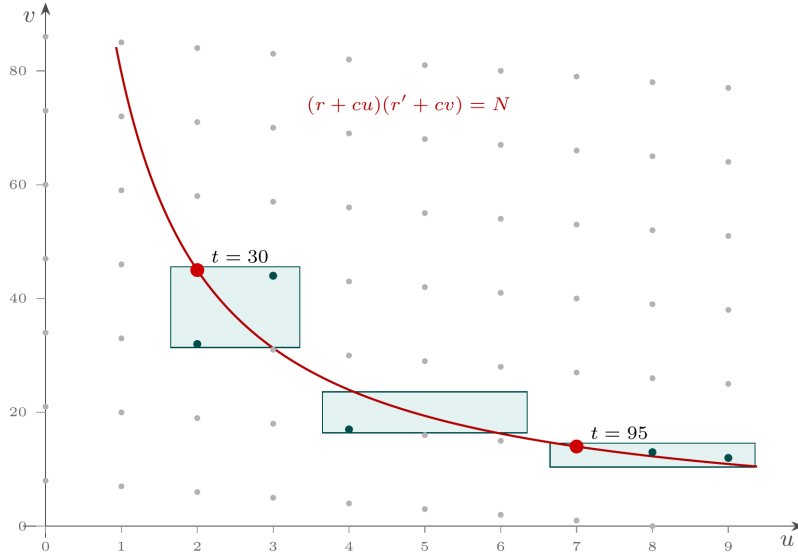


FIGURE 7. The divisors of  $N = 17670$  in the class 4 modulo 13 up to  $\sqrt{N}$ , found as in Lemma 4.2. The grey points form the coset (4.2), here  $v \equiv 8 - u \pmod{13}$ ; the curve is the hyperbola (4.1). The boxes cover the curve for  $2 \leq u \leq 9$  (the values  $u = 0, 1$  are tested directly), and only the points inside a box are examined. Two of them lie on the curve and give the divisors  $t = 30$  and  $t = 95$ .

**Theorem 4.4.** *Let  $(p_1, \dots, p_{k-3})$  be a node at depth  $k - 3$  of the search of Section 3. Every  $n = p_1 \cdots p_k$  with  $n + \varepsilon = 2\varphi(n)$  extending this node is found by running  $s = p_{k-2}$  over the admissible primes of its interval and, for each  $s$ , listing the divisors of Proposition 4.1 by Lemmas 4.2 and 4.3 or by a complete factorisation of  $N$ .*

*Proof.* The prime  $p_{k-2}$  lies in the interval and passes the prune, as in the proof of Theorem 3.6. For  $s = p_{k-2}$ , Proposition 4.1 shows that  $t = cp_{k-1} - 2B'$  is one of the divisors listed: Lemma 4.2 shows that  $u = (t - r)/c$  lies in the set (4.3) of the box that contains it, and Lemma 4.3 lists every element of that set, or, if  $N$  is factored,  $t$  is read off from its factorisation. The converse part of Proposition 4.1 then recovers  $p_{k-1}$  and  $p_k$ .  $\square$

Every divisor that is found yields integers  $p \leq q$  with  $A_{k-3}spq + \varepsilon = 2B_{k-3}(s-1)(p-1)(q-1)$ , by the converse part of Proposition 4.1, and  $n = p_1 \cdots p_{k-3}spq$  is a solution exactly when  $s < p < q$  and  $p$  and  $q$  are prime. As before, the correctness of a negative answer depends on no primality proof except in the factoring route, where every prime factor of  $N$  is proved prime: below  $\psi_{13}$  by the strong probable-prime test to the first thirteen prime bases, which is a proof there by the work of Sorenson and Webster [34], and above it by the primality prover of FLINT [13], which uses the  $n - 1$  and  $n + 1$  methods of Brillhart, Lehmer and Selfridge [5] when enough of  $n - 1$  or  $n + 1$  can be factored, and otherwise the Jacobi sum test of Adleman, Pomerance and Rumely [1] in the form given by Cohen and Lenstra [9].

**4.4. The sum of the two factors.** The classes of Proposition 4.1 are special:  $t$  and  $N/t$  lie in the same class modulo  $c$ , because  $N \equiv r^2$ . This allows a third treatment which uses neither boxes nor factoring. Write  $t = r + cu$  and  $N/t = r + cv$  with  $0 \leq u \leq v$ . Since  $r' = r$ , the integer  $w$  of Section 4.2 is  $w = (N - r^2)/c$ , and (4.1) becomes

$$cuv + r(u + v) = w, \quad (4.4)$$

so the sum  $S = u + v$  lies in the single class  $S \equiv wr^{-1} \pmod{c}$ , and  $u$  and  $v$  are the roots of  $X^2 - SX + (w - rS)/c$ . Conversely, if  $S \equiv wr^{-1} \pmod{c}$ , then  $(w - rS)/c$  is an integer, and if  $S^2 - 4(w - rS)/c = e^2$  with an integer  $e \geq 0$  of the parity of  $S$ , then  $u = (S - e)/2$  and  $v = (S + e)/2$  are integers satisfying (4.4), so that  $(r + cu)(r + cv) = N$ . Fix  $u_d \geq 1$  and put  $t_d = r + cu_d$ ; if  $t_d > \sqrt{N}$  there is no divisor with  $u \geq u_d$  to look for. Since  $x + N/x$  decreases on  $(0, \sqrt{N}]$ , every divisor  $t \leq \sqrt{N}$  with  $u \geq u_d$  has  $2r + cS = t + N/t \leq t_d + N/t_d$ , while  $t + N/t \geq 2\sqrt{N}$ . So  $S$  runs over at most  $N/(c^3 u_d) + 1$  values of one residue class, and each value needs a single test whether  $S^2 - 4(w - rS)/c$  is a perfect square. Every divisor with  $u \geq u_d$  is found in this way, and by the converse every square found gives a factorisation  $N = (r + cu)(r + cv)$ . The divisors with  $u < u_d$  are found by trial division. With  $u_d$  close to  $(N/c^3)^{1/2}$  the work is about  $2(N/c^3)^{1/2} + 2$  operations. This is comparable with the number of boxes above, but unlike the work in the boxes it does not depend on where the divisors lie.

The same argument bounds the number of divisors.

**Proposition 4.5.** *In the situation of Proposition 4.1, suppose that  $c^3 \geq N + 2c^2$ . Then at most one divisor  $t \leq \sqrt{N}$  of  $N$  with  $t \equiv r \pmod{c}$  is different from  $r$ , and  $N$  has at most four positive divisors congruent to  $r$  modulo  $c$ .*

*Proof.* By Proposition 4.1, every divisor  $t$  of  $N$  in the class has  $N/t \equiv r^2 r^{-1} = r \pmod{c}$ . Let  $t_1 < t_2 \leq \sqrt{N}$  be divisors of  $N$  in the class, both different from  $r$ . The hypothesis gives  $c^3 > 2c^2$ , so  $c > 2$  and  $0 < r < c$ . Hence  $t_i = r + cu_i$  with  $u_i \geq 1$ , and  $N/t_i = r + cv_i$  with  $v_i \geq u_i$ . By (4.4) the sums  $u_i + v_i$  lie in one class modulo  $c$ , so the numbers  $t_i + N/t_i = 2r + c(u_i + v_i)$  differ by a multiple of  $c^2$ . The difference is positive, because  $x + N/x$  decreases on  $(0, \sqrt{N}]$ , and both numbers lie between  $2\sqrt{N}$  and  $r + c + N/(r + c) < 2c + N/c$ . Hence  $c^2 < 2c + N/c - 2\sqrt{N} < 2c + N/c$ , that is,  $c^3 < N + 2c^2$ , contrary to the hypothesis. Finally, every divisor of  $N$  above  $\sqrt{N}$  in the class is  $N/t$  for a divisor  $t < \sqrt{N}$  in the class, and so there are at most four divisors in the class:  $r$ ,  $N/r$ ,  $t$  and  $N/t$ .  $\square$

For classes of this special kind, and at the price of the term  $2c^2$ , Proposition 4.5 sharpens Lenstra's bound of 11 to four.

We ran this route, too, over every pair of a node at depth  $k - 3$  and an admissible  $s$ . It takes its primes  $s$  from a sieve of its own and uses no factorisation, so its negative answers depend on no primality test at all.

## 5. LEHMER'S EQUATION: PROOF OF THEOREM 1.1

The last two assertions of Theorem 1.1 were proved in Section 2.1. For the first, let  $n$  be composite with  $\varphi(n) \mid n - 1$  and suppose  $\omega(n) \leq 15$ . By Lemma 2.2,  $3 \nmid n$ ; by Lemma 2.1(i),  $n = p_1 \cdots p_k$  with  $5 \leq p_1 < \cdots < p_k$  and  $k \leq 15$ ; and by Lemma 2.3,  $n - 1 = 2\varphi(n)$ . If  $k \leq 6$  then  $\prod p_i/(p_i - 1) \leq \prod_{i \leq 6} p_i/(p_i - 1) = 1.9490\dots$ , which contradicts (3.2). It therefore suffices to run the search with  $\varepsilon = -1$  and  $p_{\min} = 5$  for  $7 \leq k \leq 15$ , and to find nothing. For  $k \leq 14$  we use the search of Section 3, and for  $k = 15$  the method of Section 4.

**5.1. Up to fourteen primes.** Table 2 records the outcome. The column *terminal* counts the nodes at depth  $k - 2$  whose interval for  $p_{k-1}$  is non-empty, and the last column gives the largest width  $U_{k-2} - L_{k-2}$  of such an interval. No solution was found for any  $k \leq 14$ , so by Theorem 3.6 there is none. The case  $k \leq 13$  recovers the theorem of Cohen and Hagis, and the case  $k = 14$  is the bound  $\omega(n) \geq 15$  announced by Renze and by Pinch.

For  $k = 14$  the numbers of nodes at depths  $0, 1, \dots, 12$  are

$$1, 6, 16, 27, 29, 23, 22, 33, 58, 127, 335, 1654, 29631,$$

TABLE 2. The search for  $n - 1 = 2\varphi(n)$  with  $5 \leq p_1$ . Both implementations visit the same number of nodes for every  $k$ .

| $k$ | nodes  | terminal | largest width for $p_{k-1}$ |
|-----|--------|----------|-----------------------------|
| 7   | 8      | 0        | —                           |
| 8   | 18     | 1        | 2                           |
| 9   | 29     | 1        | 3                           |
| 10  | 42     | 1        | 2                           |
| 11  | 73     | 7        | 49                          |
| 12  | 187    | 61       | 298                         |
| 13  | 1 092  | 725      | 25 854                      |
| 14  | 31 962 | 29 549   | 54 962 375                  |

as in the row  $k = 14$  of Figure 8, and 82 of the 29631 nodes at depth 12 have an empty interval for  $p_{13}$ . The largest prime occurring at depth 12 is 77563. Most of the work lies in a few terminal nodes. Exactly fifteen have an interval for  $p_{13}$  of width more than three million; they are listed in Table 3. Each of these intervals was enumerated in full, and independently the integer  $D = 2AB - C$  was factored completely for each of the fifteen; the factorisations are part of the archived data.

TABLE 3. The fifteen terminal nodes of the case  $k = 14$  with an interval for  $p_{13}$  of width more than  $3 \cdot 10^6$ , with the number of digits of  $D = 2AB - C$ .

| $p_1, \dots, p_{12}$                             | width      | digits of $D$ |
|--------------------------------------------------|------------|---------------|
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25867 | 54 962 375 | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25889 | 19 589 597 | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 73, 317, 5483  | 14 242 697 | 40            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25903 | 13 903 362 | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25913 | 11 517 582 | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25919 | 10 443 048 | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25933 | 8 577 304  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25939 | 7 967 703  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 83, 97, 107, 3343  | 6 121 542  | 39            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 97, 199, 577   | 6 086 511  | 38            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 25969 | 5 880 810  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 26003 | 4 537 275  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 26017 | 4 147 802  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, 26029 | 3 863 803  | 41            |
| 5, 7, 13, 17, 19, 23, 37, 59, 83, 109, 173, 199  | 3 320 179  | 37            |

Eleven of the fifteen share the prefix 5, 7, 13, 17, 19, 23, 37, 59, 67, 89, 173, whose product  $\prod p/(p-1)$  is already within  $8 \cdot 10^{-5}$  of 2; the twelfth prime then closes most of the remaining gap. For the node ending in 25867 one has  $T - 1 = C/A = 1.8 \cdot 10^{-8}$ , and for the node ending in 5483 one has  $T - 1 = 7.0 \cdot 10^{-8}$ . These near misses are the reason the width of the last interval grows so quickly with  $k$ : by Proposition 3.2 with  $m = 2$  the interval for  $p_{k-1}$  runs from about  $A/C$  to about  $2A/C$ , and a small  $C$  means a long interval. For the node ending in 25867 this

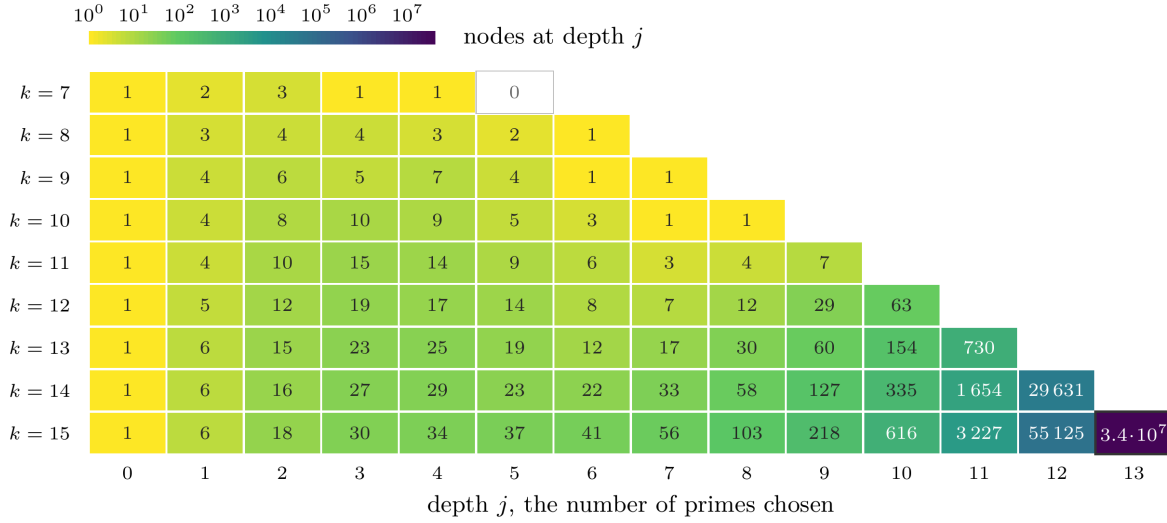


FIGURE 8. Number of nodes at each depth  $j$  of the search for  $k = 7, \dots, 15$  primes, with the colour on a logarithmic scale. The counts are the same for  $n - 1 = 2\varphi(n)$  and  $n + 1 = 2\varphi(n)$  with  $p_1 \geq 5$ . The tree stays small until the last two levels. The last cell of each row counts the nodes at depth  $k - 2$ : for  $k \leq 14$  these are the nodes where  $p_{k-1}$  is enumerated, and for  $k = 15$  (framed) they are the 33865004 choices of  $p_{13}$ , each completed by the method of Section 4.

gives exactly the interval  $[54962376, 109924751]$ . Among the fifteen integers  $D$  one, for the node ending in 25913, is itself a prime of 41 digits<sup>9</sup>.

**5.2. Fifteen primes.** For  $k = 15$  the traversal of Section 3 is stopped at depth 12. The numbers of nodes at depths  $0, 1, \dots, 12$  are

$$1, 6, 18, 30, 34, 37, 41, 56, 103, 218, 616, 3227, 55125,$$

and 54985 of the nodes at depth 12 have a non-empty interval for  $p_{13}$ . Both implementations of the search produce the same 54985 prefixes with the same intervals, and so do the bounds for  $\varepsilon = +1$ . The intervals have total length  $1.005 \cdot 10^9$  and contain 33865004 primes  $s$  that pass the congruence prune. For each of these pairs of a prefix and a prime  $s = p_{13}$  we solved the two-prime problem of Proposition 4.1, in which  $N$  has between 40 and 58 digits, three times: by the two implementations of the lattice method (Section 4.2) and by the route of Section 4.4. The first lattice implementation factored  $N$  completely in the 473 cases where the estimated number of boxes exceeded its threshold and took 19 minutes of processor time in all; the second factored  $N$  in 1772 cases and took 45 minutes; the route of Section 4.4, which factors nothing, took 13 minutes. The three agree on each of the 55443 pieces into which the work was divided.

Figure 9 shows how the pairs are distributed according to the ratio  $c^3/N$ , which by Section 4 decides how hard the two-prime problem is. For 33786783 of them, all but 0.23 per cent, the modulus satisfies  $c^3 > N$ , so that Lenstra's bound applies and a few boxes suffice. The remaining 78221 pairs come from values of  $s$  just above the lower end of long intervals, where  $c = C_{12}s - 2B_{12}$  is small. The most extreme is the prefix ending in 25867 of Table 3, whose

<sup>9</sup>For this node (3.5) has only the trivial factorisations  $D = 1 \cdot D$  and  $D = D \cdot 1$ , and neither gives a pair of primes in the required range. The enumeration of the  $1.15 \cdot 10^7$  candidates for  $p_{13}$  reaches the same conclusion without knowing that  $D$  is prime.

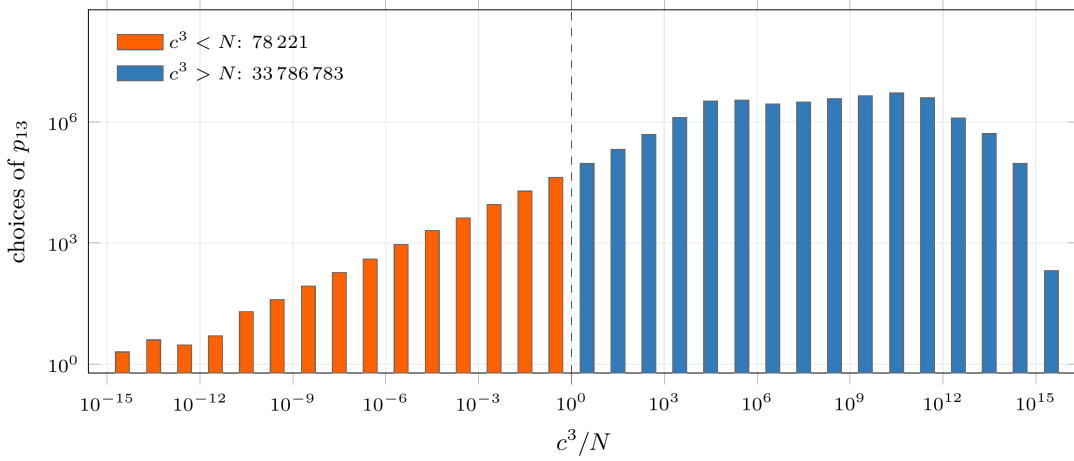


FIGURE 9. The 33865004 two-prime problems of the case  $k = 15$ , by the ratio  $c^3/N$  on a logarithmic scale (bins of one decade). To the right of the dashed line is Lenstra's range, where  $N$  has at most 11 divisors in the class and a few boxes find them. The left tail comes from the lower ends of the longest intervals for  $p_{13}$ .

first admissible  $s$  gives  $c^3/N \approx 2 \cdot 10^{-15}$ . For these pairs the route of Section 4.4 needs about  $2(N/c^3)^{1/2}$  operations.

For none of the 33865004 pairs does  $N$  have a divisor in the class of Proposition 4.1 that yields integers  $s < p < q$ , and the three computations agree on this. So there is not even a completion by composite integers, and in particular no solution with  $k = 15$ . Together with the case  $k \leq 14$  this proves Theorem 1.1.

## 6. THE SIZE OF A SOLUTION

This section proves Theorem 1.2. Of what precedes it uses only Section 2, and unlike Sections 3 to 5 it does not assume  $M = 2$ . Throughout,  $n = p_1 \cdots p_k$  is a composite solution of (1.1) with  $p_1 < \cdots < p_k$ , so that  $\varepsilon = -1$  and  $M = (n-1)/\varphi(n)$ , and  $A_j, B_j$  and  $P_j$  are as in the Notation of Section 1; thus  $B_j = \varphi(A_j)$ . By Lemma 2.1,  $n$  is odd and squarefree,  $k \geq 2$ ,  $M \geq 2$ , and no prime factor of  $n$  divides  $p-1$  for a prime factor  $p$  of  $n$ . By Lemma 2.2, if  $3 \mid n$  then  $p_1 = 3$  and  $M \equiv 1 \pmod{3}$ , so that  $M \geq 4$ .

**Lemma 6.1.** (i)  $P_1 < P_2 < \cdots < P_{k-1} < M < P_k = M + 1/\varphi(n)$ .

(ii)  $p_k = (MB_{k-1} - 1)/(MB_{k-1} - A_{k-1})$ .

*Proof.* The equality  $P_k = M + 1/\varphi(n)$  is Lemma 2.1(iii), and  $P_j$  increases with  $j$ . Writing  $n-1 = M\varphi(n)$  as  $A_{k-1}p_k - 1 = MB_{k-1}(p_k - 1)$  gives  $p_k(MB_{k-1} - A_{k-1}) = MB_{k-1} - 1 > 0$ . Hence  $MB_{k-1} > A_{k-1}$ , that is  $P_{k-1} < M$ , and (ii) follows.  $\square$

**6.1. A sharp product lemma.** For integers  $m \geq 1$  and real  $y \geq 1$  put

$$E_m(y) = y^{2^m} - y^{2^{m-1}}.$$

The function  $E_m$  is increasing on  $y \geq 1$ , satisfies  $E_m(y) < y^{2^m}$ , and

$$E_m(y^{2^u}) = E_{m+u}(y) \quad (m \geq 1, u \geq 0). \quad (6.1)$$

**Theorem 6.2.** *Let  $a < b$  be positive integers,  $d = b - a$ ,  $m \geq 2$ , and let  $2 \leq x_1 \leq x_2 \leq \cdots \leq x_m$  be integers with*

$$\prod_{i=1}^m \left(1 - \frac{1}{x_i}\right) \leq \frac{a}{b} < \prod_{i=1}^{m-1} \left(1 - \frac{1}{x_i}\right). \quad (6.2)$$

Then

$$a x_1 x_2 \cdots x_m \leq E_{m-1} \left( \frac{b(a+1)}{d} \right).$$

If  $d$  divides  $b + 1$ , equality holds for a suitable choice of the  $x_i$ .

For  $d = 1$  the bound is  $E_{m-1}((a+1)^2) = E_m(a+1)$ , which is the lemma of Cook and Nielsen (Lemma 6.5 below) behind the bound of Burek and Żmija. For  $d \geq 2$  it is smaller, and it improves the bound  $E_{m-1}(b^2/d)$  of Stone [35, Corollary 1], since  $b(a+1) < b^2$ . Applied to the prime factors of  $n$  that follow the first  $j$ , Theorem 6.2 is strong whenever  $P_j$  is still well below  $M$ .

Throughout this subsection  $a < b$  are positive integers,  $d = b - a$ , and  $x_1 \leq \cdots \leq x_m$  are integers at least 2 satisfying (6.2). We write

$$X_u = x_1 \cdots x_u, \quad Y_u = (x_1 - 1) \cdots (x_u - 1),$$

with  $X_0 = Y_0 = 1$ , and use the same notation  $Z_u = z_1 \cdots z_u$  for other sequences.

**Lemma 6.3.** *For  $0 \leq u \leq m - 1$  the tail  $x_{u+1}, \dots, x_m$  satisfies (6.2) with  $(a, b)$  replaced by  $(aX_u, bY_u)$ , and  $aX_u < bY_u$ .*

*Proof.* Divide (6.2) by  $\prod_{i \leq u} (1 - 1/x_i) = Y_u/X_u$ . Since  $u \leq m - 1$ , we have  $Y_u/X_u \geq \prod_{i \leq m-1} (1 - 1/x_i) > a/b$ .  $\square$

The next lemma is a majorisation inequality; only the sequence  $x$  needs to be ordered.

**Lemma 6.4.** *Let  $x_1 \leq \cdots \leq x_m$  and  $z_1, \dots, z_m$  be real numbers greater than 1 with  $X_u \geq Z_u$  for  $1 \leq u \leq m$ . Then*

$$\prod_{i=1}^m \left(1 - \frac{1}{x_i}\right) \geq \prod_{i=1}^m \left(1 - \frac{1}{z_i}\right),$$

with strict inequality if  $X_m > Z_m$ .

*Proof.* The function  $f(t) = \log(1 - e^{-t})$  on  $t > 0$  has derivative  $f'(t) = 1/(e^t - 1)$ , which is positive and decreasing, so  $f$  is concave. Put  $\tau_i = \log x_i$  and  $\sigma_i = \log z_i$ , so that  $f(\tau_i) = \log(1 - 1/x_i)$  and  $f(\sigma_i) = \log(1 - 1/z_i)$ . By concavity

$$f(\sigma_i) \leq f(\tau_i) + c_i(\sigma_i - \tau_i), \quad c_i = f'(\tau_i) = \frac{1}{x_i - 1},$$

and  $c_1 \geq c_2 \geq \cdots \geq c_m > 0$  because the  $x_i$  are nondecreasing. The partial sums  $\Delta_u = \sum_{i \leq u} (\sigma_i - \tau_i) = \log(Z_u/X_u)$  are at most 0 by hypothesis, and summation by parts gives

$$\sum_{i=1}^m c_i(\sigma_i - \tau_i) = \sum_{u=1}^{m-1} (c_u - c_{u+1})\Delta_u + c_m\Delta_m \leq c_m\Delta_m \leq 0,$$

with  $c_m\Delta_m < 0$  when  $X_m > Z_m$ . Hence  $\sum f(\sigma_i) \leq \sum f(\tau_i)$ , strictly in that case, and the lemma follows on exponentiating.  $\square$

**Lemma 6.5** (Cook, Nielsen). *Under (6.2), with  $m \geq 1$ , we have  $aX_m \leq E_m(a+1)$ .*

This is the lemma of Cook [10] and Nielsen [26] mentioned in the introduction. We include its proof, on which the proof of Theorem 6.2 builds.

*Proof.* We use induction on  $m$ . Define

$$z_u = (a+1)^{2^{u-1}} + 1 \quad (1 \leq u \leq m-1), \quad z_m = (a+1)^{2^{m-1}}.$$

Induction on  $u$  gives  $aZ_u = (a+1)^{2^u} - 1$  for  $0 \leq u \leq m-1$ , and then  $aZ_m = E_m(a+1)$ . Writing  $c_u = (a+1)^{2^{u-1}}$ , so that  $c_{u+1} = c_u^2$  and  $1 - 1/z_u = c_u(c_u - 1)/(c_{u+1} - 1)$  for  $u < m$ , the product telescopes:

$$\prod_{u=1}^m \left(1 - \frac{1}{z_u}\right) = \frac{(a+1)^{2^{m-1}-1} a}{c_m - 1} \cdot \frac{c_m - 1}{c_m} = \frac{a}{a+1}. \quad (6.3)$$

These identities use only that  $a$  is a positive real number, and we shall apply them once more with a value of  $a$  that need not be an integer.

Suppose first that  $X_u \leq Z_u$  for some  $1 \leq u \leq m-1$ , so that  $aX_u + 1 \leq (a+1)^{2^u}$ . By Lemma 6.3 and the induction hypothesis applied to the  $m-u$  terms  $x_{u+1}, \dots, x_m$  with numerator  $aX_u$ , and by (6.1),

$$aX_m = aX_u \prod_{i>u} x_i \leq E_{m-u}(aX_u + 1) \leq E_{m-u}((a+1)^{2^u}) = E_m(a+1).$$

Otherwise  $X_u > Z_u$  for every  $u \leq m-1$  (for  $m=1$  this is vacuous). If also  $X_m > Z_m$ , Lemma 6.4 and (6.3) give  $\prod(1 - 1/x_i) > a/(a+1) \geq a/b$ , contrary to (6.2). Hence  $X_m \leq Z_m$  and  $aX_m \leq E_m(a+1)$ .  $\square$

The proof compares  $x_1, \dots, x_m$  with an extremal sequence  $z_1, \dots, z_m$  through the partial products  $X_u$  and  $Z_u$ , and splits according to whether  $X_u \leq Z_u$  for some  $u \leq m-1$  (Figure 10).

*Proof of Theorem 6.2.* Put  $G = b(a+1)/d$ , so that  $G-1 = a(b+1)/d$ , and define real numbers

$$z_1 = \frac{b+1}{d}, \quad z_u = G^{2^{u-2}} + 1 \quad (2 \leq u \leq m-1), \quad z_m = G^{2^{m-2}}.$$

Then  $aZ_1 = G-1$ , induction gives  $aZ_u = G^{2^{u-1}} - 1$  for  $1 \leq u \leq m-1$ , and  $aZ_m = (G^{2^{m-2}} - 1)G^{2^{m-2}} = E_{m-1}(G)$ . The numbers  $z_2, \dots, z_m$  are the sequence of the proof of Lemma 6.5 with  $G-1$  in place of  $a$  and  $m-1$  terms, so by (6.3)

$$\prod_{u=1}^m \left(1 - \frac{1}{z_u}\right) = \frac{a+1}{b+1} \cdot \frac{G-1}{G} = \frac{a+1}{b+1} \cdot \frac{a(b+1)}{b(a+1)} = \frac{a}{b}.$$

Suppose first that  $X_u \leq Z_u$  for some  $1 \leq u \leq m-1$ , so that  $aX_u + 1 \leq G^{2^{u-1}}$ . Lemmas 6.3 and 6.5, applied to the  $m-u \geq 1$  terms  $x_{u+1}, \dots, x_m$ , and (6.1) give

$$aX_m \leq E_{m-u}(aX_u + 1) \leq E_{m-u}(G^{2^{u-1}}) = E_{m-1}(G).$$

Otherwise  $X_u > Z_u$  for all  $u \leq m-1$ . If also  $X_m > Z_m$ , then Lemma 6.4, which applies because every  $z_u$  exceeds 1 (for  $z_1$  because  $b+1 > d$ ), gives  $\prod(1 - 1/x_i) > \prod(1 - 1/z_i) = a/b$ , contrary to (6.2). Hence  $X_m \leq Z_m$ , that is  $aX_m \leq E_{m-1}(G)$ .

If  $d \mid b+1$ , then  $z_1$  is an integer greater than 1, so  $z_1 \geq 2$ , and  $G-1 = az_1$ ,  $G \geq 3$  and all  $z_u$  are integers. They are nondecreasing:  $z_2 \geq G \geq z_1$ , since  $G \geq z_1$  amounts to  $ab \geq 1$ ; the middle terms increase; and for  $m \geq 3$  we have  $z_m = (z_{m-1} - 1)^2 \geq z_{m-1}$  because  $z_{m-1} - 1 \geq G \geq 3$ . The product of  $1 - 1/z_u$  is  $a/b$  over all  $m$  terms and larger over the first  $m-1$ , so the  $z_u$  satisfy (6.2) and attain  $aZ_m = E_{m-1}(G)$ .  $\square$

*Remark 6.6.* Since  $G = (a+1)/(1 - a/b)$ , the bound of Theorem 6.2 depends on  $b$  only through the distance of  $a/b$  from 1. Its base is smaller than the base  $(a+1)^2$  of Lemma 6.5 by the factor  $(a+1)d/b$ , which is at least 1 because  $(a+1)d - b = a(d-1) \geq 0$ ; Figure 11 shows the two bases. Stone's bound  $E_{m-1}(b^2/d)$  exceeds ours by the factor  $b/(a+1)$  in the base.

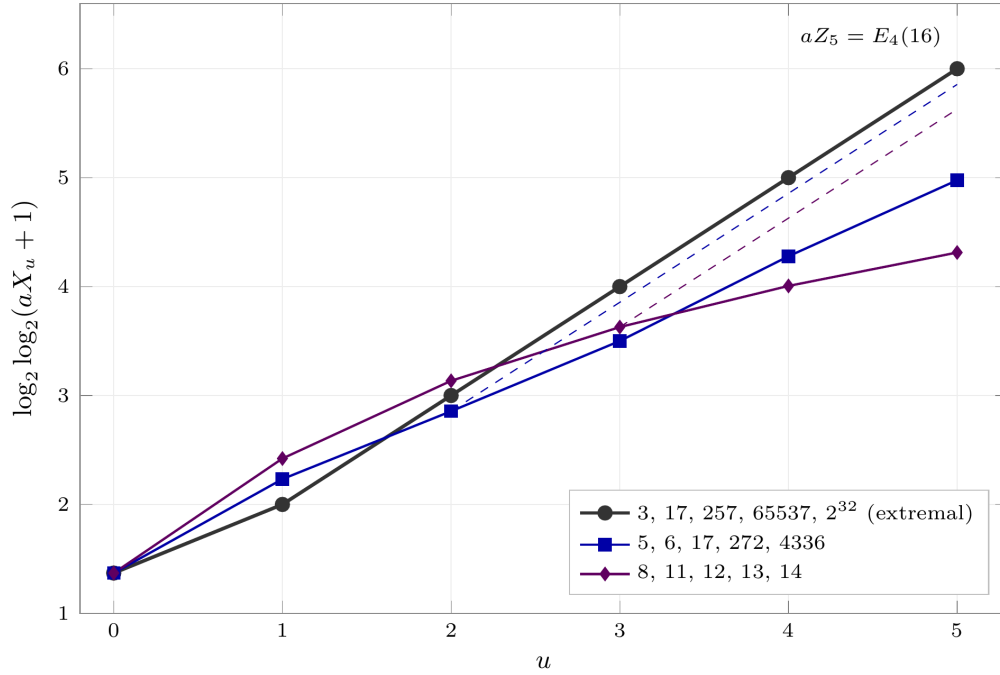


FIGURE 10. The proof of Theorem 6.2 for  $a/b = 5/8$  and  $m = 5$ , on the scale  $\log_2 \log_2(aX_u + 1)$ . The extremal sequence  $3, 17, 257, 65537, 2^{32}$  rises by exactly 1 per step from  $u = 1$  on and ends at  $aZ_5 = E_4(16)$ . The two other sequences satisfy (6.2), start above it, and fall below it at  $u = 2$  and  $u = 3$ . From that point Lemmas 6.3 and 6.5 give  $aX_5 < (aX_u + 1)^{2^{5-u}}$ , the dashed line of slope 1, which ends below  $E_4(16)$ . A sequence that stayed above the extremal one up to  $u = 4$  would end below it at  $u = 5$ , by Lemma 6.4.

In the application below  $a = A_j$  and  $b = MB_j$ , and then the base  $G = (A_j + 1)/(1 - P_j/M)$  decreases as  $M$  increases, so the smallest admissible value of  $M$  yields an upper bound. The base  $b^2/d = M^2 B_j^2 / (MB_j - A_j)$  of Stone has no such property: it decreases for  $M < 2P_j$  and increases for  $M > 2P_j$ .

*Remark 6.7.* The extremal sequences of Theorem 6.2 are explicit. For  $a/b = 5/8$  we have  $d = 3$ ,  $z_1 = 3$  and  $G = 16$ , and with  $m = 4$  the sequence is  $3, 17, 257, 65536$ :

$$\frac{2}{3} \cdot \frac{16}{17} \cdot \frac{256}{257} \cdot \frac{65535}{65536} = \frac{5}{8}, \quad 5 \cdot 3 \cdot 17 \cdot 257 \cdot 65536 = 2^{32} - 2^{16} = E_3(16).$$

The first three terms are Fermat primes. For  $a = 1$  the sequence of the proof of Lemma 6.5 consists of the Fermat numbers  $F_0, \dots, F_{m-2}$  followed by  $2^{2^{m-1}}$ , and the products  $F_0 F_1 \dots F_{r-1} = 2^{2^r} - 1$  with  $1 \leq r \leq 5$  are the solutions  $3, 15, 255, 65535$  and  $4294967295$  of (1.2) in (1.3). Such products have  $n/\varphi(n)$  just below 2, and so do the deepest nodes of the search of Section 6.3 (Section 6.6).

**6.2. Heavy solutions.** For  $0 \leq j \leq k - 1$  put

$$V_j = V_j(M) = \frac{MB_j(A_j + 1)}{MB_j - A_j} = \frac{A_j + 1}{1 - P_j/M}. \quad (6.4)$$

By Lemma 6.1(i) the denominator is positive, and  $V_j(M)$  decreases as  $M$  increases.

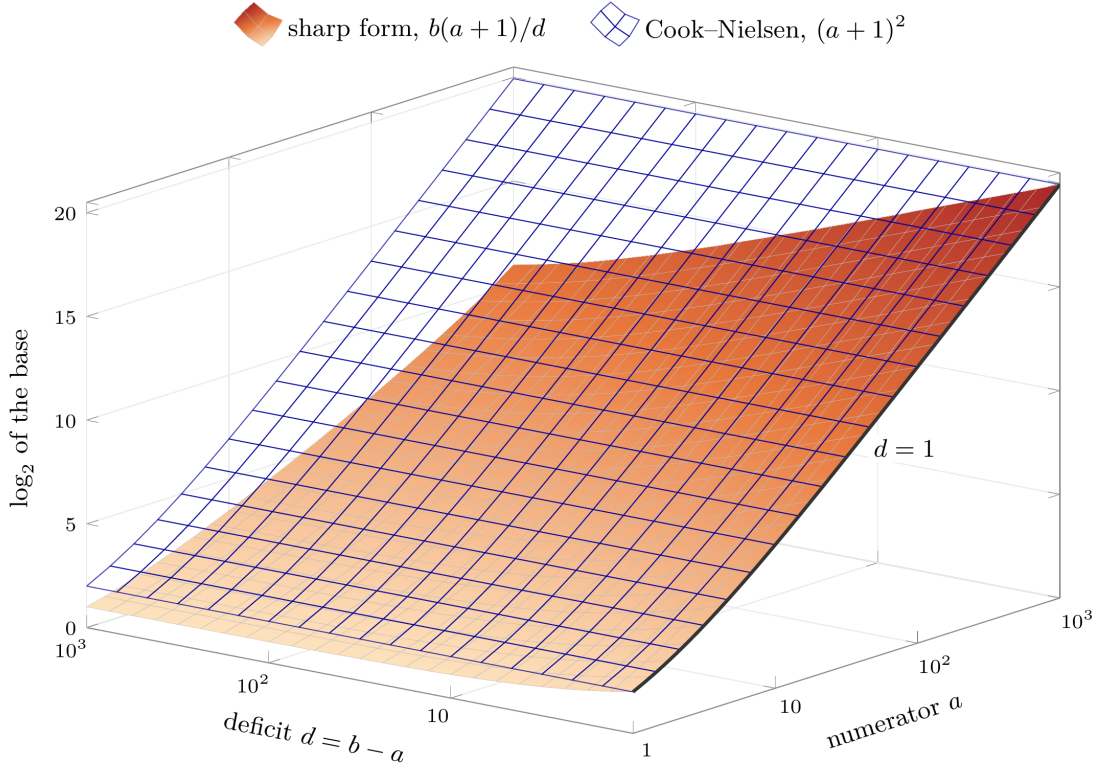


FIGURE 11. The base of the bound for  $aX_m$  written as  $E_{m-1}(\cdot)$ , on a  $\log_2$  scale, over the numerator  $a$  and the deficit  $d = b - a$ . The wire mesh is the base  $(a+1)^2$  of the lemma of Cook and Nielsen, the shaded sheet below it the base  $b(a+1)/d$  of Theorem 6.2. They meet along  $d = 1$ , where the two bounds coincide, and separate as  $d$  grows.

**Proposition 6.8.** For  $0 \leq j \leq k-1$ ,

$$n < (A_j + 1)^{2^{k-j}} \quad \text{and} \quad n < V_j^{2^{k-j-1}}.$$

*Proof.* Take  $m = k - j$ ,  $x_i = p_{j+i}$ ,  $a = A_j$  and  $b = MB_j$ . Then

$$\prod_{i=j+1}^k \left(1 - \frac{1}{p_i}\right) = \frac{P_j}{P_k} < \frac{P_j}{M} = \frac{a}{b} < \frac{P_j}{P_{k-1}} = \prod_{i=j+1}^{k-1} \left(1 - \frac{1}{p_i}\right)$$

by Lemma 6.1(i), so (6.2) holds,  $a < b$ , and  $ax_1 \cdots x_m = n$ . Lemma 6.5 gives  $n \leq E_{k-j}(A_j + 1)$ , which implies the first bound. For  $j \leq k-2$ , Theorem 6.2 gives  $n \leq E_{k-j-1}(V_j)$ , because  $b(a+1)/d = V_j$ . For  $j = k-1$ , Lemma 6.1(ii) gives  $n = A_{k-1}p_k < A_{k-1}MB_{k-1}/(MB_{k-1} - A_{k-1}) < V_{k-1}$ .  $\square$

**Definition 6.9.** Let  $s \geq 0$  be an integer, and put  $\theta_i = 2^{2^{i-s}}$  for  $i \geq s$  and  $\theta_i = 1$  for  $i < s$ . A composite solution  $n$  of (1.1) is  $s$ -heavy if for every  $j$  with  $0 \leq j \leq k-1$

(H1)  $A_j \geq \theta_j$ , and

(H2)  $V_j \geq \theta_{j+1}$ .

**Corollary 6.10.** A composite solution of (1.1) with  $k$  prime factors that is not  $s$ -heavy satisfies  $n < 2^{2^{k-s}}$ .

*Proof.* If (H1) fails at  $j$ , then  $j \geq s$  and  $A_j + 1 \leq 2^{2^{j-s}}$ , and Proposition 6.8 gives  $n < (A_j + 1)^{2^{k-j}} \leq 2^{2^{k-s}}$ . If (H2) fails at  $j$ , then  $j + 1 \geq s$  because  $V_j > 1$ , and  $n < V_j^{2^{k-j-1}} < 2^{2^{k-s}}$ .  $\square$

For  $s = 2$  condition (H1) alone contradicts Lemma 6.1(i), and the argument can be carried out by hand.

**Proposition 6.11.** *No composite solution of (1.1) is 2-heavy. Consequently every composite solution with  $k$  prime factors satisfies  $n < 2^{2^{k-2}}$ .*

*Proof.* Let  $n$  be 2-heavy, and let  $1 \leq i \leq k - 1$ . Condition (H1) at  $j = i$  and  $A_i \leq p_i^i$  give  $p_i \geq \theta_i^{1/i} = 2^{2^{i-2}/i}$ . Hence  $p_7 \geq 24$ ,  $p_8 \geq 256$ ,  $p_9 > 19000$  and  $p_i > 4^i$  for  $i \geq 10$ , so that

$$\sum_{8 \leq i \leq k-1} \frac{1}{p_i - 1} < \frac{1}{256} + \frac{1}{19000} + \sum_{i \geq 10} \frac{1}{4^i - 1} < 0.004.$$

Put  $f(x) = x/(x - 1)$ , which decreases for  $x > 1$  and satisfies  $f(x) \leq e^{1/(x-1)}$ . Then  $P_k = \prod_{i \leq k} f(p_i)$  and  $f(p_k) < f(p_{k-1})$ . Let  $b_1, \dots, b_7$  be lower bounds for  $p_1, \dots, p_7$  and  $\beta_j = \prod_{i \leq j} f(b_i)$ . If  $k \leq 8$ , then  $P_k < \beta_{k-1} f(b_{k-1})$ . If  $k \geq 9$ , the prime factors  $p_8, \dots, p_k$  contribute less than  $e^{0.004} f(p_{k-1}) < e^{0.008}$ , and  $P_k < \beta_7 e^{0.008}$ . We distinguish four cases.

- (a) If  $3 \mid n$ , then  $p_1 = 3$  and  $M \geq 4$  by Lemma 2.2, and the next prime factors are at least 5, 7, 11, 13, 17 and, as  $p_7 \geq 24$ , 29.
- (b) If  $p_1 = 5$  and  $p_2 = 7$ , then no other prime factor is congruent to 1 modulo 5 or 7 by Lemma 2.1(ii), which excludes 11, 29 and 31.
- (c) If  $p_1 = 5$  and  $p_2 \neq 7$ , then  $p_2 \geq 13$ , and 11 and 31 are excluded because they are congruent to 1 modulo 5.
- (d) If  $p_1 \geq 7$ , the prime factors are at least the primes from 7 on, and  $p_7 \geq 29$ .

In cases (b) to (d),  $3 \nmid n$  and  $M \geq 2$ . Table 4 lists the bounds  $b_1, \dots, b_7$  in each case and the larger of  $\max_{j \leq 7} \beta_j f(b_j)$  and  $\beta_7 e^{0.008}$ . In every case it is less than  $M$ , so that  $P_k < M$ , contrary to Lemma 6.1(i). Hence no composite solution is 2-heavy, and Corollary 6.10 with  $s = 2$  gives the bound.  $\square$

TABLE 4. The four cases in the proof of Proposition 6.11: lower bounds  $b_1, \dots, b_7$  for the seven smallest prime factors and the resulting upper bound for  $P_k$ , rounded up.

| case                      | $b_1, \dots, b_7$         | bound for $P_k$ | $M \geq$ |
|---------------------------|---------------------------|-----------------|----------|
| (a) $p_1 = 3$             | 3, 5, 7, 11, 13, 17, 29   | 2.972           | 4        |
| (b) $p_1 = 5, p_2 = 7$    | 5, 7, 13, 17, 19, 23, 37  | 1.957           | 2        |
| (c) $p_1 = 5, p_2 \neq 7$ | 5, 13, 17, 19, 23, 29, 37 | 1.738           | 2        |
| (d) $p_1 \geq 7$          | 7, 11, 13, 17, 19, 23, 29 | 1.749           | 2        |

The proof uses only condition (H1), whose failure gives the bound through the lemma of Cook and Nielsen, and the fact that no prime factor divides another minus one. For  $s = 1$  the same argument needs only the cases  $3 \mid n$  and  $3 \nmid n$ , with the primes from 5 on as lower bounds and  $p_6 \geq 41$ ,  $p_7 \geq 569$ ,  $p_8 \geq 65537$  from (H1); it shows that no composite solution is 1-heavy, which gives  $n < 2^{2^{k-1}}$ , half the number of binary digits allowed by the bound of Burek and Žmija. For larger  $s$  the estimates of this proof are too weak, and Theorem 1.2(i) follows once we know that no solution is 7-heavy, which Section 6.3 establishes with exact bounds at every node and with condition (H2) and Theorem 6.2 as well.

We first record how fast the prime factors of a heavy solution grow. Fix  $s$ , an index  $j \geq 0$  and the product  $A_j$  of  $j$  odd primes. For  $i > j$  let  $\lambda_i$  be the least positive integer  $\lambda$  with  $A_j \lambda^{i-j} \geq \theta_i$ , and put  $y_i = (\theta_i/A_j)^{1/(i-j)}$ . Then  $\lambda_i = \max\{1, \lceil y_i \rceil\}$ , so  $\lambda_i \geq y_i$ , and  $\lambda_i < y_i + 1$  whenever  $\lambda_i > 1$ .

**Lemma 6.12.** *Let  $i > j$ .*

- (i) *If  $i \geq j + 2$  and  $y_i > 1$ , then  $y_{i+1} \geq y_i^{4/3}$ .*
- (ii) *If  $n$  is an  $s$ -heavy solution,  $A_j$  is the product of its  $j$  smallest prime factors and  $i \leq k - 1$ , then  $p_i \geq \lambda_i$ .*

*Proof.* (i) If  $y_i > 1$ , then  $\theta_i > A_j \geq 1$ , so  $i \geq s$  and  $\theta_{i+1} = \theta_i^2$ . Hence

$$(i + 1 - j) \log y_{i+1} = 2 \log \theta_i - \log A_j \geq 2(\log \theta_i - \log A_j) = 2(i - j) \log y_i,$$

so  $\log y_{i+1} \geq \frac{2(i-j)}{i+1-j} \log y_i \geq \frac{4}{3} \log y_i$ , since  $i - j \geq 2$  and  $\log y_i > 0$ .

(ii) By (H1) at  $i$ ,  $\theta_i \leq A_i \leq A_j p_i^{i-j}$ , so  $p_i \geq \lambda_i$ .  $\square$

The search needs an upper bound for  $P_k/P_j$  that uses only  $p_1, \dots, p_j$  and a lower bound for  $p_{j+1}$ . For an integer  $q \geq 3$  and  $i > j$  let  $\pi_i(q)$  be the  $(i - j)$ -th smallest prime that is at least  $q$ , if this prime is less than  $2 \cdot 10^7$ , and  $\pi_i(q) = q + 2(i - j - 1)$  otherwise. Put  $b_i = \max\{\pi_i(q), \lambda_i\}$ , let  $i_0$  be the least index  $i \geq j + 3$  with  $\lambda_i > 10^{30}$ , and define

$$\Gamma_j(q) = \max \left\{ \frac{q}{q-1}, \max_{j+1 \leq i < i_0} \frac{b_i + 2}{b_i + 1} \prod_{l=j+1}^i \frac{b_l}{b_l - 1}, \left(1 + \frac{16}{\lambda_{i_0} - 2}\right) \prod_{l=j+1}^{i_0-1} \frac{b_l}{b_l - 1} \right\}. \quad (6.5)$$

The quantity  $\Gamma_j(q)$  depends only on  $j$ ,  $A_j$ ,  $q$  and  $s$ , and it is a finite computation because  $\lambda_i$  grows doubly exponentially in  $i$ .

**Lemma 6.13.** *Let  $n$  be an  $s$ -heavy solution,  $0 \leq j \leq k - 1$ , and let  $q$  be an integer with  $3 \leq q \leq p_{j+1}$ . Then  $P_k \leq P_j \Gamma_j(q)$ .*

*Proof.* The quotient  $P_k/P_j$  is the product of  $p_i/(p_i - 1)$  over  $j < i \leq k$ , and  $x/(x - 1)$  decreases in  $x$ . For  $j < i \leq k - 1$  we have  $p_i \geq \pi_i(q)$ , since  $p_{j+1} < \dots < p_i$  are odd primes at least  $q$ , and  $p_i \geq \lambda_i$  by Lemma 6.12(ii); so  $p_i \geq b_i$ . For the last prime,  $p_k \geq q$  if  $k = j + 1$  and  $p_k \geq p_{k-1} + 2 \geq b_{k-1} + 2$  otherwise. This bounds  $P_k/P_j$  by the first term of (6.5) when  $k = j + 1$ , and by the second term, with  $i = k - 1$ , when  $j + 2 \leq k \leq i_0$ .

Suppose  $k > i_0$ . The primes  $p_{j+1}, \dots, p_{i_0-1}$  contribute at most  $\prod_{l=j+1}^{i_0-1} b_l/(b_l - 1)$ , as before. Put  $Y = y_{i_0}$ . Since  $\lambda_{i_0} > 10^{30}$ , we have  $Y > \lambda_{i_0} - 1 \geq 10^{30}$ , and Lemma 6.12 gives  $p_i \geq \lambda_i \geq y_i \geq Y^{(4/3)^{i-i_0}}$  for  $i_0 \leq i \leq k - 1$ . As  $p_k > p_{k-1}$ , the factor of  $p_k$  is at most that of  $p_{k-1}$ , so

$$\prod_{i=i_0}^k \frac{p_i}{p_i - 1} \leq \prod_{i=i_0}^{k-1} \left(1 + \frac{1}{p_i - 1}\right)^2 \leq \exp \left(2 \sum_{u \geq 0} \frac{1}{Y^{(4/3)^u} - 1}\right).$$

Because  $(4/3)^u \geq 1 + u/3$  and  $Y \geq 10^{30}$ , the terms with  $u \geq 1$  add up to less than  $1/(Y - 1)$ , so the exponent is less than  $4/(Y - 1) \leq 1$ . With  $e^x \leq 1 + 2x$  for  $0 \leq x \leq 1$  the product is at most  $1 + 8/(Y - 1) < 1 + 8/(\lambda_{i_0} - 2)$ , which is smaller than the factor  $1 + 16/(\lambda_{i_0} - 2)$  of the third term of (6.5).  $\square$

**6.3. The search.** Fix  $s$ . Let  $S = (p_1, \dots, p_j)$ ,  $j \geq 0$ , be an increasing sequence of odd primes in which no  $p_i$  divides any  $p_l - 1$ , and write  $A_j$ ,  $B_j$  and  $P_j$  as before. Let  $\mu(S)$  be the least integer  $M > P_j$  with  $M \geq 2$  and, if  $p_1 = 3$ ,  $M \equiv 1 \pmod{3}$ , and let  $\mu'(S)$  be the next such integer. We say that a solution  $n$  extends  $S$  if  $j \leq k - 1$  and  $S$  consists of the  $j$  smallest prime

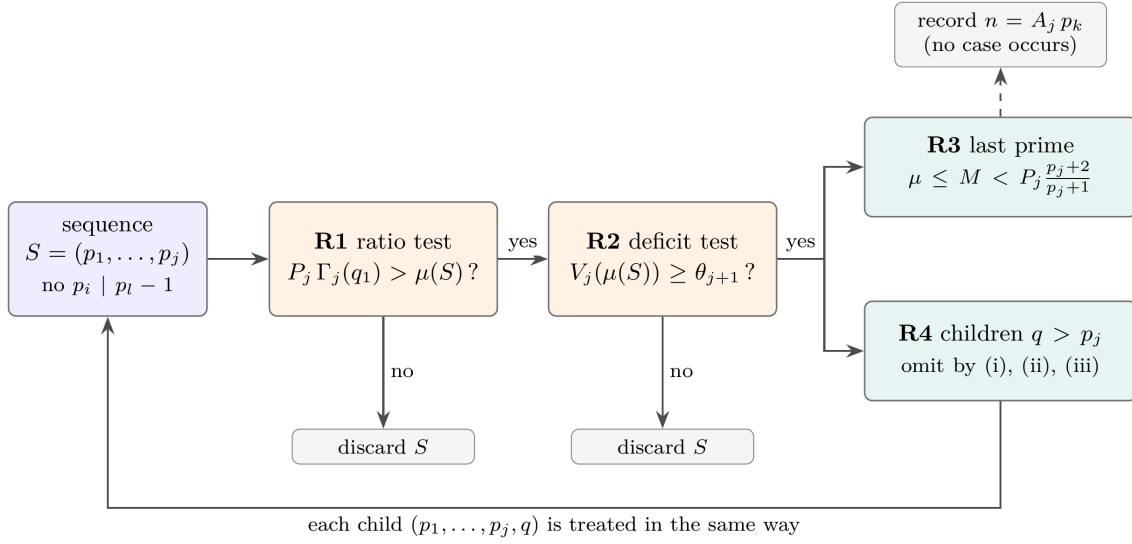


FIGURE 12. The treatment of a sequence  $S = (p_1, \dots, p_j)$  in the tree  $\mathcal{T}_s$ . A sequence that fails the ratio test or the deficit test cannot be extended by an  $s$ -heavy solution. A node is tested for a last prime factor and passes to its children every prime that the three omission rules allow.

factors of  $n$ . By Lemmas 2.1, 2.2 and 6.1(i), the quotient  $M$  of such an  $n$  satisfies  $M \geq \mu(S)$ . For  $M > P_j$  we use  $V_j(M)$  as in (6.4), computed from  $S$ .

The tree  $\mathcal{T}_s$  is built from the empty sequence by the following rules, applied to each sequence  $S$  of length  $j$  that is generated; Figure 12 summarises them.

(R1) *Ratio test*. Let  $q_1$  be the least prime greater than  $p_j$  ( $q_1 = 3$  if  $j = 0$ ). If  $P_j \Gamma_j(q_1) \leq \mu(S)$ , the sequence  $S$  is discarded.

(R2) *Deficit test*. If  $V_j(\mu(S)) < \theta_{j+1}$ , the sequence  $S$  is discarded.

A sequence that survives both tests is a *node* of  $\mathcal{T}_s$ . At a node:

(R3) *Last prime*. If  $j \geq 1$ , then for each integer  $M$  with  $\mu(S) \leq M < P_j(p_j + 2)/(p_j + 1)$ , and  $M \equiv 1 \pmod{3}$  if  $p_1 = 3$ , test whether  $(MB_j - 1)/(MB_j - A_j)$  is a prime greater than  $p_j$  that is not 1 modulo any  $p_i$ .

(R4) *Children*. Put  $\mu = \mu(S)$  and  $q_c = \lfloor \mu B_j / (\mu B_j - A_j) \rfloor$ . Generate the sequences  $(p_1, \dots, p_j, q)$  for the primes  $q > p_j$  with  $q \not\equiv 1 \pmod{p_i}$  for  $i \leq j$  and with  $A_j q \geq \theta_{j+1}$ , taking  $q$  in increasing order and omitting  $q$  in the following cases.

- (i) If  $P_j \Gamma_j(q) \leq \mu$ , omit  $q$  and every larger prime.
- (ii) If  $q \leq q_c$  and  $P_j \Gamma_j(q) \leq \mu'(S)$ , omit every prime in  $[q, q_c]$ .
- (iii) If  $q > q_c$  and  $V_{j+1}(\mu)$ , computed for the sequence  $(p_1, \dots, p_j, q)$ , is less than  $\theta_{j+2}$ , omit  $q$ .

A prime  $q$  satisfies  $q \leq q_c$  exactly when  $P_j q / (q - 1) \geq \mu$ , so in (ii) a solution extending  $(p_1, \dots, p_j, q)$  has  $P_{j+1} \geq \mu$  and hence  $M \geq \mu'(S)$ . For  $q > q_c$  we have  $P_{j+1} < \mu$ , so  $V_{j+1}(\mu)$  is defined, and after multiplication by its positive denominator the inequality in (iii) becomes a quadratic inequality in  $q$  with positive leading coefficient;<sup>10</sup> the primes omitted by (iii) therefore form an interval, which is computed once per node.

<sup>10</sup>With  $D = \mu B_j - A_j$  and  $\theta = \theta_{j+2}$  one has  $V_{j+1}(\mu) = \mu B_j(q - 1)(A_j q + 1)/(Dq - \mu B_j)$ , and  $V_{j+1}(\mu) < \theta$  reads  $\mu A_j B_j q^2 + (\mu B_j - \mu A_j B_j - \theta D)q + (\theta - 1)\mu B_j < 0$ .

**Proposition 6.14.** *Let  $n$  be an  $s$ -heavy solution. Then  $(p_1, \dots, p_{k-1})$  is a node of  $\mathcal{T}_s$ , and rule (R3) at that node finds  $p_k$ .*

*Proof.* Write  $S_j = (p_1, \dots, p_j)$ . We show by induction on  $j$  that  $S_j$  is generated and is a node, for  $0 \leq j \leq k-1$ . The empty sequence  $S_0$  is generated by definition. Suppose that  $S_j$  is generated. As  $n$  extends  $S_j$ , we have  $M \geq \mu(S_j)$ . Lemma 6.13 with  $q = q_1 \leq p_{j+1}$  gives  $P_j \Gamma_j(q_1) \geq P_k > M \geq \mu(S_j)$ , so (R1) does not discard  $S_j$ , and  $V_j(\mu(S_j)) \geq V_j(M) \geq \theta_{j+1}$  by (H2), so (R2) does not either. Thus  $S_j$  is a node.

Now let  $j \leq k-2$ . The prime  $p_{j+1}$  is a candidate of (R4) at  $S_j$ , by Lemma 2.1(ii) and (H1) at  $j+1$ . Lemma 6.13 gives  $P_j \Gamma_j(q) \geq P_k > M \geq \mu(S_j)$  for every integer  $q$  with  $3 \leq q \leq p_{j+1}$ , so rule (i) applies at no prime  $q \leq p_{j+1}$ . If  $p_{j+1} \leq q_c$ , then  $P_{j+1} \geq \mu(S_j)$ , and  $M > P_{j+1}$  forces  $M \geq \mu'(S_j)$ ; hence  $P_j \Gamma_j(q) > \mu'(S_j)$  for  $q \leq p_{j+1}$ , rule (ii) applies at no prime  $q \leq p_{j+1}$ , and an application at a larger prime does not reach  $p_{j+1}$ . If  $p_{j+1} > q_c$ , rule (ii) omits only primes up to  $q_c$ , and  $V_{j+1}(\mu(S_j)) \geq V_{j+1}(M) \geq \theta_{j+2}$  by (H2) at  $j+1$ , so rule (iii) does not omit  $p_{j+1}$ . Hence  $S_{j+1}$  is generated.

At the node  $S_{k-1}$  we have  $\mu(S_{k-1}) \leq M < P_k$  and  $P_k = P_{k-1}p_k/(p_k - 1) \leq P_{k-1}(p_{k-1} + 2)/(p_{k-1} + 1)$ , so  $M$  is one of the values tested in (R3), and  $p_k$  is the quotient given by Lemma 6.1(ii).  $\square$

**6.4. The computation.** All quantities in the rules are rational numbers, and the search evaluates them exactly. The primes  $\pi_i(q)$  below  $2 \cdot 10^7$  are taken from a sieve. Primes above that are reached by stepping to the next probable prime; such a step never skips a prime, so a composite number may occasionally be treated as a prime, which can only enlarge the tree. It is equally harmless to replace  $q_1$  in (R1) by the smaller number  $p_j + 2$ , as one of the two programs does beyond the sieve, because the proof of Proposition 6.14 uses only Lemma 6.13, which holds for every integer  $q$  with  $3 \leq q \leq p_{j+1}$ . In rule (R3) the quotient was never an integer, so no primality proof is needed.

TABLE 5. The tree  $\mathcal{T}_7$  by depth  $j$ : the number of nodes and the number of values of  $M$  tested by rule (R3).

| $j$   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7  | 8  | 9  | 10  | 11   | 12    | 13   |
|-------|---|---|---|---|---|---|---|----|----|----|-----|------|-------|------|
| nodes | 1 | 1 | 1 | 3 | 6 | 7 | 7 | 13 | 31 | 79 | 266 | 1501 | 29073 | 2689 |
| tests | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 16  | 215  | 11894 | 2689 |

**Theorem 6.15.** *The tree  $\mathcal{T}_7$  has 33678 nodes and depth 13. Rule (R3) tests 14814 values of  $M$ , and for none of them is the quotient  $(MB_j - 1)/(MB_j - A_j)$  an integer. Consequently, by Proposition 6.14, no composite solution of (1.1) is 7-heavy.*

Table 5 lists the nodes by depth. Every node has  $\mu(S) = 2$ , so every node has  $p_1 \neq 3$  and  $P_j < 2$ ; in particular the sequence (3), for which  $\mu = 4$ , is discarded at once by the ratio test. The first branch of the tree is 5, 7, 13, 17, 19, 23, 37, 59, 67, 73, the beginning of the independent set of Section 2.1 chosen one prime at a time, and the deepest nodes have thirteen entries.

*Proof of Theorem 1.2(i).* By Theorem 6.15 no composite solution of (1.1) is 7-heavy, and Corollary 6.10 with  $s = 7$  gives  $n < 2^{2^{k-7}}$ .  $\square$

For  $s = 1, \dots, 7$  the trees  $\mathcal{T}_s$  have 1, 3, 7, 12, 24, 118 and 33678 nodes. The deficit test does most of the work: without rules (R2) and (iii), the tree  $\mathcal{T}_6$  already has 4469 nodes and depth 12.

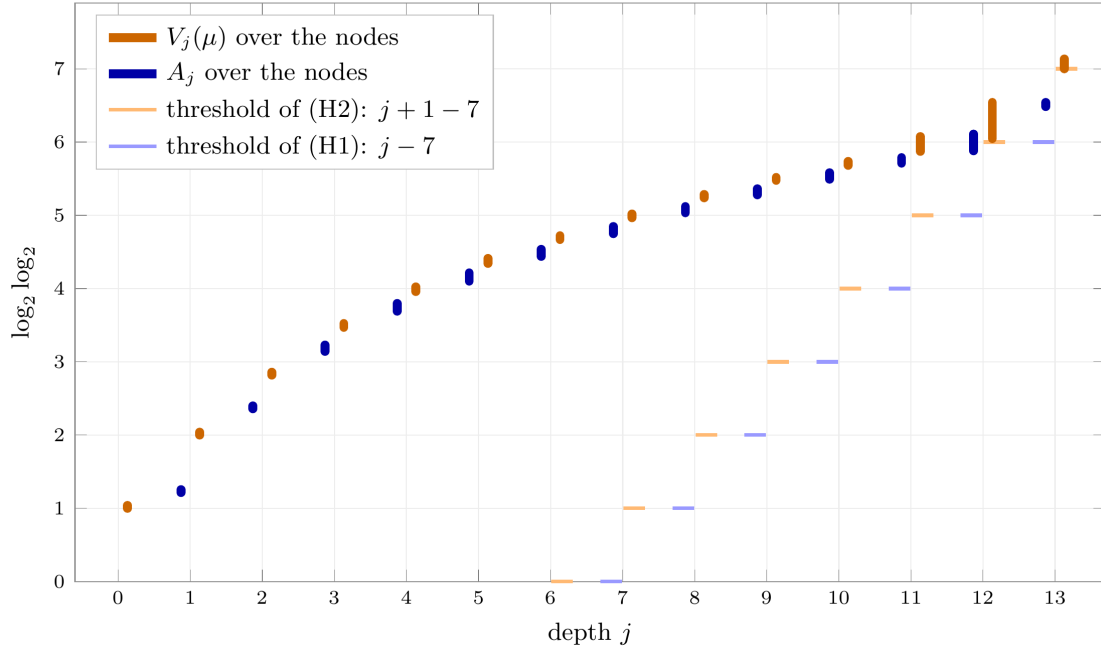


FIGURE 13. The nodes of  $\mathcal{T}_7$  by depth  $j$ . The bars span the range of  $\log_2 \log_2 V_j(\mu)$  (right of each depth) and of  $\log_2 \log_2 A_j$  (left of each depth) over the nodes of depth  $j$ , and the short marks below them are the thresholds  $j + 1 - 7$  of (H2) and  $j - 7$  of (H1). The deficit condition (H2) is the one that binds: at depth 13 the smallest value of  $\log_2 \log_2 V_{13}$  is 7.0000075...

**6.5. Larger quotients.** When  $M \geq 3$ , condition (H1) for a suitable  $s$  is incompatible with Lemma 6.1(i) and the bounds of Section 2.1, and no further search is needed. Let  $\ell_1 < \ell_2 < \dots$  be integers with  $p_i \geq \ell_i$  for  $i \leq k$ , let  $J \geq 1$ , and let  $V^*$  be a number with  $P_{\min\{k, J\}} < V^*$ . If (H1) holds for all  $j \leq k - 1$ , then  $A_i \leq p_i^i$  gives  $p_i \geq \rho_i$  for  $i \leq k - 1$ , where  $\rho_i$  is the least positive integer  $x$  with  $x^i \geq \theta_i$ . With  $w_i = \max\{\ell_i, \rho_i\}$  we obtain, as in the proof of Lemma 6.13,

$$P_k < R_s = V^* \max\left\{1, \max_{J \leq i < i_0} \frac{w'_i + 2}{w'_i + 1} \prod_{l=J+1}^i \frac{w_l}{w_l - 1}, \left(1 + \frac{16}{\rho_{i_0} - 2}\right) \prod_{l=J+1}^{i_0-1} \frac{w_l}{w_l - 1}\right\}, \quad (6.6)$$

where  $w'_i = \max\{w_i, \ell_{i+1} - 2\}$  accounts for  $p_{i+1} \geq \max\{\ell_{i+1}, w_i + 2\}$ , and  $i_0 \geq \max\{J + 1, s + 4\}$  is an index with  $\rho_{i_0} > 10^{30}$ . The three terms cover  $k \leq J$ , then  $J < k \leq i_0$  with  $k = i + 1$ , and  $k > i_0$ . The tail estimate is that of Lemma 6.13, applied to  $y_i = \theta_i^{1/i}$ , for which  $p_i \geq \rho_i \geq y_i > \rho_i - 1$  when  $\rho_i > 1$ , and  $\log y_{i+1} = \frac{2i}{i+1} \log y_i \geq \frac{4}{3} \log y_i$  when  $i \geq \max\{s, 2\}$ . We evaluate  $R_s$  in exact rational arithmetic, with integer lower bounds in place of the  $\rho_i$ , which can only enlarge it.

*Proof of Theorem 1.2(ii) and (iii).* Suppose  $3 \mid n$ . Then  $M \geq 4$ , every prime factor other than 3 is congruent to 2 modulo 3, by Lemma 2.2, and  $k \geq 10^8 + 2$  by Corollary 2.6(ii). Take  $J = 10^8 + 1$ ,  $\ell_i = 2i + 1$ , which is at most the  $i$ -th odd prime, and  $V^* = 4(1 - 10^{-6})$ . The  $J$  smallest prime factors are 3 and an independent set of  $10^8$  primes  $p \equiv 2 \pmod{3}$ , so  $P_J < \frac{3}{2} \cdot \frac{8}{3}(1 - 10^{-6}) = V^*$  by Proposition 2.5(ii). The computation gives  $R_{10^8} = 3.9999966 \dots < 4$ . If (H1) held for all  $j \leq k - 1$  with  $s = 10^8$ , then (6.6) would give  $P_k < 4 \leq M$ , contrary to Lemma 6.1(i). So  $n$  is not  $10^8$ -heavy, and Corollary 6.10 gives (iii).

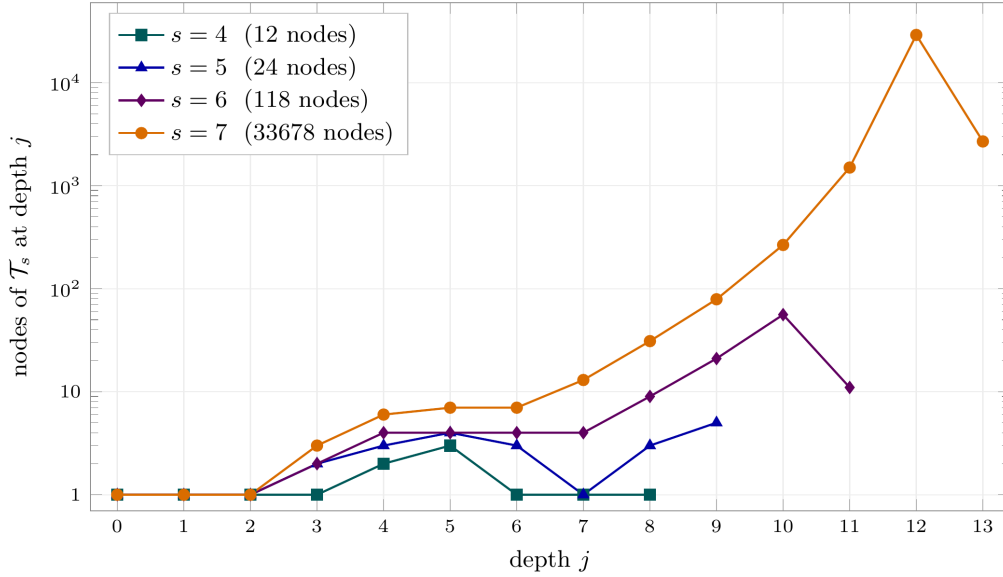


FIGURE 14. The number of nodes of  $\mathcal{T}_s$  at each depth, for  $s = 4, 5, 6, 7$ , on a logarithmic scale.

Suppose  $M \geq 3$  and  $3 \nmid n$ . Then  $k \geq 16001$  by Corollary 2.6(i). Take  $J = 16000$ ,  $\ell_i = r_i$ , the primes from 5 on, and  $V^* = 2.999999$ ; the  $J$  smallest prime factors form an independent set, so  $P_J < V^*$  by Proposition 2.5(i). Then  $R_{15981} = 2.99999900041 \dots < 3$ , and the same argument gives  $n < 2^{k-15981}$ . If  $M \geq 3$  and  $3 \mid n$ , part (iii) gives a stronger bound.  $\square$

The exponent 15981 is the largest that this argument gives with  $J = 16000$ , since  $R_{15982} = 3.000016 \dots$ . Everything is decided by the prime factors around the 16000-th. For  $s = 15981$ , (H1) at  $j = 16000$  gives  $p_{16000} \geq \rho_{16000} > 7 \cdot 10^9$ , so that the remaining prime factors raise  $P_{16000} < 2.999999$  by a factor less than  $1 + 10^{-9}$ . For  $s = 15982$  the corresponding bound  $\rho_{16000} < 10^5$  lies below  $r_{16000} = 176089$ , and the next prime factor alone may raise the product by the factor  $1 + 1/176122$ .

**6.6. Where the search is hard.** The nodes that survive longest are prefixes with  $P_j$  just below 2, the same kind of prefix that makes the next cases expensive (Section 9). Every node at depth 13 has  $2 - P_{13} < 10^{-10}$ : there  $A_{13}$  has between 27 and 29 digits, while  $C_{13} = 2B_{13} - A_{13}$  has between 14 and 18. In this respect the deep nodes behave like the products of Fermat primes of Remark 6.7, for which  $2\varphi(n) - n = 1$ . For such a prefix  $1 - P_j/2$  is tiny, so  $V_j$  is large and the deficit test cannot discard it. Its children are few for the same reason: a prime  $q \leq q_c = \lfloor 1/(1 - P_j/2) \rfloor$ , which exceeds  $3 \cdot 10^{10}$  at depth 13, would raise  $P_{j+1}$  to 2 or more and force  $M \geq 3$ , and rule (ii) removes these primes. Figure 13 shows how close to the threshold of (H2) the deepest nodes come.

The tree grows quickly with  $s$  (Figure 14), and for  $s = 7$  more than nine tenths of its nodes lie at depths 12 and 13. Each further unit of  $s$  halves the number of binary digits allowed by Theorem 1.2(i). The method proves  $n < 2^{2^{k-s}}$  for every  $s$  for which the tree  $\mathcal{T}_s$  can be searched, and the cost of the search is dominated by the deep prefixes with  $P_j$  close to 2. We have not completed the search for  $s = 8$ : at the node 5, 7, 13, 17, 19, 23, 37, 59, 67, 73, 317, 5483 of depth 12, where  $2/P_{12} - 1 = 7.0 \cdot 10^{-8}$ , rule (R4) already generates more than  $10^6$  sequences.

**6.7. Comparison with other bounds.** Stone [35, (52), (53)] treated tuples of primes with  $\prod(1 - 1/p_i) \leq \frac{1}{2} < \prod_{i < k}(1 - 1/p_i)$ , which include the prime factors of solutions with  $M = 2$ . For  $p_1 = 5$  he obtained  $\prod p_i \leq E_k(\beta)$  with  $\beta \approx 1.66127$ , and with  $\beta \approx 1.50188$  when also  $p_2 = 7$ . Theorem 1.2(i) gives the base  $2^{1/128} = 1.00543 \dots$  for solutions of (1.1). Table 6 compares, along the first branch of  $\mathcal{T}_7$ , the exponents  $e$  for which the three bases give  $n < 2^{2^{k-e}}$ . The deficit gains about three quarters of a unit of  $e$  over Lemma 6.5, and Theorem 6.2 adds a further gain over Stone's form that is largest for short prefixes.

TABLE 6. Along the branch 5, 7, 13, ... of  $\mathcal{T}_7$  with  $M = 2$ : the exponent  $e$  such that the bound for  $n$  obtained at depth  $j$  is  $2^{2^{k-e}}$ , for the base  $(A_j + 1)^2$  of Lemma 6.5, the base  $b^2/d$  of Stone and the base  $V_j$  of Theorem 6.2.

| $j$ | $p_j$ | $P_j$   | Cook–Nielsen | Stone  | Theorem 6.2 |
|-----|-------|---------|--------------|--------|-------------|
| 1   | 5     | 1.25000 | −0.370       | −0.142 | 0.000       |
| 2   | 7     | 1.45833 | −0.370       | 0.099  | 0.181       |
| 3   | 13    | 1.57986 | −0.143       | 0.486  | 0.530       |
| 4   | 17    | 1.67860 | 0.309        | 1.017  | 1.041       |
| 5   | 19    | 1.77186 | 0.899        | 1.644  | 1.657       |
| 6   | 23    | 1.85240 | 1.561        | 2.324  | 2.330       |
| 7   | 37    | 1.90385 | 2.251        | 3.030  | 3.033       |
| 8   | 59    | 1.93668 | 2.965        | 3.759  | 3.761       |
| 9   | 67    | 1.96602 | 3.720        | 4.516  | 4.517       |
| 10  | 73    | 1.99333 | 4.507        | 5.265  | 5.265       |

## 7. THE COMPANION EQUATION

**7.1. Up to seven prime factors.** Let  $n$  satisfy (1.2) with  $\omega(n) \leq 7$ . If  $n \leq 3$  there is nothing to prove. Otherwise  $n$  is odd, squarefree and composite by Lemma 2.1, and  $n + 1 = 2\varphi(n)$  by Corollary 2.4. Since 3 may divide  $n$ , the search now starts from  $p_{\min} = 3$ , and we run it with  $\varepsilon = +1$  for  $2 \leq k \leq 7$ . The outcome is in Table 7.

TABLE 7. The search for  $n + 1 = 2\varphi(n)$  with  $3 \leq p_1$  and  $k \leq 7$ .

| $k$ | nodes | terminal | largest width for $p_{k-1}$ | solutions found      |
|-----|-------|----------|-----------------------------|----------------------|
| 2   | 1     | 1        | 1                           | 15                   |
| 3   | 3     | 2        | 2                           | 255                  |
| 4   | 5     | 2        | 15                          | 65535                |
| 5   | 11    | 4        | 256                         | 83623935, 4294967295 |
| 6   | 60    | 44       | 65 536                      | 6992962672132095     |
| 7   | 6 914 | 6 826    | $2^{32}$                    | none                 |

For  $k \leq 6$  the search returns exactly the known solutions, each at its own depth, and this is also a test of the programs. For  $k = 7$  it returns nothing. Here the enumeration route is too slow, because one terminal node, the prefix 3, 5, 17, 257, 65537, has an interval for  $p_6$  of width  $2^{32}$ ; for this prefix  $C = 1$  and  $D = 2AB + C = 2^{64} - 2^{32} + 1$ , which is prime<sup>11</sup>. We therefore

<sup>11</sup>Indeed  $A = 2^{32} - 1$ ,  $B = 2^{31}$  and  $C = 2B - A = 1$ , so  $D = 2^{32}(2^{32} - 1) + 1 = \Phi_6(2^{32})$ , the sixth cyclotomic polynomial at  $2^{32}$ . Its primality is well known and is confirmed by a deterministic test, since  $D < 3.3 \cdot 10^{24}$ .

use the divisor route throughout, which is rigorous in this range: the largest integer  $D$  factored in the case  $k \leq 7$  has 21 digits, so every factorisation is certified. For  $k \leq 6$  the enumeration route was also run and agrees node for node. This proves Theorem 1.3 for solutions with at most seven prime factors; the case of eight is treated in Section 7.4.

**7.2. Solutions prime to three.** Let  $n > 3$  satisfy (1.2) with  $3 \nmid n$  and  $\omega(n) \leq 15$ . Then  $n = p_1 \cdots p_k$  with  $5 \leq p_1$ , and  $n + 1 = 2\varphi(n)$  by Lemma 2.3. For  $k \leq 6$  the product  $\prod p_i/(p_i - 1)$  is at most  $1.9490\dots$ , whereas (3.2) requires it to be  $2 - 1/\varphi(n)$ . For  $k \geq 2$  this is at least  $2 - \frac{1}{24} > 1.95$ , as  $\varphi(n) \geq 4 \cdot 6$ , and for  $k = 1$  it is at least  $\frac{7}{4}$ , whereas  $p_1/(p_1 - 1) \leq \frac{5}{4}$ . For  $7 \leq k \leq 14$  we ran both implementations with  $\varepsilon = +1$  and  $p_{\min} = 5$ . Neither found a solution, and the node counts, the numbers of terminal nodes and the largest widths coincide with those of Table 2 for every  $k$ . For  $k = 15$  the nodes at depth 12 and their intervals are again those of Section 5.2, and the method of Section 4, run with  $\varepsilon = +1$  by both implementations and by the route of Section 4.4 over the same 33865004 pairs, finds no completion in integers. This proves Theorem 1.4.

The coincidence with Table 2 has a simple cause. The two sets of bounds in (3.3) differ only by the terms  $1/(A_j W)$ , which in this range are far smaller than the gaps between consecutive admissible values of  $p/(p - 1)$  near the endpoints, and the computation confirms that no endpoint moves. The tests (3.4) at the last level do differ, but by then the tree is fixed. For  $3 \nmid n$  and  $k \leq 15$  the two equations therefore lead to the same search.

### 7.3. Extensions of Fermat-type solutions.

*Proof of Theorem 1.6.* Let  $2\varphi(n_0) = n_0 + 1$  and let  $Q > 1$  be squarefree and coprime to  $n_0$ . Then  $n = n_0 Q$  satisfies  $2\varphi(n) = n + 1$  if and only if  $2\varphi(n_0)\varphi(Q) = n_0 Q + 1$ , that is,  $(n_0 + 1)\varphi(Q) = n_0 Q + 1$ , or

$$n_0(Q - \varphi(Q)) = \varphi(Q) - 1. \quad (7.1)$$

(i) For  $Q = q$  prime, (7.1) reads  $n_0 = q - 2$ .

(ii) For  $Q = pq$  with primes  $p < q$  we have  $Q - \varphi(Q) = p + q - 1$  and  $\varphi(Q) - 1 = pq - p - q$ , so (7.1) becomes  $pq - (n_0 + 1)(p + q) + n_0 = 0$ , and adding  $(n_0 + 1)^2 - n_0$  to both sides gives

$$(p - n_0 - 1)(q - n_0 - 1) = n_0^2 + n_0 + 1.$$

The two factors on the left have the same sign, and they are positive: if both were negative we would have  $2 \leq p < q \leq n_0$ , and the left-hand side would be at most  $(n_0 - 1)^2 < n_0^2 + n_0 + 1$ . Conversely, a factorisation  $n_0^2 + n_0 + 1 = de$  with  $d < e$  gives  $p = n_0 + 1 + d$  and  $q = n_0 + 1 + e$ , and if both are prime then  $n_0 pq$  is a Fermat-type solution; the conditions  $p, q \nmid n_0$  hold because  $p, q > n_0$ . The divisor  $d = 1$  gives  $p = n_0 + 2$  and  $q = (n_0 + 1)^2 + 1 = n_0(n_0 + 2) + 2$ , which is the composition of two one-prime extensions.

(iii) Factoring  $n_0^2 + n_0 + 1$  and testing  $n_0 + 2$  for each of the eight numbers gives Table 8. The one-prime extensions are  $1 \rightarrow 3 \rightarrow 15 \rightarrow 255 \rightarrow 65535 \rightarrow 4294967295$  and  $83623935 \rightarrow 6992962672132095$ . The two-prime extensions are those listed in the table, and the only one that is not a composition of one-prime extensions is  $255 \rightarrow 83623935$ . Every extension of each of the eight numbers is therefore again one of the eight, and each of them is reached from 1, so the eight numbers are exactly the closure of  $\{1\}$ . The rows for 65535, 83623935, 4294967295 and 6992962672132095 show where the chains stop. Every factor in Table 8 is below  $2^{64}$ , where the primality tests we use are proofs, and a test that declares a candidate  $p$  or  $q$  composite is never wrong.  $\square$

The chains stop for the following reasons. For  $n_0 = 65535$  the trivial divisor gives  $p = 65537$  and  $q = 2^{32} + 1 = 641 \cdot 6700417$ , which is Euler's factorisation of  $F_5$ , and the divisor 193 gives  $p = 65729$ , a prime, but  $q = 22318913 = 3037 \cdot 7349$ . For  $n_0 = 4294967295$  the one-prime

TABLE 8. One- and two-prime extensions of the Fermat-type solutions.

| $n_0$            | $n_0 + 2$ prime | factorisation of $n_0^2 + n_0 + 1$ ; two-prime extensions $(p, q)$       |
|------------------|-----------------|--------------------------------------------------------------------------|
| 1                | yes             | 3; (3, 5)                                                                |
| 3                | yes             | 13; (5, 17)                                                              |
| 15               | yes             | 241; (17, 257)                                                           |
| 255              | yes             | $97 \cdot 673$ ; (257, 65537) and (353, 929)                             |
| 65535            | yes             | $193 \cdot 22253377$ ; none                                              |
| 83623935         | yes             | $67 \cdot 180097 \cdot 579535339$ ; none                                 |
| 4294967295       | no              | $2^{64} - 2^{32} + 1$ , prime; none                                      |
| 6992962672132095 | no              | $13 \cdot 37 \cdot 109 \cdot 30802180789 \cdot 30280941915943441$ ; none |

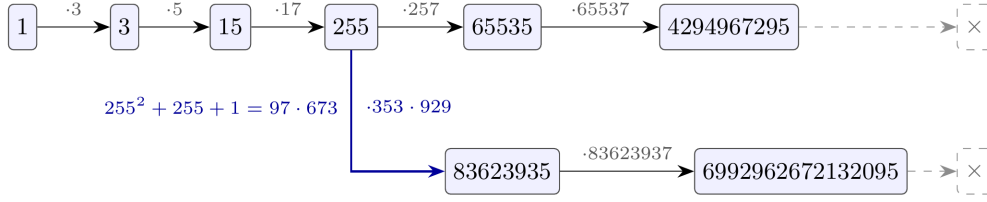


FIGURE 15. The Fermat-type solutions of  $n + 1 = 2\varphi(n)$  as the closure of 1 under extensions. Horizontal arrows are one-prime extensions  $n_0 \mapsto n_0(n_0 + 2)$ , labelled by the prime  $n_0 + 2$ . The bent arrow is the two-prime extension of 255 by (353, 929), coming from the factorisation  $65281 = 97 \cdot 673$ . Dashed arrows end the two chains: 4294967297 and 6992962672132097 are composite, and neither  $n_0^2 + n_0 + 1$  has a factorisation that yields two primes.

extension fails because  $n_0 + 2 = 2^{32} + 1$  is composite once more. Figure 16 shows the hyperbolas for  $n_0 = 255$  and  $n_0 = 65535$  with their lattice points.

*Remark 7.1.* For three new primes,  $Q = spq$  with  $s < p < q$ , write  $x = s - 1$ ,  $y = p - 1$ ,  $z = q - 1$ . Equation (7.1) becomes

$$xyz = n_0(xy + yz + zx) + n_0(x + y + z) + n_0 + 1,$$

and dividing by  $n_0xyz$  shows that  $1/x + 1/y + 1/z$  falls short of  $1/n_0$  by  $(n_0(x + y + z) + n_0 + 1)/(n_0xyz)$ , which is less than  $4/x^2$  because  $x < y < z$ . Hence  $x$ , the smallest of the three, satisfies  $3/x > 1/n_0 - 4/x^2$  and  $1/x < 1/n_0$ , that is,  $n_0 < x < 3n_0 + 2$ . For fixed  $x$  the equation is bilinear in  $y$  and  $z$ :

$$((x - n_0)y - n_0(x + 1))((x - n_0)z - n_0(x + 1)) = (x - n_0)(n_0x + n_0 + 1) + n_0^2(x + 1)^2,$$

so the pair  $(y, z)$  comes from the divisors of an integer of size about  $n_0^2x^2$ . For  $n_0 = 4294967295$  there are about  $3.8 \cdot 10^8$  primes  $s$  in the range, and  $1.3 \cdot 10^8$  of them pass the congruence prune:  $s \equiv 2 \pmod{3}$  as Lemma 2.2 requires, and  $s \not\equiv 1$  modulo 5, 17, 257 and 65537. Each of them leaves an integer of about forty digits to factor. This is the prefix 3, 5, 17, 257, 65537 that dominates the case  $k = 8$  in Section 7.4.

**7.4. Eight prime factors.** Let  $n$  satisfy (1.2) with  $\omega(n) = 8$ . Then  $n > 3$ , and  $n$  is odd and squarefree by Lemma 2.1. If  $3 \nmid n$ , then  $\omega(n) \geq 16$  by Theorem 1.4; hence  $3 \mid n$ . By Lemma 2.2 the quotient  $M$  is either 2 or at least 5, and in the second case  $\omega(n) \geq 1540$ . Hence  $n + 1 = 2\varphi(n)$  and  $p_1 = 3$ , and it suffices to run the search with  $\varepsilon = +1$  and  $k = 8$  over the prefixes that begin with 3.

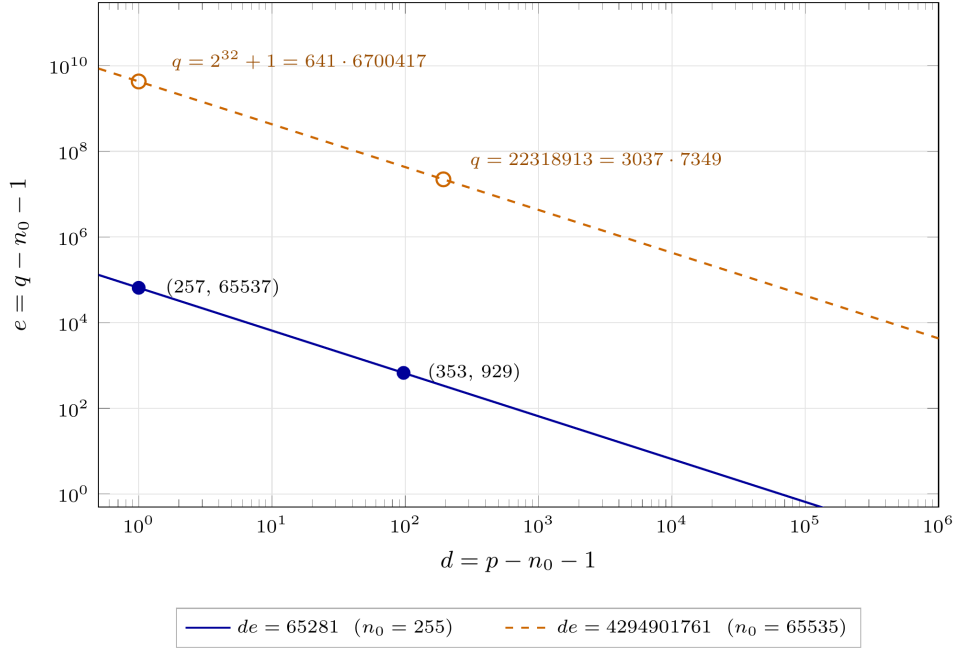


FIGURE 16. The two-prime extensions of  $n_0$  are the lattice points  $(d, e) = (p - n_0 - 1, q - n_0 - 1)$  on  $de = n_0^2 + n_0 + 1$  with both  $p$  and  $q$  prime, shown on logarithmic axes. For  $n_0 = 255$  both lattice points with  $d \leq e$  give primes (filled). For  $n_0 = 65535$  both lattice points give a prime  $p$  and a composite  $q$  (open).

The traversal down to depth 5 takes a few seconds. It yields 10459 prefixes  $(p_1, \dots, p_5)$  with a non-empty interval for  $p_6$ , and 10458 of them begin with 3. Their intervals have total length  $1.3045 \cdot 10^{10}$ , and most of it belongs to a few prefixes (Figure 18, left):  $8.59 \cdot 10^9$  to 3, 5, 17, 257, 65537, that is, to the solution 4294967295 itself, where  $C_5 = 1$ ;  $1.23 \cdot 10^9$  to 3, 5, 17, 257, 65543; and  $1.67 \cdot 10^8$  to 3, 5, 17, 353, 929, the solution 83623935, where again  $C_5 = 1$ . Thirty-four prefixes have an interval longer than  $10^7$ , and 326 one longer than  $10^6$ . The intervals contain 228260561 primes  $s$  that pass the congruence prune, 131788942 of them for the prefix 3, 5, 17, 257, 65537. Each pair of a prefix and such a prime  $s = p_6$  leaves the two-prime problem of Proposition 4.1, in which  $N$  has between 15 and 40 digits.

These problems lie far to the left of those of Figure 9. For the prefix 3, 5, 17, 257, 65537 and  $p_6 = s$  one has  $A' = (2^{32} - 1)s$  and  $B' = 2^{31}(s - 1)$ , so that the modulus is  $c = s - 2^{32}$ , and  $N = 2A'B' + c$  is close to  $2^{64}s^2$ . As  $s$  runs over  $[2^{32} + 1, 3 \cdot 2^{32} - 1]$ , the ratio  $c^3/N$  stays below  $2.1 \cdot 10^{-10}$ , and the number  $(N/c^3)^{1/2}$  that governs the methods of Section 4 grows from  $7 \cdot 10^4$  at the upper end to  $4 \cdot 10^{16}$  at the first admissible prime,  $s = 4294967357$ , where  $c = 61$ . We therefore combine the sum route with trial division and with a complete factorisation, as follows.

Fix a prefix and an admissible prime  $s$ , define  $A'$ ,  $B'$ ,  $c$ ,  $N$  as in Section 4, and let  $t = cp - 2B'$  be the divisor of Proposition 4.1. Since  $p \geq s + 2$ , it satisfies  $t \geq t_0 = c(s + 2) - 2B'$ , and we choose a split point  $t_d$  with  $t_0 \leq t_d$ .

- (i) The divisors with  $t < t_d$  are found by running  $p$  over the integers with  $t_0 \leq cp - 2B' < t_d$  and testing whether  $t$  divides  $A'p + \varepsilon$ , as in Proposition 3.5. Since  $c(A'p + \varepsilon) = A't + N$  and  $\gcd(t, c) = 1$  by Proposition 4.1, this is the same as testing whether  $t$  divides  $N$ , and the program does that when  $A'p$  is too large for 128-bit arithmetic.

- (ii) For the divisors with  $t_d \leq t \leq \sqrt{N}$  we use the sum  $\sigma = p + q$  and the product  $\pi = pq$ . By (3.5),

$$c\pi = 2B'\sigma - 2B' + \varepsilon, \quad (q - p)^2 = \sigma^2 - 4\pi, \quad (7.2)$$

so  $\sigma$  lies in the class of  $(2B' - \varepsilon)(2B')^{-1}$  modulo  $c$ , and  $t + N/t = c\sigma - 4B'$ . As  $x + N/x$  decreases on  $(0, \sqrt{N}]$ , the sum runs over its class in the range  $2\sqrt{N} + 4B' \leq c\sigma \leq t_d + N/t_d + 4B'$ , and each value is tested by asking whether  $\sigma^2 - 4\pi$  is a perfect square. This is the route of Section 4.4, written in terms of  $p$  and  $q$  instead of  $u$  and  $v$ : the two sums differ by the constant  $(2r + 4B')/c$ .

- (iii) Both loops are sieved by congruences. Let  $p < q$  be integers with  $A'pq + 1 = 2B'(p-1)(q-1)$ . No  $p_i$  divides  $p-1$  or  $q-1$ , since it would then divide 1. If moreover  $p$  and  $q$  have no prime factor up to 61, which holds for primes larger than  $s$  because every admissible  $s$  exceeds 61, then modulo every prime  $\ell \leq 61$  they avoid the residue 0, and also the residue 1 when  $\ell$  is one of the  $p_i$ ; in particular they are odd and congruent to 2 modulo 3. Call these residues excluded. Since  $X^2 - \sigma X + \pi = (X-p)(X-q)$ , the roots of this polynomial modulo a prime  $\ell$  are  $p$  and  $q$  modulo  $\ell$ . The sieve discards only values of  $\sigma$  that are odd, or not congruent to 1 modulo 3, or for which  $\sigma^2 - 4\pi$  is not a square modulo one of  $2^8$ ,  $3^2$ ,  $5^2$  and  $7^2$ , or for which, for some prime  $\ell \leq 43$ , the polynomial has no root modulo  $\ell$  or a root at an excluded residue; and in the trial division it discards only values of  $p$  at an excluded residue modulo some prime  $\ell \leq 61$ . Every other value is tested exactly. So the sieve keeps every value that comes from integers  $p$  and  $q$  without prime factors up to 61, and in particular every value that comes from primes.
- (iv) If the estimated work of (i) and (ii) exceeds four milliseconds, the time we allow for factoring  $N$ , then  $N$  is factored completely with PARI/GP [29], every prime factor is proved prime,<sup>12</sup> and all divisors  $t$  of  $N$  with  $t_0 \leq t \leq \sqrt{N}$  in the class of Proposition 4.1 are listed.

Since  $c$  is odd and prime to 3, the admissible values of  $\sigma$  form one class modulo  $6c$ . The work of (i) and (ii) is therefore about  $(t_d - t_0)/c$  values of  $p$  and  $(t_d + N/t_d - 2\sqrt{N})/(6c^2)$  values of  $\sigma$ , and  $t_d$  is chosen to minimise it (Figure 17), with the two kinds of steps weighted by their measured costs. With equal weights the minimum lies near  $t_d = (N/6c)^{1/2}$  and amounts to about  $2(N/6c^3)^{1/2}$  steps, as in Section 4.4. By the argument of that section, (i) and (ii) together find every divisor  $t \geq t_0$  of  $N$  in the class up to  $\sqrt{N}$ , and so does (iv); the sieve (iii) loses none that comes from integers  $p$  and  $q$  without prime factors up to 61. Where  $\sigma$  and  $\pi$  are too large for 128-bit arithmetic, (ii) is carried out in multiprecision arithmetic. Now let  $n$  be a solution with  $p_1 = 3$  and  $k = 8$ . By Theorem 3.6 its first five primes form one of the prefixes and  $p_6$  is one of the admissible primes  $s$  of its interval, and by Proposition 4.1 its last two primes give a divisor  $t \geq t_0$  in the class. Hence  $n$  is found. Every completion  $s < p < q$  in integers that is found is checked against  $A'pq + 1 = 2B'(p-1)(q-1)$  exactly, and  $p$  and  $q$  are tested for primality.

The computation examined all 228260561 pairs; in 1967265 of them  $N$  was factored, with up to 39 digits. The computation was cut into 10897 pieces, each a range of  $s$  under one prefix, and run on four processors; it made  $3.4 \cdot 10^{11}$  trial divisions, the sums ranged over  $6.7 \cdot 10^{13}$  values of  $\sigma$ , the sieve (iii) passed  $2.1 \cdot 10^8$  of them on to the exact test, and the whole run took 6.9 hours of processor time, 3.2 of them in PARI/GP. It found 21 completions  $s < p < q$  in

<sup>12</sup>With the option `factor_proven` set, PARI/GP certifies every prime factor it returns. A factor below  $2^{64}$  is certified by its Baillie–PSW test, which no composite number below  $2^{64}$  passes; the PARI/GP documentation records that this was checked against the table of base-2 pseudoprimes below  $2^{64}$  of Feitsma and Galway [12]. A larger factor is certified by the  $n-1$  method [5] when enough of  $n-1$  is known, and otherwise by the test of Adleman, Pomerance and Rumely [1] in the form of Cohen and Lenstra [9].

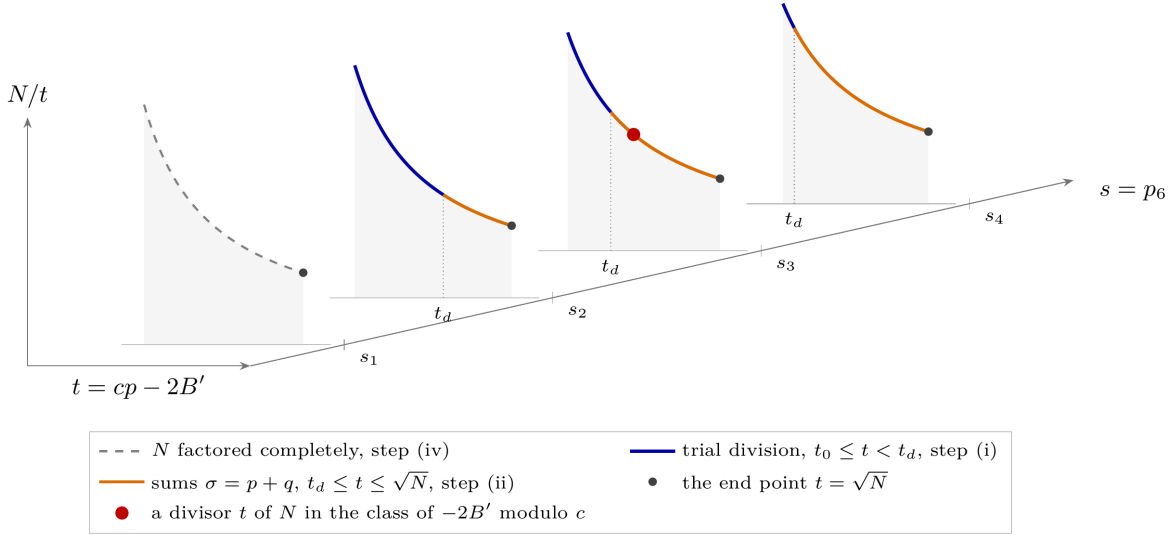


FIGURE 17. Steps (i), (ii) and (iv) for four admissible values  $s_1 < s_2 < s_3 < s_4$  of  $p_6$ , not to scale. The divisors  $t = cp - 2B'$  of  $N$  with  $t_0 \leq t \leq \sqrt{N}$  lie on the branch  $t \mapsto N/t$ . Trial division covers  $t < t_d$  and the sums cover  $t_d \leq t \leq \sqrt{N}$ . The split point  $t_d$  is chosen for each  $s$  to minimise the work, and where both routes would be slow, as for  $s_1$ , the integer  $N$  is factored completely.

integers and no solution. Sixteen of the completions lie under the prefix 3, 5, 17, 353, 929 with  $s = 83623937$ , where  $3 \cdot 5 \cdot 17 \cdot 353 \cdot 929 \cdot s$  is the solution  $n_0 = 6992962672132095$ . They are the pairs  $p = n_0 + 1 + d$ ,  $q = n_0 + 1 + e$  of Theorem 1.6(ii), one for each factorisation  $n_0^2 + n_0 + 1 = de$  with  $d < e$ , and by Table 8 none of them consists of two primes. In each of the other five, listed in Table 9,  $p$  or  $q$  is divisible by one of  $p_1, \dots, p_5$ . So (1.2) has no solution with eight prime factors, and together with Section 7.1 this proves Theorem 1.3.

TABLE 9. The completions  $s < p < q$  in integers of the case  $k = 8$  other than the sixteen two-prime extensions of 6992962672132095, with the factorisations of  $p$  and  $q$ .

| $s = p_6$                   | $p$                                        | $q$                                                             |
|-----------------------------|--------------------------------------------|-----------------------------------------------------------------|
| prefix 3, 5, 17, 257, 65537 |                                            |                                                                 |
| 4294967459                  | $17 \cdot 43 \cdot 673 \cdot 230521189693$ | $5 \cdot 409642859 \cdot 26363459459$                           |
| 4294967969                  | $27419312914203719$                        | $5 \cdot 19 \cdot 5011 \cdot 7699 \cdot 21393991009$            |
| 4596195557                  | $65533391633$                              | $17 \cdot 20923232569978145771052877$                           |
| prefix 3, 5, 17, 353, 929   |                                            |                                                                 |
| 83625827                    | $3^2 \cdot 5^3 \cdot 13 \cdot 252862031$   | $3 \cdot 5 \cdot 7 \cdot 13 \cdot 113 \cdot 794656817253868193$ |
| 83760029                    | $3 \cdot 5 \cdot 3431155289$               | $3 \cdot 87389293 \cdot 45294677621$                            |

## 8. CHECKS ON THE COMPUTATIONS

A search that finds nothing proves something only if it would have found what it looks for. Every search of this paper was therefore carried out by at least two programs written independently, and every count reported here is the same for all of them. For Theorems 1.1 and 1.4 one of the two programs never factors an integer, so that its negative answers depend

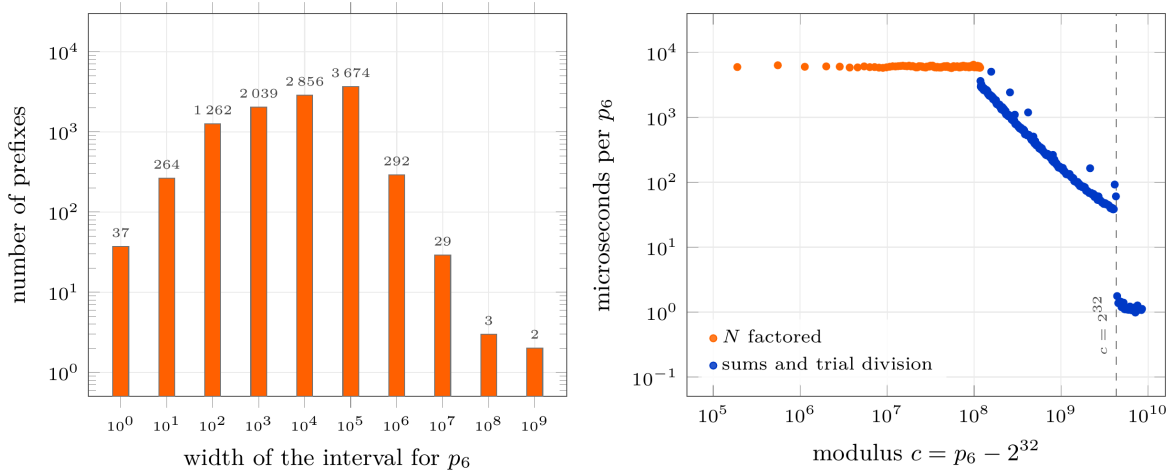


FIGURE 18. The case  $k = 8$  of the companion equation. Left: the 10458 prefixes  $(3, p_2, \dots, p_5)$  by the width of the interval for  $p_6$ , on logarithmic axes; the two in the last bar end in 65537 and 65543. Right: processor time per admissible  $p_6$  under the prefix 3, 5, 17, 257, 65537, one point for each piece of the run, against the modulus  $c = p_6 - 2^{32}$ . For  $c$  below about  $10^8$  the integer  $N$  is factored, and each  $p_6$  costs about six milliseconds in all; above that the sums and the trial division are faster, and their cost falls as  $c$  grows. Beyond  $c = 2^{32}$ , that is for  $p_6 > 2^{33}$ , the condition  $p > p_6$  leaves only a short range of divisors.

on no primality proof. The programs were also run on equations that do have solutions: on the equation  $2^k(n-1) = (2^k + h)\varphi(n)$ , whose 56 solutions with  $k = 4$  and  $k = 5$  in the ranges of Appendix A they all return, and on (1.2), where they return every known solution with at most six prime factors at the right depth and nothing else. The divisor routines of Section 4 were compared with complete factorisations and tested on planted divisors, and the program of Section 7.4 was tested in the same ways and partly run twice. The programs, their recorded output and the details of these tests are in the archive [4].

## 9. THE ASSERTION THAT REMAINS OPEN

The methods of Section 4 settle the case  $k = 15$  of Lehmer's equation in a quarter of an hour on a single processor and the case of eight primes of the companion equation in 6.9 hours. They do not reach the next cases: for (1.1) with sixteen prime factors the number of two-prime problems grows from  $3.4 \cdot 10^7$  to about  $9 \cdot 10^{13}$ , and for (1.2) with nine prime factors and  $3 \mid n$  the Fermat primes alone leave some  $10^{17}$  of them. In both cases most of the cost comes from a few prefixes whose product  $\prod p/(p-1)$  differs from 2 by less than  $10^{-7}$ , the prefixes that Section 6.6 describes. We end with the statement, equivalent to Lehmer's conjecture, that a proof for all  $k$  would have to establish.

**Proposition 9.1.** *Equation (1.1) has no composite solution if and only if there are no finite set  $S$  of odd primes and no integer  $M \geq 2$  such that, with  $A = \prod_{p \in S} p$  and  $B = \prod_{p \in S} (p-1)$ , the number  $D = MB - A$  is positive, divides  $A - 1$ , and makes  $(MB - 1)/D$  a prime larger than every element of  $S$ .*

*Proof.* Let  $n$  be a composite solution of (1.1), let  $q$  be its largest prime factor,  $S$  the set of its other prime factors and  $M = (n-1)/\varphi(n)$ ; by Lemma 2.1 the elements of  $S$  are odd. Then

$n = Aq$ ,  $\varphi(n) = B(q - 1)$ , and  $n - 1 = M\varphi(n)$  reads  $qD = MB - 1$ ; so  $D \geq 1$ ,  $D$  divides  $MB - 1 = D + (A - 1)$  and hence  $A - 1$ , and  $q = (MB - 1)/D$  is a prime larger than every element of  $S$ . Conversely, for  $S$  and  $M$  as in the statement the number  $n = Aq$  with  $q = (MB - 1)/D$  is composite and satisfies  $n - 1 = M\varphi(n)$ , since  $qD = MB - 1$  is the same equation.  $\square$

By Theorem 1.1 a set  $S$  with these properties has at least 15 elements, at least 16000 if  $M \geq 3$ , and at least  $10^8$  if  $3 \in S$ . The assertion asks that one number attached to each set  $S$  never be a prime above the elements of  $S$ . For the companion equation the same construction with  $D = 1$  gives the Fermat numbers: the Fermat primes  $S = \{3, 5, 17, 257, 65537\}$  give  $2^{32} + 1$ . For Lehmer's equation  $D = 1$  would give  $(MB - 1)/D = A$ , which is not a prime larger than the elements of  $S$ ; so  $D > 1$ , and the assertion is not a question about Fermat numbers. Two further constraints on  $D$  follow from the proof. Since  $qD = D + A - 1$ , the divisor is  $D = (A - 1)/(q - 1)$ , so  $q - 1$  divides  $A - 1$  as well and the assertion asks that  $1 + (A - 1)/(MB - A)$  never be a prime larger than the elements of  $S$ ; and since  $2^{|S|}$  divides  $B$  while  $A$  is odd,  $D$  is odd and  $D \equiv -A \pmod{2^{|S|}}$ , so  $D \geq 3$ .

#### APPENDIX A. SOLUTIONS OF THE TEST EQUATION

The following 56 integers  $n = p_1 \cdots p_k$  satisfy  $2^k(n - 1) = (2^k + h)\varphi(n)$  for the indicated  $h$ . They are the complete output of both implementations for  $k = 4$ ,  $1 \leq h \leq 40$ , and  $k = 5$ ,  $1 \leq h \leq 85$ , with  $p_1 \geq 3$ , listed in increasing order of  $n$  down the columns. Each was verified by direct arithmetic.

TABLE 10. Solutions with  $k = 4$ ,  $1 \leq h \leq 40$ .

| $p_1, \dots, p_4$ | $h$ | $p_1, \dots, p_4$    | $h$ |
|-------------------|-----|----------------------|-----|
| 3, 5, 47, 89      | 15  | 13, 29, 433, 2969    | 2   |
| 5, 19, 29, 163    | 6   | 7, 97, 139, 8581     | 3   |
| 7, 17, 229, 251   | 4   | 13, 29, 383, 28879   | 2   |
| 3, 29, 263, 521   | 9   | 11, 47, 1109, 15497  | 2   |
| 13, 37, 151, 271  | 2   | 29, 41, 1291, 15199  | 1   |
| 3, 29, 197, 1559  | 9   | 11, 47, 1049, 77477  | 2   |
| 17, 23, 83, 1709  | 2   | 7, 61, 859, 183397   | 3   |
| 29, 71, 163, 193  | 1   | 31, 37, 2377, 66499  | 1   |
| 41, 43, 97, 491   | 1   | 19, 157, 9697, 15511 | 1   |
| 37, 41, 137, 829  | 1   | 19, 157, 8353, 20887 | 1   |

#### DECLARATIONS

**Data and code availability.** The programs that carry out the searches, the threshold computations and the checks of Section 8, their recorded output, the factorisations of the fifteen integers of Table 3, the logs of the three-prime searches of Section 5.2, the program and the journal of the eight-prime search of Section 7.4, the programs of Sections 2.1 and 6.3 and their output, the programs that check the numbers of Table 1 and of the proof of Proposition 6.11, Lean 4 proofs [25], checked with Mathlib [21], of many statements that do not depend on a search (among them Theorem 6.2 and Proposition 6.8), and the sources of this paper are available at <https://github.com/creelie/eLhmr>, released as version 1.4.0 [4].

**Funding.** No funding was received for conducting this study.

**Declaration of competing interest.** The author has no competing interests to declare that are relevant to the content of this article.

TABLE 11. Solutions with  $k = 5$ ,  $1 \leq h \leq 85$ .

| $p_1, \dots, p_5$          | $h$ | $p_1, \dots, p_5$                        | $h$ |
|----------------------------|-----|------------------------------------------|-----|
| 5, 7, 13, 139, 653         | 19  | 5, 47, 317, 40253, 499027                | 9   |
| 11, 13, 17, 109, 379       | 9   | 19, 269, 443, 1747, 1852699              | 2   |
| 3, 17, 29, 317, 4643       | 21  | 23, 71, 983, 1031, 103437589             | 2   |
| 11, 17, 107, 173, 1373     | 6   | 19, 199, 757, 5059, 23698783             | 2   |
| 5, 19, 37, 73, 19739       | 12  | 31, 43, 233, 318629, 12310321            | 2   |
| 19, 73, 97, 103, 613       | 3   | 19, 313, 389, 1277, 590838019            | 2   |
| 11, 17, 127, 131, 2879     | 6   | 11, 47, 1051, 67967, 102302009           | 4   |
| 5, 7, 269, 317, 4283       | 15  | 7, 59, 1721, 41813, 958695209            | 6   |
| 5, 53, 233, 857, 5689      | 9   | 41, 167, 14411, 322633, 1891541          | 1   |
| 31, 53, 149, 569, 3407     | 2   | 31, 37, 2411, 47681, 5732949913          | 2   |
| 13, 17, 271, 1567, 5573    | 5   | 31, 37, 2297, 2639701, 1379629439        | 2   |
| 7, 13, 17, 1553, 480499    | 11  | 37, 463, 827, 1667389, 4325665973        | 1   |
| 17, 23, 131, 199, 167099   | 4   | 37, 353, 1867, 25109743, 844718249       | 1   |
| 11, 83, 163, 239, 210461   | 4   | 43, 139, 61729, 212167, 32068788511      | 1   |
| 13, 167, 349, 1747, 8237   | 3   | 37, 367, 1553, 21098893, 41563512479     | 1   |
| 19, 181, 1439, 3767, 17683 | 2   | 43, 139, 47857, 57253657, 72147837793    | 1   |
| 3, 11, 269, 17837, 3861929 | 21  | 37, 353, 1867, 24384917, 20504256743083  | 1   |
| 43, 317, 331, 953, 166357  | 1   | 43, 139, 47819, 1145259251, 653813324977 | 1   |

**Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.** During the preparation of this work the author used Claude (Anthropic) to write and run the programs, to write the Lean proofs and to draft text. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## REFERENCES

- [1] L. M. Adleman, C. Pomerance, and R. S. Rumely, *On distinguishing prime numbers from composite numbers*, Ann. of Math. (2) **117** (1983), no. 1, 173–206. doi:10.2307/2006975
- [2] R. Baillie and S. S. Wagstaff, Jr., *Lucas pseudoprimes*, Math. Comp. **35** (1980), no. 152, 1391–1417. doi:10.1090/S0025-5718-1980-0583518-6
- [3] W. D. Banks, C. W. Nevans, and C. Pomerance, *A remark on Giuga's conjecture and Lehmer's totient problem*, Albanian J. Math. **3** (2009), no. 2, 81–85. doi:10.51286/albjm/1245207560
- [4] D. Bhattacharjee, *Lehmer's totient problem with fewer than sixteen prime factors: programs, data and Lean proofs*, version 1.4.0, 2026. <https://github.com/creelie/eLhmr/releases/tag/v1.4.0>
- [5] J. Brillhart, D. H. Lehmer, and J. L. Selfridge, *New primality criteria and factorizations of  $2^m \pm 1$* , Math. Comp. **29** (1975), no. 130, 620–647. doi:10.1090/S0025-5718-1975-0384673-1
- [6] P. Burcsi, S. Czirbusz, and G. Farkas, *Computational investigation of Lehmer's totient problem*, Ann. Univ. Sci. Budapest. Sect. Comput. **35** (2011), 43–49. doi:10.71352/ac.35.043
- [7] D. Burek and B. Žmija, *A new upper bound for numbers with the Lehmer property and its application to repunit numbers*, Int. J. Number Theory **15** (2019), no. 7, 1463–1468. doi:10.1142/S1793042119500830
- [8] G. L. Cohen and P. Hagi, Jr., *On the number of prime factors of  $n$  if  $\varphi(n) \mid (n-1)$* , Nieuw Arch. Wisk. (3) **28** (1980), 177–185.
- [9] H. Cohen and H. W. Lenstra, Jr., *Primality testing and Jacobi sums*, Math. Comp. **42** (1984), no. 165, 297–330. doi:10.1090/S0025-5718-1984-0726006-X
- [10] R. J. Cook, *Bounds for odd perfect numbers*, in: Number Theory (Ottawa, 1996), CRM Proc. Lecture Notes **19**, Amer. Math. Soc., Providence, RI, 1999, 67–71.
- [11] D. Coppersmith, N. Howgrave-Graham, and S. V. Nagaraj, *Divisors in residue classes, constructively*, Math. Comp. **77** (2008), no. 261, 531–545. doi:10.1090/S0025-5718-07-02007-8
- [12] J. Feitsma and W. Galway, *Tables of pseudoprimes and related data*, <https://www.cecm.sfu.ca/Pseudoprimes/index-2-to-64.html>.
- [13] The FLINT team, *FLINT: Fast Library for Number Theory*, version 3.6.0, <https://flintlib.org>.

- [14] S. W. Golomb, *Sets of primes with intermediate density*, Math. Scand. **3** (1955), 264–274. [doi:10.7146/math.scand.a-10445](https://doi.org/10.7146/math.scand.a-10445)
- [15] A. Grytczuk and M. Wójtowicz, *On a Lehmer problem concerning Euler’s totient function*, Proc. Japan Acad. Ser. A Math. Sci. **79** (2003), no. 8, 136–138. [doi:10.3792/pjaa.79.136](https://doi.org/10.3792/pjaa.79.136)
- [16] R. K. Guy, *Unsolved Problems in Number Theory*, 3rd ed., Problem Books in Mathematics, Springer, New York, 2004. [doi:10.1007/978-0-387-26677-0](https://doi.org/10.1007/978-0-387-26677-0)
- [17] P. Hagis, Jr., *On the equation  $M \cdot \varphi(n) = n - 1$* , Nieuw Arch. Wisk. (4) **6** (1988), no. 3, 255–261.
- [18] D. R. Heath-Brown, *Odd perfect numbers*, Math. Proc. Cambridge Philos. Soc. **115** (1994), no. 2, 191–196. [doi:10.1017/S0305004100072030](https://doi.org/10.1017/S0305004100072030)
- [19] D. H. Lehmer, *On Euler’s totient function*, Bull. Amer. Math. Soc. **38** (1932), no. 10, 745–751. [doi:10.1090/S0002-9904-1932-05521-5](https://doi.org/10.1090/S0002-9904-1932-05521-5)
- [20] H. W. Lenstra, Jr., *Divisors in residue classes*, Math. Comp. **42** (1984), no. 165, 331–340. [doi:10.1090/S0025-5718-1984-0726007-1](https://doi.org/10.1090/S0025-5718-1984-0726007-1)
- [21] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, New York, 2020, 367–381. [doi:10.1145/3372885.3373824](https://doi.org/10.1145/3372885.3373824)
- [22] H. G. Meijer, *Sets of primes with intermediate density*, Math. Scand. **34** (1974), 37–43. [doi:10.7146/math.scand.a-11502](https://doi.org/10.7146/math.scand.a-11502)
- [23] H. L. Montgomery, *A note on the large sieve*, J. London Math. Soc. **43** (1968), 93–98. [doi:10.1112/jlms/sl-43.1.93](https://doi.org/10.1112/jlms/sl-43.1.93)
- [24] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), no. 2, 119–134. [doi:10.1112/S0025579300004708](https://doi.org/10.1112/S0025579300004708)
- [25] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction – CADE 28, Lecture Notes in Comput. Sci., Springer, Cham, 2021, 625–635. [doi:10.1007/978-3-030-79876-5\\_37](https://doi.org/10.1007/978-3-030-79876-5_37)
- [26] P. P. Nielsen, *An upper bound for odd perfect numbers*, Integers **3** (2003), #A14.
- [27] M. Norman, *A group-theoretical approach to Lehmer’s totient problem*, preprint, 2021. [arXiv:2106.11781](https://arxiv.org/abs/2106.11781)
- [28] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, sequences A050474 and A203966, <https://oeis.org>, accessed September 2026.
- [29] The PARI Group, *PARI/GP*, version 2.15.4, Univ. Bordeaux, <https://pari.math.u-bordeaux.fr>.
- [30] R. G. E. Pinch, *A note on Lehmer’s totient problem*, poster, Seventh Algorithmic Number Theory Symposium (ANTS VII), Berlin, 2006. [https://antsmath.org/ANTSVII/Poster/Pinch\\_Poster3.pdf](https://antsmath.org/ANTSVII/Poster/Pinch_Poster3.pdf)
- [31] C. Pomerance, *On composite  $n$  for which  $\varphi(n) \mid n - 1$ , II*, Pacific J. Math. **69** (1977), no. 1, 177–186. [doi:10.2140/pjm.1977.69.177](https://doi.org/10.2140/pjm.1977.69.177)
- [32] C. Pomerance, J. L. Selfridge, and S. S. Wagstaff, Jr., *The pseudoprimes to  $25 \cdot 10^9$* , Math. Comp. **35** (1980), no. 151, 1003–1026. [doi:10.1090/S0025-5718-1980-0572872-7](https://doi.org/10.1090/S0025-5718-1980-0572872-7)
- [33] J. Renze, *Computational evidence for Lehmer’s totient conjecture*, Mathematica notebook, Wolfram Library Archive, MathSource item 5483, 2004. <https://library.wolfram.com/infocenter/MathSource/5483/>
- [34] J. Sorenson and J. Webster, *Strong pseudoprimes to twelve prime bases*, Math. Comp. **86** (2017), no. 304, 985–1003. [doi:10.1090/mcom/3134](https://doi.org/10.1090/mcom/3134)
- [35] A. Stone, *Improved upper bounds for odd perfect numbers – Part I*, Integers **24** (2024), #A114. [doi:10.5281/zenodo.14547960](https://doi.org/10.5281/zenodo.14547960)
- [36] E. Wong, *Computations on normal families of primes*, M.Sc. thesis, Simon Fraser University, 1997. [https://summit.sfu.ca/\\_flysystem/fedora/sfu\\_migrate/7394/b18771816.pdf](https://summit.sfu.ca/_flysystem/fedora/sfu_migrate/7394/b18771816.pdf)
- [37] T. Yamada, *On Lehmer’s problem and related problems*, preprint, 2023; also RIMS Kôkyûroku No. 2285. [arXiv:2303.16853](https://arxiv.org/abs/2303.16853)

FORMERLY, ELECTRO-GRAVITATIONAL SPACE PROPULSION LABORATORY (EGSPL), BHUBANESWAR, ODISHA 751030, INDIA

Email address: [d.bhattacharjee@erl-forschung.de](mailto:d.bhattacharjee@erl-forschung.de)

Email address: [itsdeep@live.com](mailto:itsdeep@live.com)

URL: ORCID [0000-0003-0466-750X](https://orcid.org/0000-0003-0466-750X)