A FACTORIZATION ALGORITHM BASED ON COHN'S IRREDUCIBILITY CRITERION

EOIN MACKALL

Abstract. We consider an algorithm which, for an integer $n > 1$ as input, outputs either a proper divisor d of n or determines that n is prime. The algorithm is based on an old theorem attributed to Cohn and works by writing n in a small base, forming an integer polynomial from the digits, and factoring this polynomial.

On the one hand, this algorithm simply iterates through all integers from 2 to the square root of n, similar to trial division, and so requires $\mathcal{O}(\sqrt{n} \cdot \text{Poly}(\log(n)))$ bit operations in a worst case. On the other hand, for composite n, this algorithm is likely to terminate in less steps than needed in trial division.

1. Introduction

The problem of either finding an efficient factorization algorithm for large integers or, showing such an algorithm doesn't exist, is difficult. The large computational cost of all currently known factorization algorithms forms the basis for the security of RSA cryptographic primitives, which are still in widespread use today. The best known algorithms for finding a proper divisor of a composite integer, e.g. the Elliptic Curve Method [Len87] or the General Number Field Sieve, are known to run with subexponential time complexity.

As an initial step in factoring algorithms where sieving over number field elements is required, one needs to produce an irreducible integral polynomial $f(x)$ which has a known solution modulo n for the integer n that will be factored. To construct such a polynomial, one searches over integers $2 \leq b \leq \sqrt{n}$, one writes n in base-b, and then one takes the polynomial whose coefficients are the digits of this base-b representation of n. Finding criteria for choosing an adequate polynomial $f(x)$ (i.e. a polynomial which may considerably speed up the factoring algorithm, see [Mur98], [MB98], and [DZ20, Cox15]) in these algorithms is a topic of modern research.

Date: March 15, 2026.

2020 *Mathematics Subject Classification.* 11Y16.

Key words and phrases. integer factorization; polynomial factorization.

The idea behind considering polynomials formed from the digits of a base-b expansion of n in this way is likely based on *Cohn's Irreducibility Criterion* for integer polynomials. Namely, Cohn's theorem implies that the polynomials formed in this way are irreducible if they take on a prime value for some input $x > b$, which seems to happen frequently. As we explain below, even moreso, a proof of this theorem due to Murty [RM02] shows that if such a polynomial $f(x)$ obtained from the base-b representation in this way factors, then any proper polynomial factor will evaluate at $x = b$ to a proper divisor of n.

The goal of this article is to consider the process of factoring n using only Murty's theorem, and to experimentally analyze the distribution of those integers $2 \leq b \leq \sqrt{n}$ that produce a proper factor of n. To do so, we consider the algorithm which – iteratively, for each b in this range – constructs the given polynomial from the digits of n in base-b and factors it, if possible. Note that there is no hope for this algorithm to factor any particular large composite number as its run-time complexity requires an exponential $\mathcal{O}(\sqrt{n} \cdot \text{Poly}(\log(n)))$ number of bit operations in a worst case. Still, the exceptional integers $2 \leq b \leq \sqrt{n}$ that are not proper divisors of n, yet still factor n in this manner, are interesting.

To this end, we give a conjectural description for the average order of the arithmetic function counting these "generalized divisors" of n, analogous to the well-known average order of the divisor function. We call this function $\tau_{\geq 2}\beta(x)$ below.

Conjecture. *There exists constants $c, \varepsilon > 0$ such that the average order of $\tau_{\geq 2}\beta$ is given as*

$$\sum_{n \leq x} \tau_{\geq 2}\beta(n) = cx^{1+\varepsilon} + o(x^{1+\varepsilon}).$$

The average order of the function $\tau_{\geq 2}\beta(x)$ was computed directly over the range of integers $2 \leq x \leq 42,000,000$ (see Figures 1 and 3). Heuristics are provided that seem to explain the behavior of the function $\tau_{\geq 2}\beta(x)$, for general $x \in \mathbb{Z}$, in more detail in Section 3. We also prove some facts about the function $\tau_{\geq 2}\beta(n)$ along the way: for example, it is unbounded on the restricted set of semiprime integers.

Data Collection. Data collection and data processing for this project was done using the source code available at `https://github.com/eoinmackall/CohnFactorization`. The algorithms that were used to generate the data for Figures 1-5 were written using the FLINT C library. The main data collection computation was accomplished by an AMD Ryzen Threadripper 7960X 24-Core CPU making use of OpenMP

to parallelize the runtime over 48 threads; the total wall-clock time for the collection was about 30 minutes.

2. Cohn's Irreducibility Criterion

Let $n > 0$ be an integer. Given a second integer $b > 0$, we write $n = (b_r b_{r-1} \cdots b_1 b_0)_b$ to denote the base b representation of n, so there is an equality

$$n = b_r \cdot b^r + b_{r-1} \cdot b^{r-1} + \cdots + b_1 \cdot b + b_0$$

for integers $0 \leq b_0, ..., b_r < b$.

Definition 2.1. Let $\Phi : \mathbb{N} \times \mathbb{Z}_{>1} \to \mathbb{Z}[x]$ be the function defined as follows: for any tuple (n, b) we have

$$\Phi(n, b)(x) = b_r x^r + b_{r-1} x^{r-1} + \cdots + b_1 x + b_0$$

where $n = (b_r b_{r-1} \cdots b_1 b_0)_b$ is the base b representation of n.

Cohn's Irreducibility Criterion, in its original form, is the statement that $\Phi(n, 10)(x)$ is irreducible if n is prime. More recently, the phrase *Cohn's Irreducibility Criterion* may also refer to the following theorem, generalizing this original statement, due to Murty [RM02, Theorem 2].

Theorem 2.2. *Let $n > 0$ be an integer and let $b \geq 2$ be a second integer. If n is prime, then $\Phi(n, b)(x)$ is irreducible.*

While the above theorem immediately implies the statement that if $\Phi(n, b)(x)$ is reducible then n is composite, a careful reading of the proof of the theorem actually shows the stronger statement that any proper polynomial factor of $\Phi(n, b)(x)$ gives a proper divisor of n on evaluation at b. Stated precisely:

Theorem 2.3. *Let $n > 0$ be an integer and let $b \geq 2$ be any integer. Suppose that there is a factorization*

$$\Phi(n, b)(x) = f(x)g(x) \qquad \deg(f), \deg(g) < \deg \Phi(n, b)(x)$$

for integer polynomials $f(x)$ and $g(x)$. Then $1 < f(b) < n$ is a proper divisor of n.

Proof. We explain how the results of [RM02] allow one to deduce this result directly. Since $\Phi(n, b)(x)$ is nonnegative for any $x \geq 0$ (and in fact $\Phi(n, b)(x) = 0$ if and only if both $x = 0$ and b divides n), we can assume that both $f(x)$ and $g(x)$ are also nonnegative for any $x \geq 0$. Since also $f(b)g(b) = \Phi(n, b)(b) = n$, we're reduced to showing that an equality $f(b) = 1$, for $f(x)$ as above, is impossible.

By [RM02, Lemma 2], any root $\alpha \in \mathbb{C}$ of $\Phi(n,b)(x)$ satisfies

$$|\alpha| < \frac{1+\sqrt{1+4(b-1)}}{2}.$$

We can write

$$f(x) = c\prod_{i=1}^{m}(x-\alpha_i)$$

for the leading coefficient $c \in \mathbb{Z}$ and α_i a subset of the roots of $\Phi(n,b)(x)$. For any integer $b > 2$, we have

$$|\alpha_i| < \frac{1+\sqrt{1+4(b-1)}}{2} \leq b-1$$

so that $|b-\alpha_i| > 1$. In particular, for all $b > 2$, this implies $f(b) > 1$.

For the case when $b = 2$, a more precise bound and a more accurate argument is needed. In [RM02, Lemma 3] it's shown that any root $\alpha \in \mathbb{C}$ of a polynomial such as $\Phi(n,2)(x)$ satisfies $\mathrm{Re}(\alpha) < 3/2$. Thus

$$|f(2)| = c\prod_{i=1}^{m}|2-\alpha_i| > c\prod_{i=1}^{m}|1-\alpha_i| = |f(1)|.$$

But $f(1) \in \mathbb{Z}$ is positive, so $|f(2)| > |f(1)| \geq 1$. □

3. An Algorithm For Integer Factorization

Theorem 2.3 suggests the following algorithm (Algorithm 1 below) for either finding a proper divisor of an integer $n > 1$ or concluding that n is prime.

For each integer $2 \leq b < \lfloor\sqrt{n}\rfloor$, the coefficients of $\Phi(n,b)(x)$ can be calculated using at most $\mathcal{O}(\log(n)^3)$ bit operations. Further, since the polynomial $\Phi(n,b)(x)$ has degree $\lfloor\log_b(n)\rfloor$, it can be decided whether or not $\Phi(n,b)(x)$ is irreducible, and if $\Phi(n,b)(x)$ is found to be reducible then $\Phi(n,b)(x)$ can subsequently be factored, in at most $\mathcal{O}(\mathrm{Poly}(\log(n)))$ bit operations, see [LLL82]. Thus, the simple algorithm described has time complexity as stated in the abstract.

Corollary 3.1. *The above algorithm correctly returns either a divisor of $n > 1$ or returns that n is prime using at most $\mathcal{O}(\sqrt{n}\cdot\mathrm{Poly}(\log(n)))$ bit operations.*

Proof. The algorithm runs at most $\sqrt{n}$ iterations. Each iteration uses $\mathcal{O}(\mathrm{Poly}(\log(n)))$ bit operations. To see that it terminates correctly, note that a composite integer n has a divisor $d \leq \sqrt{n}$ and $\Phi(n,d)(x)$ is reducible for such d. □

Algorithm 1 A factorization algorithm for $n \in \mathbb{Z}$ and $n > 1$

function IS_IRREDUCIBLE($f(x)$)
    if $f(x)$ is irreducible **then**
        return true
    else
        return false
    end if
end function

function POLY_FACTOR($f(x)$)
    if Is_irreducible($f(x)$)== **false then**
        return $g(x)$ dividing $f(x)$ with $\deg(g) < \deg(f)$
    else
        return $f(x)$
    end if
end function

for $b = 2$ **to** $b = \lfloor \sqrt{n} \rfloor$ **do**
    if Is_irreducible($\Phi(n, b)(x)$)==**false then**
        $g(x) \leftarrow$ Poly_factor($\Phi(n, b)(x)$)
        return $g(b)$
    end if
end for
return n is prime

Remark 3.2. The algorithm above can be improved in efficiency for the range between $\lfloor \sqrt[5]{n} \rfloor + 1$ and $\lfloor \sqrt{n} \rfloor$ using explicit formulas for the roots of such polynomials. For example, if $n > 1$ and $b \geq 2$ are integers such that $\Phi(n, b)(x) = cx^2+dx+e$ for integers $c, d, e > 0$ then $\Phi(n, b)(x)$ factors in $\mathbb{Z}$ if and only if both $d^2 - ce = z^2$ is a square of an integer $z \geq 0$ and if $d \equiv z \pmod 2$.

The main interest in the above algorithm is that it often terminates at an iterate $2 \leq k \leq \lfloor \sqrt{n} \rfloor$ where k is not a proper divisor of n.

Example 3.3. Let $n = 207314063849$. Then

$$n = 207314063849 = 323131 \cdot 641579$$

is a prime factorization of n. But

$$\Phi(207314063849, 40098) = 128x^2+37640x+25817 = (8x+347)(16x+11).$$

Evaluating at $x = 40098$ yields $323131 = 8(40098) + 347$.

To quantify the number of successes obtained through the above algorithm, we introduce the following arithmetic functions.

Definition 3.4. For any integer $n \geq 2$, define the following set

$$B(n) = \{b \in \mathbb{Z} : 1 < b \leq n,\ \Phi(n,b)(x) \text{ is not primitive or is reducible}\}.$$

Then define $\beta(n) = \#B(n)$ to be the cardinality of this set. Similarly, for any $k \in \mathbb{N}$ define $(\tau_{\geq k}\beta)(n)$ to be the function with

$$(\tau_{\geq k}\beta)(n) = \#\left(B(n) \cap \{1, ..., \lfloor \sqrt[k]{n} \rfloor\}\right).$$

It follows almost immediately from the above that $\beta(n) \geq \tau(n) - 1$. We include a plot, in Figure 1 below, of points $(n, \tau_{\geq 2}\beta(n))$ for a random sample of one million integers from the interval $2 \leq n \leq 42,000,000$. For comparison we include a plot, Figure 2 below, of the points $(n, \tau(n))$ for the divisor counting function τ over the same interval of integers.

A least squares regression of the cumulative average for the function $\tau_{\geq 2}\beta$ computed over the set $2 \leq n \leq 42,000,000$ is given in Figure 3. The graphic is suggestive of the following conjecture:

Conjecture 3.5. *There exists constants $c, \varepsilon > 0$ such that the average order of $\tau_{\geq 2}\beta$ is given as*

$$\sum_{n \leq x} \tau_{\geq 2}\beta(n) = cx^{1+\varepsilon} + o(x^{1+\varepsilon}).$$

We don't know how one might go about proving Conjecture 3.5. Instead, we focus the remainder of this paper on analyzing the behavior of the function $\tau_{\geq 2}\beta$ on semiprimes, i.e. integers n which factor $n = pq$ as a product of two primes p, q. These integers are interesting not only due to their practical use in RSA encryption schemes but, because they form the class of composite integers where τ takes the minimal value 4. In contrast, Figure 4, which displays points $(n, \tau_{\geq 2}\beta(n))$ over a sample of $100,000$ semiprimes n from the same interval $2 \leq n \leq 42,000,000$, shows that $\tau_{\geq 2}\beta$ is very much nonconstant on semiprimes.

The striations in the graph of the function $\tau_{\geq 2}\beta$ on semiprimes are comprised of integers sharing the same smallest prime divisor (the band in the graph which seems to be on top is made up of even numbers; below that is a band of numbers divisible by 3 and so on). Heuristically, for fixed n and varying b, the polynomials $\Phi(n,b)(x)$ seem to behave like essentially random polynomials with a bias. The value of $\tau_{\geq 2}\beta$ is then mostly determined by the number of bases b whose digits are all a multiple of a divisor of n; strangely, however, this often under-counts the value of $\tau_{\geq 2}\beta$, sometimes even by a large margin.

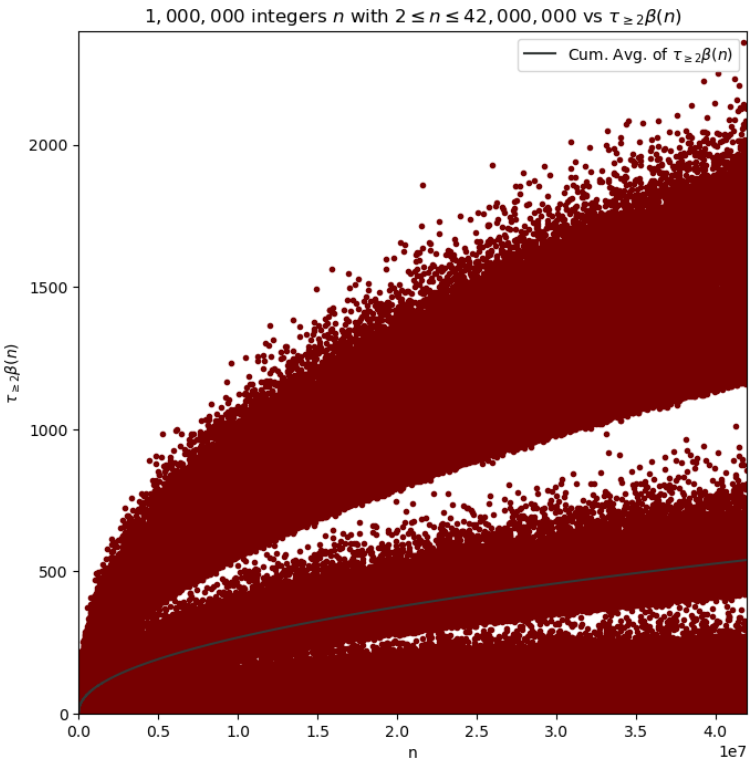

FIGURE 1. A plot of $1,000,000$ random points $(n, \tau_{\geq 2}\beta(n))$ sampled from $2 \leq n \leq 42,000,000$. The cumulative average was computed using all integers in this interval.

Example 3.6. Even semiprime numbers $n = 2p$ with $n \approx 25,000,000$ typically have a value of $\tau_{\geq 2}\beta(n) \approx 800$ according to Figure 4. For this value of n, we have $\sqrt{n} \approx 5000$ and $\sqrt[3]{n} \approx 292$. Most of the $\Phi(n, b)(x)$ for values of b in $2 \leq b \leq \sqrt{n}$ are therefore quadratic. If each coefficient of these quadratic polynomials has around a 50% chance at being even, then there is about a $12.5\% = 1/8$ chance that any given polynomial $\Phi(n, b)(x)$ has content a multiple of 2. Altogether this accounts for approximately $5000/8 \approx 625$ of the 800 "generalized divisors" of n.

Additionally, restricted to semiprime integers $n = pq$ of the same approximate magnitude, the function $\tau_{\geq 2}\beta$ seems to depend on the *relative* size of the prime divisors p and q (in a sense made precise in Figure 5). Experimental evidence suggests that $\tau_{\geq 2}\beta$ takes on a larger

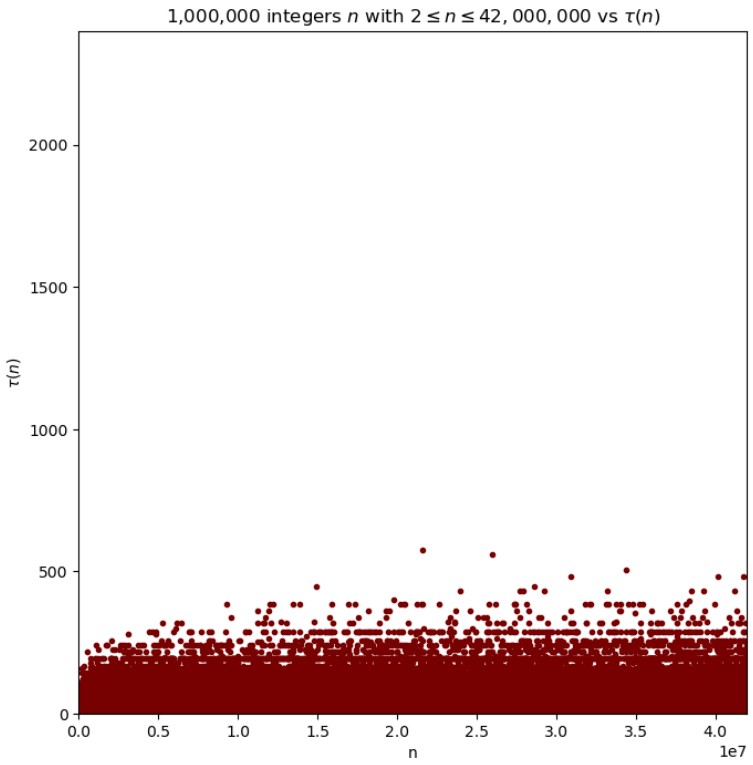

FIGURE 2. A plot of $1,000,000$ random points $(n, \tau(n))$ sampled from $2 \le n \le 42,000,000$.

value for semiprimes $n = pq$ where p and q are similar in size, $p \approx q$. This observation may be mostly attributable to the following result.

Theorem 3.7. *Let $\mathcal{S} \subset \mathbb{N}$ denote the set of semiprime integers. Then*

$$\sup\{\tau_{\geq 2}\beta(s) : s \in \mathcal{S}\} = \infty.$$

Proof. We use the fact [Zha14] that there exists an integer $k > 1$ and an infinite sequence of primes $p_1 < p_2, p_3 < p_4, ...$ satisfying the inequalities $p_{2i} - p_{2i-1} < k$ for all integers $i \geq 1$. We set $n_i = p_{2i-1}p_{2i}$ and we set $j_i = \lfloor p_{2i-1}/2 \rfloor + 1$. Note that for any integer $p_{2i-1} > b \geq j_i + \frac{k}{2}$ we have

$$p_{2i} = b + m_{2i}(b) \quad \text{and} \quad p_{2i-1} = b + m_{2i-1}(b)$$

for integers $1 \leq m_{2i}(b), m_{2i-1}(b) < b$ with $m_{2i}(b) - m_{2i-1}(b) < k$.

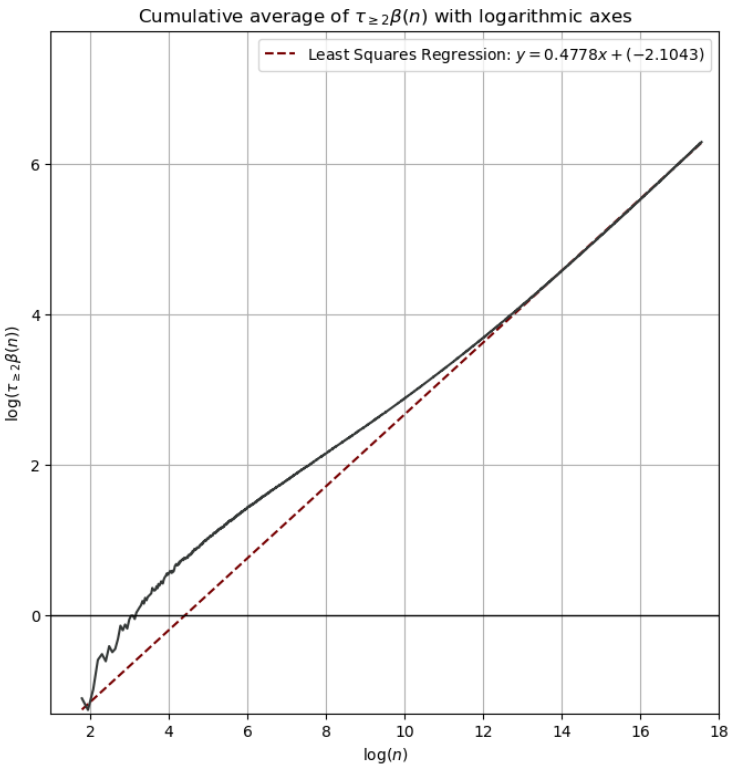

FIGURE 3. A least squares regression of the cumulative average of the function $\tau_{\geq 2}\beta(n)$ for all integers $2 \leq n \leq 42,000,000$ with logarithmic axes suggests $\varepsilon \approx 0.4778$ and $c \approx 0.1219$ in Conjecture 3.5.

It follows from the above that there is an equality

$$\begin{aligned} n_i = p_{2i-1}p_{2i} &= (b + m_{2i-1}(b))(b + m_{2i}(b)) \\ &= b^2 + (m_{2i-1}(b) + m_{2i}(b))b + m_{2i-1}(b)m_{2i}(b). \end{aligned} \tag{1}$$

For any $N \in \mathbb{N}$, we may choose $i \gg 0$ sufficiently large so that for all b with $k+1 \leq m_{2i}(b) < N$ we have

$$m_{2i-1}(b) + m_{2i}(b) < j_i \quad \text{and} \quad m_{2i-1}(b)m_{2i}(b) < j_i.$$

Hence the equation (1) yields a factorization of $\Phi(n_i, b)(x)$ for all such integers b. It follows that $\tau_{\geq 2}\beta(n_i) \geq N - k - 2$ for $i \gg 0$ large enough. Since N was arbitrary, this implies the claim. □

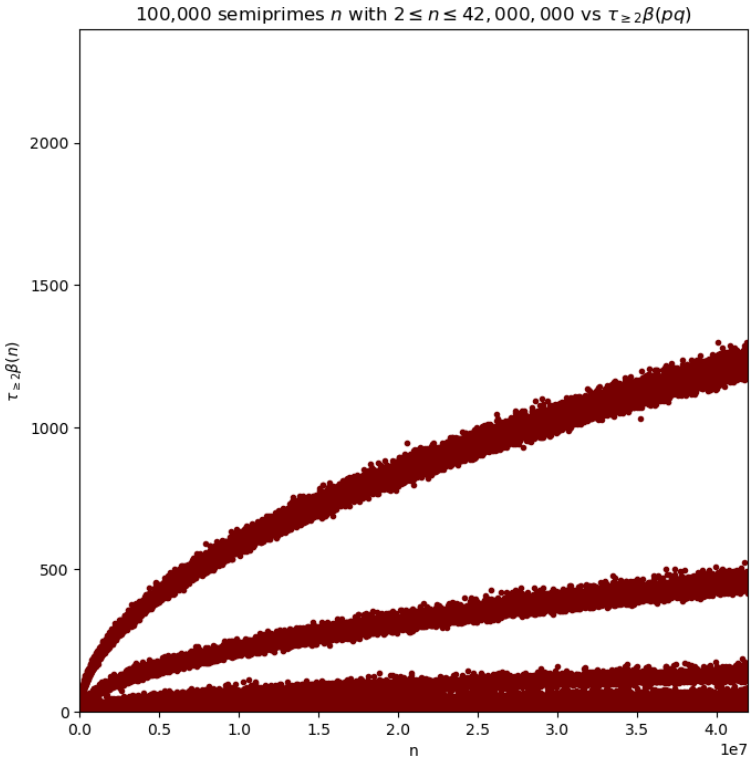

FIGURE 4. A plot of $100,000$ random points $(n, \tau_{\geq 2}\beta(n))$ sampled from all semiprimes $n = pq$ with distinct prime factors $p \neq q$ in the interval $2 \leq n \leq 42,000,000$.

Example 3.8. For the semiprime

$$n = 30674101 = 331 \cdot 92671$$

we have $\tau_{\geq 2}\beta(n) = 1$ so that $\Phi(n, b)(x)$ either factors or $\Phi(n, b)(x)$ is not primitive for $2 \leq b < \sqrt{n}$ if and only if $b = 331$.

On the other hand, for the semiprime

$$m = 250000070000004899 = 500000069 \cdot 500000071$$

we have $\tau_{\geq 2}\beta(m) = 56706$.

REFERENCES

[Cox15] Nicholas Coxon, *Montgomery's method of polynomial selection for the number field sieve*, Linear Algebra Appl. **485** (2015), 72–102. MR 3394139

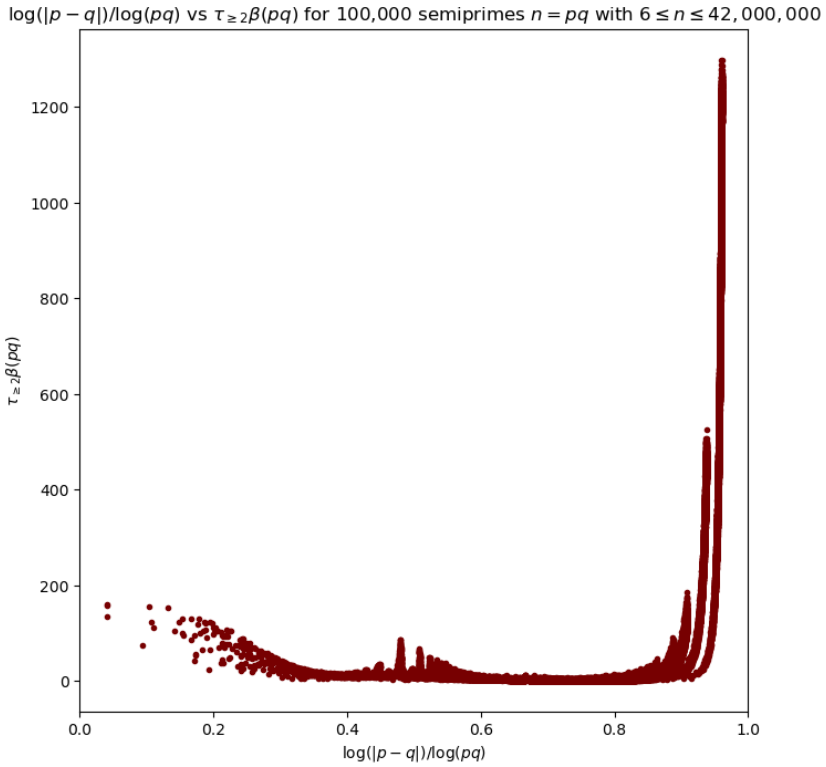

FIGURE 5. A plot of $\log|p-q|/\log(pq)$ vs. $\tau_{\geq 2}\beta(pq)$ for the semiprimes $n = pq$ considered in Figure 4.

[DZ20] Nicolas David and Paul Zimmermann, *A new ranking function for polynomial selection in the number field sieve*, 75 years of mathematics of computation, Contemp. Math., vol. 754, Amer. Math. Soc., [Providence], RI, [2020] ©2020, pp. 315–325. MR 4132128

[Len87] H. W. Lenstra, Jr., *Factoring integers with elliptic curves*, Ann. of Math. (2) **126** (1987), no. 3, 649–673. MR 916721

[LLL82] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), no. 4, 515–534. MR 682664

[MB98] Brian Murphy and Richard P. Brent, *On quadratic polynomials for the number field sieve*, Computing theory '98 (Perth), Aust. Comput. Sci. Commun., vol. 20.3, Springer, Singapore, 1998, pp. 199–213. MR 1723947

[Mur98] Brian Murphy, *Modelling the yield of number field sieve polynomials*, Algorithmic number theory (Portland, OR, 1998), Lecture Notes in Comput. Sci., vol. 1423, Springer, Berlin, 1998, pp. 137–150. MR 1726067

[RM02] M. Ram Murty, *Prime numbers and irreducible polynomials*, Amer. Math. Monthly **109** (2002), no. 5, 452–458. MR 1901498

[Zha14] Yitang Zhang, *Bounded gaps between primes*, Ann. of Math. (2) **179** (2014), no. 3, 1121–1174. MR 3171761

Department of Mathematics, University of California Santa Cruz, Santa Cruz, CA USA

Email address: eoinmackall@gmail.com