On the proportion of numbers coprime to a given integer

Paul Erdős

Florian Luca
Instituto de Matemáticas
Universidad Nacional Autonoma de México
C.P. 58089, Morelia, Michoacán, México
fluca@matmor.unam.mx

Carl Pomerance
Department of Mathematics
Dartmouth College
Hanover, NH 03755–3551, USA
carl.pomerance@dartmouth.edu

December 29, 2006

Abstract

For a positive integer n and its Euler function $\phi(n)$ we write $\phi(n)/n = a/b$, where $a = a(n)$ and $b = b(n)$ are coprime. For a fixed integer a, we consider the number of integers b for which the above relation holds for some n, and we also fix b and count corresponding a's. We discuss the greatest common divisor of n and $\phi(n)$, applying it to the relation $\phi(n) \mid f(n)$ for f a polynomial.

Mathematics Subject Classification: 11N37, 11N56
Key Words: Euler function

1 Introduction

For a positive integer n we write $\phi(n)$ for the Euler function of n, namely the number of integers in $[1, n]$ coprime to n. The fraction $\phi(n)/n$ is thus the asymptotic density of the set of the positive integers relatively prime to n. This proportion has been extensively studied. For example, it was known to Euler that $\{\phi(n)/n\}_{n\geq 1}$ is dense in $[0, 1]$, and one of the earliest results concerning the distribution of values of arithmetic functions is Schoenberg's 1928 theorem (see [16]) to the effect that $\phi(n)/n$ possesses a continuous distribution in $[0, 1]$. That is, $D(u)$, defined as the asymptotic density of the set of those positive integers n such that $\phi(n)/n \leq u$, exists for every real number u. In addition, $D(u)$ is continuous and strictly increasing on $[0, 1]$ (and clearly $D(0) = 0$ and $D(1) = 1$). This result might be argued to mark the dawn of probabilistic number theory.

This paper discusses several arithmetic functions that are directly related to the ratio $\phi(n)/n$. First, it is natural to reduce this fraction to its lowest terms: $\phi(n)/n = a/b$. We write $a = a(n)$ and $b = b(n)$, so that

$$\phi(n)/n = a(n)/b(n), \quad \gcd(a(n), b(n)) = 1.$$

Let $\mathrm{rad}(n)$ denote the largest squarefree number that divides n. Since

$$\phi(n)/n = \prod_{p|n,\ p \text{ prime}} (1 - 1/p) = \phi(\mathrm{rad}(n))/\mathrm{rad}(n),$$

we have

$$a(n) = a(\mathrm{rad}(n)), \quad b(n) = b(\mathrm{rad}(n))$$

for every positive integer n. For positive integers a, b, we put

$$f(a) = \#\{n \text{ squarefree} : a(n) = a\},$$

$$g(b) = \#\{n \text{ squarefree} : b(n) = b\}.$$

For example, $f(1)$ is the number of squarefree n with $\phi(n) \mid n$. It is easy to see that $n = 1, 2, 6$ are the only such numbers, so that $f(1) = 3$, a fact recorded in [17], p. 232.

As we shall see in the next section, the function $n \mapsto \phi(n)/n$ is one-to-one when restricted to squarefree numbers, so we have the alternate definitions

$$f(a) = \#\{b \ :\ \gcd(a, b) = 1 \text{ and } a/b = \phi(n)/n \text{ for some } n\},$$

$$g(b) = \#\{a \ : \ \gcd(a, b) = 1 \text{ and } a/b = \phi(n)/n \text{ for some } n\}.$$

For any given numbers a, b, we present simple finite procedures for computing the values $f(a), g(b)$. We discuss the maximal orders of the arithmetic functions $f(a), g(b)$, and we show these functions are normally 0. We also discuss $\gcd(n, \phi(n)) = n/b(n) = \phi(n)/a(n)$. We show that normally it is the largest divisor of n composed of primes at most $\log\log n$, and on average it is bounded by $n^{o(1)}$. We also consider its maximal order when restricted to squarefree values of n. Our result on the average order of $\gcd(n, \phi(n))$ has an application in the counting of solutions of certain polynomial congruences.

Throughout this paper, we use the order symbols $\gg$, $\ll$, $\asymp$, O, and o with their ususal meanings in analytic number theory. They are all absolute except where specified differently as in Theorem 11 and in Section 6. For a positive real number x, we use $\log x$ for the natural logarithm of x, $\log_1 x = \max\{1, \log x\}$, and if $k \geq 2$ we use $\log_k x$ for the k-fold iterated composition of the function $\log_1$ evaluated at x. We use p and q with or without subscripts for prime numbers. We use $\pi(x)$ and $\pi(x; b, a)$ for the number of primes $p \leq x$ and the number of primes $p \leq x$ in the arithmetic progression $p \equiv a \pmod b$, respectively. We use the notation $v_p(n)$ for the exponent on the prime p in the prime factorization of the natural number n. (In particular, if $p \mid n$, then $p^{v_p(n)} \| n$, and if $p \nmid n$, then $v_p(n) = 0$.) We let $P(n)$ denote the greatest prime factor of n if $n > 1$, and we let $P(1) = 1$. We let $\tau(n)$ denote the number of divisors of n, and $\omega(n)$ denotes the number of these divisors that are prime. We let p_i denote the i-th prime. We also use $c_0,\ c_1,\ \ldots$ for positive computable constants.

Acknowledgements. The functions $f(a)$ and $g(b)$ were discussed by the first and third authors about 20 years ago; in particular much of the material in Sections 3 and 4 dates from that period. The current attack on these problems started during a very enjoyable visit of the third author at the Mathematical Institute of the UNAM in Morelia, Mexico as a Distinguished Visiting Professor of the Mexican Academy of Sciences in February of 2006. He would like to thank these institutions for their hospitality and support. The second author was supported in part by grants PAPIIT IN104505, SEP-CONACyT 46755 and a Guggenheim Fellowship. The third author was also supported in part by NSF grant DMS-0401422.

2 An algorithm

In this section we give simple procedures for the evaluation of the functions $f(a)$ and $g(b)$. There is no emphasis on efficiency, only on the existence of a deterministic procedure for the evaluations. We begin with a simple lemma.

Lemma 1. *Let a, b be coprime natural numbers. If there is a natural number m with $\phi(m)/m = a/b$, then*

(i) $0 < a/b \le 1$;

(ii) *b is squarefree;*

(iii) *there is a unique squarefree number n with $\phi(n) = a/b$;*

(iv) $P(b) = P(n)$;

(v) $\gcd(b, \phi(b)) = 1$;

(vi) $\omega(n) \le v_2(a) + 2$.

Proof. The first assertion follows immediately from $0 < \phi(m) \le m$ for all m. As we saw in the Introduction, if $\phi(m)/m = a/b$, then $n = \mathrm{rad}(m)$ is squarefree and $\phi(n)/n = a/b$. Thus, $b \mid n$, so that b is squarefree. Suppose n_1, n_2 are squarefree numbers with $\phi(n_i)/n_i = a/b$ for $i = 1, 2$. Since $\phi(n) < n$ for $n > 1$, it follows that $n_1 = n_2 = 1$ if $a/b = 1$. Suppose that $a/b < 1$, so that $n_1, n_2 > 1$. As $P(n_i) \nmid \phi(n_i)$, we have $P(n_i) \mid b$. But $b \mid n_i$ implies $P(b) \le P(n_i)$ for $i = 1, 2$, so that $P(n_1) = P(n_2) = P(b)$. That is, if n_1, n_2 are squarefree and $\phi(n_1)/n_1 = \phi(n_2)/n_2$, then $P(n_1) = P(n_2)$. We thus may replace n_i with $n_i/P(n_i)$ and iterate, coming to the conclusion that $n_1 = n_2$. This proves the uniqueness assertion in (iii), and in the course of the proof, we saw (iv).

To prove (v), assume not and let p be a common prime factor of b and $\phi(b)$. Let n be squarefree with $a/b = \phi(n)/n$, so that $b \mid n$, and $\phi(b) \mid \phi(n)$. It follows that $v_p(\phi(n)) \ge 1 = v_p(n)$, so that in the reduction $\phi(n)/n$ to a/b, the denominator b is not divisible by p after all. Thus it must be that $\gcd(b, \phi(b)) = 1$.

To see the last assertion, let $k = \omega(n)$. Then n is divisible by at least $k - 1$ odd prime factors, so that $v_2(\phi(n)) \ge k - 1$. But n is squarefree, so that $v_2(n) \le 1$ and $v_2(a) = v_2(\phi(n)) - v_2(n) \ge k - 2$. □

So, given coprime natural numbers a, b we might ask how we might determine if a/b is in the range of $\phi(n)/n$, and if it is, how we might find the unique squarefree pre-image n. The following algorithm gives such a procedure.

Algorithm A. `Let` a, b `be coprime natural numbers. If there is a squarefree number` n `with` $\phi(n)/n = a/b$`, this algorithm finds` n`. If there is no such number` n`, this algorithm reports NONE.`

1. `Let` $n = 1$;
2. `If` $a = b$, `report` n;
3. `If` $a > b$, `report NONE;`
4. `If` b `is not squarefree, report NONE;`
5. `Let` $p = P(b)$, $d = \gcd(a, p-1)$, $a = a/d$, $b = (p-1)b/pd$, $n = pn$, `and go to step 2;`

It is clear from Lemma 1 that the algorithm correctly reports NONE when it does so. The iteration in the last step is based on the fact that if a squarefree number n exists with $\phi(n)/n = a/b$, then Lemma 1 implies that $P(n) = P(b)$, so that if $p = P(n)$ and $N = n/p$, we have

$$\frac{\phi(N)}{N} = \frac{a/(p-1)}{b/p} = \frac{a/\gcd(a, p-1)}{(p-1)/\gcd(a,(p-1)) \cdot b/p} = \frac{a'}{b'},$$

say. We thus may reduce the problem to the new pair a', b', justifying the last step of the algorithm.

We may use Algorithm A to compute $f(a)$, $g(b)$ as follows. Given a natural number b, run Algorithm A for each pair a, b with $1 \leq a \leq b$ and $\gcd(a, b) = 1$. Count 1 for each time the algorithm reports a value of n with $\phi(n)/n = a/b$, the total count being $g(b)$. For the $f(a)$ computation, note that if $\phi(n)/n = a/b$ for some n and b with b coprime to a, then from Lemma 1, we have $k := v_2(a) + 2 \geq \omega(n)$. If such a number n exists, then $\phi(n)/n \geq \prod_{i=1}^{k}(1 - 1/p_i)$. Let this product be denoted $z = z(a)$. Then $b = an/\phi(n) \leq a/z$. So, to compute $f(a)$, for each integer b coprime to a with $a \leq b \leq a/z$, run Algorithm A and count 1 for each such b where the algorithm reports a value for n with $\phi(n)/n = a/b$. The total count is $g(b)$.

We remark that the algorithms in this section have not been optimized, our point merely being that there exists a deterministic procedure. The issue of computing the solutions n to $\phi(n) = m$ has been discussed from the point of view of complexity in [5]. We also remark that a procedure similar to Algorithm A was presented by Anderson [2].

3 The function $f(a)$

We start with the maximal order of $f(a)$.

Theorem 2. *The inequality $f(a) \le (1+o(1))a \log_2 a/\log_3 a$ holds as $a \to \infty$. On the other hand, there exists a positive constant c_0 such that $f(a) > a^{c_0}$ holds for infinitely many a.*

Proof. Assume that n is a positive integer such that $\phi(n)/n = a/b$ for some integer b coprime to a. As we saw at the end of the last section, we have

$$b \le B := a \prod_{i=1}^{k} (1 - 1/p_i)^{-1},$$

where $k = v_2(a) + 2$. But $v_2(a) \le \log a/\log 2$, so that by the prime number theorem, $p_k \le (1 + o(1)) \log a \log\log a/\log 2$. Thus, by Mertens's theorem,

$$b \le B \le (e^{\gamma} + o(1))a \log\log a, \tag{1}$$

where γ is the Euler–Mascheroni constant. But, by Lemma 1,

$$f(a) \le \#\{b \le B \ : \ \gcd(b, \phi(b)) = 1\}.$$

By a result of Erdős (see [7]), the cardinality of the set

$$\{n \le y \ : \ \gcd(n, \phi(n)) = 1\}$$

is $(e^{-\gamma} + o(1))y/\log_3 y$ as $y \to \infty$. Applying this with $y = B$ and using our upper bound for B, we get the first claim of the theorem.

For the second claim, it is known that there exists a positive constant c_1 such that for infinitely many positive integers m, the inequality

$$\mathcal{A}_m = \#\{n \ : \ \mu^2(n) = 1 \text{ and } \phi(n) = m\} \ge m^{c_1}$$

holds. Here, $\mu(n)$ is the Möbius function of n. The above result is due to Erdős and appears in [6]. Let m be one of these integers, and let $n \in \mathcal{A}_m$.

Then $\phi(n)/n = m/n = a/b$ for some divisor a of m. Thus, there are at least $\#\mathcal{A}_m/\tau(m) \ge m^{c_1}/\tau(m)$ values of n corresponding to the same value of a. Since for each such n we have that $b = na/m$, we get that the values of b are distinct. Since $\tau(m) = m^{o(1)}$ as m tends to infinity, the second claim of the theorem follows with any constant c_0 smaller than c_1. $\square$

The best (largest) known constant c_1 above is $0.7067\ldots$ and is due to Baker and Harman [3]. We conjecture that $f(a) \ge a^{1-(1+o(1))\log_3 a/\log_2 a}$ holds for infinitely many positive integers a. In fact, we conjecture that this function of a is also an upper bound for $f(a)$ as $a \to \infty$, but we cannot even prove that the inequality $f(a) < a$ holds for all sufficiently large values of a.

Theorem 3. *We have $f(a) = 0$ for almost all positive integers a.*

Proof. We prove more. Namely, we show that if we put

$$\mathcal{A}(x) = \{a \le x \ : \ f(a) \ne 0\},$$

then

$$\#\mathcal{A}(x) \le x \exp\big(-(\log 2 + o(1)) \log_2 x/\log_3 x\big) \qquad \text{as } x \to \infty. \tag{2}$$

We let x be large, let $k < \log_2 x$ be a positive integer tending to infinity with x in a way which will be specified later, and let $K = \lfloor 2\log_2 x \rfloor$. Let $\mathcal{A}_1$ be the set of $a \le x$ such that $2^{k-1} \mid a$ and $\mathcal{A}_2$ be the set of $a \le x$ such that $\omega(a) > K$. It is obvious that

$$\#\mathcal{A}_1 \le \frac{x}{2^{k-1}}. \tag{3}$$

Further, it follows by a result of Hardy and Ramanujan [11] that

$$\#\mathcal{A}_2 \ll \frac{x}{(\log x)^{2\log 2 - 1}}. \tag{4}$$

Let $\mathcal{A}_3$ be the set of all $a \le x$ not in $\mathcal{A}_1 \cup \mathcal{A}_2$ for which $f(a) \ne 0$. Let $a \in \mathcal{A}_3$ and assume that $a/b = \phi(n)/n$ for some squarefree integers b, n, where a and b are coprime. Write $n = uv$ where $u = \gcd(n, \phi(n))$, so that $a = \phi(v)\phi(u)/u$. Since $\omega(u) \le \omega(n) \le k < \log_2 x$, we have

$$\phi(v) \le xu/\phi(u) \ll x \log_3 x. \tag{5}$$

We fix a value $w = \phi(v)$ of the Euler function that satisfies (5) and ask how many values of u with $w\phi(u)/u \in \mathcal{A}_3$ can correspond to it. We will show that for a given w there are at most $O((K+k)^k)$ choices for u that can work, and it will remain to multiply this quantity by the number of choices for w.

Write $u = q_1 \dots q_l$ where $q_1 < \cdots < q_l$. Since $v_2(a) \leq k-2$ and $a = w\phi(u)/u$, it follows that $v_2(\phi(u)) \leq k-1$ and $l = \omega(u) \leq k$. Note too that for any valid choice for u, we have $\omega(w\phi(u)) = \omega(au) \leq \omega(a)+\omega(u) \leq K+k$. Since $q_l \nmid \phi(u)$, we must have $q_l \mid w$, so there are at most $K+k$ choices for q_l. Once q_l is chosen, note that $q_{l-1} \nmid \phi(u/q_l)$, so $q_{l-1} \mid w\phi(q_l)$ and there are at most $K+k$ choices for q_{l-1}. In general, $q_{l-i} \mid w\phi(q_{l-i+1} \dots q_l)$, so that once $q_{l-i+1}, \dots, q_l$ are chosen, there are at most $K+k$ choices for q_{l-i}. Thus the number of choices for u is at most $\sum_{l=0}^{k}(K+k)^l \ll (K+k)^k$.

For any positive real number y put

$$V(y) = \{\phi(m) \ : \ m \text{ is a positive integer}, \ \phi(m) \leq y\}.$$

In 1988, Maier and Pomerance [15] showed that that there is a positive constant c_2 with

$$\#V(y) = \frac{y}{\log y} \exp\left((c_2 + o(1))(\log_3 y)^2\right) \qquad \text{as } y \to \infty.$$

(The exact order of magnitude of $V(y)$ has been determined by Ford in [9].) Therefore, from (5),

$$\#\mathcal{A}_3 \leq \frac{x}{\log x} \exp\left(O((\log_3 x)^2)\right)(K+k)^k.$$

We equate this bound for $\#\mathcal{A}_3$ with the bound for $\#\mathcal{A}_1$ in (3), leading us to choose

$$k = \left\lfloor \frac{\log_2 x}{\log_3 x} - 2\log 2 \frac{\log_2 x}{(\log_3 x)^2} \right\rfloor .$$

With this value of k we have (2) and the theorem. □

We remark that it may be the case that there is some positive constant c such that $\#\mathcal{A}(x) \ll x/(\log x)^c$, but we have not been able to prove this. Since $p-1 \in \mathcal{A}(x)$ for every prime $p \leq x+1$, we have $\#\mathcal{A}(x) \gg x/\log x$.

4 The function $g(b)$

The first result here addresses the maximal order of the function $g(b)$.

Theorem 4. *We have $g(b) \leq b^{(1+o(1))\log_3 b/\log_2 b}$ as $b \to \infty$.*

Proof. Let $a/b = \phi(n)/n$ for some squarefree integer n. Then $an = b\phi(n)$, therefore $n \mid b\phi(n)$. Let $r(n) = \mathrm{rad}(\phi(n))$ and let r_k be the k-fold iteration of r, with $r_1 = r$ and r_0 the identity. Note that if u is squarefree and $u \mid vw$, then $r(u) \mid r(v)r(w)$. Since $n \mid br(n)$, it follows by induction on $k \geq 0$, that $r_k(n) \mid r_k(b)r_{k+1}(n)$. Let $k(b)$ be the smallest positive integer k such that $r_k(b) = 1$. The above divisibility relation shows that $r_{k(b)}(n) \mid r_{k(b)+1}(n)$, which easily leads to the conclusion that $r_{k(b)}(n) = 1$. Hence, writing

$$F(b) = \prod_{0\leq k\leq k(b)} r_k(b), \tag{6}$$

we get

$$n \,\Big|\, br(n) \,\Big|\, br(b)r_2(n) \,\Big|\, \ldots \,\Big|\, F(b).$$

Thus, $a \mid \phi(n) \mid \phi(F(b))$ and also $a \leq b$. By a generalization of a result of Pratt, see Theorem 4.6 in [8], we have that $\omega(F(b)) < \log b/\log 2 + 1$. Put $t = \lfloor \log b/\log 2 \rfloor + 1$ and assume that b is large. Let $\mathcal{P}_b$ be the set of all prime factors of $F(b)$ and $\Psi_{\mathcal{P}_b}(x)$ denote the number of positive integers $a \leq x$ all whose prime factors are in $\mathcal{P}_b$. Then $g(b) \leq \Psi_{\mathcal{P}_b}(b)$. The prime number theorem (or estimates of Chebyshev) imply that $p_j \leq 2j\log j$ holds for all sufficiently large j. Put $P(n)$ for the largest prime factor of n. The above argument shows that if b is large enough, then

$$g(b) \leq \Psi_{\mathcal{P}_b}(b) \leq \Psi(b, p_t) \leq \Psi(b, 2t\log t), \tag{7}$$

where we use

$$\Psi(x, y) = \#\{n \leq x \ : \ P(n) \leq y\}.$$

By the de Bruijn estimates for the function $\Psi(x, y)$ (see, for example, Theorem 2 on p. 359 in [18]), we have

$$\Psi(b, 2t\log t) \leq \exp\left((1+o(1))\frac{\log b}{\log t}\log\left(1+\frac{2t\log t}{\log b}\right)\right). \tag{8}$$

Comparing estimates (7) and (8) and recalling the definition of t, we get the conclusion of the theorem as $b \to \infty$. $\square$

A natural question is the average order of $g(b)$, but we have not been able to substantially improve on the estimate afforded by Theorem 4. Using this theorem, we can get the following result for the average order of $f(a)$.

Corollary 5. *As $x \to \infty$, we have $\sum_{a \le x} f(a) \le x^{1+(1+o(1)) \log_3 x / \log_2 x}$.*

Proof. If $\gcd(a, b) = 1$ and $a/b = \phi(n)/n$ for some integer n, then for a large, (1) implies that $b \le 2a \log_2 a$. Thus, for x large,

$$\sum_{a \le x} f(a) \le \sum_{b \le 2x \log_2 x} g(b),$$

and so the result follows from Theoerem 4. □

Theorem 6. *We have $g(b) = 0$ for almost all positive integers b. In fact the number of integers $b \le x$ with $g(b) > 0$ is $\sim e^{-\gamma} x / \log_3 x$ as $x \to \infty$.*

Proof. From Lemma 1, if $g(b) > 0$, then $\gcd(b, \phi(b)) = 1$. Conversely, if $\gcd(b, \phi(b)) = 1$, then $\phi(b)/b$ is already reduced, so $g(b) > 0$. Thus the theorem follows immediately from the result of Erdős [7] quoted in the proof of Theorem 2. □

We remark that Theorem 6 is implicit in [2].

5 The greatest common divisor of n and $\phi(n)$

Our first result in this direction addresses the maximal order of the greatest common divisor of n and $\phi(n)$.

Theorem 7. *The inequality*

$$\gcd(n, \phi(n)) \le 2n \exp\left(-\sqrt{\log 2 \log n}\right)$$

holds for all squarefree $n \ge 1$. On the other hand, there is an infinite set $\mathcal{S}$ of squarefree numbers n such that

$$\gcd(n, \phi(n)) \ge n^{1-(1+o(1)) \log_3 n / \log_2 n}$$

as $n \to \infty$, $n \in \mathcal{S}$.

Proof. Assume that $n \geq 3$ is squarefree. Write $d = \gcd(n, \phi(n))$. Since n is squarefree, we have that $P(n) \nmid \phi(n)$, which shows that $d \leq n/P(n)$. Thus, the first inequality follows immediately if $P(n) > \exp(\sqrt{\log 2 \log n})$ even without the factor of 2 on the right hand side. On the other hand, if $P(n) \leq \exp(\sqrt{\log 2 \log n})$, then, using that n is squarefree, we get

$$\omega(n) \geq \frac{\log n}{\log P(n)} \geq \sqrt{\log n / \log 2}. \tag{9}$$

Let $\beta = v_2(\phi(n))$. If n is odd we have $\beta \geq \omega(n)$ and $d \mid \phi(n)/2^{\beta}$, so that $d \leq n/2^{\omega(n)}$. Thus (9) gives the first inequality, again without the factor of 2. Finally, if n is even, we have $d \mid \phi(n)/2^{\beta-1}$, since n is squarefree. But in this case $\beta \geq \omega(n) - 1$ and $\phi(n) \leq n/2$, so that $d \leq n/2^{\beta} \leq n/2^{\omega(n)-1}$. Using (9) gives the first inequality for d.

For the lower bound, note that from Theorem 3 in [14], there is a set of numbers $\mathcal{T}$ having asymptotic density 1, such that for $t \to \infty$, $t \in \mathcal{T}$, we have $\operatorname{rad}(F(t)/t) > t^{(1+o(1)) \log_2 t / \log_3 t}$, where $F(t)$ is defined in (6). Note too that for any squarefree number t we have $\gcd(\operatorname{rad}(F(t)), \phi(\operatorname{rad}(F(t))))$ equal to $\operatorname{rad}(F(t)/t)$. Let $\mathcal{S}$ be the set of numbers $n = \operatorname{rad}(F(t))$ for $t \in \mathcal{T}$ with t squarefree. The second inequality of the theorem follows. ☐

Our next result addresses the normal order of the greatest common divisor of n and $\phi(n)$.

Theorem 8. *For almost all n, $\gcd(n, \phi(n))$ is the largest divisor of n supported on the prime divisors of n in the interval $[1, \log\log n]$.*

Proof. Let x be a large positive real number. We first note that most numbers $n \leq x$ are "nearly" squarefree, in that $v_p(n) \leq 1$ for all primes $p \geq \log_3 x$. Indeed if $\mathcal{E}_0(x)$ is the set of integers n which violate this property, then

$$\#\mathcal{E}_0(x) \;\leq\; \sum_{p \geq \log_3 x} \frac{x}{p^2} \;\sim\; \frac{x}{\log_3 x \log_4 x} \;=\; o(x).$$

Assume $n \leq x$ and $n \notin \mathcal{E}_0(x)$. Let $\mathcal{E}_1(x)$ be the set of such n with $p \mid \gcd(n, \phi(n))$ for some prime $p \geq \log_2 x$. To bound $\#\mathcal{E}_1(x)$, we fix a prime $p \geq \log_2 x$ and look at the number of $n \leq x$ such that $p \mid \gcd(n, \phi(n))$. It is clear that either $p^2 \mid n$ or $pq \mid n$ for some prime $q \equiv 1 \pmod{p}$. Since

$n \notin \mathcal{E}_0(x)$ the first choice does not occur. Thus,

$$\begin{aligned} \#\mathcal{E}_1(x) &\le \sum_{p \ge \log_2 x} \sum_{\substack{q \le x/p \\ q \equiv 1 \pmod p}} \left\lfloor \frac{x}{pq} \right\rfloor \le x \sum_{p \ge \log_2 x} \frac{1}{p} \sum_{\substack{q \le x \\ q \equiv 1 \pmod p}} \frac{1}{q} \\ &\ll x \sum_{p \ge \log_2 x} \frac{\log_2 x}{p(p-1)} \ll \frac{x}{\log_3 x} = o(x), \end{aligned} \tag{10}$$

where we used the estimate

$$\sum_{\substack{q \le x \\ q \equiv 1 \pmod b}} \frac{1}{q} \ll \frac{\log_2 x}{\phi(b)}, \tag{11}$$

which is uniform for $2 \le b \le x$ (see, for example, inequality (3.1) in [8]). Thus we may assume that $n \notin \mathcal{E}_1(x)$.

In Lemma 2 in [13] it is shown that there exists a positive constant c_3 such that if we put $M(x)$ for the least common multiple of all numbers $m \le B := c_3 \log_2 x / \log_3 x$, then all numbers $n \le x$ have the property that $M(x) \mid \phi(\mathrm{rad}(n)) \mid \phi(n)$ except for a set $\mathcal{E}_2(x)$ of them of cardinality $o(x)$.

Consider the properties:

(i) there is a prime factor p of n in the interval $I := [B, \log_2 x]$;

(ii) there is a prime $p \le B$ and a positive integer a such that $p^a > B$ and $p^a \mid n$.

Let $\mathcal{E}_3(x)$ and $\mathcal{E}_4(x)$ be the subsets of integers $n \le x$ not in $\mathcal{E}_0(x)$ satisfying (i) and (ii), respectively. It is easy to see that if $e^{e^B} \le n \le x$ and n is not in $\mathcal{E}_j(x)$ for $j = 0, \ldots, 4$, then the largest divisor of n supported on the primes in $[1, \log_2 n]$ is $\gcd(n, M(x))$ and this divisor is equal to $\gcd(n, \phi(n))$. Since $e^{e^B} = o(x)$ it will suffice to show then that $\#\mathcal{E}_j(x) = o(x)$ for $j = 3, 4$.

A simple computation with Mertens's theorem shows that $\sum_{p \in I} 1/p = o(1)$, so $\#\mathcal{E}_3(x) \le \sum_{p \in I} x/p = o(x)$.

We now bound $\#\mathcal{E}_4(x)$. Note that $a \ge 2$, so that $n \notin \mathcal{E}_0(x)$ implies that $p < \log_3 x$. Then $B < p^a < (\log_3 x)^a$, which implies that $a \ge K := \lceil \log B / \log_4 x \rceil$ if x is sufficiently large. Thus, n is a multiple of p^K for some prime p, so the number of such $n \le x$ does not exceed

$$\sum_{p \ge 2} \frac{x}{p^K} \le x(\zeta(K) - 1) \ll \frac{x}{2^K} = o(x).$$

Thus $\#\mathcal{E}_4(x) = o(x)$, which, together with our estimates for the other counts $\#\mathcal{E}_j(x)$ for $j = 0, 1, 2, 3$, completes the proof of the theorem. □

Theorem 8 has the following consequences.

Corollary 9. *The normal order of* $\omega(\gcd(n, \phi(n)))$ *is* $\log_4 n$.

Proof. This result follows from Theorem 8 and the fact that the normal order for the number of prime factors of n below $\log_2 n$ is $\log_4 n$, see Theorem 8 on p. 312 of [18]. □

Note that Corollary 9 was stated without proof in [7].

Corollary 10. *For each real number* $u > 0$*, the asymptotic density of the set of natural numbers* n *with*

$$\gcd(n, \phi(n)) > (\log\log n)^u$$

is $e^{-\gamma} \int_u^\infty \rho(t)\, dt$*, where* ρ *is the Dickman–de Bruijn function.*

Proof. In light of Theorem 8 we may replace the function $\gcd(n, \phi(n))$ in the corollary with the function $D(n)$, defined as the largest divisor of n supported on the prime factors of n in the interval $[1, \log\log n]$. Let $y = y(x) = \log\log x$ and let $D_y(n)$ be the largest y-smooth divisor of n. That is, $D_y(n)$ is the largest divisor of n supported on the prime factors of n in $[1, y]$. Since the function $\log\log x$ grows so slowly, it suffices to show that the number of n in $[1, x]$ with $D_y(x) > y^u$ is $\sim \delta_u x$, where $\delta_u = e^{-\gamma} \int_u^\infty \rho(t)\, dt$.

First note that the number of $n \le x$ with $D_y(n) \ge x^{1/2}$ is $o(x)$. Fix a positive real number u. Then the number of integers $n \le x$ with $D_y(n) > y^u$ is

$$\sum_{m > y^u,\, P(m) \le y} \ \sum_{n \le x,\, D_y(n) = m} 1 = \sum_{x^{1/2} > m > y^u,\, P(m) \le y} \ \sum_{n \le x,\, D_y(n) = m} 1 \ + \ o(x).$$

Let L denote the product of the primes in $[1, y]$. The inner sum above is

$$\sum_{k \le x/m,\, \gcd(k,L) = 1} 1 = \left(\frac{\phi(L)}{L} + o(1) \right) \frac{x}{m}$$

uniformly for all m in consideration, where we use a complete inclusion-exclusion over the primes in L. By the theorem of Mertens, we have $\phi(L)/L \sim e^{-\gamma}/\log y$, so that our count is

$$(1+o(1))\frac{x}{e^{\gamma}\log y}\sum_{x^{1/2}>m>y^u,\, P(m)\le y}\frac{1}{m}.$$

Our corollary then follows by partial summation and standard results on the distribution of y-smooth integers m. □

Our last result in this section addresses the average value of $\gcd(n,\phi(n))$.

Theorem 11. *Let*

$$A(x)=\frac{1}{x}\sum_{n\le x}\gcd(n,\phi(n)).$$

Then, for any $k>0$ we have

$$(\log x)^k \le A(x) \le x^{(1+o(1))\log_3 x/\log_2 x}$$

as $x\to\infty$.

Proof. We start with the upper bound since it is easier. We have

$$\sum_{n\le x}\gcd(n,\phi(n)) = \sum_{\substack{r\le x\\ r \text{ squarefree}}}\ \sum_{\substack{m\le x/r\\ \mathrm{rad}(m)\mid r}}\gcd(rm,\phi(rm))$$

$$= \sum_{\substack{r\le x\\ r \text{ squarefree}}}\gcd(r,\phi(r))\sum_{\substack{m\le x/r\\ \mathrm{rad}(m)\mid r}} m \ \le\ x\sum_{\substack{r\le x\\ r \text{ squarefree}}}\frac{\gcd(r,\phi(r))}{r}\sum_{\substack{m\le x\\ \mathrm{rad}(m)\mid r}}1.$$

We estimate the inner sum, call it $S(r)$. It is majorized by replacing r with $P_j=p_1\dots p_j$ where $j=\omega(r)$, so we have $S(r)\le S(P_j)=\Psi(x,p_j)$. Since $p_j\le(1+o(1))\log x$, it follows from Theorem 2 (of de Bruijn) on p. 359 of [18] that $S(r)\le x^{(\log 4+o(1))/\log_2 x}$ as $x\to\infty$, uniformly in r. Using this, we have for large x,

$$\begin{aligned}\sum_{n\le x}\gcd(n,\phi(n)) &\le x^{1+2/\log_2 x}\sum_{\substack{r\le x\\ r\ \text{squarefree}}}\frac{\gcd(r,\phi(r))}{r}\\ &= x^{1+2/\log_2 x}\sum_{b\le x}\frac{1}{b}\sum_{\substack{r\le x\\ r\ \text{squarefree}\\ r/\gcd(r,\phi(r))=b}} 1\\ &= x^{1+2/\log_2 x}\sum_{b\le x}\frac{g(b)}{b} \le x^{1+(1+o(1))\log_3 x/\log_2 x},\end{aligned}$$

where for the last estimate we used Theorem 4. Dividing by x, we have the asserted upper bound for $A(x)$.

We now deal with the lower bound. Let $\mathcal{D}$ be the set of positive integers d such that

$$\sum_{\substack{p\le d^{10}\\ p\equiv 1 \pmod d}}\frac{1}{p}\ge\frac{1}{2d}.$$

We state the following lemma for future use.

Lemma 12. *The set $\mathcal{D}$ contains all positive integers except for at most $O(x/\log x)$ of them.*

Proof. Indeed, let x be large and let $d\in(x/\log x,x)$. Theorem 2.1 in [1] shows that there exists a constant c_4 such that the inequality

$$\pi(y;1,d)\ge\frac{\pi(y)}{2\phi(d)}$$

holds for large x and for all positive integers $d\in(x/\log x,x)$ uniformly in the range $y\in(x^{2.6},x^{10})$ except for a set $\mathcal{D}'(x)$ of such d, where $\mathcal{D}'(x)$ consists of all positive integers $d\in(x/\log x,x)$ which are divisible by one of at most c_4 positive integers $d_i=d_i(x)$, all of which exceed $\log x$. Certainly,

$$\#\mathcal{D}'(x)\le\sum_{i\le c_4}\left\lfloor\frac{x}{d_i}\right\rfloor\ll\frac{x}{\log x}.$$

Assume now that $d \notin \mathcal{D}'(x)$. Then $d^3 \ge (x/\log x)^3 \ge x^{2.6}$ when x is large, therefore

$$\sum_{\substack{p \le d^{10} \\ p \equiv 1 \pmod d}} \frac{1}{p} \ \ge \ \int_{d^3}^{d^{10}} \frac{d\pi(y;1,d)}{y} \ \ge \ \frac{\pi(y)}{2\phi(d)y}\Big|_{y=d^3}^{y=d^{10}} + \int_{d^3}^{d^{10}} \frac{\pi(y)dy}{2\phi(d)y^2}$$

$$= \ \frac{1}{2\phi(d)} \sum_{d^3 < p \le d^{10}} \frac{1}{p} \ = \ \frac{\log(10/3) + o(1)}{2\phi(d)} \ \ge \ \frac{1}{2d},$$

which shows that $(x/\log x, x)\backslash\mathcal{D}'(x) \subset \mathcal{D} \cap [1, x]$. This completes the proof. □

Assume that k is a positive integer. We put $\alpha_i = 2^{-1}11^{-2(i+2)}$ for $i = 1, \ldots, k+1$. Set $y = x^{\alpha_{k+1}}$. For each $i = 1, \ldots, k$, we put $\mathcal{I}_i = [x^{\alpha_i}, x^{10\alpha_i}]$. We consider k-tuples of positive integers $(d_1, \ldots, d_k)$ such that

(i) $d_i \in \mathcal{D} \cap \mathcal{I}_i$ for all $i = 1, \ldots, k$;

(ii) $P(d_i) \le y$ for all $i = 1, \ldots, k$;

(iii) $\omega(d_i) \le 2\log_2 x$ for all $i = 1, \ldots, k$;

(iv) $\gcd(d_i, d_j) = 1$ for all $i \ne j$;

(v) the smallest prime factor of d_i exceeds $(\log_2 x)^4$ for all $i = 1, \ldots, k$.

We now let $\mathcal{E}$ be the set of such k-tuples $(d_1, \ldots, d_k)$. For each k-tuple $\mathbf{d} = (d_1, \ldots, d_k)$ in $\mathcal{E}$ we consider the set $\mathcal{F}_{\mathbf{d}}$ of k-tuples of primes $\mathbf{p} = (p_1, \ldots, p_k)$ with $p_i \in \mathcal{P}_{\mathbf{d},i}$, where $\mathcal{P}_{\mathbf{d},i}$ consists of primes with the following properties:

(i) $p_i \le d_i^{10}$;

(ii) $p_i \equiv 1 \pmod{d_i}$;

(iii) $\omega(p_i - 1) \le 10\log_2 x$;

(iv) $p_i - 1$ is coprime to $\prod_{j \ne i} d_j$, for all $i = 1, \ldots, k$. Finally, for a fixed k-tuple $\mathbf{d}$ in $\mathcal{E}$ and a fixed k-tuple of primes $\mathbf{p}$ in $\mathcal{F}_{\mathbf{d}}$ consider the set $\mathcal{G}_{\mathbf{d},\mathbf{p}}$ of all positive integers

$$m \in \left[\frac{x}{2\prod_{i=1}^{k} d_i p_i}, \frac{x}{\prod_{i=1}^{k} d_i p_i} \right].$$

which are coprime to $(p_1 - 1) \ldots (p_k - 1)$ and have $P(m) \leq y$.

We shall first prove that

Lemma 13. *The estimate*

$$\#\mathcal{G}_{\mathbf{d},\mathbf{p}} \gg_k \frac{x}{\log_2 x \prod_{i=1}^{k} d_i p_i}.$$

holds for all $\mathbf{d} \in \mathcal{E}$ *and* $\mathbf{p} \in \mathcal{F}_{\mathbf{d}}$.

Proof. For a positive integer Q let

$$\Psi_Q(x, y) = \sum_{\substack{n \leq x \\ P(n) \leq y \\ \gcd(n,Q)=1}} 1$$

denote the number of y-smooths in $[1, x]$ coprime to Q. We fix $\mathbf{d}$ and $\mathbf{p}$ and put $P = \prod_{i=1}^{k} d_i p_i$ and $Q = \prod_{i=1}^{k} (p_i - 1)$. Then

$$\#\mathcal{G}_{\mathbf{d},\mathbf{p}} = \Psi_Q(x/P, y) - \Psi_Q(x/2P, y).$$

Note that

$$Q \leq P \leq (d_1 \ldots d_k)^{11} \leq x^{11(\alpha_1 + \ldots + \alpha_k)} = x^{(1 + 11^{-2} + \ldots + 11^{-2(k-1)})/242} < x^{1/240}.$$

It thus follows from Theorem 1 in [10] that

$$\#\mathcal{G}_{\mathbf{d},\mathbf{p}} = \frac{\phi(Q)}{Q} (\Psi(x/P, y) - \Psi(x/2P, y)) \left(1 + O_k\left((\log_2 x)^2 / \log x\right)\right).$$

We have $\phi(Q)/Q \gg 1/\log_2 x$ and $\Psi(x/P, y) - \Psi(x/2P, y) \gg_k x/P$, so the result follows. □

We now look at numbers of the form $n = d_1 \dots d_k p_1 \dots p_k m$, where $\mathbf{d} = (d_1, \dots, d_k) \in \mathcal{E}$, $\mathbf{p} \in \mathcal{F}_{\mathbf{d}}$ and $m \in \mathcal{G}_{\mathbf{d},\mathbf{p}}$. Clearly, $n \in (x/2, x)$. We show that each such n arises from a unique triple $(\mathbf{d}, \mathbf{p}, m)$. Note first that since

$$p_i \le d_i^{10} < x^{100\alpha_i} < x^{\alpha_{i-1}} \le d_{i-1} < p_{i-1}$$

holds for all $i \ge 2$, while $y = x^{1/2(11)^{2(k+3)}} < d_k$, it follows that $p_1 > p_2 > \dots > p_k > y \ge P(d_1 \dots d_k m)$. Now assume that the same n arises also from $\mathbf{d}' = (d_1', \dots, d_k')$, $\mathbf{p}' = (p_1', \dots, p_k')$ and m'. Then $d_1 \dots d_k p_1 \dots p_k m = d_1' \dots d_k' p_1' \dots p_k' m'$. By identifying the k largest prime factors of n, we get that $p_i = p_i'$ for $i = 1, \dots, k$. Once $p_1, \dots, p_k$ are known, then $d_1 \dots d_k = \gcd(n, (p_1 - 1) \dots (p_k - 1))$, so $m = m'$. Thus, $d_1 \dots d_k = d_1' \dots d_k'$ and the unicity of the d_i's follows from the fact that $d_i = \gcd(d_1, \dots, d_k, p_i - 1)$. Hence, each such n arises from a unique triple $(\mathbf{d}, \mathbf{p}, m)$.

Note now that if n arises from $(\mathbf{d}, \mathbf{p}, m)$, then $d_1 \dots d_k \mid \gcd(n, \phi(n))$. Thus, we get

$$\begin{aligned} A(x) \quad &\ge \quad x^{-1} \sum_{\mathbf{d} \in \mathcal{E}} d_1 \dots d_k \sum_{\mathbf{p} \in \mathcal{F}_{\mathbf{d}}} \sum_{m \in \mathcal{G}_{\mathbf{d},\mathbf{p}}} 1 \;=\; x^{-1} \sum_{\mathbf{d} \in \mathcal{E}} d_1 \dots d_k \sum_{\mathbf{p} \in \mathcal{F}_{\mathbf{d}}} \#\mathcal{G}_{\mathbf{d},\mathbf{p}} \\ &\gg_k \quad \frac{1}{\log_2 x} \sum_{\mathbf{d} \in \mathcal{E}} \sum_{\mathbf{p} \in \mathcal{F}_{\mathbf{d}}} \frac{1}{p_1 \dots p_k} \;=\; \frac{1}{\log_2 x} \sum_{\mathbf{d} \in \mathcal{E}} \prod_{i=1}^{k} \left(\sum_{p \in \mathcal{P}_{\mathbf{d},i}} \frac{1}{p} \right), \end{aligned} \tag{12}$$

where in the above chain of inequalities we used Lemma 13.

We shall also need the following two lemmas whose proofs we postpone for the moment.

Lemma 14. *There exists x_k such that the inequality*

$$\sum_{p \in \mathcal{P}_{\mathbf{d},i}} \frac{1}{p} \ge \frac{1}{3d_i}$$

holds for all $i = 1, \dots, k$ and all $\mathbf{d} \in \mathcal{E}$ provided that $x \ge x_k$.

Lemma 15. *There exists x_k and $\beta(k) > 0$ such that the inequality*

$$\sum_{\mathbf{d} \in \mathcal{E}} \frac{1}{d_1 \dots d_k} \ge \frac{1}{2} \left(\beta(k) \frac{\log x}{\log_3 x} \right)^k$$

holds for all $x \ge x_k$.

Clearly, estimate (12) and Lemmas 14 and 15 show that if x is sufficiently large (with respect to k), then

$$\begin{aligned} A(x) &\gg_k \frac{1}{\log_2 x}\sum_{\mathbf{d}\in\mathcal{E}}\prod_{i=1}^{k}\left(\sum_{p\in\mathcal{P}_{\mathbf{d},i}}\frac{1}{p}\right) && \text{(by (12))}\\ &\gg \frac{1}{3^k\log_2 x}\sum_{\mathbf{d}\in\mathcal{E}}\frac{1}{d_1\dots d_k} && \text{(by Lemma 14)}\\ &\gg_k \frac{1}{\log_2 x}\left(\beta(k)\frac{\log x}{\log_3 x}\right)^k && \text{(by Lemma 15)}\\ &\geq (\log x)^{k/2} \end{aligned}$$

if x is sufficiently large with respect to k, which completes the proof of the lower bound asserted by Theorem 11, since k was arbitrary. □

It remains to prove Lemmas 14 and 15.

Proof of Lemma 14. Fix $\mathbf{d}$ in $\mathcal{E}$ and $i = 1, \dots, k$. We let $\mathcal{P}_1$ be the set of primes satisfying conditions (i) and (ii) from the definition of the p_i's, $\mathcal{P}_2$ be the subset of $\mathcal{P}_1$ failing (iii) and $\mathcal{P}_3$ be the subset of $\mathcal{P}_1$ failing (iv). From the fact that $d_i \in \mathcal{D}$, we have that

$$S_1 = \sum_{p\in\mathcal{P}_1}\frac{1}{p} \geq \frac{1}{2d_i}.$$

Since

$$\sum_{p\in\mathcal{P}_{\mathbf{d},i}}\frac{1}{p} \geq S_1 - S_2 - S_3,$$

where

$$S_j = \sum_{p\in\mathcal{P}_j}\frac{1}{p}$$

for $j = 2,\ 3$, it suffices to show that both $S_2 = o(d_i^{-1})$ and $S_3 = o(d_i^{-1})$ hold as x goes to infinity. If $p \in \mathcal{P}_2$, it then follows that $p - 1 = d_i m$, where $\omega(m) \geq \omega(p-1) - \omega(d_i) \geq 6\log_2 x$ because of condition (iii) on the d_i. Putting $K = \lfloor 6\log_2 x\rfloor$ and using the multinomial formula, unique factorization, the Stirling formula and the known fact that

$$\sum_{p^\alpha\leq t}\frac{1}{p^\alpha} = \log_2 t + O(1)$$

uniformly in $t \geq 3$, we get that

$$\begin{aligned} S_2 &= \sum_{p \in \mathcal{P}_2} \frac{1}{p} \leq \sum_{p \in \mathcal{P}_2} \frac{1}{p-1} \leq \frac{1}{d_i} \sum_{\substack{m \leq x \\ \omega(m) \geq K}} \frac{1}{m} \\ &= \frac{1}{d_i} \sum_{\ell \geq K} \sum_{\substack{m \leq x \\ \omega(m) = \ell}} \frac{1}{m} \leq \frac{1}{d_i} \sum_{\ell \geq K} \frac{1}{\ell!} \left(\sum_{p^\alpha \leq x} \frac{1}{p^\alpha} \right)^\ell \\ &= \frac{1}{d_i} \sum_{\ell \geq K} \left(\frac{e \log_2 x + O(1)}{\ell} \right)^\ell \leq \frac{1}{d_i} \sum_{\ell \geq K} \frac{1}{2^\ell} \ll \frac{1}{2^K d_i} = o(d_i^{-1}) \end{aligned}$$

as $x \to \infty$, where we used also the fact that $6 > 2e$, therefore $(e \log_2 x + O(1))/K < 1/2$ if x is sufficiently large.

For S_3, let $\mathcal{Q}_i$ be the set of prime factors of $\prod_{j \neq i} d_j$. Note that $\#\mathcal{Q}_i \leq 2(k-1) \log_2 x$ (by the property (iii) on the d_j's), and that if q_1 is the smallest element in Q_i then $q_1 \geq (\log_2 x)^4$ (by the property (v) on the d_j's). If $p \in \mathcal{P}_3$, there exists $q \in \mathcal{Q}_i$ such that $p \equiv 1 \pmod{q}$. Since q divides d_j for some $j \neq i$, it does not divide d_i (by the property (iv) of the d_i's), so $p \equiv 1 \pmod{d_i q}$. Hence,

$$\begin{aligned} S_3 &\leq \sum_{q \in \mathcal{Q}_i} \sum_{\substack{p \leq x \\ p \equiv 1 \pmod{d_i q}}} \frac{1}{p} \ll \sum_{q \in \mathcal{Q}_i} \frac{\log_2 x}{\phi(d_i q)} = \frac{\log_2 x}{\phi(d_i)} \sum_{q \in \mathcal{Q}_i} \frac{1}{q-1} \\ &\leq \frac{(\log_2 x)^2 \#\mathcal{Q}_i}{d_i(q_1 - 1)} \ll_k \frac{(\log_2 x)^3}{d_i (\log_2 x)^4} = \frac{1}{d_i \log_2 x} = o(d_i^{-1}) \end{aligned}$$

as $x \to \infty$, where we used aside from the minimal order $\phi(d_i)/d_i \geq 1/\log_2 x$ of the Euler function for $d_i \in [1, x]$ also the estimate (11). □

Proof of Lemma 15. We let $\mathcal{D}_i$ be the set of all positive integers d_i satisfying all the properties (i)–(v) *except* (iv). We shall later show that

$$T_i := \sum_{d \in \mathcal{D}_i} \frac{1}{d} \asymp_k \frac{\log x}{\log_3 x}. \tag{13}$$

Assuming estimate (13), let $\mathcal{E}_1 = \times_{i=1}^k \mathcal{D}_i$ and $\mathcal{E}_2$ be the subset of $\mathcal{E}_1$ consisting in k-tuples $\mathbf{d} = (d_1, \ldots, d_k)$ such that there exist $i \neq j$ with $\gcd(d_i, d_j) \neq 1$.

It is clear that $\mathcal{E} = \mathcal{E}_1 \backslash \mathcal{E}_2$, so

$$\sum_{\mathbf{d}\in\mathcal{E}} \frac{1}{d_1 \dots d_k} = S_1 - S_2, \tag{14}$$

where $S_j = \sum_{\mathbf{d}\in\mathcal{E}_j} (d_1 \dots d_k)^{-1}$ for $j = 1,\ 2$. By estimate (13), we have

$$S_1 = \prod_{i=1}^{k} T_i \asymp_k \left(\frac{\log x}{\log_3 x}\right)^k.$$

Furthermore, note that

$$\begin{aligned}
S_2 &\le \sum_{i\ne j}\Bigg(\prod_{\substack{1\le \ell\le k\\ \ell\ne i,j}} T_\ell\Bigg)\Bigg(\sum_{p\ge(\log_2 x)^4}\ \sum_{\substack{d_i\le x\\ p\mid d_i}}\ \sum_{\substack{d_j\le x\\ p\mid d_j}} \frac{1}{d_i d_j}\Bigg)\\
&\ll_k \left(\frac{\log x}{\log_3 x}\right)^{k-2}\Bigg(\sum_{p\ge(\log_2 x)^4}\frac{1}{p^2}\Bigg)\Bigg(\sum_{u\le x}\frac{1}{u}\Bigg)^2\\
&\ll \left(\frac{\log x}{\log_3 x}\right)^{k-2}\left(\frac{1}{(\log_2 x)^4\log_3 x}\right)(\log x)^2\\
&= \frac{(\log x)^k}{(\log_3 x)^{k-1}(\log_2 x)^4} = O\left(\frac{S_1 \log_3 x}{(\log_2 x)^4}\right) = o(S_1),
\end{aligned}$$

as $x \to \infty$, which together with inequality (14) completes the proof of the lemma. Hence, it suffices to prove estimate (13).

Let $i = 1, \dots, k$ be fixed. Let $R = \prod_{q\le(\log_2 x)^3} q$. As in the proof of Lemma 13, one checks that if we put $\mathcal{D}'_i(t)$ for the set of positive integers in $[1, t]$ satisfying (ii) and (v), then

$$\#\mathcal{D}'_i(t) = \rho(v_t)\frac{\phi(R)}{R}t\left(1 + O\left(\frac{(\log_2 x)^3}{\log x}\right)\right)$$

uniformly for $t \in \mathcal{I}_i$, where $v_t = (\log t)/\log y)$. Of these numbers, the number of numbers that fail (iii) are $O_k(x/(\log x)^{c_3})$ with $c_3 = 1 - 2\log(2/e) = 0.38629436122\dots$, while the number of numbers that fail (i) is, by Lemma 12, $O(t/\log t) = O_k(t/\log x)$. Hence,

$$\#\mathcal{D}_i(t) = \rho(v_t)\frac{\phi(R)}{R}t\left(1 + O_k\left(\frac{(\log_2 x)^3}{\log x}\right)\right) \asymp_k \frac{t}{\log_3 x},$$

where the last estimate above follows from Mertens's formula. By partial summation, we get that

$$\begin{aligned} T_i &= \sum_{d \in \mathcal{D}_i} \frac{1}{d} = \int_{x^{\alpha_i}}^{x^{10\alpha_i}} \frac{d\#\mathcal{D}(t)}{t} \\ &\geq -1 + \int_{x^{\alpha_i}}^{x^{10\alpha_i}} \frac{\#\mathcal{D}(t)dt}{t^2} \asymp_k \frac{\log t}{\log_3 x}\Big|_{t=x^{\alpha_i}}^{t=x^{10\alpha_i}} \asymp_k \frac{\log x}{\log_3 x}, \end{aligned}$$

which completes the proof of Lemma 15 and so the proof of Theorem 11. □

We remark that an examination of the proof (and the tool we used from [10]) shows that we may take k in the theorem of size $c \log_3 x$ for some small positive constant c, and so obtain the lower bound $A(x) > (\log x)^{c \log_3 x}$. We are not sure what to suggest for the true order of $A(x)$.

6 An application

For a nonzero polynomial $f(X) \in \mathbb{Z}[X]$, let $\mathcal{L}_f(x)$ denote the set of integers $n \leq x$ with $\phi(n) \mid f(n)$. In the recent paper [4], it was shown that there are certain "canonical" solutions to this relation of the form pm where m is a positive integral root of f and p is a prime in one of a few prescribed residue classes mod $\phi(m)$, and that non-canonical solutions are few in number. In particular, it was shown that

Theorem 16. *Let $f(X) \in \mathbb{Z}[X]$ be a polynomial of degree k, with $f(0) \neq 0$, whose roots all have multiplicity at most ν. For each positive integral root m of $f(X)$, there exist certain residue classes $\{\alpha_{j,m} \pmod{\phi(m)} : j = 1, \ldots, r_m\}$ for which the following estimate holds:*

$$\begin{aligned} \#\mathcal{L}_f(x) = \sum_{\substack{m \in \mathbb{N} \\ f(m)=0}} \sum_{j=1}^{r_m} \pi(x/m; \phi(m), \alpha_{j,m}) \\ + O\left(x^{1-1/(2\nu+1)+o(1)} + x^{1-1/(k+1)+o(1)}\right), \end{aligned} \tag{15}$$

where the functions implied by $o(1)$ and the constant implied by O depend only on f.

In the case $f(0) = 0$ a weaker error estimate of the shape $x^{1-o(1)}$ was obtained in [4]. Here, we use the results from Section 5 to show that Theorem 16 holds without the assumption that $f(0) \neq 0$.

Theorem 17. *The conclusion of Theorem 16 holds even when $f(0) = 0$.*

Proof. We follow the proof of Theorem 5 in [4]. Put

$$\alpha = \frac{\nu}{2\nu + 1}.$$

It is shown in the proof of Theorem 5 in [4] that the number of $n \le x$ with $P(n) > x^\alpha$ for which $\phi(n) \mid f(n)$ is given by the right hand side of formula (15), and this argument from [4] does not use the fact that $f(0) \ne 0$. Thus, the theorem will be proved if we show that the number $N(x)$ of $n \le x$ with $P(n) \le x^\alpha$ for which $\phi(n) \mid f(n)$ satisfies

$$N(x) \le x^{1-\alpha/\nu+o(1)}. \tag{16}$$

Such a number n clearly has a divisor m with $x^\alpha < m \le x^{2\alpha}$. We fix a number m in this interval and ask how many integers $r \le x/m$ there are with $\phi(mr) \mid f(mr)$. Let this count be denoted $N_m(x)$.

We assume that $f(0) = 0$. Let $F(X)$ denote the product of the distinct irreducible factors (in $\mathbb{Z}[X]$) of $f(X)$ and the content of f (the gcd of the coefficients of f). Let

$$t(m) = \prod_{p^k \| \phi(m)} p^{\lceil k/\nu \rceil},$$

so that $\phi(m) \mid \phi(mr)$ and $\phi(mr) \mid f(mr)$ imply that $t(m) \mid F(mr)$. Also, let $F(X) = XG(X)$, where $G(X) \in \mathbb{Z}[X]$, and let

$$s(m) = t(m)/\gcd(m, t(m)).$$

Then $t(m) \mid mrG(mr)$ implies that $s(m) \mid rG(mr)$. For $d \mid s(m)$, we consider separately those r with $\gcd(r, s(m)) = d$. For these values of r we have $s(m) \mid rG(mr)$ if and only if $s(m)/d \mid G(mr)$. We let $r = du$, where $u \le x/md$ and u is coprime to $s(m)/d$. Dropping the coprimality condition, let

$$N_{m,d}(x) \ = \ \#\{u \le x/md \ : \ s(m)/d \mid G(mdu)\},$$

so that

$$N_m(x) \le \sum_{d \mid s(m)} N_{m,d}(x).$$

We use the Nagell–Ore theorem, a strong form of it appearing in [12]. It asserts that for a squarefree polynomial $g(X)$ in $\mathbb{Z}[X]$, the number of solutions to the congruence $g(u) \equiv 0 \pmod{n}$ in a complete residue system mod n is bounded above by $\deg(g)^{\omega(n)} D(g)^2$, where $D(g)$ is the discriminant of g. Say p is a prime and $p^a \| s(m)/d$. If p does not divide md, then the number of solutions to $G(mdu) \equiv 0 \pmod{p^a}$ in a complete residue system mod p^a is equal to the number of solutions of $G(u) \equiv 0 \pmod{p^a}$, which is $O(1)$, by the Nagell–Ore theorem, the constant depending on the polynomial $G(X)$ (which in turn depends on $f(X)$). So, say p does divide md, say $p^b \| md$. If $G(mdu) \equiv 0 \pmod{p^a}$ has any solutions at all, we must have $p^{\min\{a,b\}} \mid G(0)$. Since $G(0) \neq 0$, we have that p is in a finite set depending on $G(X)$, and that $\min\{a, b\}$ is bounded as well. If $a = \min\{a, b\}$, we take p^a as the (trivial) upper bound for the number of solutions to $G(mdu) \equiv 0 \pmod{p^a}$. If $b = \min\{a, b\}$, we consider the polynomial $H(X) = G(p^b X)$ and again use the Nagell–Ore theorem. The discriminant of H is $p^{b(l^2-l)}$ times the discriminant of G, where $l = \deg G$, so that the number of solutions to $H(u) \equiv 0 \pmod{p^a}$ is bounded by $O(p^{2bl^2}) = O(1)$, again with the constant depending ultimately on f. But mdp^{-b} is coprime to p, so the number of solutions to $H(mdp^{-b}u) \equiv 0 \pmod{p^a}$ is exactly the same quantity as with $H(u)$, and it only remains to note that $H(mdp^{-b}X) = G(mdX)$. So, in each case, the number of solutions to $G(mdu) \equiv 0 \pmod{p^a}$ is at most some constant C that depends only on f. Thus, by the Chinese remainder theorem,

$$N_{m,d}(x) \;\le\; C^{\omega(s(m)/d)} \left(\frac{x}{mds(m)/d} + 1 \right) \;=\; C^{\omega(s(m)/d)} \left(\frac{x}{ms(m)} + 1 \right).$$

Thus,

$$N_m(x) \;\le\; \tau(s(m)) C^{\omega(s(m))} \left(\frac{x}{ms(m)} + 1 \right) \;=\; m^{o(1)} \left(\frac{x}{ms(m)} + 1 \right),$$

where we use the well-known maximal orders for the functions τ and ω.

Note that

$$s(m) \;\ge\; \frac{\phi(m)^{1/\nu}}{\gcd(m, \phi(m))} \;=\; \frac{m^{1/\nu+o(1)}}{\gcd(m, \phi(m))}.$$

Thus,

$$\begin{aligned} N(x) \;\le\; \sum_{x^\alpha < m \le x^{2\alpha}} N_m(x) \;&\le\; x^{o(1)} \sum_{x^\alpha < m \le x^{2\alpha}} \left(\frac{x \gcd(m, \phi(m))}{m^{1+1/\nu}} + 1 \right) \\ &=\; x^{1+o(1)} \sum_{x^\alpha < m \le x^{2\alpha}} \frac{\gcd(m, \phi(m))}{m^{1+1/\nu}} + x^{2\alpha+o(1)}. \end{aligned}$$

Let S denote the last sum above. To compute S, we use Theorem 11 and partial summation and get that it is equal to

$$x^{-2\alpha(1+1/\nu)} \sum_{x^\alpha < m \le x^{2\alpha}} \gcd(m,\phi(m)) + \int_{x^\alpha}^{x^{2\alpha}} \frac{1+1/\nu}{t^{2+1/\nu}} \sum_{x^\alpha < m \le t} \gcd(m,\phi(m))\, dt$$

$$\le\ x^{2\alpha+o(1)-2\alpha(1+1/\nu)} + x^{o(1)} \int_{x^\alpha}^{x^{2\alpha}} \frac{dt}{t^{1+1/\nu}} \ =\ x^{-\alpha/\nu+o(1)}.$$

Thus,

$$N(x)\ \le\ x^{1-\alpha/\nu+o(1)} + x^{2\alpha+o(1)}\ =\ x^{1-\alpha/\nu+o(1)}.$$

This last estimate establishes (16) and so completes the proof of the theorem.

□

References

[1] W. R. Alford, A. Granville and C. Pomerance, 'There are infinitely many Carmichael numbers', *Ann. Math.* **140** (1994), 703–722.

[2] C. W. Anderson, 'The solutions of $\Sigma(n) = \frac{\sigma(n)}{n} = \frac{a}{b}$, $\Phi(n) = \frac{\varphi(n)}{n} = \frac{a}{b}$, and related considerations', unpublished manuscript, 1974.

[3] R. C. Baker and G. Harman, 'Shifted primes without large prime factors', *Acta Arith.* **83** (1998), 331–361.

[4] W. D. Banks, F. Luca, and I. E. Shparlinski, 'Some divisibility properties of the Euler function', *Glasg. Math. J.* **47** (2005), 517–528.

[5] S. Contini, E. Croot, and I. E. Shparlinski, 'Complexity of inverting the Euler function', *Math. Comp.* **75** (2006), 983–996.

[6] P. Erdős, 'On the normal number of prime factors of $p-1$ and some other related problems concerning Euler's ϕ function', *Quart. J. Math. (Oxford Ser.)* **6** (1935), 205–213.

[7] P. Erdős, 'Some asymptotic formulas in number theory', *J. Indian Math. Soc. (N.S.)* **12** (1948), 75–78.

[8] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, 'On the normal behavior of the iterates of some arithmetic functions', *Analytic Number Theory*, Birkhäuser, Boston, 1990, 165–204.

[9] K. Ford, 'The distribution of totients', *Paul Erdős (1913–1996), Ramanujan J.* **2** (1998), 67–151.

[10] E. Fouvry and G. Tenenbaum, 'Entiers sans grand facteur premier en progressions arithmétiques', *Proc. London Math. Soc.* **63** (1991), 449–494.

[11] G. H. Hardy and S. Ramanujan, 'The normal number of prime factors of an integer', *Quart. J. Math. (Oxford Ser.)* **48** (1917), 79–92.

[12] M. N. Huxley, 'A note on polynomial congruences', *Recent Progress in Analytic Number Theory, Vol.1*, Academic Press, 1981, 193–196.

[13] F. Luca and C. Pomerance, 'On some problems of Mąkowski–Schinzel and Erdős concerning the arithmetical functions ϕ and σ', *Coll. Math.* **92** (2002), 111-130.

[14] F. Luca and C. Pomerance, 'Irreducible radical extensions and Euler-function chains,' *Integers*, to appear.

[15] H. Maier and C. Pomerance, 'On the number of distinct values of Euler's ϕ-function', *Acta Arith.* **49** (1988), 263–275.

[16] I. J. Schoenberg, 'Über die asymptotische Verteilung reeller Zahlen mod 1', *Math. Z.* **28** (1928), 171–199.

[17] W. Sierpiński, *Elementary Theory of Numbers*, North Holland, 1988.

[18] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.