Fibonacci Integers

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

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

Stephan Wagner
Department of Mathematical Sciences,
Stellenbosch University,
Private Bag X1,
Matieland 7602, South Africa
swagner@sun.ac.za

June 7, 2010

Abstract

A *Fibonacci integer* is an integer in the multiplicative group generated by the Fibonacci numbers. For example, $77 = 21 \cdot 55/(3 \cdot 5)$ is a Fibonacci integer. Using some results about the structure of this multiplicative group, we determine a near-asymptotic formula for the counting function of the Fibonacci integers, showing that up to x the number of them is between $\exp(c(\log x)^{1/2} - (\log x)^{\epsilon})$ and $\exp\left(c(\log x)^{1/2} + (\log x)^{1/6+\epsilon}\right)$, for an explicitly determined constant c. The proof is based on both combinatorial and analytic arguments.

1 Introduction

Let $(F_n)_{n\ge 1}$ be the Fibonacci sequence given by $F_1 = F_2 = 1$ and $F_{n+2} = F_{n+1} + F_n$ for all $n \ge 1$ and let $\mathcal{G}_F$ be the set of integers in the multiplicative group generated by $(F_n)_{n\ge 1}$ inside $\mathbb{Q}^*$. Hence, $\mathcal{G}_F$ consists of all integers which can be represented as a ratio of products of Fibonacci numbers. We call the members of $\mathcal{G}_F$ *Fibonacci integers*. The smallest positive integer n that is not a Fibonacci integer is 37, and the number of Fibonacci integers in $[1, 100]$ is 88. One might think then that most integers are Fibonacci integers, but this is not the case. For a positive real number x let $\mathcal{G}_F(x) = \mathcal{G}_F \cap [1, x]$ be the set of Fibonacci integers in $[1, x]$. In [5], it was shown that the estimate

$$\#\mathcal{G}_F(x) \ll_A \frac{x}{(\log x)^A} \qquad \text{holds for all } x \ge 2$$

with any constant A, where the implied constant above depends on A. Applying this with any $A > 1$, it was deduced in [5] that

$$\sum_{n\in\mathcal{G}_F} \frac{1}{n} < \infty.$$

Here, we improve on this estimate.

Theorem 1. *For each fixed $\epsilon > 0$, the estimate*

$$\exp\left(c(\log x)^{1/2} - (\log x)^{\epsilon}\right) \le \#\mathcal{G}_F(x) \le \exp\left(c(\log x)^{1/2} + (\log x)^{1/6+\epsilon}\right)$$

holds for all sufficiently large x, with

$$c = 2\zeta(2)\sqrt{\frac{\zeta(3)}{\zeta(6)\log\alpha}} = 5.15512\ldots,$$

where ζ is the Riemann zeta-function and $\alpha = (1+\sqrt{5})/2$ is the golden mean.

Our method is general and can be applied to any Lucas sequence of general term

$$u_n = \frac{a^n - b^n}{a - b} \quad \text{or} \quad v_n = a^n + b^n \qquad \text{for all} \quad n \ge 1,$$

where $a + b$, ab are nonzero integers and a/b is not a root of 1. Write $\mathcal{G}_u$ and $\mathcal{G}_v$ for the positive integers in the multiplicative groups generated by $\{u_n\}_{n\geq 1}$ and $\{v_n\}_{n\geq 1}$ inside $\mathbb{Q}^*$, respectively, and for a positive real number x write $\mathcal{G}_u(x)$ and $\mathcal{G}_v(x)$ for the intersection with $[1,x]$ of $\mathcal{G}_u$ and $\mathcal{G}_v$, respectively. Then, at least assuming that a and b are real, both $\#\mathcal{G}_u(x)$ and $\#\mathcal{G}_v(x)$ obey estimates of the same shape as what is shown in Theorem 1, where in the formula for the constant c we need to replace $(1+\sqrt{5})/2$ by $\max\{|a|,|b|\}$. Perhaps this is still true even when a and b are complex conjugates but we have not worked out the details for this situation. To simplify the presentation, we shall deal only with the Fibonacci sequence.

2 Arithmetic considerations

Let $\alpha = (1+\sqrt{5})/2$ and $\beta = (1-\sqrt{5})/2$ be the two roots of the characteristic equation $X^2 - X - 1 = 0$ of the Fibonacci sequence. Then it is well-known that

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} \qquad \text{holds for all} \quad n \geq 1.$$

For a positive integer m write

$$\Phi_m(X) = \prod_{\substack{1\leq k\leq m\\ (k,m)=1}} (X - \exp(2\pi ik/m)) \in \mathbb{Z}[X]$$

for the m-cyclotomic polynomial and let

$$\Phi_m(X,Y) = \prod_{\substack{1\leq k\leq m\\ (k,m)=1}} (X - \exp(2\pi ik/m)Y) \in \mathbb{Z}[X,Y]$$

be its homogenization. Further, let Φ_m stand for $\Phi_m(\alpha,\beta)$. Note that for $m > 1$, Φ_m is an integer. We have

$$F_m = \prod_{\substack{d|m\\ d>1}} \Phi_d \tag{1}$$

and by Möbius inversion, we have, for $m > 1$,

$$\Phi_m = \prod_{d|m} F_{m/d}^{\mu(d)}, \tag{2}$$

where μ is the Möbius function. In particular, $\Phi_m \in \mathcal{G}_F$ when $m > 1$. Formula (1) shows that the numbers Φ_m for $m > 1$ generate the same group as the Fibonacci numbers. It turns out that this group is almost freely generated by the numbers $(\Phi_m)_{m>1}$. This is not exactly so because of the exceptions

$$\Phi_2 = 1, \quad \Phi_6 = \frac{F_6}{F_2 F_3} = 2^2 = \Phi_3^2, \quad \Phi_{12} = \frac{F_{12} F_2}{F_6 F_4} = 6 = \Phi_3 \cdot \Phi_4. \tag{3}$$

The numbers Φ_m capture the so-called *primitive* prime divisors of F_m, namely those primes p dividing F_m that do not divide F_n for any $n < m$. It is known that each F_m, when $m \neq 1, 2, 6, 12$, has at least one primitive prime factor. Let Ψ_m be the product of the primitive prime factors of F_m with the corresponding exponents as they appear in F_m. For $m > 1$, we have $\Psi_m \mid \Phi_m$. To investigate the quotient $\delta_m = \Phi_m/\Psi_m$, note that for every positive integer k there exists some integer n such that $k \mid F_n$. We write $z(k)$ for the smallest such n (the *index of appearance* of k in the Fibonacci sequence). Thus, we can rephrase the condition that p is primitive for F_m as $z(p) = m$ and the definition of Ψ_m as

$$\Psi_m = \prod_{\substack{p^{a_p} \| F_m \\ z(p) = m}} p^{a_p}.$$

Further, $\delta_m = \Phi_m/\Psi_m = 1$ except if

$$m = p^k z(p) \text{ for some } k \geq 1 \text{ and prime } p.$$

If $m = p^k z(p)$ and $m \neq 12$, then $\delta_m = p$.

All of the above properties can be found in either Section 2 of [1], or in [6]. If p is prime and $z(p) = m$, then $p \equiv 0, \pm 1 \pmod{m}$, and the sign in fact equals the Legendre symbol $\left(\frac{p}{5}\right)$. In particular, we see that if $\delta_m = p$ is an odd prime, then $p = P(m)$, where we write $P(m)$ for the largest prime factor of m. Thus, if m ($\neq 12$) is of the form $p^k z(p)$, then it is uniquely of this form; that is, p and k are determined.

Let $\mathcal{M} = \mathbb{N} \setminus \{1, 2, 6, 12\}$. The result that F_m has a primitive prime factor for each $m \in \mathcal{M}$ shows that the group generated by $(F_n)_{n \geq 1}$ is freely generated by $(\Phi_m)_{m \in \mathcal{M}}$. This group is also freely generated by $(F_m)_{m \in \mathcal{M}}$, but the cyclotomic numbers Φ_m almost freely generate the Fibonacci integers as a mutiplicative *semi*group. However, there are some Fibonacci integers not of this form, for example

$$\frac{\Phi_{24}}{\Phi_3}, \ \frac{\Phi_{25}}{\Phi_5}, \ \text{and} \ \frac{\Phi_{37 \cdot 19} \Phi_{113 \cdot 19}}{\Phi_{19}}.$$

The following lemma sheds some light on the structure of $\mathcal{G}_F$.

Lemma 1. *Assume that $\mathcal{I}, \mathcal{J}$ are finite multisets of indices with $m_i, n_j \in \mathcal{M}$, $n_i \neq m_j$ for all $i \in \mathcal{I}$, $j \in \mathcal{J}$ and*

$$n = \prod_{i \in \mathcal{I}} \Phi_{n_i} \prod_{j \in \mathcal{J}} \Phi_{m_j}^{-1} \in \mathbb{N}.$$

There exists an injection f from the multiset of prime factors of

$$\prod_{j \in \mathcal{J}} \Psi_{m_j} = p_1 p_2 \dots p_k$$

into the multiset $\{n_i\}_{i \in \mathcal{I}}$ where $f(p_l) = p_l^{k_l} z(p_l)$ for some $k_l \in \mathbb{N}$.

Proof. If p is a prime factor of $\prod \Psi_{m_j}$, then $z(p) = m_j$ for some $j \in \mathcal{J}$. Since $m_j \notin \{n_i\}_{i \in \mathcal{I}}$, it follows that $p = \delta_{n_i}$ for some $i \in \mathcal{I}$ and $n_i = p^k z(p)$ for some positive integer k. In particular, if $p^a \| \prod \Psi_{m_j}$, then there are at least a values of $i \in \mathcal{I}$ with n_i of the form $p^k z(p)$ (perhaps with different values of k), so that we may assign each factor of p to a different n_i. Further, since no n_i has two different representations in the form $p^k z(p)$ with p prime and $k > 0$, we may continue this mapping for each prime factor of $\prod \Psi_{m_j}$. □

Remark 1. Lemma 1 does not tell the whole story. What is not being accounted for in the result is the contribution of the primes in $\prod_{j \in \mathcal{J}} \delta_{m_j}$ which also need to occur in the product $\prod_{i \in \mathcal{I}} \Phi_{n_i}$. It turns out that these primes are negligible in our counting problem.

Remark 2. We remark that a prime p is a Fibonacci integer if and only if $\Psi_{z(p)} = p$ and $\delta_{z(p)}$ is a Fibonacci integer. Since $\delta_{z(p)}$ is either 1 or a prime (much) smaller than p, we thus have a simple algorithm for determining if a given prime is a Fibonacci integer. The first few primes which are *not* Fibonacci integers are 37, 43, and 53. Since Ψ_n is exponentially large in n, it is easy to see that the number of prime Fibonacci integers in $[1, x]$ is $O(\log x)$. Probably there are infinitely many of them, but we do not know how to prove this. It seems to be a slightly easier assertion than the conjecture that there are infinitely many prime Fibonacci numbers, but that doesn't seem to be of much help.

Yuri Bilu asked us if, in general, it is decidable whether a given natural number n is a Fibonacci integer. The arguments in the next section, and in partiqular (10), show that this is indeed the case.

Remark 3. Say that a Fibonacci integer is an *atom* if it exceeds 1 and it is not the product of two smaller Fibonacci integers. Let $\Xi_n = \Psi_n$ if δ_n is a Fibonacci integer, and otherwise, let $\Xi_n = \Phi_n$ (cf. Remark 2). Using the thoughts behind Lemma 1 it is possible to characterize the atoms as the union of $\{\Xi_n : n \in \mathcal{M}\}$ and

$$\left\{\frac{\prod_{i=1}^{l} \Phi_{p_i^{k_i} m}}{\Xi_m} \;:\; m \in \mathcal{M},\ \Xi_m = p_1 \dots p_l,\ l \geq 2, \text{ and } p_1^{k_1} m, \dots, p_l^{k_l} m \in \mathcal{M}\right\},$$

where $p_1, \dots, p_l$ denote primes. The Fibonacci integers do not enjoy unique factorization into atoms. Here are three examples based on the fact that $\Phi_{19} = 37 \cdot 113$: Let $n(j,l)$ denote the atom $\Phi_{37^j \cdot 19} \Phi_{113^l \cdot 19} / \Phi_{19}$, where $j, l > 0$, and note that

$$\begin{aligned} \Phi_{37 \cdot 19} \times \Phi_{113 \cdot 19} &= \Phi_{19} \times n(1,1), \\ \Phi_{37^2 \cdot 19} \times n(1,1) &= \Phi_{37 \cdot 19} \times n(2,1), \\ n(1,1) \times n(2,2) &= n(1,2) \times n(2,1). \end{aligned}$$

Such redundancies complicate the possible attainment of an asymptotic formula for the distribution of Fibonacci integers.

Lemma 2. *The inequality*

$$\alpha^{\phi(m)-1} \leq \Phi_m < \alpha^{\phi(m)+1} \tag{4}$$

holds for all integers $m > 1$.

Proof. The lemma holds with equality at the lower bound when $m = 2$, so assume $m \geq 3$. From (2), we get

$$\Phi_m = \prod_{d|m} F_{m/d}^{\mu(d)} = \prod_{d|m} \left(\alpha^{m/d} - \beta^{m/d}\right)^{\mu(d)} = \alpha^{\phi(m)} \prod_{d|m} \left(1 - \left(\frac{\beta}{\alpha}\right)^{m/d}\right)^{\mu(d)},$$

so that since $\beta/\alpha = -\alpha^{-2}$,

$$L_m := \log\left(\Phi_m / \alpha^{\phi(m)}\right) = \sum_{d|m} \mu(d) \log\left(1 - (-\alpha^{-2})^{m/d}\right).$$

It suffices to show that $|L_m| < \log \alpha$. Note that if $k \in \mathbb{N}$,

$$\begin{aligned} \sum_{j>k} \left|\log\left(1 - (-\alpha^{-2})^j\right)\right| &< \sum_{j>k} \left|\log\left(1 - \alpha^{-2j}\right)\right| \\ &= \sum_{i \geq 1} \frac{1}{i} \alpha^{-2ik} \frac{1}{\alpha^{2i} - 1} < \left|\log\left(1 - \alpha^{-2k}\right)\right|. \end{aligned}$$

If m is not squarefree, we have

$$|L_m| < \sum_{j>1} \left|\log\left(1-(-\alpha^{-2})^j\right)\right| < \left|\log\left(1-\alpha^{-2}\right)\right| = \log\alpha.$$

Suppose m is squarefree and p is the smallest odd prime factor of m. Then $p+1 \nmid m$. If m is odd,

$$\begin{aligned}|L_m| &< \left|\log\left(1+\alpha^{-2}\right) - \log\left(1+\alpha^{-2p}\right)\right| + \sum_{j>p+1} \left|\log\left(1-(-\alpha^{-2})^j\right)\right| \\ &< \left|\log\left(1+\alpha^{-2}\right) - \log\left(1+\alpha^{-2p}\right)\right| + \left|\log\left(1-\alpha^{-2p-2}\right)\right| \\ &= \log\left(1+\alpha^{-2}\right) - \log\left(1+\alpha^{-2p}\right) - \log\left(1-\alpha^{-2p-2}\right) \\ &< \log\left(1+\alpha^{-2}\right) < \log\alpha.\end{aligned}$$

If m is even,

$$\begin{aligned}|L_m| &< \left|\log\left(1+\alpha^{-2}\right) - \log\left(1-\alpha^{-4}\right) - \log\left(1+\alpha^{-2p}\right)\right| \\ &\qquad + \sum_{j>p+1} \left|\log\left(1-(-\alpha^{-2})^j\right)\right| \\ &< \left|-\log\left(1-\alpha^{-2}\right) - \log\left(1+\alpha^{-2p}\right)\right| + \left|\log\left(1-\alpha^{-2p-2}\right)\right| \\ &= -\log\left(1-\alpha^{-2}\right) - \log\left(1+\alpha^{-2p}\right) - \log\left(1-\alpha^{-2p-2}\right) \\ &< -\log\left(1-\alpha^{-2}\right) = \log\alpha.\end{aligned}$$

This completes the proof. □

3 Combinatorial arguments

Let $\mathcal{G}_1$ be the multiplicative semigroup freely generated by $\{\Phi_m\}_{m\in\mathcal{M}}$ inside the set $\mathbb{N}$ of natural numbers and for a positive real number x let $\mathcal{G}_1(x) = \mathcal{G}_1 \cap [1,x]$. In the next section, with tools specific to complex analysis, we will prove the following theorem.

Theorem 2. *For each fixed $\epsilon > 0$, the estimate*

$$\exp\left(c(\log x)^{1/2} - (\log x)^{\epsilon}\right) \le \#\mathcal{G}_1(x) \le \exp\left(c\sqrt{\log x} + (\log x)^{\epsilon}\right) \tag{5}$$

holds for all sufficiently large x.

In this section, we shall show how to use this theorem and the results of the previous section to complete the proof of our main result, Theorem 1.

It suffices to deal with the upper bound on $\#\mathcal{G}_F(x)$, since the lower bound is an immediate consequence of Theorem 2 and the fact that each Φ_m for $m \in \mathcal{M}$ is a Fibonacci integer.

Let x be large and assume that

$$N = \prod_{i\in\mathcal{I}} \Phi_{n_i} \prod_{j\in\mathcal{J}} \Phi_{m_j}^{-1} \le x$$

is an integer. We need to bound from above the number of such possible N's. For each $m \in \mathcal{M}$, let $l(m) = \Omega(\Psi_m)$ be the number of primitive prime factors of Φ_m counted with multiplicity. Let f be an injection from the multiset of prime factors of

$$\prod_{j\in\mathcal{J}} \Psi_{m_j}$$

into the multiset $\{n_i\}_{i\in\mathcal{I}}$ as guaranteed by Lemma 1. For $j \in \mathcal{J}$, let $p_{j,1} \ge p_{j,2} \ge \cdots \ge p_{j,l(m_j)}$ be the multiset of prime factors of Ψ_{m_j}, and for $1 \le l \le l(m_j)$, let $n_{j,l} = f(p_{j,l})$, so that $n_{j,l} = p_{j,l}^{k_{j,l}} m_j$ for some positive integer $k_{j,l}$. Thus, by a change in notation, we wish to count the number of numbers of the form

$$N = \left(\prod_{j\in\mathcal{J}} \frac{1}{\Phi_{m_j}} \prod_{l=1}^{l(m_j)} \Phi_{n_{j,l}} \right) \prod_{i\in\mathcal{I}} \Phi_{n_i} \le x. \tag{6}$$

(Note that numbers N in (6) are not necessarily integers since if some Φ_{m_j} has a non-primitive prime factor, we have not necessarily arranged for it to be cancelled with some corresponding prime among the Φ_n's, see Remark 1.) We thus have

$$\#\mathcal{G}_F(x) \le \sum_w \#\mathcal{G}_1(x/w), \tag{7}$$

where w ranges over rationals of the form of the parenthetical double product in (6).

Fix some j, l and look at $m := m_j$ and $n := n_{j,l} = p^k m$, where $p = p_{j,l}$ is a primitive prime factor of Φ_m and $k \ge 1$. By Lemma 2, we have

$$\frac{\Phi_n}{\Phi_m} \ge \alpha^{\phi(n)-\phi(m)-2}. \tag{8}$$

We claim that

$$\phi(n) - \phi(m) - 2 \ge \frac{1}{2}\phi(n). \tag{9}$$

Indeed, if $p = 2$, then $m = 3$ and $k \ge 3$, so that

$$\phi(n) - \phi(m) - 2 = \frac{1}{3}n - 4 \ge \frac{1}{6}n = \frac{1}{2}\phi(n).$$

If $p = 3$, then $m = 4$ and $k \geq 2$, so that

$$\phi(n) - \phi(m) - 2 = \frac{1}{3}n - 4 \geq \frac{2}{9}n = \frac{2}{3}\phi(n).$$

In the remaining cases, we have $p \geq 5$ and $\phi(m) \geq 4$, so that

$$\phi(n) - \phi(m) - 2 \geq \phi(p^k)\phi(m) - \frac{3}{2}\phi(m) \geq \frac{5}{8}\phi(p^k)\phi(m) = \frac{5}{8}\phi(n).$$

In each case we have (9).

It follows from (6), (8), and (9) that

$$\sum_{j \in \mathcal{J}} \sum_{l=1}^{l(m_j)} \phi(n_{j,l}) \leq \frac{2}{\log \alpha} \log x,$$

and so using the minimal order of ϕ, we have

$$\sum_{j \in \mathcal{J}} \sum_{l=1}^{l(m_j)} n_{j,l} \leq K, \tag{10}$$

where $K = \lfloor \kappa \log x \log\log\log x \rfloor$ and κ is an absolute computable constant. Since $n_{j,l} = p_{j,l}^{k_{j,l}} m_j$, we also have each $p_{j,1} m_j \leq K$. So, the inequality

$$\frac{\Phi_{m_j}}{m_j} \leq \Psi_{m_j} \leq p_{j,1}^{l(m_j)}$$

and Lemma 2 imply that $l(m_j) \gg m_j/(\log\log x \log\log\log x)$. Thus,

$$l(m_j) \geq \frac{m_j}{(\log\log x)^2} \tag{11}$$

for sufficiently large x, each $j \in \mathcal{J}$, and all $\mathcal{J}$ as above.

Let $L = (\log x)^{1/6}$. We say a multiset $\{n_{j,l}\}$ is *good* if for each j there are at most L distinct primes $p_{j,l}$ with exponents $k_{j,l}$ at least 2. We say $\{n_{j,l}\}$ is *bad* if for each j, there are more than L distinct primes $p_{j,l}$ with exponents $k_{j,l}$ at least 2. Each multiset in our problem can be partitioned into a good multiset and a bad multiset. With T, U, M positive integers, let $\mathcal{N}_T$ be the set of good multisets with T distinct m_j's and let $\mathcal{N}_{U,M}$ be the set of bad multisets with U distinct m_j's, where the sum of the distinct m_j's is M. Let $\mathcal{W}_T$ denote the set of rationals w in $[1, x]$ of the form

$$w = \prod_{j \in \mathcal{J}} \frac{1}{\Phi_{m_j}} \prod_{l=1}^{l(m_j)} \Phi_{n_{j,l}},$$

where the multiset $\{n_{j,l}\}$ is in $\mathcal{N}_T$ and let w_T be the least member of $\mathcal{W}_T$, with $w_T = 1$ if $\mathcal{W}_T$ is empty. Similarly define $\mathcal{W}_{U,M}$ and $w_{U,M}$ for multisets in $\mathcal{N}_{U,M}$. We have from (7) that

$$\#\mathcal{G}_F(x) \le \sum_{T,U,M} \sum_{\substack{w\in\mathcal{W}_T\\ w'\in\mathcal{W}_{U,M}}} \#\mathcal{G}_1\left(\frac{x}{ww'}\right) \le \sum_{T,U,M} \#\mathcal{G}_1\left(\frac{x}{w_T w_{U,M}}\right) N_T N_{U,M}, \tag{12}$$

where T, U, M run up to K, and $N_T = 1 + \#\mathcal{N}_T$, $N_{U,M} = 1 + \#\mathcal{N}_{U,M}$.

We wish to count the number of multisets $\{n_{j,l}\}$ arising with certain constraints. Such a multiset uniquely determines the corresponding multiset $\{m_j\}$, and so we count by first choosing this simpler multiset and then extending to $\{n_{j,l}\}$. The number of ways of choosing a multiset $\{m_j\}$ with $\sum m_j \le K$ and with T distinct m_j's is at most $K^{2T} \le \exp(3T\log\log x)$ for x large. Given some m_j, the number of corresponding multisets $\{n_{j,l}\}$ is at most $(2\log K)^{l(m_j)} \le \exp(2m_j \log\log x)$ for large x, since $l(m_j) < m_j$ and the number of choices for an exponent $k_{j,l}$ for the prime $p_{j,l}$ is at most $\log K/\log p_{j,l} < 2\log K$. In the case where we know that the number of distinct primes $p_{j,l}$ which have exponents at least 2 is at most L, the number of ways of choosing these distinct primes is at most K^L. Fixing one such prime p, the number of ways of choosing exponents for all of its copies is at most $K^{2\log K}$ for large x. Indeed, with $Z = \lfloor 2\log K\rfloor$, we are counting integer vectors $(v_1, \ldots, v_Z)$ where v_k is the number of copies of p with exponent k. If there are s copies of p in all, then each $v_k \le s$. So, there are at most $(s+1)^Z$ choices, and it remains to note that since $s \le K/2$, the above quantity is smaller than $K^{2\log K}$ for large values of x. So, the total number of ways to choose these L distinct primes and exponents for them and their copies is at most $K^{L+2L\log K} \le \exp(3L(\log\log x)^2)$. We conclude that

$$N_T N_{U,M} \le \exp\left(3(T+U+M+L)(\log\log x)^2\right). \tag{13}$$

Thus, to use (12), we wish to have upper bounds for T, U, M and lower bounds for the numbers w_T and $w_{U,M}$. We will see that these tasks are related. Suppose w_T arises from the multiset pair $\{m_j\}$, $\{n_{j,l}\}$. Without loss of generality, we may assume that $m_1, \ldots, m_T$ are distinct. Since each

$p_{j,l} \geq m_j - 1$ and $m_j \geq 3$, we have by (11)

$$\sum_{j\in\mathcal{J}} \sum_{l=1}^{l(m_j)} n_{j,l} \geq \sum_{j=1}^{T} \sum_{l=1}^{l(m_j)} p_{j,l} m_j \geq \sum_{j=1}^{T} l(m_j)(m_j-1)m_j$$

$$\geq \frac{1}{(\log\log x)^2} \sum_{j=3}^{T+2} j^2(j-1) \geq \frac{1/4}{(\log\log x)^2} T^4.$$

We conclude from Lemma 2, (9), (10) and the minimal order of ϕ that

$$w_T \geq \exp\left(\frac{T^4}{(\log\log x)^3}\right) \tag{14}$$

for all large x.

We achieve two lower bounds for $w_{U,M}$ as follows. For $\{n_{j,l}\}$ in $\mathcal{N}_{U,M}$, with corresponding multiset $\{m_j\}$, we assume that $m_1, \ldots, m_U$ are distinct. For each $j \leq U$ there are more than L distinct primes $p_{j,l}$ with exponent at least 2 in $n_{j,l}$. Each of these primes satisfies $p_{j,l} \equiv 0, \pm 1 \pmod{m_j}$, so that $\sum_l p_{j,l}^2 \gg L^3 m_j^2$ and $\sum\sum n_{j,l} \gg L^3 \sum m_j^3$. As above, we deduce that

$$w_{U,M} \geq \exp\left(\frac{L^3 \sum m_j^3}{\log\log x}\right) \tag{15}$$

for all large x. In addition, the primes associated with m_j are all different from the primes associated with $m_{j'}$ when $m_j \neq m_{j'}$, so there are more than UL distinct primes $p_{j,l}$ with exponents at least 2 among the various $n_{j,l}$'s. The sum of their squares is at least of order U^3L^3, and since each $m_j > L$ (using $l(m_j) > L$), we have $\sum\sum n_{j,l} \gg U^3L^4$. We thus deduce that

$$w_{U,M} \geq \exp\left(\frac{U^3L^4}{\log\log x}\right) \tag{16}$$

for all large x. By Hölder's inequality, we have

$$M = \sum_{j=1}^{U} m_j \leq \left(\sum_{j=1}^{U} m_j^3\right)^{1/3} U^{2/3},$$

so that from (16) and then (15), we obtain

$$U \leq \frac{(\log w_{U,M})^{1/3}}{L^{4/3}}(\log\log x)^{1/3}, \quad M \leq \frac{(\log w_{U,M})^{5/9}}{L^{17/9}}(\log\log x)^{5/9} \tag{17}$$

for all large x.

Note that for $1 \le w \le x$,

$$\left(\log \frac{x}{w}\right)^{1/2} = (\log x)^{1/2}\left(1 - \frac{\log w}{\log x}\right)^{1/2} \le (\log x)^{1/2}\left(1 - \frac{\log w}{2\log x}\right)$$
$$= (\log x)^{1/2} - \frac{\log w}{2(\log x)^{1/2}}.$$

Thus, from Theorem 2, we have

$$\#\mathcal{G}_1\left(\frac{x}{w_T w_{U,M}}\right) \le \exp\left(c(\log x)^{1/2} - \frac{c\log(w_T w_{U,M})}{2(\log x)^{1/2}} + (\log x)^{\epsilon}\right)$$

for all large x. With (12) we thus get,

$$\frac{\#\mathcal{G}_F(x)}{\exp\left(c(\log x)^{1/2} + (\log x)^{\epsilon}\right)} \le \sum_{T,U,M} \exp\left(-\frac{c\log(w_T w_{U,M})}{2(\log x)^{1/2}}\right) N_T N_{U,M}. \tag{18}$$

Since the sum in (18) has at most $(K+1)^3$ terms, it suffices to show that each term is at most $\exp((\log x)^{1/6+\epsilon/2})$. Suppose $2/3 < a \le 1$ and we have shown that all of the terms in (18) with $w_T w_{U,M} > \exp((\log x)^a(\log\log x)^5)$ are negligible. (We definitely have this for $a = 1$ since no term satisfies this inequality.) Then (14) gives $T \le (\log x)^{a/4}(\log\log x)^2$ and (17) gives both $U \le (\log x)^{a/3-2/9}(\log\log x)^2$ and $M < (\log x)^{5a/9-17/54}(\log\log x)^4$. Note that $a/3 - 2/9 < 5a/9 - 17/54 < a/4$, so from (13),

$$N_T N_{U,M} \le \exp\left(4(\log x)^{a/4}(\log\log x)^4\right).$$

Since $a/4+1/2 > 2/3$, we thus may replace a with $a/4+1/2$ in the argument, since those terms with $w_T w_{U,M} > \exp((\log x)^{a/4+1/2}(\log\log x)^5)$ are now seen to be negligible in (18) because $a/4 < a - 1/2$. In a finite number of steps, we reach a value of a with $2/3 < a < 2/3 + \epsilon$, and then all remaining terms in (18) are smaller than $\exp((\log x)^{1/6+\epsilon/2})$. This completes the proof of Theorem 1. □

Remark 4. To improve on the exponent 1/6 in Theorem 1 it would seem that a finer knowledge of the prime factors of the numbers Φ_m would be needed. It seems reasonable that for all large m there is a prime factor of Φ_m that is larger than any fixed power of m, and if this is the case, the exponent 1/6 can be replaced with 0. Assuming a strong form of the *abc* conjecture, it follows that for each fixed $\epsilon > 0$ the number Φ_m has a squarefree divisor larger than $\Phi_m^{1-\epsilon}$ for all sufficiently large m (see [7]). Using this, our argument would give exponent 1/8 in place of 1/6 in Theorem 1.

4 Analytic arguments

In the previous section, we reduced the problem to counting the number of positive integers in $[1, x]$ that belong to the semigroup $\mathcal{G}_1$ freely generated by $\{\Phi_m\}_{m\in\mathcal{M}}$. We need to prove that this number satisfies the estimate given by Theorem 2.

Note first that the Dirichlet series associated with $\mathcal{G}_1$ is given by

$$D(z) := \sum_{n\in\mathcal{G}_1} n^{-z} = \prod_{m\in\mathcal{M}} (1-\Phi_m^{-z})^{-1}.$$

In view of the exponential growth of Φ_m, the series $D(z)$ converges for $\Re(z) > 0$. Now we apply the following well-known variant of Perron's formula (see, e.g., [4]):

$$G(x) := \sum_{\substack{n\le x\\ n\in\mathcal{G}_1}} \left(1-\frac{n}{x}\right) = \frac{1}{2\pi i}\int_{r-i\infty}^{r+i\infty} \frac{D(z)x^z}{z(z+1)}\,dz \tag{19}$$

for any $r > 0$. It will later turn out to be advantageous to work with this variant rather than Perron's formula itself. The integral in (19) is estimated by means of the saddle-point method. In order to choose r appropriately, we have to study the behavior of $D(z)$ as $z \to 0$. We have

$$\log D(z) = -\sum_{m\in\mathcal{M}} \log\left(1-\Phi_m^{-z}\right) = -\sum_{m\in\mathcal{M}} \log\left(1-\exp(-(\log\Phi_m)z)\right). \tag{20}$$

Now we apply the Mellin transform to this harmonic sum; the Mellin transform of $-\log(1-e^{-z})$ is given by $\Gamma(s)\zeta(s+1)$, which implies that the Mellin transform of $\log D(z)$ is $\Gamma(s)\zeta(s+1)C(s)$, where

$$C(s) = \sum_{m\in\mathcal{M}} (\log\Phi_m)^{-s}.$$

Next we need information on the analytic behavior of the Dirichlet series $C(s)$, which is provided in the following lemma:

Lemma 3. *The Dirichlet series $C(s)$ satisfies*

$$C(s) = (\log\alpha)^{-s}\sum_{m\ge1}\phi(m)^{-s} + A(s),$$

where $A(s)$ is analytic on $\Re(s) > 0$ and satisfies $A(s) = O(|s|)$ for $\epsilon \le \Re(s) \le \epsilon^{-1}$, where the implied constant only depends on ϵ.

Proof. Write $L_m = \log \Phi_m - \phi(m) \log \alpha$ as in the proof of Lemma 2, and recall that

$$|L_m| \leq \log \alpha$$

for arbitrary $m > 1$. Thus, we have

$$\begin{aligned} A(s) &= C(s) - (\log \alpha)^{-s} \sum_{m \geq 1} \phi(m)^{-s} \\ &= -(\log \alpha)^{-s}(2 + 2^{-s} + 4^{-s}) + \sum_{m \in \mathcal{M}} \left((\log \Phi_m)^{-s} - (\log \alpha)^{-s} \phi(m)^{-s} \right). \end{aligned}$$

The first part is analytic and bounded in the indicated region, so it remains to consider the sum over $\mathcal{M}$, which we denote $A_0(s)$. We have

$$A_0(s) = (\log \alpha)^{-s} \sum_{m \in \mathcal{M}} \phi(m)^{-s} \left(\left(1 + \frac{L_m}{\phi(m) \log \alpha} \right)^{-s} - 1 \right),$$

so that

$$\begin{aligned} |A_0(s)| &\ll |(\log \alpha)^{-s}| \left(\sum_{\substack{m \in \mathcal{M} \\ \phi(m) < |s|}} |\phi(m)^{-s}| + \sum_{\substack{m \in \mathcal{M} \\ \phi(m) \geq |s|}} |\phi(m)^{-s}| \cdot \frac{|sL_m|}{\phi(m) \log \alpha} \right) \\ &\leq |(\log \alpha)^{-s}| \left(\sum_{\substack{m \in \mathcal{M} \\ \phi(m) < |s|}} \phi(m)^{-\epsilon} + |s| \sum_{\substack{m \in \mathcal{M} \\ \phi(m) \geq |s|}} \phi(m)^{-1-\epsilon} \right) \\ &\ll |s|. \end{aligned}$$

This shows that the sum converges absolutely and uniformly on compact subsets of the half-plane $\Re(s) > 0$, and so we have that $A_0(s)$, and hence $A(s)$, is analytic on this half-plane. This completes the proof of the lemma. □

The Dirichlet series $\sum_{m \geq 1} \phi(m)^{-s}$ was studied, for instance, in [2]. Since $\phi(m)$ is multiplicative, we have the Euler product

$$\begin{aligned} \sum_{m \geq 1} \phi(m)^{-s} = \prod_p \sum_{\alpha \geq 0} \phi(p^\alpha)^{-s} &= \prod_p \left(1 + \frac{1}{(p-1)^s (1 - p^{-s})} \right) \\ &= \prod_p (1 - p^{-s})^{-1} \cdot \prod_p \left(1 + (p-1)^{-s} - p^{-s} \right) \\ &= \zeta(s) \cdot \prod_p \left(1 + (p-1)^{-s} - p^{-s} \right). \end{aligned}$$

The second factor converges for $\Re(s) = \sigma \geq \epsilon > 0$, and it is easy to show that it grows subexponentially as $|\Im(s)| \to \infty$ in this region. Indeed, noting that $(1-1/p)^{-s} = 1 + O(|s|p^{-1})$ if $p > |s|$, we find

$$\begin{aligned}
\log \prod_p \left(1 + (p-1)^{-s} - p^{-s}\right) &= \sum_p \log\left(1 + (p-1)^{-s} - p^{-s}\right) \\
&= \sum_{p \leq |s|} \log\left(1 + (p-1)^{-s} - p^{-s}\right) + \sum_{p > |s|} \log\left(1 - p^{-s}\left(1 - (1-1/p)^{-s}\right)\right) \\
&= \sum_{p \leq |s|} \log\left(1 + O(p^{-\sigma})\right) + \sum_{p > |s|} \log\left(1 + O(|s|p^{-\sigma-1})\right) \\
&\ll \sum_{p \leq |s|} p^{-\sigma} + |s| \sum_{p > |s|} p^{-\sigma-1} \ll |s|^{1-\sigma} \ll |s|^{1-\epsilon}.
\end{aligned}$$

Since $|\Gamma(s)|$ decreases exponentially as $|\Im(s)| \to \infty$, this is therefore also the case for $\Gamma(s)\zeta(s+1)C(s)$, which allows us to apply the Mellin inversion formula: we have

$$\log D(r) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \Gamma(s)\zeta(s+1)C(s)r^{-s}\,ds.$$

If we shift the path of integration to $\Re(s) = \epsilon$ and pick up the residue at $s = 1$ (see [3] for details), we obtain

$$\log D(r) = \frac{A}{r} + O(r^{-\epsilon}),$$

where the constant A is given by

$$\begin{aligned}
\frac{\Gamma(1)\zeta(2)}{\log \alpha} \cdot \prod_p \left(1 + (p-1)^{-1} - p^{-1}\right) &= \frac{\zeta(2)}{\log \alpha} \prod_p \frac{p^2 - p + 1}{p(p-1)} \\
&= \frac{\zeta(2)}{\log \alpha} \prod_p \frac{1 - p^{-6}}{(1-p^{-2})(1-p^{-3})} = \frac{\zeta(2)^2\zeta(3)}{\zeta(6)\log \alpha}.
\end{aligned}$$

By similar arguments, we find that

$$\frac{d}{dz} \log D(z) = -\sum_{m \in \mathcal{M}} \frac{\log \Phi_m}{\Phi_m^z - 1}$$

has Mellin transform $-\Gamma(s)\zeta(s)C(s-1)$, which yields

$$\frac{d}{dz} \log D(z)\Big|_{z=r} = -\frac{A}{r^2} + O(r^{-1-\epsilon}),$$

and in the same manner

$$\frac{d^2}{dz^2}\log D(z)\Big|_{z=r} = \frac{2A}{r^3} + O(r^{-2-\epsilon}).$$

Finally, we have, if $z = r + it$,

$$\begin{aligned}\left|\frac{d^3}{dz^3}\log D(z)\right| &= \left|\sum_{m\in\mathcal{M}} \frac{(\log\Phi_m)^3\Phi_m^z(\Phi_m^z+1)}{(\Phi_m^z-1)^3}\right| \\ &\le \sum_{m\in\mathcal{M}} \frac{(\log\Phi_m)^3\Phi_m^r(\Phi_m^r+1)}{(\Phi_m^r-1)^3} = O(r^{-4}),\end{aligned}$$

uniformly in t. Now we can expand $\log D(z)$ into the series:

$$\log D(r) - it\left(\frac{A}{r^2} + O(r^{-1-\epsilon})\right) - \frac{t^2}{2}\left(\frac{2A}{r^3} + O(r^{-2-\epsilon})\right) + O\left(\frac{|t|^3}{r^4}\right).$$

If we restrict ourselves to the central part $|t| \le T = r^{7/5}$, then this gives us

$$\log D(z) = \log D(r) - \frac{iAt}{r^2} - \frac{At^2}{r^3} + O(r^{1/5}),$$

when $0 < r < 1$. Using

$$\frac{1}{z} = \frac{1}{r}\left(1 + O\Big(\frac{|t|}{r}\Big)\right), \quad \frac{1}{z+1} = \frac{1}{r+1}(1 + O(|t|)) = 1 + O(r),$$

we consequently have

$$\frac{D(z)x^z}{z(z+1)} = \frac{D(r)x^r}{r}\exp\left(-\frac{iAt}{r^2} - \frac{At^2}{r^3} + it\log x + O(r^{1/5})\right).$$

Now choose r in such a way that the linear terms in the exponent cancel, i.e., $r = \sqrt{A/\log x}$. Then the central part in the integral in (19) is

$$\frac{1}{2\pi i}\int_{r-iT}^{r+iT} \frac{D(z)x^z}{z(z+1)}\,dz = \frac{D(r)x^r}{r}\cdot\frac{1}{2\pi}\int_{-T}^{T}\exp\left(-\frac{At^2}{r^3}\right)\left(1 + O(r^{1/5})\right)\,dt.$$

Completing the integral on the right to the entire interval $(-\infty,\infty)$ only yields an exponentially small error term, hence we have

$$\begin{aligned}\frac{1}{2\pi i}\int_{r-iT}^{r+iT} \frac{D(z)x^z}{z(z+1)}\,dz &= \frac{D(r)x^r}{r}\cdot\frac{1}{2\sqrt{\pi A}}r^{3/2}\left(1 + O(r^{1/5})\right) \\ &= \frac{D(r)x^r\sqrt{r}}{2\sqrt{\pi A}}\left(1 + O(r^{1/5})\right).\end{aligned}$$

It remains to estimate the part with $|t| \geq r^{7/5}$ in the integral in (19). The part $|t| \geq r^{-1}$ can be trivially estimated, since $|D(z)| \leq |D(r)|$:

$$\left| \frac{1}{2\pi i} \int_{r+ir^{-1}}^{r+i\infty} \frac{D(z)x^z}{z(z+1)}\, dz \right| \leq D(r)x^r \cdot \frac{1}{2\pi} \int_{r+ir^{-1}}^{r+i\infty} \frac{1}{|z(z+1)|}\, dz$$
$$\ll rD(r)x^r,$$

and likewise for $t \leq -r^{-1}$. Here it is essential that we were using the modified version of Perron's formula to obtain a convergent integral.

For the part $r^{7/5} \leq |t| \leq r^{-1}$, we need another estimate for $D(z)$.

Lemma 4. *Write $z = r + it$, and suppose that $r^{7/5} \leq |t| \leq r^{-1}$. Then we have the estimate*

$$\log D(r) - \Re(\log D(z)) \gg r^{-1/5}$$

uniformly for $0 < r \leq 1/3$.

Proof. First of all, we reduce the task to an estimate for certain exponential sums. We have

$$\log D(z) = -\sum_{m\in\mathcal{M}} \log\left(1 - \Phi_m^{-z}\right) = \sum_{m\in\mathcal{M}} \sum_{k=1}^{\infty} \frac{1}{k} \Phi_m^{-kz}$$
$$= \sum_{m\in\mathcal{M}} \sum_{k=1}^{\infty} \frac{1}{k} \Phi_m^{-kr} \left(\cos(kt\log\Phi_m) - i\sin(kt\log\Phi_m)\right).$$

Hence,

$$\log D(r) - \Re(\log D(z)) = \sum_{m\in\mathcal{M}} \sum_{k=1}^{\infty} \frac{1}{k} \Phi_m^{-kr} \left(1 - \cos(kt\log\Phi_m)\right)$$
$$\geq \sum_{m\in\mathcal{M}} \Phi_m^{-r} \left(1 - \cos(t\log\Phi_m)\right).$$

By Lemma 2, $\Phi_m < \alpha^{\phi(m)+1} \leq \alpha^m$ for $m \in \mathcal{M}$, which implies $\Phi_m^{-r} \geq 1/\alpha$ for $m \leq r^{-1}$, $m \in \mathcal{M}$. So we obtain the following estimate:

$$\log D(r) - \Re(\log D(z)) \geq \alpha^{-1} \sum_{m\leq r^{-1}} \left(1 - \cos(t\log\Phi_m)\right),$$

which allows us to restrict our attention to a somewhat simpler sum. We distinguish several cases.

If $|t| \leq r$, we then have

$$|t \log \Phi_m| \leq |t| m \log \alpha \leq rm \log \alpha \leq \log \alpha < \frac{\pi}{2}$$

for $m \leq r^{-1}$ and thus $1 - \cos(t \log \Phi_m) \gg t^2 (\log \Phi_m)^2$. Hence, we obtain

$$\log D(r) - \Re(\log D(z)) \gg t^2 \sum_{m \leq r^{-1}} (\log \Phi_m)^2 \gg t^2 \sum_{m \leq r^{-1}} \phi(m)^2.$$

Now an elementary argument shows that

$$\sum_{m \leq M} \phi(m)^2 \asymp M^3,$$

so that using $|t| \geq r^{7/5}$, we have

$$\log D(r) - \Re(\log D(z)) \gg t^2 r^{-3} \geq r^{-1/5}.$$

If $r \leq |t| \leq r^{1/5}$, we can apply the same argument to obtain

$$\begin{aligned}\log D(r) - \Re(\log D(z)) &\geq \alpha^{-1} \sum_{m \leq r^{-1}} (1 - \cos(t \log \Phi_m)) \\ &\geq \alpha^{-1} \sum_{m \leq |t|^{-1}} (1 - \cos(t \log \Phi_m)) \\ &\gg t^2 \sum_{m \leq |t|^{-1}} (\log \Phi_m)^2 \gg t^2 \cdot |t|^{-3} \gg r^{-1/5}.\end{aligned}$$

For $|t| \geq r^{1/5}$, we need different arguments. Write $X = r^{-1}$. Clearly,

$$\sum_{m \leq X} (1 - \cos(t \log \Phi_m)) \geq \sum_{m \in \mathcal{N}} (1 - \cos(t \log \Phi_m))$$

for any set $\mathcal{N}$ of integers in $[1, X]$. We focus on the following two cases:
I. Take the set $\mathcal{N}_1$ that consists of all numbers p and $2p$ where p is a prime in the interval $[\frac{X}{4}, \frac{X}{2}]$. We have

$$\log(\Phi_p) = \log F_p = \log \left(\frac{\alpha^p - \beta^p}{\alpha - \beta} \right) = p \log \alpha - \log(\alpha - \beta) + O\left(\alpha^{-X/2}\right),$$

using $\beta/\alpha = -\alpha^{-2}$, and

$$\log(\Phi_{2p}) = \log(F_{2p}/F_p) = \log(\alpha^p + \beta^p) = p \log \alpha + O\left(\alpha^{-X/2}\right).$$

Thus,

$$\begin{aligned}&\sum_{m\in\mathcal{N}_1}\exp(it\log\Phi_m)\\&\qquad=(1+\exp(-it\log(\alpha-\beta)))\sum_{X/4\le p\le X/2}\exp(itp\log\alpha)+O\left(X\alpha^{-X/2}\right).\end{aligned}$$

Since $\#\mathcal{N}_1$ is twice the number of primes in $[X/4,X/2]$, we now have the estimate

$$\begin{aligned}\sum_{m\in\mathcal{N}_1}\cos(t\log\Phi_m)&=\Re\left(\sum_{m\in\mathcal{N}_1}\exp(it\log\Phi_m)\right)\le\left|\sum_{m\in\mathcal{N}_1}\exp(it\log\Phi_m)\right|\\&\le\frac{\#\mathcal{N}_1}{2}\left|1+\exp(-it\log(\alpha-\beta))\right|+o(1).\end{aligned}$$

II. Similarly, for an odd prime $q<\log X$, we consider the set $\mathcal{N}_q$ of all integers of the form pq or $2pq$, where $p\in[\frac{X}{4q},\frac{X}{2q}]$ is prime. We have

$$\begin{aligned}\log(\Phi_{pq})&=\log\left(\frac{(\alpha^{pq}-\beta^{pq})(\alpha-\beta)}{(\alpha^p-\beta^p)(\alpha^q-\beta^q)}\right)\\&=(pq-p)\log\alpha+\log(\alpha-\beta)-\log\left(\alpha^q-\beta^q\right)+O\left(\alpha^{-2p}\right)\end{aligned}$$

and

$$\begin{aligned}\log(\Phi_{2pq})&=\log\left(\frac{\alpha^{pq}+\beta^{pq}}{(\alpha^p+\beta^p)(\alpha^q+\beta^q)}\right)\\&=(pq-p)\log\alpha-\log\left(\alpha^q+\beta^q\right)+O\left(\alpha^{-2p}\right).\end{aligned}$$

Then the same reasoning as above shows that

$$\begin{aligned}&\sum_{m\in\mathcal{N}_q}\cos(t\log\Phi_m)\\&\qquad\le\frac{\#\mathcal{N}_q}{2}\left|1+\exp\left(it\log(\alpha-\beta)+it\log\left(\frac{\alpha^q+\beta^q}{\alpha^q-\beta^q}\right)\right)\right|+o(1).\end{aligned}$$

Now suppose that X is large enough that there exists a prime q in the interval

$$\left[\frac{\frac{6}{7}\log Y}{2\log\alpha},\frac{\log Y}{2\log\alpha}\right]$$

for all $Y \geq X^{1/10}$, which is guaranteed by the prime number theorem. Choose q in such a way that it lies inside this interval with $Y = |t|X^{1/3}$; by our assumptions on t, we have $Y \geq X^{-1/5} \cdot X^{1/3} > X^{1/10}$. It follows that

$$Y^{-1} \leq \alpha^{-2q} \leq Y^{-6/7}$$

and thus

$$X^{-1/3} = |t|Y^{-1} \leq |t|\alpha^{-2q} \leq |t|Y^{-6/7} = |t|^{1/7}X^{-2/7} \leq X^{-1/7}.$$

Using $\beta/\alpha = -\alpha^{-2}$, we thus have

$$\left|\exp\left(it\log\left(\frac{\alpha^q+\beta^q}{\alpha^q-\beta^q}\right)\right) - 1\right| \gg X^{-1/3}. \tag{21}$$

Now consider the two expressions

$$A_1 = \frac{1}{2}\left|1+\exp\left(it\log\left(\alpha-\beta\right)\right)\right|$$

and

$$A_2 = \frac{1}{2}\left|1+\exp\left(it\log\left(\alpha-\beta\right) + it\log\left(\frac{\alpha^q+\beta^q}{\alpha^q-\beta^q}\right)\right)\right|.$$

Trivially, $A_1, A_2 \leq 1$, and the estimate (21) now shows that either $A_1 \leq 1 - C_1X^{-2/3}$ or $A_2 \leq 1 - C_1X^{-2/3}$ for some positive constant C_1. In the first case, we obtain the estimate

$$\begin{aligned}\sum_{m\in\mathcal{N}_1}(1-\cos(t\log\Phi_m)) &\geq \#\mathcal{N}_1 - (1-C_1X^{-2/3})\#\mathcal{N}_1 \\ &= C_1X^{-2/3}\#\mathcal{N}_1 \gg \frac{X^{1/3}}{\log X}.\end{aligned}$$

In the second case we have, analogously,

$$\sum_{m\in\mathcal{N}_q}(1-\cos(t\log\Phi_m)) \gg \frac{X^{1/3}}{\log^2 X}.$$

In either case, the proof of the lemma is completed. □

Making use of this lemma, it is now easy to estimate the remaining part of the integral:

$$\begin{aligned}\left|\frac{1}{2\pi i}\int_{r+ir^{7/5}}^{r+ir^{-1}}\frac{D(z)x^z}{z(z+1)}\,dz\right| &\leq D(r)x^r\exp(-C_2r^{-1/5})\int_{r+ir^{7/5}}^{r+ir^{-1}}\frac{1}{|z(z+1)|}\,dz \\ &\ll |\log r|\exp(-C_2r^{-1/5})D(r)x^r\end{aligned}$$

for some positive constant C_2.

Putting all three parts of the integral together, we obtain

$$\begin{aligned} G(x) = \sum_{\substack{n \le x \\ n \in \mathcal{G}_1}} \left(1 - \frac{n}{x}\right) &= \frac{D(r)x^r\sqrt{r}}{2\sqrt{\pi A}} \left(1 + O(r^{1/5})\right) \\ &= \exp\left(2\sqrt{A \log x} + O((\log x)^{\epsilon/2})\right) \end{aligned}$$

for any fixed $\epsilon > 0$. The quantity $\#\mathcal{G}_1(x)$ that we are actually interested in can be estimated in terms of $G(x)$ as follows: trivially,

$$\#\mathcal{G}_1(x) \ge G(x).$$

On the other hand, however, we have

$$\begin{aligned} \#\mathcal{G}_1(x) &\le \left(1 - \frac{1}{\log x}\right)^{-1} \sum_{\substack{n \le x \log x \\ n \in \mathcal{G}_1}} \left(1 - \frac{n}{x \log x}\right) \\ &= \left(1 - \frac{1}{\log x}\right)^{-1} G(x \log x) = \exp\left(2\sqrt{A \log x} + O((\log x)^{\epsilon/2})\right), \end{aligned}$$

which proves Theorem 2 and thus also completes the proof of Theorem 1. We remark that it is possible to obtain an asymptotic formula for $G(x)$ (and also for $\#\mathcal{G}_1(x)$) by further studying the behavior of $C(s)$ (near $s = 0$).

Acknowledgments. F. L. and S. W. thank Arnold Knopfmacher for useful discussions. Part of this work was done during a visit of F. L. and S. W. at the School of Mathematics of the University of the Witwatersrand in February of 2010. They thank the people of this institution for their hospitality. During the preparation of this paper, F. L. was also supported in part by grants SEP-CONACyT 79685 and PAPIIT 100508 and C. P. was supported in part by NSF grant DMS-0703850. S. W. was supported financially by the National Research Foundation of South Africa under grant number 70560.

References

[1] Y. Bilu, G. Hanrot, and P. M. Voutier. Existence of primitive divisors of Lucas and Lehmer numbers. *J. Reine Angew. Math.*, 539:75–122, 2001. With an appendix by M. Mignotte.

[2] H. G. Diamond. The distribution of values of Euler's phi function. In *Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972)*, pages 63–75. Amer. Math. Soc., Providence, R.I., 1973.

[3] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, 144(1-2):3–58, 1995. Special volume on mathematical analysis of algorithms.

[4] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Mellin transforms and asymptotics: digital sums. *Theoret. Comput. Sci.*, 123(2):291–314, 1994.

[5] F. Luca and Š. Porubský. The multiplicative group generated by the Lehmer numbers. *Fibonacci Quart.*, 41(2):122–132, 2003.

[6] C. L. Stewart. On divisors of Fermat, Fibonacci, Lucas, and Lehmer numbers. *Proc. London Math. Soc. (3)*, 35(3):425–447, 1977.

[7] C. L. Stewart and G. Tenenbaum. A refinement of the *abc* conjecture. preprint, 2005.