# UNIT FRACTIONS AND THE CLASS NUMBER OF A CYCLOTOMIC FIELD

Ernest S. Croot III and Andrew Granville

Abstract. We further examine Kummer's incorrect conjectured asymptotic estimate for the size of the first factor of the class number of a cyclotomic field.

## 1. Introduction

Let $h(p)$ be the class number of the cyclotomic field $\mathbb{Q}(\zeta_p)$ (where $\zeta_p$ is a primitive $p$th root of unity) and $h_2(p)$ be the class number of the real subfield $\mathbb{Q}(\zeta_p + \zeta_p^{-1})$. Kummer proved that the ratio $h_1(p) = h(p)/h_2(p)$ is an integer which he called the first factor of the class number, and in 1850 wrote in [8], "*La loi asymptotique des valeurs de ce premier facteur du nombre des classes H est exprimée par la formule*

$$h_1(p) \sim 2p\left(\frac{p}{4\pi^2}\right)^{\frac{p-1}{4}} = G(p), \tag{1}$$

*dont je me reserve la demonstration ...*" This proof never appeared and in [3] the second author showed that (1) is false if we assume two well-known conjectures of analytic number theory: First, Hardy and Littlewood's conjecture [5] that there are $\gg x/\log^2 x$ primes $p \le x$ for which $2p+1$ is also prime, and secondly the following conjecture on the distribution of primes in arithmetic progression "on average". Let $\pi(x;q,a)$ denote the number of primes $p \le x$ with $p \equiv a \pmod q$.

**Conjecture EH** (Elliott and Halberstam [1])**.** *For any fixed $\varepsilon > 0, A > 0$,*

$$\sum_{q<x^{1-\varepsilon}} |\pi(x;q,1) - \pi(x;q,-1)| \ll_{\varepsilon,A} \frac{x}{\log^A x}.$$

Despite this result, Kummer's law (1) does actually hold for almost all primes $p$, as was proved by Murty and Petridis in [9], assuming just Conjecture EH. In this paper we carefully study the set of exceptions to (1), under the assumption of suitable conjectures. We again use conjecture EH, but here we will need a more precise and more general version of the Hardy-Littlewood conjecture:

The second author is supported, in part, by the National Science Foundation

Typeset by $\mathcal{AMS}$-TEX

**Conjecture HL2** (Hardy and Littlewood [5]). *Suppose that $a_ix+b_i$ are distinct polynomials with integer coefficients, $1 \le i \le k$. Define $\omega(p)$ to be the number of distinct solutions $r \pmod p$ to $(a_1r+b_1)(a_2r+b_2)\dots(a_kr+b_k) \equiv 0 \pmod p$, and suppose $\omega(p) < p$ for all primes $p$. Then*

$$\#\{x < n \le 2x : \text{ Each } a_ip+b_i \text{ prime}\} \sim \prod_p \left(1-\frac{\omega(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \frac{x}{\log^k x}$$

*as $x \to \infty$.*

Note that if *any* $w(p) = p$ then, for every integer $n$, some $a_in+b_i$ is divisible by $p$, and so the $k$-tuple are only finitely often simultaneously prime. A set of polynomials for which every $\omega(p) < p$ is called "admissible".

**Theorem 1.** *Assume Conjectures EH and HL2.*

(i) *If $\alpha$ is a real number for which $\log\alpha$ is rational then there exists an integer $B(\alpha) \ge 1$ and a constant $C_\alpha > 0$ such that there are $\sim C_\alpha x/\log^{B(\alpha)} x$ primes $p \le x$ for which $h_1(p) \sim \alpha G(p)$.*

(Here, and throughout, "log" means log to base $e$)

(ii) *Fix $A > 0$. For any real number $\alpha$ for which $\log\alpha$ is irrational there are $\ll_\alpha x/\log^A x$ primes $p \le x$ for which $h_1(p) \sim \alpha G(p)$.*

Theorem 1 implies that $\{h_1(p)/G(p)\}$ is dense in $[0,\infty]$, which was also proved in Theorem 4 of [3]. The case $\alpha = 1$ here implies that (1) holds for "almost all" primes, which is Theorem 1.2 of [9]: Our proof of Theorem 1 was inspired by and, restricted to this case is essentially the same as, the proof of Theorem 1.2 in [9].

Conjecture A in [3] claims that

$$(\log\log p)^{-1/2+o(1)} \le h_1(p)/G(p) \le (\log\log p)^{1/2+o(1)}, \tag{2}$$

and that these bounds are best possible. This gives some idea of how large and small this ratio can be. It is also of interest to determine how often $h_1(p)/G(p)$ is large and small.

**Theorem 2.** *Assume Conjectures EH and HL2. For any given $A > 1$ there exists a constant $\kappa_A > 0$ and an integer $\beta(A) \ge 1$ such that there are $\sim \kappa_A x/\log^{\beta(A)} x$ primes $p \le x$ for which $h_1(p) \gtrsim AG(p)$. Similarly there are $\sim \kappa_A x/\log^{\beta(A)} x$ primes $p \le x$ for which $h_1(p) \lesssim G(p)/A$.*

**Corollary 1.** *Assume Conjectures EH and HL2. There exist constants $c_1, C > 0$ such that $G(p)/A < h_1(p) < AG(p)$ for all but $O(x/\log^{e^{c_1A^C}} x)$ primes $p \le x$, for any fixed $A > 1$.*

The conjecture in (2) is suggested if Corollary 1 holds uniformly in a wide range for $A$ with $C = 2$.

There are just two ideas here that do not appear in [3]. First, Hooley's version of the Brun-Titchmarsh theorem (see Lemma 3.1 below) as used in [9]; and second, the following seemingly unrelated result on unit fractions:

**Proposition 2.** *For every rational number $r$ there exists a finite set of integers $M$ with $r = \sum_{m\in M} 1/m$ such that, for every prime $p$, there exists a nonzero residue class mod $p$ which does not contain any element of the set $M$.*

Another main ingredient, which appears in other works, is the following Proposition.

**Proposition 1.** *Assume Conjecture HL2 and fix an $L > 1$. Then, if $M = \{m_1, ..., m_k\}$ is any set of integers so that*

$$\{x, m_1x + 1, ..., m_kx + 1\}$$

*is an admissible set of polynomials, then for $\gg_L x/\log^{|M|+1} x$ primes $p \le x$, we have that*

$$\{m \le L \ : \ |mp + 1| \text{ is prime}\} \ = \ M.$$

Using various techniques from analytic number theory, we will show that

$$2\log(h_1(p)/G(p)) = p\left(\sum_{q\equiv 1 \pmod p} 1/q - \sum_{q\equiv -1 \pmod p} 1/q\right) + O(\eta),$$

for all but $O(x/\log^A x)$ primes $p \le x$, where both sums are over the primes $q \le p/\eta$, and $\eta > 0$ is small. Thus if $M = \{m \in \mathbb{Z} : \ |mp + 1| \text{ is prime and } < p/\eta\}$ then $2\log(h_1(p)/G(p)) = \sum_{m\in M} 1/m + O(\eta)$. We now see how, from Proposition 2 and Conjecture HL2 we might deduce Theorem 1(i). Similarly since $\sum_{m\in M} 1/m$ is always rational, we see how we might deduce Theorem 1(ii).

We now discuss, in terms of unit fractions, how we determine the constants $B(\alpha), C_\alpha, \kappa_A, \beta(A)$ which appear in the two theorems above:

Define $\mathbb{M}$ to be those finite sets of distinct nonzero integers $M$ such that for every prime $p$ there exists a nonzero residue class mod $p$, which does not contain any element of the set $M$. If we define $\omega_M(p)$ to equal the number of distinct residue classes $\pmod p$ containing at least one element of $\{0\} \cup M$, then $\omega_M(p) < p$ for every $M \in \mathbb{M}$. For any $M \in \mathbb{M}$ let $n(M)$ be the number of elements in $M$, let $\Sigma(M) = \sum_{m\in M} 1/m$, and define $\Pi(M) = \prod_p \left((1 - \omega_M(p)/p)/(1 - 1/p)^{n(M)+1}\right)$.

Proposition 2 implies that $\Sigma(\mathbb{M}) = \{\Sigma(M) : \ M \in \mathbb{M}\} = \mathbb{Q}$. For any $r \in \mathbb{Q}$ define (by various abuses of notation) $n(r)$ to be the smallest $n$ such that there exists $M \in \mathbb{M}$ with $n(M) = n$ for which $\Sigma(M) = r$. We define $\Pi(r) = \sum_M \Pi(M)$ where the sum is over those $M \in \mathbb{M}$ for which $n(M) = n(r)$ and $\Sigma(M) = r$. Note that $n(0) = 0$ and $\Pi(0) = 1$. The constants in Theorem 1(i) are defined, when $\alpha$ is a real number for which $r = 2\log\alpha$ is rational, by $C_\alpha = \Pi(r)$ and $B(\alpha) = n(r) + 1$. To obtain the constants used in Theorem 2, for any given $A > 1$ let $\beta(A)$ denote one more than the size of the smallest set $M \in \mathbb{M}$ such that $\Sigma(M) \ge 2\log A$. Then let

$\kappa_A = \sum_M \Pi(M)$ where the sum is over those $M \in \mathbb{M}$ for which $n(M) = \beta(A) - 1$ and $\Sigma(M) \geq 2\log A$.

Let us note that that $\beta(A) = 2$ for $1 < A \leq e^{1/2}$. Evidently $\beta(A)$ is a non-decreasing function, and we now discuss its rate of growth for large $A$. What is obvious is that if $\beta(A) = N + 1$ then $2\log A \leq 1/1 + 1/2 + \cdots + 1/N \leq \log N + 1$ so that $\beta(A) \geq 1 + A^2/e$. In fact since any set $M \in \mathbb{M}$ has no elements in one congruence class mod $p$, for each prime $p$, we see by the sieve that $M$ has $\ll x/\log x$ elements $\leq x$. Partial summation then reveals that $\Sigma(M) \ll \log\log n(M)$; and so $2\log A \ll \log\log\beta(A)$; that is, $\log\beta(A) \gg A^C$, for some constant $C > 0$. Inserting this in Theorem 2 implies Corollary 1.

It would be nice to get some idea of the size of $n(r)$, how it relates to the size and height of $r$ (and thus $B(\alpha)$). Certainly our *ad hoc* construction used to prove Proposition 2 seems unlikely to reveal the truth. We do know, from the argument in the paragraph immediately above, that there exists a constant $c_1 > 0$ such that $n(r) \gg \exp(\exp(c_1|r|))$. On the other hand, from the arguments in section 7 making our ad hoc construction explicit, we find that there exists a constant $c_2 > 0$ such that $n(r) \ll \exp(O(q))\exp(\exp(\exp(c_2|r|)))$ where $q$ is the largest prime power divisor of the denominator of $r$. Presumably this upper bound can be improved though we hesitate to guess to what.

On a lighter note, we note that prime twin conjectures are essentially equivalent to certain conjectures about $h_1(p)$:

**Theorem 3A.** *Assume Conjecture EH, and let $m$ be an even positive integer, with $\delta = -1$ or $1$. There are $\gg_m x/\log^2 x$ primes $p \leq x$ for which $mp + \delta$ is also prime if and only if $h_1(p) \sim G(p)e^{\delta/2m}$ for $\gg_m x/\log^2 x$ primes $p \leq x$*

One can even get an unconditional result along these lines:

**Theorem 3B.** *Fix $c > e^2$. If $h_1(p) > cG(p)$ for $\gg x/\log^A x$ primes $p \leq x$ for all sufficiently large $x$, for some $A > 0$, then there exists an integer $m$ for which there are $\gg x/\log^A x$ primes $p \leq x$ with $mp + 1$ also prime. Analogously if $h_1(p) > G(p)/c$ for $\gg x/\log^A x$ primes $p \leq x$ then there exists an integer $m$ for which there are $\gg x/\log^A x$ primes $p \leq x$ with $mp - 1$ also prime.*

We remark that it would be awkward to show a general result exactly analogous to Theorem 3A for prime triplets since $\Sigma(\{4\}) = \Sigma(\{6, 12\})$, amongst many other examples.

It is of interest to try to find an unconditional proof that (1) is false. This seems unlikely in the near future since the ideas that go into Theorems 3 suggest that for primes $p$ for which (1) fails there must be primes $mp \pm 1$ with $m$ small, something that seems far from being unconditionally provable. (Indeed, it is not even known that there are infinitely many primes $p$ for which there is a prime $mp \pm 1$ with $m < p^{1/10}$.) However, by the argument of Theorem 3B we have found that we

can dispense with the assumption of Conjecture EH, at the expense of assuming a stronger prime $k$-tuple conjecture than in [3]:

**Theorem 4.** *Suppose that there exists a set $M$ of $k$ distinct positive integers with $|\sum_{m\in M} 1/m| > 4$, such that there are $\gg x/\log^{k+1} x$ primes $p \le x$ for which $mp+1$ is also prime for every $m \in M$. Then (1) fails for $\gg x/\log^{k+1} x$ primes $p \le x$.*

Note that by Proposition 2 and Conjecture HL2 we believe that there are many sets $M$ which satisfy the hypotheses of Theorem 4.

## 2. An analytic expression for $h_1(p)$

Hasse [6] showed that the value of $h_1(p)$ is equal to $G(p)$ times the product of the $L$-functions of the odd characters $\chi \pmod p$ at $s = 1$. By considering the value of this expression as $s$ goes towards 1 from above, one can deduce that

$$h_1(p) = G(p)\exp\left\{\frac{p-1}{2}f_p\right\}, \quad \text{where } f_p = \lim_{x\to\infty} f_p(x) \text{ and}$$

$$f_p(x) = \sum_{m\ge 1}\frac{1}{m}\left(\sum_{\substack{q \text{ prime}, q^m\le x\\ q^m\equiv 1 (\mathrm{mod} p)}}\frac{1}{q^m} - \sum_{\substack{q \text{ prime}, q^m\le x\\ q^m\equiv -1 (\mathrm{mod} p)}}\frac{1}{q^m}\right).$$

Note that the statement $h_1(p) \sim G(p)e^{r/2}$ is equivalent to $f_p \sim r/p$. Moreover Kummer's conjecture, (1), may be restated as $f_p = o(1/p)$.

The expression for $f_p$ is a little unwieldy but by employing a number of results of analytic number theory in [3] we showed how to simplify it:

- In Proposition 1 of [3] we showed that the prime powers in (3) contribute $o(1/p)$ to the total sum for all $O(\sqrt{x}\log^2 x)$ primes $p \le x$.
- At the start of section 3 of [3] we saw that a simple application of the Siegel-Walfisz Theorem implies that $f_p = f_p(2^p) + o(1/p)$, so that we can restrict our attention to the finite sum $f_p(2^p)$.

A similar argument using the Bombieri-Vinogradov Theorem would allow us to restrict our attention to the much smaller sum $f_p(p^{2+\delta})$ for any $\delta > 0$, for all but $O_A(x/\log^A x)$ primes $p \le x$. However, assuming conjecture EH, we have $f_p = f_p(p^{1+\delta}) + o(1/p)$, for all but $O_A(x/\log^A x)$ primes $p \le x$, by the argument in Proposition 2 of [3]. Therefore

$$f_p = \sum_{\substack{q\le p^{1+\delta}\\ q\equiv 1 \pmod p}}\frac{1}{q} - \sum_{\substack{q\le p^{1+\delta}\\ q\equiv -1 \pmod p}}\frac{1}{q} + o\left(\frac{1}{p}\right) \tag{3}$$

for all but $O(x/\log^A x)$ primes $p \le x$.

As in [9], the Brun-Titchmarsh Theorem ([4], Theorem 3.8) allows us to bound the contribution to $f_p$ of primes $q$ between $p^{1+\delta}$ and $p^{2+\delta}$ by an absolute constant. Combining this with the rest of our arguments one can recover Theorem 1.1 of [9], which states that $h_1(p) \asymp G(p)$ for almost all primes $p$. Actually we use this observation to prove Theorem 3B in section 8.

## 3. Lemmata old and new

We will use the following useful uniform version of the Brun-Titchmarsh theorem:

**Lemma 3.1** (Hooley [7]). *Let $\ell$ be a fixed nonzero integer. Suppose $A > B + 30$. If $\sqrt{Q} \le x < Q/\log^A Q$ then $\pi(Q; k, l) \le \{4 + o(1)\}Q/\phi(k)\log x$ for all but at most $O(x/\log^B Q)$ integers $k$, $x < k \le 2x$.*

**Corollary 3.2.** *Let $\ell$ be a nonzero integer and $A > C + 32$. For all but at most $O(x/\log^C Q)$ integers $k$, $x < k \le 2x$, we have $\pi(Q; k, l) \le \{4 + o(1)\}Q/\phi(k)\log x$ for all $Q$ in the range $x\log^A x \le Q \le x^2$.*

*Proof.* Let $Q_j = x\log^A x(1 + 1/\log x)^j$ for $j = 0, 1, \dots J$, where $J$ is the smallest integer for which $Q_J \ge x^2$. Note that $J \asymp \log^2 x$. We apply Lemma 3.1 for each $Q_j$ leading to $O(x/\log^C Q)$ exceptional $k$, where $C = B-2$. If $Q_{j-1} \le Q \le Q_j$ then, for any non-exceptional $k$, we have $\pi(Q; k, l) \le \pi(Q_j; k, l) \le \{4 + o(1)\}Q_j/\phi(k)\log x \le \{4 + o(1)\}Q/\phi(k)\log x$ since $Q_j \le Q(1 + 1/\log x)$.

**Lemma 3.3.** *Let $M$ be a finite set of distinct non-zero integers. Consider the set of polynomials $\{x\} \cup \{mx + 1 : m \in M,\ m > 0\} \cup \{-mx - 1 : m \in M,\ m < 0\}$. This set of polynomials is admissible (that is, each $\omega(p) < p$) if and only if for every prime $p$ there exists a nonzero residue class mod $p$, which does not contain any element of the set $M$.*

*Proof.* Note that, for this set of polynomials, $\omega(p) < p$ if and only if there is some integer $r$ such that $r\prod_{m\in M}(mr + 1) \not\equiv 0 \pmod p$. In other words $r \not\equiv 0 \pmod p$ and $r \not\equiv -1/m \pmod p$ for any $m \in M$. Taking $s \equiv -1/r \pmod p$ we have $s \not\equiv 0 \pmod p$ and $\not\equiv m \pmod p$ for any $m \in M$.

By the fundamental lemma of the sieve ([4], Theorem 2.6) one knows that

$$(4)\quad \#\{x < n \le 2x : \text{ Each } a_ip + b_i \text{ prime}\} \ll_k \prod_p \left(1 - \frac{\omega(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \frac{x}{\log^k x}$$

uniformly.

**Lemma 3.4.** *Suppose that $b_1, \dots, b_k$ are given integers, where only $b_1$ can possibly be zero. For any $1/2 \le A_1 \le A_2 \le \cdots \le A_k$ we have*

$$\sum_{A_1 < a_1 \le 2A_1} \sum_{A_2 < a_2 \le 2A_2} \cdots \sum_{A_k < a_k \le 2A_k} \prod_p \left(1 - \frac{\omega(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \asymp A_1A_2 \dots A_k,$$

*where the sums are restricted to avoid any cases where $a_ib_j = a_jb_i$ for some $1 \le i < j \le k$, and $A_1 = 1/2$ if $b_1 = 0$.*

*Proof.* We proceed by induction on $k$. Fix each $a_i, b_i$ for $1 \le i \le k-1$, and let $\omega'(p)$ denote the number of solutions $n \pmod p$ to $(a_1n+b_1)(a_2n+b_2)\dots(a_{k-1}n+b_{k-1}) \equiv 0 \pmod p$. Then

$$\prod_p \left(1-\frac{\omega(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \asymp \prod_p \left(1-\frac{\omega'(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-(k-1)} \prod_{p:\ a\in R_p}\left(1+\frac{1}{p}\right).$$

where $R_p$ is the set of integers $a_k \pmod p$ such that $\omega(p) = \omega'(p)$.

Now, if $\omega(p) = \omega'(p)$ then $p$ must divide $a_k\prod_{1\le j\le k-1}(a_kb_j - a_jb_k)$, which is $\ll A_k^{O(1)}$, and so the number of such $p > \varepsilon \log A_k$ is $\ll \log A_k/\log\log A_k$. Since each of these contribute no more than $1 + 1/p = 1 + O(1/\log A_k)$ to the last Euler product, they contribute, altogether $1 + O(1/\log\log A_k)$ which is irrelevant. Thus we can restrict the last Euler product to primes $p \le \varepsilon \log A_k$ so that its value only depends upon of $a_k \pmod m$ where $m$ is the product of the primes $\le \varepsilon \log A_k$, (and so $m < A_k^{2\varepsilon}$). Allowing $a_k$ to run through $m$ consecutive integers, this last Euler product is thus, on average,

$$\sim \prod_{p|m}\frac{1}{p}\left(\sum_{\substack{0\le a\le p-1\\ a\notin R_p}} 1 + \sum_{\substack{0\le a\le p-1\\ a\in R_p}}\left(1+\frac{1}{p}\right)\right) = \prod_{p|m}\left(1+\frac{|R_p|}{p^2}\right) \asymp 1,$$

since $|R_p| \le k$. Splitting the interval for $a_k$ into subintervals of length $m$, the result follows.

## 4. Unit fractions

To prove Proposition 2 we need the following result:

**Proposition 3.** *There exist integers $a_p$, with $1 \le a_p \le p-1$, for every prime $p$ such that $(a_p/p) = -1$ for all odd $p$; and if $M_0$ is the set of all positive integers for which $n \not\equiv a_p \pmod p$ for every $p$, then $\sum_{n\in M_0} 1/n$ diverges. In fact $\sum_{n\le x,\ n\in M_0} 1/n \gg \log\log x$.*

Note that in the introduction we saw that $\sum_{n\le x,\ n\in M_0} 1/n \ll \log\log x$, so that Proposition 3 is "best possible" up to the values of the implicit constant.

To prove Proposition 3 we will use an old result of Ron Graham [2]:

**Lemma 4.1** (Graham [2])**.** *A rational number $r$ can be written as the sum of the reciprocals of a finite number of distinct squares if and only if $0 \le r < \pi^2/6 - 1$ or $1 \le r < \pi^2/6$.*

In section 7 we will revisit this lovely result and make it a little more explicit.

*Deduction of Proposition 2.* Let $M_1 = M_0 \setminus \{(2b)^2 : b \in \mathbb{Z}\}$; since the sum of the squares converges, thus $\sum_{n\in M_1} 1/n$ diverges. Therefore, given $r > 0$ we can select a finite subset $M_2$ of $M_1$ such that $s = \sum_{n\in M_2} 1/n \in (r - 1/10, r)$. But then $0 < 4(r - s) < 2/5 < \pi^2/6 - 1$, and so by Lemma 4.1 there exists a finite set of squares $M_3$ such that $4(r - s) = \sum_{n\in M_3} 1/n$. Let $M_4 = \{4n : n \in M_3\}$. Since every element in $M_4$ is an even square, thus $M_4$ and $M_2$ are disjoint subsets of $M_0$. Therefore $M = M_2 \cup M_4 \subset M_0$ with $\sum_{n\in M} 1/n = s + 4(r - s)/4 = r$, so is the set required in the result.

We have shown in the above paragraph that for every rational $r > 0$ there is such a set $M_r$ with $r = \Sigma(M_r)$. For rational $-r < 0$ we simply take $M_r = -M_{-r}$. Finally let $M_0 = \emptyset$.

To prove Proposition 3, we will require yet another consequence of the fundamental lemma of sieve theory ([4], Theorem 2.5):

**Lemma 4.2.** *If $u$ is sufficiently large then, for* any *selection of $a_p \pmod p$ for $p \le x^{1/u}$ (with $x$ sufficiently large) there are $\gg_u x/\log x$ integers $n \in (x, 2x]$ such that $n \not\equiv a_p \pmod p$ for all primes $p \le x^{1/u}$.*

By splitting the interval $[y^u, y^{2u}/2]$ into dyadic intervals we deduce

**Corollary 4.3.** *Select $u$ sufficiently large. If $y$ is sufficiently large then, for* any *selection of $a_p \pmod p$ for $p \le y$, we have*

$$\sum_{\substack{y^u < n \le y^{2u}/2 \\ n\not\equiv a_p \pmod p \text{ for all } p\le y}} \frac{1}{n} \gg_u 1$$

*Proof of Proposition 3.* Select $u$ and $y$ sufficiently large as in corollary 4.3, and $y$ so that for every prime $p > y$ there are $\ge p/5$ integers $a \in (p/2, p)$ with $(a/p) = -1$ (there are in fact $\sim p/4$ such integers). We shall suppose that $a_2 = 1$, $a_3 = 2$ and $a_p \pmod p$ are all chosen with $(a_p/p) = -1$ for each $p \le y$. Let $y_0 = y, y_1 = y^{2u}$ and $y_k = y^{(2u)^k}$ for all $k \ge 0$.

We will select the values of $a_p$ for each successive prime $p > y$: For each $k = 0, 1, 2, \ldots$ we pick the $a_p$ for each $p$ from $y_k$ to $y_{k+1}$ so that $(a_p/p) = -1$ with $a_p \in (p/2, p)$ and

$$\sum_{\substack{y_k^u < n \le y_k^{2u}/2 \\ n\not\equiv a_q \pmod q \text{ for all } q<p \\ n\equiv a_p \pmod p}} \frac{1}{n}$$

minimal. Since there are at least $p/5$ feasible values of $a$, we know that one of these sums is

$$\le \frac{5}{p} \sum_{\substack{y_k^u < n \le y_k^{2u}/2 \\ n\not\equiv a_q \pmod q \text{ for all } q<p}} \frac{1}{n}.$$

Iterating this gives that

$$\sum_{\substack{y_k^u<n\le y_{k+1}/2\\ n\not\equiv a_q \pmod q \text{ for all } q\le y_{k+1}}} \frac{1}{n} \ge \prod_{y_k<p\le y_k^{2u}} \left(1-\frac{5}{p}\right) \sum_{\substack{y_k^u<n\le y_k^{2u}/2\\ n\not\equiv a_q \pmod q \text{ for all } q\le y_k}} \frac{1}{n} \gg_u 1$$

by Mertens' Theorem and Corollary 4.3. Since we choose all $a_p > p/2$ we see that for the $n$ counted in the first sum here, $n \not\equiv a_p \pmod p$ for all $p > y_{k+1}$ since then $n \le y_{k+1}/2 < p/2 < a_p < p$. Therefore every $n$ counted in this sum is in $M_0$. So, summing up over such sums we deduce that, if $y_{K+1} < x < y_{K+2}$ then

$$\sum_{\substack{n\le x\\ n\in M_0}} \frac{1}{n} \ge \sum_{k=1}^{K} \sum_{\substack{y_k^u<n\le y_{k+1}/2\\ n\in M_0}} \frac{1}{n} \gg_u K \gg \log\log x.$$

This gives our result, and is in fact best possible, up to the value of the implicit constant, by the remarks at the end of the introduction.

## 5. Further calculations: Small primes

**Proposition 1.** *Assume Conjecture EH. Fix integer $n \ge 2$ and select $\eta > 0$ arbitrarily small. All but $O(x/\log^{n+1/2} x)$ of the primes $x < p \le 2x$ have the following properties:*

(i) *There are no more than $n-1$ primes which are $\equiv \pm 1 \pmod p$ and $< p/\eta$.*

(ii) $f_p = \sum_{q\le p/\eta,\ q\equiv 1 \pmod p} 1/q - \sum_{q\le p/\eta,\ q\equiv -1 \pmod p} 1/q + O(\eta/p)$

*Proof.* We could take any value of $\alpha \in (0, 1/n)$ in the proof below, though we take $\alpha = 1/2n$ to get the result above. Let $L = (\log x)^\alpha$.

(i) The number of primes $x < p \le 2x$ for which there are $\ge n$ distinct primes $q \le pL$ with each $q \equiv \pm 1 \pmod p$ is less than or equal to the sum, over all possible integers $1 \le a_1 \le a_2 \cdots \le a_n \le L$ and $b_1, b_2, \dots, b_n \in \{-1, 1\}$ with the forms $a_i x + b_i$ all distinct, of

$$\#\{x < p \le 2x : \ p \text{ and each } a_i p + b_i \text{ prime, for } 1 \le i \le n\}.$$

By (4) and Lemma 3.4 this is $\ll_n xL^n/\log^{n+1} x \ll x/\log^{n+1/2} x$.

(ii) By (i) all but $O(x/\log^{n+1/2} x)$ of the primes $x < p \le 2x$ have no more than $n-1$ primes $q \le pL$ which are $\equiv \pm 1 \pmod p$, and also satisfy (3) (taking $\delta = \eta$ and $A = n+1$ there). Therefore (ii) follows if we can show that

$$\sum_{x/\eta\le q\le x^{1+\eta},\ q\equiv \pm 1 \pmod p} \frac{1}{q} \ll \eta/p$$

for all but $O(x/\log^{n+1} x)$ primes $x < p \le 2x$. First note that, for these primes $p$,

$$\sum_{x/\eta \le q \le pL,\ q \equiv \pm 1 \pmod p} \frac{1}{q} \le n/(x/\eta) \ll \eta/x \ll \frac{\eta}{p}.$$

Taking $C = n+1$ and $A = n+35$ in Corollary 3.2 we find, by partial summation, that

$$\sum_{\substack{x\log^A x < q < x^{1+\eta} \\ q \equiv \delta \pmod p}} \frac{1}{q} \ll \frac{\eta}{p},$$

for all but $O(x/\log^{n+1} x)$ primes $p,\ x < p \le 2x$.

To cover the remaining range, let $U = x^{1/L}$ and $\delta = \pm 1$ and let $P$ be the set of primes $p,\ x < p \le 2x$ for which

$$\sum_{\substack{Lp < q < Up \\ q \equiv \delta \pmod p}} \frac{1}{q} \ge \frac{\eta}{x}.$$

Take $m$ to be an integer $\ge n/\alpha$, so that

$$\sum_{x<p\le 2x} \left( \sum_{\substack{Lp<q<Up \\ q\equiv\delta \pmod p}} \frac{1}{q} \right)^m = \sum_{L<k_1,k_2,\dots,k_m<U} \sum_{\substack{x<p\le 2x \\ \text{each } k_ip+\delta \text{ prime}}} \prod_i \frac{1}{k_ip+1}$$

$$\ll \frac{1}{x^m} \sum_{r=1}^{m} \sum_{\substack{e_1,\dots,e_r \ge 1 \\ e_1+\cdots+e_r = m}} S_{e_1,\dots e_r}(L,U) \frac{x}{\log^{r+1} x}$$

by (3), where

$$S_{e_1,\dots e_r}(L,U) = \sum_{L<k_1<k_2<\cdots<k_r<U} \frac{1}{k_1^{e_1}\dots k_r^{e_r}} \prod_p \left(1 + \frac{r-\omega(p)}{p}\right)$$

and $\omega(p)$ is the number of distinct solutions $j \pmod p$ to $j(k_1j+\delta)\dots(k_rj+\delta) \equiv 0 \pmod p$. Lemma 3.4 and partial summation reveal that

$$S_{e_1,\dots e_r}(L,U) \ll (\log U)^{\#\{i:e_i=1\}}/L^{m-r} \ll (L\log U)^r/L^m = \log^r x/L^m.$$

We deduce that

$$|P|\left(\frac{\eta}{x}\right)^m \le \sum_{x<p\le 2x} \left( \sum_{\substack{Lp<q<Up \\ q\equiv\delta \pmod p}} \frac{1}{q} \right)^m \ll \frac{x}{\log x}\left(\frac{1}{Lx}\right)^m;$$

so that $|P| \ll x/\log^{1+m\alpha} x \ll x/\log^{n+1} x$.

## 6. Proofs of the theorems

When we apply Proposition 1 and study those primes satisfying (i) and (ii), we note that we can write all the primes $q \equiv \pm 1 \pmod p$ and $< p/\eta$ either as $mp+1$ with $0 < m < 2/\eta$, or $-mp-1$ with $-2/\eta < m < 0$. Letting $M$ be the set of such integers $m$, we see that $M$ has no more than $n-1$ elements and that $pf_p = \Sigma(M) + O(\eta)$.

**Lemma 6.1.** *The numbers in* $\Sigma(\mathbb{M}, n) := \{\Sigma(M) : \ M \subset \mathbb{Z}^*, \ |M| < n\}$ *are all rational. The limit points of the set are just* $\Sigma(\mathbb{M}, n-1)$. *Thus*

(i) *If* $r \notin \mathbb{Q}$ *then, for every* $n$ *there exists a constant* $\nu = \nu(r,n) > 0$ *such that there are no elements* $s \in \Sigma(\mathbb{M}, n)$ *with* $|r-s| < \nu$.

(ii) *If* $r \in \mathbb{Q}$ *then there exists a constant* $\nu(r) > 0$ *such that if* $s \in \Sigma(\mathbb{M}, n(r)+1)$ *with* $|r-s| < \nu$ *then* $r = s$.

*Proof.* Once we know that the limit points of $\Sigma(\mathbb{M}, n)$ are just $\Sigma(\mathbb{M}, n-1)$ then (i) follows since $r \notin \Sigma(\mathbb{M}) = \mathbb{Q}$, and (ii) follows since $r \notin \Sigma(\mathbb{M}, n(r))$.

To prove this, first note that if $M$ is a set of integers with $< n-1$ elements then $\lim_{a\to\infty} \Sigma(M \cup \{a\}) = \Sigma(M)$. On the other hand suppose that $r$ is a limit point of $\Sigma(\mathbb{M}, n)$ and select $n$ minimally so that this is so. Then $r = \lim_{i\to\infty} \Sigma(M_i)$ where the $M_i$ are distinct sets of $< n$ elements. Since the $M_i$ are distinct thus if $m_i$ is the $m \in M_i$ with largest absolute value, then $|m_i| \to \infty$ as $i \to \infty$. Therefore if $N_i = M_i \setminus \{m_i\}$ then $\lim_{i\to\infty} \Sigma(N_i) = \lim_{i\to\infty} \Sigma(M_i) - \lim_{i\to\infty} 1/m_i = r$. Since each $N_i$ has $< n-1$ elements there can only be finitely many distinct $N_i$, by the definition of $n$, and so $r = \Sigma(N_i) \in \Sigma(\mathbb{M}, n-1)$ for some $i$.

*Proof of Theorem 1 when* $r \notin \mathbb{Q}$. We apply Proposition 1 and Lemma 6.1(ii) with $n > A$ and $\eta > 0$ sufficiently small (with $\eta \ll \nu(r,n)$). Thus (i) and (ii) of Proposition 1 hold for all but $O(x/\log^A x)$ of the primes $x < p \le 2x$. For these $p$, if $h_1(p) \sim G(p)e^{r/2}$ then $r \sim pf_p = \Sigma(M) + O(\eta)$, which is impossible by Lemma 6.1(i).

*Proof of Theorem 1 when* $r \in \mathbb{Q}$. If $r = 0$ then apply Proposition 1 with $n = 2$, and $\eta$ small. We see that the number of primes $x < p \le 2x$ for which $|f_p| \gg \eta/p$ is

$$\ll \frac{x}{\log^{5/2} x} + \sum_{1\le m\le 2/\eta} \#\{x < p \le 2x : \ p, mp \pm 1 \text{ both prime}\}$$

$$\ll \frac{x}{\log^2 x} \left\{ \sum_{1\le m\le 2/\eta} \prod_{p\mid m} \left(1 - \frac{1}{p}\right)^{-1} \right\} \ll \frac{x}{\eta \log^2 x}$$

by Lemma 3.4. The result follows.

By Proposition 2 we know that if $r \in \mathbb{Q}$ then $r \in \Sigma(\mathbb{M})$. For $r \ne 0$ take $n = n(r)+1$ in Proposition 1, with $\eta$ far smaller than $|r|$, and far smaller than

$\nu(r)$ in Lemma 6.1(ii). Proposition 1 and Lemma 6.1(ii) imply that for all but $O(x/\log^{n(r)+3/2} x)$ of the primes $x < p \le 2x$, we have that $f_p \sim \{r + O(\eta)\}/p$ if and only if the set of primes $\equiv \pm 1 \pmod p$ which are $< p/\eta$, must be of the form $\{|mp+1| : \ m \in M\}$ where $n(M) = n(r)$ and $\Sigma(M) = r$. The result now follows immediately from Conjecture HL2.

*Proof of Theorem 2.* We wish to determine how often $f_p > 2\log A + o(1)$ (and, analogously, how often $f_p < -2\log A + o(1)$). We select $n = \beta_A$ in Proposition 1 with $\eta$ sufficiently small. Evidently if $f_p \ge 2\log A + O(\eta)$ for some prime $p$ in $(x, 2x]$ satisfying (i) and (ii) then the set of primes $\equiv \pm 1 \pmod p$ which are $< p/\eta$, must be of the form $\{|mp+1| : \ m \in M\}$ where $n(M) = \beta_A - 1$ and $\Sigma(M) \ge 2\log A$. The result now follows immediately from Conjecture HL2.

## 7. Unit fractions, revisited: Bounds on $n(r)$

Lemma 4.1 is a consequence of a rather more general result on unit fractions by Graham [2]. However the proof does not easily reveal *how many* reciprocals of distinct squares one needs to represent a given rational in the appropriate range. However we need to estimate this in order to bound $n(r)$. In this section we will give a proof of a version of Lemma 4.1, inspired by the proof in [2], which will allow us to bound these quantities:

Our first observation is about integers which can be written as the sum of distinct squares.

**Lemma 7.1.** *If $N \ge 9$ then for any integer $r$ in the range $129 \le r \le \sum_{n=1}^{N} n^2 - 129$ there exists a set $M$ of distinct integers from $[1, N]$ such that $r = \sum_{m\in M} m^2$.*

*Proof.* We proceed by induction. One can verify this by direct computation for $N = 9$ and 10. Otherwise, if it is true for $N-1$, then for any $r \in [129, \sum_{n=1}^{N-1} n^2 - 129]$ we can use the same set $M$ as before; whereas for any $r \in [N^2+129, \sum_{n=1}^{N} n^2 - 129]$ we can let $M = M' \cup \{N\}$ where $M'$ was the set used at the $N-1$st stage for representing $r - N^2$. These two intervals combine to give what was desired provided that $\sum_{n=1}^{N-1} n^2 - 129 + 1 \ge N^2 + 129$, which is true for $N \ge 11$.

By the same inductive proof we have the following more general result:

**Corollary 7.2.** *Suppose that $1 \le u_1 < u_2 < \cdots < u_N$ are a sequence of positive integers with $u_j = j$ for all $j \le 10 \le N$, with the property that $u_{k+1}^2 + 257 \le \sum_{j=1}^{k} u_j^2$ for each $k \ge 10$. Then for any integer $r$ in the range $129 \le r \le \sum_{n=1}^{N} u_n^2 - 129$ there exists a set $M$ of distinct integers from $[1, N]$ such that $r = \sum_{m\in M} u_m^2$.*

Let $N_y$ be the least common multiple of the integers $\le y$.

**Lemma 7.3.** *For $y \ge 96$, the set of proper divisors of $N_y$ satisfy the hypothesis of Corollary 7.2.*

*Proof.* We proceed by induction on $k$. The set $\{m : m|N_y\}$ contains all of the integers $\le y$ and these satisfy the hypothesis as we saw in the proof of Lemma 7.1. If we can show that $v < u_{k+1} \le \sqrt{2}v$ for some $v$ dividing $N_y$ then $v = u_r$ for some $r \le k$ so that $v \le u_k$ and thus $u_{k+1} \le \sqrt{2}u_k$ which leads to

$$\sum_{j=1}^{k} u_j^2 = u_k^2 + \sum_{j=1}^{k-1} u_j^2 \ge u_k^2 + (u_k^2 + 257) \ge u_{k+1}^2 + 257,$$

and we are done. We will use the fact that if $x$ is an integer then there is a prime in the interval $(x, x\sqrt{2})$ unless $x = 1, 2, 3$ or 7. This is easily proved using the explicit bounds in [10] and a little computation.

We now show how to find $v$ as above whenever $u_{k+1} < N_y/2$, provided $k \ge y \ge 96$. Let $p_\ell$ be the largest prime factor of $U = u_{k+1}$, and assume for now that $p_\ell \ge 11$. Suppose that there exists prime $p_j$, with $11 \le p_j < p_\ell$ such that $p_j^{e_j} \not| U$. Pick the largest such $j$, and then we can take $v = p_j U/p_{j+1}$ above. Thus we may assume that $p_j^{e_j}$ divides $U$ for $5 \le j \le \ell - 1$.

Similarly $2^{e_2-2}$ and $3^{e_3-1}$ divide $U$ else we can take $v = 8U/11$ or $9U/11$, respectively. Since $y \ge 9$ we know that 2 and 3 divide $U$ and so $5^{e_5}$ divides $U$ else we can take $v = 5U/6$. But then $2^{e_2-1}$ divides $U$ else we can take $v = 4U/5$; so that 4 divides $U$ and therefore $3^{e_3}$ divides $U$ else we can take $v = 3U/4$. But now we know that 9 divides $U$ and so $7^{e_7}$ divides $U$ else we can take $v = 7U/9$.

If $p_\ell = 7$ then $5^{e_5}$ divides $U$ else we can take $v = 5U/7$. In this case, or if $p_\ell = 5$, then $2^{e_2-1}$ divides $U$ else we can take $v = 4U/5$; so that 4 divides $U$ and therefore $3^{e_3}$ divides $U$ else we can take $v = 3U/4$.

We have thus proved that if $p_\ell \ge 5$ then $\prod_{i=1}^{\ell} p_i^{e_i}/U$ is one or two times a power of $p_\ell$. If $p_\ell^{e_\ell}$ does not divide $U$ then let $q = p_\ell$; otherwise $p_{\ell+1} \le y$ since $U < N_y/2$ and we let $q = p_{\ell+1}$. Now either $q+1$ or $q+3$ is divisible by 2 but not 4, and so can be written in the form $2a$ where $a$ is an odd number dividing $U$, so that we can take $v = qU/2a$. Therefore $p_\ell \le 3$.

If $p_\ell = 3$ then 9 divides $U$ else 32 divides $U$ (since $k \ge y \ge 96$), and so 9 divides $3^{e_3-2}$ divides $U$ else we can take $v = 27U/32$. Therefore $2^{e_2-2}$ divides $U$ else we can take $v = 8U/9$. But then we can take $v = 5U/6$. If $p_\ell = 2$ then $2^2$ divides $U$ by the induction hypothesis, so we can take $v = 3U/4$.

Finally if $u_{k+1} = N_y/2$ then $\sum_{j=3}^{7}(N_y/j)^2 > (N_y/2)^2$ and $\sum_{i=1}^{9} i^2 > 257$, as required.

**Theorem 7.4.** *For any rational number $p/q \in (0, 1/2]$ let $y$ equal the largest prime power divisor of $129q$. There exists a set $M$ of divisors of $N_y$ such that $p/q = \sum_{m\in M} 1/m^2$.*

*Proof of Theorem 7.4.* Let $r = pN_y^2/q$, so that $129 \le r \le \sum_{m|N_y,\ m<N_y} m^2 - 129$. Therefore, by Lemma 7.3 and Corollary 7.2, there exists a set $D$ of divisors of $N_y$,

with each $d \in D$ satisfying $d < N_y$ such that $r = \sum_{d \in D} d^2$. Dividing through by $N_y^2$ and writing $M = \{N_y/d : \ d \in D\}$ gives our result.

Combining Theorem 7.4 with Proposition 3 we can now get an effective version of Proposition 2:

**Proposition 2(ii).** *Let $r = p/q$. There exists a constant $c > 0$ such that if $y$ is the larger of $e^{e^{c|r|}}$ and the largest prime power divisor of $129q$ then we can select $M$ so that the integers in $M$ all divide the square of the least common multiple of the integers $\le y$. In particular $|M| \le e^{O(y)}$.*

*Remark.* A calculation reveals that all multiples of 13, up to 129, can be written as the sum of distinct squares of integers. This allows us to replace 129 by 13 in the two results immediately above, and this is the smallest such number.

## 8. A surprising equivalence: Class numbers and prime pairs

*Proof of Theorem 3A.* Take $n = 2$ in Proposition 1, so that assuming Conjecture EH, with $\eta > 0$ arbitrarily small, we have the following: For all but $o(x/\log^2 x)$ primes $x < p \le 2x$, there is no more than one prime of the form $mp \pm 1$ with $m < 1/\eta$. If there is such a prime then $pf_p = \delta/m + O(\eta)$; otherwise $pf_p = O(\eta)$. Therefore, letting $\eta \to 0$ as $x \to \infty$, we see that, for such primes $p$, we have that $mp \pm 1$ is also prime if and only if $h_1(p) \sim G(p)e^{\delta/2m}$. The result follows.

*Proof of Theorem 3B.* If we do not assume Conjecture EH then a version of Proposition 2 still holds, by the same proof. The only change that needs to be made is in the statement for $f_p$ in (ii), where we need to add the contribution for those primes $q \equiv \pm 1 \pmod p$ in the range $p^{1+\delta} < q < p^{2+\delta}$. As noted at the end of section 2, this extra contribution is bounded by $O(1/p)$: more explicitly we can bound this contribution by $\{4+o(1)\}/p$, for all but $O(x/\log^C x)$ primes $p \le x$, using Corollary 3.2.

Thus if $h_1(p) > cG(p)$ for $\gg x/\log^A x$ primes $p \le x$ then, by Proposition 1 with $\eta$ sufficiently small,

$$\sum_{\substack{q \le p/\eta \\ q \equiv 1 \pmod p}} \frac{1}{q} - \sum_{\substack{q \le p/\eta \\ q \equiv -1 \pmod p}} \frac{1}{q} \ge f_p - \frac{\{4+o(1)\}}{p} + O\left(\frac{\eta}{p}\right)$$
$$> \frac{\{2\log(c/e^2) + O(\eta)\}}{p} \gg \frac{1}{p}.$$

Thus the result follows for some $m < 1/\eta$.

*Proof of Theorem 4.* Take $n = k+1$ in the unconditional Proposition 1 (as described in the proof of Theorem 3B above), and then the result follows for those primes $p$ with $mp+1$ also prime for every $m \in M$ which satisfy both (i) and (ii).

## References


[1] P.D.T.A. Elliott and H. Halberstam, *A conjecture in prime number theory*, Symp. Math **4** (1968-69), 59-72.

[2] R.L. Graham, *On finite sums of unit fractions*, Proc. London Math. Soc **14** (1964), 193-207.

[3] A. Granville, *On the size of the first factor of the class number of a cyclotomic field*, Invent. Math **100** (1990), 321-338.

[4] H. Halberstam and H.E. Richert, *Sieve Methods*, Academic Press, New York, 1974.

[5] G. Hardy and J.E. Littlewood, *Some problems of 'partitio numerorum', III. On the expression of a number as a sum of primes*, Acta Math **44** (1923), 1-70.

[6] H. Hasse, *Über die Klassenzahl abelscher Zahlkörper*, Akademie-Verlag, Berlin (1952.).

[7] C. Hooley, *On the Brun-Titchmarsh Theorem, II*, Proc. London Math. Soc **30** (1975), 114-128.

[8] E.E. Kummer, *Mémoire sur la théorie des nombres complexes composés de racines de l'unité et des nombres entiers*, J. de math. pures et appl **16** (1851), 377-498; Collected Works, Vol.I., p.459.

[9] M.R. Murty and Y.N. Petridis, *On Kummer's conjecture*, J. Number Theory (to appear).
[10] J.B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math **6** (1962), 64-94..

Department of Mathematics, The University of California, Berkeley, California 94720-3840, USA
*E-mail address*: ecroot@math.berkeley.edu

Department of Mathematics, University of Georgia, Athens, Georgia 30602, USA
*E-mail address*: andrew@math.uga.edu