PRIME RECIPROCALS AND PRIMES IN ARITHMETIC PROGRESSION

DANIEL LITT

Abstract. This paper is an expository account of some (very elementary) arguments on sums of prime reciprocals; though the statements in Propositions 5 and 6 are well known, the arguments, to my knowledge, are original.

Dirichlet's theorem on primes in arithmetic progression states that if a, b are relatively prime positive integers, then there are infinitely many primes p satisfying $p \equiv a \bmod b$. We present a well-known methods of proving the special case $a = 1$, and alter it to obtain an elementary estimate on the sum

$$\sum_{p \equiv 1 \bmod b,\ p \text{ prime}} \frac{1}{p}$$

in some cases.

1. Introduction

Dirichlet's theorem on primes in arithmetic progression states

Theorem 1 (Dirichlet)**.** *Let a, b be relatively prime integers. Then there are infinitely many primes p satisfying*

$$p \equiv a \bmod b.$$

Consider the degenerate case $a = 0,\ b = 1$; then this is simply the claim that there are infinitely many primes. To prove this, we might consider the sum

$$\sum_{p \text{ prime}} \frac{1}{p}.$$

If this sum diverges, then there are certainly infinitely many primes.

Proposition 1. *The sum*

$$\sum_{p \text{ prime}} \frac{1}{p}$$

diverges.

Proof. The proof here is similar in spirit to that in Apostol [1], albeit slightly altered. Assume to the contrary that the sum converges; then there exists N such that

$$c := \sum_{p > N, p \text{ prime}} \frac{1}{p} < 1.$$

For a fixed prime p, let

$$G_p = 1 + \frac{1}{p} + \frac{1}{p^2} + \cdots = \frac{1}{1 - \frac{1}{p}}.$$

Now consider the expression

$$(1 + c + c^2 + c^3 + \ldots) \prod_{p \le N, p \text{ prime}} G_p.$$

As $c < 1$, the expression converges absolutely; but by absolute convergence, we may (rearranging terms) write

$$(1 + c + c^2 + c^3 + \ldots) \prod_{p \le N, p \text{ prime}} G_p \ge \sum_{n \in \mathbb{N}} \frac{1}{n}$$

which diverges by e.g. the integral test. So we have a contradiction. $\square$

This proposition suggests another approach—to consider the asymptotics of the sum

$$\sum_{p \text{ prime}, p<N} \frac{1}{p}.$$

We may easily show the following:

Proposition 2. *There exists a constant c, independent of N, such that*

$$\sum_{p \text{ prime},\ p<N} \frac{1}{p} \geq \ln \ln(N) + c.$$

Proof. We have

$$\begin{aligned}
\frac{\pi^2}{6} \cdot e^{\sum_{p<N} \frac{1}{p}} &= \left(\sum_{n \in \mathbb{N}} \frac{1}{n^2} \right) \prod_{p \text{ prime}, p<N} e^{\frac{1}{p}} \\
&\geq \left(\sum_{n \in \mathbb{N}} \frac{1}{n^2} \right) \prod_{p \text{ prime}, p<N} \left(1 + \frac{1}{p}\right) \\
&\geq \sum_{n<N} \frac{1}{n} \\
&\geq \int_1^N \frac{1}{x} dx \\
&= \ln N.
\end{aligned}$$

where the second line follows from the Talor series for e^x. Taking logarithms on both sides gives the desired claim. □

How might we generalize these proofs to the situation $(a, b) \neq (0, 1)$? Consider the following argument:

Proposition 3. *There are infinitely many primes p satisfying*

$$p \equiv 1 \bmod n$$

for any $n > 1$.

Before proceeding we need a lemma:

Lemma 1. *Let $f(x) \in \mathbb{Z}[x]$ be a non-constant polynomial. Let*

$$P_f := \{p \text{ prime} \mid \exists n \in \mathbb{N} \text{ s.t. } p | f(n) \neq 0\}.$$

Then P_f is infinite.

Proof of Lemma 1. Assume the contrary, and let $p_1, ..., p_k$ be an enumeration of P_f. Choose an integer s so that $f(s) = t \neq 0$; such an s exists as f is non-constant. Now note that

$$f(s + tp_1 \cdots p_k x) = f(s) + tp_1 \cdots p_k g(x) = t(1 + p_1 \cdots p_k g(x))$$

for some $g(x) \in \mathbb{Z}[x]$; in particular, $f(s + tp_1 \cdots p_k x)$ is divisible by t for any $x \in \mathbb{Z}$. Now consider $h(x) := \frac{1}{t} f(s + tp_1 \cdots p_k x) = 1 + p_1 \cdots p_k g(x)$. But h is non-constant, so we may choose $u \in \mathbb{Z}$ with $h(u) \neq 1$. So $h(u) \equiv 1 \mod p_1 \cdots p_k$, and thus $h(u)$ is divisible by some prime $p \neq p_i$ for $i = 1, ..., k$. But then $p \in P_f$, which is a contradiction. □

We now prove the proposition.

Proof of Proposition 3. Let $\Phi_n(x) \in \mathbb{Z}[x]$ be the n-th cyclotomic polynomial, that is, the minimal polynomial of a primitive n-th root of unity ζ_n over $\mathbb{Q}$.

Let $a \in \mathbb{Z}$ and consider p prime with $p \mid \Phi_n(a) \neq 0$, where $p \nmid n$. Let m be the order of a mod p; we claim that $n = m$. Indeed, $\Phi_n \mid (x^n - 1)$, so $p \mid a^n - 1$ and thus $m \mid n$. Assume $m < n$. But then $p \mid \Phi_n(a), a^m - 1$; but both $\Phi_n(x), x^m - 1$ divide $x^n - 1$, and the two polynomials are relatively prime mod p (indeed, the former is irreducible and does not divide the latter), so $x^n - 1$ has a double root mod p at a. But the discriminant of of $x^n - 1$ is n^n, which is non-zero mod p (as $p \nmid n$), so this is a contradiction. So we must have $m = n$.

But note that $a^{p-1} \equiv 1 \bmod p$, so $n \mid p-1$, and thus $p \equiv 1 \bmod n$. So any prime in $P_{\Phi_n(x)}$ either divides n or satisfies $p \equiv 1 \bmod n$.

But by Lemma 1, there are infinitely many primes in $P_{\Phi_n(x)}$, and only finitely many primes divide n, so there are infinitely many primes satisfying $p \equiv 1 \bmod n$. □

2. Sums of Reciprocals

Unfortunately, Proposition 3 does not answer the following question: Does

$$\sum_{p \equiv 1 \bmod n, p \text{ prime}} \frac{1}{p}$$

converge or diverge? Consider the case $n = 4$.

Our approach to Proposition 3 suggests looking at the number field $\mathbb{Q}[i]$, with ring of integers $\mathbb{Z}[i]$. Let N be the norm map $N : \mathbb{Z}[i] \to \mathbb{Z}$ given by $a + bi \mapsto a^2 + b^2$, with $a, b \in \mathbb{Z}$. Consider the set $N(\mathbb{Z}[i]) \subset \mathbb{Z}$. Let $r_2(n) = |N^{-1}(n)|$ We claim the following:

Proposition 4. *Let $n \in \mathbb{Z}$, $n = 2^s a_1 a_3$ where all the prime factors p of a_j satisfy $p \equiv j \bmod 4$. Then*

1. *$n \in N(\mathbb{Z}[i])$ only if a_3 is a square.*
2. *Let $b = p_1^{q_1} p_2^{q_2} \cdots p_k^{q_k}$ be the prime factorization of b. Then*

$$r_2(n) \le 2^{q_1+q_2+\cdots+q_k+2}.$$

Proof. (1) First, let p be an odd prime, and assume

$$p = x^2 + y^2$$

for $x, y \in \mathbb{Z}$. Noting that the quadratic residues mod 4 are $0, 1$, this implies that $p \equiv 0, 1, 2 \bmod 4$; as p is odd we have $p \equiv 1 \bmod 4$.

Now consider $n \in N(\mathbb{Z}[i])$, that is, $n = x^2 + y^2$. As $\mathbb{Z}[i]$ is a PID (and thus a UFD), $x + iy$ factors as $(a_1 + ib_1)^{q_1} \cdots (a_k + ib_k)^{q_k}$ for some primes $a_j + ib_j$. Note that N is multiplicative, so $x^2 + y^2 = (a_1^2 + b_1^2) \cdots (a_k^2 + b_k^2)$. So we may reduce to the case where $n = x^2 + y^2$ with $x + iy$ a prime. But then $N(x + iy) = (x + iy)(x - iy) = x^2 + y^2$, so $x + iy$ divides $N(x + iy)$. Let $p_1 \cdots p_k$ be the prime factorization of $N(x + iy)$ in $\mathbb{Z}$; as $x + iy$ is a prime, it must divide one of the p_j. But then $x - iy = \overline{x + iy}$ must divide $\overline{p_j} = p_j$. So $N(x + iy) = (x + iy)(x - iy) \mid p_j^2$, so $x^2 + y^2$ is either a prime or the square of a prime.

But then by the first paragraph, we have that if $x^2 + y^2 = p$ an odd prime, $p \equiv 1 \bmod 4$. So by writing $n = x^2 + y^2$, $x + iy = (a_1 + ib_1)^{q_1} \cdots (a_k + ib_k)^{q_k}$ we must have that the odd squarefree part of n is divisible only by primes $p \equiv 1 \bmod 4$, as desired.

(2) Note that the units of $\mathbb{Z}[i]$ (e.g. by analysis of the norm) are $\{1, -1, i, -i\}$. Note that for p a prime, $p \equiv 3 \bmod 4$, we have $r_2(p) = 0$ and $r_2(p^2) = 4$; that is, the preimages are $\{p, -p, ip, -ip\}$ (as such a prime cannot split, by the analysis above). For $p \equiv 1 \bmod 4$, we have $r_2(p) \le 8$ as such a prime may split into at most two primes (as $\mathrm{Gal}(\mathbb{Q}[i]/\mathbb{Q}) = \mathbb{Z}/2\mathbb{Z}$) of the form $x - iy, x + iy$. So the preimages of p are the four units multiplied by these two primes.

Using multiplicativity of the norm and the fact that $Z[i]$ is a UFD, we may write a preimage of N as $x = u(1-i)^s a_1' a_3'$ where u is a unit and a_1', a_3' are the preimages of a_1, a_3 respectively. We have four choices for u and one choice for the preimage of 2; having chosen a unit already, each prime factor of a_3' gives us no choice. Finally, for each prime factor of a_1' we may choose one of the at most two primes (up to a unit) lying above that prime in $\mathbb{Z}[i]$; let $t = q_1 + \cdots + q_k$ be the total number of primes (with multiplicity) dividing a_1. Then this analysis gives that $r_2(n) \le 4 \cdot 2^t = 2^{q_1+\cdots+q_k+2}$, as desired.

□

Remark 1. *Note that Proposition 4(1) is actually an if and only if; we omit the proof of the other direction, though it follows easily from an analysis of the splitting of primes in $\mathbb{Z}[i]$.*

We may use this argument to show:

Proposition 5. *The sum*

$$\sum_{p \text{ prime}, p\equiv 1 \bmod 4} \frac{1}{p}$$

diverges.

Proof. For convenience, let $P_{1,4}$ denote the set of primes $p \equiv 1 \bmod 4$. Assume the theorem is false; then there exists N such that

$$c := \sum_{p\in P_{1,4}, p>N} \frac{1}{p} < \frac{1}{2}.$$

Let G_p be as in the proof of Proposition 1, let

$$G'_p = 1 + \frac{2}{p} + \frac{2^2}{p^2} + \cdots = \frac{1}{1-\frac{2}{p}}$$

and consider the expression

$$D := 4 \cdot \left(\sum_{n\in\mathbb{N}} \frac{1}{n^2}\right) \cdot G_2 \cdot (1 + 2c + (2c)^2 + (2c)^3 + \cdots) \cdot \prod_{p\leq N, p\in P_{1,4}} G'_p.$$

Note that this expression converges absolutely, by our choice of N. For $n = 2^s a_1 a_3$ as in Proposition 4, with $a_1 = p_1^{q_1} \cdots p_k^{q_k}$, let $s_2(n) = \sum_j q_j$, and let $t(n) = a_1$. Then by absolute convergence of D, we may rearrange terms to achieve

$$\begin{aligned} D &\geq 4 \cdot \left(\sum_{n\in\mathbb{N}, t(n) \text{a square}} \frac{2^{s_2(n)}}{n}\right) \\ &= \sum_{n\in\mathbb{N}, t(n) \text{a square}} \frac{2^{s_2(n)+2}}{n} \\ &\geq \sum_{n\in\mathbb{N}} \frac{r_2(n)}{n} \\ &= \sum_{z\in\mathbb{Z}[i]^\times} \frac{1}{N(z)} \\ &= \sum_{(x,y)\in\mathbb{Z}^2-\{(0,0)\}} \frac{1}{x^2+y^2} \end{aligned}$$

where the inequality on the third line comes from Proposition 4(2). But this last expression diverges, as

$$\sum_{(x,y)\in\mathbb{Z}^2-\{0,0\}} \frac{1}{x^2+y^2} \geq \sum_{x\geq 0, y>0} \frac{1}{(x+y)^2} = \sum_{n\in\mathbb{N}} \frac{n}{n^2} = \sum_{n\in\mathbb{N}} \frac{1}{n} = \infty,$$

contradicting the claim that D converged absolutely. So we have the desired divergence. □

Remark 2. *This argument can easily be extended to the case $n = 3$, replacing the Gaussian integers with the Eisenstein integers; it proceeds essentially identically. Unfortunately, we cannot use an identical argument for primes $p \equiv 1 \bmod n$, $n > 5$ as the estimate in Proposition 4(2) relies heavily on the finiteness of the unit group of $\mathbb{Z}[i]$. The argument goes through, however, by comparing an expression analogous to D above to the Dedekind Zeta function of the number field $\mathbb{Q}[\zeta_n]$, where ζ_n is a primitive n-th root of unity—the Dedekind Zeta function diverges at 1, giving the desired comparison, but the proof of this is non-elementary and thus we will not exposit it here.*

We can, however, find a lower bound on the partial sums of

$$\sum_{p\in P_{1,4}} \frac{1}{p},$$

and an analogous argument works for $p \equiv 1 \bmod 3$. The proof follows similarly to that of Proposition 2.

Proposition 6. *There exists a constant c, independent from N, such that*

$$\sum_{p \in P_{1,4}, p<N} \frac{1}{p} \geq \frac{1}{2} \ln \ln(N) + c.$$

Proof. We have

$$\begin{aligned}
\frac{2\pi^2}{3} \cdot e^{2\sum_{p<N} \frac{1}{p}} &= 4 \left(\sum_{n \in \mathbb{N}} \frac{1}{n^2} \right) \prod_{p \text{ prime}, p<N} e^{\frac{2}{p}} \\
&\geq 4 \left(\sum_{n \in \mathbb{N}} \frac{1}{n^2} \right) \prod_{p \text{ prime}, p<N} \left(1 + \frac{2}{p}\right) \\
&\geq \sum_{n<N} \frac{2^{s_2(n)+2}}{n} \\
&\geq \sum_{n<N} \frac{r_2(n)}{n} \\
&\geq \sum_{0<x^2+y^2<N} \frac{1}{x^2+y^2} \\
&\geq \sum_{0<(x+y)^2<N;\ x\geq 0, y>0} \frac{1}{(x+y)^2} \\
&\geq \sum_{0<n<N} \frac{n}{n^2} \\
&\geq \ln N
\end{aligned}$$

where the second line follows from the Taylor series for e^x and the last line follows by bounding by an integral, as in the proof of Proposition 2. Taking logarithms on both sides gives the desired claim. □

3. Further Remarks

Unfortunately, it seems that generalizing this argument to large n as in Remark 1 would require estimates on the partial sums of the Dedekind Zeta function for $\mathbb{Q}[i]$; these estimates are far from elementary. We might ask how likely such methods are to work for bounding the sum of reciprocals of primes $p \equiv a \bmod b$ with $a \not\equiv 1 \bmod b$.

There are several negative results in this direction:

- First, to have a hope of using norms from the ring of integers of a number field to analyze primes $p \equiv a \bmod b$, we must be working in a number field with prime splitting controlled by some congruence conditions. But by results of Murty in [2], such number fields exist only if $a^2 \equiv 1 \bmod b$.
- Number fields whose prime splitting is controlled by congruence conditions are Abelian extensions of $\mathbb{Q}$, by Artin reciprocity; by the Kronecker-Weber Theorem, such fields are subfields of cyclotomic fields. So $a \equiv 1 \bmod b$ will often split, and by our own analysis the primes given by this splitting diverge, dominating or obscuring divergence by primes in other congruence classes mod b. (This is of course heuristic.)

Note also that Mertens' Theorem implies that the estimate in Proposition 2 is asymptotically sharp. Consider the following quantitative form of Dirichlet's theorem:

Theorem 2 (Dirichlet)**.** *Let $P_{a,b}$ be the set of primes congruent to $a \bmod b$, with $(a,b) = 1$. Then*

$$\frac{|P_{a,b} \cap \{1, ..., n\}|}{|P_{0,1} \cap \{1, ..., n\}|}$$

tends to $1/\phi(b)$ as n tends to infinity.

Together with Proposition 2, this implies that the estimate in Proposition 6 is also asymptotically sharp, as $\phi(4) = 2$.

References

[1] T. Apostol, *Introduction to Analytic Number Theory (Undergraduate Texts in Mathematics)*, Springer (May 28, 1998). Pgs. 18-19.

[2] M.R. Murty, *Primes in certain arithmetic progressions*, J. Madras Univ. (1988), 161-169.