

# An Elementary Proof of Dirichlet's Theorem on Primes in Arithmetic Progression

Jordi Gutiérrez Hermoso

March 30, 2002

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The Estimate <math>\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)</math></b>	<b>6</b>
2.1	A Couple of Integral Identities . . . . .	7
2.2	Dirichlet Convolution . . . . .	8
2.3	Generalized Convolution . . . . .	9
2.4	More Asymptotic Relations . . . . .	11
<b>3</b>	<b>Dirichlet characters</b>	<b>16</b>
3.1	A Few Preliminary Results . . . . .	16
3.2	Orthogonality Relation for Characters . . . . .	18
3.3	Dirichlet Characters . . . . .	18
3.4	Non-vanishing of $L(s, \chi)$ for Real Nonprincipal $\chi$ . . . . .	19
<b>4</b>	<b>The Extraction</b>	<b>22</b>
<b>5</b>	<b>Acknowledgments</b>	<b>27</b>
<b>A</b>	<b>Appendix</b>	<b>28</b>

# 1 Introduction

This paper is concerned with a proof of the following

**Theorem (Dirichlet's Theorem).** *Let  $U_m = (\mathbb{Z}/m\mathbb{Z})^\times$  be the multiplicative group of units modulo  $m$ . Suppose  $k \in U_m$ , that is,  $\gcd(k, m) = 1$ . Then there exist infinitely many primes  $p$  such that  $p \equiv k \pmod{m}$ .*

The statement of the theorem is simple, yet all known proofs use analytic methods. Dirichlet's theorem states that if  $(a + bn)_{n \in \mathbb{N}}$  is an arithmetic sequence of numbers where  $\gcd(a, b) = 1$  then there are infinitely many primes in this sequence. As a simple consequence of the proof we will see, it will also become apparent that the distribution of primes in every arithmetic sequence of this kind is rather uniform. This is reassuring, in a sense, for intuitively we feel that primes are more or less randomly spread out over the integers. Furthermore, there are many results that depend on the existence of infinitely many primes of a particular kind, and Dirichlet's theorem gives us these primes in certain cases.

Certain special cases of Dirichlet's theorem can be dispatched with relative ease.

**Theorem 1.1.** *There are infinitely many primes congruent to 3 modulo 4.*

*Proof.* Suppose not. Let  $p_1, p_2, \dots, p_k$  be all primes congruent to 3 modulo 4. Put  $N = 4 \prod_{n=1}^k p_n - 1$ . Thus,  $N \equiv 3 \pmod{4}$ . We have that  $N > p_k$ , so it cannot be prime, and no prime less than  $p_k$  divides  $N$ . This means that all of the prime factors of  $N$  are congruent to 1 modulo 4. This is a contradiction, as  $1 \cdot 1 \equiv 1 \pmod{4}$ , but  $N \equiv 3 \pmod{4}$ .  $\square$

**Theorem 1.2.** *There are infinitely many primes congruent to 1 modulo 4.*

*Proof.* Say  $N$  be any any integer greater than 1. We will produce a prime  $p \equiv 1 \pmod{4}$ . Let  $m = (N!)^2 + 1$ . Notice that  $m$  is odd. Let  $p$  be the smallest prime divisor of  $m$ . We have that  $p > N$ . So,

$$\begin{aligned}(N!)^2 &\equiv -1 \pmod{p} \\ (N!)^{p-1} &\equiv (-1)^{\frac{p-1}{2}}\end{aligned}$$

By the Euler-Fermat theorem,  $(N!)^{p-1} \equiv 1 \pmod{p}$ , which yields

$$(-1)^{\frac{p-1}{2}} \equiv 1 \pmod{p}$$

So  $(-1)^{\frac{p-1}{2}} - 1$  is either 0 or 2, but it cannot be 2, as  $p$  divides it. This implies that  $(-1)^{\frac{p-1}{2}}$  is 1, so  $(p-1)/2$  is even, i.e.  $p \equiv 1 \pmod{4}$ .  $\square$

These simple kind of arguments, however, admit no simple generalization. There are other known simple proofs for other special cases, but the general one requires analytic methods. The proof we will present here makes no usage of complex analysis, nor of any other heavy machinery. It is, in fact, an elementary proof. It is by no means simple, however, so we will present an outline of the proof.

## Outline of the proof

To give a flavour of the sort of proof we will see, consider the following simple example on the infinitude of primes.

**Theorem.** *There are infinitely many primes.*

*Proof.* We will show that

$$\sum_{n=1}^{\infty} \frac{1}{p_n}$$

diverges, where  $p_n$  denotes the  $n$ th prime. This will suffice, for if there were finitely many primes, the sum would then converge. Suppose that this is so, i.e. the sum converges. Then we would have some  $k \in \mathbb{Z}$  such that

$$\sum_{m=k+1}^{\infty} \frac{1}{p_m} < \frac{1}{2}$$

Let  $Q = \prod_{m=1}^k p_m$ . Consider the sequence  $(1 + nQ)_{n \in \mathbb{N}}$ . We have that  $p_n$  does not divide  $1 + nQ$  for  $n \leq k$ . Thus, all prime factors of  $1 + nQ$  occur among the primes  $p_{k+1}, p_{k+2}, \dots$ . So, for each  $r \geq 1$ , we have

$$\sum_{n=1}^r \frac{1}{1 + nQ} \leq \sum_{t=1}^{\infty} \left( \sum_{m=k+1}^{\infty} \frac{1}{p_m} \right)^t$$

This follows because the sum on the right hand side contains all the terms of the one the left. This can be seen by expanding the sum and distributing with the binomial theorem accordingly. However, the sum on the right is bounded by the convergent geometric series  $\sum_{t=1}^{\infty} (1/2)^t$ , so it converges, from which it follows that  $\sum_{n=1}^{\infty} \frac{1}{1+nQ}$  converges as well, as it has bounded partial sums. This is a contradiction, as a limit comparison test with  $\sum_{n=1}^{\infty} \frac{1}{n}$  shows that  $\sum_{n=1}^{\infty} \frac{1}{1+nQ}$  diverges.  $\square$

The proof above conveys some of the spirit of the proof of Dirichlet's theorem that we will follow. We will prove Dirichlet's theorem by establishing the divergence of a certain series. Our first step will be to establish the following asymptotic relation:

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1) \tag{1}$$

The notation means that the summation is extended over all the primes less than  $x$ . The  $O(1)$  term denotes an unspecified function of  $x$  that remains bounded as  $x \rightarrow \infty$ . Establishing the above relation will be no small task. In our way to do so, we will build many tools that will later prove useful.

Notice that relation (1) already implies that there are infinitely many primes, for in the right hand side we have  $\log x$ , which is unbounded. This could not happen if there were only finitely many primes.

The next step will be a sort of “extraction”. We will use a certain family of arithmetic functions, the Dirichlet characters, for this purpose. This will lead us to develop a few properties of Dirichlet characters.

Once we have done this, we will be ready for the extraction. This will yield the following related asymptotic relation

$$\sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + O(1) \quad (2)$$

This time, the notation says that the sum is extended only over the primes congruent to  $h$  modulo  $k$ . It is remarkable that the right-hand side of equation (2) is independent of  $h$ , and that there is a  $1/\varphi(k)$  factor in front of  $\log x$ . This suggests that the distribution of the primes over the congruence classes modulo  $k$  is more or less uniform, and indeed it is.

Lastly, we should mention a crucial step in the extraction. Every Dirichlet character  $\chi$  gives rise to one of Dirichlet’s  $L$ -series:

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

For the purposes of this paper, we will only be concerned with the family of series where  $s = 1$ . If  $\chi$  is a nonprincipal character, meaning that it takes on values other than 0 and 1, we need to know that  $L(1, \chi) \neq 0$ , since we will need to divide by this value at one point.

Let us begin, then, on the path to establish relation (1). We parallel very closely the development in Apostol’s *Introduction to Analytic Number Theory* [1]

## 2 The Estimate $\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$

Some notation first.

**Notation.** Throughout this paper,  $p$  will denote a positive prime number,  $n, m, h, k$  will be natural numbers (zero is not a natural number for us), and  $x, y$  are real numbers.  $[x]$  is the greatest integer less than or equal to  $x$ . An expression of the form

$$\sum_{n \leq x} f(x)$$

means

$$\sum_{n=1}^{[x]} f(x)$$

and

$$\sum_{p \leq x} f(x)$$

means that the sum is extended over all positive primes less than or equal to  $x$ . All these sums are 0 if  $x < 1$ .

A couple of definitions are now in order.

**Definition 2.1.** A function  $f : \mathbb{N} \rightarrow \mathbb{C}$  is an *arithmetical function*.

**Definition 2.2 (Big-Oh notation).** We write, for any function  $f : \mathbb{R} \rightarrow \mathbb{C}$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ .

$$f(x) = O(g(x))$$

to mean that the quotient  $|f(x)|/g(x)$  remains bounded for all  $x$  greater than some  $a$ , that is to say, there exists a constant  $M$  such that

$$|f(x)| \leq M g(x) \quad \forall x \geq a$$

In this case, we say that “ $f$  is big-oh of  $g$ ”. Furthermore, an expression of the form

$$f(x) = g(x) + O(h(x))$$

means

$$f(x) - g(x) = O(h(x))$$

So, for example,  $f(x) = O(1)$  means that  $f$  remains bounded as  $x \rightarrow \infty$ . The symbol  $O(f(x))$  is subject to the following relations, on which we will not elaborate. Notice that we are making usage of the equality sign in a non-symmetric manner.

1.  $O(O(g(x))) = O(g(x))$
2.  $O(g(x)) \pm O(g(x)) = O(g(x))$
3.  $f(x)O(g(x)) = O(f(x)g(x))$
4.  $O(f(x)) = O(g(x))$  if  $f(x) \leq g(x)$  for all  $x \geq a$  (In other words, we may always give a weaker estimate.)

We will be making extensive use of the above relations and similar others, without making explicit reference to them. They can be found in Chapter 6 of LeVeque [2]. The reader is advised to become thoroughly familiarized with them; they all follow easily from the definition.

## 2.1 A Couple of Integral Identities

In this section, we collect some of the tools that we will need later. Not everything here will be used immediately, but should be kept in mind for future reference.

The following fundamental identity leads to many other useful conclusions.

**Theorem 2.3 (Abel's Identity).** *For any arithmetical function  $a(n)$ , define  $A(x) = \sum_{n \leq x} a(n)$ . Suppose  $f$  has a continuous derivative on the closed interval  $[y, x]$ , with  $0 < y < x$ . Then we have,*

$$\sum_{y < n \leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t) dt$$

*Proof.* This is nothing more than integration by parts using the Riemann-Stieltjes integral. Notice that  $A(x)$  is a step function with jump  $a(n)$  at each integer  $n$ , so

$$\sum_{y < n \leq x} a(n)f(n) = \int_y^x f(t) dA(t)$$

Integration by parts gives

$$\begin{aligned} \int_y^x f(t) dA(t) &= f(t)A(t)|_y^x - \int_y^x A(t) df(t) \\ &= f(x)A(x) - f(y)A(y) - \int_y^x A(t)f'(t) dt \end{aligned}$$

□

Sometimes we need to approximate a sum with an integral. The following corollary tells us what the error will be with such an approximation.

**Corollary 2.4 (Euler's Summation Formula).**

$$\sum_{y < n \leq x} f(n) = \int_y^x f(t) dt + \int_y^x (t - [t])f'(t) dt + f(x)([x] - x) - f(y)([y] - y)$$

*Proof.* Put  $a(n) = 1$  for all  $n$  in Abel's identity, so  $A(x) = [x]$ . This gives

$$\sum_{y, n \leq x} f(n) = f(x)[x] - f(y)[y] - \int_y^x [t]f'(t) dt$$

We also have, by integration by parts,

$$0 = \int_y^x tf'(t) dt - xf(x) + yf(y) + \int_y^x f(t) dt$$

Adding the two equations gives the result.  $\square$

Euler's summation formula leads to many interesting asymptotic relations. We need not concern ourselves with them for now. The few that are needed for this paper are collected in the appendix.

## 2.2 Dirichlet Convolution

In this section we will assume a few well-known results concerning Dirichlet convolution. They are summarised below for convenience. They can be found in Anthony Gioia's *The Theory of Numbers*.

**Definition 2.5.** Let  $f$  and  $g$  be arithmetical functions. Their *Dirichlet convolution* is another arithmetical function  $h$  denoted  $h = f * g$  defined by

$$h(n) = \sum_{d|n} f(d)g\left(\frac{n}{d}\right)$$

where the sum is extended over all divisors  $d$  of  $n$ .

We also will find need for the following notion.

**Definition 2.6.** An arithmetical function  $f$  is said to be *multiplicative* if  $f(mn) = f(m)f(n)$  whenever  $\gcd(m, n) = 1$ . It is *completely multiplicative* if  $f(mn) = f(m)f(n)$  for all  $m, n \in \mathbb{N}$ . Also, we insist that  $f \equiv 0$  is not a multiplicative function.

Arithmetical functions under Dirichlet convolution behave as follows:

- Dirichlet convolution is associative and commutative.
- An arithmetical function  $f$  has a Dirichlet inverse whenever  $f(1) \neq 0$ . In other words, the arithmetical functions that are nonzero at 1 form an abelian group. The identity is the function  $\varepsilon(n) = [1/n]$ .

- The Dirichlet convolution of two multiplicative functions is multiplicative.
- The Dirichlet inverse of a multiplicative function is also multiplicative. In other words, the multiplicative functions form a normal subgroup.

We now restrict our attention to particular arithmetical functions that will become useful in establishing the asymptotic relations we seek.

**Definition 2.7.** The *Möbius function*  $\mu$  is the inverse under Dirichlet convolution of the function  $u \equiv 1$ . The *von Mangoldt function*  $\Lambda$  is the Dirichlet convolution of  $\mu$  with  $\log$ .

From this definition we can infer the following:

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ is divisible by the square of some prime} \end{cases}$$

and

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^m \text{ for some prime } p \text{ and some } m \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

The Möbius function gives us a useful identity, known as the *Möbius inversion formula*.

**Proposition (Möbius Inversion Formula).**

$$g(n) = \sum_{d|n} f(d)$$

if and only if

$$f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right)$$

We mention another result without proof.

**Proposition (Inverse of Completely Multiplicative Function).** for  $f$  completely multiplicative, we have that  $f^{-1}(n) = \mu(n)f(n)$ .

## 2.3 Generalized Convolution

We will now establish the generalized Möbius inversion formula that will play a prominent role in many of the results that will come.

**Definition 2.8.** Let  $F$  denote a real or complex-valued function on  $\mathbb{R}^+$  such that  $F(x) = 0$  for  $0 < x < 1$ , and let  $\alpha$  be an arithmetical function. The sum

$$G(x) = \sum_{n \leq x} \alpha(n) F\left(\frac{x}{n}\right)$$

is a *generalized convolution* of  $\alpha$  and  $F$  and is denoted here by  $G = \alpha \odot F$ .

Notice that  $G = \alpha \odot F$  also vanishes on the open interval  $(0, 1)$ . The reason why this is called a *generalized* convolution becomes clear if we consider what happens if  $F(x)$  is zero for all non-integral  $x$ . In this case,  $\alpha \odot F = \alpha * F$ , and we obtain Dirichlet convolution.

Generalized convolution does not behave as nicely as Dirichlet convolution does. It is not commutative, nor even associative. However, we do have the following useful replacement for associativity.

**Theorem 2.9 (Associative Rule).** *Let  $\alpha, \beta$  be arithmetical functions, and  $F$  as above. Then we have  $\alpha \odot (\beta \odot F) = (\alpha * \beta) \odot F$ .*

*Proof.*

$$\begin{aligned}
(\alpha \odot (\beta \odot F))(x) &= \sum_{n \leq x} \alpha(n) \sum_{m \leq x/n} \beta(m) F\left(\frac{x}{mn}\right) \\
&= \sum_{mn \leq x} \alpha(n) \beta(n) F\left(\frac{x}{mn}\right) \\
&= \sum_{k \leq x} \left( \sum_{n|k} \alpha(n) \beta\left(\frac{k}{n}\right) F\left(\frac{x}{k}\right) \right) \\
&= \sum_{k \leq x} (\alpha * \beta)(k) F\left(\frac{x}{k}\right) \\
&= ((\alpha * \beta) \odot F)(x)
\end{aligned}$$

□

Notice that the same function  $\varepsilon(n) = [1/n]$  that is the identity under Dirichlet convolution works as an identity under generalized convolution. From this we infer the generalized inversion formula.

**Theorem 2.10 (Generalized Inversion Formula).** *If  $\alpha$  has a Dirichlet inverse  $\alpha^{-1}$ , we then have*

$$G(x) = \sum_{n \leq x} \alpha(n) F\left(\frac{x}{n}\right)$$

*if and only if*

$$F(x) = \sum_{n \leq x} \alpha^{-1}(n) G\left(\frac{x}{n}\right)$$

*Proof.*  $G = \alpha \odot F$ , so  $\alpha^{-1} \odot G = (\alpha^{-1} \odot (\alpha \odot F)) = (\alpha^{-1} * \alpha) \odot F = F$ . The converse is similar. □

In particular, if  $\alpha$  is completely multiplicative we have the generalized Möbius inversion formula.

**Theorem 2.11 (Generalized Möbius Inversion Formula).** *If  $\alpha$  is completely multiplicative, we have*

$$G(x) = \sum_{n \leq x} \alpha(n) F\left(\frac{x}{n}\right) \iff F(x) = \sum_{n \leq x} \mu(n) \alpha(n) G\left(\frac{x}{n}\right)$$

## 2.4 More Asymptotic Relations

Using the tools we have developed so far, we now build more. Their usefulness will be proven shortly.

**Theorem 2.12.** *If  $h := f * g$ , and if*

$$F(x) := \sum_{n \leq x} f(n), G(x) := \sum_{n \leq x} g(n), H(x) := \sum_{n \leq x} h(n)$$

then

$$H(x) = \sum_{n \leq x} f(n) G\left(\frac{x}{n}\right) = \sum_{n \leq x} g(n) F\left(\frac{x}{n}\right)$$

*Proof.* Put

$$U(n) \begin{cases} 0 & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \tag{3}$$

Then  $F = f \odot U$  and  $G = g \odot U$ , so

$$f \odot G = f \odot (g \odot U) = (f * g) \odot U = H$$

$$g \odot F = g \odot (f \odot U) = (g * f) \odot U = H$$

□

In particular, if  $g \equiv 1$  (so that  $G(x) = [x]$ ), we obtain

**Corollary 2.13.**

$$\sum_{n \leq x} \sum_{d|n} f(d) = \sum_{n \leq x} f(n) \left[\frac{x}{n}\right] = \sum_{n \leq x} F\left(\frac{x}{n}\right)$$

Now, put  $f$  above to be the von Mangoldt function. This gives us another identity.

**Corollary 2.14.** For  $x \geq 1$  we have

$$\sum_{n \leq x} \Lambda(n) \left[\frac{x}{n}\right] = \log[x]!$$

*Proof.* Put  $f$  to be the von Mangoldt function in Corollary 2.13

$$\begin{aligned}\sum_{n \leq x} \Lambda(n) \left[ \frac{x}{n} \right] &= \sum_{n \leq x} \sum_{d|n} \Lambda(d) \\ &= \sum_{n \leq x} \log(n) = \log[x]!\end{aligned}$$

□

**Lemma 2.15.**

$$\sum_{n \leq x} \Lambda(n) \left[ \frac{x}{n} \right] = x \log x - x + O(\log x)$$

*Proof.* Put  $f(t) = \log t$  in Euler's summation formula. We get

$$\begin{aligned}\sum_{n \leq x} \log n &= \int_1^x \log t \, dt + \int_1^x \frac{t - [t]}{t} \, dt - (x - [x]) \log x \\ &= x \log x - x + 1 + \int_1^x \frac{t - [t]}{t} \, dt + O(\log x)\end{aligned}$$

But

$$\int_1^x \frac{t - [t]}{t} \, dt = O\left(\int_1^x \frac{1}{t} \, dt\right) = O(\log x)$$

and this proves the result, as  $\sum_{n \leq x} \log n = \log[x]! = \sum_{n \leq x} \Lambda(n) \left[ \frac{x}{n} \right]$  by Corollary 2.14. □

Our next theorem is one step away from the asymptotic relation from which we will extract our primes in  $U_k$ .

**Theorem 2.16.** *For  $x \geq 2$  we have*

$$\sum_{p \leq x} \left[ \frac{x}{p} \right] \log p = x \log x + O(x)$$

where the sum is over all primes  $\leq x$ .

*Proof.* Since  $\Lambda(n) = 0$  unless  $n$  is a prime power, we have

$$\sum_{n \leq x} \left[ \frac{x}{n} \right] \Lambda(n) = \sum_p \sum_{\substack{m=1 \\ p^m \leq x}}^{\infty} \left[ \frac{x}{p^m} \right] \Lambda(p^m)$$

Now,  $p^m \leq x$  implies  $p \leq x$ . Also,  $[x/p^m] = 0$  if  $p > x$ , so write

$$\sum_{p \leq x} \sum_{m=1}^{\infty} \left[ \frac{x}{p^m} \right] \log p = \sum_{p \leq x} \left[ \frac{x}{p} \right] \log p + \sum_{p \leq x} \sum_{m=2}^{\infty} \left[ \frac{x}{p^m} \right] \log p$$

We will show that the term on the right in the above sum is  $O(x)$ . We have

$$\begin{aligned} \sum_{p \leq x} \log p \sum_{m=2}^{\infty} \left[ \frac{x}{p^m} \right] &\leq \sum_{p \leq x} \log p \sum_{m=2}^{\infty} \frac{x}{p^m} = x \sum_{p \leq x} \log p \sum_{m=2}^{\infty} \left( \frac{1}{p} \right)^m \\ &= x \sum_{p \leq x} \log p \frac{1}{p^2(1-1/p)} = x \sum_{p \leq x} \frac{\log p}{p(p-1)} \\ &\leq x \sum_{n=2}^{\infty} \frac{\log n}{n(n-1)} = O(x) \end{aligned}$$

since the last sum converges. So,

$$\sum_{n \leq x} \left[ \frac{x}{n} \right] \Lambda(n) = \sum_{p \leq x} \left[ \frac{x}{p} \right] \log p + O(x)$$

and by Lemma (2.15), since  $O(x)$  is a weaker bound than  $O(\log x)$ ,

$$\sum_{p \leq x} \left[ \frac{x}{p} \right] \log p = x \log x + O(x)$$

□

Our result would be complete if we could somehow justify dropping the square brackets in Theorem (2.16). This justification is essentially the content of what follows.

**Theorem 2.17.** *Let  $a(n)$  be a nonnegative arithmetical function such that*

$$\sum_{n \leq x} \left[ \frac{x}{n} \right] a(n) = x \log x + O(x) \text{ for all } x \geq 1$$

*Then,*

1. *There is a  $B > 0$  such that  $\sum_{n \leq x} a(n) \leq Bx$  for all  $x \geq 1$ .*
2. *For  $x \geq 1$  we have  $\sum_{n \leq x} \frac{a(n)}{n} = \log x + O(1)$*

*Proof.* Put

$$S(x) = \sum_{n \leq x} a(n) \text{ and } T(x) = \sum_{n \leq x} a(n) \left[ \frac{x}{n} \right]$$

We will first establish the following inequality

$$S(x) - S\left(\frac{x}{2}\right) \leq T(x) - T\left(\frac{x}{2}\right)$$

Write

$$\begin{aligned} T(x) - 2T\left(\frac{x}{2}\right) &= \sum_{n \leq x} a(n) \left[\frac{x}{n}\right] - 2 \sum_{n \leq x/2} a(n) \left[\frac{x}{2n}\right] \\ &= \sum_{n \leq x/2} \left(\left[\frac{x}{n}\right] - 2\left[\frac{x}{2n}\right]\right) a(n) + \sum_{x/2 < n \leq x} \left[\frac{x}{n}\right] a(n) \end{aligned}$$

but  $\left[\frac{2x}{2n}\right] - 2\left[\frac{x}{2n}\right]$  is 0 or 1. So the left term in the right-hand side is nonnegative. We drop it, to obtain

$$T(x) - 2T\left(\frac{x}{2}\right) \geq \sum_{x/2 < n \leq x} \left[\frac{x}{n}\right] a(n) = \sum_{x/2 < n \leq x} a(n) = S(x) - S\left(\frac{x}{2}\right)$$

which is the desired inequality. But the hypothesis of the theorem says

$$T(x) - 2T\left(\frac{x}{2}\right) = x \log x + O(x) - 2\left(\frac{x}{2} \log \frac{x}{2} - O(x)\right) = O(x)$$

So inequality says  $S(x) - S\left(\frac{x}{2}\right) = O(x)$ . So, by definition, there is some  $K > 0$  such that

$$S(x) - S\left(\frac{x}{2}\right) \leq Kx \text{ for all } x \geq 1$$

Replace  $x$  successively above by  $\frac{x}{2}, \frac{x}{4}, \dots$ , to get

$$\begin{aligned} S\left(\frac{x}{2}\right) - S\left(\frac{x}{4}\right) &\leq K\frac{x}{2} \\ S\left(\frac{x}{4}\right) - S\left(\frac{x}{8}\right) &\leq K\frac{x}{4} \end{aligned}$$

etc. Add up all such inequalities to get  $S(x) \leq Kx(1 + 1/2 + 1/4 + \dots) = 2Kx$ . This proves part (1) with  $B = 2K$ .

Now notice that  $\left[\frac{x}{n}\right] = \frac{x}{n} + O(1)$ . We obtain

$$T(x) = \sum_{n \leq x} \left[\frac{x}{n}\right] a(n) = \sum_{n \leq x} \left(\frac{x}{n} + O(1)\right) a(n) = x \sum_{n \leq x} \frac{a(n)}{n} + O\left(\sum_{n \leq x} a(n)\right)$$

But the right term is  $O(x)$ , by part (1). This finally gives

$$\sum_{n \leq x} \frac{a(n)}{n} = \frac{1}{x} T(x) + O(1) = \log x + O(1)$$

Which follows as  $T(x) = x \log x + O(1)$  by hypothesis. □

The hard work is done. We now immediately obtain

**Corollary 2.18 (Our Asymptotic Relation).**

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$$

*Proof.* Let

$$a(n) = \begin{cases} \log p & \text{if } n = p \text{ (p is prime)} \\ 0 & \text{otherwise} \end{cases}$$

Theorem (2.16) assures us that the hypothesis of Theorem (2.17) are satisfied.  $\square$

### 3 Dirichlet characters

Now we focus our attention on something of a different nature. In this section, we develop the properties of the arithmetical functions that will allow us to extract our primes. We briefly adopt a more algebraic spirit as we investigate characters.

#### 3.1 A Few Preliminary Results

We will need a few elementary results from group theory. We will be mostly concerned with finite *abelian* groups, but the following definition makes sense in general.

**Definition 3.1.** Let  $H$  and  $G$  be finite groups with  $H \subseteq G$ , and  $g \in G$ . The smallest  $n \in \mathbb{Z}$  such that  $g^n \in H$  is called the *indicator* of  $g$  in  $H$ .

The indicator is well-defined, as it follows from Lagrange's theorem that if  $\#G = n$  then  $g^n = id_G = id_H \in H$ .

The following theorem furnishes a method to construct a subgroup  $G''$  such that if  $G' \triangleleft G$ , then  $G' \triangleleft G'' \triangleleft G$ .

**Theorem 3.2.** Let  $G' \triangleleft G$  be finite abelian groups,  $a \in G \setminus G'$ , and  $h$  is the indicator of  $a$  in  $G'$ . Then, if we define

$$G'' = \{xa^k : x \in G' \text{ and } k = 0, 1, \dots, h-1\}$$

we have  $G' \triangleleft G'' \triangleleft G$ . Furthermore,  $\#G'' = h\#G'$ .

*Proof.* We first check that  $G''$  is in fact a subgroup of  $G$ . Suppose  $x_1a^l, x_2a^k \in G''$ . Then the product  $x_1x_2a^{l+k} \in G''$ , where we have used the fact that  $G$  is abelian. As for inverses, say  $xa^k \in G''$ . Then  $(xa^k)^{-1} = x^{-1}a^{-k} = (x^{-1}a^{-h})(a^{h-k}) \in G''$ , where we again used the commutativity of  $G$ . Thus,  $G'' \triangleleft G$ , as necessary.

As for counting  $\#G''$ , we note that for  $k < h$ , and all  $x \in G'$ ,  $xa^k \notin G'$ , for otherwise  $xa^k = g' \in G'$  so  $a^k = x^{-1}g' \in G'$ , as  $G'$  is a subgroup, which contradicts the definition of the indicator  $h$ .

Moreover, if  $0 \leq l < k < h$ , we cannot have  $x_1a^l = x_2a^k$  for  $x_1, x_2 \in G'$ , as this implies  $x_2^{-1}x_1 = a^{k-l} \in G'$ , again contradicting the minimality of  $h$ . Hence,  $\#G'' = h\#G'$ , as claimed.  $\square$

There is a special kind of homomorphism from a group  $G$  into  $\mathbb{C}$  that will interest us. We single them out, as they we will later specialize them further to define Dirichlet characters.

**Definition 3.3.** Let  $G$  be a group. A homomorphism  $f : G \rightarrow \mathbb{C}^\times$  is a *character* of  $G$  if  $f(c) \neq 0$  for some  $c \in G$ . The character  $f$  such that  $f(a) = 1 \forall a \in G$  is called the *principal character*.

**Theorem 3.4.** *Say  $f$  is a character of a finite abelian group  $G$ ,  $\#G = n$ . Then,*

- 1)  $f(e) = 1$  where  $e := id_G$ .
- 2)  $f(a)^n = 1$ , for all  $a \in G$ , so all the values of the character are  $n$ th roots of unity.
- 3)  $G$  has exactly  $n$  distinct characters.

*Proof.* Property (1) is the well-known result that a homomorphism fixes the identity element. Property (2) follows easily, since we have

$$f(a)^n = f(a^n) = f(e) = 1$$

which follows from property (1) and the fact that  $f$  is a multiplicative homomorphism.

We prove property (3) by induction. To this effect, we will use the result of Theorem 3.2. We will build an ascending chain of subgroups of  $G$ . We will denote by  $\langle G', a \rangle$  the subgroup  $G''$  in Theorem 3.2

$$\langle e \rangle = G_1 \subset G_2 \subset \dots \subset G_n = G$$

where  $\langle e \rangle$  is the trivial subgroup, and  $G_{n+1} = \langle G_n, a \rangle$  for some  $a \in G \setminus G_n$ .

For  $G_1 = \langle e \rangle$ , there is clearly only one character. Suppose  $\#G_r = m$  and that there are exactly  $m$  indicators for  $G_r$ . Consider  $G_{r+1} = \langle G_r, a_r \rangle$ , where  $h$  is the indicator of  $a_r$  in  $G_r$ . Will show that there are exactly  $h$  ways to extend a character of  $G_r$  to one in  $G_{r+1}$  and that each character of  $G_{r+1}$  is the extension of some character in  $G_r$ . This will work, since we have  $hm$  elements in  $G_{r+1}$ .

Let  $xa_r^k \in G_{r+1}$ ;  $0 \leq k < h$ . Say  $f$  is a character of  $G_r$ , and say it's possible to extend it to a character  $\tilde{f}$  of  $G_{r+1}$ . We have

$$\tilde{f}(xa_r^k) = \tilde{f}(x)\tilde{f}(a_r^k) = f(x)\tilde{f}(a_r)^k \text{ since } x \in G_r$$

Hence  $\tilde{f}$  is determined by its value at  $a_r$ . Let  $c = a_r^h$ .  $\tilde{f}(c) = f(c)$  and  $\tilde{f}(c) = \tilde{f}(a_r)^h = f(c)$ , id est,  $\tilde{f}(a_r)$  is one of the  $h$ th roots of  $f(c) \in \mathbb{C}$ . There are exactly  $h$  choices for  $\tilde{f}(a_r)$ , which gives us all the  $h$  characters as claimed. Thus, we can now define  $\tilde{f}(xa_r^k) := f(x)\tilde{f}(a_r)^k$ . Hence, every one of the  $m$  characters of  $G_r$  can be extended to one of  $G_{r+1}$  in one of  $h$  ways.

Moreover, if  $\varphi$  is a character of  $G_{r+1}$ , then its restriction to  $G_r$  is a character, so the above method of extension gives us all the characters.  $\square$

We can now define the following operation. By the above theorem, this will give us a group dual to any finite abelian group.

**Definition 3.5.** Let  $f_1 f_2$  be two characters on a finite abelian group  $G$ . Define pointwise multiplication between any two characters. That is,  $f_1 \cdot f_2$  is the homomorphism defined by  $(f_1 \cdot f_2)(a) = f_1(a)f_2(a) \forall a \in G$ .

**Theorem 3.6.** *The  $n$  characters on a finite abelian group  $G$  of order  $n$  form themselves an abelian group. We denote this group by  $\hat{G}$ .*

**Remark.**  $|f(a)| = 1$  for all characters in  $\hat{G}$ , for all  $a \in G$ . Hence,  $f^{-1} = \overline{f(a)}$ , where the bar denotes complex conjugation. Also,  $\overline{f(a)} = f(a^{-1})$ .

### 3.2 Orthogonality Relation for Characters

Let  $A = (a_{ij})$  be the matrix whose entry in the  $i$ th row  $j$ th column is  $a_{ij} = f_i(a_j)$ , where  $f_i$  is the  $i$ th character on  $G$  and  $a_j \in G$  the  $j$ th element.  $f_1$  is the principal character.

**Theorem 3.7.** *The row sums of  $A$  are all zero, except for the first row that sums up to  $n$ .*

*Proof.* The first row of  $A$  contains nothing but ones, as it is the row of the principal character. There are  $n$  entries in every row, so the first row adds up to  $n$ .

Now, in any other row, there is an entry  $f_i(b) \neq 1$ . As  $r$  runs over  $1, \dots, n$ ,  $a_r$  runs over all elements of  $G$  and so does  $ba_r$ . Hence,

$$S := \sum_{r=1}^n f_i(a_r) = \sum_{r=1}^n f_i(ba_r) = f_i(b) \sum_{r=1}^n f(a_r) = f_i(b)S$$

So  $S(1 - f_i(b)) = 0$ , but  $f_i(b) \neq 1$ , so  $S = 0$ . □

**Theorem 3.8.**  *$AA^* = nI$ , where  $A^*$  is the adjoint of  $A$  and  $I$  is the identity matrix.*

*Proof.* Let  $(b_{ij}) = AA^*$ . We thus have

$$b_{ij} = \sum_{r=1}^n f_i(a_r) \overline{f_j(a_r)} = \sum_{r=1}^n (f_i \cdot \overline{f_j})(a_r) = \sum_{r=1}^n f_k(a_r)$$

But  $f_k = f_1$  iff  $i = j$ . So by theorem 3.7,  $B = nI$ . □

The fact that a matrix commutes with its inverse and that  $A^{-1} = (1/n)A^*$  immediately gives us the following orthogonality relation.

**Theorem 3.9.**

$$\sum_{r=1}^n \overline{f_r(a_i)} f_r(a_j) = \begin{cases} n & \text{if } a_i = a_j \\ 0 & \text{otherwise} \end{cases}$$

*Proof.*  $AA^* = nI$ , so  $A^*A = nI$ . The entries of  $A^*A$  are the above. □

Thus,  $\sum_{r=1}^n (f_r)(a_j) = n$  if  $a_j = e$  and 0 otherwise.

### 3.3 Dirichlet Characters

**Definition 3.10.** Let  $U_k = (\mathbb{Z}/k\mathbb{Z})^\times$  be the multiplicative group of units modulo  $k$ . For each character  $f$  of  $U_k$  and coset  $n + k\mathbb{Z} \in U_k$  the arithmetical function

$$\chi(n) = \begin{cases} f(n + k\mathbb{Z}) & \text{if } n + k\mathbb{Z} \in U_k \\ 0 & \text{otherwise} \end{cases}$$

$\chi$  is a *Dirichlet character* modulo  $k$ . The principal character is  $\chi_1$  for which  $\chi_1(n) = 1$  if  $\gcd(n, k) = 1$  and 0 otherwise.

By the above considerations on general characters, we have that there are  $\#U_k = \varphi(k)$  Dirichlet characters modulo  $k$ , each of which is a completely multiplicative function, and they have period  $k$ , i.e.  $\chi(n+k) = \chi(n)$  for all  $n$ .

The Dirichlet characters, in particular, also satisfy the following orthogonality condition.

**Theorem 3.11.** *Let  $\chi_1, \chi_2, \dots, \chi_{\varphi(k)}$  be the  $\varphi(k)$  Dirichlet characters of  $U_k$ . Suppose  $\gcd(n, k) = 1$ . Then,*

$$\sum_{r=1}^{\varphi(k)} \chi_r(m) \overline{\chi_r}(n) = \begin{cases} \varphi(k) & \text{if } m \equiv n \pmod{k} \\ 0 & \text{if } m \not\equiv n \pmod{k} \end{cases}$$

### 3.4 Non-vanishing of $L(s, \chi)$ for Real Nonprincipal $\chi$

Dirichlet achieved the proof of his theorem by studying certain families of series. Besides the Riemann zeta function, his proof makes use of so-called  $L$ -functions, defined as

$$L(1, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

where  $\chi$  is a Dirichlet character and  $s$  is complex-valued. As mentioned before, a key step in the proof of Dirichlet's theorem involves showing that  $L(1, \chi)$  is nonzero for a nonprincipal character  $\chi$ . We first establish this fact in the case when  $\chi$  is a principal character that only takes on the values  $0, \pm 1$ .

**Lemma 3.12.** Let  $\chi$  be any real-valued character modulo  $k$  and let

$$A(n) = \sum_{d|n} \chi(d)$$

Then  $A(n) \geq 0$  for all  $n$ , and  $A(n) \geq 1$  if  $n$  is a square.

*Proof.* Since  $A(n)$  is multiplicative (it is the Dirichlet convolution of two multiplicative functions), we only need to determine  $A(p^a)$  for prime powers  $p^a$ . We have

$$A(p^a) = \sum_{t=0}^a \chi(p^t) = 1 + \sum_{t=1}^a (\chi(p))^t$$

where  $\chi$  can only take the values  $0, \pm 1$ . If  $\chi(p) = 0$ , then  $A(p^a) = 1$ ; if  $\chi(p) = 1$ , then  $A(p^a) = 1 + a$ , and if  $\chi(p) = -1$  then

$$A(p^a) = \begin{cases} 1 & \text{if } a \text{ is even} \\ 0 & \text{if } a \text{ is odd} \end{cases}$$

either way,  $A(p^a) \geq 1$  if  $a$  is even. Now let  $n = p_1^{a_1} \cdot p_2^{a_2} \cdot \dots \cdot p_r^{a_r}$ , so  $A(n) = A(p_1^{a_1}) \cdot A(p_2^{a_2}) \cdot \dots \cdot A(p_r^{a_r})$ . Every term is positive, and if  $n$  is a square every  $a_j$  is even, so every term is greater than or equal to 1.  $\square$

**Lemma 3.13.** Let  $f$ ,  $g$ , and  $h := f * g$  be arithmetical functions and  $F(x)$ ,  $G(x)$ , and  $H(x)$  respectively be their partial sums up to and including  $x$ . If  $a$  and  $b$  are positive numbers such that  $ab = x$ , then

$$H(x) = \sum_{\substack{c,d \\ cd \leq x}} f(c)g(d) = \sum_{n \leq x} f(n)G\left(\frac{x}{n}\right) + \sum_{n \leq x} g(n)F\left(\frac{x}{n}\right) - F(a)G(b)$$

*Proof.* The first equality follows by definition, since

$$H(x) = \sum_{n \leq x} (f * g)(n) = \sum_{n \leq x} \sum_{cd=n} f(c)g(d) = \sum_{\substack{c,d \\ cd \leq x}} f(c)g(d)$$

as for the second equality, notice that the summation is over the lattice points in the parabolic region bounded by  $cd = x$  and  $c, d \geq 1$ . The point  $(a, b)$  is on the hyperbola  $cd = x$ . Consider the region in the hyperbolic region of summation bounded by  $c \leq a$  and the region bounded by  $d \leq b$ .

If we add up the contributions from those two regions, and subtract the contribution from the portion of the hyperbolic region bounded by  $c \leq a$  and  $d \leq b$  that was counted twice, we obtain the identity

$$\begin{aligned} H(x) &= \sum_{\substack{c,d \\ cd \leq x}} f(c)g(d) = \sum_{c \leq a} \sum_{d \leq x/c} f(c)g(d) + \sum_{d \leq b} \sum_{c \leq x/d} f(c)g(d) - \sum_{d \leq a, c \leq b} f(c)g(d) \\ &= \sum_{n \leq x} f(n)G\left(\frac{x}{n}\right) + \sum_{n \leq x} g(n)F\left(\frac{x}{n}\right) - F(a)G(b) \end{aligned}$$

□

**Theorem 3.14.** For any real-valued nonprincipal character  $\chi$ , put

$$A(n) = \sum_{d|n} \chi(d) \text{ and } B(n) = \sum_{n \leq x} \frac{A(n)}{\sqrt{n}}$$

Then we have: (a)  $B(x) \rightarrow \infty$  as  $x \rightarrow \infty$ . (b)  $B(x) = 2\sqrt{x}L(1, \chi) + O(1)$ .

So,  $L(1, \chi) \neq 0$

*Proof.* Part (a) follows easily from Lemma (3.12). Drop all the terms from  $B(x)$  (which we know to be positive) except for the squares (which we know to be  $\geq 1$  so

$$B(x) \geq \sum_{\substack{n \leq x \\ n=m^2}} \frac{1}{\sqrt{n}} = \sum_{m \leq \sqrt{x}} \frac{1}{m}$$

i.e.  $B(x)$  is bounded below by a divergent harmonic series.

For part (b), write

$$B(x) = \sum_{n \leq x} \frac{1}{\sqrt{n}} \sum_{d|n} \chi(d) = \sum_{\substack{q,d \\ qd \leq x}} \frac{\chi(d)}{\sqrt{dq}}$$

Use lemma (3.13) with  $a = b = \sqrt{x}$ ,  $f(n) = \chi(n)/\sqrt{n}$ , and  $g(n) = 1/\sqrt{n}$  to obtain

$$B(x) = \sum_{n \leq \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} G\left(\frac{x}{n}\right) + \sum_{n \leq \sqrt{x}} \frac{1}{\sqrt{n}} F\left(\frac{x}{n}\right) - F(\sqrt{x})G(\sqrt{x}) \quad (4)$$

where

$$F(x) = \sum_{n \leq x} \frac{\chi(n)}{\sqrt{n}} \text{ and } G(x) = \sum_{n \leq x} \frac{1}{\sqrt{n}}$$

By Lemma A.2 in the Appendix,

$$G(x) = \sum_{n \leq x} \frac{1}{\sqrt{n}} = 2\sqrt{x} + K_1 + O(x^{-1/2}),$$

where  $K_1 = \zeta(1/2)$  is a constant. So  $G(x) = 2\sqrt{x} + O(1)$ . Also, by the third relation in Lemma A.4,

$$F(x) = \sum_{n \leq x} \frac{\chi(n)}{\sqrt{n}} = K_2 + O\left(\frac{1}{\sqrt{x}}\right),$$

where  $K_2 = \sum_{n=1}^{\infty} \frac{\chi(n)}{\sqrt{n}}$ . So  $F(x) = O(1)$ , since it converges.

This gives  $F(\sqrt{x})G(\sqrt{x}) = 2K_2x^{1/4} + O(1)$ . From equation (4), we have

$$\begin{aligned} B(x) &= \sum_{n \leq \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} \left[ 2\sqrt{\frac{x}{n}} + K_1 + O\left(\sqrt{\frac{n}{x}}\right) \right] \\ &\quad + \sum_{n \leq \sqrt{x}} \frac{1}{\sqrt{n}} \left[ K_2 + O\left(\sqrt{\frac{n}{x}}\right) \right] - 2K_2 + O(1) \\ &= 2\sqrt{x} \sum_{n \leq \sqrt{x}} \frac{\chi(n)}{n} + K_1 \sum_{n \leq \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}} \sum_{n \leq \sqrt{x}} |\chi(n)|\right) \\ &\quad + K_2 \sum_{n \leq \sqrt{x}} \frac{1}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}} \sum_{n \leq \sqrt{x}} 1\right) - 2K_2x^{1/4} + O(1) \end{aligned}$$

This is not really as horrendous as it looks. All of the  $O$ -terms are  $O(1)$ ; the constant  $K_1$  gets absorbed into  $O(1)$  as well, and the combination of the  $1/\sqrt{(n)}$  minus  $2x^{1/4}$  converges, as it is bounded above by the corresponding integral minus  $x^{1/4}$  which is constantly equal to  $-2$ . Hence, everything condenses into

$$B(x) = 2\sqrt{x}L(1, \chi) + O(1)$$

But since  $B(x)$  diverges by part (a),  $L(1, \chi) \neq 0$ . □

## 4 The Extraction

We are on the final stretch now. We will put together the estimate from Section 2 and the Dirichlet characters from Section 3 to form a new estimate from which to estimate from which Dirichlet's theorem will be a simple corollary.

**Notation.** In this section,  $k$  is a fixed modulus and  $h$  is coprime to  $k$ .  $\chi_r$  will denote a nonprincipal Dirichlet character modulo  $k$  and we reserve, as usual, the symbol  $\chi_1$  for the principal character.

The following series for a nonprincipal character  $\chi$  are of concern to us in this section. Their convergence is established by the first and second relations in Lemma A.4.

$$\begin{aligned} L(1, \chi) &= \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \\ L'(1, \chi) &= - \sum_{n=1}^{\infty} \frac{\chi(n) \log n}{n} \end{aligned}$$

As mentioned in the introduction, we seek

$$\sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + O(1) \quad (5)$$

The first lemma in that direction is

**Lemma 4.1.** For  $x > 1$

$$\sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{r=2}^{\varphi(k)} \bar{\chi}_r(h) \sum_{p \leq x} \frac{\chi_r(p) \log p}{p} + O(1)$$

*Proof.* Starting from the main result of Section 2, Corollary 2.18,

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$$

we will use the orthogonality condition of Dirichlet characters from Theorem 3.11

$$\sum_{r=1}^{\varphi(k)} \chi_r(m) \bar{\chi}_r(n) = \begin{cases} \varphi(k) & \text{if } m \equiv n \pmod{k} \\ 0 & \text{if } m \not\equiv n \pmod{k} \end{cases}$$

which is valid for  $\gcd(n, k) = 1$ . Let  $m := p$  and  $n := h$  above where  $\gcd(h, k) = 1$ . Multiply both sides by  $(\log p)/p$  and sum over  $p \leq x$  to get

$$\sum_{p \leq x} \sum_{r=1}^{\varphi(k)} \chi_r(p) \bar{\chi}_r(h) \frac{\log p}{p} = \varphi(k) \sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p}$$

The congruence condition on the sum on the right-hand side comes from the orthogonality relations. Now, take out the terms from left-hand side that correspond to  $\chi_1$  and rewrite

$$\varphi(k) \sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p} = \bar{\chi}_1(h) \sum_{p \leq x} \chi_1(p) \frac{\log p}{p} + \sum_{r=2}^{\varphi(k)} \bar{\chi}_r(h) \sum_{p \leq x} \chi_r(p) \frac{\log p}{p} \quad (6)$$

in the left term of the right-hand side,  $\chi_1(h)$  is the principal character, so it is 1, as  $h$  is coprime to the modulus  $k$ . The  $\chi_1(p)$  factor is 0 if  $p|k$  and 1 otherwise. Hence, this whole term becomes

$$\sum_{\substack{p \leq x \\ \gcd(p,k)=1}} \frac{\log p}{p} = \sum_{p \leq x} \frac{\log p}{p} - \sum_{\substack{p \leq x \\ p|k}} \frac{\log p}{p} = \sum_{p \leq x} \frac{\log p}{p} + O(1) \quad (7)$$

since only finitely many primes divide  $k$ . Putting equation (7) back into equation (6) we see

$$\varphi(k) \sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} \frac{\log p}{p} = \sum_{p \leq x} \frac{\log p}{p} + \sum_{r=2}^{\varphi(k)} \bar{\chi}_r(h) \sum_{p \leq x} \chi_r(p) \frac{\log p}{p} + O(1)$$

Now we conclude by applying Corollary 2.18 and dividing by  $\varphi(k)$ .  $\square$

Lemma 4.1 would immediately imply equation (5) if we could only show that the sum

$$\sum_{p \leq x} \frac{\chi(p) \log p}{p} \quad (8)$$

is  $O(1)$ . Our next step is to express this sum in a form that is not extended over all primes.

**Lemma 4.2.**

$$\sum_{p \leq x} \frac{\chi(p) \log p}{p} = -L'(1, \chi) \sum_{n \leq x} \frac{\mu(n) \chi(n)}{n} + O(1)$$

*Proof.* Commence with the sum

$$\sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n}$$

and express this sum in two ways. First, the definition of the von Mangoldt function gives us

$$\sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n} = \sum_{p \leq x} \frac{\chi(p) \log p}{p} + \sum_{p \leq x} \sum_{\substack{a=2 \\ p^a \leq x}}^{\infty} \frac{\chi(p^a) \log p}{p^a}$$

because  $|\chi(p^a)| \leq 1$ , the right sum on the right-hand side is majorized by

$$\sum_p \log p \sum_{a=2}^{\infty} \frac{1}{p^a} = \sum_p \frac{\log p}{p(p-1)} < \sum_{n=2}^{\infty} \frac{\log n}{n(n-1)} = O(1)$$

again, because the last sum converges. Thus we have

$$\sum_{p \leq x} \frac{\chi(p) \log p}{p} = \sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n} + O(1) \quad (9)$$

On the other hand, Möbius inversion gives us that  $\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d)$ , so

$$\sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n} = \sum_{n \leq x} \frac{\chi(n)}{n} \sum_{d|n} \mu(d) \log \left( \frac{n}{d} \right)$$

write  $n = cd$  and use the multiplicative property of  $\chi$  to arrive at

$$\begin{aligned} \sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n} &= \sum_{d \leq x} \frac{\mu(d) \chi(d)}{d} \sum_{c \leq x/d} \frac{\chi(c) \log(c)}{c} \text{ use 2nd of Lemma A.4} \\ &= \sum_{d \leq x} \frac{\mu(d) \chi(d)}{d} \left( -L'(1, \chi) + O \left( \frac{\log x/d}{x/d} \right) \right) \text{ but } |\mu(d) \chi(d)| \leq 1 \\ &= -L'(1, \chi) \sum_{d \leq x} \frac{\mu(d) \chi(d)}{d} + O \left( \sum_{d \leq x} \frac{1}{d} \frac{\log x/d}{x/d} \right) \end{aligned}$$

The sum in the  $O$ -term is

$$\frac{1}{x} \sum_{d \leq x} (\log x - \log d) = \frac{1}{x} \left( [x] \log x - \sum_{d \leq x} \log d \right)$$

and the bracketed quantity above is, by Corollary 2.14 and Lemma 2.15.

$$[x] \log x - \log[x]! = O(x)$$

so upon dividing by  $x$  we see that the entire  $O$ -term is  $O(1)$ .

Thus,

$$\sum_{n \leq x} \frac{\chi(n) \Lambda(n)}{n} = -L'(1, \chi) \sum_{d \leq x} \frac{\mu(d) \chi(d)}{d} + O(1)$$

Combine this with equation (9) to conclude

$$\sum_{p \leq x} \frac{\chi(p) \log p}{p} = -L'(1, \chi) \sum_{n \leq x} \frac{\mu(n) \chi(n)}{n} + O(1)$$

as necessary.  $\square$

Lemma 4.2 will imply that the sum (8) is  $O(1)$  if we show that the sum

$$\sum_{n \leq x} \frac{\mu(n)\chi(n)}{n}$$

is  $O(1)$ . To this end we work in this direction:

**Lemma 4.3.** For  $x > 1$  and  $\chi \neq \chi_1$ , we have

$$L(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} = O(1)$$

**Remark.** This is the most delicate step of the proof. We would be done and Dirichlet's theorem would be proven if we could just show that  $L(1, \chi)$  is nonzero for nonprincipal  $\chi$ . If we had that, we could divide by  $L(1, \chi)$  and the proof goes through.

*Proof.* Use the generalized Möbius inversion formula. Put  $\alpha(n) = \chi(n)$  and  $F(x) = x$  in Theorem 2.11 to get

$$x = \sum_{n \leq x} \mu(n)\chi(n)G\left(\frac{x}{n}\right) \quad (10)$$

where

$$G(x) := \sum_{n \leq x} \chi(n) \frac{x}{n} = x \sum_{n \leq x} \frac{\chi(n)}{n}$$

By the first relation in Lemma A.4, write  $G(x) = xL(1, \chi) + O(1)$ . Substitute this back into equation (10). We discover

$$x = \sum_{n \leq x} \mu(n)\chi(n) \left( \frac{x}{n} L(1, \chi) + O(1) \right) = xL(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} + O(x)$$

Now divide by  $x$  to arrive at the result.  $\square$

Our upcoming move is rather ingenious. We need to show that the  $L$ -series of interest does not vanish. We have already established this fact in Section 3 for nonprincipal real characters. So, assume  $\chi$  is nonprincipal, and complex-valued. Say  $L(1, \chi) = 0$  for some  $\chi$ . We could then also have that  $L(1, \bar{\chi}) = 0$ , yet  $\chi \neq \bar{\chi}$ . If we let  $N(k)$  denote the number of characters for which the  $L$ -series vanishes, this argument shows that  $N(k)$  is *even*, because the nasty characters (if any!) occur in conjugate pairs. Our goal is to show that  $N(k) = 0$ , of course.

Before that, an intermediate step on a necessary condition for the vanishing of  $L(1, \chi)$ .

**Lemma 4.4.** If  $\chi \neq \chi_1$  and  $L(1, \chi) = 0$ , we have

$$L'(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} = \log x + O(1)$$

*Proof.* This time put  $F(x) = x \log x$  and again  $\alpha(n)\chi(n) =$  in the generalized Möbius inversion formula:

$$x \log x = \sum_{n \leq x} \mu(n)\chi(n)G\left(\frac{x}{n}\right) \quad (11)$$

where

$$G(x) := \sum_{n \leq x} \chi(n) \frac{x}{n} \log \frac{x}{n} = x \log x \sum_{n \leq x} \frac{\chi(n)}{n} - x \sum_{n \leq x} \frac{\chi(n) \log(n)}{n}$$

Next use Lemma A.4 part 1 to get

$$\begin{aligned} G(x) &= x \log x \left( L(1, \chi) + O\left(\frac{1}{x}\right) \right) + x \left( L'(1, \chi) + O\left(\frac{\log x}{x}\right) \right) \\ &= xL'(1, \chi) + O(\log x) \text{ since } L(1, \chi) = 0 \text{ by assumption.} \end{aligned}$$

Put this back into equation (11). Thus,

$$\begin{aligned} x \log x &= \sum_{n \leq x} \mu(n)\chi(n) \left( \frac{x}{n} L'(1, \chi) + O\left(\log \frac{x}{n}\right) \right) \\ &= xL'(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} + O\left( \sum_{n \leq x} (\log x - \log n) \right) \end{aligned}$$

In the course of the proof of Lemma 4.2 we already showed that the  $O$ -term is  $O(x)$ . Therefore, upon dividing by  $x$ , we get

$$\log x = L'(1, \chi) \sum_{n \leq x} \frac{\mu(n)\chi(n)}{n} + O(1)$$

which is what we sought.  $\square$

... and finally,

**Lemma 4.5.** Let  $N(k)$  denote the number of nonprincipal characters  $\chi$  modulo  $k$  such that  $L(1, \chi) = 0$ . If  $x > 1$ ,

$$\sum_{\substack{p \leq x \\ p \equiv 1 \pmod{k}}} \frac{\log p}{p} = \frac{1 - N(k)}{\varphi(k)} \log x + O(1)$$

**Remark.** This result concludes the proof of Dirichlet's theorem, for if  $N(k)$  is not zero, we have a contradiction, since then the estimate on the right-hand side tends to  $-\infty$ , while the left-hand side contains only positive terms.

*Proof.* Take  $h = 1$  in Lemma 4.1. Obtain

$$\sum_{\substack{p \leq x \\ p \equiv 1 \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{r=2}^{\varphi(k)} \sum_{p \leq x} \frac{\chi_r(p) \log p}{p} + O(1) \quad (12)$$

Apply Lemma 4.2 to the rightmost sum, yielding

$$\sum_{\substack{p \leq x \\ p \equiv 1 \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{r=2}^{\varphi(k)} \left( -L'(1, \chi_r) \sum_{n \leq x} \frac{\mu(n) \chi_r(n)}{n} + O(1) \right)$$

and the sum  $\sum_{n \leq x} (\mu(n) \chi(n))/n$  is  $O(1)$  if  $L(1, \chi_r) \neq 0$ .

However, for all  $r$  such that  $L(1, \chi_r) = 0$ , Lemma 4.4 implies

$$L'(1, \chi_r) \sum_{n \leq x} \frac{\mu(n) \chi_r(n)}{n} = \log x + O(1)$$

So the sum in equation (12) reads

$$\frac{1}{\varphi(k)} \log x + \frac{1}{\varphi(k)} \sum_{\forall r (L(1, \chi_r) = 0)} -\log x + O(1) = \frac{1 - N(k)}{\varphi(k)} \log(x) + O(1)$$

which concludes our proof.  $\square$

Hence, every arithmetic series generated by coprime elements contains infinitely many primes. *Quod Erat Demonstrandum.*

## 5 Acknowledgments

I wish to thank the Undernet IRC Network's #math channel for their help and support in writing this paper, in particular Kimberly Schneider's contributions, suggestions, and encouragement.

## A Appendix

Here we collect miscellaneous results used in the body of this paper.

First, a definition of a very famous function, which we tried to avoid as much as possible in this paper, in an interest to keep the proof of Dirichlet's theorem elementary.

**Definition A.1.** The *Riemann zeta function*  $\zeta : \mathbb{C} \rightarrow \mathbb{C}$  is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

where  $s = \sigma + it$  ( $\sigma, t \in \mathbb{R}$ ) and  $\sigma > 1$  or by

$$\zeta(s) = \lim_{x \rightarrow \infty} \left( \sum_{n \leq x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s} \right)$$

if  $0 < \sigma < 1$ .

We made use of the following asymptotic result when we discussed sums of Dirichlet characters.

**Lemma A.2.**

$$\sum_{n \leq x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + \zeta(s) + O(x^{-s}) \text{ if } s \in \mathbb{R}^+, s \neq 1$$

*Proof.* Put  $f(x) = x^{-s}$  and  $y = 1$  in Euler's summation formula, for  $s > 0, s \neq 1$ . This gives

$$\begin{aligned} \sum_{n \leq x} \frac{1}{n^s} &= \int_1^x \frac{dt}{t^s} - s \int_1^x \frac{t - [t]}{t^{s+1}} dt + 1 - \frac{x - [x]}{x^s} \\ &= \frac{x^{1-s}}{1-s} + \left( 1 - \frac{1}{1-s} - s \int_1^{\infty} \frac{t - [t]}{t^{s+1}} dt \right) + O(x^{-s}) \end{aligned}$$

The last term is  $O(x^{-s})$  because  $x - [x]$  is bounded by 1. If  $s > 1$ , the left hand-side of the equation approaches  $\zeta(s)$  as  $x \rightarrow \infty$ , while every other term on the right-hand side goes to zero. Thus, the middle bracketed term is  $\zeta(s)$ .

In case  $0 < s < 1$ , then  $x^{-s} \rightarrow 0$  and the alternate definition of  $\zeta$  for  $0 < s < 1$  gives that the bracketed term is again  $\zeta(s)$ .  $\square$

The following proof is valid for any periodic arithmetical function with bounded partial sums, but we focus on a nonprincipal Dirichlet character  $\chi$ .

**Lemma A.3.** Let  $\chi$  be a nonprincipal character modulo  $k$ , and let  $f$  be a nonnegative function such that  $f'(x)$  is continuous for all  $x \geq x_0$ . Then if  $y > x \geq x_0$ , we have

$$\sum_{x < n \leq y} \chi(n) f(n) = O(f(x))$$

If, in addition,  $f(x) \rightarrow 0$  as  $x \rightarrow \infty$ , then  $\sum_{n=1}^{\infty} \chi(n)f(n)$  converges, and we have, for  $x \geq x_0$ ,

$$\sum_{n \leq x} \chi(n)f(n) = \sum_{n=1}^{\infty} \chi(n)f(n) + O(f(x))$$

*Proof.* Let  $A(x) = \sum_{n \leq x} \chi(n)$ . Since  $\chi$  is nonprincipal, the orthogonality conditions give us that

$$A(k) = \sum_{n=1}^k \chi(n) = 0$$

By periodicity,  $A(nk) = 0$  for all  $n \in \mathbb{N}$ . Hence,  $|A(x)| < \varphi(k)$ , i.e.  $A(k) = O(1)$ . Next, use Abel's identity

$$\begin{aligned} \sum_{x < n \leq y} \chi(n)f(n) &= A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t) dt \\ &= O(f(x)) + O(f(y)) + O\left(\int_x^y (-f'(t)) dt\right) = O(f(x)) \end{aligned}$$

This proves the first part.

If  $f(x) \rightarrow 0$  as  $x \rightarrow \infty$ , then the Cauchy criterion gives us convergence of the series. For the last part, put

$$\sum_{n=1}^{\infty} \chi(n)f(n) = \sum_{n \leq x} \chi(n)f(n) + \lim_{y \rightarrow \infty} \sum_{x < n \leq y} \chi(n)f(n)$$

and the last limit is  $O(f(x))$ , by the first part, which concludes the proof.  $\square$

We call upon the next results during several spots in the proof of Dirichlet's theorem. We collect them here as lemmata, upon putting  $f(x)$  equal to  $1/x$ ,  $(\log x)/x$ , and  $1/\sqrt{x}$  in Lemma A.3.

**Lemma A.4.** We have

1.

$$\sum_{n \leq x} \frac{\chi(n)}{n} = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} + O\left(\frac{1}{x}\right)$$

2.

$$\sum_{n \leq x} \frac{\chi(n) \log(n)}{n} = \sum_{n=1}^{\infty} \frac{\chi(n) \log(n)}{n} + O\left(\frac{\log x}{x}\right)$$

3.

$$\sum_{n \leq x} \frac{\chi(n)}{\sqrt{n}} = \sum_{n=1}^{\infty} \frac{\chi(n)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}}\right)$$

## References

- [1] Tom H. Apostol *Introduction to Analytic Number Theory* Springer-Verlag, New York, 1976
- [2] William J. LeVeque *Fundamentals of Number Theory* Dover Publications Inc, New York 1996
- [3] Anthony H. Gioia *The Theory of Numbers* Dover Publications Inc, New York, 2001
- [4] Walter Rudin *Principles of Mathematical Analysis* McGraw-Hill Inc, New York, 1976