ANALYTIC NUMBER THEORY — LECTURE NOTES BASED ON DAVENPORT'S BOOK

ANDREAS STRÖMBERGSSON

These lecture notes follow to a large extent Davenport's book [15], but with things reordered and often expanded. The point of these notes is not in the first place to serve as an outline of the actual lectures which I will give, but instead to (hopefully!) collect all the details for those results which I wish to mention, so that in the lectures I can focus more on the main points and ideas.

Contents

.

1. Primes in Arithmetic Progressions (Ch. 1, 4 in [15])

In this first lecture we will prove Dirichlet's Theorem from 1837-40:

Theorem 1.1. *If a, q are positive integers with $(a, q) = 1$, then there are infinitely many primes in the arithmetic progression*

$$a, a + q, a + 2q, a + 3q, \ldots.$$

The proof which I will give is complete except for 3-4 technical facts, which I will not have time to prove in detail but which I can hopefully convince you are reasonable to believe. I will come back to these facts and prove them in the next few lectures.

The proof which I will give does *not* follow all steps of the proof which Dirichlet originally gave, instead it is shorter and makes use of more complex analysis; the key new step is a trick by de la Vallée Poussin from 1896 which is presented on pp. 32–34 in Davenport's book. However the original proof by Dirichlet is interesting in its own right because of its connection with quadratic forms and class numbers, and I will come back to this in later lectures.

To get started, we introduce the so called *Riemann Zeta Function*:

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \qquad (s \in \mathbb{C},\ \mathrm{Re}\, s > 1). \tag{1}$$

We will follow standard notation in analytic number theory and write $s = \sigma + it$ $(\sigma, t \in \mathbb{R})$. Thus, for instance, $\{s : \sigma > 1\}$ is the set of all s which have real part greater than one.

Lemma 1.2. *The series $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is absolutely convergent for all $s \in \mathbb{C}$ with $\sigma > 1$, and uniformly absolutely convergent in any compact subset of $\{s : \sigma > 1\}$. In particular, by Weierstrass's Theorem*[1]*, $\zeta(s)$ is an analytic function in the set $\{s : \sigma > 1\}$.*

Proof. This follows directly by comparison with the (positive) series $\sum_{n=1}^{\infty} n^{-c}$ for $c > 1$, which is well-known to be convergent. Details: Since any compact subset of $\{s : \sigma > 1\}$ is contained in $\{s : \sigma \geq c\}$ for some $c > 1$, it suffices to prove that $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is uniformly absolutely convergent for all s with $s \geq c$, where c is some fixed number > 1. Let $\varepsilon > 0$ be given. Since the positive series $\sum_{n=1}^{\infty} n^{-c}$ is convergent, there is some $N \in \mathbb{Z}^+$ such that $\sum_{n=N}^{\infty} n^{-c} < \varepsilon$. Now take any $s = \sigma + it \in \mathbb{C}$ with $\sigma \geq c$. Then for each $n \geq N$ we have

$$|n^{-s}| = |e^{-s \log n}| = |e^{-\sigma \log n - it \log n}| = e^{-\sigma \log n} = n^{-\sigma} \leq n^{-c}, \tag{2}$$

[1] Cf. Ahlfors [1, Sec. 5.1.1] or Priestley [56, Ex. 14.7] or Rudin [60, Thm. 10.28]. Weierstrass Theorem states that if $\{f_k\}$ is a sequence of analytic functions in an open set $\Omega \subset \mathbb{C}$ and if f is a function on Ω such that $f_k \to f$ uniformly on any compact subset of Ω, then f is analytic in Ω, and also $f_k^{(n)} \to f^{(n)}$ uniformly on compact subsets of Ω.

and hence

$$\sum_{n=N}^{\infty} |n^{-s}| \le \sum_{n=N}^{\infty} n^{-c} < \varepsilon. \tag{3}$$

Now we have proved that for each $\varepsilon > 0$ there exists some $N \in \mathbb{Z}^+$ such that (3) holds for all $s \in \mathbb{C}$ with $\operatorname{Re} s \ge c$. This is exactly the desired statement. □

The reason why $\zeta(s)$ is important in the study of primes is the following identity, the so called *Euler's identity* or *Euler product:*

Lemma 1.3.

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} \tag{4}$$

for all $s \in \mathbb{C}$ with $\sigma > 1$.

Here in the product in the right hand side, p runs over all primes, and the lemma in particular contains the fact that this infinite product is convergent (when $\sigma > 1$). *I will not give a complete proof of this lemma here,* but will come back to it in the next lecture where I will discuss infinite products in general (see Section 2.2 below). But I will give an outline of the main idea of the proof: By the formula for a geometric sum we have $(1 - p^{-s})^{-1} = 1 + p^{-s} + p^{-2s} + p^{-3s} + \dots$ for each prime p. Hence the infinite product in the right hand side of (4) is:

$$\begin{aligned} &\big(1 + 2^{-s} + 2^{-2s} + 2^{-3s} + \dots\big)\big(1 + 3^{-s} + 3^{-2s} + 3^{-3s} + \dots\big) \\ &\quad \cdot \big(1 + 5^{-s} + 5^{-2s} + 5^{-3s} + \dots\big)\big(1 + 7^{-s} + 7^{-2s} + 7^{-3s} + \dots\big) \\ &\quad\quad \cdot \big(1 + 11^{-s} + 11^{-2s} + 11^{-3s} + \dots\big) \cdots \end{aligned}$$

When expanding this product completely (using the distributive law) it turns out that because of convergence properties we only pick up those products which take "1" from all except a finite number of the factors above; thus we get

$$\begin{aligned} &= 1 + 2^{-s} + 3^{-s} + (2^{-s}3^{-s}) + (2^{-s}5^{-s}) + 2^{-2s} + (2^{-2s}3^{-s}) + 5^{-s} + \dots \\ &= 1 + 2^{-s} + 3^{-s} + (2 \cdot 3)^{-s} + (2 \cdot 5)^{-s} + (2^2)^{-s} + (2^2 \cdot 3)^{-s} + 5^{-s} + \dots \end{aligned}$$

The terms above are not chosen in any systematic order, but we see that we get exactly one term for each distinct prime factorization $p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$, the corresponding term being $(p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m})^{-s}$. But by the fundamental theorem of arithmetic, *each positive integer has one unique prime factorization* (we here include the number 1 which has the "empty" prime factorization); hence the above sum actually contains exactly one term n^{-s} for each positive integer n, i.e. the sum equals $\sum_{n=1}^{\infty} n^{-s}$, which was to be proved.

Note that none of the factors in the right hand side of (4) vanishes, since $|p^{-s}| = p^{-\sigma} < 1$ when $\sigma > 1$. Hence it seems reasonable that we have:

$$\zeta(s) \neq 0 \qquad \text{for all } s \in \mathbb{C} \text{ with } \sigma > 1. \tag{5}$$

(We will prove this rigorously in the next lecture. See Theorem 2.2.) It follows that $\log \zeta(s)$ can be defined for each $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$. Note that this is in principle a multivalued function for each s. However, there is one choice of $\log \zeta(s)$ which is the most natural: Motivated by the Euler product (4), let us *define* $\log \zeta(s)$ by

$$\log \zeta(s) = -\sum_p \log(1 - p^{-s}), \qquad (s \in \mathbb{C},\ \sigma > 1), \tag{6}$$

where each logarithm on the right is taken with the principal branch (this is ok since $|p^{-s}| < 1$). *We will see in the next lecture,* in connection with our discussion of infinite products (see Example 2.2 below) that the sum in (6) is absolutely convergent for all s with $\sigma > 1$, and indeed gives a logarithm of $\zeta(s)$ (i.e., $e^{-\sum_p \log(1-p^{-s})} = \zeta(s)$). We also note that when s is *real*, $s > 1$, then $\zeta(s)$ is real and positive by definition, and in this case $\log \zeta(s)$ is just the usual, real valued, logarithm (since the sum in the right hand side of (6) is real valued in this case).

Using the Taylor expansion $-\log(1 - z) = z + \frac{z^2}{2} + \frac{z^3}{3} + \ldots$, valid for all $z \in \mathbb{C}$ with $|z| < 1$, we can write (6) in the form

$$\log \zeta(s) = \sum_p \sum_{m=1}^{\infty} m^{-1} p^{-ms}. \tag{7}$$

The double sum in the right hand side of (7) is absolutely convergent for any $s \in \mathbb{C}$ with $\sigma > 1$. [Proof: The convergence of the right hand side in (6) and Taylor's formula combine to show that the iterated sum $\sum_p \big(\sum_{m=1}^{\infty} m^{-1} p^{-ms}\big)$ converges for any $s \in \mathbb{C}$ with $\sigma > 1$. Applying this for *real* $s > 1$ we have a double sum with positive terms, and hence the convergence automatically implies absolute convergence of the double sum. For general $s \in \mathbb{C}$ with $\sigma > 1$ the absolute convergence now follows by using $|p^{-ms}| = p^{-m\sigma}$.]

Now we restrict to considering *real* $s > 1$. Note that directly from the definition of $\zeta(s)$ we get

$$\lim_{s \to 1^+} \zeta(s) = +\infty, \ \text{ and thus } \ \lim_{s \to 1^+} \log \zeta(s) = +\infty. \tag{8}$$

Also note that, for each $s > 1$,

$$\begin{aligned} \sum_p \sum_{m=2}^{\infty} m^{-1} p^{-ms} < \sum_p \sum_{m=2}^{\infty} p^{-m} = \sum_p \frac{1}{p(p-1)} = \sum_p \Big(\frac{1}{p-1} - \frac{1}{p}\Big) \\ \leq \sum_{n=2}^{\infty} \Big(\frac{1}{n-1} - \frac{1}{n}\Big) = 1. \end{aligned} \tag{9}$$

Hence, using (8) and (7), we conclude that

$$\lim_{s\to 1^+} \big(\sum_p p^{-s}\big) = +\infty. \tag{10}$$

This implies that there are infinitely many prime numbers. It even implies the stronger fact that

$$\sum_p p^{-1} = \infty. \tag{11}$$

[Proof: Given any $A > 0$ there exists some $s > 1$ such that $\sum_p p^{-s} > A$, by (10). Hence there exists a finite set of primes $p_1 < p_2 < \ldots < p_N$ such that $\sum_{k=1}^N p_k^{-s} > A$. Now $\sum_{k=1}^N p_k^{-1} > \sum_{k=1}^N p_k^{-s} > A$. Since this can be obtained for any $A > 0$ we conclude that (11) holds.]

Dirichlet's aim in his 1837 memoir was to prove the corresponding fact for the set of prime numbers in an arithmetic sequence, i.e. to prove:

Theorem 1.4. *If $a \in \mathbb{Z}$, $q \in \mathbb{Z}^+$ and $(a, q) = 1$, then*

$$\sum_{p\equiv a \,(mod\ q)} p^{-1} = \infty.$$

Note that Theorem 1.4 in particular implies Theorem 1.1, i.e. the fact that there are infinitely many primes in the arithmetic sequence $a, a + q, a + 2q, a + 3q, \ldots$.

The main tool in the proof of Theorem 1.4 is a generalization of the Riemann Zeta function and Euler's identity to *Dirichlet L-functions*, which are sums involving *Dirichlet characters*.

To motivate their introduction, let us note that a naive way to try to generalize the above proof of $\sum_p p^{-1} = \infty$ to the case of Theorem 1.4 would be to replace $\zeta(s)$ with the series

$$\sum_{\substack{n=1 \\ n\equiv a \,(\mathrm{mod}\ q)}} n^{-s}. \tag{12}$$

However this immediately runs into the problem that *the Euler product* (4) *does not generalize to this series!* Studying the proof sketch of (4) we see that in order to have something like Euler's identity, we should generalize $\zeta(s) = \sum_{n=1}^\infty n^{-s}$ to sums of the form $\sum_{n=1}^\infty c_n n^{-s}$, where the c_n's are *multiplicative*. For our purpose of proving Theorem 1.4 we then wish to go (using linear combinations) from multiplicative coefficients to the case of (12), viz. "$c_n = 1$ if $n \equiv a \,(\mathrm{mod}\ q)$, $c_n = 0$ otherwise", and for this we might hope that it suffices to consider sequences $c_1, c_2, c_3, \ldots$ which are *periodic with period q*. This leads to the definition of a Dirichlet character (notation $c_n = \chi(n)$):

Definition 1.1. Let q be a positive integer. A *Dirichlet character of period* q (or "modulo q") is a function $\chi : \mathbb{Z} \to \mathbb{C}$ which is periodic with period q, i.e.

$$\chi(n+q) = \chi(n), \qquad \forall n \in \mathbb{Z}. \tag{13}$$

and *multiplicative without restriction*, i.e.

$$\chi(nm) = \chi(n)\chi(m), \qquad \forall n, m \in \mathbb{Z}, \tag{14}$$

and which also satisfies

$$\chi(1) = 1 \tag{15}$$

and

$$\chi(n) = 0 \qquad \text{whenever } (n,q) > 1. \tag{16}$$

Remark 1.1. The condition (14) is called "multiplicativity without restrictions"[2] since in number theory the term "multiplicativity" is reserved for a weaker notion: A function $f : \mathbb{Z}^+ \to \mathbb{C}$ is said to be *multiplicative* if $f(mn) = f(m)f(n)$ holds for all $m, n \in \mathbb{Z}^+$ with $(m,n) = 1$.

Remark 1.2. Given conditions (13) and (14), the condition $\chi(1) = 1$ is equivalent with saying that χ is not identically zero, and this is also equivalent with saying that $\chi(n) \neq 0$ for all n with $(n,q) = 1$.

Proof. If χ is not identically zero then there is some n with $\chi(n) \neq 0$, and (14) with $m = 1$ gives $\chi(n) = \chi(n)\chi(1)$ which implies $\chi(1) = 1$. Conversely $\chi(1) = 1$ clearly implies that χ is not identically zero. Next assume $\chi(n) = 0$ for some n with $(n,q) = 1$. By Euler's Theorem $n^{\phi(q)} \equiv 1 \pmod q$, and hence by repeated use of (14) and then (13) we get $0 = \chi(n)^{\phi(q)} = \chi(n^{\phi(q)}) = \chi(1)$, which contradicts $\chi(1) = 1$. Hence if $\chi(1) = 1$ then $\chi(n) \neq 0$ for all n with $(n,q) = 1$. □

Remark 1.3. If χ is a Dirichlet character of period q, then for all $n \in \mathbb{Z}$ with $(n,q) = 1$ we have $\chi(n)^{\phi(q)} = \chi(n^{\phi(q)}) = \chi(1) = 1$, since $n^{\phi(q)} \equiv 1 \pmod q$. In other words, $\chi(n)$ is a $\phi(q)$th root of unity for each $n \in \mathbb{Z}$ with $(n,q) = 1$, and in particular $|\chi(n)| = 1$. It follows that $|\chi(n)| \leq 1$ for *all* $n \in \mathbb{Z}$.

Example 1.1. There is exactly one Dirichlet character of period $q = 1$: $\chi \equiv 1$.

There is exactly one Dirichlet character of period $q = 2$:

$n =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\chi_0(n) =$	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	$\cdots$

There are exactly two Dirichlet characters of period $q = 3$:

[2]Other names for the same thing are "total multiplicativity" and "complete multiplicativity".

$n=$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\chi_0(n)=$	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	$\cdots$
$\chi_1(n)=$	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	$\cdots$

There are exactly two Dirichlet characters of period $q=4$:

$n=$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\chi_0(n)=$	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	$\cdots$
$\chi_1(n)=$	1	0	-1	0	1	0	-1	0	1	0	-1	0	1	0	-1	$\cdots$

There are exactly four Dirichlet characters of period $q=5$:

$n=$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$\chi_0(n)=$	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	$\cdots$
$\chi_1(n)=$	1	i	-i	-1	0	1	i	-i	-1	0	1	i	-i	-1	0	$\cdots$
$\chi_2(n)=$	1	-1	-1	1	0	1	-1	-1	1	0	1	-1	-1	1	0	$\cdots$
$\chi_3(n)=$	1	-i	i	-1	0	1	-i	i	-1	0	1	-i	i	-1	0	$\cdots$

If we study the above tables it is not hard to see that at least for $q=1,2,3,4,5$, for every $a\in\mathbb{Z}$ with $(a,q)=1$, the sequence

$$c_n=\begin{cases}1 & \text{if } n\equiv a\\ 0 & \text{otherwise,}\end{cases}$$

can be expressed as a linear combination of the Dirichlet characters modulo q. To give an example, for $q=5$ and $a=3$ this means that the following sequence;

$n=$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\cdots$
$c_n=$	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	$\cdots$

,

can be expressed as a linear combination of the four Dirichlet characters $\chi_0,\chi_1,\chi_2,\chi_3$ of period 5 (cf. the previous table), and this is indeed the case: We see that

$$c_n=\frac{1}{4}\chi_0(n)+\frac{i}{4}\chi_1(n)-\frac{1}{4}\chi_2(n)-\frac{i}{4}\chi_3(n),\qquad \forall n\in\mathbb{Z}^+.$$

This is a special case of the following lemma, which shows that the same thing works for arbitrary q:

Lemma 1.5. *Let X_q be the set of all Dirichlet characters modulo q. Then for any $a,n\in\mathbb{Z}$ with $(a,q)=1$ we have*

$$\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\chi(n)=\begin{cases}1 & \textit{if } n\equiv a\ (\textit{mod } q),\\ 0 & \textit{otherwise.}\end{cases}\tag{17}$$

We postpone the proof of this to the fourth lecture, when we discuss Dirichlet characters in more detail. (We will also see that $\#X_q = \phi(q)$.)

We will next see that if we generalize the Riemann Zeta function using Dirichlet characters then we indeed still have an Euler product.

Definition 1.2. If χ is any Dirichlet character, we define the corresponding *Dirichlet L-function* by

$$L(s,\chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}, \tag{18}$$

for all $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$.

Lemma 1.6. *The series $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)n^{-s}$ is absolutely convergent for all $s \in \mathbb{C}$ with $\sigma > 1$, and uniformly absolutely convergent in any compact subset of $\{s : \sigma > 1\}$. In particular, by Weierstrass's Theorem, $L(s,\chi)$ is an analytic function in the set $\{s : \sigma > 1\}$.*

Proof. This is exactly as the proof of Lemma 1.2, using $|\chi(n)| \leq 1$ for all n (see Remark 1.3). □

(Note that $\zeta(s)$ is a special case of a Dirichlet L-function; we have $\zeta(s) \equiv L(s,\chi)$ when $\chi \equiv 1$, the unique Dirichlet character of period $q = 1$.)

The *Euler product* for the Dirichlet L-function looks as follows:

Lemma 1.7.

$$L(s,\chi) = \prod_{p} (1 - \chi(p)p^{-s})^{-1}. \tag{19}$$

for all $s \in \mathbb{C}$ with $\sigma > 1$.

Here, again, in the product in the right hand side, p runs over all primes. The proof of (19) is deferred to the next lecture, but we mention that, just as for the Riemann Zeta function, the main step in the proof uses the fact that each positive integer has one unique prime factorization, together with the unrestricted multiplicativity of χ: The right hand

side of (19) equals

$$\begin{aligned}
&\big(1+\chi(2)2^{-s}+\chi(2)^2 2^{-2s}+\chi(2)^3 2^{-3s}+\ldots\big)\\
&\qquad\cdot\big(1+\chi(3)3^{-s}+\chi(3)^2 3^{-2s}+\chi(3)^3 3^{-3s}+\ldots\big)\\
&\qquad\qquad\cdot\big(1+\chi(5)5^{-s}+\chi(5)^2 5^{-2s}+\chi(5)^3 5^{-3s}+\ldots\big)\cdots\\
&=1+\chi(2)2^{-s}+\chi(3)3^{-s}+(\chi(2)2^{-s}\chi(3)3^{-s})+(\chi(2)2^{-s}\chi(5)5^{-s})\\
&\qquad+\chi(2)^2 2^{-2s}+(\chi(2)^2 2^{-2s}\chi(3)3^{-s})+\chi(5)5^{-s}+\ldots\\
&=1+\chi(2)2^{-s}+\chi(3)3^{-s}+\chi(2\cdot 3)(2\cdot 3)^{-s}+\chi(2\cdot 5)(2\cdot 5)^{-s}\\
&\qquad+\chi(2^2)(2^2)^{-s}+\chi(2^2\cdot 3)(2^2\cdot 3)^{-s}+\chi(5)5^{-s}+\ldots\\
&=\sum_{n=1}^{\infty}\chi(n)n^{-s}=L(s,\chi).
\end{aligned}$$

Note that none of the factors in the right hand side of (19) vanishes, since $|\chi(p)p^{-s}|\le |p^{-s}|=p^{-\sigma}<1$ for all $s\in\mathbb{C}$ with $\sigma>1$. Hence (by Theorem 2.2 which we will prove in the next lecture)

$$L(s,\chi)\neq 0 \qquad \text{for all } s\in\mathbb{C} \text{ with } \sigma>1. \tag{20}$$

It follows that $\log L(s,\chi)$ can be defined for each $s\in\mathbb{C}$ with $\sigma>1$. Note that this is in principle a multivalued function for each s. However, there is one choice of $\log L(s,\chi)$ which is the most natural: Let us *define* $\log L(s,\chi)$ by

$$\log L(s,\chi)=-\sum_p \log(1-\chi(p)p^{-s}), \qquad (s\in\mathbb{C},\ \sigma>1), \tag{21}$$

where each logarithm on the right is taken with the principal branch (this is ok since $|\chi(p)p^{-s}|<1$). We will see in the next lecture, in connection with our discussion of infinite products (see Example 2.2 below) that the sum in (21) is absolutely convergent, uniformly on compact subsets of $\{s\in\mathbb{C}:\sigma>1\}$; hence this sum defines an analytic function in $\{s\in\mathbb{C}:\sigma>1\}$, and this indeed gives a logarithm of $L(s,\chi)$ (i.e., $e^{-\sum_p \log(1-\chi(p)p^{-s})}=L(s,\chi)$).

Inserting the Taylor expansion of the logarithm in (21) we obtain, for any $s\in\mathbb{C}$ with $\sigma>1$:

$$\log L(s,\chi)=\sum_p\sum_{m=1}^{\infty} m^{-1}\chi(p^m)p^{-ms}, \tag{22}$$

where the double sum in the right hand side is absolutely convergent for any $s\in\mathbb{C}$ with $\sigma>1$ (this is seen by comparison with (7) applied with σ in place of s).

Now let a be a fixed integer with $(a,q)=1$. In order to use Lemma 1.5 we multiply (22) with $\phi(q)^{-1}\overline{\chi(a)}$, and then add over all $\chi \in X_q$. This gives

$$\begin{aligned}\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\log L(s,\chi) &= \frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\sum_{p}\sum_{m=1}^{\infty} m^{-1}\chi(p^m)p^{-ms}\\ &= \frac{1}{\phi(q)}\sum_{p}\sum_{m=1}^{\infty}\sum_{\chi\in X_q}\overline{\chi(a)}m^{-1}\chi(p^m)p^{-ms} = \sum_{\substack{p\\ p^m\equiv a \,(\mathrm{mod}\ q)}}\sum_{m=1}^{\infty} m^{-1}p^{-ms},\end{aligned} \tag{23}$$

where in the last step we used Lemma 1.5. Using comparison with (9) to treat all terms with $m \geq 2$, we obtain[3]

$$\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\log L(s,\chi) = \sum_{p\equiv a \,(\mathrm{mod}\ q)} p^{-s} + O(1), \tag{24}$$

for all $s \in \mathbb{C}$ with $\sigma > 1$. The essential idea of Dirichlet's memoir (1837) is to prove that the left side of (24) tends to $+\infty$ as $s \to 1^+$ (i.e. keeping s real and > 1). This will imply that there are infinitely many primes $p \equiv a \,(\mathrm{mod}\ q)$ (viz., Theorem 1.1) and further that the series $\sum_{p\equiv a \,(\mathrm{mod}\ q)} p^{-1}$ is divergent (viz., Theorem 1.4). Cf. our proof of (11) above.

To prove that the left side of (24) tends to $+\infty$ as $s \to 1^+$, we discuss each $\chi \in X_q$ individually. First of all, there is always one *trivial* or *principal* character in X_q; this is denoted by $\chi = \chi_0$ and is defined by $\chi_0(n) = 1$ if $(n,q) = 1$ and $\chi_0(n) = 0$ if $(n,q) > 1$. The corresponding L-function is

$$L(s,\chi_0) = \sum_{n=1}^{\infty}\chi_0(n)n^{-s} = \Big(\prod_{p\mid q}(1-p^{-s})\Big)\zeta(s). \tag{25}$$

(The last identity follows from the Euler product formula for $L(s,\chi_0)$ and for $\zeta(s)$: We have $\zeta(s) = \prod_p(1-p^{-s})^{-1}$ and $L(s,\chi_0) = \prod_p(1-\chi_0(p)p^{-s})^{-1} = \prod_{p\nmid q}(1-p^{-s})^{-1}$, since $\chi_0(p) = 0$ if $p \mid q$ and $\chi_0(p) = 1$ if $p \nmid q$.)

Using $\lim_{s\to 1^+}(1-p^{-s}) = (1-p^{-1}) > 0$ for each of the finitely many primes which divide q, and the fact that $\lim_{s\to 1^+}\zeta(s) = +\infty$, we conclude that $\lim_{s\to 1^+} L(s,\chi_0) = +\infty$, and thus also

$$\log L(s,\chi_0) \to +\infty \qquad \text{as } s \to 1^+. \tag{26}$$

[3]Recall the "big O"-notation, which you are hopefully familiar with e.g. from discussions involving Taylor expansions: If $a \geq 0$ is a *non-negative* number, the symbol "$O(a)$" is used to denote any number b for which $|b| \leq Ca$, where C is a positive constant, called *the implied constant*. When using this notation, it should always be clear for which variable ranges the bound holds. For example: "$f(x) = O(x^3)$ as $x \to \infty$" means that there is some constant $C > 0$ such that *for all sufficiently large* x we have $|f(x)| \leq Cx^3$. On the other hand, "$f(x) = O(x^3)$ for $x \geq 1$" means that there is some constant $C > 0$ such that $|f(x)| \leq Cx^3$ *holds for all* $x \geq 1$.

Since we also have $\overline{\chi_0(a)} = 1$, to complete the proof that the left side of (24) tends to $+\infty$ as $s \to 1^+$ it now suffices to show that for each choice of $\chi \in X_q$ other than $\chi = \chi_0$, $\log L(s,\chi)$ is *bounded* as $s \to 1^+$.

At this point it clarifies the situation if we note that when $\chi \neq \chi_0$, the series

$$L(s,\chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s} \tag{27}$$

is convergent not only for $\sigma > 1$ but for all s with $\sigma > 0$, and defines an analytic function of s in this region. This follows from general facts about *Dirichlet series* which we will discuss in the third lecture (see Example 3.5). I remark that a *Dirichlet series* is any series of the form $\sum_{n=1}^{\infty} c_n n^{-s}$.

(The key reason why (27) converges for all $\sigma > 0$ when $\chi \neq \chi_0$ is that then the sequence $\chi(1), \chi(2), \chi(3), \dots$ is very oscillating and in particular it has *average* 0; in fact, as we will see in the third lecture, we have $\sum_{n=k}^{k+q-1} \chi(n) = 0$ for all $\chi \in X_q \setminus \{\chi_0\}$ and all $k \in \mathbb{Z}^+$.)

Hence, our task to prove that for each $\chi \in X_q \setminus \{\chi_0\}$, $\log L(s,\chi)$ is bounded as $s \to 1^+$ is equivalent to proving that

$$L(1,\chi) \neq 0. \tag{28}$$

The proof splits into two cases:

Case 1: χ is *complex*, i.e. there is some $n \in \mathbb{Z}$ for which $\chi(n) \notin \mathbb{R}$. If we take $a = 1$ in (23) we get

$$\frac{1}{\phi(q)} \sum_{\chi \in X_q} \log L(s,\chi) = \sum_{\substack{p \\ p^m \equiv 1 \pmod q}} \sum_{m=1}^{\infty} m^{-1} p^{-ms} \tag{29}$$

for any $s \in \mathbb{C}$ with $\sigma > 1$. Take $s > 1$ (thus s real) in (29); then the right hand side is clearly real and non-negative; hence

$$\sum_{\chi \in X_q} \log L(s,\chi) \geq 0, \qquad \forall s > 1. \tag{30}$$

Exponentiating this we get

$$\prod_{\chi \in X_q} L(s,\chi) \geq 1, \qquad \forall s > 1. \tag{31}$$

Now if there is some complex $\chi \in X_q$ for which $L(1,\chi) = 0$, then if $\overline{\chi}$ denotes conjugate of χ (viz., $\overline{\chi}(n) = \overline{\chi(n)}$ for all n) then we have

$$L(1,\overline{\chi}) = \sum_{n=1}^{\infty} \overline{\chi}(n) n^{-1} = \overline{\sum_{n=1}^{\infty} \chi(n) n^{-1}} = \overline{L(1,\chi)} = 0. \tag{32}$$

Furthermore χ and $\overline{\chi}$ are *different* Dirichlet characters, since χ is complex. Hence *two* of the factors in the product $\prod_{\chi \in X_q} L(s,\chi)$ are zero at $s=1$. We now take the following fact on trust: *The function $L(s,\chi_0)$ has a meromorphic continuation to $\sigma > 0$, with one simple pole at $s=1$.* It then follows that also the product $\prod_{\chi \in X_q} L(s,\chi)$ has a meromorphic continuation to $\sigma > 0$, and at $s=1$ there is one factor which has a simple pole, *two* factors which are zero, and the other factors are analytic (zero or non-zero); hence $\prod_{\chi \in X_q} L(s,\chi)$ has a removable singularity at $s=1$ and extends to an analytic function which is *zero* at $s=1$! In particular we have

$$\lim_{s \to 1^+} \prod_{\chi \in X_q} L(s,\chi) = 0,$$

and this contradicts (31).

Hence there *cannot* exist any complex $\chi \in X_q$ with $L(1,\chi) = 0$; viz. we have proved that (28), $L(1,\chi) \neq 0$, holds for all complex $\chi \in X_q$!

Before moving on to the case of real χ, we comment on the fact which was taken on trust above: *The function $L(s,\chi_0)$ has a meromorphic continuation to $\sigma > 0$, with one simple pole at $s=1$.* To prove this, in view of the formula $L(s,\chi_0) = \big(\prod_{p|q}(1-p^{-s})\big)\zeta(s)$ (see (25)) it suffices to prove the corresponding fact for $\zeta(s)$, i.e. that *the function $\zeta(s)$ has a meromorphic continuation to $\sigma > 0$, with one simple pole at $s=1$.* A proof of this will be given in the third lecture, when we study Dirichlet series in more detail, see Example 3.6 below. (In fact we will see later that much more is true: $\zeta(s)$ has a meromorphic continuation to *the whole complex plane*, and the *only* pole is at $s=1$!)

We should also stress the difference between $L(s,\chi_0)$ and $L(s,\chi)$ with $\chi \in X_q \setminus \chi_0$: For $\chi \in X_q \setminus \chi_0$ the series $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$ actually converges for all s with $\sigma > 0$ (although we do not have *absolute* convergence when $0 < \sigma \leq 1$), and this can be used to see that $L(s,\chi)$ is analytic in the whole region $\sigma > 0$. By contrast, the series $L(s,\chi_0) = \sum_{\substack{n=1 \\ (n,q)=1}}^{\infty} n^{-s}$ does *not* converge for any s with $\sigma \leq 1$, and it is only by other means that we are able to show that it has a meromorphic continuation. This difference is also reflected in the fact that $L(s,\chi_0)$ has a *pole* at $s=1$!

Case 2: χ is *real*, i.e. $\chi(n) \in \mathbb{R}$ for all $n \in \mathbb{Z}$. Then the above argument is inapplicable, since now $\overline{\chi} = \chi$. Suppose that $L(1,\chi) = 0$. We will show that this leads to a contradiction. (We now follow Davenport pp. 33–34.)

Since $L(s,\chi)$ has a zero at $s=1$ and $L(s,\chi_0)$ has a simple pole at $s=1$, the product

$$L(s,\chi)L(s,\chi_0)$$

is analytic at $s=1$ and therefore analytic for $\sigma>0$. Since $L(2s,\chi_0)$ is analytic and $\neq 0$ in the region $\sigma>\frac{1}{2}$, the function

$$\psi(s)=\frac{L(s,\chi)L(s,\chi_0)}{L(2s,\chi_0)} \tag{33}$$

is analytic for $\sigma>\frac{1}{2}$. We also have

$$\lim_{s\to\frac{1}{2}^+}\psi(s)=0,$$

since $\lim_{s\to\frac{1}{2}^+}L(2s,\chi_0)=+\infty$. Applying the Euler product formula for the three L-functions we get (when $\sigma>1$):

$$\psi(s)=\frac{\prod_p(1-\chi(p)p^{-s})^{-1}\prod_p(1-\chi_0(p)p^{-s})^{-1}}{\prod_p(1-\chi_0(p)p^{-2s})^{-1}} \tag{34}$$

$$=\prod_p\frac{(1-\chi(p)p^{-s})^{-1}(1-\chi_0(p)p^{-s})^{-1}}{(1-\chi_0(p)p^{-2s})^{-1}}. \tag{35}$$

Here $\chi_0(p)=1$ if $p\nmid q$ and $\chi_0(p)=0$ if $p\mid q$, and since χ is real we also know that $\chi(p)=\pm1$ if $p\nmid q$, and $\chi(p)=0$ if $p\mid q$. Hence we see that if $p\mid q$ then $\frac{(1-\chi(p)p^{-s})^{-1}(1-\chi_0(p)p^{-s})^{-1}}{(1-\chi_0(p)p^{-2s})^{-1}}=1$, and also when $\chi(p)=-1$ we get $\frac{(1-\chi(p)p^{-s})^{-1}(1-\chi_0(p)p^{-s})^{-1}}{(1-\chi_0(p)p^{-2s})^{-1}}=\frac{(1+p^{-s})^{-1}(1-p^{-s})^{-1}}{(1-p^{-2s})^{-1}}=1$. Hence

$$\psi(s)=\prod_{\substack{p\\ \chi(p)=1}}\frac{(1-p^{-s})^{-1}(1-p^{-s})^{-1}}{(1-p^{-2s})^{-1}}=\prod_{\substack{p\\ \chi(p)=1}}\Big(\frac{1+p^{-s}}{1-p^{-s}}\Big).$$

This holds for $\sigma>1$. If there were no primes with $\chi(p)=1$ then we conclude $\psi(s)=1$ for all s with $\sigma>1$, and therefore by analytic continuation $\psi(s)=1$ for all s with $\sigma>\frac{1}{2}$, contradicting the fact $\lim_{s\to\frac{1}{2}^+}\psi(s)=0$.

We have for $\sigma>1$:

$$\psi(s)=\prod_{\substack{p\\ \chi(p)=1}}\Big((1+p^{-s})(1+p^{-s}+p^{-2s}+p^{-3s}+\dots)\Big), \tag{36}$$

and if this is expanded we obtain a Dirichlet series

$$\psi(s)=\sum_{n=1}^{\infty}a_n n^{-s} \tag{37}$$

with absolute convergence for all s with $\sigma>1$ (uniform absolute convergence in any compact subset of $\{s:\sigma>1\}$), where $a_1=1$ and $a_n\geq 0$ for all n. (Here we again used a fact that an infinite product over all primes may be expanded in a "formal" way. We will prove this in the second lecture; see Example 2.3 below.)

Since $\psi(s)$ is analytic for $\sigma > \frac{1}{2}$, it has an expansion in powers of $s-2$ with a radius of convergence at least $\frac{3}{2}$. This power series is

$$\psi(s) = \sum_{m=0}^{\infty} \frac{1}{m!} \psi^{(m)}(2)(s-2)^m.$$

We can calculate $\psi^{(m)}(2)$ from the Dirichlet series (37) by termwise differentiation (this is ok by Weierstrass Theorem, cf. footnote 1), and we obtain

$$\psi^{(m)}(2) = (-1)^m \sum_{n=1}^{\infty} a_n (\log n)^m n^{-2} = (-1)^m b_m,$$

say, where $b_m \geq 0$. Hence

$$\psi(s) = \sum_{m=0}^{\infty} \frac{1}{m!} b_m (2-s)^m,$$

and this holds for $|s-2| < \frac{3}{2}$. If $\frac{1}{2} < s < 2$ then since all the terms are nonnegative we have

$$\psi(s) \geq \psi(2) \geq 1,$$

and this contradicts the fact that $\lim_{s \to \frac{1}{2}^+} \psi(s) = 0$. Thus the hypothesis that $L(1,\chi) = 0$ is disproved.

This concludes the proof of Dirichlet's Theorem 1.4.

□ □ □

Let us recall what facts we haven't proved completely:

• Infinite products; manipulating them to prove the Euler product formula $L(s,\chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}$ (the formula for $\zeta(s)$ is a special case), and to get to the logarithm, $\log L(s,\chi)$. Also to see that the product formula (36) can indeed be expanded to give (37) with $a_1 = 1$ and all $a_n \geq 0$.

• Dirichlet series: Proving that if $\chi \neq \chi_0$ then $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$ converges for all $\sigma > 0$ and defines an analytic function in this region. Also proving that $\zeta(s)$ has a meromorphic continuation to $\sigma > 0$ with one simple pole at $s = 1$.

• Fact about linear combinations of Dirichlet characters: Lemma 1.5.

2. Infinite products

2.1. **Infinite products.** We review some facts and definitions about infinite products. (We borrow from Rudin, "Real and Complex Analysis", [60, §15.1-5].)

Definition 2.1. Suppose $\{u_n\}$ is a sequence of complex numbers,

$$r_n = (1+u_1)(1+u_2)\cdots(1+u_n) = \prod_{k=1}^{n}(1+u_k) \tag{38}$$

and $r = \lim_{n\to\infty} r_n$ exists. Then we write

$$r = \prod_{k=1}^{\infty}(1+u_k). \tag{39}$$

The r_n are the *partial products* of the *infinite product* (39). We shall say that the infinite product (39) converges if the sequence $\{r_n\}$ converges, i.e. if $\lim_{n\to\infty} r_n$ exists.

In the study of infinite series $\sum a_n$ it is of significance whether the a_n approach 0 rapidly. Analogously, in the study of infinite products it is of interest whether the factors are or are not close to 1. This accounts for the above notation: $1+u_n$ is close to 1 if u_n is close to 0.

Lemma 2.1. *If $u_1, \ldots, u_N$ are complex numbers, and if*

$$r_N = \prod_{n=1}^{N}(1+u_n), \qquad r_N^* = \prod_{n=1}^{N}(1+|u_n|),$$

then

$$r_N^* \le e^{|u_1|+\ldots+|u_N|} \tag{40}$$

and

$$|r_N - 1| \le r_N^* - 1. \tag{41}$$

Proof. For $x \ge 0$ the inequality $1+x \le e^x$ is an immediate consequence of $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$. Replace x by $|u_1|, \ldots, |u_N|$ and multiply the resulting inequalities. This gives (40).

To prove (41) we note that when completely expanding the product r_N using the distributive law we get

$$r_N = \prod_{n=1}^{N}(1+u_n) = \sum_{M\subset\{1,2,\ldots,N\}} \prod_{n\in M} u_n.$$

(The sum is taken over all subsets M of $\{1,\dots,N\}$, and as usual we interpret $\prod_{n\in\emptyset} u_n$ as 1.) Hence

$$|r_N - 1| = \Big|\Big(\sum_{M\subset\{1,\dots,N\}} \prod_{n\in M} u_n\Big) - 1\Big| = \Big|\sum_{\substack{M\subset\{1,\dots,N\}\\ M\neq\emptyset}} \prod_{n\in M} u_n\Big|.$$

Applying the triangle inequality and then mimicking the above computation backwards we get

$$\Big|r_N - 1\Big| \le \sum_{\substack{M\subset\{1,\dots,N\}\\ M\neq\emptyset}} \prod_{n\in M} |u_n| = \Big|\prod_{n=1}^{N}(1+|u_n|) - 1\Big| = |r_N^* - 1|.$$

□

Theorem 2.2. *Suppose $\{u_n\}$ is a sequence of complex numbers and that $\sum_{n=1}^{\infty}|u_n|$ converges. Then the product*

$$r = \prod_{n=1}^{\infty}(1+u_n) \tag{42}$$

converges, and $r = 0$ if and only if $u_n = -1$ for some n. Furthermore, if $\{n_1, n_2, n_3, \dots\}$ is any permutation of $\{1, 2, 3, \dots\}$ then we also have

$$r = \prod_{k=1}^{\infty}(1+u_{n_k}). \tag{43}$$

Definition 2.2. An infinite product satisfying the assumption in the first sentence of Theorem 2.2 is said to be *absolutely convergent*.

Proof. Write

$$r_N = (1+u_1)(1+u_2)\cdots(1+u_N) = \prod_{k=1}^{N}(1+u_k), \tag{44}$$

as before. Using Lemma 2.1 and the fact that $\sum_{n=1}^{\infty}|u_n|$ converges we conclude that $|r_N| \le e^C$ for all N, where $C = \sum_{n=1}^{\infty}|u_n| < \infty$.

Choose ε, $0 < \varepsilon < \frac{1}{2}$. There exists an N_0 such that

$$\sum_{n=N_0}^{\infty}|u_n| < \varepsilon. \tag{45}$$

Let $\{n_1, n_2, n_3, \dots\}$ be a permutation of $\{1, 2, 3, \dots\}$. If $N \ge N_0$, if M is so large that

$$\{1, 2, \dots, N\} \subset \{n_1, n_2, \dots, n_M\}, \tag{46}$$

and if s_M denotes the Mth partial product of (43) then

$$s_M - r_N = r_N\Big(\prod_{\substack{1\le k\le M\\ n_k>N}}(1+u_{n_k})-1\Big). \tag{47}$$

Hence, using Lemma 2.1, (45) and $e^\varepsilon - 1 < 2\varepsilon$ for $0<\varepsilon<\frac{1}{2}$,

$$|s_M - r_N| \le |r_N|(e^\varepsilon - 1) \le 2|r_N|\varepsilon \le 2e^C\varepsilon. \tag{48}$$

If $n_k = k$ $(k = 1,2,3,\ldots)$ then $s_M = r_M$, and (48) holds for all $M \ge N \ge N_0$; thus the sequence $\{r_n\}$ is Cauchy, so that the limit $r = \lim_{n\to\infty} r_n$ exists. Also (48) shows that

$$|r_M - r_{N_0}| \le 2|r_{N_0}|\varepsilon \quad \text{for all } \ M \ge N_0, \tag{49}$$

so that $|r_M| \ge (1-2\varepsilon)|r_{N_0}|$, and this implies that $r = 0$ if and only if $r_{N_0} = 0$, which happens if and only if $u_n = -1$ for some n. Finally, (48) also shows that $\{s_M\}$ converges to the same limit as $\{r_N\}$. □

Corollary 2.3. *Suppose $\{u_n\}$ is a sequence of bounded complex functions on a set S, such that $\sum_{n=1}^\infty |u_n(s)|$ converges uniformly on S. Then the product*

$$f(s) = \prod_{n=1}^\infty (1+u_n(s)) \tag{50}$$

converges uniformly on S, and $f(s_0) = 0$ at some $s_0 \in S$ if and only if $u_n(s_0) = -1$ for some n. Furthermore, if $\{n_1, n_2, n_3, \ldots\}$ is any permutation of $\{1,2,3,\ldots\}$ then we also have

$$f(s) = \prod_{k=1}^\infty (1+u_{n_k}(s)). \tag{51}$$

Proof. Except for the uniform convergence of (50), all the statements follow directly by applying Theorem 2.2 to the sequence $\{u_n(s)\}$, for each individual $s \in S$.

To prove the uniformity in (50), we just have to check that the argument in proof of Theorem 2.2 extends in a uniform way to the present setting. This is straightforward: Write

$$r_N(s) = \prod_{k=1}^N (1+u_k(s)). \tag{52}$$

Since each function u_n is bounded and $\sum_{n=1}^\infty |u_n(s)|$ is uniformly convergent, there exists a constant $C < \infty$ such that $\sum_{n=1}^\infty |u_n(s)| \le C$ for all $s \in S$. (Proof: Since $\sum_{n=1}^\infty |u_n(s)|$ is uniformly convergent there is some N_0 such that $\sum_{n>N_0} |u_n(s)| \le 1$ for all $s \in S$. Now $\sum_{n=1}^{N_0} |u_n(s)|$ is a finite sum of bounded functions on S, hence is itself a bounded function on S, i.e. there is some $B > 0$ such that $\sum_{n=1}^{N_0} |u_n(s)| \le B$ for all $s \in S$. Now take $C = B+1$;

then $\sum_{n=1}^{\infty} |u_n(s)| \le C$ holds for all $s \in S$.) Hence by Lemma 2.1, $|r_N(s)| \le e^C$ for all N and all $s \in S$. Choose ε, $0 < \varepsilon < \frac{1}{2}$. There exists an N_0 such that

$$\sum_{n=N_0}^{\infty} |u_n(s)| < \varepsilon, \qquad \forall s \in S. \tag{53}$$

Now for all $M \ge N \ge N_0$ and all $s \in S$ we have, using Lemma 2.1,

$$|r_M(s) - r_N(s)| = |r_N(s)| \Big| \prod_{n=N+1}^{M} (1 + u_n(s)) - 1 \Big| \le |r_N(s)|(e^{\varepsilon} - 1)$$

$$\le 2|r_N(s)|\varepsilon \le 2e^C\varepsilon. \tag{54}$$

Hence $\{r_n(s)\}$ is uniformly Cauchy, and thus uniformly convergent. □

In the above setting it is also easy to get hold of "$\log f(s)$":

Corollary 2.4. *Suppose $\{u_n\}$ is a sequence of bounded complex functions on a set S, satisfying $u_n(s) \notin (-\infty, -1]$ for all n and s, and such that $\sum_{n=1}^{\infty} |u_n(s)|$ converges uniformly on S. Then the sum $g(s) = \sum_{n=1}^{\infty} \log(1 + u_n(s))$ (principal branch of each logarithm!) is absolutely convergent, uniformly over S, and $e^{g(s)} = \prod_{n=1}^{\infty}(1 + u_n(s))$ for all s.*

Proof. Since $\sum_{n=1}^{\infty} |u_n(s)|$ converges uniformly on S there is some N such that $\sum_{n=N}^{\infty} |u_n(s)| \le \frac{1}{2}$ for all $s \in S$, and hence $|u_n(s)| \le \frac{1}{2}$ for all $n \ge N$ and all $s \in S$. Note that for all $|z| \le \frac{1}{2}$ we have, since $\log(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \dots$ (principal branch!):

$$\Big|\log(1 + z)\Big| \le \sum_{k=1}^{\infty} \frac{|z|^k}{k} \le |z| \sum_{k=1}^{\infty} \frac{2^{1-k}}{k} \le |z| \sum_{k=1}^{\infty} 2^{1-k} = 2|z|. \tag{55}$$

Now let $\varepsilon > 0$ be given. Then there is some M such that $\sum_{n=M}^{\infty} |u_n(s)| \le \varepsilon/2$ for all $s \in S$, and hence

$$\sum_{n=\max(M,N)}^{\infty} \Big|\log(1 + u_n(s))\Big| \le \sum_{n=\max(M,N)}^{\infty} 2|u_n(s)| \le \varepsilon$$

for all $s \in S$. This proves that the sum $g(s) = \sum_{n=1}^{\infty} \log(1+u_n(s))$ is absolutely convergent, uniformly over S.

Now fix $s \in S$. Using the continuity of the exponential function we have

$$e^{g(s)} = \lim_{N\to\infty} e^{\sum_{n=1}^{N} \log(1+u_n(s))} = \lim_{N\to\infty} \prod_{n=1}^{N} e^{\log(1+u_n(s))} = \lim_{N\to\infty} \prod_{n=1}^{N} (1 + u_n(s)) = \prod_{n=1}^{\infty} (1 + u_n(s)).$$

□

We end with two propositions which show that the assumption of absolute convergence in Theorem 2.2 is often necessary:

Proposition 2.5. *Suppose $u_n \geq 0$. Then*

$$\prod_{n=1}^{\infty}(1+u_n) \text{ converges if and only if } \sum_{n=1}^{\infty} u_n < \infty.$$

Proof. If $\sum_{n=1}^{\infty} u_n < \infty$ then Theorem 2.2 implies that $\prod_{n=1}^{\infty}(1+u_n)$ converges. Conversely, if $\prod_{n=1}^{\infty}(1+u_n)$ converges then for each N we have $\sum_{n=1}^{N} u_n \leq (1+u_1)(1+u_2)\cdots(1+u_N) \leq \prod_{n=1}^{\infty}(1+u_n)$, which implies that $\sum_{n=1}^{\infty} u_n \leq \prod_{n=1}^{\infty}(1+u_n) < \infty$. □

Proposition 2.6. *Suppose $0 \leq u_n < 1$. Then*

$$\prod_{n=1}^{\infty}(1-u_n) > 0 \quad \text{if and only if} \quad \sum_{n=1}^{\infty} u_n < \infty.$$

Proof. If $r_N = (1-u_1)\cdots(1-u_N)$, then $r_1 \geq r_2 \geq \cdots$ and $r_N > 0$, hence $r = \lim_{N\to\infty} r_N$ exists. If $\sum_{n=1}^{\infty} u_n < \infty$ then Theorem 2.2 implies $r > 0$. On the other hand,

$$r \leq r_N = \prod_{n=1}^{N}(1-u_n) \leq e^{-u_1-u_2-\ldots-u_N},$$

and the last expression tends to 0 as $N \to \infty$, if $\sum_{n=1}^{\infty} u_n = \infty$. □

2.2. **Infinite products over the primes.** (To a large extent we borrow the following presentation from [34, Thm. 5].) The following theorem was used (formally) in a number of special cases by Euler. Recall the definition of multiplicativity (with or without restrictions), Def. 1.1.

Proposition 2.7. *Let $f : \mathbb{Z}^+ \to \mathbb{C}$ be a multiplicative function which is not identically zero. Then*

$$\sum_{n=1}^{\infty} f(n) = \prod_{p}\big(1 + f(p) + f(p^2) + \ldots\big), \tag{56}$$

provided that the series on the left is absolutely convergent, in which case the product is also absolutely convergent.

If f is multiplicative without restrictions, then also

$$\sum_{n=1}^{\infty} f(n) = \prod_{p}(1-f(p))^{-1} \tag{57}$$

Proof. We observe first that $f(1) = 1$; for $f(n)f(1) = f(n)$ for all $n \in \mathbb{Z}^+$ by the multiplicativity, and n may be chosen so that $f(n) \neq 0$ since f is not identically zero.

Let $p_1, p_2, p_3, \ldots$ be a fixed ordering of the set of all primes, and consider the partial products of the right hand side in (56);

$$r_N = \prod_{k=1}^{N} \big(1 + f(p_k) + f(p_k^2) + \ldots\big). \tag{58}$$

The number of factors is finite, and each factor is an absolutely convergent series since $\sum |f(n)|$ is convergent. Hence, by Cauchy's theorem on multiplication of series, we may multiply out, and arrange the terms of the formal product in any order, the resulting sum being absolutely convergent. Using the fact that f is multiplicative this gives

$$= \sum_{v_1=0}^{\infty} \sum_{v_2=0}^{\infty} \cdots \sum_{v_N=0}^{\infty} \Big(f(p_1^{v_1}) f(p_2^{v_2}) \cdots f(p_N^{v_N}) \Big) = \sum_{v_1=0}^{\infty} \sum_{v_2=0}^{\infty} \cdots \sum_{v_N=0}^{\infty} f(p_1^{v_1} p_2^{v_2} \cdots p_N^{v_N}). \tag{59}$$

Note that in the last sum, $p_1^{v_1} p_2^{v_2} \cdots p_N^{v_N}$ runs through exactly those positive integers which only have prime factors $p_1, p_2, \ldots, p_N$ (or a subset of these) in their prime factorization. Thus we have proved:

$$r_N = \sum_{n \in M_N} f(n), \tag{60}$$

where M_N is the set of all positive integers whose prime factors all lie in $\{p_1, p_2, \ldots, p_N\}$ (in particular $1 \in M_N$ since 1 has no prime factors). Note that here we have used the fact that no positive integer can be prime factored in more than one way (this was needed to see that $p_1^{v_1} p_2^{v_2} \cdots p_N^{v_N}$ in (59) never visits the same positive integer twice).

Note that $M_1 \subset M_2 \subset M_3 \subset \ldots$. Furthermore, since each positive integer has a prime factorization, the sets M_N eventually exhaust $\mathbb{Z}^+$, i.e. for each positive integer n there is some N such that $n \in M_N$. Hence, since $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent, we have

$$r_N = \sum_{n \in M_N} f(n) \to \sum_{n=1}^{\infty} f(n) \qquad \text{as } N \to \infty. \tag{61}$$

This proves (56). The product is absolutely convergent, since

$$\sum_p \big| f(p) + f(p^2) + \ldots \big| \le \sum_p \big(|f(p)| + |f(p^2)| + \ldots \big) \le \sum_{n=2}^{\infty} |f(n)| < \infty.$$

If f is multiplicative without restrictions then

$$1 + f(p) + f(p^2) + f(p^3) + \ldots = 1 + f(p) + f(p)^2 + f(p)^3 + \ldots = \frac{1}{1 - f(p)},$$

the series being already known to be absolutely convergent; this gives (57). □

Example 2.1. Proof of Euler's product formula for $L(s,\chi)$, Lemma 1.7: Let χ be a Dirichlet character and take $s \in \mathbb{C}$ arbitrary with $\sigma > 1$. Set $f(n) = \chi(n)n^{-s}$. This function is multiplicative without restrictions, and we also know that $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent, by Lemma 1.6. Hence (57) in Proposition 2.7 applies, giving that $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)n^{-s} = \prod_p (1-\chi(p)p^{-s})^{-1}$, where the product is absolutely convergent.

Example 2.2. Proof of the claims about $\log L(s,\chi)$ which we needed in the proof of Dirichlet's theorem (see (21)): We wish to prove that the sum $-\sum_p \log(1-\chi(p)p^{-s})$ is absolutely convergent for $\sigma > 1$, uniformly in any compact subset of $\{s \in \mathbb{C} \,:\, \sigma > 1\}$, so that it defines an analytic function for $\sigma > 1$. We also wish to prove $e^{-\sum_p \log(1-\chi(p)p^{-s})} = L(s,\chi)$.

By the previous example we have $L(s,\chi) = \prod_p (1-\chi(p)p^{-s})^{-1}$ for all $\sigma > 1$. Let K be any compact subset of $\{s \in \mathbb{C} \,:\, \sigma > 1\}$. Then there is some $\sigma_0 > 1$ such that $\sigma \geq \sigma_0$ for all $s \in K$. Then for all $s \in K$ we have $\left|\chi(p)p^{-s}\right| \leq p^{-\sigma} \leq p^{-\sigma_0}$; also $\sum_p p^{-\sigma_0} < \infty$; hence $\sum_p \left|\chi(p)p^{-s}\right|$ is uniformly convergent for $s \in K$. Hence Corollary 2.4 applies (with $u_n(s) = -\chi(p_n)p_n^{-s}$, where $p_1, p_2, p_3, \ldots$ is some enumeration of all the prime numbers). This corollary implies that the sum $\sum_p \log((1-\chi(p)p^{-s})^{-1}) = -\sum_p \log(1-\chi(p)p^{-s})$ is absolutely convergent, uniformly over $s \in K$, and that $e^{-\sum_p \log(1-\chi(p)p^{-s})} = \prod_p (1-\chi(p)p^{-s})^{-1} = L(s,\chi)$. Done!

Example 2.3. Proof that the product formula (36) can indeed be expanded to give (37) with $a_1 = 1$ and all $a_n \geq 0$. Recall that, for $\sigma > 1$,

$$\psi(s) = \prod_{\substack{p \\ \chi(p)=1}} \Big((1+p^{-s})(1+p^{-s}+p^{-2s}+p^{-3s}+\ldots) \Big), \tag{62}$$

where χ is a character modulo q. When $\sigma > 1$ we have $|p^{-s}| < 1$ for each p and we compute

$$\psi(s) = \prod_{\substack{p \\ \chi(p)=1}} \big(1+2p^{-s}+2p^{-2s}+2p^{-3s}+\ldots \big), \tag{63}$$

where the inner series is absolutely convergent for each individual p. Hence to apply Proposition 2.7 we let (for some *fixed* $s \in \mathbb{C}$ with $\sigma > 1$) $f : \mathbb{Z}^+ \to \mathbb{C}$ be the unique multiplicative function such that for p a prime and $k \geq 1$ we have

$$f(p^k) = \begin{cases} 0 & \text{if } \chi(p) \neq 1 \\ 2p^{-ks} & \text{if } \chi(p) = 1. \end{cases}$$

It is an easy consequence of the unique prime factorization theorem that a multiplicative function can be completely and uniquely specified by telling its values at all prime powers.

In the present case, for general $n \in \mathbb{Z}^+$ we have $f(n) = \prod_{p|n} \begin{cases} 0 & \text{if } \chi(p) \neq 1 \\ 2p^{-(\mathrm{ord}_p n)s} & \text{if } \chi(p) = 1 \end{cases} = n^{-s} \prod_{p|n} \begin{cases} 0 & \text{if } \chi(p) \neq 1 \\ 2 & \text{if } \chi(p) = 1 \end{cases}$.

In order to apply Proposition 2.7 we need to see that $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent. One way to do this is as follows: Let X be some large integer and let Y be the number of primes $\leq X$. Take $n \in \mathbb{Z}^+$ and assume that n has prime factorization $n = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$ (with all $r_j \geq 1$). Then $|f(n)| = |n^{-s}| \prod_{k=1}^{m} \begin{cases} 0 & \text{if } \chi(p_k) \neq 1 \\ 2 & \text{if } \chi(p_k) = 1 \end{cases} \leq n^{-\sigma} 2^m$. On the other hand we have $n \geq p_1 p_2 \cdots p_m$ and if $m > Y$ then at least $m - Y$ primes among $p_1, p_2, \ldots, p_m$ must be $> X$, hence $n > X^{m-Y}$, and hence we always have $m < Y + \log_X n$. Hence

$$|f(n)| \leq n^{-\sigma} 2^m < n^{-\sigma} 2^{Y + \log_X n} = 2^Y n^{\frac{\log 2}{\log X} - \sigma}. \tag{64}$$

Since $\sigma > 1$ we can fix X so large that $\frac{\log 2}{\log X} - \sigma < -1$. For such fixed X, Y, the above inequality shows that $\sum_{n=1}^{\infty} f(n)$ is indeed absolutely convergent.

Hence Proposition 2.7 applies and shows that

$$\begin{aligned} \psi(s) &= \prod_{\substack{p \\ \chi(p)=1}} \left(1 + 2p^{-s} + 2p^{-2s} + 2p^{-3s} + \ldots\right) = \prod_{p} (1 + f(p) + f(p^2) + \ldots) \\ &= \sum_{n=1}^{\infty} f(n) = \sum_{n=1}^{\infty} n^{-s} \prod_{p|n} \begin{cases} 0 & \text{if } \chi(p) \neq 1 \\ 2 & \text{if } \chi(p) = 1 \end{cases}, \end{aligned}$$

with both the sum and the product being absolutely convergent when $\sigma > 1$. In other words $\psi(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ where $a_n = \prod_{p|n} \begin{cases} 0 & \text{if } \chi(p) \neq 1 \\ 2 & \text{if } \chi(p) = 1 \end{cases}$; thus $a_1 = 1$ and $a_n \geq 0$ for all n. Note that the above argument, using (64), also shows that $\sum_{n=1}^{\infty} a_n n^{-s}$ is *uniformly* absolutely convergent on any compact subset of $\{s : \sigma > 1\}$.

□

In fact the above analysis to prove the absolute convergence of $\sum_{n=1}^{\infty} f(n)$ could have been avoided, by noting the following general principle on going from absolute convergence of the product to absolute convergence of the sum:

Lemma 2.8. *Let $f : \mathbb{Z}^+ \to \mathbb{C}$ be a multiplicative function which is not identically zero and assume that the product*

$$\prod_{p} (1 + |f(p)| + |f(p^2)| + \ldots)$$

is absolutely convergent (in particular we assume that each sum $1+|f(p)|+|f(p^2)|+\ldots$ *is convergent). Then also the sum* $\sum_{n=1}^{\infty} f(n)$ *is absolutely convergent, and hence Proposition 2.7 applies to give*

$$\sum_{n=1}^{\infty} f(n) = \prod_{p} \big(1+f(p)+f(p^2)+\ldots\big), \tag{65}$$

The proof of this lemma is left as an exercise, see Problem 2.5.

Example 2.4. Using Lemma 2.8 the discussion in Example 2.3 can been simplified as follows: Take s with $\sigma>1$ as before. *We know from its construction* that the infinite product

$$\psi(s) = \prod_{\substack{p\\ \chi(p)=1}} \Big((1+p^{-s})(1+p^{-s}+p^{-2s}+p^{-3s}+\ldots)\Big) \tag{66}$$

converges! Indeed $\psi(s)$ was obtained by multiplying together three convergent Euler products with all factors being non-zero, see (34). In particular this holds if s is *real*, viz. $s=\sigma>1$, viz. the product

$$\prod_{\substack{p\\ \chi(p)=1}} \Big((1+p^{-\sigma})(1+p^{-\sigma}+p^{-2\sigma}+p^{-3\sigma}+\ldots)\Big) = \prod_{\substack{p\\ \chi(p)=1}} \big(1+2p^{-\sigma}+2p^{-2\sigma}+2p^{-3\sigma}+\ldots\big) \tag{67}$$

converges. Hence (by Proposition 2.5) we have $\sum_{\substack{p\\ \chi(p)=1}} \big(2p^{-\sigma}+2p^{-2\sigma}+2p^{-3\sigma}+\ldots\big)<\infty$. Defining now $f(n)$ as in Example 2.3 (for some fixed, arbitrary $s\in\mathbb{C}$ with $\sigma>1$) we conclude

$$\sum_{\substack{p\\ \chi(p)=1}} \big(|f(p)|+|f(p^2)|+\ldots\big) = \sum_{\substack{p\\ \chi(p)=1}} \big(2p^{-\sigma}+2p^{-2\sigma}+2p^{-3\sigma}+\ldots\big) < \infty.$$

This means by definition that the product

$$\prod_{\substack{p\\ \chi(p)=1}} \big(1+|f(p)|+|f(p^2)|+\ldots\big)$$

is absolutely convergent, and hence by Lemma 2.8 also the sum $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent. The proof can now be concluded as in Example 2.3.

2.3. **Problems.**

Problem 2.1. **The Möbius μ-function.** The Möbius function $\mu:\mathbb{Z}^+\to\{-1,0,1\}$ is defined as follows. Let $n\in\mathbb{Z}^+$ have the prime factorization $n=p_1^{\alpha_1}\cdots p_r^{\alpha_r}$ ($p_1,\ldots,p_r$ distinct primes and $\alpha_1,\ldots,\alpha_r\in\mathbb{Z}^+$). Then $\mu(n)=1$ if $\alpha_1=\ldots=\alpha_r=1$ and r even;

$\mu(n) = -1$ if $\alpha_1 = \ldots = \alpha_r = 1$ and r odd, and $\mu(n) = 0$ otherwise.
(a). Prove that

$$\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \qquad \sigma > 1.$$

(b). Prove that for any Dirichlet character χ,

$$L(s,\chi)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)\chi(n)}{n^s}, \qquad \sigma > 1.$$

Problem 2.2. **Euler's ϕ-function.** Recall that Euler's ϕ-function is defined as $\phi(n) = \#(\mathbb{Z}/n\mathbb{Z})^\times$ for $n \in \mathbb{Z}^+$, where $(\mathbb{Z}/n\mathbb{Z})^\times$ denotes the group of invertible elements in $\mathbb{Z}/n\mathbb{Z}$, the ring of all integer residue classes modulo n. Also recall that

$$\phi(n) = n \prod_{p|n} \Big(1 - \frac{1}{p}\Big).$$

(Cf. Section 4.1 and in particular Example 4.2.) Prove that

$$\sum_{n=1}^{\infty} \phi(n) n^{-s} = \frac{\zeta(s-1)}{\zeta(s)} \qquad (\sigma > 2).$$

Problem 2.3. For $n \in \mathbb{Z}^+$ with prime factorization $n = \prod_{j=1}^{r} p_j$ (where the primes p_j need not be distinct) we set $\lambda(n) = (-1)^r$. (In particular $\lambda(1) = 1$ corresponding to $r = 0$.)
(a). Prove that $\frac{\zeta(2s)}{\zeta(s)} = \sum_{n=1}^{\infty} \lambda(n) n^{-s}$ when $\sigma > 1$.
(b). Prove that $\frac{\zeta(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} |\mu(n)| n^{-s}$ when $\sigma > 1$. (Note here that $|\mu(n)| = 1$ if n is *squarefree*, i.e. if n has no repeated prime in its prime factorization, and otherwise $|\mu(n)| = 0$.)

Problem 2.4. **Euler product of degree 2.** (The following problem illustrates an Euler product of degree 2; this type of L-functions appear from several sources in number theory, such as modular forms, elliptic curves, Galois representations.) Let $A \in \mathbb{R}$ and $\alpha \in \mathbb{C}$ be given constants, and let $\{a_n\}$ be a sequence of complex numbers satisfying $|a_n| = O(n^A)$ for all $n \in \mathbb{Z}^+$. Assume that the sequence $\{a_n\}$ is multiplicative, not identically zero, and that $a_p a_{p^k} = a_{p^{k+1}} + p^\alpha a_{p^{k-1}}$ holds for every prime p and every $k \geq 1$. Then prove that

$$\sum_{n=1}^{\infty} a_n n^{-s} = \prod_{p} (1 - a_p p^{-s} + p^\alpha p^{-2s})^{-1} \qquad (\sigma > A + 1).$$

Problem 2.5. Prove Lemma 2.8.

Problem 2.6. Let $\rho_1, \rho_2, \ldots$ be a sequence of non-zero complex numbers, let k be an integer ≥ 0, and assume that $\sum_{n=1}^{\infty} |\rho_n|^{-1-k} < \infty$. Then prove that the following product is

absolutely convergent for each $z \in \mathbb{C}$:

$$f(z) = \prod_{n=1}^{\infty} \Big\{ \Big(1 - \frac{z}{\rho_n}\Big) \exp\Big(\frac{z}{\rho_n} + \frac{1}{2}\big(\frac{z}{\rho_n}\big)^2 + \ldots + \frac{1}{k}\big(\frac{z}{\rho_n}\big)^k \Big) \Big\}.$$

Furthermore prove that $f(z)$ is an analytic function of $z \in \mathbb{C}$ which has a zero at each point $z = \rho_j$ and no other zeros in the plane. More precisely prove that if α occurs m times in the sequence $\{\rho_1, \rho_2, \ldots\}$ then f has a zero of order (exactly) m at α.

[Hint: Write $E(w) = (1-w)\exp(w + \frac{1}{2}w^2 + \ldots + \frac{1}{k}w^k)$ so that the above product is $\prod_{n=1}^{\infty} E(z/\rho_n) = \prod_{n=1}^{\infty}(1+u_n)$ with $u_n = E(z/\rho_n) - 1$. Now start by proving that if $u = E(w) - 1$ then $u \ll |w|^{k+1}$ for all $w \in \mathbb{C}$ sufficiently near 0. This can be done by studying the power series of $(1-w)\exp(w + \frac{1}{2}w^2 + \ldots + \frac{1}{k}w^k)$ at $w = 0$.]

3. Partial summation and Dirichlet series

3.1. **Integration by parts.**

Lemma 3.1. *Assume $A < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_m \le B$; let $c_1, c_2, \ldots, c_m \in \mathbb{C}$, and set*

$$f(x) = \sum_{\lambda_n \le x} c_n$$

(the notation indicates a summation over the finite set of positive integers n for which $\lambda_n \le x$). Then, if $g \in C^1([A, B])$, we have

$$\sum_{n=1}^{m} c_n g(\lambda_n) = f(B)g(B) - \int_A^B f(x)g'(x)\,dx. \tag{68}$$

Proof. Note that $f(B) = \sum_{n=1}^{m} c_n$. Hence

$$\begin{aligned} f(B)g(B) - \sum_{n=1}^{m} c_n g(\lambda_n) &= \sum_{n=1}^{m} c_n (g(B) - g(\lambda_n)) = \sum_{n=1}^{m} c_n \int_{\lambda_n}^{B} g'(x)\,dx \\ &= \int_A^B \sum_{\lambda_n \le x} c_n g'(x)\,dx = \int_A^B f(x)g'(x)\,dx. \end{aligned}$$

This proves (68).

(Regarding the change of summation and integration in the above computation: We may express $\sum_{n=1}^{m} c_n \int_{\lambda_n}^{B} g'(x)\,dx$ as $\sum_{n=1}^{m} \int_A^B I(\lambda_n \le x) c_n g'(x)\,dx$, where $I(\cdot)$ is the indicator function. Here the sum is finite and the range of integration is a compact interval $[A, B]$; furthermore each integrand is a piecewise continuous function; hence we may change order of summation, obtaining $\int_A^B \sum_{n=1}^{m} I(\lambda_n \le x) c_n g'(x)\,dx = \int_A^B \sum_{\lambda_n \le x} c_n g'(x)\,dx$, as desired.) □

We will next explain how the left side of (68) may be expressed as the Riemann-Stieltjes integral $\int_A^B g(x)df(x)$. In this way (68) will be seen to be nothing more than a variant of the usual formula for integration by parts. The Riemann-Stieltjes notation is convenient for remembering Lemma 3.1, and it also makes it easier to apply Lemma 3.1 in a flexible way.

Definition 3.1. (Definition of a Riemann-Stieltjes integral in special cases.) Let $A < B$ and $g \in C([A, B])$. Then if f is a piecewise constant function on $[A, B]$, that is, if there are numbers $A = x_0 < x_1 < x_2 < \ldots < x_n = B$ such that f is constant on each open interval

(x_j, x_{j+1}), $j = 0, 1, \ldots, n-1$, we define the Riemann-Stieltjes integral $\int_A^B g(x)\, df(x)$ as

$$\int_A^B g(x)\, df(x) = \big(f(x_0+) - f(x_0)\big)g(x_0) + \sum_{j=1}^{n-1}\big(f(x_j+) - f(x_j-)\big)g(x_j) + \big(f(x_n) - f(x_n-)\big)g(x_n), \tag{69}$$

where

$$f(x+) = \lim_{t\to x^+} f(t) \qquad \text{and} \qquad f(x-) = \lim_{t\to x^-} f(t). \tag{70}$$

Also, if $f \in C^1([A,B])$ then we define the Riemann-Stieltjes integral $\int_A^B g(x)\, df(x)$ as

$$\int_A^B g(x)\, df(x) = \int_A^B g(x) f'(x)\, dx, \tag{71}$$

where the right hand side is an ordinary Riemann integral of a continuous function.

Hopefully you agree, after drawing a picture, that the above definition is natural. The general Riemann-Stieltjes integral is discussed in Section 3.2 (not required for this course).

Now Lemma 3.1 can be reformulated (or slightly generalized) as follows:

Lemma 3.2. *If $A < B$, $g \in C^1([A,B])$ and f is a piecewise constant function on $[A,B]$, then*

$$\int_A^B g(x)\, df(x) = \Big(f(B)g(B) - f(A)g(A)\Big) - \int_A^B f(x)g'(x)\, dx. \tag{72}$$

Proof. We prove this as a consequence of Lemma 3.1. Since f is piecewise constant on $[A,B]$ there are some $A < \lambda_1 < \lambda_2 < \ldots < \lambda_n = B$ such that f is constant on each open interval (A, λ_1) and $(\lambda_j, \lambda_{j+1})$, $j = 1, 2, \ldots, n-1$. Set $c_j = f(\lambda_j+) - f(\lambda_j-)$ for $j = 1, 2, \ldots, n-1$ and $c_n = f(B) - f(B-)$, and define $f_0(x) := \sum_{\lambda_j \le x} c_j$. Then we have $f(x) = f_0(x) + f(A+)$ for all $x \in [A,B] \setminus \{A, \lambda_1, \ldots, \lambda_{n-1}\}$. We also get from the definition (69) that

$$\int_A^B g(x)\, df(x) = \big((f(A+) - f(A)\big)g(A) + \sum_{j=1}^{n} c_j g(\lambda_j). \tag{73}$$

Applying Lemma 3.1 we have

$$\begin{aligned}\sum_{j=1}^{n} c_j g(\lambda_j) &= f_0(B)g(B) - \int_A^B f_0(x)g'(x)\,dx \\ &= \big(f(B) - f(A+)\big)g(B) - \int_A^B \big(f(x) - f(A+)\big)g'(x)\,dx \\ &= \big(f(B) - f(A+)\big)g(B) + f(A+)\big(g(B) - g(A)\big) - \int_A^B f(x)g'(x)\,dx \\ &= f(B)g(B) - f(A+)g(A) - \int_A^B f(x)g'(x)\,dx. \qquad (74)\end{aligned}$$

Combining (73) and (74) we obtain (72). □

In order to make the notation really flexible we also need the following definition of generalized Riemann-Stieltjes integrals.

Definition 3.2. We define the generalized Riemann-Stieltjes integral

$$\int_{A+}^{B} g(x)\,df(x) := \lim_{a\to A^+} \int_a^B g(x)\,df(x), \qquad (75)$$

provided that $\int_a^B g(x)\,df(x)$ is well-defined by Definition 3.1 for all $a > A$ sufficiently near A. (Then $\int_{A+}^{B} g(x)\,df(x)$ is said to *exist* or *not exist* according to as the limit $\lim_{a\to A^+} \int_{A_1}^{B} g(x)\,df(x)$ exists or does not exist.)

Similarly we define

$$\int_{A-}^{B} g(x)\,df(x) := \lim_{a\to A^-} \int_a^B g(x)\,df(x); \qquad (76)$$

$$\int_{-\infty}^{B} g(x)\,df(x) := \lim_{a\to -\infty} \int_a^B g(x)\,df(x), \qquad (77)$$

Also, the generalized Riemann-Stieltjes integrals $\int_A^{B-} g(x)\,df(x)$, $\int_A^{B+} g(x)\,df(x)$ and $\int_A^{\infty} g(x)\,df(x)$ are defined in the analogous way.

Finally generalized Riemann-Stieltjes integrals with limits on both end-points are defined in the natural way, i.e.

$$\int_{A-}^{B+} g(x)\,df(x) := \lim_{b\to B^+} \lim_{a\to A^-} \int_a^b g(x)\,df(x); \qquad (78)$$

$$\int_{-\infty}^{B-} g(x)\,df(x) := \lim_{b\to B^-} \lim_{a\to -\infty} \int_a^b g(x)\,df(x), \qquad (79)$$

etc.

Remark 3.1. In (78) (and similarly in any of the other cases with limits on both endpoints) it does not matter if the limit is considered as an iterated limit (in either order) or as a simultanous limit in a, b; if one of these limits exist (as a finite real number) then so do the other ones. This follows by fixing an arbitrary number $C \in (A, B)$ and using $\int_a^b g(x)\,df(x) = \int_a^C g(x)\,df(x) + \int_C^b g(x)\,df(x)$ inside the limit.

We now give several examples to illustrate the use and flexibility of our new notation:

Example 3.1. Let $a_1, a_2, \ldots$ be any sequence of complex numbers, and set $f(x) = \sum_{1\le n<x} a_n$. Also let $g \in C(\mathbb{R}^+)$. We then have, for any integers $1 \le M \le N$:

$$\sum_{n=M}^{N} a_n g(n) = \int_M^{N+} g(x)\,df(x) = \int_M^{N+\frac{1}{2}} g(x)\,df(x). \tag{80}$$

Hence also

$$\sum_{n=M}^{\infty} a_n g(n) = \int_M^{\infty} g(x)\,df(x). \tag{81}$$

On the other hand, if we set $f_1(x) = \sum_{1\le n\le x} a_n$ (thus $f_1(x) = f(x)$ except when x is an integer) then

$$\sum_{n=M}^{N} a_n g(n) = \int_{M-}^{N} g(x)\,df_1(x) = \int_{M-\frac{1}{2}}^{N} g(x)\,df_1(x) \tag{82}$$

and

$$\sum_{n=M}^{\infty} a_n g(n) = \int_{M-}^{\infty} g(x)\,df_1(x). \tag{83}$$

Example 3.2. A counting function of fundamental importance in number theory is

$$\pi(x) = \#\big\{p \,:\, p \text{ is a prime number} \le x\big\}.$$

(cf. Definition 6.1 below). In terms of this function we can write, e.g.

$$\sum_{p\le A} \frac{1}{p} = \int_1^A \frac{1}{x}\,d\pi(x).$$

Thus, as we saw in the first lecture, we have $\int_1^\infty \frac{1}{x}\,d\pi(x) = \infty$.

Example 3.3. We proved Lemma 3.2 as a consequence of the initial Lemma 3.1. Let us now check that Lemma 3.1 is in fact a special case of Lemma 3.2: Recall that Lemma 3.1 says that if $A < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_m \le B$, $f(x) = \sum_{\lambda_n\le x} c_n$ and $g \in C^1([A, B])$, then

$$\sum_{n=1}^{m} c_n g(\lambda_n) = f(B)g(B) - \int_A^B f(x)g'(x)\,dx. \tag{84}$$

But the left hand side is the same as the Riemann-Stieltjes integral $\int_A^B g(x)\,df(x)$ (since $A < \lambda_1$!), and hence (84) is just a special case of our partial integration identity in Lemma 3.2 (since $f(A) = 0$).

Note that Lemma 3.2 together with Definition 3.2 immediately implies that, e.g.,

$$\int_{A-}^{\infty} g(x)\,df(x) = \lim_{b\to\infty}\lim_{a\to A^-}\Big\{\Big(f(b)g(b) - f(a)g(a)\Big) - \int_a^b f(x)g'(x)\,dx\Big\}, \tag{85}$$

provided that either the left or the right side is convergent.

Example 3.4. Suppose that $a_1, a_2, \dots$ are complex numbers such that $\sum_{n=1}^N a_n n^{-\frac12} = N + O(N^{\frac23})$ as $N\to\infty$. Then what can we say about the asymptotic behaviour of $\sum_{n=1}^N a_n$ as $N\to\infty$?

Solution: Write $A(x) = \sum_{1\le n\le x} a_n n^{-\frac12}$; then we know that $A(x) = x + O(x^{\frac23})$ as $x\to\infty$. (In detail: $A(x) = A(\lfloor x\rfloor) = \lfloor x\rfloor + O(\lfloor x\rfloor^{\frac23}) = x + O(1) + O(x^{\frac23}) = x + O(x^{\frac23})$ as $x\to\infty$.) Hence (by possibly taking a larger implied constant) we have

$$A(x) = x + O(x^{\frac23}), \qquad \forall x \ge 1. \tag{86}$$

Now for $N \ge 1$ we have

$$\sum_{n=1}^N a_n = \sum_{n=1}^N n^{\frac12}(a_n n^{-\frac12}) = \int_{1-}^N x^{\frac12}\,dA(x). \tag{87}$$

By Lemma 3.2 (together with a similar type of limit identity as in (85)) this is

$$= N^{\frac12}A(N) - \frac12\int_1^N x^{-\frac12}A(x)\,dx. \tag{88}$$

Using now (86) we get, as $N\to\infty$,

$$\begin{aligned} &= N^{\frac12}\big(N + O(N^{\frac23})\big) - \frac12\int_1^N x^{-\frac12}\big(x + O(x^{\frac23})\big)\,dx \\ &= N^{\frac32} + O(N^{\frac76}) - \frac12\int_1^N x^{\frac12}\,dx + O(1)\int_1^N x^{\frac16}\,dx \\ &= N^{\frac32} + O(N^{\frac76}) - \frac12\big(\tfrac23 N^{\frac32} - \tfrac23\big) + O\big(N^{\frac76}\big) \\ &= \frac23 N^{\frac32} + O(N^{\frac76}). \end{aligned} \tag{89}$$

Answer: We have $\sum_{n=1}^N a_n = \frac23 N^{\frac32} + O(N^{\frac76})$ as $N\to\infty$.

(It is easy to see that this is the best possible error term that can be obtained under the given assumption. Indeed, let us take for example $a_n = n^{\frac12} + n^{\frac16}$; then $\sum_{n=1}^N a_n n^{-\frac12} = \sum_{n=1}^N 1 + \sum_{n=1}^N n^{-\frac13} = N + \sum_{n=1}^N n^{-\frac13}$, and by a standard integral estimate (use $n^{-\frac13} <$

$\int_{n-1}^{n} x^{-\frac{1}{3}}\,dx$ for each $n \geq 1$) this is $\leq N + \int_0^N x^{-\frac{1}{3}}\,dx = N + \frac{3}{2}N^{\frac{2}{3}}$; in particular the assumption $\sum_{n=1}^N a_n n^{-\frac{1}{2}} = N + O(N^{\frac{2}{3}})$ is fulfilled. On the other hand, for the same sequence we have $\sum_{n=1}^N a_n = \sum_{n=1}^N n^{\frac{1}{2}} + \sum_{n=1}^N n^{\frac{1}{6}}$. Here we have $\int_0^N x^{\frac{1}{2}}\,dx \leq \sum_{n=1}^N n^{\frac{1}{2}} \leq \int_1^N x^{\frac{1}{2}}\,dx + N^{\frac{1}{2}}$, thus $\frac{2}{3}N^{\frac{3}{2}} \leq \sum_{n=1}^N n^{\frac{1}{2}} \leq \frac{2}{3}N^{\frac{3}{2}} + N^{\frac{1}{2}}$, and in an entirely similar way, $\frac{6}{7}N^{\frac{7}{6}} \leq \sum_{n=1}^N n^{\frac{1}{6}} \leq \frac{6}{7}N^{\frac{7}{6}} + N^{\frac{1}{6}}$. In particular:

$$\sum_{n=1}^{N} a_n \geq \tfrac{2}{3}N^{\frac{3}{2}} + \tfrac{6}{7}N^{\frac{7}{6}}. \tag{90}$$

On the other hand if we take another example; $a_n = n^{\frac{1}{2}} - n^{\frac{1}{6}}$ then by an entirely similar argument one checks that the assumption $\sum_{n=1}^N a_n n^{-\frac{1}{2}} = N + O(N^{\frac{2}{3}})$ is again fulfilled, and this time we get $\sum_{n=1}^N a_n \leq \frac{2}{3}N^{\frac{3}{2}} + N^{\frac{1}{2}} - \frac{6}{7}N^{\frac{7}{6}}$ for all $N \geq 1$, thus[4]

$$\sum_{n=1}^{N} a_n \leq \tfrac{2}{3}N^{\frac{3}{2}} - \tfrac{6}{7}N^{\frac{7}{6}}(1 + o(1)) \qquad \text{as } N \to \infty. \tag{91}$$

Since both (90) and (91) can hold under the given assumption, it follows that the error term in "$\sum_{n=1}^N a_n = \frac{2}{3}N^{\frac{3}{2}} + O(N^{\frac{7}{6}})$" cannot be improved.)

3.2. *** The general Riemann-Stieltjes integral.** This section is external reading and not required for the course.

We follow [48, Appendix A].

The Riemann-Stieltjes Integral $\int_A^B g(x)\,df(x)$ is defined as a limit of Riemann sums $\sum_n g(\xi_n)\Delta f(x_n)$. More precisely:

Definition 3.3. Let numbers $A < B$ and two functions $f, g : [A, B] \to \mathbb{C}$ be given. For any partition

$$A = x_0 \leq x_1 \leq \ldots \leq x_N = B \tag{92}$$

and any choices of numbers $\xi_j \in [x_{j-1}, x_j]$ for $j = 1, 2, \ldots, N$, we form the sum

$$S(\{x_n\}, \{\xi_n\}) = \sum_{n=1}^{N} g(\xi_n)\big(f(x_n) - f(x_{n-1})\big). \tag{93}$$

We say that the Riemann-Stieltjes integral $\int_A^B g(x)\,df(x)$ exists and has the value I if for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\big|S(\{x_n\}, \{\xi_n\}) - I\big| < \varepsilon \tag{94}$$

[4]The "little o"-notation: We write "$f(x) = o(g(x))$ as $x \to a$" to denote that $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$; we will only use this notation when $g(x) > 0$ for all x sufficiently near a! Thus in the situation in (91), "$o(1)$" denotes some function $f(N)$ which satisfies $\lim_{N\to\infty} f(N) = 0$. Note that, unlike the "big O"-notation (cf. footnote 3 on page 13), *the "little o"-notation can only be used when we are taking a limit.*

whenever $\{x_n\}$ and $\{\xi_n\}$ are as above and

$$\operatorname{mesh}\{x_n\} := \max_{1\le n\le N}(x_n - x_{n-1}) \le \delta. \tag{95}$$

To give an overview we will now state some results about the Riemann-Stieltjes integral, many without proofs.

Definition 3.4. If f is a function $f : [A, B] \to \mathbb{C}$, then the *variation of f over* $[A, B]$, $\operatorname{Var}_{[A,B]}(f)$, is defined by

$$\operatorname{Var}_{[A,B]}(f) = \sup \sum_{n=1}^{N} |f(x_n) - f(x_{n-1})|, \tag{96}$$

where the supremum is taken over all $\{x_n\}$ satisfying $A = x_0 \le x_1 \le \ldots \le x_N = B$. The function f is said to be of *bounded variation* if $\operatorname{Var}_{[a,b]}(f) < \infty$.

Theorem 3.3. *(Cf.* [48, Appendix A, Thm. 1]*.) The Riemann-Stieltjes integral $\int_a^b g(x)\,df(x)$ exists if g is continuous on $[a, b]$ and f is of bounded variation on $[a, b]$.*

Our preliminary Definition 3.1 is now justified with:

Lemma 3.4. *If $A < B$, $g \in C([A, B])$, and $f : [A, B] \to \mathbb{C}$ is either piecewise constant or C^1, then the Definition 3.1 gives the same $\int_A^B g(x)\,df(x)$ as Definition 3.3.*

There is a formula of partial integration which holds in a general case, i.e. a generalization of Lemma 3.2:

Theorem 3.5. *For arbitrary functions $f : [A, B] \to \mathbb{C}$ and $g : [A, B] \to \mathbb{C}$, if $\int_A^B g(x)\,df(x)$ exists then $\int_A^B f(x)\,dg(x)$ also exists, and*

$$\int_A^B g(x)\,df(x) = \Big(f(B)g(B) - f(A)g(A)\Big) - \int_A^B f(x)\,dg(x). \tag{97}$$

(To see that this is a generalization of Lemma 3.2 we have to use a variant of Lemma 3.4 to see that if f is piecewise constant and $g \in C^1([A, B])$ then in the right hand side of (97) we have $\int_A^B f(x)\,dg(x) = \int_A^B f(x)g'(x)\,dx$. Cf., e.g., [48, Appendix A, Thm. 3].)

The proof of Theorem 3.5 is very simple:

Proof. For any partition $A = x_0 \le x_1 \le \ldots \le x_N = B$ and any choices of numbers $\xi_n \in [x_{n-1}, x_n]$ for $j = 1, 2, \ldots, N$, we have the following identity, if we set $\xi_0 = A$ and $\xi_{N+1} = B$:

$$\sum_{n=1}^{N} g(\xi_n)\big(f(x_n) - f(x_{n-1})\big) = f(b)g(b) - f(a)g(a) - \sum_{n=1}^{N+1} f(x_{n-1})\big(g(\xi_n) - g(\xi_{n-1})\big). \tag{98}$$

Here the sum on the right hand sum is a Riemann-Stieltjes sum $S(\{\xi_n\},\{x_{n-1}\})$ approximating $\int_A^B f(x)\,dg(x)$, since $x_{n-1}\in[\xi_{n-1},\xi_n]$. Moreover, $\mathrm{mesh}\{\xi_n\}\le\mathrm{mesh}\{x_n\}$, so that the sum on the right tends to $\int_A^B f(x)\,dg(x)$ as $\mathrm{mesh}\{x_n\}$ tends to 0. □

Remark 3.2. One has to be careful when working with the general Riemann-Stieltjes integral, since some rules which are familiar from ordinary integrals may fail to hold in general! For example, it is *not* always true that if $A<C<B$ then $\int_A^B g(x)\,df(x)=\int_A^C g(x)\,df(x)+\int_C^B g(x)\,df(x)$! An example of this is the following: Suppose that

$$f(x)=\begin{cases}1 & \text{if } 0\le x\le 1\\ 0 & \text{otherwise;}\end{cases}\qquad\qquad g(x)=\begin{cases}1 & \text{if } 0<x\le 1\\ 0 & \text{otherwise.}\end{cases}\tag{99}$$

Then $\int_{-1}^0 g\,df$ and $\int_0^1 g\,df$ both exist, but $\int_{-1}^1 g\,df$ does not exist! (Prove this as an exercise!)

We stress however, that the rule $\int_A^B g(x)\,df(x)=\int_A^C g(x)\,df(x)+\int_C^B g(x)\,df(x)$ is always true when g is continuous and f is bounded on $[A,B]$, and in particular it is always true in the special cases which are covered by Definition 3.1!

3.3. **Dirichlet series; convergence properties.** (To a large extent we follow Montgomery and Vaughan [48, §1.2] in this section.)

A series of the form $\sum_{n=1}^\infty a_n n^{-s}$ is called a *Dirichlet series*.

Theorem 3.6. *Let $a_1,a_2,\ldots\in\mathbb{C}$ and suppose that the Dirichlet series $\alpha(s)=\sum_{n=1}^\infty a_n n^{-s}$ is convergent for $s=s_0=\sigma_0+it_0$, and let $H>0$ be an arbitrary constant. Then the series $\alpha(s)$ is uniformly convergent in the sector*

$$\mathcal{S}=\{s=\sigma+it\,:\,\sigma\ge\sigma_0,\ |t-t_0|\le H(\sigma-\sigma_0)\}.\tag{100}$$

By taking H large, we see that the series $\alpha(s)$ converges for all s in the halfplane $\sigma>\sigma_0$, and hence that the domain of convergence is a halfplane. More precisely, we have

Corollary 3.7. *Any Dirichlet series $\alpha(s)=\sum_{n=1}^\infty a_n n^{-s}$ has an* ***abscissa of convergence*** *$\sigma_c\in\mathbb{R}\cup\{\pm\infty\}$ with the property that $\alpha(s)$ converges for all s with $\sigma>\sigma_c$, and for no s with $\sigma<\sigma_c$. Furthermore $\alpha(s)$ converges uniformly in any compact subset of $\{s\,:\,\sigma>\sigma_c\}$.*

In extreme cases a Dirichlet series may converge throughout the plane ($\sigma_c=-\infty$), or nowhere ($\sigma_c=+\infty$). When the abscissa of convergence is finite, the series may converge everywhere on the line σ_c+it, it may converge at some but not all points on this line, or nowhere on the line.

Proof of Theorem 3.6. The fact that $\sum_{n=1}^{\infty} a_n n^{-s_0}$ is convergent means that for any $\varepsilon > 0$ there exists some $A_0 > 1$ such that the partial sum

$$f_A(x) := \sum_{A<n\le x} a_n n^{-s_0} \tag{101}$$

satisfies $\big|f_A(x)\big| < \varepsilon$ whenever $A_0 \le A \le x$. We now study similar partial sums for $\alpha(s) = \sum_{n=1}^{\infty} a_n n^s$, $s \in \mathcal{S} \setminus \{s_0\}$: By integration by parts (Lemma 3.2) we have for any $1 \le A < B$:

$$\begin{aligned}\sum_{A<n\le B} a_n n^{-s} &= \sum_{A<n\le B} a_n n^{-s_0} n^{s_0-s} = \int_A^B x^{s_0-s}\, df_A(x)\\ &= \Big(f_A(B)B^{s_0-s} - f_A(A)A^{s_0-s}\Big) - \int_A^B f_A(x)(s_0-s)x^{s_0-s-1}\,dx\\ &= f_A(B)B^{s_0-s} + (s-s_0)\int_A^B f_A(x)x^{s_0-s-1}\,dx.\end{aligned}$$

Hence if $A \ge A_0$ and $s \in \mathcal{S} \setminus \{s_0\}$ (see (100))

$$\begin{aligned}\Big|\sum_{A<n\le B} a_n n^{-s}\Big| &\le \big|f_A(B)B^{s_0-s}\big| + \big|s-s_0\big|\int_A^B \big|f_A(x)x^{s_0-s-1}\big|\,dx\\ &\le \varepsilon B^{\sigma_0-\sigma} + \varepsilon\big|s-s_0\big|\int_A^B x^{\sigma_0-\sigma-1}\,dx\\ &= \varepsilon B^{\sigma_0-\sigma} + \varepsilon\big|s-s_0\big|\Big(\frac{A^{\sigma_0-\sigma}}{\sigma-\sigma_0} - \frac{B^{\sigma_0-\sigma}}{\sigma-\sigma_0}\Big).\end{aligned} \tag{102}$$

Note here that $s \in \mathcal{S} \setminus \{s_0\}$ implies $\sigma - \sigma_0 > 0$ and also

$$\frac{|s-s_0|}{\sigma-\sigma_0} \le \frac{\sigma-\sigma_0+|t-t_0|}{\sigma-\sigma_0} \le 1 + \frac{|t-t_0|}{\sigma-\sigma_0} \le H+1. \tag{103}$$

Hence we can continue as follows:

$$\Big|\sum_{A<n\le B} a_n n^{-s}\Big| \le \varepsilon B^{\sigma_0-\sigma} + \varepsilon(H+1)(A^{\sigma_0-\sigma} - B^{\sigma_0-\sigma}) \le \varepsilon(H+1)A^{\sigma_0-\sigma} \le \varepsilon(H+1).$$

The last number can be made arbitrarily small by chosing $A_0 > 1$ sufficiently large, and note that this bound holds for all $B > A \ge A_0$ and all $s \in \mathcal{S} \setminus \{s_0\}$. This proves that $\alpha(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ is uniformly convergent for all $s \in \mathcal{S} \setminus \{s_0\}$. Since also $\alpha(s_0)$ is convergent (by assumption) it follows that $\alpha(s)$ is in fact uniformly convergent for all $s \in \mathcal{S}$. □

Proof of Corollary 3.7. Let M be the set of those $s \in \mathbb{C}$ for which $\alpha(s)$ converges. If $M = \emptyset$ then the claim holds with $\sigma_c = +\infty$. Otherwise, if $M \ne \emptyset$, then we set $\sigma_c := \inf\{\operatorname{Re} s \,:\, s \in M\} \in \mathbb{R} \cup \{-\infty\}$. Then by definition of infimum, $\alpha(s)$ does not converge for any

s with $\operatorname{Re} s < \sigma_c$. It remains to prove that if C is a compact subset of $\{s : \sigma > \sigma_c\}$ then $\alpha(s)$ converges uniformly for $s \in C$. To prove this we set $\sigma_1 = \inf\{\operatorname{Re} s : s \in C\}$. Since C is compact there is some $s_1 \in C$ such that $\sigma_1 = \operatorname{Re} s_1$. Hence $\sigma_1 > \sigma_c$, and by definition of infimum there is some $s_0 \in M$ with $\operatorname{Re} s_0 < \sigma_1$. Now for all $s \in C$ we have $\operatorname{Re} s_0 < \sigma_1 \le \operatorname{Re} s$, and thus

$$C \ni s \mapsto \frac{|\operatorname{Im} s - \operatorname{Im} s_0|}{\operatorname{Re} s - \operatorname{Re} s_0} \in \mathbb{R}_{\ge 0}$$

is a continuous function on C. Since C is compact this function is bounded from above, i.e. there is some $H > 0$ such that $\frac{|\operatorname{Im} s - \operatorname{Im} s_0|}{\operatorname{Re} s - \operatorname{Re} s_0} \le H$ for all $s \in C$. Now if $\mathcal{S}$ is the sector defined in (100) for our s_0, H it follows that $C \subset \mathcal{S}$. Theorem 3.6 says that $\alpha(s)$ is uniformly convergent in $\mathcal{S}$; hence in particular $\alpha(s)$ is uniformly convergent in C. □

Corollary 3.8. *The series $\alpha(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ is an analytic function for $\sigma > \sigma_c$, and we have*

$$\alpha'(s) = -\sum_{n=1}^{\infty} a_n (\log n) n^{-s}$$

with uniform convergence in any compact subset of $\{s : \sigma > \sigma_c\}$.

Proof. This follows from Corollary 3.7 by Weierstrass Theorem, cf. footnote 1 on p. 5. □

To discuss absolute convergence versus conditional convergence we also introduce the following.

Definition 3.5. We define the *abscissa of absolute convergence*, σ_a, of a Dirichlet series $\alpha(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ as the infimum of those σ for which $\sum_{n=1}^{\infty} |a_n| n^{-\sigma} < \infty$. (By Corollary 3.7, σ_a equals the abscissa of convergence of the Dirichlet series $\sum_{n=1}^{\infty} |a_n| n^{-s}$.)

Since $|a_n n^{-s}| = |a_n| n^{-\sigma}$ we immediately see that the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ is absolutely convergent for all s with $\sigma > \sigma_a$, but not for any s with $\sigma < \sigma_a$. Hence if $\sigma_c < \sigma_a$ then $\sum_{n=1}^{\infty} a_n n^{-s}$ is *conditionally convergent* for all s with $\sigma_c < \sigma < \sigma_a$.

Proposition 3.9. *For every Dirichlet series we have $\sigma_c \le \sigma_a \le \sigma_c + 1$.*

Proof. The first inequality is obvious. To prove the second, suppose that $\varepsilon > 0$. Since the series $\sum_{n=1}^{\infty} a_n n^{-\sigma_c - \varepsilon}$ is convergent, the summands tend to 0 and hence $|a_n| \ll n^{\sigma_c + \varepsilon}$ for all $n \ge 1$ (the implied constant may depend on ε and on the given sequence $\{a_n\}$). Hence the series $\sum_{n=1}^{\infty} a_n n^{-\sigma_c - 1 - 2\varepsilon}$ is absolutely convergent by comparison with the series $\sum_{n=1}^{\infty} n^{-1-\varepsilon}$, proving $\sigma_a \le \sigma_c + 1 + 2\varepsilon$. This is true for all $\varepsilon > 0$ and hence we obtain $\sigma_a \le \sigma_c + 1$. □

It is an important fact that the coefficients of a Dirichlet series are uniquely determined from the resulting function:

Proposition 3.10. *If $\sum_{n=1}^{\infty} a_n n^{-s} = \sum_{n=1}^{\infty} b_n n^{-s}$ for all s with $\sigma > \sigma_0$ then $a_n = b_n$ for all $n \geq 1$.*

Proof. We put $c_n = a_n - b_n$, and consider $\sum_{n=1}^{\infty} c_n n^{-s}$. Suppose that $c_n = 0$ for all $n < N$. Since $\sum_{n=1}^{\infty} c_n n^{-\sigma} = 0$ for $\sigma > \sigma_0$ we have $c_N = -N^{\sigma} \sum_{n>N} c_n n^{-\sigma}$. By Proposition 3.9 this sum is absolutely convergent for $\sigma > \sigma_0 + 1$; thus for these σ we get

$$c_N \leq \left| N^{\sigma} \sum_{n>N} c_n n^{-\sigma} \right| = \sum_{n=N+1}^{\infty} |c_n| (N/n)^{\sigma}. \tag{104}$$

Since each term in the last sum is a (non-negative) decreasing function of σ and tends to 0 as $\sigma \to \infty$, we get by a standard argument that the whole sum tends to 0 as $\sigma \to \infty$. Hence $c_N = 0$, and by induction we deduce that this holds for all N.

(Details for the "standard argument" two lines up: Given any $\varepsilon > 0$, since $\sum_{n=N+1}^{\infty} |c_n|(N/n)^{\sigma_0+2}$ converges there is some $N_0 > N$ such that $\sum_{n>N_0} |c_n|(N/n)^{\sigma_0+2} < \varepsilon$. Furthermore, since $\lim_{\sigma\to\infty} |c_n|(N/n)^{\sigma} = 0$ for each $n \in \{N+1, N+2, \ldots, N_0\}$, and there are only finitely many such n's, we can choose $S > \sigma_0 + 2$ such that $|c_n|(N/n)^{\sigma} < \frac{\varepsilon}{N_0 - N}$ for all $\sigma \geq S$ and all $n \in \{N+1, N+2, \ldots, N_0\}$. We then have, for all $\sigma \geq S$: $\sum_{n>N} |c_n|(N/n)^{\sigma} < \sum_{n=N+1}^{N_0} \frac{\varepsilon}{N_0 - N} + \sum_{n=N_0+1}^{\infty} |c_n|(N/n)^{\sigma_0+2} < \varepsilon + \varepsilon = 2\varepsilon$. Since this holds for arbitrarily small ε we obtain the desired convergence.) □

We next express a convergent Dirichlet series as an absolutely convergent integral.

Theorem 3.11. *Let $A(x) = \sum_{n \leq x} a_n$, and let σ_c be the abscissa of convergence of the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$, as before. If $\sigma_c < 0$ then $A(x)$ is a bounded function, and*

$$\sum_{n=1}^{\infty} a_n n^{-s} = s \int_1^{\infty} A(x) x^{-s-1}\, dx \tag{105}$$

for $\sigma > 0$, the integral being absolutely convergent. If $\sigma_c \geq 0$ then

$$\limsup_{x\to\infty} \frac{\log |A(x)|}{\log x} = \sigma_c, \tag{106}$$

and (105) *holds for $\sigma > \sigma_c$, again with the integral being absolutely convergent.*

Proof. By partial summation we have

$$\begin{aligned} \sum_{n=1}^{N} a_n n^{-s} = \int_{1^-}^{N} x^{-s}\, dA(x) = \int_{\frac{1}{2}}^{N} x^{-s}\, dA(x) &= A(N)N^{-s} + s \int_{\frac{1}{2}}^{N} A(x) x^{-s-1}\, dx \\ &= A(N)N^{-s} + s \int_1^{N} A(x) x^{-s-1}\, dx, \end{aligned} \tag{107}$$

since $A(x) = 0$ for $x < 1$. Note that for any number $\theta > \limsup_{x\to\infty} \frac{\log|A(x)|}{\log x}$ we have $A(x) \ll x^\theta$ for all $x \geq 1$,[5] where the implied constant may depend on the a_n and on θ (but of course not on x). Hence if $\operatorname{Re} s > \theta$ then the integral in (105) is absolutely convergent, and we obtain the equality (105) by letting $N \to \infty$ in (107), since $\lim_{N\to\infty} A(N)N^{-s} = 0$.

Hence it follows that (105) holds for all s with $\operatorname{Re} s > \limsup_{x\to\infty} \frac{\log|A(x)|}{\log x}$, with the integral being absolutely convergent.

First assume $\sigma_c < 0$. Then by Corollary 3.7 we know that $\sum_{n=1}^\infty a_n n^{-s}$ converges for $s = 0$; in other words $A(x)$ tends to a finite limit as $x \to \infty$. This implies $\limsup_{x\to\infty} \frac{\log|A(x)|}{\log x} \leq 0$, and hence (107) holds for all s with $\operatorname{Re} s > 0$.

We now turn to the remaining case, $\sigma_c \geq 0$. By Corollary 3.7 we know that the left hand side in (105) diverges when $\operatorname{Re} s < \sigma_c$; hence $\limsup_{x\to\infty} \frac{\log|A(x)|}{\log x} \geq \sigma_c$. On the other hand, taking any real $s_0 > \sigma_c$ say, we know that the series $f(x) = \sum_{1\leq n\leq x} a_n n^{-s_0}$ tends to a finite limit as $x \to \infty$, and by partial summation we have

$$A(x) = \sum_{1\leq n\leq x} a_n n^{-s_0} n^{s_0} = \int_{\frac{1}{2}}^x u^{s_0}\, df(u) = f(x)x^{s_0} - s_0 \int_{\frac{1}{2}}^x f(u)u^{s_0-1}\, du. \tag{108}$$

Since $f(x)$ is a bounded function it follows that $A(x) \ll x^{s_0}$ for all $x \geq 1$, where the implied constant may depend on the a_n and on s_0. Hence $\limsup_{x\to\infty} \frac{\log|A(x)|}{\log x} \leq s_0$. Since this holds for any real $s_0 > \sigma_c$ we conclude that $\limsup_{x\to\infty} \frac{\log|A(x)|}{\log x} \leq \sigma_c$. Hence we have proved that (106) holds. □

We highlight the following important special case of Theorem 3.11.

Corollary 3.12. *If $A(x) = \sum_{n\leq x} a_n$ is a bounded function then $\sigma_c \leq 0$ and the formula (105) holds for all s with $\sigma > 0$, the integral being absolutely convergent.*

Proof. If $A(x)$ is bounded then $\limsup_{x\to\infty} \frac{\log|A(x)|}{\log x} \leq 0$, and hence the claims follow from Theorem 3.11. □

Example 3.5. Using Corollary 3.12 we can now prove the fact used in the proof of Dirichlet's theorem (Theorem 1.4) that *if χ is a non-principal Dirichlet character, then the series for the Dirichlet L-function $L(s,\chi) = \sum_{n=1}^\infty \chi(n)n^{-s}$ is convergent for all $\sigma > 0$ (and uniformly convergent in any compact subset of $\{s \in \mathbb{C} : \sigma > 0\}$).*

Let $A(x) = \sum_{1\leq n\leq x} \chi(n)$. By Corollary 3.12 we will be done if we can prove that the function $A(x)$ is bounded! To do this we will first prove the following lemma:

[5]The symbol "$\ll$": "$b \ll a$" means that there is a constant $C > 0$ (called the *implied constant*) such that $b \leq Ca$; *we will only use this notation when both $a \geq 0$ and $b \geq 0$!* Thus, recalling the "big O"-notation (cf. footnote 3 on page 13), we may note that $b = O(a)$ is equivalent to $|b| \ll a$. In this vein we also introduce the symbol "$\gg$": "$b \gg a$" is (again) only used when both $a \geq 0$ and $b \geq 0$, and it denotes that there is a constant $C > 0$ such that $b \geq Ca$.

Lemma 3.13. *If χ is a non-principal Dirichlet character modulo q then $\sum_{n=1}^{q}\chi(n)=0$.*

Proof. Since χ is periodic with period q we may just as well consider χ as a function from $\mathbb{Z}/q\mathbb{Z}$ to $\mathbb{C}$, and then our task is to prove $\sum_{n\in\mathbb{Z}/q\mathbb{Z}}\chi(n)=0$. Also $\chi(n)=0$ whenever $(n,q)=1$, thus $\chi(n)\neq 0$ can only hold when $n\in(\mathbb{Z}/q\mathbb{Z})^\times$, and hence our task is to prove $\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(n)=0$. Since χ is non-principal there is some $m\in(\mathbb{Z}/q\mathbb{Z})^\times$ with $\chi(m)\neq 1$. Since χ is multiplicative without restrictions we have

$$\chi(m)\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(n)=\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(m)\chi(n)=\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(mn).$$

But here mn visits each element in $(\mathbb{Z}/q\mathbb{Z})^\times$ exactly once; hence we conclude

$$\chi(m)\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(n)=\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(n).$$

Since $\chi(m)\neq 1$ this implies $\sum_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}\chi(n)=0$. □

The lemma implies that $A(q)=\sum_{1\leq n\leq q}\chi(n)=0$. Hence since χ has period q we obtain $A(2q)=A(3q)=\ldots=0$ and also $A(x+q)=A(x)$ for all $x\geq 1$. Hence $A(x)$ is bounded. This concludes the proof that $L(s,\chi)=\sum_{n=1}^{\infty}\chi(n)n^{-s}$ has abscissa of convergence $\sigma_c\leq 0$. □ □

Example 3.6. Using Corollary 3.12 together with a trick we can now also prove the fact used in the proof of Dirichlet's theorem (Theorem 1.4), that $\zeta(s)$ *has a meromorphic continuation to $\sigma>0$ with one simple pole at $s=1$.*

First of all note that Theorem 3.11 easily implies that $\zeta(s)=\sum_{n=1}^{\infty}n^{-s}$ has abscissa of convergence exactly $\sigma_c=1$; hence we need to *manipulate* the series $\sum_{n=1}^{\infty}n^{-s}$ in some way before we can obtain the meromorphic continuation.

Note that for any s with $\sigma>1$ we have

$$\begin{aligned}
&1^{-s}-2^{-s}+3^{-s}-4^{-s}+5^{-s}-\ldots\\
&=\big(1^{-s}+2^{-s}+3^{-s}+4^{-s}+5^{-s}+\ldots\big)-2\big(2^{-s}+4^{-s}+6^{-s}+\ldots\big)\\
&=\big(1^{-s}+2^{-s}+3^{-s}+4^{-s}+5^{-s}+\ldots\big)-2\cdot 2^{-s}\big(1^{-s}+2^{-s}+3^{-s}+4^{-s}+\ldots\big)\\
&=\zeta(s)-2\cdot 2^{-s}\zeta(s)\\
&=(1-2^{1-s})\zeta(s).
\end{aligned}$$

Hence:

$$L(s):=\sum_{n=1}^{\infty}(-1)^{n+1}n^{-s}=1^{-s}-2^{-s}+3^{-s}-4^{-s}+5^{-s}-\ldots=(1-2^{1-s})\zeta(s).$$

(This identity can alternatively be deduced from Proposition 2.7 with the multiplicative function $f(n) = (-1)^{n+1}n^{-s}$.) But the Dirichlet series $L(s)$ has abscissa of convergence $\sigma_c \leq 0$ by Corollary 3.12, and in particular $L(s)$ is an analytic function for $\sigma > 0$.

Hence

$$\zeta(s) = \frac{L(s)}{1 - 2^{1-s}},$$

a meromorphic function for $\sigma > 0$! To see that $\zeta(s)$ has a simple pole at $s = 1$ we need now only note that $L(1) \neq 0$ (since $L(1) = (1^{-1} - 2^{-1}) + (3^{-1} - 4^{-1}) + \ldots > 0$) and that the denominator $f(s) = 1 - 2^{1-s}$ has a simple zero at $s = 1$ (since $f(1) = 0$ but $f'(1) = \log 2 \neq 0$). □ □

In fact we will see later that much more is true: ζ has a meromorphic continuation to *the whole complex plane*, and the *only* pole is at $s = 1$!

Example 3.7. Applying Corollary 3.8 and Theorem 3.11 to the formula for $\log L(s, \chi)$ which we proved in the last lecture, leads to a very important formula (it plays a key role in the first proof of the prime number theorem which we will give). Recall from (22) (cf. also Example 2.2) that we have

$$\log L(s, \chi) = \sum_p \sum_{m=1}^{\infty} m^{-1} \chi(p^m) p^{-ms} \qquad (\sigma > 1), \tag{109}$$

with absolute convergence of the double sum for any s with $\sigma > 1$. The right hand side is really a Dirichlet series:

$$\log L(s, \chi) = \sum_{n=1}^{\infty} a_n n^{-s} \qquad (\sigma > 1), \tag{110}$$

where

$$a_n = \chi(n) \begin{cases} m^{-1} & \text{if } n = p^m \\ 0 & \text{otherwise,} \end{cases} \tag{111}$$

and as we have noted this Dirichlet series has abscissa of convergence $\sigma_c \leq 1$. Hence we can apply Corollary 3.8 to get hold of the derivative! Note here that

$$a_n \log n = \chi(n)\Lambda(n), \text{ where } \Lambda(n) := \begin{cases} \log p & \text{if } n = p^m \\ 0 & \text{otherwise.} \end{cases} \tag{112}$$

Hence we get:

$$\frac{L'(s, \chi)}{L(s, \chi)} = -\sum_{n=1}^{\infty} \chi(n)\Lambda(n) n^{-s} \qquad (\sigma > 1). \tag{113}$$

(Justification: Recall that we really *defined* $\log L(s,\chi)$ by the formula (109), but we proved in Example 2.2 that this gives a branch of the logarithm of $L(s,\chi)$, and hence the derivative of $\log L(s,\chi)$ is indeed $\frac{L'(s,\chi)}{L(s,\chi)}$.) Finally, by Theorem 3.11, (113) can be expressed as

$$\frac{L'(s,\chi)}{L(s,\chi)} = -s\int_1^\infty \psi(x,\chi)x^{-s-1}\,dx \qquad (\sigma>1), \tag{114}$$

where

$$\psi(x,\chi) = \sum_{n\le x}\chi(n)\Lambda(n). \tag{115}$$

(the function $\psi(x,\chi)$ will play a very important role in our further developments, cf. Definition 14.1 below).

In the special case $\chi = 1$ ($q=1$) the last formula reads

$$\frac{\zeta'(s)}{\zeta(s)} = -s\int_1^\infty \psi(x)x^{-s-1}\,dx \qquad (\sigma>1), \tag{116}$$

where

$$\psi(x) = \sum_{n\le x}\Lambda(n). \tag{117}$$

3.4. **Problems.**

* *Problem* 3.1. Give proofs of (some or all of) the results stated without proof in §3.2.

Problem 3.2. Let $0<\beta<1$. Suppose that $a_1, a_2, \dots$ are complex numbers such that $\sum_{n=1}^N a_n n^\beta = N + O(N^{\frac12})$ as $N\to\infty$. Then prove that

$$\sum_{n=1}^N a_n = \frac{1}{1-\beta}N^{1-\beta} + \left\{\begin{matrix} O(N^{\frac12-\beta}) & \text{if } \beta<\frac12 \\ O(\log N) & \text{if } \beta=\frac12 \\ O(1) & \text{if } \beta>\frac12 \end{matrix}\right\} \qquad \text{as } N\to\infty. \tag{118}$$

Problem 3.3. Suppose that $a_1, a_2, \dots$ are complex numbers such that $\sum_{n=1}^N a_n n^{-\frac12} \sim N$ as $N\to\infty$. [6] Then prove that $\sum_{n=1}^N a_n \sim \frac23 N^{\frac32}$ as $N\to\infty$. (Hint: Try to carry over the method from Example 3.4.)

Problem 3.4. Let $\rho_1, \rho_2, \dots$ be a sequence of complex numbers. Set $N(r) := \#\{j : |\rho_j| \le r\}$ and

$$A = \limsup_{r\to\infty}\frac{\log N(r)}{\log r}. \tag{119}$$

[6] The "$\sim$" symbol: We write "$f(x)\sim g(x)$ as $x\to a$" to denote that $\lim_{x\to a}\frac{f(x)}{g(x)} = 1$. Thus this notation can only be used when $g(x)\neq 0$ for all x sufficiently near a. [And we may note that *if* $g(x)\neq 0$ for all x sufficiently near a, then "$f(x)\sim g(x)$ as $x\to a$" is equivalent with "$f(x) = g(x)(1+o(1))$ as $x\to a$".]

Also set

$$\tau = \inf\big\{\alpha > 0 \,:\, \sum_{j=1}^{\infty}(1+|\rho_j|)^{-\alpha} < \infty\big\}. \tag{120}$$

(a). Prove that $\tau \le A$. [Hint. Note that $\sum_{j=1}^{\infty}(1+|\rho_j|)^{-\alpha}$ can be written as a Riemann-Stieltjes integral involving $N(r)$. Also note/show that for any $A_1 > A$ we have $N(r) \ll r^{A_1}$ as $r \to \infty$. From these facts show that $\sum_{j=1}^{\infty}(1+|\rho_j|)^{-\alpha}$ converges if $\alpha > A_1$.]
(b). Prove that $A \le \tau$, i.e. $A = \tau$.

Problem 3.5. Suppose that $a_1, a_2, \dots$ are complex numbers such that $\sum_{n=1}^{N} a_n \sim N^2$ as $N \to \infty$. Then prove that $\sum_{n=1}^{N} a_n(N-n)^2 \sim \frac{1}{6}N^4$ as $N \to \infty$. (Hint: Compare Problem 3.3.)

Problem 3.6. The following is a step in some proofs of the prime number theorem: Write $\vartheta(x) = \sum_{p \le x} \log p$ (i.e. sum over all prime numbers p with $p \le x$). Prove that if $\vartheta(x) \sim x$ as $x \to \infty$ then $\pi(x) \sim \frac{x}{\log x}$ as $x \to \infty$. (Hint: Compare Example 3.4 and Problem 3.3.)

Problem 3.7. **Möbius inversion formula.**
(a). Let $a_1, a_2, \dots$ be complex numbers such that the Dirichlet series $\alpha(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ has abscissa of convergence $< \infty$. Set $b_n = \sum_{d|n} a_d$ for $n = 1, 2, \dots$. Then prove that $\beta(s) = \sum_{n=1}^{\infty} b_n n^{-s}$ has abscissa of convergence $< \infty$, and $\beta(s) = \zeta(s)\alpha(s)$ for σ sufficiently large.
(b). Combining (a) with Problem 2.1 and Proposition 3.10, prove that if $\{a_n\}$ and $\{b_n\}$ are as in (a) then $a_n = \sum_{d|n} \mu(n/d) b_d$ for all $n = 1, 2, \dots$.
(c). Remove the assumption that $\sum_{n=1}^{\infty} a_n n^{-s}$ has abscissa of convergence $< \infty$, i.e. prove that for *all* sequences $\{a_n\}$ of complex numbers, if we set $b_n = \sum_{d|n} a_d$ for $n = 1, 2, \dots$, then $a_n = \sum_{d|n} \mu(n/d) b_d$ for all $n = 1, 2, \dots$.

Problem 3.8. Prove that $\phi(n) = \sum_{d|n} \mu(\frac{n}{d}) d$ for all $n \in \mathbb{Z}^+$,
(a). using Dirichlet series;
(b). directly.

Problem 3.9. (a). Let $d(n)$ be the number of divisors of n, for each $n \in \mathbb{Z}^+$. Prove that $\sum_{n=1}^{\infty} d(n) n^{-s} = \zeta(s)^2$ when $\sigma > 1$.
(b). (Generalization of (a).) For any $\alpha \in \mathbb{C}$ we set $\sigma_\alpha(n) = \sum_{d|n} d^\alpha$. Prove that $\sum_{n=1}^{\infty} \sigma_\alpha(n) n^{-s} = \zeta(s)\zeta(s-\alpha)$ when $\sigma > \max(1, 1 + \operatorname{Re}\alpha)$.

Problem 3.10. Let $a_1, a_2, \dots$ be an arbitrary sequence of complex numbers, and set

$$b_n = \sum_{\substack{d|n \\ d \text{ squarefree}}} a_{n/d}, \qquad n = 1, 2, \dots. \tag{121}$$

Find a formula for a_n in terms of $b_1, b_2, \dots$. [Hint. Show that (121) may be written as $b_n = \sum_{d|n} |\mu(n/d)| a_d$; then try to mimic Problem 3.7, making use of Problem 2.3.]

Problem 3.11. **Euler product of degree 2, again** (cf. Problem 2.4). Let $\alpha \in \mathbb{C}$ be a given constant and let $\{a_n\}$ be a sequence of complex numbers. Prove that the following two assertions are equivalent:
(i). The sequence $\{a_n\}$ is multiplicative, not identically zero, satisfies $|a_n| \ll n^A$ for all $n \in \mathbb{Z}^+$ and some constant $A \in \mathbb{R}$; and for every prime p and every $k \geq 1$ we have

$$a_p a_{p^k} = a_{p^{k+1}} + p^\alpha a_{p^{k-1}}.$$

(ii). There is some $B \in \mathbb{R}$ such that

$$\sum_{n=1}^{\infty} a_n n^{-s} = \prod_p (1 - a_p p^{-s} + p^\alpha p^{-2s})^{-1}$$

holds for all $s \in \mathbb{C}$ with $\sigma > B$ (in particular both the sum and the product converge when $\sigma > B$).

Remark 3.3. From the relations in (i) above one can derive the following general multiplication formula:

$$a_m a_n = \sum_{d|(m,n)} d^\alpha a_{mn/d^2}, \qquad \forall m, n \in \mathbb{Z}^+. \tag{122}$$

(You may like to prove this as an exercise, but I don't think this has much to do with the methods introduced in the present section...)

Remark 3.4. Note that for any $\alpha \in \mathbb{C}$, the sequence $a_n = \sigma_\alpha(n)$ satisfies conditions (i)$\Longleftrightarrow$(ii) above, cf. Problem 3.9(b). However, as already mentioned, similar types of L-functions also arise from more advanced sources, e.g. modular forms, elliptic curves and Galois representations.

Problem 3.12. (In connection with Theorem 3.11.) Prove that for any continuous function $A : [1, \infty) \to \mathbb{C}$ we have

$$\limsup_{x\to\infty} \frac{\log |A(x)|}{\log x} = \inf\big\{\theta \in \mathbb{R} \,:\, |A(x)| \ll x^\theta,\ \forall x \geq 1\big\}.$$

4. More on Dirichlet characters

This lecture corresponds to Davenport, Chapters 4,5. Below we review several things from more basic number theory which I hope that most of you are aware of, and which I will not have time to discuss in class. In particular this concerns the structure of the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^\times$ and quadratic reciprocity. Here are some recommended textbook references, very different in style, but each containing the necessary material:
• Nagell, "Introduction to Number Theory", [50], chapters III and IV.
• Ireland and Rosen, "A Classical Introduction to Modern Number Theory", [35], chapters 3, 4, 5.
• Niven, Zuckerman and Montgomery, "An Introduction to the Theory of Numbers", [53], chapters 2, 3.

4.1. **[Review: Some basic facts about $\mathbb{Z}/q\mathbb{Z}$ and $(\mathbb{Z}/q\mathbb{Z})^\times$].** For $q \in \mathbb{Z}^+$ we denote by $\mathbb{Z}/q\mathbb{Z}$ the ring of all integer residue classes modulo q. This is a finite ring with q elements, it has a multiplicative identity $1 \in \mathbb{Z}/q\mathbb{Z}$, and it is commutative. For any commutative ring R with multiplicative identity $1 \in R$, we denote by $R^\times$ the group of invertible elements of R. Thus as a set,

$$R^\times = \{a \in R \,:\, \exists x \in R : ax = 1\}, \tag{123}$$

and $R^\times$ is a group under the multiplication coming from R. Returning to $R = \mathbb{Z}/q\mathbb{Z}$, it is a well-known fact that *an integer x gives an element in $(\mathbb{Z}/q\mathbb{Z})^\times$ if and only if $(x, q) = 1$.* The number of elements of $(\mathbb{Z}/q\mathbb{Z})^\times$ equals $\phi(q)$, where $\phi : \mathbb{Z}^+ \to \mathbb{Z}^+$ is *Euler's ϕ-function.* It is given by the formula

$$\phi(q) = q \prod_{p|q} \Big(1 - \frac{1}{p}\Big). \tag{124}$$

See Example 4.2 below for a proof.

Theorem 4.1. ***Euler's Theorem (or Fermat-Euler Theorem).*** *For each integer x with $(x, q) = 1$ we have $x^{\phi(q)} \equiv 1 \,(mod\ q)$.*

Proof. The assumption $(x, q) = 1$ means that x gives an element in $(\mathbb{Z}/q\mathbb{Z})^\times$. But $(\mathbb{Z}/q\mathbb{Z})^\times$ is a finite group with $\phi(q)$ elements; hence by Lagrange's Theorem, $g^{\phi(q)} = 1$ holds for all $g \in (\mathbb{Z}/q\mathbb{Z})^\times$. □

Theorem 4.2. ***Chinese Remainder Theorem.*** *Let $q_1, q_2, \ldots, q_m$ be positive integers which are pairwise coprime, i.e. $(q_j, q_k) = 1$ for all $1 \le j < k \le m$. Then for any given $x_1, \ldots, x_m \in \mathbb{Z}$ there is some $x \in \mathbb{Z}$ such that $x \equiv x_j \,(mod\ q_j)$ for all $j = 1, \ldots, m$. This x is uniquely determined modulo $q_1 q_2 \cdots q_m$.*

(See almost any book on basic number theory for a proof.)

Corollary 4.3. *If $q = q_1 q_2 \cdots q_m$ where $q_1, q_2, \ldots, q_m$ are positive integers which are pairwise coprime, then there is a canonical isomorphism between the ring $\mathbb{Z}/q\mathbb{Z}$ and the ring $(\mathbb{Z}/q_1\mathbb{Z}) \times (\mathbb{Z}/q_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})$.*

Proof. (Sketch) For each j, since $q_j \mid q$, every congruence class modulo q determines a congruence class modulo q_j. Thus we have an (obvious) homomorphism $J : \mathbb{Z}/q\mathbb{Z} \to (\mathbb{Z}/q_1\mathbb{Z}) \times (\mathbb{Z}/q_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})$. The Chinese Remainder Theorem gives that J is surjective (from the existence part of the Chinese Remainder Theorem) and also injective (from the uniqueness part of the Chinese Remainder Theorem). □

Corollary 4.4. *If $q = q_1 q_2 \cdots q_m$ where $q_1, q_2, \ldots, q_m$ are positive integers which are pairwise coprime, then there is a canonical isomorphism between the group $(\mathbb{Z}/q\mathbb{Z})^\times$ and the group $(\mathbb{Z}/q_1\mathbb{Z})^\times \times (\mathbb{Z}/q_2\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})^\times$.*

Proof. Let $J : \mathbb{Z}/q\mathbb{Z} \to (\mathbb{Z}/q_1\mathbb{Z}) \times (\mathbb{Z}/q_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})$ be the canonical ring isomorphism from above. J restricts to an isomorphism between the groups of invertible elements,

$$J^\times : (\mathbb{Z}/q\mathbb{Z})^\times \to \big((\mathbb{Z}/q_1\mathbb{Z}) \times (\mathbb{Z}/q_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})\big)^\times.$$

But here

$$\big((\mathbb{Z}/q_1\mathbb{Z}) \times (\mathbb{Z}/q_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})\big)^\times = (\mathbb{Z}/q_1\mathbb{Z})^\times \times (\mathbb{Z}/q_2\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/q_m\mathbb{Z})^\times,$$

and this completes the proof. □

Example 4.1. If we denote $\mathbb{Z}/10\mathbb{Z} = \{0,1,2,3,4,5,6,7,8,9\}$, $\mathbb{Z}/5\mathbb{Z} = \{0,1,2,3,4\}$ and $\mathbb{Z}/2\mathbb{Z} = \{0,1\}$ then the canonical ring isomorphism $J : (\mathbb{Z}/10\mathbb{Z}) \to (\mathbb{Z}/5\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ maps as follows:

$x =$	0	1	2	3	4	5	6	7	8	9
$J(x) =$	(0,0)	(1,1)	(2,0)	(3,1)	(4,0)	(0,1)	(1,0)	(2,1)	(3,0)	(4,1)

The groups of invertible elements are $(\mathbb{Z}/10\mathbb{Z})^\times = \{1,3,7,9\}$, $(\mathbb{Z}/5\mathbb{Z})^\times = \{1,2,3,4\}$ and $(\mathbb{Z}/2\mathbb{Z})^\times = \{1\}$, with group isomorphism $J^\times : (\mathbb{Z}/10\mathbb{Z})^\times \to (\mathbb{Z}/5\mathbb{Z})^\times \times (\mathbb{Z}/2\mathbb{Z})^\times$:

$x =$	1	3	7	9
$J(x) =$	(1,1)	(3,1)	(2,1)	(4,1)

Example 4.2. Note that Corollary 4.4 immediately implies that *Euler's ϕ-function is multiplicative*. Indeed, if $m, n \in \mathbb{Z}^+$ are coprime, then Corollary 4.4 gives $\#(\mathbb{Z}/nm\mathbb{Z})^\times = \#\big((\mathbb{Z}/n\mathbb{Z})^\times \times (\mathbb{Z}/m\mathbb{Z})^\times\big) = \#(\mathbb{Z}/n\mathbb{Z})^\times \cdot \#(\mathbb{Z}/m\mathbb{Z})^\times$, viz. $\phi(nm) = \phi(n)\phi(m)$.

From this it is easy to prove the formula (124), $\phi(q) = q\prod_{p|q}\big(1 - \frac{1}{p}\big)$. To wit, note that if q is a prime power, $q = p^\alpha$ ($\alpha \geq 1$), then an integer x is coprime with q if and only if $p \nmid x$, and since exactly $p^{\alpha-1}$ among the integers $0, 1, 2, \ldots, p^\alpha - 1$ are divisible by p we have

$$\phi(p^\alpha) = p^\alpha - p^{\alpha-1} = p^\alpha\Big(1 - \frac{1}{p}\Big).$$

Hence by multiplicativity, if q has prime factorization $q = p_1^{\alpha_1} \cdots p_m^{\alpha_m}$ then

$$\phi(q) = \prod_{j=1}^{m} \phi(p_j^{\alpha_j}) = \prod_{j=1}^{m} \Big(p_j^{\alpha_j} \Big(1 - \frac{1}{p_j}\Big)\Big) = q \prod_{j=1}^{m} \Big(1 - \frac{1}{p_j}\Big).$$

□

4.2. **[Review: The structure of $(\mathbb{Z}/q\mathbb{Z})^\times$].** In this section we will describe the structure of $(\mathbb{Z}/q\mathbb{Z})^\times$ as an abstract group.

Definition 4.1. Any element $g \in (\mathbb{Z}/q\mathbb{Z})^\times$ such that $(\mathbb{Z}/q\mathbb{Z})^\times = \{g^0, g^1, g^2, \dots, g^{\phi(q)-1}\}$ is called a *primitive root* modulo q.

Thus a primitive root modulo q exists if and only if $(\mathbb{Z}/q\mathbb{Z})^\times$ is a cyclic group, and then a primitive root is the same as a *generator* of this cyclic group.

Let us make the general observation that for any $g \in (\mathbb{Z}/q\mathbb{Z})^\times$, if $g^a \equiv g^b \pmod q$ with $0 \le a < b$, then $g^{b-a} \equiv 1 \pmod q$, since $(\mathbb{Z}/q\mathbb{Z})^\times$ is a group. Hence if v is the smallest positive integer for which $g^v \equiv 1 \pmod q$, then all the elements $1 = g^0, g^1, g^2, \dots, g^{v-1}$ are *distinct* in $(\mathbb{Z}/q\mathbb{Z})^\times$. In particular g is a primitive root if and only if $v = \phi(q)$.

Our first result is that when $q = p$ a prime, $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic:

Lemma 4.5. *If p is an odd prime then there is a primitive root modulo p.*

Proof. For each $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ we let $v(a)$ be the smallest positive integer such that $a^{v(a)} = 1$; since $a^{p-1} = 1$ by Theorem 4.1 we have $v(a) \mid p-1$. For each $v \mid p-1$ we set

$$n(v) := \#\{a \in (\mathbb{Z}/p\mathbb{Z})^\times \,:\, v(a) = v\}.$$

Since every $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ is a zero of the polynomial $X^{p-1} - 1$ over the field $\mathbb{Z}/p\mathbb{Z}$, and $\deg(X^{p-1} - 1) = \#(\mathbb{Z}/p\mathbb{Z})^\times = p - 1$, we conclude by polynom division that

$$X^{p-1} - 1 = \prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} (X - a)$$

as a polynomial identity over $\mathbb{Z}/p\mathbb{Z}$. Now if d is any divisor of $p-1$ then

$$X^{p-1} - 1 = (X^d - 1)(X^{(e-1)d} + X^{(e-2)d} + \ldots + 1)$$

where $e = \frac{p-1}{d}$, and hence exactly d among the factors in $\prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times}(X - a)$ divide $X^d - 1$ and all the other factors divide $X^{(e-1)d} + X^{(e-2)d} + \ldots + 1$. Hence $X^d - 1$ has exactly d zeros in $(\mathbb{Z}/p\mathbb{Z})^\times$. On the other hand $a^d \equiv 1 \pmod p$ holds if and only if $v(a) \mid d$; thus

$$\sum_{v \mid d} n(v) = d, \qquad \forall d \mid p-1.$$

Hence by Möbius inversion, for each $d \mid p-1$ we have (cf. Problems 3.7 and 3.8)

$$n(d) = \sum_{v|d} \mu\Big(\frac{d}{v}\Big) v = \phi(d).$$

(Here we applied Möbius inversion to the "sequence" $\{n(v)\}$, although $n(v)$ is only defined for $v \mid p-1$. However, we may set e.g. $n(v) = 0$ for all other v, and apply Möbius inversion to *this* sequence $n(1), n(2), n(3), \ldots$; compare the solution to Problem 3.7(c).) In particular $n(p-1) = \phi(p-1)$, i.e. we have proved that there are exactly $\phi(p-1) > 0$ primitive roots modulo p. □

Next we will prove that if p is an odd prime then also $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is cyclic for any $\alpha \geq 2$.

Lemma 4.6. *If p is an odd prime and $\alpha \geq 1$ then there is a primitive root modulo p^α.*

Proof. Let $g \in \mathbb{Z}$ be a primitive root modulo p (this exists by Lemma 4.5). Then $g^{p-1} \equiv 1 \pmod p$ and hence $g^{p-1} = ap + 1$ for some $a \in \mathbb{Z}$. *If* $p \mid a$, then we set $g' = g + p$ and note that g' is also a primitive root modulo p, since $g' \equiv g \pmod p$, and also

$$g'^{p-1} \equiv (p+1)^{p-1} = \sum_{k=0}^{p-1} \binom{p-1}{k} p^k \equiv \sum_{k=0}^{1} \binom{p-1}{k} p^k \equiv 1 - p \pmod{p^2}.$$

Hence by possibly replacing g by g', we may from now on assume that $g \in \mathbb{Z}$ is a primitive root modulo p and $g^{p-1} = ap + 1$, $p \nmid a$.

We now claim that this *g is a primitive root modulo p^α for all $\alpha \geq 1$*. We will prove this by induction over α; in fact we will prove the more precise statement that *for each $\alpha \geq 1$, g is a primitive root modulo p^α and $g^{(p-1)p^{\alpha-1}} = ap^\alpha + 1$ for some $a \in \mathbb{Z}$ with $p \nmid a$*. This statement is true for $\alpha = 1$ as we have just noted. Next assume that $\alpha \geq 2$ and that the statement is true with α replaced by $\alpha - 1$. Then since g is a primitive root modulo $p^{\alpha-1}$, for any $v \in \mathbb{Z}$ which is not divisible with $\phi(p^{\alpha-1}) = p^{\alpha-2}(p-1)$ we have $g^v \not\equiv 1 \pmod{p^{\alpha-1}}$ and thus $g^v \not\equiv 1 \pmod{p^\alpha}$. Hence if $v > 0$ is such that $g^v \equiv 1 \pmod{p^\alpha}$ then $v = np^{\alpha-2}(p-1)$ for some $n > 0$, and now since we are also assuming $g^{(p-1)p^{\alpha-2}} = ap^{\alpha-1} + 1$, we have

(125)

$$g^v = (ap^{\alpha-1}+1)^n = \sum_{k=0}^{n} \binom{n}{k} a^k p^{k(\alpha-1)} \equiv \sum_{k=0}^{1} \binom{n}{k} a^k p^{k(\alpha-1)} = nap^{\alpha-1} + 1 \pmod{p^\alpha}.$$

Since $p \nmid a$ the above is $\equiv 1 \pmod{p^\alpha}$ if and only if $p \mid n$. Hence $g^v \equiv 1 \pmod{p^\alpha}$ holds if and only if v is divisible by $p^{\alpha-1}(p-1) = \phi(p^\alpha)$, and thus *$g$ is a primitive root modulo p^α*. Also, in the special case $n = p$ the computation in (125) actually holds *modulo $p^{\alpha+1}$*, since $p^{k(\alpha-1)} \equiv 0 \pmod{p^{\alpha+1}}$ for all $k \geq 3$ and also $\binom{p}{k} p^{k(\alpha-1)} \equiv 0 \pmod{p^{\alpha+1}}$ for $k = 2$ (here we use the fact that $p \neq 2$; thus $p \mid \binom{p}{2}$). Hence we have $g^{(p-1)p^{\alpha-1}} = a'p^\alpha + 1$ *for some $a' \in \mathbb{Z}$ with $p \nmid a'$*. This concludes the induction step, and thus the proof of the lemma. □

The following observation is also important:

Lemma 4.7. *If $g \in \mathbb{Z}$ is a primitive root modulo p^α (some $\alpha \geq 2$) then g is also a primitive root modulo p^k for each $1 \leq k \leq \alpha - 1$.*

Proof. In the proof of Lemma 4.6 we constructed an integer – let's call it g_0 here – which is a primitive root modulo p^β for *all* $\beta \geq 1$. In particular g_0 is a primitive root modulo p^α and hence there is some $v \geq 1$ such that $g \equiv g_0^v \pmod{p^\alpha}$. Set $d = (v, (p-1)p^{\alpha-1})$. If $d > 1$ then $g^{(p-1)p^{\alpha-1}/d} \equiv g_0^{v(p-1)p^{\alpha-1}/d} \equiv 1^{v/d} = 1 \pmod{p^\alpha}$, and since $1 \leq (p-1)p^{\alpha-1}/d < (p-1)p^{\alpha-1}$ this contradicts the fact that g is a primitive root modulo p^α. Thus we must have $(v, (p-1)p^{\alpha-1}) = d = 1$.

Now take any $1 \leq k \leq \alpha - 1$. Then also $(v, (p-1)p^{k-1}) = 1$ and thus for every integer w we have $vw \equiv 0 \bmod (p-1)p^{k-1}$ if and only if $w \equiv 0 \bmod (p-1)p^{k-1}$. But g_0 is a primitive root modulo p^k and thus for an integer $n \geq 0$ we have $g_0^n \equiv 1 \pmod{p^k}$ if and only if $n \equiv 0 \bmod (p-1)p^{k-1}$. Hence, for $w \geq 0$ we have the following chain of equivalent statements:

$$\begin{aligned} g^w \equiv 1 \pmod{p^k} \iff g_0^{vw} \equiv 1 \pmod{p^k} &\iff vw \equiv 0 \bmod (p-1)p^{k-1} \\ &\iff w \equiv 0 \bmod (p-1)p^{k-1}. \end{aligned}$$

Hence g is a primitive root modulo p^k. □

We next turn to the case $p = 2$. Here $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$ is cyclic for $\alpha = 1, 2$ (of order 1 and 2 respectively), but it turns out to be non-cyclic for $\alpha \geq 3$. For example, regarding $(\mathbb{Z}/8\mathbb{Z})^\times$, it is well-known that the square of *any* odd number is $\equiv 1 \pmod 8$, and in fact the group $(\mathbb{Z}/8\mathbb{Z})^\times$ is isomorphic to the product of two cyclic groups of order 2. The following lemma shows that, more generally, $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$ for $\alpha \geq 3$ is isomorphic to the product of a cyclic group of order $2^{\alpha-2}$ and a cyclic group of order 2.

Lemma 4.8. *If $q = 2^\alpha$ with $\alpha \geq 2$ then $5^{2^{\alpha-2}} \equiv 1 \pmod{2^\alpha}$; $\pm 5^0, \pm 5^1, \pm 5^2, \ldots, \pm 5^{2^{\alpha-2}-1}$ are distinct elements in $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$, and hence*

$$(\mathbb{Z}/2^\alpha\mathbb{Z})^\times = \big\{(-1)^\nu 5^{\nu'} \,:\, \nu \in \{0,1\},\, \nu' \in \{0, 1, \ldots, 2^{\alpha-2} - 1\}\big\}. \tag{126}$$

Proof. Note that $5^v \equiv 1 \pmod 4$ while $-5^v \equiv 3 \pmod 4$ for all $v \geq 0$; hence it suffices to prove that if we let v be the smallest positive integer for which $5^v \equiv 1 \pmod{2^\alpha}$, then $v = 2^{\alpha-2}$. This is true if $\alpha = 2$; hence from now on we assume $\alpha \geq 3$.

By Theorem 4.1 (note $\phi(2^\alpha) = 2^{\alpha-1}$) we have $5^{2^{\alpha-1}} \equiv 1 \pmod{2^\alpha}$; hence $v \mid 2^{\alpha-1}$. But we cannot have $v = 2^{\alpha-1}$ since then the sequence $5^0, 5^1, 5^2, \ldots, 5^{v-1}$ would contain all elements in $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$, while we saw above that it actually misses all elements which are $\equiv 3 \pmod 4$. Hence $v < 2^{\alpha-1}$ and thus $v \mid 2^{\alpha-2}$. Now $v = 2^{\alpha-2}$ will follow if we can only prove $5^{2^{\alpha-3}} \not\equiv 1 \pmod{2^\alpha}$. In fact we will prove by induction that $5^{2^{\alpha-3}} \equiv 2^{\alpha-1}+1 \pmod{2^\alpha}$

for all $\alpha \geq 3$. This is clearly true for $\alpha = 3$. Now assume that it is true for some given $\alpha \geq 3$; this means that $5^{2^{\alpha-3}} = 2^{\alpha-1}u + 1$ for some *odd* integer u, and hence

$$5^{2^{\alpha-2}} = (2^{\alpha-1}u + 1)^2 = 2^{2(\alpha-1)}u^2 + 2^\alpha u + 1 = 2^\alpha(2^{\alpha-2}u^2 + u) + 1.$$

Here $2^{\alpha-2}u^2 + u$ is odd since $\alpha \geq 3$; hence $5^{2^{\alpha-2}} \equiv 2^\alpha + 1 \pmod{2^{\alpha+1}}$, i.e. the claim is true also for $\alpha + 1$. This concludes the proof. □

Combining Lemma 4.6 and Lemma 4.8 with the Chinese Remainder Theorem (see Cor. 4.4) we obtain the following structure theorem for $(\mathbb{Z}/q\mathbb{Z})^\times$. Let us write $Z(m)$ to denote a cyclic group of order m.

Theorem 4.9. *Let $q \geq 2$ have the prime factorization $q = 2^{\alpha_0}p_1^{\alpha_1}\cdots p_m^{\alpha_m}$ (where $p_1, \ldots, p_m$ are distinct odd primes, $\alpha_1, \ldots, \alpha_m \geq 1$, and $\alpha_0 \geq 0$). Then $(\mathbb{Z}/q\mathbb{Z})^\times$ is isomorphic to the direct product*

$$\begin{Bmatrix} Z(1) & \text{if } \alpha_0 \in \{0,1\} \\ Z(2) \times Z(2^{\alpha_0-2}) & \text{if } \alpha_0 \geq 2 \end{Bmatrix} \times Z\big(p_1^{\alpha_1-1}(p_1-1)\big) \times \cdots \times Z\big(p_m^{\alpha_m-1}(p_m-1)\big).$$

We remark that when applying Theorem 4.9 in practice, one often wants to use an *explicit* isomorphism between $(\mathbb{Z}/q\mathbb{Z})^\times$ and the above direct product, i.e. fix choices of primitive roots in each $(\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$ and work directly with the Chinese Remainder Theorem.

4.3. **Explicit list of all Dirichlet characters modulo q.** Recall the definition of a Dirichlet character, Definition 1.1 on page 9, and Remarks 1.1–1.3. We make some further remarks. (Notation: $\mathbb{C}^\times$ is the group $\mathbb{C} \setminus \{0\}$ with ordinary multiplication, in accordance with (123).)

Remark 4.1. There is a natural bijective correspondence between the set of Dirichlet characters χ of period q, and the set of group homomorphisms f from $(\mathbb{Z}/q\mathbb{Z})^\times$ to $\mathbb{C}^\times$. Indeed, if χ is a Dirichlet character of period q then we get a corresponding group homomorphism $f : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$ by letting $f([n]) = \chi(n)$ for any $n \in \mathbb{Z}$ with $(n, q) = 1$, where $[n] \in \mathbb{Z}/q\mathbb{Z}$ denotes the residue class of $n \bmod q$. Conversely, if $f : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$ is a group homomorphism then we get a Dirichlet character χ of period q by letting

$$\chi(n) = \begin{cases} f([n]) & \text{if } (n,q) = 1 \\ 0 & \text{if } (n,q) > 1. \end{cases} \tag{127}$$

In future we will often use the same letter χ both to denote a Dirichlet character modulo q and the corresponding group homomorphism from $(\mathbb{Z}/q\mathbb{Z})^\times$ to $\mathbb{C}^\times$.

We write X_q for the set of all Dirichlet characters modulo q.

Remark 4.2. X_q is a *group*, if we define the product $\chi_1\chi_2$ of any two characters $\chi_1, \chi_2 \in X_q$ by the formula

$$[\chi_1\chi_2](n) := \chi_1(n)\chi_2(n), \qquad \forall n \in \mathbb{Z}. \tag{128}$$

The group axioms are all easy to verify. In particular the identity element in X_q is the principal character χ_0 (recall that the principal character $\chi_0 \in X_q$ is defined by $\chi_0(n) = 1$ if $(n, q) = 1$, $\chi_0(n) = 0$ if $(n, q) > 1$). Furthermore, using Remark 1.3 and (16) we see that the inverse of any $\chi \in X_q$ is given by the *conjugate* character $\overline{\chi}$, defined by $\overline{\chi}(n) := \overline{\chi(n)}$.

We will not use the group structure of X_q until later. Instead, we will start by showing how to make an explicit list of all the Dirichlet characters modulo q. In particular we will see that $\#X_q = \phi(q)$.

Definition 4.2. Assume that $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic and *fix* a primitive root $g \in (\mathbb{Z}/q\mathbb{Z})^\times$ modulo q. Then the *index* $\nu = \nu(x)$ of any $x \in (\mathbb{Z}/q\mathbb{Z})^\times$ is defined to be the unique number $\nu \in \{0, 1, \dots, \phi(q) - 1\}$ such that $x = g^{\nu(x)}$. We also write $\nu(n) := \nu([n])$ for any $n \in \mathbb{Z}$ with $(n, q) = 1$.

Lemma 4.10. *Let $q \geq 2$ be such that $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic, and let ν be the index function corresponding to a fixed primitive root $g \in (\mathbb{Z}/q\mathbb{Z})^\times$. Let ω be any $\phi(q)$th root of unity, i.e. any complex number satisfying $\omega^{\phi(q)} = 1$. Then we get a Dirichlet character χ modulo q by the formula*

$$\chi(n) = \begin{cases} \omega^{\nu(n)} & \text{if } (n, q) = 1 \\ 0 & \text{if } (n, q) > 1. \end{cases} \tag{129}$$

Conversely every Dirichlet character modulo q can be expressed in this way. Different choices of ω gives different χ's, and hence there are exactly $\phi(q)$ distinct Dirichlet characters of period q.

Proof. The function defined by (129) is easily seen to satisfy all the conditions in Definition 1.1 (the multiplicativity relation (14) follows from $\nu(nm) \equiv \nu(n) + \nu(m) \pmod{\phi(q)}$ if $(n, q) = (m, q) = 1$, whereas if $(n, q) > 1$ or $(m, q) > 1$ then $\chi(nm) = 0 = \chi(n)\chi(m)$). Hence (129) gives a Dirichlet character modulo q.

Conversely, assume that χ is a given Dirichlet character modulo q. Set $\omega = \chi(g) \in \mathbb{C}^\times$. Then $\omega^{\phi(q)} = \chi(g)^{\phi(q)} = \chi(g^{\phi(q)}) = \chi(1) = 1$, i.e. ω is a $\phi(q)$th root of unity. Furthermore, for each $n \in \mathbb{Z}$ with $(n, q) = 1$ we have $\chi(n) = \chi(g^{\nu(n)}) = \chi(g)^{\nu(n)} = \omega^{\nu(n)}$, whereas $\chi(n) = 0$ if $(n, q) > 1$, by (16). Hence (129) holds.

Different choices of ω give different Dirichlet characters, since $\chi(g) = \omega$. Note also that there are exactly $\phi(q)$ distinct $\phi(q)$th roots of unity, namely $e(0), e(\frac{1}{\phi(q)}), e(\frac{2}{\phi(q)}), \dots, e(\frac{\phi(q)-1}{\phi(q)})$.[7] Hence there are exactly $\phi(q)$ distinct Dirichlet characters of period q. □

[7]Here and later we use the standard notation $e(x) := e^{2\pi i x}$.

We will later apply Lemma 4.10 when $q = p^\alpha$ with p an odd prime number; this is ok since $(\mathbb{Z}/q\mathbb{Z})^\times$ is known to be cyclic in this case, cf. Lemma 4.6. We now turn to the case $q = 2^\alpha$ with $\alpha \geq 2$. In this case we have seen that $(\mathbb{Z}/q\mathbb{Z})^\times$ is non-cyclic (unless $\alpha = 2$). In view of Lemma 4.8, we can uniquely define the index functions $\nu : (\mathbb{Z}/2^\alpha\mathbb{Z})^\times \to \{0, 1\}$ and $\nu' : (\mathbb{Z}/2^\alpha\mathbb{Z})^\times \to \{0, 1, \dots, 2^{\alpha-2} - 1\}$ by the relation

$$n = (-1)^{\nu(n)} 5^{\nu'(n)}, \qquad \forall n \in (\mathbb{Z}/2^\alpha\mathbb{Z})^\times. \tag{130}$$

We now give a complete description of the set of Dirichlet characters modulo $q = 2^\alpha$:

Lemma 4.11. *Let $q = 2^\alpha$ with $\alpha \geq 2$. Let $\omega \in \{-1, 1\}$ and let ω' be any $2^{\alpha-2}$th root of unity. Then we get a Dirichlet character χ modulo q by the formula*

$$\chi(n) = \begin{cases} \omega^{\nu(n)}(\omega')^{\nu'(n)} & \text{if } 2 \nmid n \\ 0 & \text{if } 2 \mid n. \end{cases} \tag{131}$$

Conversely every Dirichlet character of period q can be expressed in this way. Different choices of the pair $\langle \omega, \omega' \rangle$ give different χ's, and hence there are exactly $2 \cdot 2^{\alpha-2} = \phi(q)$ distinct Dirichlet characters of period $q = 2^\alpha$.

(Note that in the special case $q = 4$, ω' can only be $\omega' = 1$, and hence we get the two Dirichlet characters which we saw in Example 1.1.)

Proof. The function defined by (131) is easily seen to satisfy all the conditions in Definition 1.1 Hence (131) gives a Dirichlet character modulo q. Conversely, assume that χ is a given Dirichlet character modulo q. Set $\omega = \chi(-1)$ and $\omega = \chi(5)$. Then $\omega^2 = \chi(-1)^2 = \chi((-1)^2) = \chi(1) = 1$ and $(\omega')^{2^{\alpha-2}} = \chi(5)^{2^{\alpha-2}} = \chi(5^{2^{\alpha-2}}) = \chi(1) = 1$, i.e. $\omega \in \{-1, 1\}$ and ω' is a $2^{\alpha-2}$th root of unity. Furthermore, for each odd $n \in \mathbb{Z}$ we have $\chi(n) = \chi((-1)^{\nu(n)}5^{\nu'(n)}) = \chi(-1)^{\nu(n)}\chi(5)^{\nu'(n)} = \omega^{\nu(n)}(\omega')^{\nu'(n)}$, whereas $\chi(n) = 0$ if n is even, by (16). Hence (131) holds. Finally, different pairs $\langle \omega, \omega' \rangle$ give different χ's, since $\chi(-1) = \omega$ and $\chi(5) = \omega'$. □

Having treated all cases of prime powers, it is now easy to give an explicit list of all Dirichlet characters for a *general* q, using the Chinese Remainder Theorem:

Lemma 4.12. *Let q be an integer ≥ 2, and write its prime factorization as $q = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$ (where $p_1, \dots, p_r$ are distinct primes and $\alpha_1, \dots, \alpha_r \in \mathbb{Z}^+$). For $j = 1, \dots, r$ we let $X_{p_j^{\alpha_j}}$ be the set of all Dirichlet characters modulo $p_j^{\alpha_j}$; note that this set is explicitly described in Lemma 4.10 (if $p_j \neq 2$) or Lemma 4.11 (if $p_j = 2$). Then for each choice of an r-tuple $\langle \chi_1, \dots, \chi_r \rangle \in X_{p_1^{\alpha_1}} \times \cdots \times X_{p_r^{\alpha_r}}$, we get a Dirichlet character χ modulo q by the formula*

$$\chi(n) = \prod_{j=1}^{r} \chi_j(n). \tag{132}$$

Conversely every Dirichlet character modulo q can be expressed in this way. Different choices of r-tuples $\langle \chi_1, \dots, \chi_r \rangle \in X_{p_1^{\alpha_1}} \times \cdots \times X_{p_r^{\alpha_r}}$ give different χ's, and hence there are exactly $\prod_{j=1}^r \# X_{p_j^{\alpha_j}} = \prod_{j=1}^r \phi(p_j^{\alpha_j}) = \phi(q)$ Dirichlet characters modulo q.

Proof. The function defined by (132) is easily seen to satisfy all the conditions in Definition 1.1. (To prove the periodicity relation (13), note that for each j we have $\chi_j(n+p_j^{\alpha_j}) = \chi_j(n)$ and $p_j^{\alpha_j} \mid q$; hence $\chi_j(n+q) = \chi_j(n)$. To prove relation (16), i.e. $\chi(n) = 0$ whenever $(n,q) > 1$, note that if $(n,q) > 1$ then there is some j such that $p_j \mid n$, and then $\chi_j(n) = 0$, implying $\chi(n) = 0$.) Hence (132) gives a Dirichlet character modulo q.

Conversely, assume that χ is a given Dirichlet character modulo q. Take some $j \in \{1, \dots, r\}$. For each $n \in \mathbb{Z}$ we know from the Chinese Remainder Theorem that there is some $n_j \in \mathbb{Z}$ (uniquely determined modulo q) such that $n_j \equiv n \pmod{p_j^{\alpha_j}}$ and $n_j \equiv 1 \pmod{p_k^{\alpha_k}}$ for all $k \neq j$. Let us *define* $\chi_j(n) := \chi(n_j)$! This definition is unambiguous since n_j is uniquely determined modulo q for given n, and one checks that χ_j satisfies the relations (13)–(16) with q replaced by $p_j^{\alpha_j}$, i.e. $\chi_j \in X_{p_j^{\alpha_j}}$! Carrying this out for each j we get an r-tuple $\langle \chi_1, \dots, \chi_r \rangle \in X_{p_1^{\alpha_1}} \times \cdots \times X_{p_r^{\alpha_r}}$, and we claim that (132) holds for this tuple. For our χ_j's, the right hand side of (132) is $\prod_{j=1}^r \chi_j(n) = \prod_{j=1}^r \chi(n_j) = \chi\big(\prod_{j=1}^r n_j\big)$. Here $\prod_{j=1}^r n_j \equiv 1 \cdot \ldots \cdot 1 \cdot n \cdot 1 \cdot \ldots \cdot 1 = n \pmod{p_k^{\alpha_k}}$ for each $k \in \{1, \dots, r\}$; hence by the Chinese Remainder Theorem we have $\prod_{j=1}^r n_j \equiv n \pmod{q}$, and thus $\chi\big(\prod_{j=1}^r n_j\big) = \chi(n)$, and hence (132) holds.

Finally, to see that different choices of r-tuples $\langle \chi_1, \dots, \chi_r \rangle \in X_{p_1^{\alpha_1}} \times \cdots \times X_{p_r^{\alpha_r}}$ give different χ's, note that for any $j \in \{1, \dots, r\}$ and any given $n \in \mathbb{Z}$, if we take n_j as in the last paragraph then (132) implies $\chi(n_j) = \prod_{k=1}^r \chi_k(n_j) = \prod_{k=1}^r \begin{cases} \chi_k(n) & \text{if } k = j \\ 1 & \text{if } k \neq j \end{cases} = \chi_j(n)$. This shows that χ_j is uniquely determined from χ in (132). □

The above Lemma 4.12 shows in particular that $\# X_q = \phi(q)$. Thus X_q and $(\mathbb{Z}/q\mathbb{Z})^\times$ have the same number of elements! We will see in Section 4.5 that there exists a *duality* between these two groups: We know that X_q is the group of characters on $(\mathbb{Z}/q\mathbb{Z})^\times$ (cf. Remark 4.1), but conversely it turns out that $(\mathbb{Z}/q\mathbb{Z})^\times$ can be viewed as the group of characters on X_q.

4.4. **Some consequences.** Recall that we proved in Lemma 3.13 that if χ is a non-principal Dirichlet character modulo q then $\sum_{n=1}^q \chi(n) = 0$. Note also that, trivially, if χ equals the principal Dirichlet character χ_0 modulo q then $\sum_{n=1}^q \chi(n) = \phi(q)$. Hence, for any Dirichlet character $\chi \in X_q$:

$$\sum_{n=1}^{q} \chi(n) = \begin{cases} \phi(q) & \text{if } \chi = \chi_0 \\ 0 & \text{otherwise.} \end{cases} \tag{133}$$

We will now prove a "dual" version of this result:

Lemma 4.13. *For each $q \in \mathbb{Z}^+$ and $n \in \mathbb{Z}$ we have*

$$\sum_{\chi \in X_q} \chi(n) = \begin{cases} \phi(q) & \text{if } n \equiv 1 \ (mod\ q) \\ 0 & \text{otherwise.} \end{cases} \tag{134}$$

Proof. If $n \equiv 1 \pmod q$ then the result is trivial since $\chi(n) = 1$ for all $\chi \in X_q$, and if $(n, q) > 1$ then the result is also trivial, since then $\chi(n) = 0$ for all $\chi \in X_q$. Hence from now on we may assume $n \not\equiv 1 \pmod q$ and $(n, q) = 1$, and our task is to prove that $\sum_{\chi \in X_q} \chi(n) = 0$.

To start with we will construct some $\chi' \in X_q$ such that $\chi'(n) \neq 1$. Let q have the prime factorization $q = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$. Since $n \not\equiv 1 \pmod q$ there is some $k \in \{1, \dots, r\}$ for which $n \not\equiv 1 \pmod{p_k^{\alpha_k}}$. If $p_k \neq 2$ then if we apply Lemma 4.10 with "$q_{new} = p_k^{\alpha_k}$" and $\omega = e(\frac{1}{\phi(p_k^{\alpha_k})})$ we get a Dirichlet character χ_k modulo $p_k^{\alpha_k}$ with $\chi_k(n) \neq 1$. In the remaining case $p_k = 2$ we note that $n \not\equiv 1 \pmod{2^{\alpha_k}}$ and $2 \nmid n$ together imply $\alpha_k \geq 2$. Take $\nu \in \{0, 1\}$ and $\nu' \in \{0, 1, \dots, 2^{\alpha_k - 2} - 1\}$ so that $n \equiv (-1)^\nu 5^{\nu'} \pmod{2^{\alpha_k}}$, cf. Lemma 4.8. Then either $\nu \neq 0$ or $\nu' \neq 0$, since $n \not\equiv 1 \pmod{2^{\alpha_k}}$. If $\nu' \neq 0$ (this is only possible when $\alpha_k \geq 3$) then we take $\omega = 1$ and $\omega' = e(2^{2-\alpha_k})$ in Lemma 4.11; if $\nu' = 0$ and $\nu \neq 0$ then we instead take $\omega = -1$ and $\omega' = 1$ in Lemma 4.11; in all cases this gives a character χ_k modulo 2^{α_k} such that $\chi_k(n) \neq 1$. Finally, for each $j \neq k$ we let χ_j be the principal character modulo $p_j^{\alpha_j}$, and define $\chi' = \prod_{j=1}^r \chi_j$ as in (132); viz. in more concrete terms:

$$\chi'(m) = \begin{cases} \chi_k(m) & \text{if } (m, q) = 1 \\ 0 & \text{if } (m, q) > 1. \end{cases} \tag{135}$$

Then $\chi' \in X_q$ and $\chi'(n) \neq 1$, as desired!

Now we can use the proof method from the proof of Lemma 3.13, letting our χ' play the role of "m" in that proof, and using the group structure of X_q (cf. Remark 4.2). Note that

$$\chi'(n) \sum_{\chi \in X_q} \chi(n) = \sum_{\chi \in X_q} \chi'(n)\chi(n) = \sum_{\chi \in X_q} [\chi'\chi](n), \tag{136}$$

and here $\chi'\chi$ visits each character in X_q exactly once, since X_q is a group; hence we conclude

$$\chi'(n) \sum_{\chi \in X_q} \chi(n) = \sum_{\chi \in X_q} \chi(n). \tag{137}$$

Since $\chi'(n) \neq 1$ this implies $\sum_{\chi \in X_q} \chi(n) = 0$, and we are done. □

Using Lemma 4.13 it is now easy to prove Lemma 1.5, which we needed in the first lecture in the proof of Dirichlet's Theorem. Recall that Lemma 1.5 states that for any $a, n \in \mathbb{Z}$

with $(a,q)=1$ we have

$$\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\chi(n)=\begin{cases}1 & \text{if } n\equiv a \pmod q,\\ 0 & \text{otherwise.}\end{cases} \tag{138}$$

Proof of Lemma 1.5. Let a^{-1} be the multiplicative inverse of a in $(\mathbb{Z}/q\mathbb{Z})^\times$. Then $\chi(a)\chi(a^{-1})=\chi(aa^{-1})=\chi(1)=1$ and hence $\chi(a^{-1})=\overline{\chi(a)}$. Hence the left hand side of (138) equals

$$\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi(a)}\chi(n)=\frac{1}{\phi(q)}\sum_{\chi\in X_q}\chi(a^{-1})\chi(n)=\frac{1}{\phi(q)}\sum_{\chi\in X_q}\chi(a^{-1}n) \tag{139}$$

By Lemma 4.13 this is $=1$ if $a^{-1}n\equiv 1 \pmod q$, and otherwise $=0$. □

4.5. *** Fourier analysis and structure theory for finite abelian groups.** This section is external reading and not required for the course. The purpose is to show how many of the results and concepts in Section 4.3 and 4.4 take a slightly simpler and more elegant form when formulated in the more general context of an arbitrary finite abelian group (in place of $(\mathbb{Z}/q\mathbb{Z})^\times$). We also wish to comment on how these facts belong to an even more general framework which I believe most of you are already familiar with, at least in parts.

Throughout this section we will assume that G is a finite abelian group (except in certain expository paragraphs below where we explicitly mention that we consider a more general group G). We will use multiplicative notation, i.e. the binary operation in G is $G\ni\langle g_1,g_2\rangle\mapsto g_1g_2\in G$, the identity is denoted $1\in G$, and the inverse of $g\in G$ is g^{-1}.

Definition 4.3. Let G be a finite abelian group. A *character* of G is a homomorphism from G to $\mathbb{C}^\times$ (the group of non-zero complex numbers under multiplication). If χ_1 and χ_2 are characters of G then their *product*, $\chi_1\chi_2$, is defined by pointwise multiplication, viz.

$$[\chi_1\chi_2](g)=\chi_1(g)\chi_2(g),\qquad \forall g\in G.$$

The set of characters of G form a group under the above operation; this group is called the *dual group* of G, and denoted by $\widehat{G}$.

Remark 4.3. Note that for any $\chi\in\widehat{G}$ and $g\in G$ we actually have $|\chi(g)|=1$, for if $n=\#G$ then $g^n=1$ by Lagrange's Theorem, and hence $\chi(g)^n=\chi(g^n)=1$.

We now have the following results (we postpone the proofs until the end of this section):

Theorem 4.14. *For any $\chi_1,\chi_2\in\widehat{G}$ we have*

$$\frac{1}{\#G}\sum_{g\in G}\chi_1(g)\overline{\chi_2(g)}=\begin{cases}1 & \textit{if } \chi_1=\chi_2\\ 0 & \textit{else.}\end{cases} \tag{140}$$

Theorem 4.15. *The dual group $\widehat{G}$ is isomorphic with G. (But there is no canonical choice of isomorphism.)*

Let us denote by $L^2(G)$ the Hilbert space of all functions $f : G \to \mathbb{C}$, with inner product given by

$$\langle f_1, f_2 \rangle = \frac{1}{\#G} \sum_{g \in G} f_1(g)\overline{f_2(g)}. \tag{141}$$

Then Theorem 4.14 says that $\widehat{G}$ is an orthonormal set of vectors in $L^2(G)$. Furthermore $\dim L^2(G) = \#G$ and Theorem 4.15 implies $\#\widehat{G} = \#G$; hence $\widehat{G}$ *is in fact an orthonormal basis in* $L^2(G)$.

Note that every $g \in G$ gives a character $J(g)$ on $\widehat{G}$ defined by $J(g) : \widehat{G} \ni \chi \mapsto \chi(g) \in \mathbb{C}^\times$.

Theorem 4.16. ***Pontryagin Duality Theorem.*** *The map J is a (canonical) isomorphism between G and $\widehat{\widehat{G}}$, the dual group of the dual group of G.*

Example 4.3. In the special case of $G = (\mathbb{Z}/q\mathbb{Z})^\times$, the characters on G are exactly (after a trivial identification, cf. Remark 4.1) the Dirichlet characters modulo q, i.e. the dual group is $\widehat{G} = X_q$. Hence Theorem 4.14 implies that for any χ_1, χ_2 we have

$$\frac{1}{\phi(q)} \sum_{n=1}^{q} \chi_1(n)\overline{\chi_2(n)} = \begin{cases} 1 & \text{if } \chi_1 = \chi_2 \\ 0 & \text{else.} \end{cases} \tag{142}$$

Taking χ_2 to be the principal character this gives back the relation (133). (On the other hand (142) follows from (133) applied to the character $\chi = \chi_1\overline{\chi}_2$.)

Next let us apply Theorem 4.14 to the group $\widehat{G} = X_q$; this gives, for any two characters X_1, X_2 on $\widehat{G}$: $\frac{1}{\phi(q)} \sum_{\chi \in X_q} X_1(\chi)\overline{X_2(\chi)} = 1$ if $X_1 = X_2$ and 0 otherwise. Using now Theorem 4.16 we can write $X_1 = J(a_1)$, $X_2 = J(a_2)$ with $a_1, a_2 \in (\mathbb{Z}/q\mathbb{Z})^\times$; then $X_1(\chi) = \chi(a_1)$, $X_2(\chi) = \chi(a_2)$, and $X_1 = X_2$ holds if.f. $a_1 = a_2$. Hence

$$\frac{1}{\phi(q)} \sum_{\chi \in X_q} \chi(a_1)\overline{\chi(a_2)} = \begin{cases} 1 & \text{if } a_1 = a_2 \\ 0 & \text{else.} \end{cases}$$

Thus we have recovered Lemma 1.5.

We now comment on how the above results are special cases of more general theory. First, recall that in representation theory, if G is *any* finite group then a "character" on G is by definition a function $\chi : G \to \mathbb{C}$ obtained as $\chi(g) := \mathrm{Tr}\ \rho(g)$ where $\rho : G \to \mathrm{GL}(V)$ is some representation of G on a (finite dimensional) complex vector space V. Also χ is said to be an "irreducible character" if ρ is an irreducible representation. Note here that *in the special case* $\dim V = 1$ we may identify $\mathrm{GL}(V)$ with $\mathbb{C}^\times$ and then $\chi = \rho$, i.e. χ *is simply a homomorphism from G to $\mathbb{C}^\times$*. Now one knows that *if G is abelian* then *all* irreducible representations of G have degree 1 (cf. e.g., [64, Thm. 9]). Hence the characters of a finite abelian group G which we define in Definition 4.3 are exactly the "irreducible characters" in representation theory. Now Theorem 4.14 is seen to be a special case of the orthogonality

relations for irreducible characters on a general finite group, [64, Thm. 3]. Furthermore we know in general that the number of irreducible representations of G (up to isomorphism) is equal to the number of conjugacy classes of G [64, Thm. 7]; *if G is abelian* this number is equal to $\#G$ and thus $\#\widehat{G} = \#G$.

But the above facts are also special cases of results in *Fourier analysis on abelian groups:* Keep G abelian, but allow it to be *infinite*; more precisely assume G to be a topological group which is locally compact and abelian (LCA) (we refer to [61, App. B] for the exact definitions; our purpose here is only to give a first brief orientation). Then $\widehat{G}$, the *dual group* of G, by definition consists of all continuous homomorphisms $\chi : G \to \mathbb{C}^\times$ such that $|\chi(g)| = 1$ for all $g \in G$. This is equivalent with Definition 4.3 in the special case when G is finite[8], and for a general LCA group G there is a natural definition of a topology on $\widehat{G}$ under which $\widehat{G}$ is a LCA group [61, §1.2]. The standard Pontryagin Duality Theorem (cf. Theorem 4.16) is about this general setting: The definition of $J : G \to \widehat{\widehat{G}}$ given above turns out to be valid for a general LCA group G, and J *gives an isomorphism of topological groups* $J : G \xrightarrow{\sim} \widehat{\widehat{G}}$, [61, Thm. 1.7.2]. However Theorem 4.15 does *not* generalize; on the contrary it turns out that, for instance, if G is discrete then $\widehat{G}$ is compact, and if G is compact then $\widehat{G}$ is discrete [61, Thm. 1.2.5].[9] Finally the orthogonality relations in Theorem 4.14 generalize in a natural way to the case of an arbitrary *compact* abelian group G (cf. [61, p. 10(1)]):

$$\int_G \chi_1(x)\overline{\chi_2(x)}\,dx = \begin{cases} 1 & \text{if } \chi_1 = \chi_2 \\ 0 & \text{otherwise,} \end{cases}$$

where dx is the (so called) Haar measure on G, normalized so that $\int_G dx = 1$.

A central concept in this general theory of LCA groups is the *Fourier transform*, for which there exists general results on the Fourier inversion formula and Parseval's formula. You are probably familiar with these formulae in the special cases of $G = \mathbb{R}/\mathbb{Z}$ and $G = \mathbb{R}$. I will now write them out in our special case of G a finite abelian group. (Cf. [61, §§1.2.3, 1.5.1, 1.6.1] for the general case.)

Thus from now on we (again) assume that G is a finite abelian group.

Definition 4.4. Given any function $f : G \to \mathbb{C}$, the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by

$$\widehat{f}(\chi) = \frac{1}{\#G}\sum_{g\in G} f(g)\chi(g^{-1}) \qquad (\chi \in \widehat{G}) \tag{143}$$

is called the *Fourier transform* of f.

[8] A topological group G is always assumed to be Hausdorff [61, B4] and thus if G is finite then the topology has to be the *discrete* topology.

[9] This statement is consistent with Theorem 4.15, since G finite $\iff$ [G compact *and* discrete].

Theorem 4.17. ***Fourier inversion.*** *For every function $f : G \to \mathbb{C}$ we have*

$$f(g) = \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\chi(g) \qquad (g \in G). \tag{144}$$

(We again postpone the proofs until the end of this section.) For the next result we define an inner product on $L^2(\widehat{G})$ by

$$\langle f_1, f_2 \rangle = \sum_{\chi \in \widehat{G}} f_1(\chi)\overline{f_2(\chi)}, \qquad (f_1, f_2 \in L^2(\widehat{G})). \tag{145}$$

Theorem 4.18. ***Plancherel's Theorem; Parseval's formula.*** *The Fourier transform is a Hilbert space isomorphism from $L^2(G)$ onto $L^2(\widehat{G})$. In particular we have $\langle f_1, f_2 \rangle = \langle \widehat{f_1}, \widehat{f_2} \rangle$, viz.*

$$\frac{1}{\#G} \sum_{g \in G} f_1(g)\overline{f_2(g)} = \sum_{\chi \in \widehat{G}} \widehat{f_1}(\chi)\overline{\widehat{f_2}(\chi)} \tag{146}$$

for all $f_1, f_2 \in L^2(G)$.

Finally, we will now give the proofs of Theorems 4.14, 4.15, 4.16, 4.17, 4.18.

Proof of Theorem 4.14. (Compare the proof of Lemma 3.13.) The result is trivial if $\chi_1 = \chi_2$ since then $\chi_1(g)\overline{\chi_2(g)} = 1$ for all $g \in G$. Now assume $\chi_1 \neq \chi_2$. Set $\chi = \chi_1\chi_2^{-1}$; then for each $g \in G$ we have $\chi_1(g)\overline{\chi_2(g)} = \chi_1(g)\chi_2(g)^{-1} = \chi(g)$ and hence our task is to prove that $\sum_{g \in G} \chi(g) = 0$. Since $\chi_1 \neq \chi_2$ there is some $h \in G$ such that $\chi(h) \neq 1$. Now

$$\chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G} \chi(hg) = \sum_{g \in G} \chi(g), \tag{147}$$

for when g runs through all the elements of G then so does hg. The above equality, together with $\chi(h) \neq 1$, implies $\sum_{g \in G} \chi(g) = 0$. □

Next, to prove Theorem 4.15 ("$G \cong \widehat{G}$"), we will use the following result on the structure of a general finite abelian group (for a proof, cf. e.g., [9, §V.4]).

Theorem 4.19. ***Kronecker decomposition theorem.*** *Every finite abelian group is isomorphic to a direct product of cyclic groups.*

Let us write $Z(m)$ to denote a cyclic group of order m.

Remark 4.4. There are in general *several* ways to decompose a given group as a direct product of cyclic groups; for example, if m and n are any coprime positive integers then the two groups $Z(mn)$ and $Z(m) \times Z(n)$ are isomorphic. One way to obtain a *unique* decomposition is to require that each cyclic group should have order equal to a prime power. Thus: *Every G finite abelian group is isomorphic to a direct product of cyclic*

groups of the form $Z(p_1^{\alpha_1}) \times Z(p_2^{\alpha_2}) \times \cdots \times Z(p_m^{\alpha_m})$ *where the* p_i *are primes, not necessarily distinct, and the* α_i *are positive integers. The direct product is unique except for possible rearrangements of the factors.* Cf., e.g., [9, §§I.7, V.5]

Now let G be a given finite abelian group. Then by Theorem 4.19 there exist positive integers $n_1, n_2, \ldots, n_r$ and an isomorphism $J : Z(n_1) \times Z(n_2) \times \cdots \times Z(n_r) \xrightarrow{\sim} G$. For each $j \in \{1, \ldots, r\}$ we let h_j be a generator of $Z(n_j)$ and set

$$g_j := J\big((1, \ldots, 1, h_j, 1, \ldots, 1)\big) \in G$$

(with h_j occurring in the jth entry of "$(1, \ldots, 1, h_j, 1, \ldots, 1)$", all other entries being the identity element). It then follows that

$$G = \big\{ g_1^{a_1} g_2^{a_2} \cdots g_r^{a_r} \ : \ a_1, \ldots, a_r \in \mathbb{Z} \big\}, \tag{148}$$

where $g_1^{a_1} g_2^{a_2} \cdots g_r^{a_r} = g_1^{a_1'} g_2^{a_2'} \cdots g_r^{a_r'}$ holds if and only if $a_j \equiv a_j' \pmod{n_j}$ for all $j \in \{1, \ldots, r\}$.

Proof of Theorem 4.15. Let $\chi \in \widehat{G}$. For each $j \in \{1, \ldots, r\}$ we have $\chi(g_j)^{n_j} = \chi(g_j^{n_j}) = \chi(1) = 1$. Hence $\chi(g_j)$ is an n_jth root of unity, and thus there is an integer m_j, uniquely determined modulo n_j, so that $\chi(g_j) = e(m_j/n_j)$. Hence

$$\chi(g_1^{a_1} g_2^{a_2} \cdots g_r^{a_r}) = \prod_{j=1}^{r} \chi(g_j)^{a_j} = e\Big(\frac{m_1 a_1}{n_1} + \ldots + \frac{m_r a_r}{n_r}\Big), \tag{149}$$

for all choices of $a_1, \ldots, a_r \in \mathbb{Z}$. This formula shows that χ is uniquely determined from the m_j's. Also note that for any given $m_1, \ldots, m_r \in \mathbb{Z}$, the formula (149) gives a well-defined character $\chi : G \to \mathbb{C}^\times$, for if $g_1^{a_1} g_2^{a_2} \cdots g_r^{a_r} = g_1^{a_1'} g_2^{a_2'} \cdots g_r^{a_r'}$ then $a_j \equiv a_j' \pmod{n_j}$ for all $j \in \{1, \ldots, r\}$, and hence $e\big(\frac{m_1 a_1}{n_1} + \ldots + \frac{m_r a_r}{n_r}\big) = e\big(\frac{m_1 a_1'}{n_1} + \ldots + \frac{m_r a_r'}{n_r}\big)$. Hence we have obtained a bijection

$$\widehat{G} \ni \chi \mapsto \langle m_1, \ldots, m_r \rangle \in (\mathbb{Z}/n_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_r\mathbb{Z}).$$

Finally note that this is a group isomorphism if we take the group operation to be the standard addition in each $\mathbb{Z}/n_j\mathbb{Z}$. (Proof: Suppose $\chi \in \widehat{G}$ maps to $\langle m_1, \ldots, m_r \rangle$ and $\chi' \in \widehat{G}$ maps to $\langle m_1', \ldots, m_r' \rangle$, i.e. $\chi(g_j) = e(m_j/n_j)$ and $\chi'(g_j) = e(m_j'/n_j)$ for all j. Then $[\chi\chi'](g_j) = \chi(g_j)\chi'(g_j) = e((m_j + m_j')/n_j)$ and hence $\chi\chi' \in \widehat{G}$ maps to $\langle m_1 + m_1', \ldots, m_r + m_r' \rangle$.) To conclude the proof we need only note that $(\mathbb{Z}/n_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_r\mathbb{Z})$ is a product of cyclic groups, isomorphic with $Z(n_1) \times \ldots \times Z(n_r)$, and thus isomorphic with G. □

Proof of Theorem 4.16 (Pontryagin duality). The map $J : G \to \widehat{\widehat{G}}$ is a homomorphism, for if $g_1, g_2 \in G$ then for each $\chi \in \widehat{G}$ we have $J(g_1 g_2)(\chi) = \chi(g_1 g_2) = \chi(g_1) \cdot \chi(g_2) = J(g_1)(\chi) \cdot J(g_2)(\chi) = \big[J(g_1)J(g_2)\big](\chi)$; hence $J(g_1 g_2) = J(g_1)J(g_2)$.

Next we prove that J is injective. Let $g \in G \setminus \{1\}$. Then $g = g_1^{a_1} g_2^{a_2} \cdots g_r^{a_r}$ for some integers $a_1, \ldots, a_r$ by (148), and $g \neq 1$ means that there is some k such that $a_k \not\equiv 0 \pmod{n_k}$. Now set $m_k = 1$ and set $m_j = 0$ for all $j \neq k$, and let $\chi \in \widehat{G}$ be the corresponding character given by (149). Then $\chi(g) = e\big(0 + \ldots + \frac{m_k a_k}{n_k} + \ldots + 0\big) = e(a_k/n_k) \neq 1$. Thus $J(g)(\chi) = \chi(g) \neq 1$ and hence $J(g) \neq 1$. This proves that J is injective.

Finally it follows from Theorem 4.15 that $\#\widehat{\widehat{G}} = \#\widehat{G} = \#G$; hence J is also surjective. □

Proof of Theorem 4.17 (Fourier inversion). By linearity it suffices to prove (144) for f in a given basis of the vector space of functions from G to $\mathbb{C}$. But we noted above (just below (141)) that $\widehat{G}$ is such a basis; hence it suffices to prove (144) in the case $f \in \widehat{G}$. But in this case we see from Definition 4.4 (where we notice $\chi(g^{-1}) = \chi(g)^{-1} = \overline{\chi(g)}$) and Theorem 4.14 that $\widehat{f}(\chi) = 1$ for $\chi = f$ and $\widehat{f}(\chi) = 0$ otherwise. Hence the right hand side of (144) equals $f(g)$. □

Proof of Theorem 4.18 (Plancherel, Parseval). We first prove Parseval's formula, (146). For any $f_1, f_2 \in L^2(G)$ we have

$$\sum_{\chi \in \widehat{G}} \widehat{f_1}(\chi)\overline{\widehat{f_2}(\chi)} = \frac{1}{(\#G)^2} \sum_{\chi \in \widehat{G}} \Big(\sum_{g_1 \in G} f_1(g_1)\chi(g_1^{-1}) \Big) \Big(\sum_{g_2 \in G} \overline{f_2(g_2)\chi(g_2^{-1})} \Big)$$
$$= \frac{1}{(\#G)^2} \sum_{g_1, g_2 \in G} f_1(g_1)\overline{f_2(g_2)} \sum_{\chi \in \widehat{G}} \chi(g_1^{-1})\overline{\chi(g_2^{-1})}.$$

Here the inner sum equals $\sum_{\chi \in \widehat{G}} J(g_1^{-1})(\chi)\overline{J(g_2^{-1})(\chi)}$ where $J : G \to \widehat{\widehat{G}}$ is the Pontryagin isomorphism, and hence by Theorem 4.14 applied to $\widehat{G}$ the inner sum is $\#\widehat{G}$ if $g_1^{-1} = g_2^{-1}$, and otherwise 0. Hence the triple sum above is seen to equal $\frac{1}{\#G} \sum_{g \in G} f_1(g)\overline{f_2(g)}$, i.e. we have proved (146).

This identity shows that the Fourier transform is a Hilbert space isomorphism from $L^2(G)$ onto some subspace of $L^2(\widehat{G})$. But $\dim L^2(\widehat{G}) = \#\widehat{G} = \#G = \dim L^2(G)$; hence the Fourier transform actually maps onto $L^2(\widehat{G})$. □

4.6. **Primitive characters.** (Cf. Davenport chapter 5.)

Let χ be a nonprincipal character modulo q. Thus $\chi(n)$ is a periodic function of period q; $\chi(n) = 0$ if $(n, q) > 1$, and $\chi(n) \neq 0$ if $(n, q) = 1$. It is possible, that for values of n restricted by the condition $(n, q) = 1$, the function $\chi(n)$ may have a period less than q.

Definition 4.5. We say that χ is *imprimitive* if $\chi(n)$ restricted by $(n,q)=1$ has a period which is less than q.[10] Otherwise we say that χ is *primitive*.

If $q \geq 2$; regarding the principal character modulo q we note that Davenport chooses *not* to count this as an imprimitive character, and it is certainly not primitive either! Thus the set X_q of all characters modulo $q \geq 2$ is partitioned into three disjoint subsets: The primitive characters, the imprimitive characters, and the principal character. However for $q=1$ it is natural to agree that the principal (and only) character $\chi(n) \equiv 1$ *is primitive*.

Definition 4.6. Given $\chi \in X_q$, the *conductor of* χ, $c(\chi)$, is defined to be the smallest positive integer q_1 such that $\chi(n)$ restricted by $(n,q)=1$, has a period q_1. (Davenport calls $c(\chi)$ the "least period" of χ; however the term "conductor" is standard in modern literature.)

Thus $\chi \in X_q$ is primitive if and only if $c(\chi)=q$, and χ is principal if and only if $c(\chi)=1$.

Example 4.4. The following is a Dirichlet character χ modulo 15. Find $c(\chi)$![11]

$n=$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	$\cdots$
$\chi(n)=$	1	-i	0	-1	0	0	-i	i	0	0	1	0	i	-1	0	1	-i	$\cdots$

Lemma 4.20. *If $\chi \in X_q$ and if $q_1 \in \mathbb{Z}^+$ is a period of [$\chi(n)$ for n restricted by $(n,q)=1$], then $c(\chi) \mid q_1$. Hence in particular we have $c(\chi) \mid q$.*

Proof. Set $q' = (c(\chi), q_1)$. Then we know (fact about greatest common divisor) that $q' = xc(\chi) + yq_1$ for some $x, y \in \mathbb{Z}$. Hence since [$\chi(n)$ for n restricted by $(n,q)=1$] has period q_1 *and* period $c(q)$, it must also have period q'. But $q' \mid c(\chi)$ and $c(\chi)$ is by definition the least period of [$\chi(n)$ for n restricted by $(n,q)=1$]. Hence $c(\chi)=q'$, and thus $c(\chi) \mid q_1$. □

Lemma 4.21. *For each $\chi \in X_q$ there is a unique Dirichlet character $\chi_1 \in X_{c(\chi)}$ such that*

$$\chi(n) = \begin{cases} \chi_1(n) & \text{if } (n,q)=1 \\ 0 & \text{if } (n,q)>1. \end{cases} \tag{150}$$

This character χ_1 is primitive.

(The lemma is of course trivial if χ is primitive, since then we get $\chi_1 = \chi \in X_q$.)

Proof. Let $\chi \in X_q$ be given and set $q_1 = c(\chi)$. We define the character $\chi_1 \in X_{q_1}$ as follows. For $n \in \mathbb{Z}$ with $(n, q_1) > 1$ we set $\chi_1(n) := 0$ (of course), and for $n \in \mathbb{Z}$ with $(n, q_1) = 1$ we choose some $t \in \mathbb{Z}$ such that $(n + tq_1, q) = 1$ and then set $\chi_1(n) := \chi(n + tq_1)$.

[10] For maximal clarity, let us write out exactly what we mean by "$\chi(n)$ restricted by $(n,q)=1$ has a period $a \in \mathbb{Z}^+$". This means: For any integers m, n with $(m,q) = (n,q) = 1$ and $m \equiv n \pmod{a}$ we have $\chi(m) = \chi(n)$.

[11] Answer: $c(\chi) = 5$.

Before we continue we must prove that such an integer t exists, for any given $n \in \mathbb{Z}$ with $(n, q_1) = 1$. Note that it suffices to ensure that $p \nmid n + tq_1$ holds for each prime $p \mid q$; this is automatic if $p \mid q_1$ because of our assumption $(n, q_1) = 1$; hence it suffices to ensure that $p \nmid n + tq_1$ holds for each $p \in M$, where M is the set of primes p which divide q but not q_1. But for each $p \in M$ we have $(p, q_1) = 1$ and hence the linear congruence equation $n + tq_1 \equiv 1 \pmod{p}$ has a unique solution $t \pmod{p}$. Hence by the Chinese remainder theorem, since M is a finite set of primes, there exists some $t \in \mathbb{Z}$ such that $n + tq_1 \equiv 1 \pmod{p}$ holds for *all* $p \in M$. In particular we then have $p \nmid n + tq_1$ for all $p \in M$, as desired.

Having thus proved the existence of t for any given $n \in \mathbb{Z}$ with $(n, q_1) = 1$, we must also note that our definition $\chi_1(n) := \chi(n + tq_1)$ does not depend on the choice of t; this is clear since $[\chi(n)$ for n restricted by $(n, q) = 1]$ has period $c(\chi) = q_1$. Hence χ_1 is now a well-defined function $\chi_1 : \mathbb{Z} \to \mathbb{C}$.

One now easily checks that χ_1 satisfies the relations (13)–(16) with q replaced by q_1; thus $\chi_1 \in X_{q_1}$. Furthermore the formula (150) is clear from our definition of χ_1, since for every n with $(n, q) = 1$ we have $(n + tq_1, q) = 1$ with $t = 0$ and hence $\chi_1(n) = \chi(n + tq_1) = \chi(n)$. Also note that if χ_1 is *any* Dirichlet character in X_{q_1} satisfying (150) for our given χ, then $\chi_1(n) = 0$ for all n with $(n, q_1) = 1$, and (150) together with the q_1-periodicity of χ_1 implies $\chi_1(n) = \chi_1(n + tq_1) = \chi(n + tq_1)$ for all $n, t \in \mathbb{Z}$ with $(n + tq_1, q) = 1$. Hence χ_1 is uniquely determined by χ.

Finally we must prove that χ_1 is *primitive*. Suppose that q_2 is a positive integer such that $[\chi_1(n)$ for n restricted by $(n, q_1) = 1]$ has period q_2. Then, a fortiori, $[\chi_1(n)$ for n restricted by $(n, q) = 1]$ has period q_2, since $(n, q_1) = 1$ holds for all n with $(n, q) = 1$. But we have $\chi_1(n) = \chi(n)$ for all n with $(n, q) = 1$; hence $[\chi(n)$ for n restricted by $(n, q) = 1]$ has period q_2, and by the definition of $q_1 = c(\chi)$ this implies $q_2 \geq q_1$. Hence q_1 is the *smallest* period of $[\chi_1(n)$ for n restricted by $(n, q_1) = 1]$, and this means that $\chi_1 \in X_{q_1}$ is primitive. □

In connection with the above lemma, let us note that if $q_1 \in \mathbb{Z}^+$, if χ_1 is any character modulo q_1 (primitive or not), and if q is any multiple of q_1, then (150) defines a character χ modulo q. In this situation we say that χ_1 *induces* χ.

Lemma 4.22. *If χ is a Dirichlet character modulo q which is induced by the Dirichlet character χ_1 modulo q_1 (thus $q_1 \mid q$), then*

$$L(s, \chi) = L(s, \chi_1) \prod_{p \mid q} (1 - \chi_1(p) p^{-s}) \qquad \text{for } \sigma > 1. \tag{151}$$

Remark 4.5. The most important application of Lemma 4.22 is when χ_1 is the primitive character corresponding to χ, as in Lemma 4.21. This case of (151) shows that the study of an arbitrary Dirichlet L-function can be reduced to the study of an L-function corresponding to a *primitive* Dirichlet character.

In the special case that χ is the principal character modulo q, the corresponding primitive character χ_1 is the trivial character modulo 1; $\chi_1 \equiv 1$, thus $L(s,\chi_1) = \zeta(s)$, and Lemma 4.22 gives the relation $L(s,\chi) = \zeta(s)\prod_{p|q}(1-p^{-s})$, which we have already proved in the first lecture, see (25).

Proof of Lemma 4.22. By Euler's product formula for $L(s,\chi)$ (see Lemma 1.7, which we proved in Example 2.1) we have

$$L(s,\chi) = \prod_p (1-\chi(p)p^{-s})^{-1}; \qquad L(s,\chi_1) = \prod_p (1-\chi_1(p)p^{-s})^{-1} \tag{152}$$

for each s with $\sigma > 1$, and both products are absolutely convergent. Here for each prime $p \nmid q$ we have $\chi_1(p) = \chi(p)$, but if $p \mid q$ then we have $\chi(p) = 0$ and thus $1-\chi(p)p^{-s} = 1$ (whereas $\chi_1(p)$ may or may not be 0). Hence

$$L(s,\chi_1) = \prod_{p\nmid q}(1-\chi_1(p)p^{-s})^{-1}\prod_{p|q}(1-\chi_1(p)p^{-s})^{-1} = L(s,\chi)\prod_{p|q}(1-\chi_1(p)p^{-s})^{-1}. \tag{153}$$

□

4.7. **Quadratic reciprocity; the Legendre, Jacobi and Kronecker symbols.** In this section I borrow some material from [9, Chapters 1,2] and [39, Teil I, Kap. 6].

Definition 4.7. For p an odd prime, the values of $a \in \mathbb{Z}/p\mathbb{Z}$ for which the congruence in x,

$$x^2 \equiv a \pmod p \tag{154}$$

is solvable are called the *quadratic residues* p.

Definition 4.8. For p an odd prime and $a \in \mathbb{Z}/p\mathbb{Z}$ (or $a \in \mathbb{Z}$), the *Legendre symbol* (or *quadratic reciprocity symbol*) $\left(\frac{a}{p}\right)$ (also written (a/p)) is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } p\nmid a \text{ and } x^2 \equiv a \pmod p \text{ is solvable} \\ 0 & \text{if } p \mid a \\ -1 & \text{if } x^2 \equiv a \pmod p \text{ is unsolvable.} \end{cases} \tag{155}$$

It follows from this that the number of solutions modulo p to the equation $x^2 \equiv a \pmod p$ equals $1 + (a/p)$.

Obviously

$$\left(\frac{a_1}{p}\right) = \left(\frac{a_2}{p}\right), \qquad \text{if } a_1, a_2 \in \mathbb{Z},\ a_1 \equiv a_2 \pmod p, \tag{156}$$

and one easily proves that

$$\left(\frac{a_1}{p}\right)\cdot\left(\frac{a_2}{p}\right) = \left(\frac{a_1a_2}{p}\right), \qquad \forall a_1, a_2 \in \mathbb{Z}. \tag{157}$$

Using these two relations, the evaluation of the symbol (a/p) reduces to the evaluation of the symbols $(-1/p)$, $(2/p)$ and (q/p), where q is any odd prime $< \frac{p}{2}$.

The famous *quadratic reciprocity* relations are:

Theorem 4.23. *If p, q are any two odd primes then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}, \tag{158}$$

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}, \tag{159}$$

$$\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}. \tag{160}$$

There is also the following important relation due to Euler:

Theorem 4.24. *If p is an odd prime and $a \in \mathbb{Z}/p\mathbb{Z}$ then*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \ (mod\ p). \tag{161}$$

It is useful for many reasons to introduce the following extension of the Legendre symbol:

Definition 4.9. For any integers a, b with b *odd and positive*, the *Jacobi symbol* $\left(\frac{a}{b}\right)$ is defined as follows. Let the prime factorization of b be $b = \prod_{j=1}^{r} p_j$ (where the p_j's don't have to be distinct). Then

$$\left(\frac{a}{b}\right) := \prod_{j=1}^{r}\left(\frac{a}{p_j}\right). \tag{162}$$

(In particular if $b = 1$ then $r = 0$ and thus $\left(\frac{a}{1}\right) = 1$ for all $a \in \mathbb{Z}$.)

It follows directly from the definition that the Jacobi symbol satisfies

$$\left(\frac{a_1}{b}\right) = \left(\frac{a_2}{b}\right), \qquad \text{if } a_1, a_2 \in \mathbb{Z},\ a_1 \equiv a_2 \ (\mathrm{mod}\ b), \tag{163}$$

and

$$\left(\frac{a_1}{b}\right)\cdot\left(\frac{a_2}{b}\right) = \left(\frac{a_1 a_2}{b}\right), \qquad \forall a_1, a_2 \in \mathbb{Z}. \tag{164}$$

Furthermore, as a simple corollary of Theorem 4.23, we have:

Theorem 4.25. *If a and b are positive, odd integers, then*

$$\left(\frac{-1}{b}\right) = (-1)^{\frac{b-1}{2}}, \tag{165}$$

$$\left(\frac{2}{b}\right) = (-1)^{\frac{b^2-1}{8}}, \tag{166}$$

$$\left(\frac{a}{b}\right) = \left(\frac{b}{a}\right)(-1)^{\frac{a-1}{2}\cdot\frac{b-1}{2}}. \tag{167}$$

Note that if $a, b \in \mathbb{Z}$ with b odd but not a prime number, then it is certainly *not* true that the congruence $x^2 \equiv a \pmod{b}$ is solvable if and only if $\left(\frac{a}{b}\right) \neq -1$. However, the condition $\left(\frac{a}{b}\right) \neq -1$ is a *necessary* condition for the solvability of $x^2 \equiv a \pmod{b}$. (Prove this as an exercise!)

Finally we have the following extension of the Jacobi symbol:

Definition 4.10. For any integers a, b with b *even and positive*, and with $a \equiv 0$ or $1 \pmod 4$, we define the *Kronecker symbol* $\left(\frac{a}{b}\right)$ by

$$\left(\frac{a}{b}\right) := \begin{cases} 0 & \text{if } 4 \mid a \\ \left(\frac{b}{|a|}\right) & \text{if } a \equiv 1 \pmod 4. \end{cases} \tag{168}$$

(Here in the right hand side, $\left(\frac{b}{|a|}\right)$ is a Jacobi symbol.)

Remark 4.6. To sum up, we have now defined the symbol $\left(\frac{a}{b}\right)$ exactly when

$$a, b \in \mathbb{Z} \quad \text{and} \quad b > 0 \quad \text{and} \quad \big[2 \nmid b \ \text{ or } \ a \equiv 0 \pmod 4 \ \text{ or } \ a \equiv 1 \pmod 4\big]. \tag{169}$$

Remark 4.7. The exact range of definition for the Jacobi and the Kronecker symbols varies from book to book. The definition which we give here extends that which Davenport gives, in that we also define $\left(\frac{a}{b}\right)$ (as 0) when b is even and $4 \mid a$. In fact our range of definition of the Kronecker symbol is slightly different from each one of [9, Chapters 1,2], [39, Teil I, Kap. 6] and [15]; our main purpose here is to make our definitions agree with those of Davenport [15] while allowing some convenient extra generality.

We now collect several properties of the general Kronecker-Jacobi symbol (the proofs are given at the end of the section).

Proposition 4.26. *For any $a, b \in \mathbb{Z}$ such that $\left(\frac{a}{b}\right)$ is defined, we have $\left(\frac{a}{b}\right) = \pm 1$ if $(a, b) = 1$, and $\left(\frac{a}{b}\right) = 0$ if $(a, b) > 1$.*

Proposition 4.27. *We have $\left(\frac{1}{b}\right) = 1$ for every positive integer b.*

Proposition 4.28. *If a_1, a_2, b are integers such that $\left(\frac{a_1}{b}\right)$ and $\left(\frac{a_2}{b}\right)$ are defined, then*

$$\left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right) = \left(\frac{a_1 a_2}{b}\right). \tag{170}$$

Proposition 4.29. *If a, b_1, b_2 are integers such that $\left(\frac{a}{b_1}\right)$ and $\left(\frac{a}{b_2}\right)$ are defined, then*

$$\left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right) = \left(\frac{a}{b_1 b_2}\right). \tag{171}$$

The following is a way of stating quadratic reciprocity.

Proposition 4.30. *For any $a, b \in \mathbb{Z}$ with $a \equiv 1 \,(mod\ 4)$ and $b > 0$ we have*

$$\left(\frac{a}{b}\right) = \left(\frac{b}{|a|}\right). \tag{172}$$

The following proposition is a direct consequence of Proposition 4.30 and it corresponds to the relations stated by Davenport on p.39 (8) and below.

Proposition 4.31. *If $n, P \in \mathbb{Z}^+$ with P odd and squarefree, then*

$$\left(\frac{n}{P}\right) = \left(\frac{P'}{n}\right), \qquad \text{where } P' = (-1)^{\frac{1}{2}(P-1)} P. \tag{173}$$

We stress that the periodicity relation (163) does *not* generalize in a direct way from the case of the Jacobi symbol to the Kronecker symbol. For example, one checks that

$$\left(\frac{a}{2}\right) = \begin{cases} 0 & \text{if } 4 \mid a \\ 1 & \text{if } a \equiv 1 \,(\text{mod } 8) \\ -1 & \text{if } a \equiv 5 \,(\text{mod } 8) \\ \text{undefined} & \text{otherwise,} \end{cases} \tag{174}$$

and this function of a has period 8 and no smaller period! On the other hand, the following proposition gives a control of the periodicity of $\left(\frac{a}{b}\right)$ *as a function of* b, for special choices of a.

Proposition 4.32. *For any $a, b_1, b_2 \in \mathbb{Z}$ with $b_1, b_2 > 0$ and $a \equiv 0$ or $1 \,(mod\ 4)$, if $b_1 \equiv b_2 \,(mod\ a)$ then $\left(\frac{a}{b_1}\right) = \left(\frac{a}{b_2}\right)$.*

Finally we give the proofs of Propositions 4.26–4.32.

Proof of Proposition 4.26. If b is odd then the proposition follows directly from Definition 4.9 and the fact that for any odd prime p we have $\left(\frac{a}{p}\right) = \pm 1$ if $p \nmid a$ and $\left(\frac{a}{p}\right) = 0$ if $p \mid a$. If b is even and $4 \mid a$ then $\left(\frac{a}{b}\right) = 0$ and $2 \mid (a, b)$, thus $(a, b) > 1$, so that the claim holds. It remains to consider the case when b is even and $a \equiv 1 \,(\text{mod } 4)$. In this case we have $\left(\frac{a}{b}\right) = \left(\frac{b}{|a|}\right)$ by Definition 4.10, and now the desired claim again follows from Definition 4.9. □

Proof of Proposition 4.27. If b is odd then $\left(\frac{1}{b}\right) = 1$ follows from Definition 4.9 since $\left(\frac{1}{p}\right) = 1$ for each prime p. If b is even then $\left(\frac{1}{b}\right) = \left(\frac{b}{1}\right) = 1$, again by Definition 4.9. □

Proof of Proposition 4.30. If b is even then the proposition follows directly from Definition 4.10. Now assume that b is odd. Note that $a > 0$ or $a < 0$, since $a \equiv 1 \pmod 4$. If $a > 0$ then directly by Theorem 4.25,

$$\left(\frac{a}{b}\right) = \left(\frac{b}{a}\right)(-1)^{\frac{a-1}{2}\cdot\frac{b-1}{2}} = \left(\frac{b}{a}\right) = \left(\frac{b}{|a|}\right). \tag{175}$$

If $a < 0$ then, by (164) and Theorem 4.25, and noticing $|a| \equiv 3 \pmod 4$,

$$\left(\frac{a}{b}\right) = \left(\frac{-1}{b}\right)\left(\frac{|a|}{b}\right) = (-1)^{\frac{b-1}{2}}\left(\frac{b}{|a|}\right)(-1)^{\frac{|a|-1}{2}\frac{b-1}{2}} = (-1)^{\frac{b-1}{2}}\left(\frac{b}{|a|}\right)(-1)^{\frac{b-1}{2}} = \left(\frac{b}{|a|}\right).$$

□

Proof of Proposition 4.31. Note that $P' \equiv 1 \pmod 4$; hence Proposition 4.30 applies with $a = P'$ and $b = n$; note that then $|a| = P$ and hence Proposition 4.30 gives the desired relation. □

Proof of Proposition 4.28. If $\left(\frac{a_1}{b}\right)$ and $\left(\frac{a_2}{b}\right)$ are defined then $b > 0$. If b is odd then the proposition follows from (164). If b is even then we must have [$a_1 \equiv 0$ or $1 \pmod 4$] and [$a_2 \equiv 0$ or $1 \pmod 4$]. If $a_1 \equiv 0$ or $a_2 \equiv 0 \pmod 4$ then $\left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right) = 0 = \left(\frac{a_1a_2}{b}\right)$, and in the remaining case $a_1 \equiv a_2 \equiv 1 \pmod 4$ we also have $a_1a_2 \equiv 1 \pmod 4$, and $\left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right) = \left(\frac{b}{|a_1|}\right)\left(\frac{b}{|a_2|}\right) = \left(\frac{b}{|a_1a_2|}\right) = \left(\frac{a_1a_2}{b}\right)$. (Here the Jacobi symbol relation $\left(\frac{b}{|a_1|}\right)\left(\frac{b}{|a_2|}\right) = \left(\frac{b}{|a_1a_2|}\right)$ is a direct consequence of Definition 4.9.) □

Proof of Proposition 4.29. If $\left(\frac{a}{b_1}\right)$ and $\left(\frac{a}{b_2}\right)$ are defined then $b_1, b_2 > 0$. If b_1 and b_2 are odd then the proposition follows directly from Definition 4.9. Now assume that at least one of b_1 and b_2 are even. Then a must be $\equiv 0$ or $1 \pmod 4$, and in the case $a \equiv 0 \pmod 4$ we have $\left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right) = 0 = \left(\frac{a}{b_1b_2}\right)$. Hence we may now assume that $a \equiv 1 \pmod 4$. But then, using Proposition 4.30 and Proposition 4.28 we have

$$\left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right) = \left(\frac{b_1}{|a|}\right)\left(\frac{b_2}{|a|}\right) = \left(\frac{b_1b_2}{|a|}\right) = \left(\frac{a}{b_1b_2}\right).$$

□

Proof of Proposition 4.32. If $a \equiv 1 \pmod 4$ then by Proposition 4.30 and (163) we have $\left(\frac{a}{b_1}\right) = \left(\frac{b_1}{|a|}\right) = \left(\frac{b_2}{|a|}\right) = \left(\frac{a}{b_2}\right)$. If $a \equiv 0 \pmod 4$ and b_1 is even then also b_2 is even (because of $b_1 \equiv b_2 \pmod a$) and hence $\left(\frac{a}{b_1}\right) = 0 = \left(\frac{a}{b_2}\right)$.

It now remains to treat the case when $a \equiv 0 \pmod 4$ and b_1 is odd; then also b_2 is odd (because of $b_1 \equiv b_2 \pmod a$). Note that if $(a, b_1) > 1$ then also $(a, b_2) > 1$ and thus $\left(\frac{a}{b_1}\right) = 0 = \left(\frac{a}{b_2}\right)$; hence we may assume $(a, b_1) = 1$ (and thus $(a, b_2) = 1$). If $a = 0$ then $(a, b_j) = 1$ forces $b_j = 1$ and the claim is clear. Now assume $a \neq 0$; we can then write $a = \varepsilon 2^v a'$ for some $\varepsilon \in \{\pm 1\}$, $v \geq 2$ and an odd positive integer a'. Now by Theorem 4.25 we have

$$\left(\frac{a}{b_1}\right) = \left(\frac{\varepsilon}{b_1}\right)\left(\frac{2}{b_1}\right)^v\left(\frac{a'}{b_1}\right) = \varepsilon^{\frac{b_1-1}{2}}(-1)^{\frac{b_1^2-1}{8}v}\left(\frac{a'}{b_1}\right) = \varepsilon^{\frac{b_1-1}{2}}(-1)^{\frac{b_1^2-1}{8}v}(-1)^{\frac{a'-1}{2}\frac{b_1-1}{2}}\left(\frac{b_1}{a'}\right)$$

and similarly for b_2; thus our task is to prove

$$\varepsilon^{\frac{b_1-1}{2}}(-1)^{\frac{b_1^2-1}{8}v}(-1)^{\frac{a'-1}{2}\frac{b_1-1}{2}}\left(\frac{b_1}{a'}\right) = \varepsilon^{\frac{b_2-1}{2}}(-1)^{\frac{b_2^2-1}{8}v}(-1)^{\frac{a'-1}{2}\frac{b_2-1}{2}}\left(\frac{b_2}{a'}\right).$$

But $b_1 \equiv b_2 \pmod a$ implies $b_1 \equiv b_2 \pmod 4$ and hence $(-1)^{\frac{b_1-1}{2}} = (-1)^{\frac{b_2-1}{2}}$. Also $\left(\frac{b_1}{a'}\right) = \left(\frac{b_2}{a'}\right)$ by (163). Hence it only remains to prove $(-1)^{\frac{b_1^2-1}{8}v} = (-1)^{\frac{b_2^2-1}{8}v}$. If $v = 2$ then this is clear since $(-1)^v = 1$. In the remaining case, $v \geq 3$, we have $8 \mid a$, hence $b_1 \equiv b_2 \pmod 8$ and thus $(-1)^{\frac{b_1^2-1}{8}} = (-1)^{\frac{b_2^2-1}{8}}$, concluding the proof. □

4.8. **Characterization of the real (primitive) characters.** (This is also Davenport chapter 5.)

Definition 4.11. If d is any non-zero integer which is $\equiv 0$ or $1 \pmod 4$, then we define $\chi = \left(\frac{d}{\cdot}\right)$ to be the unique function $\chi : \mathbb{Z} \to \{-1, 0, 1\}$ with period $|d|$ such that

$$\chi(n) = \left(\frac{d}{n}\right) \qquad \text{for } n \in \mathbb{Z}^+.$$

Note that by Proposition 4.32 we have $\left(\frac{d}{n_1}\right) = \left(\frac{d}{n_2}\right)$ for any $n_1, n_2 \in \mathbb{Z}^+$ with $n_1 \equiv n_2 \pmod{|d|}$, and this shows that there indeed exists a unique function $\chi : \mathbb{Z} \to \{-1, 0, 1\}$ as in the definition.

Lemma 4.33. *For every non-zero integer $d \equiv 0$ or $1 \pmod 4$ the function $\chi = \left(\frac{d}{\cdot}\right)$ is a Dirichlet character modulo $|d|$.*

Proof. χ is periodic with period $|d|$ by definition. Furthermore by Proposition 4.29 we have $\chi(n_1 n_2) = \chi(n_1)\chi(n_2)$ for all $n_1, n_2 \in \mathbb{Z}^+$, and using the $|d|$-periodicity this is easily seen to hold for all $n_1, n_2 \in \mathbb{Z}$. Furthermore $\chi(1) = 1$ follows directly from Definition 4.9, and by Proposition 4.26 we see that $\chi(n) = \pm 1$ when $(n, |d|) = 1$ and $\chi(n) = 0$ when $(n, |d|) > 1$. Hence all the properties (13)-(16) hold, proving that χ is a Dirichlet character modulo $|d|$; viz. $\chi \in X_{|d|}$. □

Note that $\chi = \left(\frac{d}{\cdot}\right)$ is a Dirichlet character modulo q also for certain *other* values of q than $q = |d|$; for example if p is any prime dividing d then $\chi \in X_q$ for any $q = |d|p^m$, $m \geq 0$.

We now state the two main results of this section:

Theorem 4.34. ***"Dirichlet's Lemma"*** *(cf.* [9, Ch. II.7]*): Given any $q \in \mathbb{Z}^+$ and any real character $\chi \in X_q$, there is some non-zero integer $d \equiv 0$ or $1 \pmod 4$ such that $\chi = \left(\frac{d}{\cdot}\right)$.*

(We stress that d depends both on q and on χ, and that d is in general not uniquely determined.)

Definition 4.12. An integer d is called a *fundamental discriminant* if either $d \neq 1$, $d \equiv 1 \pmod 4$ and d is squarefree, or $d = 4N$ for some square-free integer $N \equiv 2$ or $3 \pmod 4$.

Theorem 4.35. *If d is a fundamental discriminant then $\chi = \left(\frac{d}{\cdot}\right)$ is a real primitive Dirichlet character modulo $|d|$. Conversely, if $q \geq 2$ and χ is a real primitive character*

modulo q then there is a unique fundamental discriminant d such that $\chi = \left(\frac{d}{\cdot}\right)$, and this d is in fact $d = q$ or $d = -q$.

In particular we see that for each $q \in \mathbb{Z}^+$ there are exactly 0 or 1 or 2 primitive real characters modulo q, this number being equal to the number of fundamental discriminants in the set $\{-q, q\}$.

We also mention two more facts regarding the characters $\left(\frac{d}{\cdot}\right)$:

Lemma 4.36. *For every non-zero integer $d \equiv 0$ or $1 \,(mod\, 4)$, the character $\left(\frac{d}{\cdot}\right)$ is a principal character if and only if d is a square.*

Lemma 4.37. *For every non-zero integer $d \equiv 0$ or $1 \,(mod\, 4)$, if $\chi = \left(\frac{d}{\cdot}\right)$ then*

$$\chi(-1) = \begin{cases} 1 & \text{if } d > 0 \\ -1 & \text{if } d < 0. \end{cases}$$

(Note that $\chi(-1)$ is defined by periodicity, cf. Definition 4.11. The symbol "$\left(\frac{d}{-1}\right)$" is *undefined* in our setup since -1 is not positive!)

The proofs of the above results are not difficult but tedious deductions from our general explicit formulas for Dirichlet characters in Lemma 4.10, Lemma 4.11 and Lemma 4.12, combined with the facts about the reciprocity symbol. The prime 2 always requires special attention in these arguments, and we write out already here the formulas for the primitive real characters which have conductor a power of 2 (cf. Problem 4.6):

$n =$	1	2	3	4	5	6	7	8	$\cdots$	
$\left(\frac{1}{n}\right) =$	1	1	1	1	1	1	1	1	$\cdots$	(period 1)
$\left(\frac{-4}{n}\right) =$	1	0	-1	0	1	0	-1	0	$\cdots$	(period 4)
$\left(\frac{8}{n}\right) =$	1	0	-1	0	-1	0	1	0	$\cdots$	(period 8)
$\left(\frac{-8}{n}\right) =$	1	0	1	0	-1	0	-1	0	$\cdots$	(period 8)

Proof of Lemma 4.36. Recall that a character $\chi \in X_q$ is principal if and only if $\chi(n) \in \{0, 1\}$ for all $n \in \mathbb{Z}$. Hence our task is to prove that $\left(\frac{d}{n}\right) \in \{0, 1\}$ holds for all $n \in \mathbb{Z}^+$ if and only if d is a square.

First assume that d is a square, say $d = d_1^2$ for some $d_1 \in \mathbb{Z} \setminus \{0\}$. If $d_1 \equiv 3 \,(\mathrm{mod}\, 4)$ then we replace d_1 with $-d_1$; hence now $d_1 \equiv 0$, 1 or $2 \,(\mathrm{mod}\, 4)$. Take $n > 0$. If both n and d are even then $\left(\frac{d}{n}\right) = 0$; in every other case the symbol $\left(\frac{d_1}{n}\right)$ is defined, and hence by Proposition 4.28 $\left(\frac{d}{n}\right) = \left(\frac{d_1}{n}\right)^2 \in \{0, 1\}$. Hence $\left(\frac{d}{\cdot}\right)$ is principal.

Conversely, assume that d is *not* a square. First assume $d \equiv 1 \pmod 4$. Then $\left(\frac{d}{n}\right) = \left(\frac{n}{|d|}\right)$ for all $n > 0$ by Proposition 4.30. If $d > 0$ then $|d| = d$, and if $d < 0$ then $|d| = -d \equiv 3 \pmod 4$; hence in all cases do we see that $|d|$ is not a square. Hence if the prime factorization of $|d|$ is $|d| = \prod_{j=1}^r p_j^{\alpha_j}$ (p_j distinct odd primes, all $\alpha_j \geq 1$) then there is some k for which α_k is odd. We know that there is some $g \in \mathbb{Z}/p_k\mathbb{Z}$ for which $\left(\frac{g}{p_k}\right) = -1$. (For example, if we take g to be a primitive root modulo p_k then $g^{\frac{p_k-1}{2}} \not\equiv 1 \pmod p$ and hence $\left(\frac{g}{p_k}\right) = -1$ by Euler's Theorem 4.24.) By the Chinese Remainder Theorem we can now find an $n > 0$ such that $n \equiv g \pmod{p_k^{\alpha_k}}$ and $n \equiv 1 \pmod{p_j^{\alpha_j}}$ for all $j \neq k$; then by Definition 4.9 we have $\left(\frac{d}{n}\right) = \left(\frac{n}{|d|}\right) = \prod_{j=1}^r \left(\frac{n}{p_j}\right)^{\alpha_j} = \left(\frac{n}{p_k}\right)^{\alpha_k} \prod_{\substack{1 \leq j \leq r \\ j \neq k}} 1^{\alpha_j} = (-1)^{\alpha_k} = -1$, thus proving that $\left(\frac{d}{\cdot}\right)$ is not a principal character.

It remains to treat the case $d \equiv 0 \pmod 4$ and d not a square. Then $d = \varepsilon 2^\alpha u$ for some $\varepsilon \in \{-1, 1\}$, $\alpha \geq 2$ and an odd positive integer u. If u is not a square then by the same argument as above (since $u > 0$) there is some $n_u > 0$ for which $\left(\frac{n_u}{u}\right) = -1$, and then by the Chinese Remainder Theorem there is some integer $n > 0$ such that $n \equiv n_u \pmod u$ and $n \equiv 1 \pmod 8$. Now by (164) and Theorem 4.25 we have $\left(\frac{d}{n}\right) = \left(\frac{\varepsilon}{n}\right)\left(\frac{2}{n}\right)^\alpha \left(\frac{u}{n}\right) = \varepsilon^{\frac{n-1}{2}}(-1)^{\frac{n^2-1}{8}\alpha}\left(\frac{u}{n}\right) = \left(\frac{u}{n}\right) = \left(\frac{n}{u}\right)(-1)^{\frac{u-1}{2}\frac{n-1}{2}} = \left(\frac{n}{u}\right) = \left(\frac{n_u}{u}\right) = -1$, proving that $\left(\frac{d}{\cdot}\right)$ is not a principal character. It remains to treat the case when u is a square, $u = u_1^2$ for some other odd integer $u_1 > 0$. Then since $d = \varepsilon 2^\alpha u_1^2$ is not a square we must have $\varepsilon = -1$ or α odd. For every odd integer $n > 0$ with $(n, u_1) = 1$ we have $\left(\frac{d}{n}\right) = \left(\frac{\varepsilon}{n}\right)\left(\frac{2}{n}\right)^\alpha \left(\frac{u_1}{n}\right)^2 = \left(\frac{\varepsilon}{n}\right)\left(\frac{2}{n}\right)^\alpha = \varepsilon^{\frac{n-1}{2}}(-1)^{\frac{n^2-1}{8}\alpha}$. Hence we get $\left(\frac{d}{n}\right) = -1$ if we take any $n > 0$ with $(n, u_1) = 1$ and

$$\begin{cases} n \equiv 7 \pmod 8 & \text{if } \varepsilon = -1 \\ n \equiv 5 \pmod 8 & \text{if } \alpha \text{ is odd.} \end{cases} \tag{176}$$

(If both $\varepsilon = -1$ and α odd then both of $n \equiv 7$ and $5 \pmod 8$ work.) This proves that χ is not a principal character. □

Proof of Lemma 4.37. $\chi = \left(\frac{d}{\cdot}\right)$ is periodic with period $|d|$; thus $\chi(-1) = \chi(|d| - 1) = \left(\frac{d}{|d|-1}\right)$. If $d \equiv 1 \pmod 4$ then by Proposition 4.30 we get $= \left(\frac{|d|-1}{|d|}\right) = \left(\frac{-1}{|d|}\right)$ and this is 1 or -1 according as $|d| \equiv 1$ or $3 \pmod 4$, i.e. according as $d > 0$ or $d < 0$. It remains to treat the case $d \equiv 0 \pmod 4$. In this case $|d| - 1$ is odd and positive and hence we can use the periodicity of the Jacobi symbol (163) to get

$$\chi(-1) = \left(\frac{d}{|d|-1}\right) = \left\{\begin{matrix} \left(\frac{1}{d-1}\right) & \text{if } d > 0 \\ \left(\frac{-1}{|d|-1}\right) & \text{if } d < 0 \end{matrix}\right\} = \begin{cases} 1 & \text{if } d > 0 \\ -1 & \text{if } d < 0, \end{cases}$$

where in the last step we used the fact that $|d| - 1 \equiv 3 \pmod 4$. □

Proof of Theorem 4.34. If $q = 1$ then χ is the trivial character $\chi \equiv 1$ and we can take $d = 1$ (cf. Proposition 4.27).

Now assume that $q > 1$. Let the prime factorization of q be $q = 2^\alpha \prod_{j=1}^r p_j^{\alpha_j}$ where $\alpha \geq 0$, $p_1, \ldots, p_r$ are distinct odd primes and $\alpha_1, \ldots, \alpha_r \in \mathbb{Z}^+$. By Lemma 4.12 combined with Lemma 4.10 and Lemma 4.11, there are complex numbers $\omega, \omega', \omega_1, \ldots, \omega_r$ where $\omega \in \{-1, 1\}$, ω' is a $2^{\alpha-2}$th root of unity (if $\alpha = 0$ or 1 then $\omega = \omega' = 1$), and ω_j is a $\phi(p_j^{\alpha_j})$th root of unity for each j, and we have the following formula for all $n \in \mathbb{Z}$:

$$\chi(n) = \begin{cases} \omega^{\nu(n)} (\omega')^{\nu'(n)} \prod_{j=1}^r \omega_j^{\nu_j(n)} & \text{if } (n,q) = 1 \\ 0 & \text{if } (n,q) > 1. \end{cases} \tag{177}$$

Here $\nu_j(n) \in \{0, 1, \ldots, \phi(p_j^{\alpha_j}) - 1\}$ is the index of n modulo $p_j^{\alpha_j}$, taken with respect to some fixed primitive root $g_j \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$, cf. Definition 4.2; furthermore if $\alpha \geq 2$ then $\nu(n)$ and $\nu'(n)$ are the unique indices $\nu(n) \in \{0, 1\}$, $\nu'(n) \in \{0, 1, \ldots, 2^{\alpha-2} - 1\}$ such that $n \equiv (-1)^{\nu(n)} 5^{\nu'(n)} \pmod{2^\alpha}$, cf. (130). (If $\alpha = 0$ or 1 then the choice of $\nu(n), \nu'(n)$ is immaterial since $\omega = \omega' = 1$; we then have $\chi(n) = \prod_{j=1}^r \omega_j^{\nu_j(n)}$ whenever $(n, q) = 1$.)

We now claim that since χ is real, all the roots of unity $\omega, \omega', \omega_1, \ldots, \omega_r$ must be real, i.e. $\omega, \omega', \omega_1, \ldots, \omega_r$ can only take the values 1 and -1. Indeed, $\omega \in \{-1, 1\}$ is built in from start. Next fix some $j \in \{1, \ldots, r\}$; then by the Chinese Remainder Theorem there is some $n \in \mathbb{Z}$ such that $n \equiv g_j \pmod{p_j^{\alpha_j}}$, $n \equiv 1 \pmod{p_k^{\alpha_k}}$ for all $k \neq j$, and also $n \equiv 1 \pmod{2^\alpha}$. We see from (177) that $\chi(n) = \omega_j$ for this n, and hence ω_j must be real. Finally we prove $\omega' \in \{-1, 1\}$: If $\alpha \leq 3$ then this is built in from start; now assume $\alpha \geq 4$. Then by the Chinese Remainder Theorem there is some $n \in \mathbb{Z}$ such that $n \equiv 5 \pmod{2^\alpha}$ and $n \equiv 1 \pmod{p_j^{\alpha_j}}$ for all $j \in \{1, \ldots, r\}$. We see from (177) that $\chi(n) = \omega'$ for this n, and hence ω' must be real.

Now we are ready to define d. Set

$$d_2 = \begin{cases} 1 & \text{if } \omega = 1,\ \omega' = 1,\ 2 \nmid q \\ 4 & \text{if } \omega = 1,\ \omega' = 1,\ 2 \mid q \\ 8 & \text{if } \omega = 1,\ \omega' = -1 \\ -4 & \text{if } \omega = -1,\ \omega' = 1 \\ -8 & \text{if } \omega = -1,\ \omega' = -1 \end{cases} \tag{178}$$

For each $j \in \{1, \ldots, r\}$ we set $p_j' = (-1)^{\frac{p_j - 1}{2}} p_j$ and define $\beta_j \in \{1, 2\}$ by the relation $\omega_j = (-1)^{\beta_j}$. Finally, we set

$$d = d_2 \prod_{j=1}^r p_j'^{\beta_j}. \tag{179}$$

This d is clearly a non-zero integer with $d \equiv 0$ or $1 \pmod 4$, since d_2 and each factor ${p'_j}^{\beta_j}$ is non-zero and $\equiv 0$ or $\equiv 1 \pmod 4$.

We will now prove that $\chi = \left(\frac{d}{\cdot}\right)$, i.e. we will prove that $\chi(n) = \left(\frac{d}{n}\right)$ for each positive integer n. It follows from (178) and (179) that the prime numbers which divide q are exactly the same as the prime numbers which divide d. Hence if $(n,q) > 1$ then $(n,d) > 1$ and thus $\chi(n) = 0 = \left(\frac{d}{n}\right)$ by (16) and Proposition 4.26. Hence from now on we may assume $(n,q) = 1$.

By Proposition 4.28 we have

$$\left(\frac{d}{n}\right) = \left(\frac{d_2}{n}\right) \prod_{j=1}^{r} \left(\frac{p'_j}{n}\right)^{\beta_j} \tag{180}$$

For each $j \in \{1, \dots, r\}$ we have $p'_j \equiv 1 \pmod 4$ and hence, by Proposition 4.30, $\left(\frac{p'_j}{n}\right)^{\beta_j} = \left(\frac{n}{|p'_j|}\right)^{\beta_j} = \left(\frac{n}{p_j}\right)^{\beta_j}$. But $n \equiv g_j^{\nu_j(n)} \pmod{p_j^{\alpha_j}}$ and thus $\left(\frac{n}{p_j}\right)^{\beta_j} = \left(\frac{g_j^{\nu_j(n)}}{p_j}\right)^{\beta_j} = \left(\frac{g_j}{p_j}\right)^{\nu_j(n)\beta_j}$. Here $\left(\frac{g_j}{p_j}\right) = g_j^{\frac{p_j-1}{2}} \pmod{p_j}$ by Euler's Theorem 4.24 and $g_j^{\frac{p_j-1}{2}} \not\equiv 1 \pmod{p_j}$ since g_j is a primitive root modulo p_j by Lemma 4.7. Hence $\left(\frac{g_j}{p_j}\right) = -1$ and it follows that

$$\left(\frac{p'_j}{n}\right)^{\beta_j} = (-1)^{\nu_j(n)\beta_j} = \omega_j^{\nu_j(n)}. \tag{181}$$

We next claim that

$$\left(\frac{d_2}{n}\right) = \omega^{\nu(n)} (\omega')^{\nu'(n)}. \tag{182}$$

The proof of this will be by a simple (but tedious) check for each of the cases in (178). If $2 \nmid q$ then $\omega = \omega' = 1$ and $d_2 = 1$ and $\left(\frac{d_2}{n}\right) = \left(\frac{1}{n}\right) = 1$ by Proposition 4.27. Hence from now on we may assume $2 \mid q$ (i.e. $\alpha \geq 1$) and thus n is odd, since $(n,q) = 1$.

If $\omega = \omega' = 1$ then $d_2 = 4$ and $\left(\frac{d_2}{n}\right) = \left(\frac{4}{n}\right) = \left(\frac{2}{n}\right)^2 = 1 = \omega^{\nu(n)}(\omega')^{\nu'(n)}$.

If $\omega = 1$ and $\omega' = -1$ then $d_2 = 8$ and $\alpha \geq 3$; thus $n \equiv (-1)^{\nu(n)} 5^{\nu'(n)} \pmod 8$. Hence $\nu'(n)$ is even if $n \equiv 1$ or $7 \pmod 8$ and odd if $n \equiv 3$ or $5 \pmod 8$; thus $\omega^{\nu(n)}(\omega')^{\nu'(n)} = (-1)^{\nu'(n)} = (-1)^{\frac{n^2-1}{8}} = \left(\frac{8}{n}\right)$ (cf. the table on p. 70). Hence (182) holds.

If $\omega = -1$ and $\omega' = 1$ then $d_2 = -4$ and $n \equiv (-1)^{\nu(n)} 5^{\nu'(n)} \pmod 4$. Thus $\nu(n)$ is even if $n \equiv 1 \pmod 4$ and odd if $n \equiv 3 \pmod 4$; hence $\omega^{\nu(n)}(\omega')^{\nu'(n)} = (-1)^{\nu(n)} = (-1)^{\frac{n-1}{2}} = \left(\frac{-4}{n}\right)$ (cf. p. 70), i.e. (182) holds.

Finally if $\omega = \omega' = -1$ then $d_2 = -8$ and $\alpha \geq 3$; thus $n \equiv (-1)^{\nu(n)} 5^{\nu'(n)} \pmod 8$; hence $\nu'(n)$ is even if $n \equiv 1$ or $7 \pmod 8$ and odd if $n \equiv 3$ or $5 \pmod 8$, and $\nu(n)$ is

even if $n \equiv 1 \pmod 4$ and odd if $n \equiv 3 \pmod 4$. Thus $\omega^{\nu(n)}(\omega')^{\nu'(n)} = (-1)^{\nu(n)+\nu'(n)} = (-1)^{\frac{n-1}{2}+\frac{n^2-1}{8}} = \left(\frac{-8}{n}\right)$ (cf. p. 70), i.e. (182) holds. This completes the proof of (182).

Combining (177) with (180), (181) and (182) we obtain $\chi(n) = \left(\frac{d}{n}\right)$, and the proof of Theorem 4.34 is complete. □

Proof of Theorem 4.35. First assume that d is a fundamental discriminant. Then we know from Lemma 4.33 that $\chi = \left(\frac{d}{\cdot}\right)$ is a Dirichlet character modulo $q = |d|$. We wish to prove that χ is primitive modulo $|d|$. From the proof of Theorem 4.34 we get a formula for $\chi(n)$ similar to (177). Indeed, let $p_1, \dots, p_r$ be the odd primes which divide d; set $p_j' = (-1)^{\frac{p_j-1}{2}} p_j$ and define d_2 so that $d = d_2 \prod_{j=1}^r p_j'$. Then since d is a fundamental discriminant we must have $d_2 \in \{1, -4, -8, 8\}$. We can then choose $\omega, \omega' \in \{-1, 1\}$ uniquely so that (178) holds. It then follows as in (180) and (181) (with all $\beta_j = 1$) that for all $n > 0$ we have

$$\chi(n) = \left(\frac{d}{n}\right) = \begin{cases} \left(\frac{d_2}{n}\right) \prod_{j=1}^r (-1)^{\nu_j(n)} & \text{if } (n,q) = 1 \\ 0 & \text{if } (n,q) > 1, \end{cases} \tag{183}$$

where $\nu_j(n) \in \{0, 1, \dots, p_j - 2\}$ is the index of n modulo p_j, taken with respect to some fixed primitive root $g_j \in (\mathbb{Z}/p_j\mathbb{Z})^\times$.

Now for each $k \in \{1, \dots, r\}$ we can argue as follows. Let us take an integer $n > 0$ with $n \equiv g_k \pmod{p_k}$, $n \equiv 1 \pmod{p_j}$ for all $j \neq k$ and $n \equiv 1 \pmod 8$; this is possible by the Chinese Remainder Theorem. Note that $n \equiv 1 \pmod 8$ implies $\left(\frac{d_2}{n}\right) = 1$ (cf. the table on p. 70); hence from (183) we get $\left(\frac{d}{n}\right) = -1$. Hence we have found an integer n with $n \equiv 1 \pmod{q/p_j}$ such that $\chi(n) = -1 \neq \chi(1)$; it follows from this that [$\chi(n)$ for n restricted by $(n,q) = 1$] does not have period q/p_k, and hence the donductor $c(\chi)$ does not divide q/p_k. On the other hand $c(\chi)$ divides q; hence we conclude that $p_k \mid c(\chi)$.

Similarly, we see in the table on p. 70 that $\left(\frac{d_2}{\cdot}\right)$ has period $|d_2|$ and no smaller period; in fact if $d_2 \neq 1$ then for $u = \frac{|d_2|}{2} + 1$ we have $\left(\frac{d_2}{u}\right) = -1$. Hence if we take $n > 0$ so that $n \equiv 1 \pmod{p_j}$ for all j and $n \equiv u \pmod 8$ then $n \equiv 1 \pmod{\frac{q}{2}}$ but by (183) we have $\chi(n) = -1 \neq \chi(1)$. Hence [$\chi(n)$ for n restricted by $(n,q) = 1$] does not have period $\frac{q}{2}$, and hence the donductor $c(\chi)$ does not divide $\frac{q}{2}$; thus $|d_2|$ divides $c(\chi)$. Since we have also seen that $c(\chi)$ is divisible by all p_k it follows that $q \mid c(\chi)$, i.e. χ is primitive. This proves the first part of Theorem 4.35.

Next suppose that $q \geq 2$ and that χ is a real primitive character modulo q. Let us follow the proof of Theorem 4.34, and introduce $\alpha, \omega, \omega', p_j, p_j', \alpha_j, \beta_j, \omega_j, d, d_2$ as there; in particular we now have $\chi = \left(\frac{d}{\cdot}\right)$. We wish to prove that d is a fundamental discriminant. Clearly $d \neq 1$ since $d = 1$ would imply $r = 0$ and $d_2 = 1$, thus $q = 1$, contrary to our assumption. In (177) we recall that $\omega_j^{\nu_j(n)} = \left(\frac{p_j'}{n}\right)^{\beta_j}$ for all $n > 0$ with $(n,q) = 1$ (cf. (181)), and by Proposition 4.32 the function $n \mapsto \left(\frac{p_j'}{n}\right)$ (for $n \in \mathbb{Z}^+$) has period p_j; thus we see

that the function $[\chi(n)$ for n restricted by $(n,q)=1]$ has the period $d_2\prod_{\substack{1\le j\le r\\(\beta_j=1)}} p_j$. Hence since χ is primitive modulo q we must have $(q, d_2\prod_{\substack{1\le j\le r\\(\beta_j=1)}} p_j) = q$; thus $\beta_j = 1$, $\omega_j = -1$ and $\alpha_j = 1$ for all $j \in \{1,\dots,r\}$. From this it also follows that (183) holds for all $n > 0$.

Next we note, using (177), that if $d_2 = 4$ then the function $[\chi(n)$ for n restricted by $(n,q)=1]$ has the period $\prod_{j=1}^r p_j$, since $\left(\frac{4}{\cdot}\right)$ looks as follows:

$n=$	1	2	3	4	5	6	7	8	$\cdots$
$\left(\frac{4}{n}\right)=$	1	0	1	0	1	0	1	0	$\cdots$

At the same time $d_2 = 4$ implies $2 \mid q$, by (178). Hence, since χ is primitive modulo q, $d_2 = 4$ cannot hold. Thus $d_2 \in \{1, 8, -4, -8\}$, and since $\prod_{j=1}^r p_j'$ is squarefree and $\equiv 1 \pmod 4$ it now follows that $d = d_2\prod_{j=1}^r p_j'$ is a fundamental discriminant.

Now by assumption we have $c(\chi) = q$, and on the other hand we saw in the first part of this proof that $\chi = \left(\frac{d}{\cdot}\right)$ is a primitive Dirichlet character modulo $|d|$; hence $q = |d|$, i.e. $d = q$ or $d = -q$.

To complete the proof we now only have to prove the uniqueness of d for given χ. Equivalently, we have to prove that if d, d' are two different fundamental discriminants then the two Dirichlet characters $\left(\frac{d}{\cdot}\right)$ and $\left(\frac{d'}{\cdot}\right)$ are different. It follows from the assumptions that $dd' \neq 0$, $dd' \equiv 0$ or $1 \pmod 4$ and that dd' is not a square. Hence by Lemma 4.36 $\left(\frac{dd'}{\cdot}\right)$ is not a principal character, i.e. there is some $n > 0$ such that $\left(\frac{dd'}{n}\right) = -1$. For this n we have, by Proposition 4.28, $\left(\frac{d}{n}\right)\left(\frac{d'}{n}\right) = -1$ and thus $\left(\frac{d}{n}\right) \neq \left(\frac{d'}{n}\right)$. Hence $\left(\frac{d}{\cdot}\right)$ and $\left(\frac{d'}{\cdot}\right)$ are different. □

4.9. **Problems.**

Problem 4.1. Consider the situation in Lemma 4.12.
(a). Prove that χ is primitive if and only if each χ_j is primitive.
(b). More generally, prove that $c(\chi) = \prod_{j=1}^r c(\chi_j)$.

Problem 4.2. Prove that the number of primitive characters modulo q, $\phi^*(q)$, is given by

$$\phi^*(q) = q\prod_{p\mid\mid q}\Big(1-\frac{2}{p}\Big)\prod_{p^2\mid q}\Big(1-\frac{1}{p}\Big)^2, \tag{184}$$

where "$p \mid\mid q$" means that the product is taken over those primes p which *divide* q *only once*, i.e. primes $p \mid q$ with $p \nmid \frac{q}{p}$. [Hint. One approach is to use Problem 4.1 to show that $\phi^*(q)$ is multiplicative; then we only have to compute $\phi^*(q)$ when q is a prime power and this can be done using Lemma 4.10 and Lemma 4.11.]

Problem 4.3. Let χ be a Dirichlet character modulo q. Prove that a given positive integer q_1 is a period of $\chi(n)$ restricted by $(n,q)=1$ if and only if $\chi(n)=1$ holds for all integers n satisfying $n \equiv 1 \pmod{q_1}$ and $(n,q)=1$.

Problem 4.4. (a). Let χ be a primitive Dirichlet character modulo q. Prove that for any $a, b \in \mathbb{Z}$ we have

$$\frac{1}{q}\sum_{c=0}^{q-1} \chi(ac+b) = \begin{cases} \chi(b) & \text{if } q \mid a \\ 0 & \text{if } q \nmid a. \end{cases} \tag{185}$$

[Hint. When $q \nmid a$, study the set of $m \in (\mathbb{Z}/q\mathbb{Z})^\times$ for which we can prove that the sum remains unchanged after multiplication with $\chi(m)$. (Cf. the proof of Lemma 3.13.)]
(b). Give an example to show that the above formula is in general not valid if χ is not primitive.

Problem 4.5. Prove (174)!

Problem 4.6. Prove that the Dirichlet characters $\left(\frac{-4}{\cdot}\right)$, $\left(\frac{-8}{\cdot}\right)$ and $\left(\frac{8}{\cdot}\right)$ are as stated in the table on p. 70.

Problem 4.7. Prove that if p is an odd prime then there is exactly one primitive real character modulo p, and this character is $\chi(n) = \left(\frac{n}{p}\right)$ $(\forall n \in \mathbb{Z})$.
(It is instructive to try to give two solutions: Give a direct proof without using Theorem 4.35. Also try to derive the result as a consequence of Theorem 4.35, wherein the character is apriori given by a different formula.)

Problem 4.8. Let $q \in \mathbb{Z}^+$. Prove that all characters modulo q are real if and only if $q \in \{1,2,3,4,6,8,12,24\}$.

5. $L(1,\chi)$ AND CLASS NUMBERS

(Cf. Davenport chapter 6.)

5.1. **Equivalence classes of quadratic forms.**

Definition 5.1. An *integral binary quadratic form* is a 2-variable function of the type

$$Q(x,y) = ax^2 + bxy + cy^2, \tag{186}$$

where $a, b, c \in \mathbb{Z}$. From now on we will call such a function a *quadratic form*, or just *form*, for short, and we will often denote the quadratic form in (5.1) by $Q = [a,b,c]$. The *discriminant* of $Q = [a,b,c]$ is given by

$$d = b^2 - 4ac. \tag{187}$$

A fundamental theme is the question of which integers are *representable* by a given quadratic form Q:

Definition 5.2. A pair of integers $\langle x, y\rangle \in \mathbb{Z}^2$ is called a *representation of n by Q* if $Q(x,y) = n$. If $\gcd(x,y) = 1$ then $\langle x, y\rangle$ is called a *proper representation of n by Q*; otherwise $\langle x, y\rangle$ is called an *improper representation of n by Q*.

[See chapter 1 in Conway's "The sensual (quadratic) form", [13] for a beautiful exposition of an algorithm which determines whether a given integer n is (properly) representable by a given quadratic form Q. Also see [7, Ch. 6] for a more detailed discussion of this question.]

We will also represent the quadratic form $Q = [a,b,c]$ by the matrix $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$. The point of this is that

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + \frac{b}{2}y & \frac{b}{2}x + cy \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + bxy + cy^2, \tag{188}$$

where in the last step we identify a 1×1 matrix with its entry. Thus we have *three* different ways to denote the same quadratic form:

$$ax^2 + bxy + cy^2; \qquad [a,b,c]; \qquad \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}. \tag{189}$$

Note that in the matrix notation the discriminant equals $d = -4 \det \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$.

Let us also note that when $a \neq 0$ we may complete the square as follows:

$$Q(x,y) = ax^2 + bxy + cy^2 = \frac{1}{4a}\big((2ax + by)^2 - dy^2\big), \tag{190}$$

and similarly if $c \neq 0$ then $Q(x,y) = \frac{1}{4c}\big((2cy + bx)^2 - dx^2\big)$; finally if $a = c = 0$ then $d = b^2$ and $Q(x,y) = bxy$. From these expression we see that if $d > 0$ then the quadratic form Q

is *indefinite* (i.e. Q can take both positive and negative values for $x, y \in \mathbb{R}$), but if $d < 0$ then Q is *positive definite* if $a > 0$ (i.e. $Q(x,y) > 0$ for all $\langle x, y\rangle \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$), and *negative definite* if $a < 0$ (i.e. $Q(x,y) < 0$ for all $\langle x, y\rangle \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$).

Note that a given integer d can be obtained as the discriminant of a quadratic form if and only if $d \equiv 0$ or $1 \pmod 4$. The following lemma shows that the case when d is a square is special. [By "square" we mean: a square of an integer].

Lemma 5.1. *A quadratic form $Q(x,y) = ax^2 + bxy + cy^2$ can be factored into two integral linear forms, $Q(x,y) = (g_1x + g_2y)(g_3x + g_4y)$ with all $g_j \in \mathbb{Z}$, if and only if d is a square.*

Proof. Note that $Q(x,y) = (g_1x + g_2y)(g_3x + g_4y)$ holds if and only $ax^2 + bx + c = (g_1x + g_2)(g_3x + g_4)$ as a polynomial identity in the variable x. If $a = 0$ then such a factorization is possible ($g_1 = 0$, $g_2 = 1$ etc.) and also $d = b^2 - 0$ is a square. Now assume $a \neq 0$. Then by [a well-known consequence of] Gauss' Lemma (cf., e.g., [20, Cor. 45.28]) a factorization of the desired type exists if and only if $ax^2 + bx + c$ is not irreducible in $\mathbb{Q}[x]$, i.e. if and only if the two zeros of $ax^2 + bx + c$ are rational. But we compute

$$ax^2 + bx + c = a(x - \theta)(x - \theta') \qquad \text{where } \theta = \frac{-b + \sqrt{d}}{2a},\ \theta' = \frac{-b - \sqrt{d}}{2a}. \tag{191}$$

Hence the zeros are rational if and only if $\sqrt{d} \in \mathbb{Q}$, i.e. if and only if d is a square. □

In view of Lemma 5.1 *we will most often assume that d is not a square.* Note that then we must have both $a \neq 0$ and $c \neq 0$.

We next come to the fundamental concept of *equivalence* between quadratic forms. For many questions we can identify two quadratic forms $Q_1(x,y) = a_1x^2 + b_1xy + c_1y^2$ and $Q_2(x,y) = a_2x^2 + b_2xy + c_2y^2$ if they are just the transforms of each other by certain special linear changes of variables,

$$\begin{pmatrix} x \\ y \end{pmatrix} = g \begin{pmatrix} x' \\ y' \end{pmatrix} \qquad \text{with } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},\ \det g \neq 0. \tag{192}$$

Then $\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} x' & y' \end{pmatrix} g^{\mathrm{tr}}$ and hence by (188), Q_1 transforms to Q_2 if and only if

$$\begin{pmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{pmatrix} = g^{\mathrm{tr}} \begin{pmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{pmatrix} g \tag{193}$$

In order for every integral form Q_1 to map to an integral form Q_2 it is natural to require that $\alpha, \beta, \gamma, \delta$ are *integers*; furthermore note that (193) is equivalent with

$$\begin{pmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{pmatrix} = (g^{-1})^{\mathrm{tr}} \begin{pmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{pmatrix} g^{-1}; \tag{194}$$

hence we also wish $g^{-1} = \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}$ to have integral entries. This makes it natural to require that g belongs to the so called *modular group*, $\mathrm{SL}(2,\mathbb{Z})$:

Definition 5.3. The *modular group*, $\mathrm{SL}(2,\mathbb{Z})$, is defined as

$$\mathrm{SL}(2,\mathbb{Z}) = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \,:\, \alpha,\beta,\gamma,\delta \in \mathbb{Z},\ \alpha\delta - \beta\gamma = 1 \right\}. \tag{195}$$

$\mathrm{SL}(2,\mathbb{Z})$ is indeed a *group* under matrix multiplication; the inverse of $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ is $g^{-1} = \left(\begin{smallmatrix} \delta & -\beta \\ -\gamma & \alpha \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$.

Definition 5.4. Two quadratic forms $Q_1 = \left(\begin{smallmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{smallmatrix}\right)$ and $Q_2 = \left(\begin{smallmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{smallmatrix}\right)$ are said to be *equivalent*, written $Q_1 \sim Q_2$, if $\left(\begin{smallmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{smallmatrix}\right) = g^{\mathrm{tr}} \left(\begin{smallmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{smallmatrix}\right) g$ for some $g \in \mathrm{SL}(2,\mathbb{Z})$.

Using the fact that $\mathrm{SL}(2,\mathbb{Z})$ one immediately checks that $\sim$ is indeed an *equivalence relation* on the set of (integral binary) quadratic forms. Note that equivalent forms have the same discriminant, for if $\left(\begin{smallmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{smallmatrix}\right) = g^{\mathrm{tr}} \left(\begin{smallmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{smallmatrix}\right) g$ with $g \in \mathrm{SL}(2,\mathbb{Z})$ then $b_2^2 - a_2c_2 = -4\det \left(\begin{smallmatrix} a_2 & b_2/2 \\ b_2/2 & c_2 \end{smallmatrix}\right) = -4\det(g^{\mathrm{tr}})\det \left(\begin{smallmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{smallmatrix}\right) \det g = -4\det \left(\begin{smallmatrix} a_1 & b_1/2 \\ b_1/2 & c_1 \end{smallmatrix}\right) = b_1^2 - a_1c_1$. Also note that if $Q_1 \sim Q_2$ then exactly the same integers are representable by Q_1 as by Q_2, and similarly for properly representable (recall Definition 5.2):

$$\begin{aligned} &\big\{Q_1(x,y) \,:\, x,y \in \mathbb{Z}\big\} = \big\{Q_2(x,y) \,:\, x,y \in \mathbb{Z}\big\}; \\ &\big\{Q_1(x,y) \,:\, x,y \in \mathbb{Z},\ \gcd(x,y) = 1\big\} = \big\{Q_2(x,y) \,:\, x,y \in \mathbb{Z},\ \gcd(x,y) = 1\big\}. \end{aligned} \tag{196}$$

This follows from the fact that if $g \in \mathrm{SL}(2,\mathbb{Z})$ then the linear map $\begin{pmatrix} x \\ y \end{pmatrix} = g \begin{pmatrix} x' \\ y' \end{pmatrix}$ is a bijection of $\mathbb{Z}^2$ onto itself, which preserves $\gcd(x,y)$.

It is useful to have the explicit expression of $g^{\mathrm{tr}} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} g$ for $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ written out:

$$\begin{aligned} g^{\mathrm{tr}} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} g &= \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha a + \gamma b/2 & \alpha b/2 + \gamma c \\ \beta a + \delta b/2 & \beta b/2 + \delta c \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \\ &= \begin{pmatrix} a\alpha^2 + b\alpha\gamma + c\gamma^2 & (2a\alpha\beta + b\beta\gamma + b\alpha\delta + 2c\gamma\delta)/2 \\ (2a\alpha\beta + b\beta\gamma + b\alpha\delta + 2c\gamma\delta)/2 & a\beta^2 + b\beta\delta + c\delta^2 \end{pmatrix} \end{aligned} \tag{197}$$

Lemma 5.2. *(Lagrange) Every quadratic form is equivalent to some quadratic form $[a,b,c]$ which satisfies $|b| \le |a| \le |c|$.*

Proof. Let us start with a given quadratic form $Q_0 = [a_0,b_0,c_0]$. Among all the integers which are representable by Q_0, we pick one which has minimal non-zero absolute value and call this integer a. Since a is representable there are some $\alpha,\gamma \in \mathbb{Z}$ with $a = Q_0(\alpha,\beta) = a_0\alpha^2 + b_0\alpha\gamma + c_0\gamma^2$. We must have $(\alpha,\gamma) = 1$, for otherwise the number $\frac{a}{(\alpha,\gamma)^2}$ would have smaller absolute value than a and also be representable (as $Q_0(\frac{\alpha}{(\alpha,\gamma)}, \frac{\gamma}{(\alpha,\gamma)})$). From $(\alpha,\gamma) = 1$

it follows that there are some integers β, δ such that $\alpha\delta - \beta\delta = 1$, i.e. $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$, and now by (197) we have

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}^{\mathrm{tr}} \begin{pmatrix}a_0 & b_0/2\\ b_0/2 & c_0\end{pmatrix} \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}a & b'/2\\ b'/2 & c'\end{pmatrix}$$

for some $b', c' \in \mathbb{Z}$, i.e. $[a_0, b_0, c_0] \sim [a, b', c']$. Next the transformation $\left(\begin{smallmatrix}1 & h\\ 0 & 1\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ ($h \in \mathbb{Z}$) takes $[a, b', c']$ to $[a, b, c]$ (thus $[a_0, b_0, c_0] \sim [a, b', c'] \sim [a, b, c]$) where $b = 2ah + b'$, again by (197). Choosing h appropriately we get $|b| \le |a|$. In this situation we also note that c is representable by $[a, b, c]$ (by $x = 0, y = 1$); hence c is representable by Q_0 (cf. (196)), and hence by our choice of a we have $|a| \le |c|$. □

Corollary 5.3. *For every integer d which is not a square there are only a **finite** number of equivalence classes of quadratic forms of discriminant d.*

Proof. By Lemma 5.2 we may choose from each equivalence class some quadratic form $[a, b, c]$ with $|b| \le |a| \le |c|$. Hence the total number of equivalence classes is

$$\le \#\big\{[a, b, c] \,:\, b^2 - ac = d,\ |b| \le |a| \le |c|\big\}. \tag{198}$$

For every $[a, b, c]$ in the above set we have $4a^2 \le 4|ac| = |d - b^2| \le |d| + a^2$, hence $|a| \le \sqrt{\frac{|d|}{3}}$. Since also $|b| \le |a|$ it follows that there are only finitely many choices of a, b, and for any such choice there is at most one integer c satisfying $b^2 - ac = d$ (this uses the fact that d is not a square). Hence the cardinality in (198) is finite. □

Definition 5.5. A quadratic form $Q(x, y) = ax^2 + bxy + cy^2$ is said to be *primitive* if $(a, b, c) = 1$. We also say that an equivalence class C of quadratic forms is *primitive* if C contains some primitive form; it is straightforward that then *all* forms in C are primitive.

Let us note a case of special interest: If d is a fundamental discriminant (cf. Definition 4.12) then *all* quadratic forms of discriminant d are primitive. Indeed, if $b^2 - 4ac = d$ and $(a, b, c) = e > 1$ then writing $a = ea_1$, $b = eb_1$, $c = ec_1$ we have $d = e^2(b_1^2 - 4a_1c_1)$, which is easily seen to contradict the fact that d is a fundamental discriminant.

Definition 5.6. If d is an integer $\equiv 0$ or $\equiv 1 \pmod 4$ which is not a square, then the number of equivalence classes of primitive quadratic forms with discriminant d and which are positive definite or indefinite, is denoted by $h(d)$. (It is called a *class number*.)

Note that for every d as above we have $h(d) \ge 1$, since the following quadratic form:

$$\begin{cases}[1, 0, -\frac{1}{4}d] & \text{if } d \equiv 0 \pmod 4,\\ [1, 1, -\frac{1}{4}(d-1)] & \text{if } d \equiv 1 \pmod 4,\end{cases} \tag{199}$$

is a primitive quadratic form with discriminant d, which is positive definite or indefinite. (This form is called the *principal form* of discriminant d.)

The class numbers for the first few fundamental discriminants:

$d=$	5	8	12	13	17	21	24	28	29	33	37	40	41	44	53	56	57	60	61	65	69	73	76	77
$h(d)=$	1	1	2	1	1	2	2	2	1	2	1	2	1	2	1	2	2	4	1	2	2	1	2	2

$d=$	-3	-4	-7	-8	-11	-15	-19	-20	-23	-24	-31	-35	-39	-40	-43	-47	-51	-52	-55	-56	-59
$h(d)=$	1	1	1	1	1	2	1	2	3	2	3	2	4	2	1	5	2	2	4	4	3

5.2. **Dirichlet's class number formula.** The central result of this lecture is *Dirichlet's class number formula* (1839), which connects $h(d)$ and $L(1, \left(\frac{d}{\cdot}\right))$:

Theorem 5.4. *Let d be an integer $\equiv 0$ or $\equiv 1 \pmod 4$ which is not a square. Then*

$$L(1, \left(\tfrac{d}{\cdot}\right)) = \begin{cases} \frac{2\pi}{w\sqrt{|d|}} h(d) & \text{if } d < 0, \\ \frac{\log \varepsilon_d}{\sqrt{d}} h(d) & \text{if } d > 0, \end{cases} \tag{200}$$

where

$$w = \begin{cases} 2 & \text{if } d < -4 \\ 4 & \text{if } d = -4 \\ 6 & \text{if } d = -3 \end{cases} \tag{201}$$

and (if $d > 0$) $\varepsilon_d = \frac{1}{2}(x + y\sqrt{d})$, where $(x, y) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ is the solution to Pell's equation $x^2 - dy^2 = 4$ for which y is minimal.

We will now give the proof of this theorem (following Landau [39, vol I, Teil IV]). From now on in this section we will always assume that d is as in the above theorem, i.e. d is an integer $\equiv 0$ or $\equiv 1 \pmod 4$ which is not a square.

Definition 5.7. A unimodular substitution $g \in \mathrm{SL}(2, \mathbb{Z})$ is called an *automorph* of a quadratic form $Q = [a, b, c]$ if $\left(\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right) = g^{\mathrm{tr}} \left(\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right) g$.

Lemma 5.5. *The set of automorphs of a given primitive quadratic form $Q = [a, b, c]$ with discriminant d equals:*

$$\left\{ \begin{pmatrix} \frac{1}{2}(t - bu) & -cu \\ au & \frac{1}{2}(t + bu) \end{pmatrix} : t, u \in \mathbb{Z},\ t^2 - du^2 = 4 \right\} \subset SL(2, \mathbb{Z}). \tag{202}$$

Proof. From (197) (and using $\beta\gamma + \alpha\delta = 1 + 2\beta\gamma$) we see that $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z})$ is an automorph of $Q = \left(\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right)$ if and only if

$$\begin{cases} a\alpha^2 + b\alpha\gamma + c\gamma^2 = a \\ 2a\alpha\beta + b(1 + 2\beta\gamma) + 2c\gamma\delta = b \\ a\beta^2 + b\beta\delta + c\delta^2 = c. \end{cases} \tag{203}$$

Here the last equation is a consequence of the first two, since we know that the $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$-transformed form has the same discriminant as Q, and c is uniquely determined from a, b, d in $b^2 - 4ac = d$ (since d is not a square). Hence $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$ is an automorph of Q if and only if

$$\begin{cases} a(\alpha^2-1) + b\alpha\gamma + c\gamma^2 = 0 \\ a\alpha\beta + b\beta\gamma + c\gamma\delta = 0. \end{cases} \tag{204}$$

Suppose that (204) holds. Then (eliminating b)

$$\begin{aligned} 0 &= \big(a(\alpha^2-1) + b\alpha\gamma + c\gamma^2\big)\beta - \big(a\alpha\beta + b\beta\gamma + c\gamma\delta\big)\alpha \\ &= -a\beta + c\gamma(\gamma\beta - \alpha\delta) = -a\beta - c\gamma. \end{aligned} \tag{205}$$

and also (eliminating c)

$$\begin{aligned} 0 &= \big(a(\alpha^2-1) + b\alpha\gamma + c\gamma^2\big)\delta - \big(a\alpha\beta + b\beta\gamma + c\gamma\delta\big)\gamma \\ &= a(\alpha(\alpha\delta - \gamma\beta) - \delta) + b\gamma(\alpha\delta - \beta\gamma) = a(\alpha - \delta) + b\gamma. \end{aligned} \tag{206}$$

Hence a divides both $c\gamma$ and $b\gamma$. Using now $(a,b,c) = 1$ it follows that a divides γ, and hence there is some $u \in \mathbb{Z}$ such that $\gamma = au$. Hence by (205) and (206) we also have $\beta = -cu$ and $\delta - \alpha = bu$. Finally we must also have $1 = \alpha\delta - \beta\gamma = \alpha(bu + \alpha) + acu^2$, viz. (completing the square) $(2\alpha + bu)^2 + (4ac - b^2)u^2 = 4$. Hence if we set $t = 2\alpha + bu \in \mathbb{Z}$ then we have $t^2 - du^2 = 4$ and $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix}\frac{1}{2}(t-bu) & -cu\\ au & \frac{1}{2}(t+bu)\end{smallmatrix}\right)$, i.e. we have proved that every automorph of Q belongs to the set in (202).

Conversely, if $t, u \in \mathbb{Z}$ and $t^2 - du^2 = 4$ then the matrix $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix}\frac{1}{2}(t-bu) & -cu\\ au & \frac{1}{2}(t+bu)\end{smallmatrix}\right)$ belongs to $\mathrm{SL}(2,\mathbb{Z})$, since modulo 2 we have $d = b^2 - 4ac \equiv b$ and thus $t \pm bu \equiv t \pm du \equiv t^2 - du^2 = 4 \equiv 0$, and also $\alpha\delta - \beta\gamma = \frac{1}{4}(t^2 - b^2u^2) - acu^2 = \frac{1}{4}(t^2 - (b^2 - 4ac)u^2) = 1$. Furthermore one verifies by direct computation that this $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is an automorph of Q. (A slightly quicker way to complete this last step is as follows: One immediately checks that $-a\beta - c\gamma = 0$ and $a(\alpha - \delta) + b\gamma = 0$; and as in (205), (206) we see that this implies $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\left(\begin{smallmatrix}-(a\alpha\beta + b\beta\gamma + c\gamma\delta)\\ a(\alpha^2-1) + b\alpha\gamma + c\gamma^2\end{smallmatrix}\right) = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$. Hence since $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is nonsingular we conclude that (204) holds, and thus $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is an automorph of Q.) □

Corollary 5.6. *If $d < 0$ then every quadratic form of discriminant d has exactly w automorphs, with w as in* (201).

Proof. Let Q be a quadratic form of discriminant $d < 0$. If $d < -4$ then the equation $t^2 - du^2 = 4$ has exactly two integer solutions; $\langle t, u\rangle = \langle \pm 2, 0\rangle$. Since $d \equiv 0$ or $1 \pmod 4$ there are only two remaining cases; $d = -3$ and $d = -4$. If $d = -3$ then the equation $t^2 - du^2 = 4$ has exactly six integer solutions; $\langle t, u\rangle \in \{\langle \pm 2, 0\rangle, \langle \pm 1, \pm 1\rangle\}$. If $d = -4$ then the equation $t^2 - du^2 = 4$ has exactly four integer solutions; $\langle t, u\rangle \in \{\langle \pm 2, 0\rangle, \langle 0, \pm 1\rangle\}$.

Hence by Lemma 5.5, the number of automorphs of Q is two if $d < -4$, six if $d = -3$ and four if $d = -4$. This agrees with w in (201). □

For $d > 0$ the situation is quite different. Let us recall that the equation $t^2 - du^2 = 4$ (in integers t, u) is called *Pell's equation*, and it has an infinite number of solutions:

Theorem 5.7. *If $d > 0$ then there are infinitely many $\langle t, u\rangle \in \mathbb{Z}^2$ satisfying $t^2 - du^2 = 4$. We set $\varepsilon_d := \frac{1}{2}(t_0 + u_0\sqrt{d})$ where $\langle t_0, u_0\rangle$ is the solution to $t^2 - du^2 = 4$ with $t_0 > 0$, $u_0 > 0$ for which u_0 is minimal. Then all solutions to $t^2 - du^2 = 4$ are given by*[12]

$$\tfrac{1}{2}(t + u\sqrt{d}) = \pm\varepsilon_d^n, \qquad n \in \mathbb{Z}. \tag{207}$$

Proof. Cf., e.g., [53, §7.8], [39, vol I (Satz 105, Satz 108, Satz 111)], or almost any textbook on basic number theory. □

Note that $\varepsilon_d > 1$, since $\varepsilon_d = \frac{1}{2}(t_0 + u_0\sqrt{d}) \geq \frac{1}{2}(1 + \sqrt{d}) > 1$.

We now turn to the question of the total number of representations of a positive integer n by a representative set of forms of given discriminant d. This question was first answered (implicitly, at least) in the classical theory of quadratic forms, developed by Lagrange and further by Gauss. Recall Definition 5.2.

Definition 5.8. A representation $\langle x, y\rangle$ of a positive integer n by a quadratic form $Q = [a, b, c]$ with $a > 0$ is said to be a *primary* representation if either $d < 0$; *or* $d > 0$ and

$$1 \leq \frac{x - \theta' y}{x - \theta y} < \varepsilon_d^2 \quad \text{and} \quad x - \theta y > 0 \qquad \big(\text{with } \theta = \tfrac{-b+\sqrt{d}}{2a},\ \theta' = \tfrac{-b-\sqrt{d}}{2a} \text{ as in (191)}\big). \tag{208}$$

Note that if $\langle x, y\rangle \in \mathbb{Z}^2$ be a representation of n by Q and if $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is an automorph of Q, then also $\langle \alpha x + \beta y, \gamma x + \delta y\rangle$ is a representation of n by Q. The point of the concept of primary representation is the following:

Lemma 5.8. *Let $\langle x, y\rangle$ be a representation of $n > 0$ by a primitive quadratic form $Q = [a, b, c]$ with $a > 0$ and $d > 0$. Then there is exactly one primary representation of n in the set*

$$\Big\{\langle \alpha x + \beta y, \gamma x + \delta y\rangle \ :\ \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \text{ is an automorph of } Q\Big\}.$$

Proof. Assume $Q = \left(\begin{smallmatrix} a & b/2\\ b/2 & c\end{smallmatrix}\right)$. By Lemma 5.5 and Theorem 5.7 the automorphs of Q are given exactly by $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix} \frac{1}{2}(t-bu) & -cu\\ au & \frac{1}{2}(t+bu)\end{smallmatrix}\right)$ where $\frac{1}{2}(t + u\sqrt{d}) = s\varepsilon_d^n$ for some $n \in \mathbb{Z}$ and

[12] When we say that t, u are determined by (207) we use the fact that if $q_1 + q_2\sqrt{d} = q_3 + q_4\sqrt{d}$ with $q_j \in \mathbb{Q}$ then $q_1 = q_3$ and $q_2 = q_4$. This follows from the fact that $\sqrt{d}$ is irrational.

some choice of sign $s \in \{1,-1\}$. For this $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ we compute that

$$\alpha x+\beta y-\theta(\gamma x+\delta y)=\tfrac{t-u\sqrt{d}}{2}x+\tfrac{du+bt-(t+bu)\sqrt{d}}{4a}y=\tfrac{t-u\sqrt{d}}{2}\big(x-\theta y\big)=s\varepsilon_d^{-n}(x-\theta y)$$

and similarly $\alpha x+\beta y-\theta'(\gamma x+\delta y)=s\varepsilon_d^{n}(x-\theta' y)$. Hence

$$\frac{\alpha x+\beta y-\theta'(\gamma x+\delta y)}{\alpha x+\beta y-\theta(\gamma x+\delta y)}=\varepsilon_d^{2n}\frac{x-\theta' y}{x-\theta y}.$$

Note also that $(x-\theta y)(x-\theta' y)=a^{-1}Q(x,y)=a^{-1}n>0$ (cf. (191)) and thus $\frac{x-\theta' y}{x-\theta y}>0$. Hence there is a unique choice of n for which $1\le \varepsilon_d^{2n}\frac{x-\theta' y}{x-\theta y}<\varepsilon_d^2$. For this n, there is a unique choice of $s\in\{1,-1\}$ which makes $\alpha x+\beta y-\theta(\gamma x+\delta y)=s\varepsilon_d^{-n}(x-\theta y)$ positive. In other words, there is a unique choice of s,n for which $\langle \alpha x+\beta y, \gamma x+\delta y\rangle$ is a primary representation of n. □

Definition 5.9. From now on in this section we fix S_d to be *a set which contains exactly one form from each equivalence class of positive definite or indefinite, primitive forms of discriminant d,* and chosen so that $a>0$ *for every form in S_d.* (The requirement $a>0$ is necessarily true if $d<0$; if $d>0$ then $a>0$ can be made to hold by choosing the forms in S_d appropriately; cf. (197) and recall that every form of discriminant $d>0$ is indefinite.)

In particular we have $\#S_d=h(d)$, cf. Definition 5.6.

Definition 5.10. For each $n\in\mathbb{Z}^+$ and any primitive quadratic form $Q=[a,b,c]$ with $a>0$ we let $R(n;Q)$ be the number of primary representations of n by Q and let $R'(n;Q)\le R(n;Q)$ be the number of *proper* such representations. We then set

$$R(n;d)=\sum_{Q\in S_d}R(n;Q)\qquad\text{and}\qquad R'(n;d)=\sum_{Q\in S_d}R'(n;Q).$$

(One easily checks, using Lemma 5.8 if $d>0$, that $R(n;Q)$ remains unchanged if Q is replaced by any equivalent form with $a>0$. Similarly for $R'(n;Q)$. Hence $R(n;d)$ and $R'(n;d)$ are independent of the choice of S_d.)

A fundamental result in the theory of quadratic forms, and our main tool in proving Dirichlet's class number formula, is the following:

Theorem 5.9. *If $n>0$ and $(n,d)=1$ then*

$$R(n;d)=w\sum_{m|n}\left(\frac{d}{m}\right) \tag{209}$$

where w is given by (201) *if $d<0$ and $w=1$ if $d>0$.*

To prove Theorem 5.9 we first prove two lemmas:

Lemma 5.10. *There is a w-to-1 map from the set*

$$\big\{\langle Q,x,y\rangle \,:\, Q\in S_d \text{ and } \langle x,y\rangle \text{ is a primary, proper representation of } n \text{ by } Q\big\},$$

onto the set

$$\big\{\ell\in\{0,1,\dots,2n-1\} \,:\, \ell^2\equiv d\,(mod\ 4n)\big\}.$$

Proof. Call the first set M_1 and the second set M_2. Let $\langle Q,x,y\rangle\in M_1$ be given, and write $Q=[a,b,c]$. Two integers r,s satisfy $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)\in \mathrm{SL}(2,\mathbb{Z})$ if and only if $xs-yr=1$. But $\gcd(x,y)=1$ (since $\langle x,y\rangle$ is a proper representation of n) and hence the equation $xs-yr=1$ is known to have a solution $\langle r,s\rangle=\langle r_0,s_0\rangle\in\mathbb{Z}^2$, and the general solution is then given by $\langle r,s\rangle=\langle r_0+hx,s_0+hy\rangle$ $(h\in\mathbb{Z})$. From (197), using $Q(x,y)=n$ we see that for any such $\langle r,s\rangle$, the map $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)$ transforms Q to $[n,\ell,m]$ where $\ell=2axr+b(xs+yr)+2cys=2axr_0+b(xs_0+yr_0)+2cys_0+2hn$. Hence there is a unique choice of h such that $0\le\ell<2n$. In other words, there is a unique choice of $r,s\in\mathbb{Z}$ such that $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)\in\mathrm{SL}(2,\mathbb{Z})$ and $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)$ transforms Q to a form $[n,\ell,m]$ with $0\le\ell<2n$. Since $[n,\ell,m]\sim Q$ we have $\ell^2-4nm=d$; thus $\ell^2\equiv d\,(\mathrm{mod}\ 4n)$, i.e. $\ell\in M_2$. Hence we have now constructed a map $F:M_1\to M_2$.

We first prove that F is *onto*. Let $\ell\in M_2$ be given. Then there is a unique integer m with $\ell^2-4nm=d$. Now the quadratic form $[n,\ell,m]$ is primitive (since $(n,d)=1$) and has discriminant d and $n>0$; hence it is equivalent to a unique form $Q\in S_d$. This equivalence means that there is some $\left(\begin{smallmatrix} x' & r'\\ y' & s'\end{smallmatrix}\right)\in\mathrm{SL}(2,\mathbb{Z})$ which transforms Q to $[n,\ell,m]$. Then by (197), $\langle x',y'\rangle$ is a representation of n by Q, and thus by Lemma 5.8 there is an automorph $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ of Q for which $\langle x,y\rangle:=\langle\alpha x'+\beta y',\gamma x'+\delta y'\rangle$ is a primary representation of n. Then $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)=\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\left(\begin{smallmatrix} x' & r'\\ y' & s'\end{smallmatrix}\right)\in\mathrm{SL}(2,\mathbb{Z})$ for some r,s, and since $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is an automorph of Q also the map $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)$ transforms Q to $[n,\ell,m]$. Now $F\big(\langle Q,x,y\rangle\big)=\ell$, and we have proved that F is onto.

Finally we prove that F is w-to-1. Let $\langle Q,x,y\rangle\in M_1$ be given and set $\ell=F\big(\langle Q,x,y\rangle\big)$. Now if $\langle Q',x',y'\rangle$ is *any* element in M_1 with $F\big(\langle Q',x',y'\rangle\big)=\ell$ then by the definition of F, both Q and Q' are equivalent with the form $[n,\ell,m]$, where $m=\frac{\ell^2-d}{4n}$. Hence $Q\sim Q'$, and since $Q,Q'\in S_d$ this implies $Q=Q'$. Furthermore in this situation, if r,s,r',s' are the unique integers such that $\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right),\left(\begin{smallmatrix} x' & r'\\ y' & s'\end{smallmatrix}\right)\in\mathrm{SL}(2,\mathbb{Z})$ transform $Q=Q'$ to $[n,\ell,m]$, then $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right):=\left(\begin{smallmatrix} x & r\\ y & s\end{smallmatrix}\right)\left(\begin{smallmatrix} x' & r'\\ y' & s'\end{smallmatrix}\right)^{-1}\in\mathrm{SL}(2,\mathbb{Z})$ is an automorph of Q. Note that $\langle x,y\rangle=\langle\alpha x'+\beta y',\gamma x'+\delta y'\rangle$ and both $\langle x,y\rangle$ and $\langle x',y'\rangle$ are primary representations of n; hence if $d>0$ then by Lemma 5.8 we must have $x'=x$, $y'=y$. We have thus proved that if $d>0$ then F is injective, i.e. 1-to-1. On the other hand if $d<0$ then we conclude that the only possible elements $\langle Q',x',y'\rangle\in M_1$ with $F\big(\langle Q',x',y'\rangle\big)=\ell$ are given by $Q'=Q$ and $\left(\begin{smallmatrix} x'\\ y'\end{smallmatrix}\right)=\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)^{-1}\left(\begin{smallmatrix} x\\ y\end{smallmatrix}\right)$ for some automorph $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ of Q. Conversely one also readily verifies that every automorph $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ of Q in through this formula really gives an element $\langle Q,x',y'\rangle\in M_1$ with $F\big(\langle Q,x',y'\rangle\big)=\ell$, and that distinct automorphs give distinct elements in M_1. Hence F is w-to-1. □

Lemma 5.11. *For every $n \in \mathbb{Z}^+$ with $(n,d)=1$ we have*

$$\#\big\{\ell \in \{0,1,\dots,2n-1\} \,:\, \ell^2 \equiv d \,(mod\ 4n)\big\} = \sum_{\substack{f|n \\ (f\ squarefree)}} \left(\frac{d}{f}\right).$$

Proof. Note $(\ell+2n)^2 = \ell^2 + 4n\ell + 4n^2 \equiv \ell^2 \pmod{4n}$; hence the formula which we wish to prove is equivalent to

$$\#\big\{\ell \in \mathbb{Z}/4n\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ 4n)\big\} = 2 \sum_{\substack{f|n \\ (f \text{ squarefree})}} \left(\frac{d}{f}\right). \tag{210}$$

Let the prime factorization of $4n$ be $4n = p_1^{\alpha_1}\cdots p_r^{\alpha_r}$. Then by the Chinese Remainder Theorem we have

$$\#\big\{\ell \in \mathbb{Z}/4n\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ 4n)\big\} = \prod_{j=1}^{r} \#\big\{\ell \in \mathbb{Z}/p_j^{\alpha_j}\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ p_j^{\alpha_j})\big\} \tag{211}$$

Now fix $j \in \{1,\dots,r\}$ and first assume that p_j is odd. We then claim

$$\#\big\{\ell \in \mathbb{Z}/p_j^{\alpha_j}\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ p_j^{\alpha_j})\big\} = 1 + \left(\tfrac{d}{p_j}\right). \tag{212}$$

Indeed, note that $(d,p_j)=1$ by assumption, and if $\left(\frac{d}{p_j}\right) = -1$ then there is no solution to the congruence equation $\ell^2 \equiv d \,(\mathrm{mod}\ p_j)$ and hence a fortiori there is no solution to $\ell^2 \equiv d \,(\mathrm{mod}\ p_j^{\alpha_j})$, and (212) holds with both sides being zero. Next assume $\left(\frac{d}{p_j}\right) = 1$. Then if $\nu : (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times \to \{0,1,\dots,\phi(p_j^{\alpha_j})-1\}$ is the index function with respect to some fixed primitive root $g \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$, modulo p_j we have $1 = \left(\frac{d}{p_j}\right) \equiv d^{\frac{p_j-1}{2}} \equiv g^{\nu(d)\frac{p_j-1}{2}}$; thus $p_j - 1$ divides $\nu(d)\frac{p_j-1}{2}$, forcing $\nu(d)$ to be even. Since also $\phi(p_j^{\alpha_j})$ is even it follows that the congruence equation $2x \equiv \nu(d) \bmod \phi(p_j^{\alpha_j})$ has exactly two solutions modulo $\phi(p_j^{\alpha_j})$, namely $x \equiv \nu(d)/2$ and $x \equiv (\nu(d)+\phi(p_j^{\alpha_j}))/2$. Writing $\ell \equiv g^x \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$ this means that the equation $\ell^2 \equiv d \bmod p_j^{\alpha_j}$ has exactly two solutions $\ell \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$, and since there are no solutions $\ell \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})$ outside $(\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^\times$ it again follows that (212) holds.

Next note that since $4n$ is divisible by 4, 2 must occur among the primes p_j with corresponding exponent $\alpha_j \geq 2$; we may thus assume $p_1 = 2$ and $\alpha_1 \geq 2$. We claim

$$\#\big\{\ell \in \mathbb{Z}/2^{\alpha_1}\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ 2^{\alpha_1})\big\} = \begin{cases} 2 & \text{if } \alpha_1 = 2 \\ 2\big(1+\big(\frac{d}{2}\big)\big) & \text{if } \alpha_1 \geq 3. \end{cases} \tag{213}$$

The case $\alpha_1 = 2$ is clear by inspection, since $d \equiv 0$ or $1 \,(\mathrm{mod}\ 4)$. (Cf. (174) concerning $\big(\frac{d}{2}\big)$.) Now assume $\alpha_1 \geq 3$. Then $d \equiv 1 \,(\mathrm{mod}\ 4)$ since $(n,d)=1$. If $d \equiv 5 \,(\mathrm{mod}\ 8)$ then the equation $\ell^2 \equiv d$ is insolvable modulo 8, so that (213) holds with both sides being zero. Finally assume $d \equiv 1 \,(\mathrm{mod}\ 8)$. Then by using facts from the proof of Lemma 4.8

we see that there is an *even* number $0 \le w < 2^{\alpha_1-2}$ such that $d \equiv 5^w \bmod 2^{\alpha_1}$, and now $((-1)^\nu 5^{\nu'})^2 \equiv d \bmod 2^{\alpha_1}$ holds if and only if $2\nu' \equiv w \bmod 2^{\alpha_1-2}$; thus there are exactly 4 solutions $\langle \nu, \nu' \rangle$ with $\nu \in \{0,1\}$ and $0 \le \nu' < 2^{\alpha_1-2}$. Hence by Lemma 4.8 the equation $\ell^2 \equiv d \bmod 2^{\alpha_1}$ has exactly 4 solutions $\ell \in (\mathbb{Z}/2^{\alpha_1}\mathbb{Z})^\times$, and since there are no solutions $\ell \in (\mathbb{Z}/2^{\alpha_1}\mathbb{Z})$ outside $(\mathbb{Z}/2^{\alpha_1}\mathbb{Z})^\times$ it again follows that (213) holds.

Now from (211), (212) and (213) we get

$$\#\big\{\ell \in \mathbb{Z}/4n\mathbb{Z} \,:\, \ell^2 \equiv d \,(\mathrm{mod}\ 4n)\big\} = 2\prod_{p|n}\big(1 + \left(\tfrac{d}{p}\right)\big) = 2 \sum_{\substack{f|n \\ (f \text{ squarefree})}} \left(\frac{d}{f}\right),$$

where in the last step we used Proposition 4.29. Hence we have proved (210). □

Proof of Theorem 5.9. Note the number of elements in the first set in Lemma 5.10 equals $R'(n;d)$ (cf. Definition 5.10). Hence by Lemma 5.10 and Lemma 5.11 we have

$$R'(n;d) = w \sum_{\substack{f|n \\ (f \text{ squarefree})}} \left(\tfrac{d}{f}\right). \tag{214}$$

To get from here to $R(n;d)$ we must consider the representations $\langle x, y\rangle$ of n by Q which are *not* proper, i.e. such that $g := \gcd(x,y) \ge 2$. For such $\langle x, y\rangle$, if $Q = [a,b,c]$ then $n = ax^2 + bxy + cy^2$ and thus $g^2 \mid n$, and $\langle \frac{x}{g}, \frac{y}{g}\rangle$ is a proper representation of $\frac{n}{g^2}$ by Q. Also $\langle \frac{x}{g}, \frac{y}{g}\rangle$ is primary if and only if $\langle x, y\rangle$ is primary, by inspection in Definition 5.8. In fact for any $g \in \mathbb{Z}^+$ with $g^2 \mid n$, the map $\langle x, y\rangle \mapsto \langle \frac{x}{g}, \frac{y}{g}\rangle$ gives a bijection from the set of primary representations $\langle x, y\rangle$ of n by Q which satisfy $\gcd(x,y) = g$, onto the set of primary proper representations of $\frac{n}{g^2}$ by Q. Hence for every $Q \in S_d$ we have

$$R(n;Q) = \sum_{g^2|n} \#\big\{\langle x, y\rangle \,:\, \langle x, y\rangle \text{ a primary repr. of } n \text{ by } Q,\ \gcd(x,y) = g\big\} = \sum_{g^2|n} R'(\frac{n}{g^2}, Q).$$

Hence using (214) with n/g^2 in place of n we get

$$R(n;Q) = w \sum_{g^2|n} \sum_{\substack{f|\frac{n}{g^2} \\ (f \text{ squarefree})}} \left(\tfrac{d}{f}\right).$$

Here for any $g \in \mathbb{Z}^+$ with $g \mid n$ we have $\left(\frac{d}{f}\right) = \left(\frac{d}{f}\right)\left(\frac{d}{g}\right)^2 = \left(\frac{d}{fg^2}\right)$, for note that $\left(\frac{d}{g}\right) = \pm 1$ since $g^2 \mid n$ and $(n,d) = 1$. Note also that when g, f run through the above double sum then $m := fg^2$ visits each positive divisor of n exactly once. (For given m, the unique way to pick f, g is to let f be the *squarefree part* of m and g the *square part* of m, viz. let f be the product of all primes $p \mid m$ for which $\mathrm{ord}_p(m)$ is odd, and $g = \sqrt{m/f} \in \mathbb{Z}^+$.) Hence

$$R(n;Q) = w \sum_{m|n} \left(\tfrac{d}{m}\right).$$

□

Having now proved the fundamental Theorem 5.9, we now turn to the proof of Dirichlet's class number formula, Theorem 5.4.

Proof of Theorem 5.4. The idea is to determine, from Theorem 5.9, the average of $R(n;d)$ as n varies. It is convenient (and it suffices for our purpose) to limit oneself to values of n that are relatively prime to d. By Theorem 5.9 we have, for any $N \geq 1$:

$$w^{-1} \sum_{\substack{1\leq n\leq N\\(n,d)=1}} R(n;d) = \sum_{\substack{1\leq n\leq N\\(n,d)=1}} \sum_{m_1|n} \left(\tfrac{d}{m_1}\right),$$

and substituting here $n = m_1 m_2$ we get

$$(215) \quad = \sum_{\substack{m_1,m_2\geq 1\\ m_1m_2\leq N\\ (m_1m_2,d)=1}} \left(\tfrac{d}{m_1}\right) = \sum_{1\leq m_1\leq\sqrt{N}} \left(\tfrac{d}{m_1}\right) \sum_{\substack{1\leq m_2\leq N/m_1\\(m_2,d)=1}} 1 + \sum_{\substack{1\leq m_2<\sqrt{N}\\(m_2,d)=1}} \sum_{\sqrt{N}<m_1\leq N/m_2} \left(\tfrac{d}{m_1}\right),$$

since the first sum comprises all pairs m_1, m_2 for which $m_1 \leq \sqrt{N}$ and the second sum all pairs for which $m_1 > \sqrt{N}$ (we also used the fact that $\left(\frac{d}{m_1}\right) = 0$ whenever $(d, m_1) \neq 1$, cf. Proposition 4.26). We have

$$\sum_{\substack{1\leq m_2\leq N/m_1\\(m_2,d)=1}} 1 = \phi(|d|)\Big(\frac{N}{m_1|d|} + O(1)\Big).$$

(Here the implied constant is absolute, but *in the following we allow the implied constants to depend on d but* ***not on*** N). Hence the first double sum in (215) is

$$N\frac{\phi(|d|)}{|d|} \sum_{1\leq m_1\leq\sqrt{N}} \tfrac{1}{m_1}\left(\tfrac{d}{m_1}\right) + O(\sqrt{N}).$$

Furthermore since $\left(\frac{d}{\cdot}\right)$ is a non-principal character modulo $|d|$ (cf. Lemma 4.36), we have $\sum_{n=1}^{|d|}\left(\frac{d}{n}\right) = 0$ (cf. Lemma 3.13) and hence $\left|\sum_{A<n\leq B}\left(\frac{d}{n}\right)\right| < |d|$ for all $A < B$. Hence the second double sum in (215) is $O(\sqrt{N})$, and we conclude:

$$w^{-1} \sum_{\substack{1\leq n\leq N\\(n,d)=1}} R(n;d) = N\frac{\phi(|d|)}{|d|} \sum_{1\leq m\leq\sqrt{N}} \tfrac{1}{m}\left(\tfrac{d}{m}\right) + O(\sqrt{N}).$$

We know from Example 3.5 that the series $L(1, \left(\frac{d}{\cdot}\right)) = \sum_{m=1}^{\infty} \frac{1}{m}\left(\frac{d}{m}\right)$ converges. In fact, using partial summation one obtains the tail estimate $\sum_{m>\sqrt{N}} \frac{1}{m}\left(\frac{d}{m}\right) = O(N^{-\frac{1}{2}})$. (Detailed proof: Using (105) with $a_m = 0$ for $m \leq \sqrt{N}$ and $a_m = \left(\frac{d}{m}\right)$ for $m > \sqrt{N}$ we get

$\left|\sum_{m>\sqrt{N}} \frac{1}{m}\left(\frac{d}{m}\right)\right| \le \int_{\sqrt{N}}^{\infty} |d|x^{-2}\,dx = O(N^{-\frac{1}{2}}).)$ Hence

$$w^{-1} \sum_{\substack{1\le n\le N\\(n,d)=1}} R(n;d) = N\frac{\phi(|d|)}{|d|} L(1,\left(\tfrac{d}{\cdot}\right)) + O(\sqrt{N}).$$

Dividing with N and letting $N\to\infty$ this gives

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{1\le n\le N\\(n,d)=1}} R(n;d) = w\frac{\phi(|d|)}{|d|} L(1,\left(\tfrac{d}{\cdot}\right)). \tag{216}$$

The next step is to evaluate the average of $R(n;d)$ from its original definition, Def. 5.10. In fact we will evaluate

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{1\le n\le N\\(n,d)=1}} R(n;Q),$$

for any fixed quadratic form $Q=[a,b,c]\in S_d$, and this limit will turn out to be independent of f. In view of Def. 5.10, comparison of the two limits will give a relation between $\#S_d = h(d)$ and $L(1,\left(\frac{d}{\cdot}\right))$.

First assume $d<0$. Then

$$\sum_{\substack{1\le n\le N\\(n,d)=1}} R(n;Q)$$

is the number of pairs of integers x,y satisfying

$$0 < ax^2+bxy+cy^2 \le N, \qquad (ax^2+bxy+cy^2, d) = 1.$$

The second condition limits x,y to certain pairs of residue classes to the modulus $|d|$. The number of such pairs is easily computed:

Lemma 5.12.

$$\#\big\{\langle x_0,y_0\rangle \in (\mathbb{Z}/d\mathbb{Z})^2 \,:\, ax_0^2+bx_0y_0+cy_0^2 \in (\mathbb{Z}/d\mathbb{Z})^\times\big\} = |d|\phi(|d|). \tag{217}$$

Proof. Let the prime factorization of $|d|$ be $|d| = p_1^{\alpha_1}\cdots p_r^{\alpha_r}$. Then by the Chinese Remainder Theorem, the above cardinality equals

$$\prod_{j=1}^{r} \#\big\{\langle x_0,y_0\rangle \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^2 \,:\, p_j \nmid ax_0^2+bx_0y_0+cy_0^2\big\}.$$

Fix some $j\in\{1,\dots,r\}$. If $a\equiv c\equiv 0 \bmod p_j$ then since p_j divides $d=b^2-4ac$ we would also get $b\equiv 0 \bmod p_j$, contradicting the fact that Q is primitive. Hence either $p_j\nmid a$ or $p_j\nmid c$.

First assume $p_j > 2$. Then if $p_j \nmid a$ we have

$$p_j \nmid ax_0^2 + bx_0y_0 + cy_0^2 \Longleftrightarrow p_j \nmid 4a(ax_0^2 + bx_0y_0 + cy_0^2) \Longleftrightarrow p_j \nmid (2ax_0 + by_0)^2 - dy_0^2 \\ \Longleftrightarrow p_j \nmid 2ax_0 + by_0,$$

where in the last step we used the fact that $p_j \mid d$. For any given $y_0 \in \mathbb{Z}/p_j^{\alpha_j}\mathbb{Z}$ the relation $p_j \nmid 2ax_0 + by_0$ holds if and only if x_0 avoids a particular residue class mod p_j (since $p_j \nmid 2a$); hence it holds for exactly $p_j^{\alpha_j - 1}(p_j - 1)$ of all $x_0 \in \mathbb{Z}/p_j^{\alpha_j}\mathbb{Z}$. Hence

$$\#\big\{\langle x_0, y_0\rangle \in (\mathbb{Z}/p_j^{\alpha_j}\mathbb{Z})^2 \,:\, p_j \nmid ax_0^2 + bx_0y_0 + cy_0^2\big\} = p_j^{\alpha_j} p_j^{\alpha_j - 1}(p_j - 1) = p_j^{\alpha_j}\phi\big(p_j^{\alpha_j}\big). \tag{218}$$

The same formula also holds in the case $p_j \nmid c$, by a completely symmetric argument.

Next assume $p_j = 2$. Then $2 \mid d$ and $2 \mid b$, and $2 \nmid ax_0^2 + bx_0y_0 + cy_0^2$ holds if and only if $ax_0 + cy_0 \equiv 1 \pmod 2$. Since at least one of a and c is odd we see that this holds for exactly $2^{\alpha_j}2^{\alpha_j - 1} = 2^{\alpha_j}\phi(2^{\alpha_j})$ pairs $\langle x_0, y_0\rangle \in (\mathbb{Z}/2^{\alpha_j}\mathbb{Z})^2$, i.e. (218) holds also when $p_j = 2$.

Multiplying (218) over all $j \in \{1, \dots, r\}$ we obtain (217). □

Continuing onward with the proof of Theorem 5.4, it now suffices to consider the number of pairs of integers x, y satisfying

$$0 < ax^2 + bxy + cy^2 \le N, \qquad x \equiv x_0,\ y \equiv y_0 \,(\mathrm{mod}\ |d|) \tag{219}$$

for some fixed integers x_0, y_0. The first inequality expresses that the point (x, y) is in an ellipse with center at the origin, and as $N \to \infty$ this ellipse expands uniformly. Using (190) one computes that the area of the ellipse is

$$\frac{\pi}{2a\sqrt{|d|}} \cdot 4aN = \frac{2\pi}{|d|^{\frac12}} N.$$

It follows that the number of points satisfying (219) is asymptotic to $|d|^{-2}\frac{2\pi}{|d|^{\frac12}}N$ as $N \to \infty$. (Cf. Remark 5.1 below.) We have to multiply this by $|d|\phi(|d|)$ to allow for the various possibilities for x_0, y_0, cf. Lemma 5.12. Thus the conclusion is that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{1 \le n \le N \\ (n,d)=1}} R(n; Q) = \frac{\phi(|d|)}{|d|} \frac{2\pi}{|d|^{\frac12}}.$$

Comparison with (215) gives, since $R(n; d) = \sum_{Q \in S_d} R(n; Q)$ and $\#S_d = h(d)$:

$$h(d) = \frac{w|d|^{\frac12}}{2\pi} L\big(1, \big(\tfrac{d}{\cdot}\big)\big) \qquad \text{for } d < 0.$$

Next assume $d > 0$. Arguing as before, we need to know the number of pairs of integers x, y satisfying

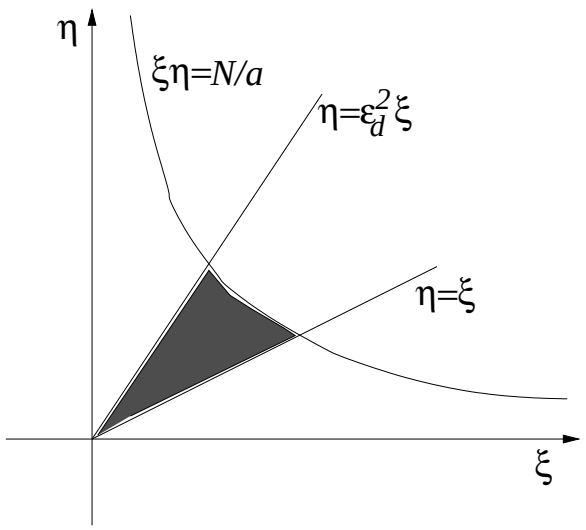
$$0 < ax^2 + bxy + cy^2 \le N, \qquad x - \theta y > 0, \qquad 1 \le \frac{x - \theta' y}{x - \theta y} < \varepsilon_d^2, \tag{220}$$

and

$$x \equiv x_0, \qquad y \equiv y_0 \qquad (\mathrm{mod}\ d). \tag{221}$$

To better understand the inequalities in (220) we make a (non-singular) linear change of variables:

$$\begin{cases} \xi = x - \theta y \\ \eta = x - \theta' y, \end{cases} \qquad \text{with} \qquad \frac{\partial(\xi,\eta)}{\partial(x,y)} = \theta - \theta' = \frac{\sqrt{d}}{a}. \tag{222}$$



Then the inequalities in (220) take the following form (cf. (191)):

$$0 < \xi\eta \le N/a, \qquad \xi > 0, \qquad \xi \le \eta < \varepsilon_d^2 \xi,$$

and we see that these conditions represent a sector of a hyperbola bounded by two fixed rays through the origin. To compute the area of this sector in the ξ, η-plane, note that the conditions are equivalent to $[0 < \xi \le (N/a)^{\frac12}$ and $\xi \le \eta < \varepsilon_d^2\xi$, $\eta \le N/a\xi]$; hence the area is[13]

$$\int_0^{\varepsilon_d^{-1}(N/a)^{\frac12}} (\varepsilon_d^2\xi - \xi)\, d\xi + \int_{\varepsilon_d^{-1}(N/a)^{\frac12}}^{(N/a)^{\frac12}} \Big(\frac{N}{a\xi} - \xi\Big)\, d\xi$$
$$= (\varepsilon_d^2 - 1)\frac12 \frac{N}{\varepsilon_d^2 a} + \frac{N}{a}\log\varepsilon_d - \frac12\Big(\frac{N}{a} - \varepsilon_d^{-2}\frac{N}{a}\Big) = \frac{N}{a}\log\varepsilon_d.$$

It follows that the area of the corresponding sector in the x, y-plane is $d^{-\frac12} N \log\varepsilon_d$, cf. (222). It follows that the number of integer points satisfying both (220) and (221) is asymptotic

[13] There seems to be a misprint in [15, p. 50, line 4].

to $d^{-2}d^{-\frac{1}{2}}N\log\varepsilon_d$ as $N\to\infty$ (cf. Remark 5.1 below). We should then multiply with $d\phi(d)$ to allow for the choices of x_0, y_0 (cf. Lemma 5.12). This gives

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\substack{1\le n\le N\\(n,d)=1}} R(n;Q)=\frac{\phi(d)}{d}\frac{\log\varepsilon_d}{d^{\frac{1}{2}}}.$$

Comparison with (215) gives (since now $w=1$):

$$h(d)=\frac{d^{\frac{1}{2}}}{\log\varepsilon_d}L(1,\left(\tfrac{d}{\cdot}\right)) \qquad \text{for } d>0.$$

□

Remark 5.1. Two times in the above proof we used the following fact: If $D\subset\mathbb{R}^2$ is a "nice" domain and x_0, y_0, d are any fixed integers with $d\neq 0$, then[14]

$$\#\Big(\big(N^{\frac{1}{2}}D\big)\bigcap\big\{(x,y)\in\mathbb{Z}^2\ :\ x\equiv x_0,\ y\equiv y_0\ (\mathrm{mod}\ |d|)\big\}\Big)\sim |d|^{-2}\mathrm{Area}(D)N \quad \text{as } N\to\infty.$$

(First we used this below (219), for $D=$ the ellipse $\{(x,y)\ :\ ax^2+bxy+cy^2\le 1\}$; note that then $(x,y)\in N^{\frac{1}{2}}D\Leftrightarrow(N^{-\frac{1}{2}}x,N^{-\frac{1}{2}}y)\in D\Leftrightarrow ax^2+bxy+cy^2\le N$; also note that the lower inequality $0<ax^2+bxy+cy^2$ in (219) can be ignored since it excludes at most a single point. Similarly, the second application was with $D=$ the hyperbola sector [(220) with $N=1$].)

Scaling down to $x=|d|x'$, $y=|d|y'$ and replacing D with $|d|^{-1}D$, the above asymptotic relation is seen to follow from the following with $(\alpha,\beta)=|d|^{-1}(x_0,y_0)$:

$$\#\Big(\big(N^{\frac{1}{2}}D\big)\bigcap\big(\mathbb{Z}^2+(\alpha,\beta)\big)\Big)\sim\mathrm{Area}(D)N \quad \text{as } N\to\infty,$$

where $\mathbb{Z}^2+(\alpha,\beta)$ denotes the shifted lattice $\{(x+\alpha,y+\beta)\ :\ (x,y)\in\mathbb{Z}^2\}$. Equivalently (writing $N=\delta^{-2}$):

$$\delta^2\cdot\#\Big(D\bigcap\delta\big(\mathbb{Z}^2+(\alpha,\beta)\big)\Big)\to\mathrm{Area}(D) \quad \text{as } \delta\to 0^+. \tag{223}$$

This can be seen as a statement belonging to the foundations of measure/integration theory and the very definition of "Area(D)": We cover the plane with a net of squares of side δ (very small), so that the central points of the squares are exactly the points in the lattice $\delta\big(\mathbb{Z}^2+(\alpha,\beta)\big)$. Now $A_\delta^-(D)\le\mathrm{Area}(D)\le A_\delta^+(D)$, where

$$A_\delta^-(D)=\text{total area of all squares which are fully inside } D;$$
$$A_\delta^+(D)=\text{total area of all squares which touch } D,$$

and if D is "nice" then both $A_\delta^-(D)$ and $A_\delta^+(D)$ tend to Area(D) as $\delta\to 0^+$. For instance one can prove that this holds whenever D is Jordan measurable, and this is certainly true

[14]Notation: For any $M\subset\mathbb{R}^2$ and $A\in\mathbb{R}$ we write $AM:=\{A(x,y)\ :\ (x,y)\in M\}$.

for our two choices of D! Since the number $\delta^2 \cdot \#\Big(D \bigcap \delta\big(\mathbb{Z}^2 + (\alpha,\beta)\big)\Big)$ is clearly also $\geq A_\delta^-(D)$ and $\leq A_\delta^+(D)$, we conclude that (223) holds.

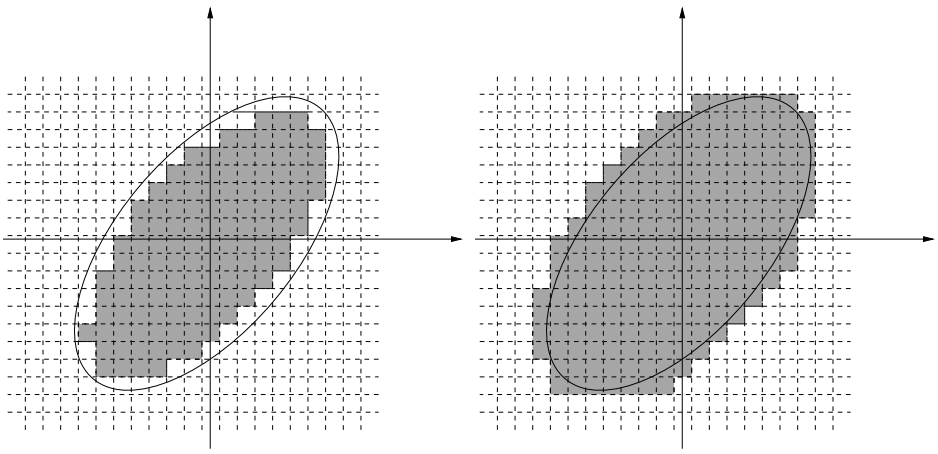


FIGURE 1. For D =the set of points inside the ellipse, $A_\delta^-(D)$ is the area of the marked squares in the left picture, and $A_\delta^+(D)$ is the area of the marked squares in the right picture.

5.3. **Gauss sums (I).** In section 5.4 we will show how to obtain a finite sum formula for $L(1, \left(\frac{d}{\cdot}\right))$ (and thus for $h(d)$). The key fact needed for this is the following formula for a certain so-called *Gauss sum*. (We will discuss more general Gauss sums in Lecture 9.)

Theorem 5.13. *Let d be a fundamental discriminant and $n \in \mathbb{Z}^+$. Then*

$$\sum_{k \in \mathbb{Z}/|d|\mathbb{Z}} \left(\frac{d}{k}\right) e\Big(\frac{nk}{|d|}\Big) = \left(\frac{d}{n}\right) \begin{cases} \sqrt{d} & \text{if } d > 0 \\ i\sqrt{|d|} & \text{if } d < 0. \end{cases} \tag{224}$$

In the following we will use the short-hand notation "$\sqrt{d}$" to denote $\begin{cases} \sqrt{d} & \text{if } d > 0 \\ i\sqrt{|d|} & \text{if } d < 0. \end{cases}$

Proof. (We borrow from Landau [39, Satz 215].) We first prove that if d_1, d_2 are two fundamental discriminants which are relatively prime, and if (224) holds for each of $d = d_1$, $d = d_2$, then (224) also holds for $d = d_1 d_2$. Indeed, if (224) holds for each of $d = d_1$, $d = d_2$, then by multiplying these two identities together we get (using the short-hand notation introduced above):

$$\sum_{\substack{k_1 \in \mathbb{Z}/|d_1|\mathbb{Z} \\ k_2 \in \mathbb{Z}/|d_2|\mathbb{Z}}} \left(\frac{d_1}{k_1}\right)\left(\frac{d_2}{k_2}\right) e\Big(\frac{nk_1}{|d_1|}\Big) e\Big(\frac{nk_2}{|d_2|}\Big) = \left(\frac{d_1}{n}\right)\left(\frac{d_2}{n}\right)\sqrt{d_1}\sqrt{d_2}$$

Here multiply both sides with $\left(\frac{d_1}{|d_2|}\right)\left(\frac{d_2}{|d_1|}\right)$ and use Proposition 4.28 and Proposition 4.29. This gives

$$\sum_{\substack{k_1\in\mathbb{Z}/|d_1|\mathbb{Z}\\ k_2\in\mathbb{Z}/|d_2|\mathbb{Z}}}\left(\frac{d_1}{k_1|d_2|}\right)\left(\frac{d_2}{k_2|d_1|}\right)e\Big(\frac{n(k_1|d_2|+k_2|d_1|)}{|d_1d_2|}\Big)=\left(\frac{d_1}{|d_2|}\right)\left(\frac{d_2}{|d_1|}\right)\left(\frac{d_1d_2}{n}\right)\sqrt{d_1}\sqrt{d_2}.$$

In the left hand side we have $\left(\frac{d_1}{k_1|d_2|}\right)\left(\frac{d_2}{k_2|d_1|}\right)=\left(\frac{d_1}{k_1|d_2|+k_2|d_1|}\right)\left(\frac{d_2}{k_1|d_2|+k_2|d_1|}\right)=\left(\frac{d_1d_2}{k_1|d_2|+k_2|d_1|}\right)$, by Proposition 4.32 and Proposition 4.28. Furthermore, since $(d_1,d_2)=1$ we have that when k_1 runs through $\mathbb{Z}/|d_1|\mathbb{Z}$ and k_2 runs through $\mathbb{Z}/|d_2|\mathbb{Z}$ then $k_1|d_2|+k_2|d_1|$ runs through $\mathbb{Z}/|d_1d_2|\mathbb{Z}$.[15] Hence the above left hand side equals $\sum_{k\in\mathbb{Z}/|d_1d_2|\mathbb{Z}}\left(\frac{d_1d_2}{k}\right)e\Big(\frac{nk}{|d_1d_2|}\Big)$. Hence to prove our claim that (224) holds for $d=d_1d_2$ it now only remains to show that $\left(\frac{d_1}{|d_2|}\right)\left(\frac{d_2}{|d_1|}\right)\sqrt{d_1}\sqrt{d_2}=\sqrt{d_1d_2}$. But since $\sqrt{d_1}\sqrt{d_2}=\begin{cases}-\sqrt{d_1d_2} & \text{if } d_1<0, d_2<0\\ \sqrt{d_1d_2} & \text{otherwise,}\end{cases}$ it suffices to show that $\left(\frac{d_1}{|d_2|}\right)\left(\frac{d_2}{|d_1|}\right)=\begin{cases}-1 & \text{if } d_1<0, d_2<0;\\ 1 & \text{otherwise.}\end{cases}$ Since $(d_1,d_2)=1$ we may without loss of generality assume that d_1 is odd, thus $d_1\equiv 1 \pmod 4$. Then by Proposition 4.30 we have $\left(\frac{d_1}{|d_2|}\right)\left(\frac{d_2}{|d_1|}\right)=\left(\frac{|d_2|}{|d_1|}\right)\left(\frac{d_2}{|d_1|}\right)$, and if $d_2>0$ this equals $\left(\frac{d_2}{|d_1|}\right)^2=1$, whereas if $d_2<0$ then it equals, by (165): $\left(\frac{|d_2|}{|d_1|}\right)(-1)^{\frac{|d_1|-1}{2}}\left(\frac{|d_2|}{|d_1|}\right)=(-1)^{\frac{|d_1|-1}{2}}=\begin{cases}1 & \text{if } d_1>0\\ -1 & \text{if } d_1<0\end{cases}$ (since $d_1\equiv 1 \pmod 4$). This completes the proof of our claim that if (224) holds for each of $d=d_1$, $d=d_2$, then (224) also holds for $d=d_1d_2$.

Now to prove (224) for a general fundamental discriminant we can do the following reduction: Let $p_1,\dots,p_r$ be the odd primes which divide d; set $p_j'=(-1)^{\frac{p_j-1}{2}}p_j$ and define d_2 so that $d=d_2\prod_{j=1}^r p_j'$. Then since d is a fundamental discriminant we must have $d_2\in\{1,-4,-8,8\}$. Note that d_2 and each p_j' are fundamental discriminants; and they are pairwise coprime. Hence by the multiplicativity fact which we proved earlier, it now suffices to prove that (224) holds for $d=d_2$ and for each $d=p_j'$.

Thus, it now suffices to prove that (224) holds for $d\in\{-4,-8,8\}$ as well as for $d=(-1)^{\frac{p-1}{2}}p$ where p is an arbitrary odd prime. We may also note that it suffices to treat the two cases $(n,d)>1$ and $n=1$, respectively. Indeed, in the remaining case, $n\neq 1$ and

[15] Cf., e.g., [39, Vol. I, Satz 73] or [30, Thm. 61]. If you did not know this fact, you may like to try to prove it as an exercise, using the Chinese Remainder Theorem.

$(n,d)=1$, we have, using $\left(\frac{d}{k}\right)=\left(\frac{d}{n}\right)^2\left(\frac{d}{k}\right)=\left(\frac{d}{n}\right)\left(\frac{d}{nk}\right)$:

$$\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{nk}{|d|}\Big)=\left(\frac{d}{n}\right)\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{nk}\right)e\Big(\frac{nk}{|d|}\Big)=\left(\frac{d}{n}\right)\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{k}{|d|}\Big),$$

where the last step follows since nk runs through $\mathbb{Z}/|d|\mathbb{Z}$ when k runs through $\mathbb{Z}/|d|\mathbb{Z}$. If we assume that (224) holds for $n=1$ then we may continue:

$$=\left(\frac{d}{n}\right)\left(\frac{d}{1}\right)\begin{Bmatrix}\sqrt{d} & \text{if } d>0\\ i\sqrt{|d|} & \text{if } d<0\end{Bmatrix}=\left(\frac{d}{n}\right)\begin{cases}\sqrt{d} & \text{if } d>0\\ i\sqrt{|d|} & \text{if } d<0,\end{cases}$$

i.e. (224) holds also for our n with $n\neq 1$, $(n,d)=1$.

The cases $d\in\{-4,-8,8\}$ are now treated by direct computation: Compare the table on p. 70. If $d=-4$ then

$$\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{nk}{|d|}\Big)=i^n-(-i)^n=\begin{cases}0 & \text{if } 2\mid n\\ 2i & \text{if } n=1,\end{cases}$$

and if $d=\pm 8$ then

$$\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{nk}{|d|}\Big)=e\Big(\frac{n}{8}\Big)\mp e\Big(\frac{3n}{8}\Big)-e\Big(\frac{5n}{8}\Big)\pm e\Big(\frac{7n}{8}\Big)=e\Big(\frac{n}{8}\Big)\big(1\mp i^n-i^{2n}\pm i^{3n}\big)$$

$$=e\Big(\frac{n}{8}\Big)\begin{Bmatrix}0 & \text{if } 2\mid n\\ 2-2i & \text{if } n=1,d=8\\ 2+2i & \text{if } n=1,d=-8\end{Bmatrix}=\begin{cases}0 & \text{if } 2\mid n\\ \sqrt{8} & \text{if } n=1,d=8\\ i\sqrt{8} & \text{if } n=1,d=-8.\end{cases}$$

In all cases this agrees with the right hand side of (224).

It now remains to treat the case $d=(-1)^{\frac{p-1}{2}}p$ where p is an odd prime. If $(n,d)>1$ then actually $p\mid n$ and thus $e(nk/|d|)=1$ for all k, and then using Lemma 3.13 we get

$$\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{nk}{|d|}\Big)=\sum_{k\in\mathbb{Z}/p\mathbb{Z}}\left(\frac{d}{k}\right)=0,$$

since $\chi=\left(\frac{d}{\cdot}\right)$ is a nonprincipal character modulo p, cf. Lemma 4.36. Hence it only remains to treat the case $n=1$. In this case we have, using Proposition 4.30,

$$\sum_{k\in\mathbb{Z}/|d|\mathbb{Z}}\left(\frac{d}{k}\right)e\Big(\frac{nk}{|d|}\Big)=\sum_{k\in\mathbb{Z}/p\mathbb{Z}}\left(\frac{k}{|d|}\right)e\Big(\frac{k}{p}\Big)=\sum_{k\in\mathbb{Z}/p\mathbb{Z}}\left(\frac{k}{p}\right)e\Big(\frac{k}{p}\Big)=\sum_{R}e\Big(\frac{R}{p}\Big)-\sum_{N}e\Big(\frac{N}{p}\Big),$$

where R runs through all the quadratic residues in $(\mathbb{Z}/p\mathbb{Z})^\times$ and N runs through all the quadratic nonresidues in $(\mathbb{Z}/p\mathbb{Z})^\times$. Since $\sum_{k\in\mathbb{Z}/p\mathbb{Z}} e\Big(\frac{k}{p}\Big) = 0$ we can continue:

$$= 1 + 2\sum_R e\Big(\frac{R}{p}\Big) = \sum_{x=0}^{p-1} e\Big(\frac{x^2}{p}\Big),$$

since x^2 takes the value 0 once and gives each quadratic residue in $(\mathbb{Z}/p\mathbb{Z})^\times$ twice. Now the desired equality (224) follows from the case $N = p$ of the following (famous) result. □

Theorem 5.14. *For any positive integer N we have*

$$\sum_{n=0}^{N-1} e\big(n^2/N\big) = \begin{cases} (1+i)N^{\frac12} & \text{if } N \equiv 0 \,(mod\ 4), \\ N^{\frac12} & \text{if } N \equiv 1 \,(mod\ 4), \\ 0 & \text{if } N \equiv 2 \,(mod\ 4), \\ iN^{\frac12} & \text{if } N \equiv 3 \,(mod\ 4). \end{cases} \tag{225}$$

The proof of this formula (for N =a prime) is one of Gauss' many great achievements, and he obtained it only after many and varied unsuccessful attempts. Since then several proofs have been given, based on a variety of different methods. The following proof is due to Dirichlet (1835) and from today's perspective it can be seen as "easy", in that it is a fairly direct application of *Poisson's summation formula,* which is an extremely useful tool in a analytic number theory, and very often used:

Lemma 5.15. ***Poisson's summation formula.*** *For any "nice" $f : \mathbb{R} \to \mathbb{C}$ we have*

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n), \qquad \text{where} \quad \widehat{f}(y) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i y x}\, dx. \tag{226}$$

For example (226) *holds if $f \in C^1(\mathbb{R})$ and both $f(x) = O(|x|^{-1-\varepsilon})$ and $f'(x) = O(|x|^{-1-\varepsilon})$ as $|x| \to \infty$.*

It is important to remember not only the formula (226) itself, but also the (quite simple!) proof, since one often has to modify the formula in some way or other to accommodate functions f which are "not quite so nice". For example this happens already in the proof of Theorem 5.14 below!

Proof of Lemma 5.15. Set

$$\varphi(x) = \sum_{n\in\mathbb{Z}} f(x+n).$$

From our assumptions it follows that both the sums $\sum_{n\in\mathbb{Z}} f(x+n)$ and $\sum_{n\in\mathbb{Z}} f'(x+n)$ are uniformly (absolutely) convergent for x in any bounded interval; hence by standard facts in analysis we have $\varphi \in C^1(\mathbb{R})$ (cf. [59, Thms. 7.17, 7.17]). From the definition we see that

φ is periodic with period 1: $\varphi(x+1) = \varphi(x)$. Hence the Fourier series of φ is absolutely convergent (cf., e.g. [37, Theorem I.6.3]) and converges to $\varphi(x)$ at every point $x \in \mathbb{R}$:

$$\varphi(x) = \sum_{m\in\mathbb{Z}} \widehat{\varphi}(m) e(mx) \qquad \Big(\text{with} \quad \widehat{\varphi}(m) = \int_0^1 \varphi(t) e(-mt)\, dt\Big). \tag{227}$$

Note here that

$$\begin{aligned} \widehat{\varphi}(m) = \int_0^1 \varphi(t) e(-mt)\, dt &= \int_0^1 \sum_{n\in\mathbb{Z}} f(n+t) e(-mt)\, dt = \sum_{n\in\mathbb{Z}} \int_0^1 f(n+t) e(-mt)\, dt \\ &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(t) e(-mt)\, dt = \int_{-\infty}^{\infty} f(t) e(-mt)\, dt = \widehat{f}(m). \end{aligned} \tag{228}$$

(In the third step we used the fact that $e(-m(t-n)) = e(-mt)$.) Here the change of order of summation and integration is justified by the uniform absolute convergence of $\sum_{n\in\mathbb{Z}} f(n+t)e(-mt)$ over $t \in [0,1]$, and all integrals and sums are absolutely convergent. Applying now the formula (227) with $x = 0$ we obtain (226). □

Proof of Theorem 5.14. (Cf. Davenport Chapter 2.) We would like to apply Poisson's summation formula with

$$f(x) = \begin{cases} e(x^2/N) & \text{if } 0 \le x < N \\ 0 & \text{else.} \end{cases}$$

However this function is not continuous, so that Lemma 5.15 does not apply directly. Instead we follow the proof of Lemma 5.15. Thus we set

$$\varphi(x) = \sum_{n\in\mathbb{Z}} f(x+n) = \sum_{-x\le n<N-x} e\big((x+n)^2/N\big).$$

This function φ is obviously periodic with period 1, and C^1 except at integer points $x \in \mathbb{Z}$ (in fact $\varphi \in C(\mathbb{R})$, but $\varphi'(x)$ has jump discontinuities at integer points x). However, since φ is *piecewise* C^1 we still have the following identity for the Fourier series, at any point $x \in \mathbb{R}$ (cf., e.g. [37, Cor. 2.2] or [69, 13.232]):

$$\frac{1}{2}(\varphi(x-) + \varphi(x+)) = \lim_{M\to\infty} \sum_{|m|\le M} \widehat{\varphi}(m) e(mx).$$

In particular, since $\varphi(0-) = \varphi(0+) = \sum_{n\in\mathbb{Z}/N\mathbb{Z}} e(n^2/N)$ we get

$$\sum_{n\in\mathbb{Z}/N\mathbb{Z}} e(n^2/N) = \lim_{M\to\infty} \sum_{|m|\le M} \widehat{\varphi}(m).$$

Furthermore the computation (228) is valid for our f (since our f is of compact support and piecewise continuous); thus

$$\widehat{\varphi}(m) = \widehat{f}(m) = \int_0^N e\big(x^2/N - mx\big)\,dx = N\int_0^1 e\big(N((y-\tfrac{1}{2}m)^2 - \tfrac{1}{4}m^2)\big)\,dy,$$

where in the last step we substituted $x = Ny$ and then completed the square. Hence

$$\sum_{n\in\mathbb{Z}/N\mathbb{Z}} e(n^2/N) = N\lim_{M\to\infty}\sum_{|m|\le M} e\big(-\tfrac{1}{4}Nm^2\big)\int_{-\frac{1}{2}m}^{1-\frac{1}{2}m} e\big(Ny^2\big)\,dy.$$

The value of $e\big(-\frac{1}{4}Nm^2\big)$ is 1 if m is even, and is i^{-N} if m is odd. We therefore divide the sum over m into two parts, according as m is even or odd, and we put $m = 2\mu$ or $2\mu+1$ ($\mu\in\mathbb{Z}$) as the case may be. This gives

$$S = N\lim_{M\to\infty}\Big(\sum_{|2\mu|\le M}\int_{-\mu}^{1-\mu} e(Ny^2)\,dy + i^{-N}\sum_{|2\mu+1|\le M}\int_{-\mu-\frac{1}{2}}^{-\mu+\frac{1}{2}} e(Ny^2)\,dy\Big) \tag{229}$$

Here each series of integrals fits together to give $\int_{-\infty}^{\infty} e(Ny^2)\,dy$. This is a (conditionally) convergent integral, regardless of whether we view it as $\lim_{Y\to\infty}\int_{-Y}^{Y} e(Ny^2)\,dy$ or in the wider sense as $\lim_{Y,Z\to\infty}\int_{-Y}^{Z} e(Ny^2)\,dy$. [Proof: If $0<Y<Y'$ then (substituting $y = t^{\frac{1}{2}}$)

$$\int_Y^{Y'} e(Ny^2)\,dy = \frac{1}{2}\int_{Y^2}^{Y'^2} t^{-\frac{1}{2}}e(Nt)\,dt = \Big[t^{-\frac{1}{2}}\frac{e(Nt)}{2\pi iN}\Big]_{Y^2}^{Y'^2} + \frac{1}{4\pi iN}\int_{Y^2}^{Y'^2} t^{-\frac{3}{2}}e(Nt)\,dt$$

and thus

$$\Big|\int_Y^{Y'} e(Ny^2)\,dy\Big| \le \frac{1}{2\pi N}\big(Y'^{-1} + Y^{-1}\big) + \frac{1}{4\pi N}\int_{Y^2}^{Y'^2} t^{-\frac{3}{2}}\,dt \le O(Y^{-1}) \qquad \text{as } Y\to\infty.$$

This proves that $\int_0^{\infty} e(Ny^2)\,dy$ is convergent. Since the integrand is even it also follows that $\int_{-\infty}^{0} e(Ny^2)\,dy$ is convergent.] Using this convergence of $\int_{-\infty}^{\infty} e(Ny^2)\,dy$ in the wider sense it follows from (229) that

$$S = N(1+i^{-N})\int_{-\infty}^{\infty} e(Ny^2)\,dy = N^{\frac{1}{2}}(1+i^{-N})\int_{-\infty}^{\infty} e(z^2)\,dz, \tag{230}$$

where we substituted $y = N^{-\frac{1}{2}}z$. Now the value of $\int_{-\infty}^{\infty} e(z^2)\,dz$ can be computed for example by taking $N=1$ in the above formula; we then have $S = \sum_{n=0}^{0} e(n^2/1) = 1$; hence we conclude $\int_{-\infty}^{\infty} e(z^2)\,dz = (1+i^{-1})^{-1}$. Hence

$$S = \frac{1+i^{-N}}{1+i^{-1}}N^{\frac{1}{2}},$$

and this agrees with the formula (225). □

5.4. **Finite sum formulas.** Before stating the main result, let us note that the computation of $L(1, \left(\frac{d}{\cdot}\right))$ can always be reduced to the case where d is a fundamental discriminant:

Lemma 5.16. *Let d be an integer $\equiv 0$ or $\equiv 1 \pmod 4$ which is not a square. Then there is a fundamental discriminant d_1 such that $d = d_1\ell^2$ for some $\ell \in \mathbb{Z}^+$, and*

$$L(1, \left(\tfrac{d}{\cdot}\right)) = L(1, \left(\tfrac{d_1}{\cdot}\right)) \prod_{p \mid d} \left(1 - \left(\tfrac{d_1}{p}\right) p^{-1}\right)$$

Proof. Let the prime factorization of d be $d = \varepsilon 2^\alpha p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ (where $\varepsilon = \pm 1$, $p_1, \dots, p_r$ are distinct odd primes, $\alpha = 0$ or $\alpha \geq 2$, and all $\alpha_j \geq 1$). Set

$$d_1 = \varepsilon \left\{ \begin{array}{ll} 1 & \text{if } 2 \mid \alpha \text{ and } d/2^\alpha \equiv 1 \pmod 4 \\ 4 & \text{if } 2 \mid \alpha \text{ and } d/2^\alpha \equiv 3 \pmod 4 \\ 8 & \text{if } 2 \nmid \alpha \end{array} \right\} \prod_{\substack{1 \leq j \leq r \\ 2 \nmid \alpha_j}} p_j.$$

Then d_1 is a fundamental discriminant (this is seen since $\varepsilon \prod_{\substack{1 \leq j \leq r \\ 2 \nmid \alpha_j}} p_j \equiv d/2^\alpha \pmod 4$), and there is a unique positive integer ℓ such that $d = d_1\ell^2$.

Now $\left(\frac{d}{\cdot}\right)$ *is the Dirichlet character modulo $|d|$ which is induced by the (primitive) Dirichlet character* $\left(\frac{d_1}{\cdot}\right) \in X_{|d_1|}$, in the sense defined just above Lemma 4.22. [Proof: We have to prove that for every $n \in \mathbb{Z}^+$ with $(n, |d|) = 1$ we have $\left(\frac{d}{n}\right) = \left(\frac{d_1}{n}\right)$. If n is odd then both the sides are Jacobi symbols, and we have $\left(\frac{d}{n}\right) = \left(\frac{d_1\ell^2}{n}\right) = \left(\frac{d_1}{n}\right)\left(\frac{\ell}{n}\right)^2 = \left(\frac{d_1}{n}\right)$ by (164). If n is even then we must have $d \equiv 1 \pmod 4$ because of $(n, |d|) = 1$; hence also $d_1 \equiv 1 \pmod 4$ and ℓ odd, and now $\left(\frac{d}{n}\right) = \left(\frac{n}{|d|}\right) = \left(\frac{n}{|d_1|\ell^2}\right) = \left(\frac{n}{|d_1|}\right) = \left(\frac{d_1}{n}\right)$ by Definition 4.10 and Definition 4.9.]

Now the stated relation between $L(1, \left(\frac{d}{\cdot}\right))$ and $L(1, \left(\frac{d_1}{\cdot}\right))$ follows from Lemma 4.22 (and analytic continuation from $\{\sigma > 1\}$ to the point $s = 1$). □

Theorem 5.17. *Let d be a fundamental discriminant. If $d < 0$ then*

$$L(1, \left(\tfrac{d}{\cdot}\right)) = -\frac{\pi}{|d|^{\frac{3}{2}}} \sum_{k=1}^{|d|} k \left(\frac{d}{k}\right). \tag{231}$$

If $d > 0$ then

$$L(1, \left(\tfrac{d}{\cdot}\right)) = -\frac{1}{d^{\frac{1}{2}}} \sum_{k=1}^{d-1} \left(\frac{d}{k}\right) \log \sin \frac{k\pi}{d}. \tag{232}$$

Remark 5.2. Hence, combining Theorem 5.17 and Theorem 5.4, we see that if $d < 0$ then

$$h(d) = -\frac{w}{2|d|} \sum_{k=1}^{|d|} k \left(\frac{d}{k}\right), \tag{233}$$

and if $d > 0$ then

$$h(d) = -\frac{1}{\log \varepsilon_d} \sum_{k=1}^{d-1} \left(\frac{d}{k}\right) \log \sin \frac{k\pi}{d}. \tag{234}$$

Proof of Theorem 5.17. We continue to use the short-hand notation "$\sqrt{d}$" for $\begin{cases} \sqrt{d} & \text{if } d > 0 \\ i\sqrt{|d|} & \text{if } d < 0. \end{cases}$

Theorem 5.13 says that $\left(\frac{d}{n}\right) = \frac{1}{\sqrt{d}} \sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) e\Big(\frac{nk}{|d|}\Big)$ for each $n \in \mathbb{Z}^+$. Hence

$$L(1, \left(\tfrac{d}{\cdot}\right)) = \sum_{n=1}^{\infty} \left(\frac{d}{n}\right) n^{-1} = \frac{1}{\sqrt{d}} \sum_{n=1}^{\infty} \Big(\sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) e\Big(\frac{nk}{|d|}\Big)\Big) n^{-1}.$$

Here we wish to change order of summation; we have to be slightly careful in doing this since we know that the outer sum is only conditionally convergent. Using the definition of an infinite sum we have

$$L(1, \left(\tfrac{d}{\cdot}\right)) = \frac{1}{\sqrt{d}} \lim_{N\to\infty} \sum_{n=1}^{N} \sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) e\Big(\frac{nk}{|d|}\Big) n^{-1} = \frac{1}{\sqrt{d}} \lim_{N\to\infty} \sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) \sum_{n=1}^{N} n^{-1} e\Big(\frac{nk}{|d|}\Big), \tag{235}$$

and here since $\sum_{k=1}^{|d|-1}$ is a finite sum we may change order between "$\lim_{N\to\infty}$" and "$\sum_{k=1}^{|d|-1}$" if we can only prove that each limit

$$\lim_{N\to\infty} \sum_{n=1}^{N} n^{-1} e\Big(\frac{nk}{|d|}\Big) = \sum_{n=1}^{\infty} n^{-1} e\Big(\frac{nk}{|d|}\Big)$$

exists. In fact, the following more general statement is true: For any $z \in \mathbb{C}$ with $|z| \le 1$ and $z \neq 1$ we have

$$\sum_{n=1}^{\infty} n^{-1} z^n = -\log(1-z) \qquad \text{(principal value of the logarithm)}, \tag{236}$$

and in particular the series in the left hand side converges. When $|z| < 1$ this is of course the well-known Taylor series for $-\log(1-z)$. The fact that the formula also holds when $|z| = 1$, $z \neq 1$ is crucial to us, and this is more difficult to prove; we leave it as an exercise; see Problem 5.4.

Let us compute the real and imaginary part of $-\log(1-z)$, if $z = e^{i\theta}$ with $0 < \theta < 2\pi$. First of all we have

$$|1-z| = \sqrt{(1-\cos\theta)^2 + \sin^2\theta} = \sqrt{2(1-\cos\theta)} = \sqrt{4\sin^2(\theta/2)} = 2\sin(\theta/2),$$

where in the last step we used the fact that $\sin(\theta/2) > 0$ for $0 < \theta < 2\pi$. Next note that $\mathrm{Re}(1-z) > 0$ and hence for the principal value of the logarithm we have $\arg(1-z) \in (-\frac{\pi}{2}, \frac{\pi}{2})$,

and thus

$$\arg(1-z) = \arcsin\Big(\frac{\operatorname{Im}(1-z)}{|1-z|}\Big) = \arcsin\Big(-\frac{\sin\theta}{2\sin(\theta/2)}\Big)$$
$$= \arcsin\big(-\cos(\theta/2)\big) = \arcsin\big(\sin\big(-\frac{\pi-\theta}{2}\big)\big) = \frac{\theta-\pi}{2},$$

since $\frac{\theta-\pi}{2} \in (-\frac{\pi}{2}, \frac{\pi}{2})$. (Alternatively the two formulas $|1-z| = 2\sin(\theta/2)$ and $\arg(1-z) = \frac{\theta-\pi}{2}$ can be deduced from a picture using a bit of Euclidean geometry.) Hence we obtain, for $z = e^{i\theta}$:

$$\sum_{n=1}^{\infty} n^{-1} z^n = -\log(1-z) = -\log|1-z| - i\arg(1-z) = -\log\big(2\sin\tfrac{\theta}{2}\big) - i\,\frac{\theta-\pi}{2}.$$

Using the above with $z = e(k/|d|)$ (thus $\theta = 2\pi k/|d|$) we see that the change of order between "$\lim_{N\to\infty}$" and "$\sum_{k=1}^{|d|-1}$" in (235) is justified, and we obtain:

$$L(1, \big(\tfrac{d}{\cdot}\big)) = -\frac{1}{\sqrt{d}} \sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) \Big[\log\big(2\sin\frac{\pi k}{|d|}\big) + i\big(\frac{\pi k}{|d|} - \frac{\pi}{2}\big)\Big].$$

Now if $d > 0$ then $\sqrt{d}$ is real, and since $L(1, \big(\frac{d}{\cdot}\big))$ also is real we must have

$$L(1, \big(\tfrac{d}{\cdot}\big)) = -\frac{1}{d^{\frac{1}{2}}} \sum_{k=1}^{d-1} \left(\frac{d}{k}\right) \log\big(2\sin\frac{k\pi}{d}\big) = -\frac{1}{d^{\frac{1}{2}}} \sum_{k=1}^{d-1} \left(\frac{d}{k}\right) \log\sin\frac{k\pi}{d},$$

where in the last equality we used $\sum_{k=1}^{d-1} \big(\frac{d}{k}\big) = 0$ (cf. Lemma 3.13 and Lemma 4.36). On the other hand if $d < 0$ then $\sqrt{d} = i\sqrt{|d|}$, and since $L(1, \big(\frac{d}{\cdot}\big))$ is real we must have

$$L(1, \big(\tfrac{d}{\cdot}\big)) = -\frac{1}{|d|^{\frac{1}{2}}} \sum_{k=1}^{|d|-1} \left(\frac{d}{k}\right) \big(\frac{\pi k}{|d|} - \frac{\pi}{2}\big) = -\frac{\pi}{|d|^{\frac{3}{2}}} \sum_{k=1}^{|d|-1} k \left(\frac{d}{k}\right),$$

where we again used $\sum_{k=1}^{d-1} \big(\frac{d}{k}\big) = 0$. This completes the proof of Theorem 5.17. □

5.5. *** Ideal classes in number fields, and the Dedekind Zeta Function.** This section is external reading. I will here give a very brief collection of definitions and facts about the so called *Dedekind Zeta Function* (for a general number field), and about the relation between ideal classes in quadratic fields and equivalence classes of quadratic forms. The purpose is merely to give a first glimpse of this very rich and beautiful area!

There exist a large number of textbooks describing various parts of this material; for example we mention Cohn [9], Lang [40] and Neukirch [52], and Ireland and Rosen [35, Ch. 12].

Let K be a number field of degree n, i.e. a field which is a finite extension of $\mathbb{Q}$, of degree n. An element $x \in K$ is called an *algebraic integer* if it is the zero of some monic polynomial with integer coefficients. Let $\mathcal{O} \subset K$ be the set of all algebraic integers; $\mathcal{O}$ is

in fact a *ring*, called "the ring of integers in K" for short. It turns out that in general $\mathcal{O}$ is not a unique factorization domain. However $\mathcal{O}$ has a property which is almost as good. Namely, every nonzero ideal of $\mathcal{O}$ can be written uniquely as a product of *prime* ideals.

The *norm* $N(A)$ of any (nonzero) ideal $A \subset \mathcal{O}$ is defined as the number of elements in the quotient ring $\mathcal{O}/A$. This is always a finite number, and $N(AB) = N(A)N(B)$ holds for any two nonzero ideals $A, B \subset \mathcal{O}$. Now the *Dedekind Zeta Function* $\zeta_K(s)$ is defined by

$$\zeta_K(s) = \sum_{A \subset \mathcal{O}} \frac{1}{N(A)^s} = \prod_P \big(1 - N(P)^{-s}\big)^{-1} \qquad (\sigma > 1), \tag{237}$$

where the sum is taken over all nonzero ideals $A \subset \mathcal{O}$, and the product (the "Euler product" for $\zeta_K(s)$) is taken over all prime ideals $P \subset \mathcal{O}$. (The Riemann Zeta function is obtained as the special case $\zeta(s) = \zeta_{\mathbb{Q}}(s)$.) The sum is in fact a Dirichlet series with abscissa of convergence $\sigma_c = \sigma_a = 1$, and the product is absolutely convergent when $\sigma > 1$.

It turns out that $\zeta_K(s)$ *has a meromorphic continuation to all of* $\mathbb{C}$*, with the only pole being a simple pole at* $s = 1$. There is a very important formula connecting the residue at $s = 1$ with various other invariants of the field K:

$$\mathrm{Res}_{s=1}\, \zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2} R_K}{\#W \cdot |d_{K/\mathbb{Q}}|^{\frac{1}{2}}} h_K. \tag{238}$$

We now describe the various numbers in the right hand side. The *deepest* invariant (in some sense) is the *class number*, h_K. To define it, two ideals $A, B \subset \mathcal{O}$ are said to be equivalent, $A \sim B$, if there exist nonzero $\alpha, \beta \in \mathcal{O}$ such that $(\alpha)A = (\beta)B$ (where (α), (β) are the principal ideals generated by α, β). This is an equivalence relation on the set of ideals of $\mathcal{O}$, and h_K is *the number of equivalence classes*. It turns out that h_K is always *finite*, and $h_K = 1$ if and only if $\mathcal{O}$ is a unique factorization domain. Thus in a sense h_K can be viewed as a measure on how far $\mathcal{O}$ is from being a unique factorization domain!

Furthermore, $d_{K/\mathbb{Q}} \in \mathbb{Z}\backslash\{0\}$ is the *discriminant* of K, defined as $d_{K/\mathbb{Q}} = \det\big(\mathrm{Tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j)\big)$ where $\alpha_1, \ldots, \alpha_n \in \mathcal{O}$ is any integral basis for $\mathcal{O}$, i.e. any basis for K over $\mathbb{Q}$ such that $\mathcal{O} = \mathbb{Z}\alpha_1 + \ldots + \mathbb{Z}\alpha_n$. (An integral basis for $\mathcal{O}$ always exists, and the determinant $\det\big(\mathrm{Tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j)\big)$ is independent of the choice of integral basis.)

Also in (238), r_1 and r_2 are the numbers of *real* and *complex embeddings of* K, respectively. That is, r_1 is the number of distinct field embeddings $K \to \mathbb{R}$ and r_2 is the number of distinct pairs of conjugate complex field imbeddings $K \to \mathbb{C}$. (We always have $r_1 + 2r_2 = n$.) We let $\rho_1, \ldots, \rho_{r_1+r_2}$ be a system of representatives of all the embeddings of K into $\mathbb{C}$ up to complex conjugation. For $x \in K$ and $j \in \{1, \ldots, r_1 + r_2\}$ we define $\|x\|_j$ as $|\rho_j(x)|$ if ρ_j is real, but $\|x\|_j := |\rho_j(x)|^2$ if ρ_j is complex.

Furthermore, W denotes the group of roots of unity in K. This is in fact a finite cyclic group, and a subgroup of $\mathcal{O}_K^\times$ (the group of invertible elements in $\mathcal{O}_K$, cf. (123)). *Dirichlet's Unit Theorem* states that $\mathcal{O}_K^\times/W$ is a free abelian group on $r = r_1 + r_2 - 1$ generators. Let

us pick some elements $u_1, \dots, u_r \in \mathcal{O}_K^\times$ which map to a set of generators for $\mathcal{O}_K^\times/W$. Then R_K, the *regulator of* K, is defined as

$$R_K := \Big|\det\big(\log \|u_k\|_j\big)_{j,k=1,\dots,r}\Big|.$$

This number is independent of the choice of $u_1, \dots, u_r$, and also independent of our choice and ordering of $\rho_1, \dots, \rho_{r+1}$, as follows from the fact that $\prod_{j=1}^{r+1} \|u\|_j = 1$ for all $u \in \mathcal{O}_K^\times$. This completes the description of the right hand side in (238).

We now turn to the special case of *quadratic number fields*, i.e. number fields K of degree $n = 2$. It turns out that each quadratic field K equals $\mathbb{Q}(\sqrt{d})$, where $d = d_{K/\mathbb{Q}}$, and the integers which occur as discriminants are exactly the *fundamental discriminants* (cf. Definition 4.12). In other words: There is a bijective correspondence between set of fundamental discriminants and the family of quadratic fields, given by $d \mapsto \mathbb{Q}(d)$, with inverse $K \mapsto d_{K/\mathbb{Q}}$.

For a quadratic number field K with $d = d_{K/\mathbb{Q}} < 0$, the other invariants in the right hand side of (238) are as follows. If $d < 0$ then $r_1 = 0$ and $r_2 = 1$, and

$$W = \begin{cases} \{e(n/6) \,:\, n \in \mathbb{Z}\} & \text{if } d = -3, \\ \{1, i, -1, -i\} & \text{if } d = -4, \\ \{1, -1\} & \text{otherwise.} \end{cases} \tag{239}$$

Thus $\#W = w$, the number in (201). Furthermore in this case $\mathcal{O}_K^\times = W$, $r = 0$, and we declare $R_K = 1$.

On the other hand if $d > 0$ then $r_1 = 2$ and $r_2 = 1$; $W = \{1, -1\}$ always, and Dirichlet's Unit Theorem says that $\mathcal{O}_K^\times = \{\pm\varepsilon^n \,:\, n \in \mathbb{Z}\}$ for some $\varepsilon \in \mathcal{O}_K^\times$. In fact we can take $\varepsilon = \varepsilon_d' := \frac{1}{2}(x + y\sqrt{d})$ where $\langle x, y\rangle \in \mathbb{Z}^+ \times \mathbb{Z}^+$ is the solution to $x^2 - dy^2 = \pm 4$ for which y is minimal. (Then $N_{K/\mathbb{Q}}(\varepsilon_d') = \frac{1}{4}(x^2 - dy^2) = \pm 1$.) In this situation it turns out that the number ε_d defined in Theorem 200 is $\varepsilon_d = \varepsilon_d'$ if $N_{K/\mathbb{Q}}(\varepsilon_d') = 1$, and $\varepsilon_d = (\varepsilon_d')^2$ if $N_{K/\mathbb{Q}}(\varepsilon_d') = -1$. The number ε_d' is called the *fundamental unit of* $\mathcal{O}_K$, and ε_d is called the *proper fundamental unit of* $\mathcal{O}_K$. We now see that the regulator of K is $R_K = \log \varepsilon_d'$, where we view $\varepsilon_d' := \frac{1}{2}(x + y\sqrt{d}) \in K$ as embedded in $\mathbb{R}$ by taking $\sqrt{d} \in \mathbb{R}^+$ to be the positive square root of d.

Hence the formula (238) says in the case of a quadratic number field K:

$$\operatorname{Res}_{s=1} \zeta_K(s) = \begin{cases} \frac{2\pi}{w\sqrt{|d|}} h_K & \text{if } d < 0, \\ \frac{2\log \varepsilon_d'}{\sqrt{d}} h_K & \text{if } d > 0. \end{cases} \tag{240}$$

The close similarity between (240) and Dirichlet's class number formula, Theorem 5.4, has the following explanation.

On the one hand side, for any quadratic number field K with discriminant d, there exists a beautiful *correspondence between the equivalence classes of ideals of K, and the equivalence classes of quadratic forms with discriminant d* (these are necessarily primitive, since d is a fundamental discriminant). More precisely, if $d<0$ or if $d>0$ and $N_{K/\mathbb{Q}}(\varepsilon_d')=-1$ then there is a canonical *bijective* correspondence between these two families. In the remaining case when $d>0$ and $N_{K/\mathbb{Q}}(\varepsilon_d')=1$ (thus $\varepsilon_d'=\varepsilon_d$), we instead get a *2-to-1 map from the set of ideal classes of K onto the set of equivalence classes of quadratic forms with discriminant d*. Hence:

$$h(d)=\left.\begin{cases}2 & \text{if } d>0 \text{ and } N_{K/\mathbb{Q}}(\varepsilon_d')=1\\ 1 & \text{otherwise}\end{cases}\right\} h_K. \tag{241}$$

On the other hand, we have a *factorization of the Dedekind Zeta function $\zeta_K(s)$:*

$$\zeta_K(s)=\zeta(s)L(s,\left(\tfrac{d}{\cdot}\right)). \tag{242}$$

The reason for this comes from a precise description of how the prime numbers $p\in\mathbb{Z}$ factor into prime ideals of $\mathcal{O}_K$: If p is a prime number with $\left(\frac{d}{p}\right)=1$ then the principal ideal $(p)\subset\mathcal{O}_K$ factors as a product of two distinct prime ideals, $(p)=P_1P_2$, and $N(P_1)=N(P_2)=p$. If $\left(\frac{d}{p}\right)=-1$ then (p) is itself a prime ideal in $\mathcal{O}_K$, with $N((p))=p^2$. Finally if $p\mid d$, i.e. $\left(\frac{d}{p}\right)=0$, then (p) factors as a square of a prime ideal; $(p)=P^2$, and $N(P)=p$. The prime ideals of $\mathcal{O}_K$ which occur in this way are all distinct, and there are no other prime ideals of $\mathcal{O}_K$. Hence for $\sigma>1$ we have

$$\begin{aligned}\zeta_K(s)=\prod_P\big(1-N(P)^{-s}\big)^{-1}&=\prod_p\left.\begin{cases}(1-p^{-s})^{-2} & \text{if } (d/p)=1\\ (1-p^{-s})^{-1} & \text{if } (d/p)=0\\ (1-p^{-2s})^{-1} & \text{if } (d/p)=-1\end{cases}\right\}\\ &=\prod_p(1-p^{-s})^{-1}\prod_p\left.\begin{cases}(1-p^{-s})^{-1} & \text{if } (d/p)=1\\ 1 & \text{if } (d/p)=0\\ (1+p^{-s})^{-1} & \text{if } (d/p)=-1\end{cases}\right\}\\ &=\prod_p(1-p^{-s})^{-1}\prod_p(1-\left(\tfrac{d}{p}\right)p^{-s})^{-1}=\zeta(s)L(s,\left(\tfrac{d}{\cdot}\right)),\end{aligned}$$

as claimed.

Since $\zeta(s)$ has a simple pole at $s=1$ with residue 1, it follows from (242) that

$$\operatorname{Res}_{s=1}\zeta_K(s)=L(1,\left(\tfrac{d}{\cdot}\right)). \tag{243}$$

Combining the facts of the last few paragraphs, we see that *for the case of a quadratic number field K, the formula* (240) *says exactly the same thing as Dirichlet's class number formula, Theorem 5.4 with $d=d_{K/\mathbb{Q}}$.*

5.6. **Problems.**

* *Problem* 5.1. Let d be a non-zero integer which is $\equiv 0$ or $1 \pmod 4$. Prove that d is a fundamental discriminant if and only if every quadratic form with discriminant d is primitive.

* *Problem* 5.2. Prove that if $d < 0$ is a fundamental discriminant, then

$$h(d) = \big\{\langle a, b, c\rangle \subset \mathbb{Z}^3 \,:\, b^2 - 4ac = d,\ [-a < b \le a < c \text{ or } 0 \le b \le a = c]\big\}.$$

* *Problem* 5.3. Give an alternative proof of Lemma 5.5 along the lines in Davenport's book, p. 45. (Note that Davenport proves one direction; that every element in the set in (202) is indeed an automorph of Q. Check the details of this argument, and then try to also prove the other direction of Lemma 5.5 working along similar lines.)

Problem 5.4. Prove that (236) holds for all $z \in \mathbb{C}$ with $|z| \le 1$ and $z \ne 1$. [Hint. Using partial summation one can prove that the series $\sum_{n=1}^{\infty} n^{-1} z^n$ converges for all z with $|z| \le 1$, $z \ne 1$, *and is continuous* for these z.]

6. The distribution of the primes

(Davenport chapter 7-8.)

6.1. **The logarithmic integral and the prime number theorem.**

Definition 6.1. We write $\pi(x)$ for the number of prime numbers not exceeding x, viz.

$$\pi(x) = \#\{p \,:\, p \text{ is a prime number} \le x\}.$$

Definition 6.2. The *logarithmic integral* is the function defined by

$$\operatorname{Li} x = \int_2^x \frac{dt}{\log t}. \tag{244}$$

In the next lecture we will prove the prime number theorem, which states that

$$\pi(x) \sim \operatorname{Li} x \qquad \text{as } x \to \infty. \tag{245}$$

Later in this course we will prove the prime number theorem with a precise error term: There is some constant $a > 0$ such that

$$\pi(x) = \operatorname{Li} x + O\Big(xe^{-a\sqrt{\log x}}\Big) \qquad \text{as } x \to \infty. \tag{246}$$

Let us also point out that if the famous *Riemann Hypothesis* about the zeros of $\zeta(s)$ is true, then the following much better estimate holds:

$$\pi(x) = \operatorname{Li} x + O\big(\sqrt{x}\log x\big) \qquad \text{as } x \to \infty. \tag{247}$$

The logarithmic integral satisfies the asymptotic relation

$$\operatorname{Li} x \sim \frac{x}{\log x} \qquad \text{as } x \to \infty$$

(this follows from taking $q = 0$ in Lemma 6.1 below), and hence the prime number theorem without error term, (245), is equivalent with the statement that

$$\pi(x) \sim \frac{x}{\log x} \qquad \text{as } x \to \infty. \tag{248}$$

However the more precise result (246) (as well as the conditional result (247)) shows that it is really the function $\operatorname{Li} x$ that is the more "correct" approximation to $\pi(x)$.

Lemma 6.1. *The logarithmic integral has the following asymptotic expansion: For any fixed integer $q \ge 0$ we have*

$$\operatorname{Li} x = \frac{x}{\log x} + \frac{1!x}{(\log x)^2} + \ldots + \frac{q!x}{(\log x)^{q+1}} + O\Big(\frac{x}{(\log x)^{q+2}}\Big) \qquad \textit{as } x \to \infty. \tag{249}$$

(The implied constant depends on q but not on x.)

Proof. Integrating by parts repeatedly we have

$$\begin{aligned}\operatorname{Li} x = \int_2^x \frac{dt}{\log t} &= \Big[\frac{t}{\log t}\Big]_{t=2}^{t=x} + \int_2^x \frac{dt}{(\log t)^2} \\ &= \Big[\frac{t}{\log t}\Big]_{t=2}^{t=x} + \Big[\frac{t}{(\log t)^2}\Big]_{t=2}^{t=x} + \int_2^x \frac{2\,dt}{(\log t)^3} \\ &= \Big[\frac{t}{\log t}\Big]_{t=2}^{t=x} + \Big[\frac{t}{(\log t)^2}\Big]_{t=2}^{t=x} + \Big[\frac{2t}{(\log t)^3}\Big]_{t=2}^{t=x} + \int_2^x \frac{3!\,dt}{(\log t)^4}.\end{aligned}$$

Continuing in the same way we get, for any integer $q \geq 0$,

$$\begin{aligned}\operatorname{Li} x &= \Big[\frac{t}{\log t}\Big]_{t=2}^{t=x} + \Big[\frac{1!t}{(\log t)^2}\Big]_{t=2}^{t=x} + \Big[\frac{2!t}{(\log t)^3}\Big]_{t=2}^{t=x} + \ldots + \Big[\frac{q!t}{(\log t)^{q+1}}\Big]_{t=2}^{t=x} + \int_2^x \frac{(q+1)!\,dt}{(\log t)^{q+2}} \\ &= \frac{x}{\log x} + \frac{1!x}{(\log x)^2} + \frac{2!x}{(\log x)^3} + \ldots + \frac{q!x}{(\log x)^{q+1}} + O\big(1\big) + O\Big(\int_2^x \frac{dt}{(\log t)^{q+2}}\Big).\end{aligned}$$

The last term may be estimated as follows, for any $x \geq 4$:

$$\int_2^x \frac{dt}{(\log t)^{q+2}} \leq \int_2^{\sqrt{x}} \frac{dt}{(\log 2)^{q+2}} + \int_{\sqrt{x}}^x \frac{dt}{(\log\sqrt{x})^{q+2}} \leq O(\sqrt{x}) + O\Big(\frac{x}{(\log x)^{q+2}}\Big).$$

This gives the stated result. □

Note that the error term in (246) decays faster than the error term in (249), i.e. for any fixed q we have $xe^{-a\sqrt{\log x}} = o\big(\frac{x}{(\log x)^q}\big)$ as $x \to \infty$.[16] Thus we cannot replace $\operatorname{Li} x$ in (246) with a finite sum as in (249) (no matter how large q we choose) if we wish to keep the good quality error term $O\big(xe^{-a\sqrt{\log x}}\big)$.

6.2. **Tchebychev's auxiliary functions ϑ and ψ.** (In this section we borrow from Ingham [34, Ch. 1].) Tchebychev introduced (1851-2) the following two auxiliary functions.

Definition 6.3. Set

$$\psi(x) = \sum_{n \leq x} \Lambda(n); \qquad \vartheta(x) = \sum_{p \leq x} \log p \qquad (x > 0). \tag{250}$$

Here recall that $\Lambda(n) = \log p$ if $n = p^m$ (p a prime, $m \in \mathbb{Z}^+$), otherwise $\Lambda(n) = 0$.

Note that the functions $\Lambda(n)$ and $\psi(x)$ are the same as we introduced in Example 3.7 (cf. (112) and (117)).

Proposition 6.2. *We have*

$$\limsup_{x\to\infty} \frac{\pi(x)}{x/\log x} = \limsup_{x\to\infty} \frac{\vartheta(x)}{x} = \limsup_{x\to\infty} \frac{\psi(x)}{x}$$

[16] Proof: Dividing with x and then letting $y = \sqrt{\log x}$ we see that the statement is equivalent with $e^{-ay} = o\big(y^{-2q}\big)$ as $y \to \infty$, and this fact is well-known from basic analysis.

and

$$\liminf_{x\to\infty} \frac{\pi(x)}{x/\log x} = \liminf_{x\to\infty} \frac{\vartheta(x)}{x} = \liminf_{x\to\infty} \frac{\psi(x)}{x}.$$

In particular this proposition shows that the three relations (as $x \to \infty$)

$$\pi(x) \sim \frac{x}{\log x}, \qquad \vartheta(x) \sim x, \qquad \psi(x) \sim x,$$

the first of which is the prime number theorem, are equivalent. Of the three functions π, ϑ, ψ, the one which arises most naturally from the analytical point of view is ψ. For this reason it is usually most convenient to work in the first instance with ψ, and to use Proposition 6.2 (or similar relations with more precise error bounds) to deduce results about π.

Proof. Let us write

$$\Lambda_1 = \limsup_{x\to\infty} \frac{\pi(x)}{x/\log x}; \qquad \Lambda_2 = \limsup_{x\to\infty} \frac{\vartheta(x)}{x}; \qquad \Lambda_3 = \limsup_{x\to\infty} \frac{\psi(x)}{x}.$$

(These numbers may possibly be $+\infty$.) For any $x > 0$ we have

$$\vartheta(x) \le \psi(x) \le \sum_{p\le x} \sum_{\substack{m\ge 1\\(p^m\le x)}} \log p = \sum_{p\le x} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p \le \sum_{p\le x} \log x = \pi(x)\log x.$$

This implies that $\Lambda_2 \le \Lambda_3 \le \Lambda_1$. On the other hand, for any $0 < \alpha < 1$ and $x > 1$,

$$\vartheta(x) \ge \sum_{x^\alpha < p\le x} \log p \ge \big(\pi(x) - \pi(x^\alpha)\big)\log(x^\alpha),$$

and hence, since $\pi(x^\alpha) < x^\alpha$,

$$\frac{\vartheta(x)}{x} > \alpha\Big(\frac{\pi(x)\log x}{x} - \frac{\log x}{x^{1-\alpha}}\Big).$$

Keep α fixed and let $x \to \infty$; since $\frac{\log x}{x^{1-\alpha}} \to 0$ we deduce that $\Lambda_2 \ge \alpha\Lambda_1$, whence $\Lambda_2 \ge \Lambda_1$ since α may be taken as near as we please to 1. This combined with the previous inequalities gives $\Lambda_2 = \Lambda_3 = \Lambda_1$. Exactly the same argument works to prove the equality between the three lim inf's. □

Remark 6.1. In Problem 3.6 we proved that $\vartheta(x) \sim x$ as $x \to \infty$ implies $\pi(x) \sim \frac{x}{\log x}$ as $x \to \infty$, using partial integration. This implication is of course also a consequence of Proposition 6.2, and the above proof of Proposition 6.2 is quite a bit shorter than our solution to Problem 3.6. However, the technique of using integration by parts is useful when deriving more precise error bounds in the translation between the three functions π, ϑ, ψ.

Remark 6.2. We point out the following relation between ψ and ϑ:

$$\psi(x) = \sum_{p \le x} \sum_{\substack{m \ge 1 \\ (p^m \le x)}} \log p = \sum_{\substack{m \ge 1 \\ (2^m \le x)}} \sum_{p \le x^{1/m}} \log p = \sum_{1 \le m \le \log_2 x} \vartheta(x^{1/m}). \tag{251}$$

From this we see that ψ and ϑ are asymptotically quite close to each other (much more than is indicated by the statement of Proposition 6.2): Recall that, trivially, $\vartheta(x) \le \psi(x)$. In the other direction we have, using the above relation and the trivial bound $\vartheta(y) \le y \log y$ for $y = x^{1/m}$, $m \ge 2$:

$$\begin{aligned} \psi(x) &\le \vartheta(x) + \sum_{2 \le m \le \log_2 x} \frac{1}{m} x^{1/m} \log x \le \vartheta(x) + \frac{1}{2}\sqrt{x} \log x + \frac{1}{3} x^{1/3} (\log_2 x) \log x \\ &\le \vartheta(x) + O\big(\sqrt{x} \log x\big) \qquad \text{as } \ x \to \infty. \end{aligned} \tag{252}$$

(In view of Theorem 6.3 below combined with Proposition 6.2 we have $\vartheta(y) = O(y)$ as $y \to \infty$, and using this above for $m = 2$ we can strengthen (252) to $\psi(x) \le \vartheta(x) + O(\sqrt{x})$ as $x \to \infty$.)

Tchebychev proved the following bound in 1852, for the first time getting in the "vicinity" of the prime number theorem:

$$(0.92...)\frac{x}{\log x} < \pi(x) < (1.105...)\frac{x}{\log x} \tag{253}$$

for all sufficiently large x. We will prove a weaker bound of the same nature, to illustrate Tchebychev's method. (See Problem 6.3 for a proof of (253).)

Theorem 6.3. *For any $\varepsilon > 0$, we have for all sufficiently large x:*

$$(\log 2 - \varepsilon)\frac{x}{\log x} < \pi(x) < (2 \log 2 + \varepsilon)\frac{x}{\log x}. \tag{254}$$

Proof. The basic idea of the proof is to consider the factorial $n!$ for large integers n: On the one hand side we can count in a precise way the prime factors appearing in $n!$:

$$\operatorname{ord}_p(n!) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \dots \qquad (\forall n \in \mathbb{Z}^+,\ p \text{ prime}); \tag{255}$$

on the other hand we have a precise knowledge of the size of $n!$, namely from Stirling's formula:

$$\log(n!) = (n + \tfrac{1}{2}) \log n - n + \log \sqrt{2\pi} + O(n^{-1}) \qquad \forall n \ge 1. \tag{256}$$

To prove (255), note that the rth term $n_r = \lfloor n/p^r \rfloor$ equals the number of factors in the product $1 \cdot 2 \cdot \ldots \cdot n$ which are divisible by p^r, and a factor which contains p exactly r times is counted exactly r times in the sum $n_1 + n_2 + \ldots$, namely once for each of the terms $n_1, n_2, \ldots, n_r$. The asymptotic formula (256) you may have seen in previous courses; anyway we will prove it in Section 8.2; and in fact for the following argument we will only need the much less precise version $\log(n!) = n \log n - n + O(\log n)$.

It is easiest to use (255) to give information about $\psi(x)$ first, rather than $\pi(x)$ directly. Indeed, note that (255) means that $n!$ has the prime factorization $n! = \prod_{p\le n} p^{\lfloor n/p\rfloor + \lfloor n/p^2\rfloor + \cdots}$, and taking the logarithm this gives

$$\log(n!) = \sum_{p\le n}\Big(\Big\lfloor\frac{n}{p}\Big\rfloor\log p + \Big\lfloor\frac{n}{p^2}\Big\rfloor\log p + \Big\lfloor\frac{n}{p^3}\Big\rfloor\log p + \dots\Big) = \sum_{p\le n}\sum_{r=0}^{\infty}\Big\lfloor\frac{n}{p^r}\Big\rfloor\log p$$
$$= \sum_{1\le m\le n}\Big\lfloor\frac{n}{m}\Big\rfloor\Lambda(m).$$

It is convenient to extend our notation to arbitrary real numbers; thus let us note that the above formula implies

$$T(x) := \log\big(\lfloor x\rfloor!\big) = \sum_{1\le m\le x}\Big\lfloor\frac{x}{m}\Big\rfloor\Lambda(m), \qquad \forall x > 0. \tag{257}$$

We remark that this formula may alternatively be proved by identifying coefficients in the Dirichlet series identity $\frac{\zeta'(s)}{\zeta(s)}\zeta(s) = \zeta'(s)$; cf. Davenport pp. 55-56 and Problem 6.2 below. We also have, from (256):

$$T(x) = x\log x - x + O\big(\log x\big) \qquad \forall x \ge 2. \tag{258}$$

(Note that the error term here is the best possible as $x\to\infty$.)

Now let us study the difference $T(x) - 2T(\frac{1}{2}x)$. Note that the function $a\mapsto \lfloor 2a\rfloor - 2\lfloor a\rfloor$ is periodic with period 1, and for $0\le a < \frac{1}{2}$ it equals 0, while for $\frac{1}{2}\le a < 1$ it equals 1. In particular we have $0\le \lfloor 2a\rfloor - 2\lfloor a\rfloor \le 1$ for all $a\in\mathbb{R}$. Hence, using (257),

$$\psi(x) - \psi(\tfrac{1}{2}x) = \sum_{\frac{1}{2}x<m\le x}\Lambda(m) \le T(x) - 2T(\tfrac{1}{2}x) \le \sum_{1\le m\le x}\Lambda(m) = \psi(x). \tag{259}$$

Also, using (258), we have

$$T(x) - 2T(\tfrac{1}{2}x) = (\log 2)x + O\big(\log x\big), \qquad \forall x\ge 4. \tag{260}$$

Hence from the second inequality in (259) we get $\psi(x)\ge (\log 2)x + O(\log x)$ as $x\to\infty$, and thus

$$\liminf_{x\to\infty}\frac{\psi(x)}{x} \ge \log 2. \tag{261}$$

On the other hand the first inequality in (259) gives

$$\psi(x) - \psi(\tfrac{1}{2}x) \le (\log 2)x + O(\log x), \qquad \forall x\ge 4,$$

and adding this applied to $x, 2^{-1}x, \dots, 2^{-(k-1)}x$ where k is determined so that $2 \le 2^{-k}x < 4$, we get

$$\begin{aligned}\psi(x) &= \sum_{j=0}^{k-1}\big(\psi(2^{-j}x) - \psi(2^{-j-1}x)\big) + \psi(2^{-k}x)\\ &\le (\log 2)\sum_{j=0}^{k-1} 2^{-j}x + O\Big(k\log x + 1\Big) \le 2(\log 2)x + O\big((\log x)^2\big), \qquad \forall x \ge 4.\end{aligned}$$

This implies

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \le 2\log 2. \tag{262}$$

By Proposition 6.2, (261) implies $\liminf_{x\to\infty} \frac{\pi(x)}{x/\log x} \ge \log 2$ and (262) implies $\limsup_{x\to\infty} \frac{\pi(x)}{x/\log x} \le 2\log 2$. These inequalities are equivalent with the statement of the theorem. □

Remark 6.3. Note that the above proof is slightly simpler than the one sketched in Davenport's book p. 56, since we make use of Proposition 6.2.

Theorem 6.4. *We have*

$$\liminf_{x\to\infty} \frac{\pi(x)}{x/\log x} \le 1 \le \limsup_{x\to\infty} \frac{\pi(x)}{x/\log x}.$$

Proof. Recall that in Example 3.7 (see formula (116)) we proved that

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^\infty \psi(x)x^{-s-1}\,dx \qquad (\sigma > 1).$$

The proof of Theorem 6.4 is based on taking $s \to 1^+$ in this formula. Let us write

$$\lambda = \liminf_{x\to\infty} \frac{\psi(x)}{x}; \qquad \Lambda = \limsup_{x\to\infty} \frac{\psi(x)}{x}.$$

Then if B is any number $> \Lambda$, there exists some $x_0 > 1$ such that $\frac{\psi(x)}{x} < B$ for all $x \ge x_0$, and we deduce that, for $s > 1$:

$$\begin{aligned}-\frac{\zeta'(s)}{\zeta(s)} &= s\int_1^\infty \frac{\psi(x)}{x^{s+1}}\,dx < s\int_1^{x_0} \frac{\psi(x)}{x^{s+1}}\,dx + s\int_{x_0}^\infty \frac{B}{x^s}\,dx\\ &= s\int_1^{x_0} \frac{\psi(x)-Bx}{x^{s+1}}\,dx + s\int_1^\infty \frac{B}{x^s}\,dx = s\int_1^{x_0} \frac{\psi(x)-Bx}{x^{s+1}}\,dx + \frac{sB}{s-1}\\ &\le s\int_1^{x_0} \frac{\psi(x)-Bx}{x^2}\,dx + \frac{sB}{s-1}.\end{aligned}$$

Thus multiplying with $(s-1)$ and letting $s \to 1^+$ we get $\limsup_{s\to 1^+}(s-1)\frac{\zeta'(s)}{\zeta(s)} \le B$. Since this is true for every $B > \Lambda$ we conclude:

$$\limsup_{s\to 1^+} \ -(s-1)\frac{\zeta'(s)}{\zeta(s)} \le \Lambda.$$

(Note that this is also true, trivially, if $\Lambda = \infty$.) Similarly one proves

$$\liminf_{s\to 1^+} \ -(s-1)\frac{\zeta'(s)}{\zeta(s)} \ge \lambda.$$

On the other hand, in Example 3.6 we proved that $\zeta(s)$ has a meromorphic continuation to $\sigma > 0$, and has one simple pole at $s = 1$. It follows from this that also $\frac{\zeta'(s)}{\zeta(s)}$ is meromorphic for $\sigma > 0$, and that $\frac{\zeta'(s)}{\zeta(s)}$ has a simple pole at $s = 1$ with residue -1, and thus

$$\lim_{s\to 1^+} \ -(s-1)\frac{\zeta'(s)}{\zeta(s)} = 1.$$

Hence $\lambda \le 1 \le \Lambda$, and this implies the theorem in view of Proposition 6.2. □

6.3. **Further asymptotic results.** Mertens proved the following in 1874:

Proposition 6.5. *There is a real constant A such that*

$$\sum_{p\le x}\frac{1}{p} = \log\log x + A + O\big((\log x)^{-1}\big) \qquad \text{as } \ x \to \infty.$$

Proof. Recall our main identity $T(x) = \log(\lfloor x\rfloor!) = \sum_{m\le x}\Lambda(m)\left\lfloor \frac{x}{m}\right\rfloor$, see (257). The contribution from those m's which are prime numbers is:

$$\sum_{p\le x}(\log p)\left\lfloor\frac{x}{p}\right\rfloor = \sum_{p\le x}(\log p)\Big(\frac{x}{p}+O(1)\Big) = x\sum_{p\le x}\frac{\log p}{p} + O(x),$$

since $\sum_{p\le x}\log p = \vartheta(x) = O(x)$ by Theorem 6.3 and Proposition 6.2. The contribution from all other m's is:

$$= \sum_{p\le x}\sum_{\substack{r\ge 2\\(p^r\le x)}}(\log p)\left\lfloor\frac{x}{p^r}\right\rfloor \le x\sum_{p\le x}(\log p)\sum_{\substack{r\ge 2\\(p^r\le x)}}p^{-r} \le x\sum_{p\le x}(\log p)\frac{p^{-2}}{1-p^{-1}} \le 2x\sum_{p\le x}\frac{\log p}{p^2}$$
$$= O(x),$$

since the last sum is convergent even if it is taken over all positive integers. Hence from $T(x) = x\log x - x + O(\log x)$ (see (258)) we conclude, after dividing with x:

$$A(x) := \sum_{p\le x}\frac{\log p}{p} = \log x + O(1), \qquad \forall x \ge 2.$$

Hence, writing $B(x) = \log x + r(x)$ where we know that $|r(x)|$ is less than a constant K for all $x \geq 2$:

$$\begin{aligned}\sum_{p\leq x}\frac{1}{p} &= \int_{2-}^{x}\frac{1}{\log y}\,dB(y) = \Big[\frac{B(y)}{\log y}\Big]_{y=2-}^{y=x} + \int_{2-}^{x}\frac{B(y)}{y(\log y)^2}\,dy \\ &= 1 + \frac{r(x)}{\log x} + \int_2^x \frac{dy}{y\log y} + \int_2^x \frac{r(y)\,dy}{y(\log y)^2} \\ &= 1 + O\big((\log x)^{-1}\big) + \big[\log\log y\big]_{y=2}^{y=x} + \int_2^\infty \frac{r(y)\,dy}{y(\log y)^2} - O\Big(\int_x^\infty \frac{dy}{y(\log y)^2}\,dy\Big) \\ &= \log\log x + A + O\big((\log x)^{-1}\big), \qquad \text{as } x\to\infty.\end{aligned}$$

where A is a certain constant. □

Remark 6.4. The above proof shows that $A = 1 - \log\log 2 + \int_2^\infty \frac{r(y)}{y(\log y)^2}\,dy$. However, it is possible to give a much simpler explicit formula for A, as follows.

Proposition 6.6. *The constant A in Proposition 6.5 is given by*

$$A = \gamma + \sum_p \Big(\log(1-p^{-1}) + p^{-1}\Big), \tag{263}$$

where $\gamma = -\Gamma'(1) = 0.577...$ is Euler's constant. (The sum is rapidly convergent since $\log(1-p^{-1}) + p^{-1} = O(p^{-2})$ as $p\to\infty$.)

Lemma 6.7.

$$Res_{s=1}\zeta(s) = 1.$$

Proof. Recall that $\zeta(s)$ has a simple pole at $s = 1$ (cf. Ex. 3.6 on p. 41). Furthermore, for $s > 1$ we have $\zeta(s) = \sum_{n=1}^\infty n^{-s} \geq \int_1^\infty x^{-s}\,dx = (s-1)^{-1}$ and $\zeta(s) = \sum_{n=1}^\infty n^{-s} \leq 1 + \int_1^\infty x^{-s}\,dx = 1 + (s-1)^{-1}$. Hence $\lim_{s\to 1}(s-1)\zeta(s) = 1$, and this proves the lemma. □

Proof of Proposition 6.6. (We borrow from Ingham [34, p. 23].) The idea is to investigate in two ways the behaviour of $g(s) = \sum_p p^{-s}$ when $s \to 1^+$. Let us write

$$A(x) = \sum_{p\leq x}\frac{1}{p} = \log\log x + A + R(x),$$

where we know from Proposition 6.5 that $R(x) = O\big((\log x)^{-1}\big)$ as $x\to\infty$, and hence in fact $R(x) = O\big((\log x)^{-1}\big)$ for all $x \geq 2$. Then on the one hand side, for any $s > 1$ we have

$$\begin{aligned}g(s) = \sum_p p^{-s} &= \int_{2-}^\infty x^{1-s}\,dA(x) = (s-1)\int_2^\infty x^{-s}A(x)\,dx. \\ &= (s-1)\int_2^\infty \frac{\log\log x}{x^s}\,dx + (s-1)\int_2^\infty \frac{A}{x^s}\,dx + (s-1)\int_2^\infty \frac{R(x)}{x^s}\,dx = I_1 + I_2 + I_3,\end{aligned}$$

say. Regarding I_3, since $\lim_{x\to\infty} R(x) = 0$ there is, for any given $\varepsilon > 0$, some $x_0 > 2$ such that $|R(x)| < \varepsilon$ for all $x > x_0$; hence

$$\big|I_3\big| < (s-1)\int_2^{x_0} |R(x)|\,dx + (s-1)\int_{x_0}^{\infty} \frac{\varepsilon}{x^s}\,dx < (s-1)\int_2^{x_0} |R(x)|\,dx + \varepsilon,$$

and this is $< 2\varepsilon$ for all $s > 1$ sufficiently close to 1. Hence $I_3 \to 0$ when $s \to 1^+$. Next, in I_1 and I_2 we replace the lower limit of integration by 1; this involves an error which tends to 0 as $s \to 0^+$. We also make the substitution $x = e^{y/(s-1)}$ in I_1. This gives:

$$g(s) = -\log(s-1) + \int_0^{\infty} e^{-y}\log y\,dy + A + o(1) = -\log(s-1) - \gamma + A + o(1) \tag{264}$$

as $s \to 1^+$. (The identity $\int_0^\infty e^{-y}\log y\,dy = \Gamma'(1) = -\gamma$ follows from $\Gamma(s) = \int_0^\infty e^{-y}y^{s-1}\,dy$ by differentiating under the integral sign. See §8.2 for more facts about γ.)

On the other hand we can use the Euler product for $\zeta(s)$ to understand $\sum_p p^{-s}$ as $s \to 1^+$. Indeed, we have already seen in the first lecture that $\log\zeta(s) = \sum_p p^{-s} + O(1)$ for all $s > 1$; see (7) and (9). In fact, as $s \to 1^+$ we have

$$\log\zeta(s) - \sum_p p^{-s} = \sum_p \sum_{m=2}^{\infty} m^{-1}p^{-ms} \to \sum_p \sum_{m=2}^{\infty} m^{-1}p^{-m} = \sum_p \Big(-\log(1-p^{-1}) - p^{-1}\Big), \tag{265}$$

since the computation in (9) really works for all $s \geq 1$, and shows that the sum $\sum_p \sum_{m=2}^\infty m^{-1}p^{-ms}$ is uniformly convergent for $s \geq 1$. Hence, as $s \to 1^+$:

$$\begin{aligned} g(s) = \sum_p p^{-s} &= \log\zeta(s) + \sum_p \Big(\log(1-p^{-1}) + p^{-1}\Big) + o(1) \\ &= -\log(s-1) + \sum_p \Big(\log(1-p^{-1}) + p^{-1}\Big) + o(1), \end{aligned} \tag{266}$$

since $\lim_{s\to 1} \log\big((s-1)\zeta(s)\big) = 0$ by Lemma 6.7. Comparing (264) and (266) we obtain (263). □

Another result of Mertens which is of interest in connection with Dirichlet's work on primes in an arithmetic progression is that the infinite series $\sum_p \frac{\chi(p)}{p}$ converges, for any nonprincipal Dirichlet character χ. Before proving this, let us note that we already know that the series $\sum_p \frac{\chi(p)}{p^s}$ is absolutely convergent for any $s > 1$, and by (22) we have (in analogy with (265)) as $s \to 1^+$:

$$\begin{aligned} \log L(s,\chi) - \sum_p \frac{\chi(p)}{p^s} = \sum_p \sum_{m=2}^{\infty} m^{-1}\chi(p^m)p^{-ms} &\to \sum_p \sum_{m=2}^{\infty} m^{-1}\chi(p^m)p^{-m} \\ &= -\sum_p \Big(\log\big(1-\chi(p)p^{-1}\big) + \chi(p)p^{-1}\Big). \end{aligned}$$

where all the sums are absolutely convergent (uniformly over $s \geq 1$). Hence

$$\sum_p \frac{\chi(p)}{p^s} = \log L(s,\chi) + \sum_p \Big(\log\big(1 - \chi(p)p^{-1}\big) + \chi(p)p^{-1} \Big) + o(1), \qquad \text{as } s \to 1^+. \tag{267}$$

However, as suggestive as this is, it does not directly imply the convergence of $\sum_p \frac{\chi(p)}{p}$!

Proposition 6.8. *If χ is any nonprincipal character mod q then $\sum_p \frac{\chi(p)}{p}$ converges, if we add over the primes p in increasing order. In fact $\sum_{p \geq x} \frac{\chi(p)}{p} = O\big((\log x)^{-1}\big)$ as $x \to \infty$, where the implied constant may depend on q.*

(Note that $\sum_p \frac{\chi(p)}{p}$ is *not* absolutely convergent, since $\sum_p \frac{1}{p} = \infty$. Hence by a well-known fact about conditionally convergent series we really need to specify how the terms are ordered; other ways of ordering the terms can lead to other values of the sum, and also to a divergent sum.)

Proof. We start by studying the Dirichlet series for $L'(s,\chi)$ and its partial sums. Writing out the identity $-L'(s,\chi) = -\frac{L'(s,\chi)}{L(s,\chi)} L(s,\chi)$ gives (cf. Corollary 3.8 and (113))

$$\sum_{n=1}^\infty \frac{\chi(n)\log n}{n^s} = \Big(\sum_{m=1}^\infty \frac{\chi(m)\Lambda(m)}{m^s} \Big) \sum_{k=1}^\infty \frac{\chi(k)}{k^s} = \sum_{m=1}^\infty \sum_{k=1}^\infty \frac{\chi(m)\Lambda(m)\chi(k)}{(mk)^s}, \tag{268}$$

where all the sums and double sums are absolutely convergent for $\sigma > 1$. Now by Proposition 3.10 the leftmost and the rightmost Dirichlet series must have identical coefficients (i.e. $\chi(n)\log n = \sum_{\substack{m,k \geq 1 \\ mk = n}} \chi(m)\Lambda(m)\chi(k)$, $\forall n \in \mathbb{Z}^+$) and hence also every partial sum must agree:

$$\sum_{n \leq x} \frac{\chi(n)\log n}{n^s} = \sum_{\substack{m \geq 1 \\ mk \leq x}} \sum_{k \geq 1} \frac{\chi(m)\Lambda(m)\chi(k)}{(mk)^s} = \sum_{m \leq x} \frac{\chi(m)\Lambda(m)}{m^s} \sum_{k \leq x/m} \frac{\chi(k)}{k^s}, \qquad \forall x > 0,\ s \in \mathbb{C}.$$

We apply this for $s = 1$. From Example 3.5 we know that $\sum_{k=1}^\infty \frac{\chi(k)}{k} = L(1,\chi)$, and the error for a finite partial sum can be estimated as follows for any $B \geq 1$ (write $A(x) = \sum_{1 \leq n \leq x} \chi(n)$ as in Example 3.5 and recall $A(x) = O(1)$ for all $x \geq 1$):

$$\sum_{k > B} \frac{\chi(k)}{k} = \int_B^\infty \frac{1}{x}\, dA(x) = \Big[\frac{A(x)}{x} \Big]_{x=B}^{x=\infty} + \int_B^\infty \frac{A(x)}{x^2}\, dx = O\Big(\frac{1}{B}\Big) + \int_B^\infty \frac{O(1)}{x^2}\, dx = O\Big(\frac{1}{B}\Big).$$

Thus $\sum_{k \leq x/m} \frac{\chi(k)}{k} = L(1,\chi) + O(\frac{m}{x})$ and hence we conclude:

$$\sum_{n \leq x} \frac{\chi(n)\log n}{n} = \sum_{m \leq x} \frac{\chi(m)\Lambda(m)}{m} \Big(L(1,\chi) + O\Big(\frac{m}{x}\Big) \Big),$$

where the big-O-terms add up to

$$O\Big(x^{-1}\sum_{m\le x}\Lambda(m)\Big) = O\Big(x^{-1}\psi(x)\Big) = O(1),$$

by Theorem 6.3. Hence, since the infinite sum $\sum_{n=1}^{\infty}\frac{\chi(n)\log n}{n}$ converges (to $-L'(1,\chi)$; cf. Corollary 3.8 and Example 3.5), and $L(1,\chi)\neq 0$, we conclude

$$\sum_{m\le x}\frac{\chi(m)\Lambda(m)}{m} = O(1), \qquad \forall x\ge 1.$$

Note that the contribution to the above sum from all non-prime m is bounded in absolute value by

$$\sum_p\sum_{r=2}^{\infty}\frac{\Lambda(p^r)}{p^r} = \sum_p(\log p)\frac{p^{-2}}{1-p^{-1}} < 2\sum_p\frac{\log p}{p^2} < \infty.$$

Hence, by subtracting off the contribution from non-prime m, we get:

$$A(x) := \sum_{p\le x}\frac{\chi(p)\log p}{p} = O(1), \qquad \forall x\ge 1.$$

From this we get by partial summation, for any $2\le M\le N$:

$$\begin{aligned}\sum_{M<p\le N}\frac{\chi(p)}{p} &= \int_M^N \frac{1}{\log x}\,dA(x) = \Big[\frac{A(x)}{\log x}\Big]_{x=M}^{x=N} + \int_M^N \frac{A(x)}{x(\log x)^2}\,dx\\ &= O\big((\log M)^{-1}\big) + O\big((\log N)^{-1}\big) + O\Big(\int_M^N\frac{dx}{x(\log x)^2}\Big)\\ &= O\big((\log M)^{-1}\big) + O\Big((\log M)^{-1} - (\log N)^{-1}\Big) = O\big((\log M)^{-1}\big).\end{aligned}$$

This tends to 0 as $M\to\infty$ (uniformly over all $N\ge M$). Hence the series $\sum_p\frac{\chi(p)}{p}$ is indeed convergent. The bound $\sum_{p\ge x}\frac{\chi(p)}{p} = O\big((\log x)^{-1}\big)$ as $x\to\infty$ also follows from the above computation. □

Corollary 6.9. *Given $q\ge 1$ and a with $(a,q)=1$ there is a real constant $A(q,a)$ such that*

$$\sum_{\substack{p\le x\\ p\equiv a\ (mod\ q)}}\frac{1}{p} = \frac{1}{\phi(q)}\log\log x + A(q,a) + O\big((\log x)^{-1}\big), \qquad as\ \ x\to\infty.$$

(The implied constant may depend on q, but not on x.)

Proof. Using first Lemma 1.5 and then Proposition 6.5 and Proposition 6.8 we have

$$\sum_{\substack{p\le x\\ p\equiv a \,(\mathrm{mod}\ q)}} \frac{1}{p} = \frac{1}{\phi(q)} \sum_{\chi\in X_q} \overline{\chi(a)}\Big(\sum_{p\le x} \frac{\chi(p)}{p}\Big)$$

$$= \frac{1}{\phi(q)}\big(\log\log x + A + O\big((\log x)^{-1}\big)\big) + \frac{1}{\phi(q)} \sum_{\chi\in X_q\setminus\{\chi_0\}} \overline{\chi(a)}\Big(\sum_{p} \frac{\chi(p)}{p} - O\big((\log x)^{-1}\big)\Big).$$

This gives the stated formula, with $A(q,a) = \frac{1}{\phi(q)}\Big\{A + \sum_{\chi\in X_q\setminus\{\chi_0\}} \overline{\chi(a)}\big(\sum_p \frac{\chi(p)}{p}\big)\Big\}$. □

6.4. **Riemann's memoir.** In his epoch-making memoir of 1860 (his only paper on the theory of numbers) Riemann showed that the key to the deeper investigation of the distribution of the primes lies in the study of $\zeta(s)$ as a function of the complex variable s. More than 30 years were to elapse, however, before any of Riemann's conjectures were proved, or any specific results about primes were established on the lines which he had indicated.

Riemann proved two main results:

(a) The function $\zeta(s)$ can be continued analytically over the whole complex plane and is then meromorphic, its only pole being a simple pole at $s = 1$ with residue 1. In other words, $\zeta(s) - (s-1)^{-1}$ is an entire function.

(b) $\zeta(s)$ satisfies the functional equation

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \pi^{-\frac{1}{2}(1-s)}\Gamma(\tfrac{1}{2}(1-s))\zeta(1-s)$$

which can be expressed by saying that the function on the left is an even function of $s - \frac{1}{2}$. The functional equation allows the properties of $\zeta(s)$ for $\sigma < 0$ to be inferred from its properties for $\sigma > 1$. In particular, the only zeros of $\zeta(s)$ for $\sigma < 0$ are the poles of $\Gamma(\frac{1}{2}s)$, that is, at the points $s = -2, -4, -6, \ldots$. These are called the *trivial zeros*. The remainder of the plane, where $0 \le \sigma \le 1$, is called *the critical strip*.

Riemann further made a number of remarkable conjectures.

(I) $\zeta(s)$ has infinitely many zeros in the critical strip. These will necessarily be placed symmetrically with respect to the real axis, and also with respect to the central line $\sigma = \frac{1}{2}$ (the latter becausee of the functional equation).

(II) The number $N(T)$ of zeros of $\zeta(s)$ in the critical strip with $0 < t \le T$ satisfies the asymptotic relation

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T) \qquad \text{as } T\to\infty. \tag{269}$$

This was proved by von Mangoldt. We shall come to the proof in §12.1.

(III) The entire function $\xi(s)$ defined by

$$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) \tag{270}$$

(entire because it has no pole for $\sigma \geq \frac{1}{2}$ and is an even function of $s - \frac{1}{2}$ has the product expansion

$$\xi(s) = e^{A+Bs}\prod_{\rho}\Big(1-\frac{s}{\rho}\Big)e^{s/\rho} \tag{271}$$

where A and B are certain constants and ρ runs through the zeros of $\zeta(s)$ in the critical strip. This was proved by Hadamard in 1893, as also (I) above. It played an important part in the proofs of the prime number theorem by Hadamard and de la Vallée Poussin. We shall come to the proof in §(10).

(IV) There is an explicit formula for $\pi(x) - \mathrm{Li}(x)$, valid for $x > 1$, the most important part of which consists of a sum over the complex zeros ρ of $\zeta(s)$. As this is somewhat complicated to state, we give instead the closely related but somewhat simples formula for $\psi(x) - x$:

$$\psi(x) - x = -\sum_{\rho}\frac{x^{\rho}}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \tfrac{1}{2}\log(1-x^{-2}). \tag{272}$$

This was proved by von Mangoldt in 1895 (as was Riemann's original formula), and we give the proof in §13. In interpreting (272) two conventions have to be observed: first, in the sum over ρ the terms ρ and $\overline{\rho}$ are to be taken together, and second, if x is an integer, the last term $\Lambda(x)$ in the sum $\psi(x) = \sum_{n\leq x}\Lambda(n)$ defining $\psi(x)$ is to be replaced with $\frac{1}{2}\Lambda(x)$.

(V) The famous Riemann Hypothesis, still undecided: that the zeros of $\zeta(s)$ in the critical strip all lie on the central line $\sigma = \frac{1}{2}$. It was proved by Hardy in 1914 [28] that infinitely many of the zeros lie on the line, and by A. Selberg in 1942 [62] that a positive proportion at least of all the zeros lie on the line. The constant was sharpened by Levinson 1974 [41] who proved that more than $\frac{1}{3}$ of the zeros lie on the line; the best known result today in this direction is that of Conrey 1989 [12], stating that more than $\frac{2}{5}$ of the zeros lie on the line.

6.5. **Problems.**

Problem 6.1. Prove (assuming (246)) that if $A(x)$ is defined by the relation $\pi(x) = \frac{x}{\log x - A(x)}$, then $\lim_{x\to\infty} A(x) = 1$. (Cf. Davenport, pp. 54-55.)

Problem 6.2. Prove the formula (257) by identifying coefficients in the Dirichlet series identity $\frac{\zeta'(s)}{\zeta(s)}\zeta(s) = \zeta'(s)$. (Cf. Davenport pp. 55-56.)

Problem 6.3. By mimicking the proof of Theorem 6.3 but using $T(x) - T(\frac{1}{2}x) - T(\frac{1}{3}x) - T(\frac{1}{5}x) + T(\frac{1}{30}x)$ instead of $T(x) - 2T(\frac{1}{2}x)$, prove that

$$0.9212\frac{x}{\log x} < \pi(x) < 1.1056\frac{x}{\log x}.$$

for all sufficiently large x.

Problem 6.4. Let χ be a nonprincipal Dirichlet character. Prove that Proposition 6.8 implies that the infinite product $\prod_p (1 - \chi(p)p^{-1})^{-1}$ converges, if we multiply over the primes p in increasing order, and that the limit equals $L(1, \chi)$.

Problem 6.5. We will later prove the prime number theorem for arithmetic progressions, which says that for any $q \geq 1$ and a with $(a, q) = 1$, if we write

$$\pi(x; q, a) = \#\big\{p \,:\, p \text{ is a prime number} \leq x \text{ and } p \equiv a \pmod q\big\},$$

then there is a constant $c > 0$ such that $\pi(x; q, a) = \frac{1}{\phi(q)} \operatorname{Li} x + O\big(xe^{-c\sqrt{\log x}}\big)$ as $x \to \infty$. (Both c and the implied constant may depend on q.)

Prove that this implies the following strengthening of Proposition 6.8: For any nonprincipal character χ mod q we have $\sum_{p>x} \frac{\chi(p)}{p} = O\big(\sqrt{\log x} \cdot e^{-c\sqrt{\log x}}\big)$ as $x \to \infty$.

Also prove that if $\pi(x; q, a) = \frac{1}{\phi(q)} \operatorname{Li} x + O\big(\sqrt{x} \log x\big)$ as $x \to \infty$ (as would follow from the Generalized Riemann Hypothesis, cf. §15), then we have $\sum_{p>x} \frac{\chi(p)}{p} = O\big(x^{-\frac{1}{2}} \log x\big)$ as $x \to \infty$.

[Hint. Start by writing $\sum_{M<p\leq N} \frac{\chi(p)}{p} = \sum_{\substack{a \bmod q \\ (a,q)=1}} \chi(a) \int_M^N \frac{1}{x}\, d\pi(x; q, a)$.]

Problem 6.6. Prove $\log(n!) = n \log n - n + O(\log n)$ for all $n \geq 2$. (This is a weak form of Stirling's formula (256); note that this weak version is sufficient for all applications in the present section.)

[Hint. Note $\log(n!) = \sum_{m=1}^n \log m$; try to estimate this sum using integration by parts.]

7. The prime number theorem

(This presentation I have mainly borrowed from Ingham [34, Ch. II].)

In this lecture we will give a proof of the prime number theorem:

Theorem 7.1. $\pi(x) \sim \frac{x}{\log x}$ *as* $x \to \infty$.

In later lectures we will prove more precise results, but today we only aim at proving the above theorem.

To start with, I give an 1-page outline of the proof. Recall Euler's identity,

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p (1-p^{-s})^{-1} \qquad (\sigma > 1). \tag{273}$$

We wish to somehow "invert" this so as to extract more explicit information on the set of prime numbers p appearing in the right hand side! We first rewrite Euler's identity by applying logarithmic differentiation, to get

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = s\int_1^{\infty} \frac{\psi(x)}{x^{s+1}}\,dx \qquad (\sigma > 1) \tag{274}$$

(cf. Example 3.7, in particular (116)). We also (for technical reasons) integrate by parts once more, to get

$$-\frac{\zeta'(s)}{\zeta(s)} = s(s+1)\int_1^{\infty} \frac{\psi_1(x)}{x^{s+2}}\,dx \qquad (\sigma > 1), \tag{275}$$

where

$$\psi_1(x) = \int_0^x \psi(u)\,du = \int_1^x \psi(u)\,du = \sum_{1\le n\le x} (x-n)\Lambda(n). \tag{276}$$

It is *this* relation (275) which we actually invert (instead of (273)), so as to get a formula for $\psi_1(x)$ in terms of the Riemann zeta function; from that formula we can eventually deduce $\psi_1(x) \sim \frac{1}{2}x^2$ as $x \to \infty$. This turns out to imply $\psi(x) \sim x$ as $x \to \infty$, and by Proposition 6.2 this implies the prime number theorem, Theorem 7.1.

Now how do we invert (275)? The trick is to note that the integral $\int_1^{\infty} \frac{\psi_1(x)}{x^{s+2}}\,dx$ can be viewed as a *Laplace transform* of the function ψ_1, and hence we can use the *inverse Laplace transform* to get a formula for ψ_1. In fact this special format (such as $\int_1^{\infty} \frac{\psi_1(x)}{x^{s+2}}\,dx$) of a Laplace transform occurs very often in number theory and has a special name; it is called a *Mellin Transform*, and the inversion formula is called the *Mellin inversion formula*. Anyway, the result of this inversion is

$$\psi_1(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\,ds \qquad (x > 0, c > 1), \tag{277}$$

where the path of integration is the straight vertical line $\sigma = c$. It turns out that $\frac{1}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)$ is nicely decaying as $t = \operatorname{Im} s \to \pm\infty$, so that the integral is absolutely convergent.

Finally, to obtain asymptotic information about $\psi_1(x)$ we use the Cauchy residue theorem to "*move the contour*" to the left. The function $\frac{x^{s+1}}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)$ has a simple pole at $s = 1$, and one checks that $\operatorname{Res}_{s=1}\frac{x^{s+1}}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big) = \frac{1}{2}x^2$, and this residue turns out to be exactly responsible for the asymptotic relation $\psi_1(x) \sim \frac{1}{2}x^2$! The central fact which one must check in order to make this argument work is that $\zeta(s) \neq 0$ for all $s \neq 1$ on the line $\sigma = 1$ (we already know that $\zeta(s) \neq 0$ when $\sigma > 1$, cf. (5) and Theorem 2.2), to ensure that there are no *other* poles of $\frac{x^{s+1}}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)$ that we have to worry about. We also have to spend some work on proving good bounds on $-\frac{\zeta'(s)}{\zeta(s)}$ as $t \to \pm\infty$, to justify the change of contour.

This ends the outline of the proof.

7.1. **Analytic continuation of $\zeta(s)$.** We proved in Example 3.6 that $\zeta(s)$ has a meromorphic continuation to $\sigma > 0$ with one simple pole at $s = 1$. We now need slightly more information about this continuation; in particular we need to know that there is no other pole on the line $\sigma = 1$. We get this by a new method of rewriting $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$.

Proposition 7.2. *The function $\zeta(s)$ has a meromorphic continuation to $\sigma > 0$ with one simple pole with residue 1 at $s = 1$, and* ***no other poles.***

Proof. For any s with $\sigma > 1$ we have

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \int_{1-}^{\infty} x^{-s}\, d\lfloor x\rfloor = \lim_{X\to\infty}\Big(\Big[x^{-s}\lfloor x\rfloor\Big]_{x=1-}^{x=X} + s\int_1^X x^{-s-1}\lfloor x\rfloor\, dx\Big) = s\int_1^{\infty}\frac{\lfloor x\rfloor}{x^{s+1}}\, dx.$$

(This is also a direct consequence of Theorem 3.11.) Writing $\lfloor x\rfloor = x - (x)$, so that $0 \leq (x) < 1$, we obtain

$$\zeta(s) = s\int_1^{\infty}\Big(x^{-s} - \frac{(x)}{x^{s+1}}\Big)\, dx = \frac{s}{s-1} - s\int_1^{\infty}\frac{(x)}{x^{s+1}}\, dx. \tag{278}$$

This formula has been proved for any s with $\sigma > 1$, but note that the last integral is in fact uniformly absolutely convergent in any half-plane $\{\sigma \geq \delta\}$ for any fixed $\delta > 0$, since $\big|(x)/x^{s+1}\big| \leq 1/x^{\sigma+1}$. Hence the last integral represents an analytic function in $\{\sigma > 0\}$, and thus the last expression represents a meromorphic function in $\{\sigma > 0\}$ which has a simple pole at $s = 1$ with residue $\operatorname{Res}_{s=1}\frac{s}{s-1} = 1$,[17] and no other poles. □

Using the formula (278) (slightly generalized) we can now also obtain some basic bounds on the size of $\zeta(s)$ in $\{\sigma > 0\}$, and in particular in the closed half-plane $\{\sigma \geq 1\}$.

[17]Thus we have an alternative proof of Lemma 6.7.

Proposition 7.3. *We have*

$$|\zeta(s)| \ll \log t, \qquad \forall \sigma \geq 1,\ t \geq 2; \tag{279}$$

$$|\zeta'(s)| \ll (\log t)^2, \qquad \forall \sigma \geq 1,\ t \geq 2; \tag{280}$$

$$|\zeta(s)| \ll_\delta t^{1-\delta}, \qquad \forall \sigma \geq \delta,\ t \geq 1; \tag{281}$$

if $0 < \delta < 1$.

(The notation "$\ll_\delta$" means that the implied constant may depend on δ. Note that in all three estimates the implied constant is of course independent of σ and t.)

Proof. Just as in the proof of Proposition 7.2 one shows that for all $X \geq 1$ and all s with $\sigma > 1$,

$$\sum_{n>X} n^{-s} = -\frac{\lfloor X \rfloor}{X^s} + s\int_X^\infty \frac{\lfloor x \rfloor}{x^{s+1}}\,dx = -X^{1-s} + \frac{(X)}{X^s} + \frac{s}{s-1}X^{1-s} - s\int_X^\infty \frac{(x)}{x^{s+1}}\,dx,$$

and hence

$$\zeta(s) = \sum_{1\leq n\leq X} n^{-s} + \frac{1}{(s-1)X^{s-1}} + \frac{(X)}{X^s} - s\int_X^\infty \frac{(x)}{x^{s+1}}\,dx. \tag{282}$$

Note that the integral converges nicely and gives an analytic function of s in the whole region $\{\sigma > 0\}$; hence the last formula in fact holds for all $s \neq 1$ with $\sigma > 0$. It follows that, for any such s with $\sigma > 0$ and $t \geq 1$:

$$|\zeta(s)| \leq \sum_{1\leq n\leq X} n^{-\sigma} + \frac{1}{tX^{\sigma-1}} + \frac{1}{X^\sigma} + |s|\int_X^\infty \frac{1}{x^{\sigma+1}}\,dx \leq \sum_{1\leq n\leq X} n^{-\sigma} + \frac{1}{tX^{\sigma-1}} + \frac{1}{X^\sigma} + \Big(1+\frac{t}{\sigma}\Big)\frac{1}{X^\sigma},$$

since $|s| \leq \sigma + t$. If $\sigma \geq 1$ then we conclude

$$|\zeta(s)| \leq \sum_{1\leq n\leq X} n^{-1} + \frac{1}{t} + \frac{1}{X} + \frac{1+t}{X} \leq \Big(1 + \int_1^{\lfloor X\rfloor}\frac{dx}{x}\Big) + 3 + \frac{t}{X} \leq \big(1+\log X\big) + 3 + \frac{t}{X}$$

since $t \geq 1$, $X \geq 1$. Taking $X = t$ we obtain (279).

If $\sigma \geq \delta$ where $0 < \delta < 1$, then

$$|\zeta(s)| \leq \sum_{1\leq n\leq X} n^{-\delta} + \frac{1}{tX^{\delta-1}} + \Big(2+\frac{t}{\delta}\Big)\frac{1}{X^\delta} < \int_0^{\lfloor X\rfloor}\frac{dx}{x^\delta} + \frac{X^{1-\delta}}{t} + \frac{3t}{\delta X^\delta} \leq \frac{X^{1-\delta}}{1-\delta} + X^{1-\delta} + \frac{3t}{\delta X^\delta}.$$

Taking $X = t$, as before, we deduce

$$|\zeta(s)| < t^{1-\delta}\Big(\frac{1}{1-\delta} + 1 + \frac{3}{\delta}\Big) \qquad (\forall \sigma \geq \delta,\ t \geq 1), \tag{283}$$

which implies (281).

The inequality (280) may be deduced in a similar way from the formula resulting from the differentiation of (282). Or we may argue as follows. Let $s_0 = \sigma_0 + t_0 i$ be any point in

the region $\sigma \geq 1$, $t \geq 2$, and C a circle with centre s_0 and radius $\rho < \frac{1}{2}$. Then by Cauchy's integral formula for $\zeta'(s_0)$ we have

$$\big|\zeta'(s_0)\big| = \Big|\frac{1}{2\pi i}\int_C \frac{\zeta(s)\,ds}{(s-s_0)^2}\Big| \leq \frac{M}{\rho}$$

where M is the maximum of $\big|\zeta(s)\big|$ on C. Now for all points s on C we have $\sigma \geq \sigma_0 - \rho \geq 1-\rho$ and $1 < t < 2t_0$, and hence by (283),

$$M \leq (2t_0)^\rho\Big(\frac{1}{\rho} + 1 + \frac{3}{1-\rho}\Big) < \frac{10t_0^\rho}{\rho},$$

since $\rho < 1-\rho < 1$ and $2^\rho < 2$. Hence

$$\big|\zeta'(s_0)\big| < \frac{10t_0^\rho}{\rho^2}.$$

Take $\rho = \frac{1}{2+\log t_0}$. Then $t_0^\rho = e^{\rho \log t_0} < e$, so that

$$\big|\zeta'(s_0)\big| < 10e\big(2 + \log t_0\big)^2.$$

This implies (280) since s_0 is any point in the region $\sigma \geq 1$, $t \geq 2$. □

Note that we certainly do not claim that the inequalities (279), (280), (281) are the best possible of their kind.

7.2. **Zeros.**

Theorem 7.4. *$\zeta(s)$ has no zeros on the line $\sigma = 1$. Furthermore there is an absolute constant $A > 0$ such that*

$$\frac{1}{\zeta(s)} = O\big((\log t)^A\big) \tag{284}$$

uniformly for $\sigma \geq 1$, as $t \to \infty$.

Proof. (Cf. Hadamard (1896), Mertens (1898).) The proof is based on the elementary inequality

$$3 + 4\cos\theta + \cos 2\theta \geq 0, \tag{285}$$

which holds for all real θ, since the left-hand side is $2(1+\cos\theta)^2$. By (110) we have, for $\sigma > 1$:

$$\log\big|\zeta(s)\big| = \operatorname{Re}\sum_{n=2}^{\infty} a_n n^{-s} = \sum_{n=2}^{\infty} a_n n^{-\sigma}\cos(t\log n); \qquad a_n = \begin{cases} m^{-1} & \text{if } n = p^m \\ 0 & \text{otherwise.}\end{cases}$$

Hence

$$\log\big|\zeta(\sigma)^3\zeta(\sigma+ti)^4\zeta(\sigma+2ti)\big| = \sum_{n=2}^{\infty} a_n n^{-\sigma}\big(3 + 4\cos(t\log n) + \cos(2t\log n)\big) \geq 0$$

by (285), since $a_n \geq 0$. Thus

$$\big((\sigma-1)\zeta(\sigma)\big)^3\Big|\frac{\zeta(\sigma+ti)}{\sigma-1}\Big|^4\big|\zeta(\sigma+2ti)\big| \geq \frac{1}{\sigma-1}, \qquad (\sigma>1). \tag{286}$$

This shows that the point $1+ti$ ($t\neq 0$) cannot be a zero of $\zeta(s)$. For, if it were, then since $\zeta(s)$ is analytic at the points $1+ti$ and $1+2ti$, and has a simple pole (with residue 1) at the point 1, the left-hand side would tend to a finite limit (namely $\big|\zeta'(1+ti)\big|^4\big|\zeta(1+2ti)\big|$), and the right-hand side to infinity, when $\sigma\to 1^+$; contradiction! This proves the first part of the theorem.

In proving the second part we may suppose $1\leq\sigma\leq 2$, for if $\sigma\geq 2$ then we have

$$\big|\zeta(s)^{-1}\big| = \Big|\prod_p(1-p^{-s})\Big| \leq \prod_p(1+p^{-\sigma}) < \zeta(\sigma) \leq \zeta(2).$$

Now if $1<\sigma\leq 2$ and $t\geq 2$, then by (286)

$$(\sigma-1)^3 \leq \big((\sigma-1)\zeta(\sigma)\big)^3\big|\zeta(\sigma+ti)\big|^4\big|\zeta(\sigma+2ti)\big| \ll \big|\zeta(\sigma+ti)\big|^4\log(2t),$$

by (279) in Proposition 7.3. Hence

$$\big|\zeta(\sigma+ti)\big| \geq \frac{(\sigma-1)^{\frac34}}{A_1(\log t)^{\frac14}} \qquad (1\leq\sigma\leq 2,\ t\geq 2) \tag{287}$$

for some absolute constant $A_1>0$ (the inequality being trivial for $\sigma=1$). Now let $1<\eta<2$. Then, if $1\leq\sigma\leq\eta$ and $t\geq 2$,

$$\big|\zeta(\sigma+ti)-\zeta(\eta+ti)\big| = \Big|\int_\sigma^\eta \zeta'(u+ti)\,du\Big| \leq A_2(\eta-1)(\log t)^2$$

for some absolute constant $A_2>0$, by (280) in Proposition 7.3. Hence

$$\big|\zeta(\sigma+ti)\big| \geq \big|\zeta(\eta+ti)\big| - A_2(\eta-1)(\log t)^2 \geq \frac{(\eta-1)^{\frac34}}{A_1(\log t)^{\frac14}} - A_2(\eta-1)(\log t)^2,$$

by (287). Because of (287), the inequality $\big|\zeta(\sigma+ti)\big| \geq \frac{(\eta-1)^{\frac34}}{A_1(\log t)^{\frac14}} - A_2(\eta-1)(\log t)^2$ *also* holds when $\eta\leq\sigma\leq 2$ and $t\geq 2$; hence it is in fact true for all s with $1\leq\sigma\leq 2$, $t\geq 2$. Now choose $\eta=\eta(t)$ so that

$$\frac{(\eta-1)^{\frac34}}{A_1(\log t)^{\frac14}} = 2A_2(\eta-1)(\log t)^2; \qquad \text{i.e. } \ \eta = 1+(2A_1A_2)^{-4}(\log t)^{-9},$$

assuming that t is large enough (say $t>t_0$) to ensure that $\eta<2$. Then

$$\big|\zeta(\sigma+ti)\big| \geq A_2(\eta-1)(\log t)^2 \gg (\log t)^{-7}$$

for $1\leq\sigma\leq 2$ and $t>t_0$. This proves the theorem (with $A=7$). □

7.3. **Fundamental formula.** We now wish to invert the relation (275);

$$-\frac{\zeta'(s)}{\zeta(s)} = s(s+1)\int_1^\infty \frac{\psi_1(x)}{x^{s+2}}\,dx \qquad (\sigma > 1). \tag{288}$$

This can be done by noticing that the right hand side is in fact a Laplace transform: If we substitute $x = e^u$ then we get

$$-\frac{\zeta'(s)}{\zeta(s)} = s(s+1)\int_0^\infty \psi_1(e^u)e^{-(s+1)u}\,du; \tag{289}$$

in other words $-\frac{1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}$ is the Laplace transform of the function $u \mapsto \psi_1(e^u)e^{-u}$. Hence by the formula for the inverse Laplace transform, we should expect that if c is any real number which is greater than the real part of all singularities of $-\frac{1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}$, then

$$\psi_1(e^u)e^{-u} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{su}\Big(\frac{-1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}\Big)\,ds,$$

where the integration is along the vertical line $\operatorname{Re} s = c$ in the complex plane. That is (again writing $x = e^u$): We should expect that

$$\psi_1(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\,ds, \qquad \forall x > 0,\ c > 1. \tag{290}$$

We will now prove that this is actually the case, by reviewing some basic facts about the Laplace transform.

In fact the Laplace transform is nothing more than a (lightly) disguised Fourier transform. Also, the special format of the Laplace transform appearing in (288) is very common in number theory, and is called the *Mellin transform*. Hence in fact we have *three* equivalent formats of the same transform, and we will briefly review all of them in the next three theorems.

Theorem 7.5. *Given $f \in L^1(\mathbb{R})$ we set $\widehat{f}(t) = \int_{-\infty}^\infty f(x)e^{-2\pi itx}\,dx$ (the **Fourier transform** of f). Then $\widehat{f} \in C_0(\mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ which vanish at infinity). In this situation, **if** $\widehat{f} \in L^1(\mathbb{R})$, then*

$$f(x) = \int_{-\infty}^\infty \widehat{f}(t)e^{2\pi ixt}\,dt \qquad \text{for almost all } x \in \mathbb{R}. \tag{291}$$

Also the right hand side is a continuous function of x (vanishing at infinity), and hence the identity (291) *holds at each point x where f is continuous.*

Proof. Cf., e.g., [60, Thm. 9.11]. □

We next give an equivalent statement about the Laplace transform and its inverse:

Theorem 7.6. *Let $c \in \mathbb{R}$ and let f be a measurable function $\mathbb{R} \to \mathbb{C}$ such that the integral $[\mathcal{L}f](s) = \int_{-\infty}^{\infty} f(x)e^{-sx}\,dx$ is absolutely convergent for $s = c$ (and hence for every $s \in \mathbb{C}$ on the line $\sigma = c$.) Also assume that $\int_{-\infty}^{\infty} \big|[\mathcal{L}f](c+it)\big|\,dt < \infty$. Then we have*

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{sx}[\mathcal{L}f](s)\,ds \qquad \text{for almost all } x \in \mathbb{R}. \tag{292}$$

Also the right hand side is a continuous function of x (vanishing at infinity), and hence the identity (292) *holds at each point x where f is continuous.*

Proof. Set $f_1(x) = f(x)e^{-cx}$; then by our assumption we have $f_1 \in L^1(\mathbb{R})$. Note also that

$$[\mathcal{L}f](c+it) = \int_{-\infty}^{\infty} f(x)e^{-(c+it)x}\,dx = \int_{-\infty}^{\infty} f_1(x)e^{-itx}\,dx = \widehat{f_1}(\tfrac{1}{2\pi}t).$$

Now the assumption $\int_{-\infty}^{\infty}\big|[\mathcal{L}f](c+it)\big|\,dt < \infty$ is the same as saying that $\widehat{f_1} \in L^1(\mathbb{R})$. Hence Theorem 7.5 applies, and gives that for almost all $x \in \mathbb{R}$ we have:

$$f(x)e^{-cx} = f_1(x) = \int_{-\infty}^{\infty} \widehat{f_1}(t)e^{2\pi ixt}\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f_1}(\tfrac{1}{2\pi}t)e^{ixt}\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}[\mathcal{L}f](c+it)e^{ixt}\,dt$$

and hence

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[\mathcal{L}f](c+it)e^{(c+it)x}\,dt = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}[\mathcal{L}f](s)e^{sx}\,ds,$$

i.e. (292) holds. Theorem 7.5 also gives the continuity of the right hand side of (292). □

Theorem 7.7. *Let $c \in \mathbb{R}$ and let f be a measurable function $\mathbb{R}^+ \to \mathbb{C}$ such that the integral $[\mathcal{M}f](s) = \int_0^{\infty} f(x)x^s\frac{dx}{x}$ is absolutely convergent for $s = c$ (and hence for every $s \in \mathbb{C}$ on the line $\sigma = c$.) Also assume that $\int_{-\infty}^{\infty}\big|[\mathcal{M}f](c+it)\big|\,dt < \infty$. Then we have*

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}[\mathcal{M}f](s)\,ds \qquad \text{for almost all } x \in \mathbb{R}^+. \tag{293}$$

Also the right hand side is a continuous function of x, and hence the identity (293) *holds at each point x where f is continuous.*

The function $\mathcal{M}f$ is called the *Mellin transform* of f.

Proof. Set $f_1(u) = f(e^{-u})$ $(u \in \mathbb{R})$. Substituting $x = e^{-u}$ in the definition of $[\mathcal{M}f](s)$ we get $[\mathcal{M}f](s) = \int_{-\infty}^{\infty} f(e^{-u})e^{-su}\,du = [\mathcal{L}f_1](s)$, and by assumption this is absolutely convergent at $s = c$, and $\int_{-\infty}^{\infty}\big|[\mathcal{L}f_1](c+it)\big|\,dt < \infty$. Hence by Theorem 7.6 we have, for almost every $u \in \mathbb{R}$:

$$f(e^{-u}) = f_1(u) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{su}[\mathcal{L}f_1](s)\,ds = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{su}[\mathcal{M}f](s)\,ds.$$

Writing here $x = e^{-u}$ we obtain (293). □

Proof of (290). The relation (288)($\Leftrightarrow$(275)) says that $-\frac{1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)} = [\mathcal{M}\psi_1](-s-1)$, with absolute convergence in the Mellin transform, for any $s \in \mathbb{C}$ with $\sigma > 1$. It also follows from (280) in Proposition 7.3 and Theorem 7.4 that $\int_{c-i\infty}^{c+i\infty}\Big|\frac{1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}\Big|\,|ds| < \infty$ holds for every $c > 1$, i.e. $\int_{c_1-i\infty}^{c_1+i\infty}\Big|[\mathcal{M}\psi_1](s_1)\Big|\,|ds_1| < \infty$ holds for every $c_1 < -2$. Hence by Theorem 7.7 we have $\psi_1(x) = \frac{1}{2\pi i}\int_{c_1-i\infty}^{c_1+i\infty} x^{-s_1}[\mathcal{M}\psi_1](s_1)\,ds_1$ for all $c_1 < -2$ and all $x > 0$. Substituting $s_1 = -s-1$ in the last integral we obtain (290) with $c = -1-c_1$ (thus c arbitrary with $c > 1$). □

To conclude this section we now also give an alternative, more direct proof of (290), not using any general facts about the Fourier transform.

Lemma 7.8. *If k is a positive integer and $c > 0$, $y > 0$, then*

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{y^s\,ds}{s(s+1)\cdots(s+k)} = \begin{cases} 0 & \text{if } y \le 1 \\ \frac{1}{k!}\big(1-y^{-1}\big)^k & \text{if } y \ge 1. \end{cases}$$

(The path of integration is the straight vertical line $\sigma = c$.)

Note that the integral is absolutely convergent, since the integrand has absolute value $\le y^c|t|^{-k-1}$ on the line of integration, and $k > 0$.

Proof. Set

$$J = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{y^s\,ds}{s(s+1)\cdots(s+k)}; \qquad J_T = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s\,ds}{s(s+1)\cdots(s+k)}$$

We use Cauchy's theorem of residues to replace the line of integration in J_T by an arc of the circle C having its centre at $s = 0$ and passing through the two points $s = c \pm iT$. If $y \ge 1$, we use the arc C_1 which lies to the left of the line $\sigma = c$, assuming T so large that $R > 2k$, where R is the radius of C. This gives

$$J_T = \frac{1}{2\pi i}\int_{C_1}\frac{y^s\,ds}{s(s+1)\cdots(s+k)} + \sum_{n=-k}^{0}\mathrm{Res}_{s=n}\Big(\frac{y^s}{s(s+1)\cdots(s+k)}\Big). \tag{294}$$

Now for s on C_1 we have $\sigma \le c$ and so $\big|y^s\big| = y^\sigma \le y^c$, since $y \ge 1$; also $|s+n| \ge R-n > \frac{1}{2}R$ for each $n = 0, 1, \dots, k$. Hence

$$\Big|\frac{1}{2\pi i}\int_{C_1}\frac{y^s\,ds}{s(s+1)\cdots(s+k)}\Big| \le \frac{1}{2\pi}\cdot\frac{y^c}{(\frac{1}{2}R)^{k+1}}\cdot 2\pi R = \frac{2^{k+1}y^c}{R^k} < \frac{2^{k+1}y^c}{T^k}.$$

Hence by (294) we have

$$J = \lim_{T\to\infty} J_T = \sum_{n=-k}^{0}\mathrm{Res}_{s=n}\Big(\frac{y^s}{s(s+1)\cdots(s+k)}\Big) = \sum_{n=0}^{k}\frac{y^{-n}}{(-1)^n n!(k-n)!} = \frac{1}{k!}(1-y^{-1})^k,$$

which proves the lemma when $y \geq 1$. The proof is similar in the case $y \leq 1$, except that the right-hand arc C_2 of C is used and no poles are passed over. □

Alternative proof of (290). By (276) and Lemma 7.8 we have (for any $c > 0$, $x > 0$)

$$\psi_1(x) = x \sum_{1 \leq n \leq x} \Big(1 - \frac{n}{x}\Big)\Lambda(n) = x \sum_{n=1}^{\infty} \frac{\Lambda(n)}{2\pi i} \int_{c-\infty i}^{c+\infty i} \frac{(x/n)^s}{s(s+1)}\, ds.$$

If $c > 1$ then the order of summation and integration may be interchanged, since

$$\sum_{n=1}^{\infty} \int_{c-\infty i}^{c+\infty i} \Big| \frac{\Lambda(n)(x/n)^s}{s(s+1)}\, ds \Big| < x^c \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^c} \int_{-\infty}^{\infty} \frac{dt}{c^2+t^2} < \infty.$$

Hence

$$\psi_1(x) = \frac{x}{2\pi i} \int_{c-\infty i}^{c+\infty i} \frac{x^s}{s(s+1)} \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}\, ds = \frac{1}{2\pi i} \int_{c-\infty i}^{c+\infty i} \frac{x^{s+1}}{s(s+1)} \Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\, ds,$$

thus completing the proof. □

7.4. **Asymptotic formula for $\psi_1(x)$.**

Theorem 7.9. *We have*

$$\psi_1(x) \sim \tfrac{1}{2}x^2 \qquad \text{as } \ x \to \infty.$$

Proof. Let us keep $x > 1$. By (290) we have for any $c > 1$:

$$\frac{\psi_1(x)}{x^2} = \int_{(c)} g(s) x^{s-1}\, ds, \tag{295}$$

where (c) denotes the line $\sigma = c$, and

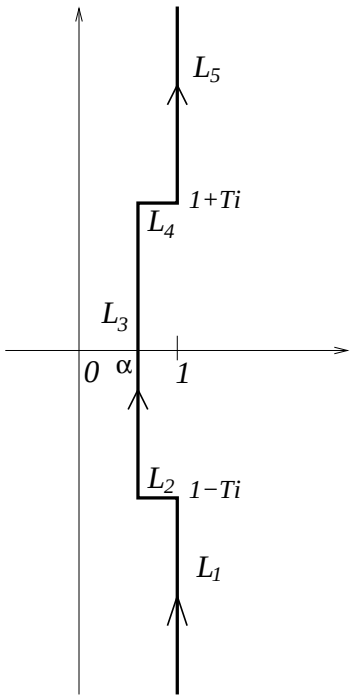
$$g(s) := \frac{1}{2\pi i} \frac{1}{s(s+1)} \Big(-\frac{\zeta'(s)}{\zeta(s)}\Big) = -\frac{1}{2\pi i} \cdot \frac{1}{s(s+1)} \cdot \zeta'(s) \cdot \zeta(s)^{-1}.$$

By Proposition 7.2, Proposition 7.3 and Theorem 7.4, the function $g(s)$ is analytic in the region $\{\sigma \geq 1\}$ except at $s = 1$, and

$$\big|g(s)\big| \ll |t|^{-2} \cdot (\log|t|)^2 \cdot (\log|t|)^A$$

for all s with $\sigma \geq 1$ and $|t|$ large. Hence there is an absolute constant $t_0 > 0$ such that

$$\big|g(s)\big| \leq |t|^{-\frac{3}{2}} \qquad (\forall \sigma \geq 1,\ |t| \geq t_0). \tag{296}$$



Take $\varepsilon > 0$ and let L be the infinite broken line

$$L_1 + L_2 + L_3 + L_4 + L_5$$

shown in the figure, where T is chosen so that

$$\int_T^\infty \big|g(1+ti)\big|\,dt < \varepsilon, \tag{297}$$

and then α $(0 < \alpha < 1)$ is chosen so that the rectangle $\alpha \le \sigma \le 1$, $-T \le t \le T$ contains no zero of $\zeta(s)$. The first choice is possible by (296), and the second because $\zeta(s)$ has no zeros on the line $\sigma = 1$ (by Theorem 7.4) and (being a meromorphic function) at most a finite number of zeros in the region $\frac{1}{2} \le \sigma \le 1$, $-T \le t \le T$. Applying Cauchy's Theorem to (295), we obtain

$$\frac{\psi_1(x)}{x^2} = \frac{1}{2} + \int_L g(s)x^{s-1}\,ds, \tag{298}$$

where the term $\frac{1}{2}$ arises from the pole at $s = 1$ (at which point $\zeta'(s)/\zeta(s)$ has a simple pole with residue -1). [Details regarding (298): By our choice of L the integrand is analytic between and on the lines (c) and L, except at $s = 1$, and if we first integrate round the closed contour bounded by portions of (c) and L and by segments of the lines $t = \pm U$, where $U > \max(t_0, T)$, the integrals along the latter segments are in absolute value

$$(c-1)\sup_{1\le\sigma\le c}\big|g(\sigma \pm Ui)\big|x^{\sigma-1} \le (c-1)U^{-\frac{3}{2}}x^{c-1},$$

by (296), and therefore tend to 0 when $U \to \infty$. Note that (296) also shows that the integral $\int_L g(s)x^{s-1}\,ds$ is absolutely convergent.]

Write

$$\int_L g(s)x^{s-1}\,ds = J_1 + J_2 + J_3 + J_4 + J_5,$$

where $J_1, \dots, J_5$ are the integrals along $L_1, \dots, L_5$, respectively. Since $g(\overline{s})x^{\overline{s}-1} = \overline{g(s)x^{s-1}}$, we have

$$\big|J_1\big| = \big|J_5\big| = \Big|\int_T^\infty g(1+ti)x^{ti}\,dt\Big| \le \int_T^\infty \big|g(1+ti)\big|\,dt < \varepsilon$$

by (297). Also, if we let M be the maximum of of $|g(s)|$ on the finite segments L_2, L_3, L_4, then (since $x > 1$)

$$\big|J_2\big| = \big|J_4\big| = \Big|\int_\alpha^1 g(\sigma+Ti)x^{\sigma+Ti-1}\,d\sigma\Big| \le M\int_\alpha^1 x^{\sigma-1}\,d\sigma < \frac{M}{\log x};$$
$$\big|J_3\big| \le Mx^{\alpha-1}\cdot 2T.$$

Hence by (298)

$$\Big|\frac{\psi_1(x)}{x^2}-\frac{1}{2}\Big| < 2\varepsilon + \frac{2M}{\log x} + \frac{2MT}{x^{1-\alpha}}.$$

Note that this is $< 3\varepsilon$ for all sufficiently large x. Since $\varepsilon > 0$ was arbitrary, this proves the desired result that $\psi_1(x)/x^2 \to \frac{1}{2}$ when $x \to \infty$. □

7.5. **Going from $\psi_1(x)$ to $\psi(x)$.**

Theorem 7.10. *We have $\psi(x) \sim x$ as $x \to \infty$.*

Proof. This follows by a standard technique from the three facts that $\psi_1(x) = \int_1^x \psi(u)\,du$, $\psi(u)$ is increasing, and $\psi_1(x) \sim \frac{x^2}{2}$ as $x \to \infty$ (cf. (276) and Theorem 7.9).

The details are as follows: Let $0 < \alpha < 1 < \beta$. Since $\psi(u)$ is an increasing function of u we have

$$\psi(x) \le \frac{1}{\beta x - x}\int_x^{\beta x} \psi(u)\,du = \frac{\psi_1(\beta x)-\psi_1(x)}{(\beta-1)x},$$

and hence

$$\frac{\psi(x)}{x} \le \frac{\psi_1(\beta x)-\psi_1(x)}{(\beta-1)x^2} = \frac{1}{\beta-1}\Big(\frac{\psi_1(\beta x)}{(\beta x)^2}\beta^2 - \frac{\psi_1(x)}{x^2}\Big).$$

Let $x \to \infty$, keeping β fixed. Then since $\lim_{x\to\infty}\frac{\psi_1(x)}{x^2} = \frac{1}{2}$, we get

$$\limsup_{x\to\infty}\frac{\psi(x)}{x} \le \frac{1}{2}\cdot\frac{\beta^2-1}{\beta-1} = \frac{1}{2}(\beta+1).$$

Similarly, by considering $\int_{\alpha x}^x \psi(u)\,du$, we prove that

$$\liminf_{x\to\infty}\frac{\psi(x)}{x} \ge \frac{1}{2}\cdot\frac{1-\alpha^2}{1-\alpha} = \frac{1}{2}(1+\alpha).$$

By taking α and β near enough to 1 we can make both $\frac{1}{2}(\beta+1)$ and $\frac{1}{2}(1+\alpha)$ be as near as we please to 1; hence

$$\lim_{x\to\infty}\frac{\psi(x)}{x} = 1.$$

□

Proof of Theorem 7.1. This follows from Theorem 7.10 and Proposition 6.2. □

7.6. **Problems.**

Problem 7.1. Let p_n denote the nth prime (thus $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc). Prove that $p_n \sim n \log n$ as $n \to \infty$.

Problem 7.2. Try to determine as explicitly as possible the coefficients of the Laurent expansion of $\zeta(s)$ at $s = 1$.
[Hint. Use the formula (278).]

Problem 7.3. By extending the method of proof of Theorem 7.4, prove that there exists constants $c > 0$ and $A > 0$ such that $\zeta(s)$ has no zero in the region

$$\sigma \geq 1 - \frac{c}{\log(|t| + 2)^A}.$$

Problem 7.4. In the following problem, by mimicking the proof of the prime number theorem given in this section, we prove that the *Mertens function* $M(x) = \sum_{1 \leq n \leq x} \mu(n)$ satisfies the bound $M(x) = o(x)$ as $x \to \infty$. (This can be interpreted as saying that on average $\mu(n) = 1$ holds as often as $\mu(n) = -1$.)
(a). Set $M_1(x) = \int_0^x M(u)\, du$ $(x > 0)$. Prove that $M_1(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)} \frac{1}{\zeta(s)}\, ds$ for any $x > 0$ and any $c > 1$.
(b). Using (a), prove that $M_1(x) = o(x^2)$ as $x \to \infty$.
(c). Using (b), prove that $M(x) = o(x)$ as $x \to \infty$. [Hint. Trying to imitate the proof of Theorem 7.10 we run into the problem that $M(u)$ is not increasing, as opposed to $\psi(u)$. However $M(u)$ has the property that $|M(u_1) - M(u_2)| \leq 1 + |u_1 - u_2|$ for any $u_1, u_2 > 0$, and this can be used as a substitute for monotonicity.]

Problem 7.5. Let $\lambda(n)$ be as in problem 2.3 and set $S(x) = \sum_{1 \leq n \leq x} \lambda(n)$. Prove that $S(x) = o(x)$ as $s \to \infty$. (This can be interpreted as saying that the asymptotic probability for a "random" large integer to have an odd number of primes in its prime factorization is 50%.)
[Hint. This is very similar to problem 7.4.]

Problem 7.6. Generalize the proof of Theorem 7.10 to prove the following: Let $a_1, a_2, \ldots$ be a given sequence of non-negative numbers, and let

$$A(x) = \sum_{n \leq x} a_n; \qquad A_1(x) = \int_0^x A(u)\, du.$$

(a). Prove that if $A_1(x) \sim Cx^a$ as $x \to \infty$, where C and a are positive constants, then $A(x) \sim Cax^{a-1}$ as $x \to \infty$.
(b). As a slightly more general case, prove that if $A_1(x) \sim Cx^a (\log x)^b$ as $x \to \infty$, where C and a are positive constants and $b \in \mathbb{R}$, then $A(x) \sim Cax^{a-1}(\log x)^b$ as $x \to \infty$.

8. The Γ-function; Integral Functions of Order 1

8.1. **Entire functions of finite order.** (Davenport chapter 11.)

Definition 8.1. An entire function $f(z)$ is said to be of *finite order* if there is some $\alpha > 0$ such that

$$f(z) = O\big(e^{|z|^\alpha}\big) \qquad \text{as } |z| \to \infty. \tag{299}$$

The infimum of all numbers $\alpha > 0$ with the property (299) is called the *order* of $f(z)$.

Lemma 8.1. *If $f(z)$ is an entire function of finite order and $f(z)$ has no zeros, then there is a polynomial $g(z)$ such that $f(z) = e^{g(z)}$.*

In fact it is not more difficult to prove the following stronger version:

Lemma 8.2. *If $f(z)$ is an entire function with no zeros, and there are some $\alpha, B > 0$ and $0 < R_1 < R_2 < R_3 < \dots$ with $\lim_{m\to\infty} R_m = \infty$ such that*

$$|f(z)| \le Be^{|z|^\alpha} \qquad \text{whenever } |z| \in \{R_1, R_2, \dots\}, \tag{300}$$

then there is a polynomial $g(z)$ such that $f(z) = e^{g(z)}$. In this case $f(z)$ is of finite order, and this order is equal to the degree of the polynomial $g(z)$.

Proof. Since $\mathbb{C}$ is simply connected, there is an entire function $g(z)$ such that $f(z) = e^{g(z)}$ for all $z \in \mathbb{C}$ (cf., e.g., [60, Thm. 13.11(h)]). Now $|f(z)| = e^{\operatorname{Re} g(z)}$, thus for any z with $|z| = R_m$ we have $\operatorname{Re} g(z) = \log|f(z)| \le \log\big(Be^{R_m^\alpha}\big) = \log B + R_m^\alpha$. If we put

$$g(z) = \sum_{n=0}^{\infty} (a_n + ib_n) z^n \qquad (a_n, b_n \in \mathbb{R})$$

then for any $R \ge 0$, $\theta \in \mathbb{R}$ we have

$$\operatorname{Re} g(Re^{i\theta}) = \sum_{n=0}^{\infty} a_n R^n \cos n\theta - \sum_{n=1}^{\infty} b_n R^n \sin n\theta. \tag{301}$$

Hence also

$$\int_0^{2\pi} \big(\operatorname{Re} g(Re^{i\theta})\big) \cos k\theta \, d\theta = (1 + \delta_{k0})\pi a_k R^k; \tag{302}$$

$$\int_0^{2\pi} \big(\operatorname{Re} g(Re^{i\theta})\big) \sin k\theta \, d\theta = -(1 - \delta_{k0})\pi b_k R^k \tag{303}$$

for all $k \ge 0$, where $\delta_{kn} := 1$ if $k = n$, otherwise $\delta_{kn} := 0$. (Proof: This is just basic Fourier analysis on $\mathbb{R}/2\pi\mathbb{Z}$. In fact for any fixed $R \ge 0$ the infinite sums in (301) are absolutely convergent, uniformly over θ. Hence after substituting (301) into the left hand side of (302) or (303) we may interchange the order of summation and integration, and the formulas now

follow upon using $\int_0^{2\pi} \cos n\theta \cos k\theta \, d\theta = \pi\delta_{nk}(1+\delta_{k0})$; $\int_0^{2\pi} \sin n\theta \sin k\theta \, d\theta = \pi\delta_{nk}(1-\delta_{k0})$; $\int_0^{2\pi} \cos n\theta \sin k\theta \, d\theta = 0$, true for all $n, k \geq 0$.)

Using (302) (twice) we get, for any $k \geq 1$:

$$\pi|a_k|R^k \leq \int_0^{2\pi} \big|\mathrm{Re}\, g(Re^{i\theta})\big| \, d\theta = -2\pi a_0 + \int_0^{2\pi} \Big(\big|\mathrm{Re}\, g(Re^{i\theta})\big| + \mathrm{Re}\, g(Re^{i\theta})\Big) \, d\theta,$$

and here we note that for any real number a we have $|a| + a = 2a$ if $a \geq 0$, but $|a| + a = 0$ if $a \leq 0$. Hence if $R = R_m$ for some m then the integrand in the last integral is everywhere $\leq 2\max(0, \log B + R_m^\alpha)$, and we thus conclude

$$\pi|a_k|R_m^k \leq -2\pi a_0 + 4\pi \max(0, \log B + R_m^\alpha).$$

Letting here $m \to \infty$ (so that $R_m \to \infty$) this implies that $a_k = 0$ whenever $k > \alpha$. In an entirely similar way (using (303)) we get that $b_k = 0$ whenever $k > \alpha$. Hence $g(z)$ is a polynomial, of degree $\leq \alpha$.

Finally, knowing that $g(z)$ is a polynomial it is obvious that $f(z) = e^{g(z)}$ is of finite order $\leq \deg g(z)$; and the conclusion above ("deg $g(z) \leq \alpha$") shows that the order of $f(z)$ is also $\geq \deg g(z)$. Hence $f(z)$ has order exactly $= \deg g(z)$. □

We next prove *Jensen's formula:*

Proposition 8.3. *Suppose that $f(z)$ is an analytic function in the disc $|z| \leq R$ which has no zero on $|z| = R$ and whose zeros in $|z| < R$ are exactly $z_1, \ldots, z_n$ (multiple zeros being repeated as appropriate). We also suppose $f(0) \neq 0$. Then*

$$\frac{1}{2\pi} \int_0^{2\pi} \log\big|f(Re^{i\theta})\big| \, d\theta - \log|f(0)| = \log \frac{R^n}{|z_1| \cdots |z_n|} = \int_0^R r^{-1} n(r) \, dr, \tag{304}$$

where $n(r)$ denotes the number of zeros in $|z| < r$.

Proof. Set

$$g(z) = f(z) \prod_{k=1}^n \frac{R^2 - \overline{z_k} z}{R(z_k - z)}.$$

Note here that each factor $\frac{R^2 - \overline{z_k} z}{R(z_k - z)}$ is a meromorphic function in the disc $|z| \leq R$ with a simple pole at $z = z_k$ and no other poles, and no zeros (since $\big|R^2 - \overline{z_k} z\big| \geq R^2 - |z_k|R > 0$ for all z with $|z| \leq R$). Hence $g(z)$ is analytic and has no zeros in the disc $|z| \leq R$. It follows that there is an analytic function $h(z)$ such that $g(z) = e^{h(z)}$ throughout the disc $|z| \leq R$.

Next note that for each z with $|z| = R$, and each $k \in \{1, \ldots, n\}$, we have

$$\Big|\frac{R^2 - \overline{z_k} z}{R(z_k - z)}\Big| = \Big|\frac{z(\overline{z} - \overline{z_k})}{R(z_k - z)}\Big| = 1.$$

Hence $|g(z)| = |f(z)|$ for all z with $|z| = R$, and it follows that the left hand side of (304) equals

$$\frac{1}{2\pi}\int_0^{2\pi} \log|g(Re^{i\theta})|\,d\theta - \log|f(0)| = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{Re} h(Re^{i\theta})\,d\theta - \log|f(0)|.$$

By the mean value property for the harmonic function $\operatorname{Re} h(z)$, this is[18]

$$= \operatorname{Re} h(0) - \log|f(0)|.$$

We note that $|g(0)| = |f(0)|\prod_{k=1}^n \frac{R}{|z_k|}$, and thus we get

$$= \Big(\log|f(0)| + \sum_{k=1}^n \log\frac{R}{|z_k|}\Big) - \log|f(0)| = \sum_{k=1}^n \log\frac{R}{|z_k|} = \log\frac{R^n}{|z_1|\cdots|z_n|}.$$

Hence the first equality in (304) is proved.

Finally note that the above equals, by integration by parts (noticing that $n(r) = 0$ for all sufficiently small r),

$$\sum_{k=1}^n \log\frac{R}{|z_k|} = \int_{0+}^R \Big(\log\frac{R}{r}\Big)\,dn(r) = \Big[\Big(\log\frac{R}{r}\Big)n(r)\Big]_{r=0+}^{r=R} + \int_0^R r^{-1}n(r)\,dr = \int_0^R r^{-1}n(r)\,dr.$$

□

Remark 8.1. The assumption that $f(z)$ should have no zero on $|z| = R$ can be removed. Cf. [60, 15.17-18].

Corollary 8.4. *If $f(z)$ is an entire function with $f(0) \neq 0$ and of finite order $\rho \geq 0$, and if $n(r)$ denotes the number of zeros in $|z| < r$, then for every $\alpha > \rho$ we have $n(R) = O(R^\alpha)$ as $R \to \infty$.*

Proof. Fix any $\alpha > \rho$; then for all sufficiently large R we have $\log|f(Re^{i\theta})| \leq R^\alpha$ for all θ, and also $\big|\log|f(0)|\big| \leq R^\alpha$. Hence Jensen's formula (Proposition 8.3) implies that, for any $R > 0$ such that $f(z) \neq 0$ for all z on the circle $|z| = R$,

$$\int_0^R r^{-1}n(r)\,dr \leq R^\alpha - \log|f(0)| \leq 2R^\alpha.$$

The same inequality must now also hold for the remaining (countably many) R-values, since both sides depend continuously on R. On the other hand we have

$$\int_R^{2R} r^{-1}n(r)\,dr \geq n(R)\int_R^{2R} r^{-1}\,dr = (\log 2)n(R).$$

This gives the desired bound. □

[18]The identity is also easy to get without mentioning harmonicity. Indeed, since $h(z)$ is analytic for $|z| \leq R$ we have, by Cauchy's Theorem, $h(0) = \frac{1}{2\pi i}\int_C \frac{h(z)}{z}\,dz$, where C is the circle $|z| = R$ with positive orientation. Parametrizing this circle as $z = Re^{i\theta}$ we conclude $h(0) = \frac{1}{2\pi}\int_0^{2\pi} h(Re^{i\theta})\,d\theta$, and thus $\operatorname{Re} h(0) = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{Re} h(Re^{i\theta})\,d\theta$, as desired.

We next remind of a fact which we gave as Problem 2.6 (see the solution for the proof):

Proposition 8.5. *Let $z_1, z_2, \dots$ be a sequence of non-zero complex numbers, let k be an integer ≥ 0, and assume that $\sum_{n=1}^{\infty} |z_n|^{-1-k} < \infty$. Then the following product is absolutely convergent for all $z \in \mathbb{C}$, uniformly on compact subsets:*

$$f(z) = \prod_{n=1}^{\infty} \Big\{ \Big(1 - \frac{z}{z_n}\Big) \exp\Big(\frac{z}{z_n} + \frac{1}{2}\big(\frac{z}{z_n}\big)^2 + \dots + \frac{1}{k}\big(\frac{z}{z_n}\big)^k\Big)\Big\}.$$

In particular $f(z)$ is an entire function. This function has a zero at each point $z = z_j$ and no other zeros in the plane. More precisely, if α occurs m times in the sequence $\{z_1, z_2, \dots\}$ then f has a zero of order (exactly) m at α.

Let us also note:

Proposition 8.6. *In the above situation, if α is any real number $\geq k$ such that $\sum_{n=1}^{\infty} |z_n|^{-\alpha}$ converges, then there is a constant $B > 0$ such that $f(z) = O\big(e^{B|z|^\alpha}\big)$ for all $z \in \mathbb{C}$.*

Proof. Note that if $f(z) = O\big(e^{B|z|^\alpha}\big)$ $(\forall z \in \mathbb{C})$ holds for one α then it also holds for any larger value of α (by possibly increasing the implied constant to accommodate the z's with $|z| \leq 1$); also recall $\sum_{n=1}^{\infty} |z_n|^{-k-1}$. Hence without loss of generality we may assume $\alpha \leq k+1$; thus $\alpha \in [k, k+1]$.

Let us write $E(w) = (1-w)\exp\big(w + \frac{1}{2}w^2 + \dots + \frac{1}{k}w^k\big)$. From the solution of Problem 2.6 we know that $\big|E(w) - 1\big| \ll |w|^{k+1} \ll |w|^\alpha$ as $|w| \to 0$; hence there is a constant $A > 0$ such that $\big|E(w)\big| \leq 1 + \big|E(w) - 1\big| \leq 1 + A|w|^\alpha \leq e^{A|w|^\alpha}$ for all w with $|w| \leq 1$. On the other hand there is a constant $A' > 0$ (which only depends on α) such that for all $|w| > 1$:

$$\big|E(w)\big| \leq (1+|w|)\exp(|w| + \frac{1}{2}|w|^2 + \dots + \frac{1}{k}|w|^k) \leq e^{A'|w|^\alpha} e^{k|w|^k} \leq e^{(A'+k)|w|^\alpha}.$$

Hence by possibly increasing A further we may assume that $\big|E(w)\big| \leq e^{A|w|^\alpha}$ holds for *all* $w \in \mathbb{C}$. Then

$$\big|f(z)\big| = \prod_{n=1}^{\infty} \Big|E\Big(\frac{z}{z_n}\Big)\Big| \leq \exp\Big(A\Big(\sum_{n=1}^{\infty} |z_n|^{-\alpha}\Big)|z|^\alpha\Big), \qquad \forall z \in \mathbb{C}.$$

□

We now prove (a particular case of) the *Weierstrass factorization Theorem:*

Theorem 8.7. *Let $f(z)$ be an entire function with $f(0) \neq 0$ and of finite order $\rho \geq 0$, and let $z_1, z_2, \dots$ be all the zeros of f, counted with multiplicity. Let k be an integer satisfying $0 \leq k \leq \rho$ and $\sum_n |z_n|^{-1-k} < \infty$; there exists at least one such k. Then there is a polynomial $g(z)$ of degree $\leq \rho$ such that*

$$f(z) = e^{g(z)} \prod_n \Big\{ \Big(1 - \frac{z}{z_n}\Big) \exp\Big(\frac{z}{z_n} + \frac{1}{2}\big(\frac{z}{z_n}\big)^2 + \dots + \frac{1}{k}\big(\frac{z}{z_n}\big)^k\Big)\Big\}. \tag{305}$$

(Note that we do not exclude the possibility that the list $z_1, z_2, \dots$ is finite, or even empty. If the list of zeros is empty then the product $\prod_n \{\cdots\}$ is of course interpreted as 1.)

Proof. Let us write $r_n = |z_n|$ for short. For any $\alpha > \rho$ we have $n(R) = O(R^\alpha)$ as $R \to \infty$, by Corollary 8.4, and thus $\sum_n r_n^{-\beta}$ converges for any $\beta > \alpha$ (cf. Problem 3.4(a)). Hence in fact $\sum_n r_n^{-\beta}$ converges for any $\beta > \rho$, and in particular there exists some integer k with $0 \le k \le \rho$ such that $\sum_n r_n^{-1-k}$ converges. We fix such an integer k for the rest of the argument.

Now by Proposition 8.5 (or trivially, if the list $z_1, z_2, \dots$ is finite) the following product is absolutely convergent, uniformly on compacta, and thus defines an entire function:

$$P(z) = \prod_n E\Big(\frac{z}{z_n}\Big); \qquad E(w) = (1-w)\exp\big(w + \tfrac{1}{2}w^2 + \ldots + \tfrac{1}{k}w^k\big).$$

Since $P(z)$ has exactly the same zeros as $f(z)$, with multiplicities (cf. Proposition 8.5), the meromorphic function

$$F(z) = f(z)/P(z)$$

is in fact an entire function without zeros! We will next prove that $F(z)$ satisfies a bound as in Lemma 8.2.

Fix α as any real number $> \rho$. Then $\sum_n r_n^{-\alpha}$ converges (as noticed above); hence the total length of all the intervals $[r_n - r_n^{-\alpha}, r_n + r_n^{-\alpha}]$ on the real line is finite, and consequently there exist arbitrarily large values of R with the property that

$$|R - r_n| > r_n^{-\alpha}, \qquad \forall n. \tag{306}$$

Let us (temporarily) fix any $R > 2$ with this property, take some z with $|z| = R$, and write

$$P(z) = \prod_{\substack{n \\ (r_n < \frac{1}{2}R)}} E\Big(\frac{z}{z_n}\Big) \prod_{\substack{n \\ (\frac{1}{2}R \le r_n \le 2R)}} E\Big(\frac{z}{z_n}\Big) \prod_{\substack{n \\ (2R < r_n)}} E\Big(\frac{z}{z_n}\Big) = P_1(z)P_2(z)P_3(z),$$

say. Now for any z_n with $|z_n| = r_n < \frac{1}{2}R$ we have

$$\begin{aligned}\Big|E\Big(\frac{z}{z_n}\Big)\Big| &= \Big|1 - \frac{z}{z_n}\Big| \cdot \Big|\exp\Big(\frac{z}{z_n} + \frac{1}{2}\big(\frac{z}{z_n}\big)^2 + \ldots + \frac{1}{k}\big(\frac{z}{z_n}\big)^k\Big)\Big| \\ &\ge \Big(\Big|\frac{z}{z_n}\Big| - 1\Big) \cdot \exp\Big(-\Big|\frac{z}{z_n}\Big| - \frac{1}{2}\Big|\frac{z}{z_n}\Big|^2 - \ldots - \frac{1}{k}\Big|\frac{z}{z_n}\Big|^k\Big) \ge \exp\big(-C_1 R^k r_n^{-k}\big).\end{aligned}$$

Here and in the following we will denote by $C_1, C_2, \dots$ certain real positive constants which do not dependent on R. Since $\alpha > \rho \ge k$ we have $\sum_{r_n < \frac{1}{2}R} r_n^{-k} \le (\frac{1}{2}R)^{\alpha-k} \sum_n r_n^{-\alpha} = C_2 R^{\alpha-k}$. Hence

$$\big|P_1(z)\big| \ge \prod_{r_n < \frac{1}{2}R} \exp\big(-C_1 R^k r_n^{-k}\big) \ge \exp\big(-C_3 R^\alpha\big).$$

Next for any z_n occurring in P_2 we have

$$\Big|E\Big(\frac{z}{z_n}\Big)\Big| \ge \frac{|z_n - z|}{2R}\exp\big(-C_4\big) > C_5 R^{-1-\alpha},$$

by (306). By possibly lowering C_5 we may require that $0 < C_5 < 1$ in the above bound; then we also have $0 < C_5 R^{-1-\alpha} < 1$. Note that the total number of factors in P_2 is less than $n(2R) \ll R^\alpha$ (cf. Corollary 8.4). We now get[19], for any fixed $\varepsilon > 0$,

$$\big|P_2(z)\big| \ge \big(C_5 R^{-1-\alpha}\big)^{C_6 R^\alpha} = \exp\Big(C_6 R^\alpha(\log C_5 - (1+\alpha)\log R)\Big) \ge \exp\Big(-C_7 R^{\alpha+\varepsilon}\Big)$$

(where we used $R > 2$, and C_7 of course depends on ε). Finally to treat P_3 we recall from the solution of Problem 2.6 that $\big|E(w) - 1\big| \ll |w|^{k+1}$ as $w \to 0$, and thus there exist constants $0 < \delta < \frac12$ and $C_8 > 0$ such that $\big|E(w)\big| \ge \exp\big(-C_8|w|^{k+1}\big)$ for all w with $|w| \le \delta$. Note also that in the region $\delta \le |w| \le \frac12$ the function $\big|E(w)\big| = \big|1-w\big| \cdot \big|e^{w+\frac12 w^2+\ldots+\frac1k w^k}\big|$ is bounded from below by a positive constant; hence by possibly increasing C_8 we have

$$\big|E(w)\big| \ge \exp\big(-C_8|w|^{k+1}\big) \qquad \text{for all } w \text{ with } |w| \le \frac12.$$

Also,

$$\text{if } \alpha \le k+1: \quad \sum_{r_n > 2R} (R/r_n)^{k+1} \le \sum_{r_n > 2R} (R/r_n)^\alpha = \Big(\sum_n r_n^{-\alpha}\Big) R^\alpha;$$
$$\text{if } \alpha > k+1: \quad \sum_{r_n > 2R} (R/r_n)^{k+1} \le \Big(\sum_n r_n^{-k-1}\Big) R^\alpha.$$

Hence

$$\big|P_3(z)\big| \ge \exp\Big(-C_8 \sum_{r_n > 2R} (R/r_n)^{k+1}\Big) \ge \exp\big(-C_9 R^\alpha\big).$$

Multiplying our bounds on $\big|P_1(z)\big|$, $\big|P_2(z)\big|$ and $\big|P_3(z)\big|$ together we obtain

$$\big|P(z)\big| \ge \exp\Big(-C_{10} R^{\alpha+\varepsilon}\Big).$$

Furthermore $\big|f(z)\big| \le \exp\big(C_{11} R^\alpha\big)$ since $\alpha > \rho$ and f is of order ρ (and $R \ge 2$); hence

$$\big|F(z)\big| = \big|f(z)/P(z)\big| \le \exp\Big(C_{12} R^{\alpha+\varepsilon}\Big) \qquad \text{for all } z \text{ with } |z| = R.$$

This holds for arbitrarily large values of R, and hence by Lemma 8.2 we must have $F(z) = e^{g(z)}$ where $g(z)$ is a polynomial of degree $\le \alpha + \varepsilon$. Finally $\alpha + \varepsilon$ can be chosen arbitrarily near ρ; thus in fact $g(z)$ must have degree $\le \rho$. □

[19] We here use the following fact: If $0 < b \le 1$, and $a_1, \ldots, a_m$ are some real numbers $\ge b$, and $m \le M$, *then* $\prod_{j=1}^m a_j \ge b^M$. Proof: $\prod_{j=1}^m a_j \ge \prod_{j=1}^m b = b^m \ge b^M$. Note that the assumption $b \le 1$ is needed in the last step!

Remark 8.2. Let us make some more observations in the situation of Theorem 8.7. Let τ be the infimum of all $\alpha > 0$ such that $\sum_n |z_n|^{-\alpha}$ converges. *Then* $\rho \geq \max(\deg g, \tau)$, *and if* k *is the* ***minimal*** *non-negative integer with* $\sum_n |z_n|^{-1-k} < \infty$ *then* $\rho = \max(\deg g, \tau)$. *Furthermore, if* $\sum_n |z_n|^{-\rho}$ *converges, then there is some constant* $B > 0$ *such that* $f(z) = O\big(e^{B|z|^\rho}\big)$ *for all* $z \in \mathbb{C}$.

Proof. We saw in the proof of Theorem 8.7 that $\deg g \leq \rho$ and that $\sum_n |z_n|^{-\alpha}$ converges for every $\alpha > \rho$, so that $\tau \leq \rho$. Thus $\max(\deg g, \tau) \leq \rho$.

Next assume that k is the *minimal* non-negative integer with $\sum_n |z_n|^{-1-k} < \infty$. Then $k \leq \tau$, and hence for any $\alpha > \tau$ the product $\prod_n E(z/z_n)$ has finite order $\leq \alpha$ by Proposition 8.6, and the same is trivially true for $e^{g(z)}$ if $\alpha \geq \deg g$. Hence by (305), for every $\alpha > \max(\deg g, \tau)$ the function $f(z)$ has order $\leq \alpha$. Thus $\rho \leq \max(\deg g, \tau)$, and this proves our claim that $\rho = \max(\deg g, \tau)$.

Finally assume that $\sum_n |z_n|^{-\rho}$ converges. Then by Proposition 8.6 (and since we always keep $k \leq \rho$) there is some $B_1 > 0$ such that $\prod_n E(z/z_n) = O\big(e^{B_1|z|^\rho}\big)$ for all $z \in \mathbb{C}$. The same type of bound obviously holds for $e^{g(z)}$ (since $\deg g \leq \rho$), and hence also for $f(z)$. □

We also note that we may compute the logarithmic derivative of $f(z)$ in a termwise way:

Proposition 8.8. *In the situation of Theorem 8.7 we have*

$$\frac{f'(z)}{f(z)} = g'(z) + \sum_n \Big(\frac{1}{z - z_n} + \frac{1}{z_n} + \frac{z}{{z_n}^2} + \frac{z^2}{{z_n}^3} + \ldots + \frac{z^{k-1}}{{z_n}^k} \Big), \tag{307}$$

where the sum is absolutely convergent for every $z \in \mathbb{C} \setminus \{z_1, z_2, \ldots\}$.

Proof. Let z_0 be any fixed complex number not in $\{z_1, z_2, \ldots\}$. Let $r = \inf_n |z_n - z_0|$ and let D be the open disc $D = \{z : |z - z_0| < r\}$, so that $z_n \notin D$ for all $n = 1, 2, \ldots$. Set

$$h(z) = g(z) + \sum_n \Big\{ \log\Big(1 - \frac{z}{z_n}\Big) + \frac{z}{z_n} + \frac{1}{2}\big(\frac{z}{z_n}\big)^2 + \ldots + \frac{1}{k}\big(\frac{z}{z_n}\big)^k \Big\}, \qquad z \in D, \tag{308}$$

where for each n we have fixed some branch of the logarithm function $D \ni z \mapsto \log\big(1 - \frac{z}{z_n}\big)$. Take $N \in \mathbb{Z}^+$ such that $|z_n| > 2(|z_0| + r)$ for all $n \geq N$. Then $\big|\frac{z}{z_n}\big| < \frac{1}{2}$ for all $z \in D$, $n \geq N$, and we make the requirement that for each $n \geq N$ we use the principal branch of the logarithm function $D \ni z \mapsto \log\big(1 - \frac{z}{z_n}\big)$ in the above sum.

For all $n \geq N$ and $z \in D$ we have, using the Taylor expansion of the logarithm,

$$\begin{aligned} \Big| \mathrm{Im} \Big\{ \log\Big(1 - \frac{z}{z_n}\Big) + \frac{z}{z_n} + \frac{1}{2}\big(\frac{z}{z_n}\big)^2 + \ldots + \frac{1}{k}\big(\frac{z}{z_n}\big)^k \Big\} \Big| &= \Big| -\mathrm{Im} \sum_{j=k+1}^{\infty} \frac{1}{j}\big(\frac{z}{z_n}\big)^j \Big| \\ &< \sum_{j=k+1}^{\infty} \frac{1}{j}\big(\frac{1}{2}\big)^j < 1 < \pi, \end{aligned}$$

and hence $\log\Big(1-\frac{z}{z_n}\Big)+\frac{z}{z_n}+\frac{1}{2}\big(\frac{z}{z_n}\big)^2+\ldots+\frac{1}{k}\big(\frac{z}{z_n}\big)^k$ equals the *principal part* logarithm $\log\Big\{\Big(1-\frac{z}{z_n}\Big)\exp\Big(\frac{z}{z_n}+\frac{1}{2}\big(\frac{z}{z_n}\big)^2+\ldots+\frac{1}{k}\big(\frac{z}{z_n}\big)^k\Big)\Big\}$. Hence by Corollary 2.4 the sum (308) is uniformly convergent on compact subsets of D, and in particular $h(z)$ is analytic on D. Corollary 2.4 also says that $e^{h(z)}=f(z)$ for all $z\in D$. [To be precise, we apply Corollary 2.4 to the $[n\geq N]$-part of the product in (305), and use the standard logarithm laws for the remaining finite product.] Differentiating the last formula we get $f'(z)=h'(z)e^{h(z)}=h'(z)f(z)$, thus $\frac{f'(z)}{f(z)}=h'(z)$ for all $z\in D$ (since $f(z)\neq 0$ for all $z\in D$). Because of the uniform convergence on compacta, we may differentiate the formula (308) termwise; hence we conclude that (307) holds for all $z\in D$ and in particular for $z=z_0$.

The absolute convergence in (307) is clear from the convergence of $\sum_n |z_n|^{-1-k}$ and the fact that for all n with $|z_n|>2|z|$ (say) we have

$$\Big|\frac{1}{z-z_n}+\frac{1}{z_n}+\frac{z}{{z_n}^2}+\frac{z^2}{{z_n}^3}+\ldots+\frac{z^{k-1}}{{z_n}^k}\Big|=\Big|\frac{z^k}{z_n^k(z-z_n)}\Big|<\frac{2|z|^k}{|z_n|^{k+1}}. \tag{309}$$

□

Finally let us note explicitly what Theorem 8.7, Remark 8.2 and Proposition 8.8 say when $f(z)$ is of order 1. This is the only case which we will need in the rest of these lectures (I think):

Corollary 8.9. *If $f(z)$ is an entire function of order 1 with $f(0)\neq 0$ then*

$$f(z)=e^{A+Bz}\prod_n\Big(1-\frac{z}{z_n}\Big)e^{z/z_n};\qquad\qquad \frac{f'(z)}{f(z)}=B+\sum_n\Big(\frac{1}{z-z_n}+\frac{1}{z_n}\Big), \tag{310}$$

where $A,B\in\mathbb{C}$ and $z_1,z_2,\ldots$ are all the zeros of $f(z)$ (counted with multiplicity). Furthermore the sum $\sum_{n=1}^{\infty}|z_n|^{-1-\varepsilon}$ converges for any $\varepsilon>0$. If also the sum $\sum_{n=1}^{\infty}|z_n|^{-1}$ converges then there is a constant $B>0$ such that

$$|f(z)|<O\big(e^{B|z|}\big)\qquad \text{for all } z\in\mathbb{C}. \tag{311}$$

Example 8.1. Let us apply the Weierstrass factorization theorem to the function $\sin\pi z$. Since $\sin\pi z$ has a (simple) zero at $z=0$ we set

$$f(z)=\frac{\sin\pi z}{z}.$$

This is (after removing the singularity at $z=0$) an entire function with $f(0)\neq 0$. Note that for all z with $|z|\geq 1$ we have

$$\big|f(z)\big|=\Big|\frac{e^{\pi iz}-e^{-\pi iz}}{2iz}\Big|\leq 2e^{\pi|z|},$$

so that $f(z)$ is of finite order ≤ 1. Since $f(yi) = \frac{\sinh \pi y}{y} \sim \frac{e^{\pi y}}{2y}$ as $y \to \infty$, the order of f is *exactly* 1. Hence by Corollary 8.9, since the zeros of $f(z)$ are exactly the non-zero integers n, and all these zeros are simple, we have

$$f(z) = \frac{\sin \pi z}{z} = e^{A+Bz} \prod_{n \in \mathbb{Z} \setminus \{0\}} \Big(1 - \frac{z}{n}\Big) e^{z/n}$$

for some constants $A, B \in \mathbb{C}$, where the product is absolutely convergent. Taking $z = 0$ (or "$z \to 0$") in this formula gives $\pi = e^A$. Also, since both $f(z)$ and $\prod_{n \in \mathbb{Z} \setminus \{0\}} \Big(1 - \frac{z}{n}\Big) e^{z/n}$ are even functions of z, also e^{Bz} must be even; thus $B = 0$. (The same fact can also be seen e.g. by taking $z = 0$ in the formula for $\frac{f'(z)}{f(z)}$, cf. (310).) Hence we have proved:

$$\sin \pi z = \pi z \prod_{n \in \mathbb{Z} \setminus \{0\}} \Big(1 - \frac{z}{n}\Big) e^{z/n} = \pi z \prod_{n=1}^{\infty} \Big(1 - \frac{z^2}{n^2}\Big). \tag{312}$$

We may also note that the formula (310) for $\frac{f'(z)}{f(z)}$ implies

$$\pi \cot \pi z = \frac{1}{z} + \sum_{n \in \mathbb{Z} \setminus \{0\}} \Big(\frac{1}{z-n} + \frac{1}{n}\Big). \tag{313}$$

8.2. **The Γ-function.** In this section we borrow from Ahlfors [1, §6.2.4] and Edwards [18, Ch. 6].

Definition 8.2. The *Gamma function*, $\Gamma(z)$, is defined by

$$\frac{1}{\Gamma(z)} = z e^{\gamma z} \prod_{n=1}^{\infty} \Big(1 + \frac{z}{n}\Big) e^{-z/n}, \tag{314}$$

where γ is *Euler's constant,* defined so that $\Gamma(1) = 1$, i.e.

$$\gamma := -\log\Big(\prod_{n=1}^{\infty} \Big(1 + \frac{1}{n}\Big) e^{-1/n}\Big) = 0.57722\ldots. \tag{315}$$

Since $\sum_{n=1}^{\infty} n^{-1-\varepsilon} < \infty$ for any $\varepsilon > 0$, the product (314) is uniformly absolutely convergent on compact subsets of $\mathbb{C}$ (cf. Proposition 8.5), and $\frac{1}{\Gamma(z)}$ is an entire function which has a simple zero at each point $z = 0, -1, -2, \ldots$, and no other zeros. Hence:

Lemma 8.10. *$\Gamma(z)$ is a meromorphic function on $\mathbb{C}$ which has a simple pole at each point $z = 0, -1, -2, \ldots$, and no other poles, and which is nowhere zero.*

The above definition of $\Gamma(z)$ can be motivated as follows: The function $\sin \pi z$ has all the integers for zeros, and it is the simplest function with this property. We now wish to

introduce functions which have only the positive or only the negative integers for zeros. The simplest function with, for instance, the negative integers for zeros, is

$$G(z) = \prod_{n=1}^{\infty} \Big(1 + \frac{z}{n}\Big) e^{-z/n}.$$

It is evident that $G(-z)$ has then the positive integers for zeros, and by comparison with the product representation (312) of $\sin \pi z$ we find at once

$$zG(z)G(-z) = \frac{\sin \pi z}{\pi}. \tag{316}$$

Because of the manner in which $G(z)$ has been constructed, it is bound to have other simple properties. We observe that $f(z) = z^{-1}G(z-1)$ is an entire function with exactly the same zeros as $G(z)$. By Proposition 8.6, $G(z)$ has order ≤ 1, and hence also $f(z)$ has order ≤ 1. Hence by Theorem 8.7 there are some $A, B \in \mathbb{C}$ such that $f(z) = e^{A+Bz}G(z)$, i.e.

$$G(z-1) = ze^{A+Bz}G(z). \tag{317}$$

In order to determine B we take the logarithmic derivatives on both sides. This gives

$$\sum_{n=1}^{\infty} \Big(\frac{1}{z-1+n} - \frac{1}{n}\Big) = \frac{1}{z} + B + \sum_{n=1}^{\infty} \Big(\frac{1}{z+n} - \frac{1}{n}\Big)$$

(cf. Proposition 8.8). Taking here $z = 1$ we get

$$0 = 1 + B + \sum_{n=1}^{\infty} \Big(\frac{1}{1+n} - \frac{1}{n}\Big) = 1 + B - 1,$$

i.e. $B = 0$. Hence A can be determined by taking $z = 1$ in (317); this gives

$$e^{-A} = G(1) = \prod_{n=1}^{\infty} \Big(1 + \frac{1}{n}\Big) e^{-1/n},$$

i.e. $A = \gamma$ (cf. (315)). Now (317) takes a somewhat simpler form if instead of $G(z)$ we consider the function $H(z) = e^{\gamma z}G(z)$; then (317) says that $H(z-1) = zH(z)$. It has been found useful to make a slight further shift, by setting $\Gamma(z) = \big(zH(z)\big)^{-1}$. Note that this agrees with the definition (314). Also the relations (317) and (316) now say:

Lemma 8.11.

$$\Gamma(z+1) = z\Gamma(z) \tag{318}$$

and

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}. \tag{319}$$

Let us note a few more basic facts.

Lemma 8.12.

$$\Gamma(n) = (n-1)!, \qquad \forall n \in \mathbb{Z}^+.$$

Proof. Note that $\Gamma(1) = 1$ directly from (314), (315). Now the formula follows by applying (318) with $z = 1, 2, 3, \ldots$. □

Lemma 8.13. *Euler's constant γ satisfies the relation*

$$\gamma = \lim_{n\to\infty} \Big(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \log n\Big).$$

Proof. In the definition of Euler's constant γ, (315), the Nth partial product can be written

$$\prod_{n=1}^{N} \Big(1+\frac{1}{n}\Big) e^{-1/n} = \frac{2}{1}\cdot\frac{3}{2}\cdot\ldots\cdot\frac{N+1}{N} e^{-\frac{1}{1}-\frac{1}{2}-\ldots-\frac{1}{N}} = (N+1)e^{-\frac{1}{1}-\frac{1}{2}-\ldots-\frac{1}{N}}.$$

Hence

$$\gamma = \lim_{N\to\infty} -\log\Big((N+1)e^{-\frac{1}{1}-\frac{1}{2}-\ldots-\frac{1}{N}}\Big) = \lim_{N\to\infty}\Big(\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{N}-\log(N+1)\Big),$$

which agrees with the stated formula since $\lim_{N\to\infty}\big(\log(N+1)-\log N\big) = 0$. □

Lemma 8.14.

$$\frac{\Gamma'(z)}{\Gamma(z)} = -\gamma - \frac{1}{z} - \sum_{n=1}^{\infty}\Big(\frac{1}{z+n} - \frac{1}{n}\Big),$$

where the sum is absolutely convergent for every $z \in \mathbb{C} \setminus \{0, -1, -2, \ldots\}$.

Proof. This follows from our definition (314) and Proposition 8.8 (note that if $f(z) = (z\Gamma(z))^{-1}$ then $\frac{f'(z)}{f(z)} = -\frac{1}{z} - \frac{\Gamma'(z)}{\Gamma(z)}$). □

Definition 8.3. For any $z \in \mathbb{C} \setminus (-\infty, 0]$ we define

$$\log\Gamma(z) := -\log z - \gamma z + \sum_{n=1}^{\infty}\Big(\frac{z}{n} - \log\Big(1+\frac{z}{n}\Big)\Big), \tag{320}$$

where the principal branch of the logarithm function is used throughout in the right hand side. Note that the right hand side is obtained by applying a (negative) logarithm factorwise in (314), and by an argument as in the proof of Proposition 8.8 one sees that the sum is absolutely convergent, uniformly in compact subsets of $\mathbb{C}\setminus(-\infty, 0]$, and that it does indeed define a logarithm of $\Gamma(z)$.

Using Lemma 8.13 the definition $\log\Gamma(z)$ can also be written in a slightly different way:

Lemma 8.15.

$$\log\Gamma(z) = \lim_{N\to\infty}\Big(z\log N - \sum_{n=0}^{N}\log(z+n) + \sum_{n=1}^{N}\log n\Big), \qquad \forall z \in \mathbb{C}\setminus(-\infty, 0].$$

Proof. The relation (320) can be written as

$$\log\Gamma(z) = -\log z - \gamma z + \lim_{N\to\infty}\sum_{n=1}^{N}\Big(\frac{z}{n} - \log(z+n) + \log n\Big).$$

By Lemma 8.13 this is

$$= -\log z + \lim_{N\to\infty}\Big(z\log N - \sum_{n=1}^{N}\log(z+n) + \sum_{n=1}^{N}\log n\Big),$$

Corollary 8.16.

$$\Gamma(z) = \lim_{N\to\infty} \frac{N^z N!}{z(z+1)\cdots(z+N)}, \qquad \forall z \in \mathbb{C} \setminus \{0,-1,-2,\ldots\}.$$

We come next to *Stirling's formula*, an asymptotic formula for $\Gamma(z)$.

Theorem 8.17. *For any fixed $\varepsilon > 0$ we have*

$$\log\Gamma(z) = \big(z-\tfrac12\big)\log z - z + \log\sqrt{2\pi} + O\big(|z|^{-1}\big), \tag{321}$$

for all z with $|z| \ge 1$ and $\big|\arg z\big| \le \pi - \varepsilon$. (The implied constant depends on ε but of course not on z.)

Proof. Given any z with $|z| \ge 1$ and $\big|\arg z\big| \le \pi - \varepsilon$ and any $N \in \mathbb{Z}^+$, we have

$$\sum_{n=0}^{N} \log(z+n) = \int_{0-}^{N} \log(z+r)\, d\lfloor r\rfloor = \Big[\lfloor r\rfloor \log(z+r)\Big]_{r=0-}^{r=N} - \int_0^N \frac{\lfloor r\rfloor}{z+r}\, dr$$
$$= N\log(z+N) + \log z - \int_0^N \frac{\lfloor r\rfloor}{z+r}\, dr.$$

Note that $\lfloor r\rfloor$ "on average equals $r-\frac12$". This motivates the following: Set

$$B_1(r) := r - \tfrac12; \qquad \overline{B}_1(r) := B_1(r) - \lfloor r\rfloor = B_1(r - \lfloor r\rfloor).$$

Then the above is (using principal branch logarithms throughout)

$$= N\log(z+N) + \log z - \int_0^N \frac{r-\frac12}{z+r}\, dr + \int_0^N \frac{\overline{B}_1(r)}{z+r}\, dr$$
$$= N\log(z+N) + \log z - \Big(N - (z+\tfrac12)\big(\log(z+N) - \log z\big)\Big) + \int_0^N \frac{\overline{B}_1(r)}{z+r}\, dr$$
$$= \big(z+N+\tfrac12\big)\log(z+N) + \big(\tfrac12 - z\big)\log z - N + \int_0^N \frac{\overline{B}_1(r)}{z+r}\, dr. \tag{322}$$

Here by construction the function $\overline{B}_1(r)$ has average 0; more precisely we have $\int_n^{n+1} \overline{B}_1(r)\, dr = 0$ for all integers n. Hence the last integral above can be expected to be fairly small, and this should become visible by integrating by parts. It is safest to first split the integration into intervals on which $\overline{B}_1(r)$ is continuous:

$$\int_0^N \frac{\overline{B}_1(r)}{z+r}\, dr = \sum_{n=0}^{N-1} \int_n^{n+1} \frac{\overline{B}_1(r)}{z+r}\, dr = \sum_{n=0}^{N-1} \int_0^1 \frac{r-\frac12}{z+r+n}\, dr$$
$$= \sum_{n=0}^{N-1} \Big(0 - 0 + \int_0^1 \frac{\frac12 r^2 - \frac12 r}{(z+r+n)^2}\, dr\Big) = \int_0^N \frac{\frac12(r-\lfloor r\rfloor)^2 - \frac12(r-\lfloor r\rfloor)}{(z+r)^2}\, dr. \tag{323}$$

Here the numerator in the last integrand is a bounded function of r and hence the integral is $O\big(\int_0^\infty |z+r|^{-2}\,dr\big)$. If $z>0$ then this is $= O\big(\int_0^\infty (z+r)^{-2}\,dr\big) = O\big(z^{-1}\big)$, but to treat our case of general complex z with $\big|\arg z\big| \le \pi-\varepsilon$ we need to be slightly more careful: Note that for all $r\ge 0$ we have $|z+r| \gg |z|$. (Proof: If $|\arg z| \le \frac{\pi}{4}$ then $\operatorname{Re} z \gg |z|$ and thus $|z+r| \ge \operatorname{Re}(z+r) \gg |z|+r \ge |z|$. On the other hand if $\frac{\pi}{4} \le |\arg z| \le \pi-\varepsilon$ then $|\operatorname{Im} z| \gg |z|$ and thus $|z+r| \ge |\operatorname{Im} z| \gg |z|$.) Also note that for $r \ge 2|z|$ we have $|z+r| \ge (r-|z|) \gg r$. Hence we get:

$$\int_0^\infty |z+r|^{-2}\,dr \ll \int_0^{2|z|} |z|^{-2}\,dr + \int_{2|z|}^\infty r^{-2}\,dr \ll |z|^{-1}.$$

Collecting our computation so far we have proved that

$$\sum_{n=0}^{N} \log(z+n) = \big(z+N+\tfrac12\big)\log(z+N) + \big(\tfrac12 - z\big)\log z - N + O\big(|z|^{-1}\big). \tag{324}$$

(The implied constant depends only on ε.) We also need to compute $\sum_{n=1}^N \log n$, and this we can do by taking $z=1$ and replacing N by $N-1$ in the above computation (let's assume $N\ge 2$). However we cannot follow the above computation all the way to (324), since this would give an error term $O(1)$ which is too imprecise for us. Instead we use (322) and (323) to get

$$\sum_{n=1}^{N} \log n = \big(N+\tfrac12\big)\log N - (N-1) + \int_0^{N-1} \frac{\frac12(r-\lfloor r\rfloor)^2 - \frac12(r-\lfloor r\rfloor)}{(1+r)^2}\,dr,$$

and here the last integral equals (writing $f(r) = \frac12(r-\lfloor r\rfloor)^2 - \frac12(r-\lfloor r\rfloor)$, and using the fact that this is a bounded function)

$$\int_1^N \frac{f(r)}{r^2}\,dr = \int_1^\infty \frac{f(r)}{r^2}\,dr - \int_N^\infty \frac{f(r)}{r^2}\,dr = \int_1^\infty \frac{f(r)}{r^2}\,dr + O(N^{-1}).$$

Combining our results so far we obtain (for any z with $|z|\ge 1$ and $\big|\arg z\big| \le \pi-\varepsilon$ and any $N\ge 2$):

$$\begin{aligned} z\log N - \sum_{n=0}^{N}\log(z+n) + \sum_{n=1}^{N}\log n \\ &= z\log N - \big(z+N+\tfrac12\big)\log(z+N) + \big(z-\tfrac12\big)\log z + N + O\big(|z|^{-1}\big) \\ &\qquad + \big(N+\tfrac12\big)\log N - (N-1) + A + O\big(N^{-1}\big) \\ &= \big(z-\tfrac12\big)\log z - \big(z+N+\tfrac12\big)\log\frac{z+N}{N} + A + 1 + O\big(|z|^{-1}\big) + O\big(N^{-1}\big), \end{aligned}$$

where $A = \int_1^\infty \frac{f(r)}{r^2}\,dr$. Note that for every $z\in\mathbb{C}$ we have

$$\lim_{N\to\infty} (z+N+\tfrac12)\log\frac{z+N}{N} = \lim_{N\to\infty} (z+N+\tfrac12)\Big(\frac{z}{N} + O_z(N^{-2})\Big) = z.$$

Hence

$$\log\Gamma(z) = \lim_{N\to\infty}\Big(z\log N - \sum_{n=0}^{N}\log(z+n) + \sum_{n=1}^{N}\log n\Big)$$
$$= \big(z-\tfrac12\big)\log z - z + A + 1 + O\big(|z|^{-1}\big). \tag{325}$$

To determine the constant A we may for example set $z = \frac12 + iy$ and let $y\to\infty$. Then by (319) we have

$$\big|\Gamma(\tfrac12+iy)\big|^2 = \Gamma(\tfrac12+iy)\Gamma(\tfrac12-iy) = \frac{\pi}{\sin(\pi(\frac12+iy))} = \frac{\pi}{\cosh(\pi y)},$$

which implies (as $y\to\infty$)

$$\log\big|\Gamma(\tfrac12+iy)\big| = \tfrac12\log\big(\frac{2\pi}{e^{\pi y}+e^{-\pi y}}\big) = -\tfrac{\pi}{2}y + \tfrac12\log\big(\frac{2\pi}{1+e^{-2\pi y}}\big) = -\tfrac{\pi}{2}y + \tfrac12\log 2\pi + O\big(e^{-2\pi y}\big);$$

and on the other hand by (325) we have

$$\log\big|\Gamma(\tfrac12+iy)\big| = \operatorname{Re}\,\log\Gamma(\tfrac12+iy) = \operatorname{Re}\,\big(iy\log(\tfrac12+iy)\big) - \tfrac12 + A + 1 + O\big(y^{-1}\big)$$
$$= -y\arg(\tfrac12+iy) + A + \tfrac12 + O\big(y^{-1}\big) = -y\big(\tfrac{\pi}{2} - \arctan\tfrac{1}{2y}\big) + A + \tfrac12 + O\big(y^{-1}\big)$$
$$= -\tfrac{\pi}{2}y + \tfrac12 + A + \tfrac12 + O\big(y^{-1}\big),$$

where we used the fact that A is real by definition. Together these two asymptotic formulas imply that $A+1 = \frac12\log 2\pi = \log\sqrt{2\pi}$. Hence (325) agrees with the stated formula, (321). □

We come next to an integral formula which is often taken as the definition of $\Gamma(z)$:

Proposition 8.18. *For every z with $\operatorname{Re} z > 0$ we have*

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}\,dt. \tag{326}$$

Proof. Set $f(z) = \int_0^\infty e^{-t}t^{z-1}\,dt$. Using the bound $\big|t^{z-1}\big| \le t^{\operatorname{Re} z-1}$ we see that the integral is uniformly absolutely convergent in any compact subset of $\{z : \operatorname{Re} z > 0\}$; hence $f(z)$ is an analytic function in $D = \{z : \operatorname{Re} z > 0\}$. Integration by parts shows that for any $z\in D$,

$$f(z+1) = \lim_{T\to\infty}\int_0^T e^{-t}t^z\,dt = \lim_{T\to\infty}\Big(\Big[-e^{-t}t^z\Big]_{t=0}^{t=T} + z\int_0^T e^{-t}t^{z-1}\,dt\Big) = zf(z).$$

Since also $\Gamma(z+1) = z\Gamma(z)$, it follows that the function $g(z) := \frac{f(z)}{\Gamma(z)}$ $(z\in D)$ satisfies $g(z+1) = g(z)$ for all $z\in D$. Note also that $g(z)$ is analytic in D. Next we set

$$h(\zeta) := g(\frac{\log\zeta}{2\pi i}) \qquad (\zeta\in\mathbb{C}\setminus\{0\}),$$

where for each ζ we choose any of the possible values of $\log\zeta$ which have positive imaginary part, so that $\frac{\log\zeta}{2\pi i} \in D$. Since different choices of $\log\zeta$ differ by an integer multiple of $2\pi i$, the resulting values for $\frac{\log\zeta}{2\pi i}$ differ by an integer; so from $g(z+1)=g(z)$ it follows that $h(\zeta)$ is well-defined, i.e. $h(\zeta)$ does not depend on the choice of $\log\zeta$. Hence $h(\zeta)$ is an analytic function of $\zeta \in \mathbb{C}\setminus\{0\}$.

We wish to bound $h(\zeta)$ as $|\zeta|\to 0$ and $|\zeta|\to\infty$, and for this it suffices to bound $g(z)$ in a vertical strip of width 1, say for $1 \le \operatorname{Re} z \le 2$. It follows from Theorem 8.17 that for $1\le x\le 2$ and $y\ge 1$ we have

$$\begin{aligned}\log\big|\Gamma(x\pm iy)\big| &= \operatorname{Re}\,\log\Gamma(x+iy) = \operatorname{Re}\,\big((x-\tfrac12+iy)\log(x+iy)\big) - x + \log\sqrt{2\pi} + O(y^{-1})\\ &= (x-\tfrac12)\log|x+iy| - y\arg(x+iy) - x + \log\sqrt{2\pi} + O(y^{-1})\\ &\ge 0 - \tfrac{\pi}{2}y - 2 + \log\sqrt{2\pi} + O(y^{-1})\end{aligned}$$

For any fixed $\varepsilon>0$ the above expression is $\ge -\frac{\pi}{2}(1+\varepsilon)y$ for all sufficiently large y, and thus $\big|\Gamma(x\pm iy)\big| \ge e^{-\frac{\pi}{2}(1+\varepsilon)y}$. On the other hand $f(x+iy)$ is bounded in the whole strip $1\le x\le 2$, since

$$\big|f(x+iy)\big| \le \int_0^\infty e^{-t}t^{x-1}\,dt = f(x).$$

Hence $\big|g(x\pm iy)\big| = O\big(e^{\frac{\pi}{2}(1+\varepsilon)y}\big)$ as $y\to\infty$. Using $\Big|\operatorname{Im}\frac{\log\zeta}{2\pi i}\Big| = \frac{1}{2\pi}\big|\log|\zeta|\big|$ we see that this implies $h(\zeta)=O(|\zeta|^{-\frac14(1+\varepsilon)})$ as $|\zeta|\to 0$ and $h(\zeta)=O(|\zeta|^{\frac14(1+\varepsilon)})$ as $|\zeta|\to\infty$. It follows that $h(\zeta)$ has a removable singularity at 0, i.e. $h(\zeta)$ extends to an entire function. Similarly $h(\zeta^{-1})$ also has a removable singularity at $\zeta=0$ and it follows that $h(\zeta)$ is a bounded entire function, hence a constant. Hence also $g(z)=\frac{f(z)}{\Gamma(z)}$ is a constant function. To compute the constant, we need only note that $\Gamma(1)=0!=1$ and $f(1)=\int_0^\infty e^{-t}\,dt=1$; hence $g(z)=1$ for all $z\in D$, and this completes the proof of the proposition. □

8.3. **Problems.**

Problem 8.1. Prove that $\gamma = -\Gamma'(1)$.

Problem 8.2. Prove that $\Gamma(\frac12)=\sqrt{\pi}$.

Problem 8.3. Prove that $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ for all $a,b\in\mathbb{C}$ with $\operatorname{Re} a>0$, $\operatorname{Re} b>0$.

Problem 8.4. Prove Legendre's duplication formula,

$$\Gamma(2z) = \pi^{-\frac12}2^{2z-1}\Gamma(z)\Gamma(z+\tfrac12).$$

Problem 8.5. Prove the following asymptotic formula for the absolute value of $\Gamma(z)$ in a vertical strip: For any fixed real numbers $a \le b$ we have

$$\big|\Gamma(x+iy)\big| = \big|\Gamma(x-iy)\big| = \sqrt{2\pi}\, y^{x-\frac12} e^{-\frac{\pi}{2}y}\big(1+O(y^{-1})\big), \tag{327}$$

uniformly over all $x \in [a,b]$ and all $y \ge 1$.

Problem 8.6. Prove that for any fixed $\varepsilon > 0$ and $\alpha \in \mathbb{C}$ we have

$$\log\Gamma(z+\alpha) = \big(z+\alpha-\tfrac12\big)\log z - z + \log\sqrt{2\pi} + O\big(|z|^{-1}\big), \tag{328}$$

for all z with $|z| \ge 1$, $|z+\alpha| \ge 1$ and $\big|\arg(z+\alpha)\big| \le \pi - \varepsilon$. (The implied constant depends on ε and α but of course not on z. Also in the right hand side we use the principal branch of the logarithm function.)

Problem 8.7. Prove that for any fixed $\varepsilon > 0$ we have

$$\frac{\Gamma'(z)}{\Gamma(z)} = \log z - \frac{1}{2z} + O\big(|z|^{-2}\big) \tag{329}$$

for all z with $|z| \ge 1$ and $\big|\arg z\big| \le \pi - \varepsilon$.
[Hint. One method of proof is to use Theorem 8.17 and Cauchy's integral formula for $\frac{\Gamma'(z)}{\Gamma(z)} = \frac{d}{dz}\log\Gamma(z)$. Compare the proof of (280) in Proposition 7.3. Note that we generally wish to use Cauchy's integral formula with a circle which is as large as possible, in order to minimize the error term.]

The procedure used in the proof of Theorem 8.17 to estimate the sum $\sum_{n=0}^{N}\log(z+n)$ can be generalized, and repeated to get successively more precise estimates; this method is called Euler-Maclaurin summation, and we outline it in the following problem.

Problem 8.8. Recall that $B_1(r) := r - \frac12$ and $\overline{B}_1(r) := B_1(r - \lfloor r \rfloor)$. We now define the *Bernoulli polynomials*, $B_n(r)$ $(n = 2,3,\ldots)$, recursively by the relations $B_n'(r) = nB_{n-1}(r)$ and $\int_0^1 B_n(r)\,dr = 0$.
(a). Prove that this determines $B_n(r)$, $n = 2,3,\ldots$ uniquely, and compute $B_n(r)$ for $n = 2,3,4$.
(b). Prove that $B_n(1-r) = B_n(r)$ for all even n, and $B_n(1-r) = -B_n(r)$ for all odd n. Also prove that $B_n(0) = B_n(1) = 0$ for all odd $n \ge 3$. Hence deduce that $\overline{B}_n(r)$ is *continuous* for $n \ge 2$.
(c). Prove that for any $h \in \mathbb{Z}^+$, any real numbers $A < B$ and any function $f \in C^h([A,B])$ we have the *Euler-Maclaurin summation formula:*

$$\sum_{\substack{n\in\mathbb{Z}\\ A<n\le B}} f(n) = \int_A^B f(x)\,dx + \sum_{r=1}^{h}\frac{(-1)^r}{r!}\Big[\overline{B}_r(x) f^{(r-1)}(x)\Big]_{x=A}^{x=B} + (-1)^{h-1}\int_A^B \frac{\overline{B}_h(x)}{h!} f^{(h)}(x)\,dx. \tag{330}$$

[Hint. Use integration by parts repeatedly. Note that in the proof of Theorem 8.17 we showed this formula for $f(n) = \log(z+n)$ and $h = 1$; see (322); also (323) is a first step

towards getting the formula for $h = 2$.]
(d). Using the above, prove the following more precise version of Stirling's formula: Write $B_n := B_n(0)$; this is called the nth *Bernoulli number;* note that $0 = B_3 = B_5 = B_7 = \dots$ by part (b). Let m be any fixed non-negative integer. Then for any fixed $\varepsilon > 0$ we have

(331)
$$\log\Gamma(z) = \big(z - \tfrac{1}{2}\big)\log z - z + \log\sqrt{2\pi} + \sum_{k=0}^{m} \frac{B_{2k+2}}{(2k+2)(2k+1)} z^{-2k-1} + O\big(|z|^{-2m-3}\big)$$

for all z with $|z| \geq 1$ and $\big|\arg z\big| \leq \pi - \varepsilon$. (The implied constant depends on m and ε but of course not on z.)

9. The functional equation

(Davenport chapters 8-9.)

9.1. The case of $\zeta(s)$.

Theorem 9.1. *The function $\zeta(s)$ can be continued analytically over the whole plane and is then meromorphic, its only pole being a simple pole at $s = 1$ with residue 1. Furthermore, $\zeta(s)$ satisfies the functional equation (for all $s \in \mathbb{C} \setminus \{0, 1\}$)*

$$\Lambda(s) = \Lambda(1-s) \quad \text{when } \Lambda(s) := \pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s). \tag{332}$$

The function $\Lambda(s)$ is also analytic in the whole plane except for simple poles at the points $s = 0$ and $s = 1$.

Remark 9.1. It is customary to define

$$\xi(s) := \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s).$$

Then the above theorem shows that $\xi(s)$ is an *entire* function which satisfies the symmetry relation $\xi(s) = \xi(1-s)$.

Proof. We start with the integral definition of the Γ-function: We have

$$\Gamma(\tfrac{1}{2}s) = \int_0^\infty e^{-t}t^{\frac{1}{2}s-1}\,dt \qquad (\sigma > 0). \tag{333}$$

cf. Proposition 8.18. Substituting $t = n^2\pi x$ we get

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)n^{-s} = \int_0^\infty x^{\frac{1}{2}s-1}e^{-n^2\pi x}\,dx. \tag{334}$$

Hence for $\sigma > 1$ we have

$$\begin{aligned}\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \sum_{n=1}^\infty \pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)n^{-s} &= \sum_{n=1}^\infty \int_0^\infty x^{\frac{1}{2}s-1}e^{-n^2\pi x}\,dx \\ &= \int_0^\infty x^{\frac{1}{2}s-1}\Big(\sum_{n=1}^\infty e^{-n^2\pi x}\Big)\,dx,\end{aligned} \tag{335}$$

where the change or order is justified by the fact that we have absolute convergence

$$\sum_{n=1}^\infty \int_0^\infty \left|x^{\frac{1}{2}s-1}e^{-n^2\pi x}\right| dx = \sum_{n=1}^\infty \int_0^\infty x^{\frac{1}{2}\sigma-1}e^{-n^2\pi x}\,dx = \sum_{n=1}^\infty \pi^{-\frac{1}{2}\sigma}\Gamma(\tfrac{1}{2}\sigma)n^{-\sigma} < \infty.$$

Let us now write

$$\omega(x) = \sum_{n=1}^\infty e^{-n^2\pi x} \qquad (x > 0)$$

so that

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \int_0^\infty x^{\frac{1}{2}s-1}\omega(x)\,dx = \int_1^\infty x^{\frac{1}{2}s-1}\omega(x)\,dx + \int_1^\infty x^{-\frac{1}{2}s-1}\omega(x^{-1})\,dx. \tag{336}$$

Note that

$$2\omega(x) = \theta(x) - 1$$

where

$$\theta(x) = \sum_{n=-\infty}^{\infty} e^{-n^2\pi x} \qquad (x > 0)$$

is a classical *theta function*, which is known to satisfy the simple symmetry relation

$$\theta(x^{-1}) = x^{\frac{1}{2}}\theta(x), \qquad \forall x > 0. \tag{337}$$

(We give a proof of this below; cf. Theorem 9.2.) It follows that

$$\omega(x^{-1}) = -\tfrac{1}{2} + \tfrac{1}{2}\theta(x^{-1}) = -\tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}}\theta(x) = -\tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}} + x^{\frac{1}{2}}\omega(x). \tag{338}$$

Hence

$$\begin{aligned}\int_1^\infty x^{-\frac{1}{2}s-1}\omega(x^{-1})\,dx &= \int_1^\infty x^{-\frac{1}{2}s-1}\big(-\tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}} + x^{\frac{1}{2}}\omega(x)\big)\,dx \\ &= -\frac{1}{s} + \frac{1}{s-1} + \int_1^\infty x^{-\frac{1}{2}s-\frac{1}{2}}\omega(x)\,dx,\end{aligned}$$

and now from (336) we get

$$\Lambda(s) := \pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = -\frac{1}{s} + \frac{1}{s-1} + \int_1^\infty \big(x^{\frac{1}{2}s-1} + x^{-\frac{1}{2}s-\frac{1}{2}}\big)\omega(x)\,dx \tag{339}$$

We have proved this under the assumption that $\sigma > 1$. But the integral on the right converges absolutely for any s, uniformly for s in any compact subset of $\mathbb{C}$, since

$$\omega(x) = O(e^{-\pi x}) \qquad \text{as } x \to \infty. \tag{340}$$

[Proof of (340): Since $n^2 \geq n$ for all $n \geq 1$, we have for every $x > 0$: $\omega(x) \leq \sum_{n=1}^\infty e^{-n\pi x} = \frac{e^{-\pi x}}{1-e^{-\pi x}}$, and this is $\ll e^{-\pi x}$ as $x \to \infty$.]

Hence the integral in (339) represents an everywhere analytic function of s, and the formula (339) gives the analytic continuation of $\zeta(s)$ over the whole plane. Since $\pi^{\frac{1}{2}s}\Gamma(\frac{1}{2}s)^{-1}$ is an entire function, the only possible singularities of $\zeta(s)$ must come from the two simple poles of the right hand side of (339) at $s = 0$ and at $s = 1$. However $\Gamma(\frac{1}{2}s)^{-1} = 0$ at $s = 0$; hence the only possible pole of $\zeta(s)$ is at $s = 1$, and indeed since $\pi^{\frac{1}{2}}\Gamma(\frac{1}{2})^{-1} = 1$ we find that $\zeta(s)$ has a simple pole at $s = 1$ with residue 1.

Note also that (339) implies the functional equation (332), since the right hand side of (339) is unchanged when s is replaced by $1 - s$. □

To complete the above proof we still have to prove the symmetry relation of the theta function $\theta(x)$, (337). We shall prove this in a more general form, which we will need shortly when we turn to the functional equation for the general Dirichlet L-function:

Theorem 9.2. *For any $\alpha \in \mathbb{C}$ and $x > 0$ we have*

$$\sum_{n=-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = x^{\frac{1}{2}} \sum_{n=-\infty}^{\infty} e^{-n^2\pi x + 2\pi i n\alpha}. \tag{341}$$

Proof. We fix x and α as above, and set

$$f(t) := e^{-(t+\alpha)^2\pi/x}.$$

Now by Poisson's summation formula we have (note that the function f satisfies all the conditions in Lemma 5.15, with flying colors):

$$\sum_{n=-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = \sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \widehat{f}(n). \tag{342}$$

Here

$$\begin{aligned} \widehat{f}(y) = \int_{-\infty}^{\infty} e^{-(t+\alpha)^2\pi/x} e^{-2\pi i y t}\, dt &= \int_{-\infty}^{\infty} \exp\Big(-\frac{\pi}{x}\big(t + ixy + \alpha\big)^2 - \pi x y^2 + 2\pi i y\alpha\Big)\, dt \\ &= e^{-\pi x y^2 + 2\pi i y\alpha} \int_{-\infty}^{\infty} e^{-\frac{\pi}{x}(t+ixy+\alpha)^2}\, dt \end{aligned} \tag{343}$$

Now for any $\delta > 0$ and $\beta \in \mathbb{C}$ we have

$$\int_{-\infty}^{\infty} e^{-\delta(t+\beta)^2}\, dt = \int_{-\infty}^{\infty} e^{-\delta t^2}\, dt. \tag{344}$$

This simply expresses a movement in the path of integration from the real axis to another line parallel to it. [Details: After performing the substitution $t_{new} = t + \operatorname{Re}\beta$ in the left hand side integral, we see that we may assume $\operatorname{Re}\beta = 0$, thus $\beta = ci$ for some $c \in \mathbb{R}$. Now for $T > 0$ we let R_T be the rectangular path in the complex plane going from $-T$ to T, then to $T + ci$, then to $-T + ci$ and then back to T. By Cauchy's Theorem we have $\int_{R_T} e^{-\delta t^2}\, dt = 0$, which gives

$$\int_{-T}^{T} e^{-\delta t^2}\, dt + \int_{S_{T,1}} e^{-\delta t^2}\, dt - \int_{-T}^{T} e^{-\delta(t+ci)^2}\, dt + \int_{S_{T,2}} e^{-\delta t^2}\, dt = 0$$

where $S_{T,1}$ is the line segment going from T to $T + ci$, and $S_{T,2}$ is the line segment going from $-T + ci$ to $-T$. Now for t on $S_{T,1}$ we have $t = T + c_1 i$ for some c_1 between 0 and c, and thus $\operatorname{Re} t^2 = T^2 - c_1^2 \geq T^2 - c^2$ and $\left|e^{-\delta t^2}\right| \leq e^{\delta c^2} e^{-\delta T^2}$. This bound implies that $\int_{S_{T,1}} e^{-\delta t^2}\, dt \to 0$ as $T \to \infty$. Similarly $\int_{S_{T,2}} e^{-\delta t^2}\, dt \to 0$ as $T \to \infty$. Hence the above

identity implies that

$$\lim_{T\to\infty}\Big(\int_{-T}^{T} e^{-\delta t^2}\,dt - \int_{-T}^{T} e^{-\delta(t+ci)^2}\,dt\Big) = 0,$$

q.e.d.] Hence we may continue the computation from (343) as follows:

$$\widehat{f}(y) = e^{-\pi x y^2+2\pi i y\alpha}\int_{-\infty}^{\infty} e^{-\frac{\pi}{x}t^2}\,dt \quad \Big\{\text{subst. } t=\sqrt{x}u\Big\} = e^{-\pi x y^2+2\pi i y\alpha}x^{\frac12}A,$$

where A is the positive constant $A=\int_{-\infty}^{\infty} e^{-\pi u^2}\,du$. Using this in (342) we get

$$\sum_{n=-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = Ax^{\frac12}\sum_{n=-\infty}^{\infty} e^{-n^2\pi x+2\pi i n\alpha}.$$

In particular taking $\alpha=0$ and applying the last formula twice we conclude $A=1$. Hence we have proved (341). □

Using the functional equation for $\zeta(s)$ we can give a basic description of where the zeros of $\zeta(s)$ are located. Recall that we have defined $\xi(s)=\frac12 s(s-1)\pi^{-\frac12 s}\Gamma(\frac12 s)\zeta(s)$ and that this is an entire function satisfying $\xi(s)=\xi(1-s)$. (cf. Remark 9.1).

Corollary 9.3. *(i) The zeros of $\xi(s)$ (if any exist) are all situated in the critical strip, $\{0\le\sigma\le 1\}$, and these zeros are placed symmetrically with respect to the real axis, and also symmetrically with respect to the central line $\sigma=\frac12$.*
(ii) The zeros of $\zeta(s)$ are identical (in position and order of multiplicity) with those of $\xi(s)$, except that $\zeta(s)$ has a simple zero at each of the points $s=-2,-4,-6,\dots$.

Proof. We know from the Euler product that $\zeta(s)$ does not have any zeros in the half-plane $\{\sigma>1\}$. We also know that $\frac12 s(s-1)\pi^{-\frac12 s}\Gamma(\frac12 s)\neq 0$ for all s in this half-plane. Hence $\xi(s)$ does not have any zeros in $\{\sigma>1\}$. Because of the symmetry relation $\xi(s)=\xi(1-s)$ it follows that $\xi(s)$ does not have any zeros in $\{\sigma<0\}$ either.

Hence all the zeros of $\xi(s)$ must lie in the critical strip $\{0\le\sigma\le 1\}$. It follows from $\xi(\overline{s})=\overline{\xi(s)}$ that these zeros are placed symmetrically with respect to the real axis, and it follows from $\xi(s)=\overline{\xi(\overline{s})}=\overline{\xi(1-\overline{s})}$ that the zeros are also placed symmetrically with respect to the line $\sigma=\frac12$. Hence (i) is proved.

Next note that the relation $\xi(s)=\frac12 s(s-1)\pi^{-\frac12 s}\Gamma(\frac12 s)\zeta(s)$ shows that the zeros of $\zeta(s)$ can differ from those of $\xi(s)$ only in so far as the function $h(s)=\frac12 s(s-1)\pi^{-\frac12 s}\Gamma(\frac12 s)$ has some zeros or poles. Note $h(s)=(s-1)\pi^{-\frac12 s}\Gamma(\frac12 s+1)$; hence the only zero of $h(s)$ is at $s=1$ (a simple zero), and this is not a zero of either $\xi(s)$ or $\zeta(s)$, since we know that $\zeta(s)$ has a simple pole at $s=1$ (and thus $\xi(1)\neq 0$). The only poles of $h(s)$ are simple poles at $s=-2,-4,-6,\dots$. Since these are points where $\xi(s)$ is analytic and not zero, it follows that they must be simple zeros of $\zeta(s)$. This proves (ii). □

Remark 9.2. In fact the zeros of $\xi(s)$ (if any exist) are all situated in the *open* critical strip, $\{0 < \sigma < 1\}$. This follows by combining the above corollary with Theorem 7.4.

Definition 9.1. The zeros $s = -2, -4, -6, \dots$ of $\zeta(s)$ are called the *trivial zeros*; the other zeros of $\zeta(s)$ are called the *non-trivial zeros*.

9.2. **Gauss sums (II).** Let χ be a Dirichlet character modulo q. In order to prove the functional equation for the Dirichlet L-function $L(s, \chi)$, we first need to express the function $n \mapsto \chi(n)$ as a linear combination of the imaginary exponentials $n \mapsto e\Big(\frac{mn}{q}\Big)$, for $m = 0, 1, \dots, q-1$ (or "$m \bmod q$")

Note that the Dirichlet characters $\chi \in X_q$ are exactly the *multiplicative* characters on $\mathbb{Z}/q\mathbb{Z}$ (i.e. characters on the group $(\mathbb{Z}/q\mathbb{Z})^\times$, cf. §4.5), while the functions $n \mapsto e\Big(\frac{mn}{q}\Big)$ are the *additive* characters on $\mathbb{Z}/q\mathbb{Z}$ (i.e. characters on the group $\mathbb{Z}/q\mathbb{Z}$ with its usual addition). Thus what we wish to do here is to express the multiplicative characters in terms of additive ones.

Definition 9.2. For any Dirichlet character χ modulo q we define the *Gaussian sum* $\tau(\chi)$ by

$$\tau(\chi) = \sum_{m \in \mathbb{Z}/q\mathbb{Z}} \chi(m) e\Big(\frac{m}{q}\Big). \tag{345}$$

Note that if $(n, q) = 1$ then, letting n^{-1} be a multiplicative inverse of n in $(\mathbb{Z}/q\mathbb{Z})^\times$,

$$\chi(n)\tau(\overline{\chi}) = \chi(n) \sum_{m \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(m) e\Big(\frac{m}{q}\Big) = \sum_{m \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(n^{-1}m) e\Big(\frac{m}{q}\Big) = \sum_{h \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h) e\Big(\frac{hn}{q}\Big). \tag{346}$$

This gives the desired expression for $\chi(n)$, provided that $(n, q) = 1$ and that $\tau(\chi) \neq 0$.

The following lemma shows that *if χ is primitive*, then the final relation in (346) holds for *all* n.

Lemma 9.4. *Let χ be a primitive Dirichlet character modulo q. Then for every $n \in \mathbb{Z}/q\mathbb{Z}$ which is not in $(\mathbb{Z}/q\mathbb{Z})^\times$ we have*

$$\sum_{h \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h) e\Big(\frac{hn}{q}\Big) = 0. \tag{347}$$

Hence for ***all*** *$n \in \mathbb{Z}/q\mathbb{Z}$ we have*

$$\chi(n)\tau(\overline{\chi}) = \sum_{h \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h) e\Big(\frac{hn}{q}\Big). \tag{348}$$

Proof. Now to prove (347), let n be any element in $\mathbb{Z}/q\mathbb{Z}$ which is not in $(\mathbb{Z}/q\mathbb{Z})^\times$. Set $a = q/(n,q)$; then we can run through all the elements $h \in \mathbb{Z}/q\mathbb{Z}$ by writing $h = ac + b$ with $c = 0, 1, \dots, (n,q) - 1$ and $b = 0, 1, \dots, a-1$, and thus

$$\sum_{h \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h) e\Big(\frac{hn}{q}\Big) = \sum_{b=0}^{a-1} \sum_{c=0}^{(n,q)-1} \overline{\chi}(ac+b) e\Big(\frac{(ac+b)n}{q}\Big) = \sum_{b=0}^{a-1} e\Big(\frac{bn}{q}\Big) \sum_{c=0}^{(n,q)-1} \overline{\chi}(ac+b),$$

where in the last step we used the fact that $e\big(\frac{(ac+b)n}{q}\big) = e\big(\frac{bn}{q}\big)$, since $\frac{acn}{q} = \frac{cn}{(n,q)} \in \mathbb{Z}$. But $\overline{\chi}$ is a primitive character since χ is, and note that $q \nmid a$, since $(n,q) > 1$. Hence by Problem 4.4 the inner sum in the last expression *vanishes* for every b. (Indeed, note that $\overline{\chi}(ac+b)$ is periodic with period (n,q) in the variable c; hence the inner sum can be written as $\frac{(n,q)}{q} \sum_{c=0}^{q-1} \overline{\chi}(ac+b)$, and this is 0 by Problem 4.4.) This proves the formula (347). □

We next prove a result which in particular entails that $\tau(\overline{\chi}) \neq 0$ for every primitive χ, so that the formula (348) indeed solves the problem of expressing χ as a linear combination of the imaginary exponentials $n \mapsto e\Big(\frac{mn}{q}\Big)$.

Lemma 9.5. *Let χ be a primitive Dirichlet character modulo q. Then*

$$\big|\tau(\chi)\big| = q^{\frac{1}{2}}.$$

Proof. The standard method to prove this type of result would be to expand $|\tau(\chi)|^2$ as a double sum. This turns out to be a slightly non-trivial but quite useful learning experience; cf. Problem 9.3 below. Here we give instead a very simple, but somewhat less direct proof: By (348) we have

$$\big|\chi(n)\big|^2 \big|\tau(\overline{\chi})\big|^2 = \sum_{h_1 \in \mathbb{Z}/q\mathbb{Z}} \sum_{h_2 \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h_1)\chi(h_2) e\Big(\frac{n(h_1-h_2)}{q}\Big).$$

We now add this relation over all $n \in \mathbb{Z}/q\mathbb{Z}$. The sum of the values of $\big|\chi(n)\big|^2$ is $\phi(q)$, and the sum of the exponentials is 0 unless $h_1 \equiv h_2 \pmod q$.[20] Hence

$$\phi(q)\big|\tau(\overline{\chi})\big|^2 = q \sum_{h \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(h)\chi(h) = q\phi(q).$$

This gives the stated formula. □

Let us note that in the special case $\chi = \big(\frac{d}{\cdot}\big) \in X_{|d|}$ (with $d \neq 0$, $d \equiv 0$ or $1 \pmod 4$ as usual) we have already proved an exact formula for $\tau(\chi)$. Indeed, by Theorem 5.13 (with

[20] To spell this out completely explicitly: We have $\sum_{n \in \mathbb{Z}/q\mathbb{Z}} e(\frac{nh}{q}) = 0$ if $q \nmid h$. (Proof: We have $e(\frac{h}{q}) \sum_{n \in \mathbb{Z}/q\mathbb{Z}} e(\frac{nh}{q}) = \sum_{n \in \mathbb{Z}/q\mathbb{Z}} e(\frac{(n+1)h}{q}) = \sum_{n \in \mathbb{Z}/q\mathbb{Z}} e(\frac{nh}{q})$. This proves the claim, since $e(\frac{h}{q}) \neq 0$ when $q \nmid h$.) Note that in the remaining case, $q \mid h$, we trivially have $\sum_{n \in \mathbb{Z}/q\mathbb{Z}} e(\frac{nh}{q}) = q$.

$n = 1$) we have

$$\tau\left(\left(\frac{d}{\cdot}\right)\right) = \begin{cases} \sqrt{d} & \text{if } d > 0 \\ i\sqrt{|d|} & \text{if } d < 0. \end{cases} \tag{349}$$

(Note also that the case of general n in Theorem 5.13 can be seen as a direct consequence of (349) together with (348).)

9.3. **The functional equation for a general Dirichlet L-function.** The following was first given by Hurwitz in 1882 (in the special case of real characters):

Theorem 9.6. *Let χ be a primitive character modulo $q \geq 3$ (thus χ is nonprincipal). Then the Dirichlet L-function $L(s,\chi)$ has an analytic continuation to an entire function. Furthermore, $L(s,\chi)$ satisfies the following functional equation:*

$$\xi(1-s,\overline{\chi}) = \frac{i^a q^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi) \quad \textit{when } \xi(s,\chi) = (\pi/q)^{-\frac{1}{2}(s+a)}\Gamma\big(\tfrac{1}{2}(s+a)\big)L(s,\chi). \tag{350}$$

Here $a = 0$ if $\chi(-1) = 1$ and $a = 1$ if $\chi(-1) = -1$. The function $\xi(s,\chi)$ is also entire.

(Note that we always have $\chi(-1) = \pm 1$, since $\chi(-1)^2 = \chi((-1)^2) = \chi(1) = 1$.)

Proof. The proof is along the same lines as the proof of Theorem 9.1, but with some new technicalities.

Substituting $t = n^2\pi x/q$ in (333) we get

$$\pi^{-\frac{1}{2}s} q^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)n^{-s} = \int_0^\infty x^{\frac{1}{2}s-1} e^{-n^2\pi x/q}\,dx \qquad (\sigma > 0). \tag{351}$$

Hence for $\sigma > 1$ we have

$$\begin{aligned} \pi^{-\frac{1}{2}s} q^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)L(s,\chi) = \sum_{n=1}^\infty \pi^{-\frac{1}{2}s} q^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\chi(n)n^{-s} &= \sum_{n=1}^\infty \chi(n)\int_0^\infty x^{\frac{1}{2}s-1} e^{-n^2\pi x/q}\,dx \\ &= \int_0^\infty x^{\frac{1}{2}s-1}\Big(\sum_{n=1}^\infty \chi(n) e^{-n^2\pi x/q}\Big)\,dx \end{aligned}$$

where the change of order is justified as in the proof of Theorem 9.1.

Let us first assume $\chi(-1) = 1$. Then we have $\chi(-n) = \chi(n)$ for all $n \in \mathbb{Z}$, and $\chi(0) = 0$, and hence we can write the last formula as

$$\pi^{-\frac{1}{2}s} q^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)L(s,\chi) = \frac{1}{2}\int_0^\infty x^{\frac{1}{2}s-1}\psi(x,\chi)\,dx \tag{352}$$

where

$$\psi(x,\chi) := \sum_{n=-\infty}^{\infty} \chi(n) e^{-n^2\pi x/q} \qquad (x > 0). \tag{353}$$

A symmetry relation between $\psi(x,\chi)$ and $\psi(x^{-1},\overline{\chi})$ can be deduced from (348) and Theorem 9.2 with x replaced by x/q:

$$\begin{aligned}\tau(\overline{\chi})\psi(x,\chi) &= \sum_{n=-\infty}^{\infty}\Big(\sum_{m=1}^{q}\overline{\chi}(m)e\Big(\frac{mn}{q}\Big)\Big)e^{-n^2\pi x/q}\\ &= \sum_{m=1}^{q}\overline{\chi}(m)\sum_{n=-\infty}^{\infty}e^{-n^2\pi x/q+2\pi inm/q}\\ &= \sum_{m=1}^{q}\overline{\chi}(m)(q/x)^{\frac12}\sum_{n=-\infty}^{\infty}e^{-(n+m/q)^2\pi q/x} \qquad (354)\\ &= (q/x)^{\frac12}\sum_{m=1}^{q}\overline{\chi}(m)\sum_{n=-\infty}^{\infty}e^{-(qn+m)^2\pi/(qx)}\\ &= (q/x)^{\frac12}\sum_{\ell=-\infty}^{\infty}\overline{\chi}(\ell)e^{-\ell^2\pi/(qx)} = (q/x)^{\frac12}\psi(x^{-1},\overline{\chi}).\end{aligned}$$

Now we split the integral in (352) into two parts and obtain

$$\begin{aligned}\xi(s,\chi) := \pi^{-\frac12 s}q^{\frac12 s}\Gamma(\tfrac12 s)L(s,\chi) &= \frac12\int_1^\infty x^{\frac12 s-1}\psi(x,\chi)\,dx + \frac12\int_1^\infty x^{-\frac12 s-1}\psi(x^{-1},\chi)\,dx\\ &= \frac12\int_1^\infty x^{\frac12 s-1}\psi(x,\chi)\,dx + \frac12\frac{q^{\frac12}}{\tau(\overline{\chi})}\int_1^\infty x^{-\frac12 s-\frac12}\psi(x,\overline{\chi})\,dx.\end{aligned}$$

This expression represents an everywhere analytic function of s, i.e. we have proved that $\xi(s,\chi)$ is an entire function. The expression also gives the analytic continuation of $L(s,\chi)$ over the whole plane; we see that $L(s,\chi)$ is an entire function since $\Gamma(\frac12 s)$ is never 0. Moreover, replacing s by $1-s$ and χ by $\overline{\chi}$ in the above formula we get

$$\xi(1-s,\overline{\chi}) = \frac12\frac{q^{\frac12}}{\tau(\chi)}\int_1^\infty x^{\frac12 s-1}\psi(x,\chi)\,dx + \frac12\int_1^\infty x^{-\frac12 s-\frac12}\psi(x,\overline{\chi})\,dx.$$

Now note that

$$\tau(\chi)\tau(\overline{\chi}) = q, \qquad (355)$$

since $\tau(\overline{\chi}) = \sum_{m\in\mathbb{Z}/q\mathbb{Z}}\overline{\chi(m)}e(m/q) = \overline{\sum_{m\in\mathbb{Z}/q\mathbb{Z}}\chi(m)e(-m/q)} = \overline{\sum_{m\in\mathbb{Z}/q\mathbb{Z}}\chi(-m)e(-m/q)} = \overline{\tau(\chi)}$, where $\chi(-m)\equiv\chi(m)$ comes from our assumption $\chi(-1)=1$, and $\tau(\chi)\overline{\tau(\chi)} = q$ by Lemma 9.5. Hence the above formulas imply

$$\xi(1-s,\overline{\chi}) = \frac{q^{\frac12}}{\tau(\chi)}\xi(s,\chi),$$

i.e. we have proved (350) in the case $\chi(-1)=1$.

We next turn to the case $\chi(-1) = -1$. The previous argument fails, since now the function $\psi(x,\chi)$ simply vanishes (since $\chi(-n) \equiv -\chi(n)$). We modify the procedure by replacing s with $s+1$ in (351), giving:

$$\pi^{-\frac{1}{2}(s+1)} q^{\frac{1}{2}(s+1)} \Gamma(\tfrac{1}{2}(s+1)) n^{-s} = \int_0^\infty n x^{\frac{1}{2}s-\frac{1}{2}} e^{-n^2\pi x/q}\, dx \qquad (\sigma > -1). \tag{356}$$

In the same way as before this yields, when $\sigma > 1$:

$$\begin{aligned} \pi^{-\frac{1}{2}(s+1)} q^{\frac{1}{2}(s+1)} \Gamma(\tfrac{1}{2}(s+1)) L(s,\chi) &= \int_0^\infty x^{\frac{1}{2}s-\frac{1}{2}} \Big(\sum_{n=1}^\infty n\chi(n) e^{-n^2\pi x/q} \Big)\, dx \\ &= \frac{1}{2} \int_0^\infty \psi_1(x,\chi) x^{\frac{1}{2}s-\frac{1}{2}}\, dx, \end{aligned} \tag{357}$$

where

$$\psi_1(x,\chi) := \sum_{n=-\infty}^\infty n\chi(n) e^{-n^2\pi x/q} \qquad (x > 0). \tag{358}$$

To prove a symmetry relation for $\psi_1(x,\chi)$ we use a *differentiated* version of Theorem 9.2. Namely, differentiating (341) (written with "y" in place of "x") with respect to α, we obtain

$$-\frac{2\pi}{y} \sum_{n=-\infty}^\infty (n+\alpha) e^{-(n+\alpha)^2\pi/y} = 2\pi i y^{\frac{1}{2}} \sum_{n=-\infty}^\infty n e^{-n^2\pi y + 2\pi i n\alpha}. \tag{359}$$

Setting here $y = x/q$ and $\alpha = m/q$ we get

$$\sum_{n=-\infty}^\infty n e^{-n^2\pi x/q + 2\pi i m n/q} = i(q/x)^{\frac{3}{2}} \sum_{n=-\infty}^\infty (n+m/q) e^{-\pi(n+m/q)^2 q/x}. \tag{360}$$

Using this we can now carry out a computation analogous to (354):

$$\begin{aligned} \tau(\overline{\chi})\psi_1(x,\chi) &= \sum_{m=1}^q \overline{\chi}(m) \sum_{n=-\infty}^\infty n e^{-n^2\pi x/q + 2\pi i n m/q} \\ &= i(q/x)^{\frac{3}{2}} \sum_{m=1}^q \overline{\chi}(m) \sum_{n=-\infty}^\infty (n+m/q) e^{-\pi(n+m/q)^2 q/x} \\ &= i q^{\frac{1}{2}} x^{-\frac{3}{2}} \sum_{m=1}^q \overline{\chi}(m) \sum_{n=-\infty}^\infty (nq+m) e^{-\pi(qn+m)^2/(qx)} \\ &= i q^{\frac{1}{2}} x^{-\frac{3}{2}} \sum_{\ell=-\infty}^\infty \ell\overline{\chi}(\ell) e^{-\pi\ell^2/(qx)} = i q^{\frac{1}{2}} x^{-\frac{3}{2}} \psi_1(x^{-1}, \overline{\chi}) \end{aligned}$$

Using this symmetry relation in (357) we obtain

$$\begin{aligned}\xi(s,\chi) &:= \pi^{-\frac{1}{2}(s+1)}q^{\frac{1}{2}(s+1)}\Gamma(\tfrac{1}{2}(s+1))L(s,\chi)\\ &= \frac{1}{2}\int_1^\infty \psi_1(x,\chi)x^{\frac{1}{2}s-\frac{1}{2}}\,dx + \frac{1}{2}\int_1^\infty \psi_1(x^{-1},\chi)x^{-\frac{1}{2}s-\frac{3}{2}}\,dx\\ &= \frac{1}{2}\int_1^\infty \psi_1(x,\chi)x^{\frac{1}{2}s-\frac{1}{2}}\,dx + \frac{1}{2}\frac{iq^{\frac{1}{2}}}{\tau(\overline{\chi})}\int_1^\infty \psi_1(x,\overline{\chi})x^{-\frac{1}{2}s}\,dx.\end{aligned}$$

This again gives the analytic continuation of $\xi(s,\chi)$ and $L(s,\chi)$ to entire functions. Furthermore, using the fact that now when $\chi(-1)=-1$ we have

$$\tau(\chi)\tau(\overline{\chi}) = -q \tag{361}$$

(the proof is exactly as for (355) except that we now have $\chi(-m)\equiv -\chi(m)$), we obtain

$$\xi(1-s,\overline{\chi}) = \frac{iq^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi),$$

i.e. we have proved (350) in the case $\chi(-1)=-1$. □

Corollary 9.7. *Let χ be a primitive character modulo $q\geq 3$.*
(i) The zeros of $\xi(s,\chi)$ (if any exist) are all situated in the critical strip $\{0\leq\sigma\leq 1\}$, with neither $s=0$ or $s=1$ being a zero. These zeros are placed symmetrically about the line $\sigma=\frac{1}{2}$.
(ii) The zeros of $L(s,\chi)$ are identical (in position and order of multiplicity) with those of $\xi(s,\chi)$, except that $L(s,\chi)$ has a simple zero at each point $s=-a,-a-2,-a-4,\ldots$.

(Here again we write $a=0$ if $\chi(-1)=1$, $a=1$ if $\chi(-1)=-1$.)

Proof. We know from the Euler product that $L(s,\chi)$ does not have any zeros in the half-plane $\{\sigma>1\}$. We also know that $(\pi/q)^{-\frac{s+a}{2}}\Gamma\big(\frac{s+a}{2}\big)\neq 0$ for all $s\in\mathbb{C}$. Hence $\xi(s,\chi)=(\pi/q)^{-\frac{s+a}{2}}\Gamma\big(\frac{s+a}{2}\big)L(s,\chi)$ does not have any zeros in $\{\sigma>1\}$. The same argument applies to show that $\xi(s,\overline{\chi})$ does not have any zeros in $\{\sigma>1\}$. Because of the symmetry relation $\xi(1-s,\overline{\chi})=\frac{i^a q^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi)$ it follows that $\xi(s,\chi)$ does not have any zeros in $\{\sigma<0\}$ either. Furthermore since $L(1,\chi)$ and $L(1,\overline{\chi})$ are non-zero (cf. (28)) we see that $\xi(1,\chi)\neq 0$ and $\xi(1,\overline{\chi})\neq 0$; hence also $\xi(0,\chi)\neq 0$ and $\xi(0,\overline{\chi})\neq 0$.

Combining the functional equation with $\xi(1-s,\overline{\chi})=\overline{\xi(1-\overline{s},\chi)}$ we obtain $\frac{i^a q^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi)=\overline{\xi(1-\overline{s},\chi)}$, and this relation shows that s is a zero of $\xi(\cdot,\chi)$ is and only if $1-\overline{s}$ is a zero with the same multiplicity. Thus we have proved (i).

Next note that the relation $\xi(s,\chi)=(\pi/q)^{-\frac{s+a}{2}}\Gamma\big(\frac{s+a}{2}\big)L(s,\chi)$ shows that the zeros of $L(s,\chi)$ can differ from those of $\xi(s,\chi)$ only in so far as the function $h(s)=(\pi/q)^{-\frac{s+a}{2}}\Gamma\big(\frac{s+a}{2}\big)$ has some zeros or poles. But $h(s)$ does not have any zeros, and the only poles of $h(s)$ are

simple poles at s with $\frac{s+a}{2} \in \{0, -1, -2, -3, \ldots\}$, in other words at $s = -a, -a-2, -a-4, \ldots$. All of these points lie in $\{\sigma < 0\}$ except the point $s = 0$ in the case $a = 0$; hence we know that $\xi(s, \chi)$ is analytic and non-zero at each of these points. It follows that each of these points must be a simple zero of $L(s, \chi)$. This proves (ii). □

Definition 9.3. The zeros of $L(s, \chi)$ with $\sigma < 0$, as well as the zero $s = 0$ in the case $\chi(-1) = 1$, are called the *trivial zeros*; the other zeros of $L(s, \chi)$ are called the *non-trivial zeros*.

9.4. **Problems.**

Problem 9.1. (a). Prove that the functional equation for $\zeta(s)$ can be written in the following form:

$$\zeta(1-s) = 2(2\pi)^{-s} \cos\big(\tfrac{\pi}{2}s\big)\Gamma(s)\zeta(s). \tag{362}$$

(b). Prove that if χ is a primitive Dirichlet character modulo $q \geq 3$ then the functional equation for $L(s, \chi)$ can be written in the following form:

$$L(1-s, \overline{\chi}) = \frac{2i^a q^s}{(2\pi)^s \tau(\chi)} \begin{Bmatrix} \cos(\frac{\pi}{2}s) & \text{if } a = 0 \\ \sin(\frac{\pi}{2}s) & \text{if } a = 1 \end{Bmatrix} \Gamma(s) L(s, \chi). \tag{363}$$

(c). Prove that the functional equation in (b) can also be expressed as

$$L(1-s, \chi) = \frac{q^{s-1}\Gamma(s)}{(2\pi)^s}\big(e^{-\pi i s/2} + \chi(-1)e^{\pi i s/2}\big)\tau(\chi) L(s, \overline{\chi}). \tag{364}$$

Problem 9.2. Prove that if χ is a Dirichlet character modulo q which is not primitive, and if the corresponding primitive Dirichlet character is χ_1 modulo $q_1 = c(\chi)$, then

$$\tau(\chi) = \mu\Big(\frac{q}{q_1}\Big)\chi_1\Big(\frac{q}{q_1}\Big)\tau(\chi_1). \tag{365}$$

[Hint: Compare Davenport p. 67.]

Problem 9.3. When trying to prove the formula $\big|\tau(\chi)\big| = \sqrt{q}$ (cf. Lemma 9.5) by expanding $\big|\tau(\chi)\big|^2$ as a double sum one encounters the so called *Ramanujan sum* $c_q(n)$;

$$c_q(n) := \sum_{m \in (\mathbb{Z}/q\mathbb{Z})^\times} e\Big(\frac{nm}{q}\Big) \qquad (q \in \mathbb{Z}^+,\ n \in \mathbb{Z}/q\mathbb{Z}). \tag{366}$$

(a). Prove that

$$c_q(n) = \sum_{d \mid (q,n)} \mu\Big(\frac{q}{d}\Big) d = \frac{\phi(q)}{\phi(\frac{q}{(q,n)})} \mu\Big(\frac{q}{(q,n)}\Big). \tag{367}$$

(Remark: the first of these two formulas is often the one which is most convenient to use in applications.)
(b). Use (a) to prove that $\big|\tau(\chi)\big| = \sqrt{q}$ for every primitive character χ modulo q.

Problem 9.4. (a). (Difficult!) Give an alternative proof of the analytic continuation and functional equation for $\zeta(s)$ (in the form (362)) by expanding $\Gamma(s)\zeta(s)$ in the same way as we expanded $\Gamma(\frac{1}{2}s)\zeta(s)$ in the proof of Theorem 9.1.

[Hint. After some computation we arrive at the integral $\int_0^\infty \frac{x^{s-1}}{e^x-1}\,dx$, which we wish to analytically continue (as a function of s) from the original region of convergence, $\{\sigma > 1\}$, to the whole complex plane. This can be done by instead studying $\int_C \frac{z^{s-1}}{e^{-z}-1}\,dz$, where C is a curve going from $-\infty$ to "$-\varepsilon$", encircling 0 once and then going back to $-\infty$. To get the functional equation one may try to change the contour C in a way to pick up contributions from the residues of $\frac{z^{s-1}}{e^{-z}-1}$ at the points $z = 2\pi i n$, $n \in \mathbb{Z} \setminus \{0\}$.]

(b). Prove (e.g. from the computations in (a)) that $\zeta(-n) = \mathrm{Res}_{z=0}\frac{z^{-n-1}}{e^z-1}$ for all nonnegative integers n.

(c). Prove that $\zeta(-n) = -\frac{B_{n+1}}{n+1}$ for all nonnegative integers n, where B_n is the Bernoulli number defined in Problem 8.8(d).

[Hint. Use (b). One way to obtain the formula is to first prove that the Bernoulli polynomials $B_n(r)$ defined in Problem 8.8 satisfy $\frac{ze^{rz}}{e^z-1} = \sum_{n=0}^\infty \frac{B_n(r)}{n!}z^n$ for $z, r \in \mathbb{C}$, $|z|$ small. In fact this relation is often taken as the definition of the Bernoulli polynomials.]

(d). Prove that $\zeta(2m) = 2^{2m-1}\pi^{2m}\frac{|B_{2m}|}{(2m)!}$ for all $m \in \mathbb{Z}^+$.

Problem 9.5. (a). Give an alternative proof of the meromorphic continuation of $\zeta(s)$ to the whole complex plane using Euler-Maclaurin summation (cf. Problem 8.8).

(b). Can you also find a proof of the functional equation using this method?

(c). Compute $\zeta(-n)$ using this method, and compare with Problem 9.4(c).

10. The Infinite Products for $\xi(s)$ and $\xi(s,\chi)$

(Davenport Chapter 12)

10.1. **The infinite products for $\xi(s)$.** Recall that we have defined

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s), \tag{368}$$

and that this is an entire function which satisfies $\xi(1-s) = \xi(s)$ (Remark 9.1). Also recall that the zeros of $\xi(s)$ (if any exist) are all situated in the open critical strip, $\{0 < \sigma < 1\}$, and these zeros are placed symmetrically with respect to the real axis, and also symmetrically with respect to the central line $\sigma = \frac{1}{2}$. Also recall that these zeros coincide exactly with the *non-trivial* zeros of $\zeta(s)$. Cf. Corollary 9.3, Remark 9.2 and Definition 9.1.

We will now apply the Weierstrass factorization theorem, Theorem 8.7, to $\xi(s)$. We first determine the order of $\xi(s)$.

Proposition 10.1. *There is a constant $C > 0$ such that*

$$|\xi(s)| < e^{C|s|\log|s|} \qquad \textit{when } |s| \textit{ is sufficiently large.} \tag{369}$$

*On the other hand there does **not** exist any choice of $C_1 > 0$ such that $|\xi(s)| < O\big(e^{C_1|s|}\big)$ as $|s| \to \infty$. Hence $\xi(s)$ has order 1.*

Proof. Since $\xi(1-s) = \xi(s)$ it suffices to prove (369) when $\sigma \geq \frac{1}{2}$. Obviously

$$\big|\tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\big| < e^{C_1|s|}$$

when $|s|$ is large (where we can take C_1 to be any fixed constant $> \frac{1}{2}\log\pi$). Also Stirling's formula Theorem 8.17 applies, since $\big|\arg(\frac{1}{2}s)\big| < \frac{1}{2}\pi$ because of $\sigma \geq \frac{1}{2} > 0$, and this gives

$$\big|\Gamma(\tfrac{1}{2}s)\big| < e^{C_2|s|\log|s|}$$

when $|s|$ is large (where we can take C_2 to be any fixed constant $> \frac{1}{2}$). Finally recall from Proposition 7.3 that $\big|\zeta(s)\big| \ll |t|^{\frac{1}{2}}$ for all s with $\sigma \geq \frac{1}{2}$ and $|t| \geq 1$; also $\zeta(s)$ is bounded in the half plane $\{\sigma \geq 2\}$, since there $\big|\zeta(s)\big| \leq \sum_{n=1}^{\infty} n^{-2}$. Hence always

$$\big|\zeta(s)\big| \leq e^{C_3|s|}$$

when $\sigma \geq \frac{1}{2}$ and $|s|$ is large (where we can take C_3 to be any fixed constant > 0). Now (369) follows by multiplying our three bounds (and we see that we can take C to be any fixed constant $> \frac{1}{2}$).

Finally to prove that $|\xi(s)| < O\big(e^{C_1|s|}\big)$ cannot holds as $|s| \to \infty$ it suffices to consider real s tending to $+\infty$. Indeed for such s we have $\zeta(s) \to 1$ while $\log\Gamma(s) \sim s\log s$ by Stirling's formula; hence certainly $\xi(s) > e^{0.99\cdot s\log s}$ for all sufficiently large (real) s. □

Theorem 10.2. *The entire function $\xi(s)$ has infinitely many zeros $\rho_1, \rho_2, \ldots$. These have the property that $\sum |\rho_n|^{-1-\varepsilon}$ converges for any $\varepsilon > 0$ but $\sum |\rho_n|^{-1}$ diverges. Furthermore there are constants A, B such that*

$$\xi(s) = e^{A+Bs} \prod_{n=1}^{\infty} \Big(1 - \frac{s}{\rho_n}\Big) e^{s/\rho_n} \tag{370}$$

and

$$\frac{\xi'(s)}{\xi(s)} = B + \sum_{n=1}^{\infty} \Big(\frac{1}{s-\rho_n} + \frac{1}{\rho_n} \Big). \tag{371}$$

Proof. This is a direct consequence of Weierstrass factorization theorem; see Theorem 8.7, Remark 8.2, Proposition 8.8 (cf. Corollary 8.9 for the present case) used together with Proposition 10.1. □

The following consequence will be the basis for much of the later work on $\zeta(s)$:

Proposition 10.3.

$$\frac{\zeta'(s)}{\zeta(s)} = B - \frac{1}{s-1} + \tfrac{1}{2}\log\pi - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+1)}{\Gamma(\frac{s}{2}+1)} + \sum_{\rho} \Big(\frac{1}{s-\rho} + \frac{1}{\rho} \Big),$$

where the sum is taken over all the non-trivial zeros ρ of $\zeta(s)$.

Proof. We write (368) in the form

$$\zeta(s) = (\tfrac{1}{2}s)^{-1}(s-1)^{-1}\pi^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)^{-1}\xi(s) = (s-1)^{-1}\pi^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s+1)^{-1}\xi(s).$$

Taking the logarithmic derivative of both sides we get

$$\frac{\zeta'(s)}{\zeta(s)} = -\frac{1}{s-1} + \tfrac{1}{2}\log\pi - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+1)}{\Gamma(\frac{s}{2}+1)} + \frac{\xi'(s)}{\xi(s)}. \tag{372}$$

Using now (371), this gives the stated formula. □

The above formula exhibits the pole of $\zeta(s)$ at $s = 1$ and the nontrivial zeros at $s = \rho$. The trivial zeros at $s = -2, -4, \ldots$ are contained in the Γ-term, since, by Lemma 8.14,

$$-\frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+1)}{\Gamma(\frac{1}{2}s+1)} = \tfrac{1}{2}\gamma + \frac{1}{s+2} + \sum_{n=1}^{\infty}\Big(\frac{1}{s+2+2n} - \frac{1}{2n}\Big) = \tfrac{1}{2}\gamma + \sum_{n=1}^{\infty}\Big(\frac{1}{s+2n} - \frac{1}{2n}\Big). \tag{373}$$

We next show that the constants A, B, though not very important, can be evaluated.

Proposition 10.4. *The constants A, B in (370) are given by*

$$e^A = \tfrac{1}{2}$$

and

$$B = -\tfrac{1}{2}\gamma - 1 + \tfrac{1}{2}\log 4\pi \approx -0.023 \tag{374}$$

where γ is Euler's constant. Furthermore if $\rho = \beta + i\gamma$ are all the nontrivial zeros of $\zeta(s)$ then

$$B = -\sum_{\rho} \frac{1}{\rho} = -2\sum_{\gamma>0} \frac{\beta}{\beta^2+\gamma^2}, \tag{375}$$

where in the first sum we group together the terms from ρ and $\overline{\rho}$.

Proof. By (368),

$$\xi(1) = \tfrac{1}{2}\pi^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}) \lim_{s\to 1}(s-1)\zeta(s) = \tfrac{1}{2}, \tag{376}$$

whence $\xi(0) = \frac{1}{2}$ and therefore $e^A = \frac{1}{2}$ by (370).

As regards B, we have

$$B = \frac{\xi'(0)}{\xi(0)} = -\frac{\xi'(1)}{\xi(1)}$$

from (371) and the functional equation $\xi(s) = \xi(1-s)$.

We next use (372) to evaluate $\frac{\xi'(1)}{\xi(1)}$. For this we note that, by (373),

$$\frac{1}{2}\frac{\Gamma'(\frac{3}{2})}{\Gamma(\frac{3}{2})} = -\tfrac{1}{2}\gamma - \sum_{n=1}^{\infty}\Big(\frac{1}{1+2n} - \frac{1}{2n}\Big) = -\tfrac{1}{2}\gamma + 1 - \log 2 \tag{377}$$

(for note that $\log 2 = \sum_{m=1}^{\infty}(-1)^{m-1}m^{-1}$, cf. (236)). Hence we obtain:

$$\begin{aligned} B = -\frac{\xi'(1)}{\xi(1)} &= \tfrac{1}{2}\log\pi - \frac{1}{2}\frac{\Gamma'(\frac{3}{2})}{\Gamma(\frac{3}{2})} - \lim_{s\to 1}\Big(\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}\Big) \\ &= \tfrac{1}{2}\gamma - 1 + \tfrac{1}{2}\log 4\pi - \lim_{s\to 1}\Big(\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}\Big). \end{aligned}$$

The last limit is easily calculated using the first two terms of the Laurent expansion of $\zeta(s)$ at $s=1$: $\zeta(s) = \frac{1}{s-1} + \gamma + O(|s-1|)$ as $s \to 1$; cf. Problem 7.2. This implies that

$$\begin{aligned} \frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1} &= \frac{-(s-1)^{-2} + O(1)}{(s-1)^{-1} + \gamma + O(|s-1|)} + \frac{1}{s-1} \\ &= -\frac{1}{s-1}\cdot\big(1 + O(|s-1|^2)\big)\cdot\big(1 - \gamma(s-1) + O(|s-1|^2)\big) + \frac{1}{s-1} \\ &= \gamma + O(|s-1|^2) \to \gamma \qquad \text{as } \ s \to 1. \end{aligned}$$

Hence we obtain the formula (374).

Finally to prove (375), take $s = 1$ in (371) and recall that we have proved $B = -\frac{\xi'(1)}{\xi(1)}$; this gives

$$-B = B + \sum_{\rho} \Big(\frac{1}{\rho} + \frac{1}{1-\rho} \Big).$$

Here (because of $\xi(s) = \xi(1-s)$) we know that when ρ runs through the zeros of $\xi(s)$ (with multiplicity) then so does $1-\rho$, and hence *if* the sum $\sum_\rho \rho^{-1}$ converges then the above relation can be rewritten into $2B = -2\sum \rho^{-1}$, as desired. However we have to be careful, since we know that $\sum |\rho|^{-1}$ diverges.

We claim that $\sum \rho^{-1}$ converges, provided one groups together the terms from ρ and $\overline{\rho}$. For if $\rho = \beta + i\gamma$ then

$$\rho^{-1} + \overline{\rho}^{-1} = \frac{2\beta}{\beta^2+\gamma^2} \le \frac{2}{|\rho|^2}, \tag{378}$$

and we know that $\sum |\rho|^{-2}$ converges. This concludes the proof of (375). □

Remark 10.1. From (374) and (375) together it is easy to see that $|\gamma| > 6$ for all nontrivial zeros of $\zeta(s)$. [Proof: Let $\rho = \beta + i\gamma$ be a nontrivial zero with $|\gamma|$ minimal. Then since the zeros are placed symmetrically both about the real axis and about $\sigma = \frac{1}{2}$, we may assume that $\gamma \ge 0$ and $\beta \ge \frac{1}{2}$. Hence by (374) we have $-B \ge \frac{2\beta}{\beta^2+\gamma^2} \ge \frac{1}{\beta^2+\gamma^2} \ge \frac{1}{1+\gamma^2}$; hence $\gamma \ge \sqrt{(-B)^{-1}-1} \approx \sqrt{0.023^{-1}-1} > 6$.]

10.2. **The infinite products for $\xi(s,\chi)$.** Next we apply similar considerations to the L-functions. Let χ be a primitive character modulo $q \ge 3$. Recall that we have defined, in Theorem 9.6:

$$\xi(s,\chi) = (\pi/q)^{-\frac{1}{2}(s+a)} \Gamma\Big(\frac{s+a}{2}\Big) L(s,\chi), \tag{379}$$

where $a = 0$ if $\chi(-1) = 1$ and $a = 1$ if $\chi(-1) = -1$; the function $\xi(s,\chi)$ is entire, and we have $\xi(1-s,\overline{\chi}) = \frac{i^a q^{\frac{1}{2}}}{\tau(\chi)} \xi(s,\chi)$, wherein $\big|\frac{i^a q^{\frac{1}{2}}}{\tau(\chi)}\big| = 1$. Also recall that the zeros of $\xi(s,\chi)$ (if any exist) are all situated in the critical strip, $\{0 \le \sigma \le 1\}$ (with no zero at $s = 0$ or $s = 1$), and these zeros are placed symmetrically with respect to the central line $\sigma = \frac{1}{2}$, but in general not symmetrically about the real axis. Also recall that these zeros coincide exactly with the *non-trivial* zeros of $L(s,\chi)$. Cf. Corollary 9.7 and Definition 9.3.

Proposition 10.5. *There is a constant $C > 0$ (in fact independent of q) such that*

$$|\xi(s,\chi)| < e^{C|s|\log|s|} \qquad \textit{when } |s| \textit{ is sufficiently large.} \tag{380}$$

On the other hand, there does ***not*** *exist any choice of $C_1 > 0$ such that $|\xi(s,\chi)| < O(e^{C_1|s|})$ as $|s| \to \infty$. Hence $\xi(s,\chi)$ has order 1.*

Proof. Since $\xi(1-s,\overline{\chi}) = \frac{i^a q^{\frac12}}{\tau(\chi)}\xi(s,\chi)$ it suffices to prove (380) when $\sigma \ge \frac12$. Obviously

$$\big|(\pi/q)^{-\frac12(s+a)}\big| < e^{C_1|s|}$$

when $|s|$ is large (where we can take C_1 to be any fixed constant $> \frac12\big|\log(\pi/q)\big|$). Also Stirling's formula (Theorem 8.17) applies, since $\big|\arg(\frac12(s+a))\big| < \frac12\pi$ because of $\sigma \ge \frac12$, and this gives

$$\big|\Gamma(\tfrac12(s+a))\big| < e^{C_2|s|\log|s|}$$

when $|s|$ is large (where we can take C_2 to be any fixed constant $> \frac12$). Finally we have the trivial bound

$$\big|L(s,\chi)\big| \le 2q|s| \qquad (\text{when } \sigma \ge \tfrac12), \tag{381}$$

which follows directly from the formula $L(s,\chi) = s\int_1^\infty A(x)x^{-s-1}\,dx$, with $A(x) = \sum_{1\le n\le x}\chi(n)$, cf. Example 3.5. (Details: Note that $\big|A(x)\big| \le q$ for all x, by Lemma 3.13 and the fact that $|\chi(m)| \le 1$ for all m. Hence we get $\big|L(s,\chi)\big| \le |s|\int_1^\infty qx^{-\sigma-1}\,dx \le |s|\int_1^\infty qx^{-\frac32}\,dx \le 2q|s|$, as claimed.) Hence

$$\big|L(s,\chi)\big| \le e^{C_3|s|}$$

when $|s|$ is large (where we can take C_3 to be any fixed constant > 0). Now (380) follows by multiplying our three bounds (and we see that we can take C to be any fixed constant $> \frac12$).

Finally the proof that $|\xi(s,\chi)| < O\big(e^{C_1|s|}\big)$ cannot hold is exactly as in the proof of Proposition 10.1. □

(We remark that we give an improvement of (381) in Problem 10.1(a) – although still very basic and far from optimal.)

Theorem 10.6. *The entire function $\xi(s,\chi)$ has infinitely many zeros $\rho_1,\rho_2,\ldots$. These have the property that $\sum|\rho_n|^{-1-\varepsilon}$ converges for any $\varepsilon > 0$ but $\sum|\rho_n|^{-1}$ diverges. Furthermore there exist constants $A = A(\chi), B = B(\chi)$ such that*

$$\xi(s,\chi) = e^{A+Bs}\prod_{n=1}^{\infty}\Big(1-\frac{s}{\rho_n}\Big)e^{s/\rho_n}, \tag{382}$$

and

$$\frac{\xi'(s,\chi)}{\xi(s,\chi)} = B + \sum_{n=1}^{\infty}\Big(\frac{1}{s-\rho_n}+\frac{1}{\rho_n}\Big). \tag{383}$$

Proof. This follows from Corollary 8.9 and Proposition 10.5. □

The following consequence will be the basis for much of the later work on $L(s,\chi)$:

Proposition 10.7.

$$\frac{L'(s,\chi)}{L(s,\chi)} = -\tfrac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})} + B(\chi) + \sum_\rho\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big).$$

Proof. This follows from (379) and (383). □

We can give explicit expressions for $A(\chi)$ and $B(\chi)$; the latter in terms of $L(1,\chi)$ and $L'(1,\chi)$ (recall that we know $L(1,\chi)\neq 0$). However it seems to be difficult to estimate $B(\chi)$ at all satisfactorily as a function of q.

Proposition 10.8. *The constants $A(\chi)$ and $B(\chi)$ in Theorem 10.6 are given by*

$$e^{A(\chi)} = \frac{i^a q^{1+\frac{1}{2}a}\pi^{-a}}{\tau(\overline{\chi})}L(1,\overline{\chi})$$

and

$$B(\chi) = -\frac{L'(1,\overline{\chi})}{L(1,\overline{\chi})} - \tfrac{1}{2}\log\frac{q}{\pi} + \tfrac{1}{2}\gamma + (1-a)\log 2 \tag{384}$$

where γ is Euler's constant. Furthermore $B(\overline{\chi}) = \overline{B(\chi)}$ and

$$\operatorname{Re} B(\chi) = -\sum_\rho \operatorname{Re}\frac{1}{\rho} \tag{385}$$

(and here $\operatorname{Re}\frac{1}{\rho} > 0$ for all ρ).

Proof. Using (382) with $s=0$ and then $\xi(1-s,\chi) = \frac{i^a q^{\frac{1}{2}}}{\tau(\overline{\chi})}\xi(s,\overline{\chi})$ and (379), we get

$$e^{A(\chi)} = \xi(0,\chi) = \frac{i^a q^{\frac{1}{2}}}{\tau(\overline{\chi})}\xi(1,\overline{\chi}) = \frac{i^a q^{\frac{1}{2}}}{\tau(\overline{\chi})}(\pi/q)^{-\frac{1}{2}(1+a)}\Gamma(\tfrac{1}{2}(1+a))L(1,\overline{\chi}) = \frac{i^a q^{1+\frac{1}{2}a}\pi^{-a}}{\tau(\overline{\chi})}L(1,\overline{\chi}).$$

As regards $B(\chi)$, using (383) with $s=0$, the functional equation $\xi(1-s,\chi) = \frac{i^a q^{\frac{1}{2}}}{\tau(\overline{\chi})}\xi(s,\overline{\chi})$, and then (379), we get

$$B(\chi) = \frac{\xi'(0,\chi)}{\xi(0,\chi)} = -\frac{\xi'(1,\overline{\chi})}{\xi(1,\overline{\chi})} = -\frac{L'(1,\overline{\chi})}{L(1,\overline{\chi})} - \tfrac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{1}{2}+\frac{a}{2})}{\Gamma(\frac{1}{2}+\frac{a}{2})} \tag{386}$$

Note here that, by (373),

$$\frac{1}{2}\frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})} = -\tfrac{1}{2}\gamma - \sum_{n=1}^{\infty}\Big(\frac{1}{2n-1}-\frac{1}{2n}\Big) = -\tfrac{1}{2}\gamma - \log 2$$

(alternative: this also follows from $\frac{\Gamma'(s+1)}{\Gamma(s+1)} = \frac{\Gamma'(s)}{\Gamma(s)} + \frac{1}{s}$ and (377)), and

$$\frac{1}{2}\frac{\Gamma'(1)}{\Gamma(1)} = -\frac{1}{2}\gamma.$$

Hence we obtain (384).

The formula $B(\overline{\chi}) = \overline{B(\chi)}$ is clear e.g. from (384).

Finally to prove (385), note that (386) and (383) give

$$B(\chi) = -\frac{\xi'(1,\overline{\chi})}{\xi(1,\chi)} = -B(\overline{\chi}) - \sum_{n=1}^{\infty}\Big(\frac{1}{1-\overline{\rho}_n} + \frac{1}{\overline{\rho}_n}\Big),$$

where $\rho_1, \rho_2, \dots$ are the zeros of $\xi(s,\chi)$ (with multiplicity), so that $\overline{\rho}_1, \overline{\rho}_2, \dots$ are the zeros of $\xi(s,\overline{\chi})$. Using $B(\overline{\chi}) = \overline{B(\chi)}$ this implies

$$2\mathrm{Re}\, B(\chi) = -\sum_{n=1}^{\infty}\Big(\mathrm{Re}\,\frac{1}{1-\overline{\rho}_n} + \mathrm{Re}\,\frac{1}{\overline{\rho}_n}\Big).$$

But we know that $0 \le \mathrm{Re}\,\rho_n \le 1$ and $\rho_n \notin \{0,1\}$; hence $\big|\arg(\overline{\rho}_n)\big| \le \frac{\pi}{2}$ and thus $\big|\arg(1/\overline{\rho}_n)\big| \le \frac{\pi}{2}$ and $\mathrm{Re}\,\frac{1}{\overline{\rho}_n} \ge 0$; similarly $\mathrm{Re}\,\frac{1}{1-\overline{\rho}_n} \ge 0$. Since all terms $\mathrm{Re}\,\frac{1}{\overline{\rho}_n}$ and $\mathrm{Re}\,\frac{1}{1-\overline{\rho}_n}$ in the above sum are non-negative we may change the order of summation arbitrarily. We use the fact that the sequence $1-\overline{\rho}_1, 1-\overline{\rho}_2, \dots$ is a permutation of $\rho_1, \rho_2, \dots$ (since the zeros $\rho_1, \rho_2, \dots$ lie symmetrically about the line $\sigma = \frac{1}{2}$). Furthermore $\mathrm{Re}\,\frac{1}{\overline{\rho}_n} = \mathrm{Re}\,\frac{1}{\rho_n}$. Hence

$$2\mathrm{Re}\, B(\chi) = -\sum_{n=1}^{\infty}\Big(\mathrm{Re}\,\frac{1}{\rho_n} + \mathrm{Re}\,\frac{1}{\overline{\rho}_n}\Big) = -2\sum_{n=1}^{\infty}\mathrm{Re}\,\frac{1}{\rho_n},$$

i.e. we have proved (385). □

Remark 10.2. If χ in Proposition 10.8 is real then $B(\chi) = -\frac{\xi'(1,\chi)}{\xi(1,\chi)}$ is real and hence (385) gives $B(\chi) = -\sum_\rho \mathrm{Re}\,\frac{1}{\rho}$. Also in this case the zeros ρ are placed symmetrically about the real axis, so that writing $\rho = \beta + i\gamma$ and using $\mathrm{Re}\,\frac{1}{\rho} = \frac{\beta}{\beta^2+\gamma^2}$ we obtain

$$B(\chi) = -\sum_{\rho}\mathrm{Re}\,\frac{1}{\rho} = -\sum_{\rho\ \mathrm{real}}\frac{1}{\rho} - 2\sum_{\gamma>0}\frac{\beta}{\beta^2+\gamma^2}.$$

10.3. **Problems.**

Problem 10.1. Let χ be a non-principal Dirichlet character modulo q.
(a). Prove that for any fixed $0 < \delta < 1$,

$$\big|L(s,\chi)\big| \ll \big(q(1+|t|)\big)^{1-\delta}, \qquad \text{for all } s \text{ with } \sigma \geq \delta, \tag{387}$$

where the implied constant depends only on δ (i.e. it is independent of q and s).

(b). Prove that for any fixed $A > 0$, $\varepsilon > 0$,

$$\big|L(s,\chi)\big| \ll \log(|t|+2) + \log q, \qquad \text{for all } s \text{ with } \sigma \geq \max\Big(\varepsilon, 1 - \frac{A}{\log(|t|+2)+\log q}\Big),$$

where the implied constant depends only on A and ε.

(c). Prove that for any fixed $A > 0$, $\varepsilon > 0$,

$$\big|L'(s,\chi)\big| \ll \big(\log(|t|+2) + \log q\big)^2, \qquad \text{for all } s \text{ with } \sigma \geq \max\Big(\varepsilon, 1 - \frac{A}{\log(|t|+2)+\log q}\Big),$$

where the implied constant depends only on A, ε.

[Hint for (a)-(c): Start with the formula $L(s,\chi) = s\int_1^\infty A(x)x^{-s-1}\,dx$ as in the discussion of (381), but generalize it by including the sum $\sum_{1\leq n\leq X}\chi(n)n^{-s}$ for some arbitrary $X \geq 1$, i.e. in an analogous way as (282) generalizes (278). Then compare the proof of Prop. 7.3.]

11. Zero-free regions for $\zeta(s)$ and $L(s,\chi)$

11.1. A zero-free region for $\zeta(s)$.

Theorem 11.1. *There exists a constant $c>0$ such that $\zeta(s)$ has no zero in the region*

$$\sigma \geq 1 - \frac{c}{\log(|t|+2)}.$$

Proof. As in Theorem 7.4 the proof is based on the elementary inequality

$$3 + 4\cos\theta + \cos 2\theta \geq 0, \qquad \forall \theta \in \mathbb{R}, \tag{388}$$

but we can now make a sharper argument since we have access to the infinite product formula for $\zeta(s)$ ((368), Theorem 10.2). It is more convenient to work with $\frac{\zeta'(s)}{\zeta(s)}$ than with $\log\zeta(s)$, since the analytic continuation of the latter to the left of $\sigma=1$ is obviously difficult. Recall that $\frac{\zeta'(s)}{\zeta(s)} = -\sum_{n=1}^{\infty}\Lambda(n)n^{-s}$, thus

$$-\mathrm{Re}\,\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty}\Lambda(n)n^{-\sigma}\cos(t\log n) \qquad (\sigma>1).$$

Hence, using (388),

$$3\Big(-\frac{\zeta'(\sigma)}{\zeta(\sigma)}\Big) + 4\Big(-\mathrm{Re}\,\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)}\Big) + \Big(-\mathrm{Re}\,\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)}\Big) \geq 0 \qquad (\sigma>1). \tag{389}$$

As in the proof of Theorem 7.4 we will now fix some t (say $t\geq 2$) and let $\sigma\to 1^+$ in the above inequality.

Since $-\frac{\zeta'(s)}{\zeta(s)}$ has a simple pole at $s=1$ with residue 1, we have for $1<\sigma\leq 2$:

$$-\frac{\zeta'(\sigma)}{\zeta(\sigma)} \leq \frac{1}{\sigma-1} + O(1).$$

The behavior of the other two functions near $\sigma=1$ is obviously much influenced by any zero that $\zeta(s)$ may have just to the left of $\sigma=1$, at a height near to t or $2t$. This influence is rendered explicit by the formula from Proposition 10.3;

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{s-1} - B - \tfrac{1}{2}\log\pi + \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+1)}{\Gamma(\frac{s}{2}+1)} - \sum_{\rho}\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big), \tag{390}$$

where the sum is taken over all the non-trivial zeros ρ of $\zeta(s)$. Here the Γ term is bounded by $O(\log t)$ if $t\geq 2$ and $1\leq\sigma\leq 2$ (this follows from the asymptotic formula in Problem 8.7). Hence, in this region,

$$-\mathrm{Re}\,\frac{\zeta'(s)}{\zeta(s)} \leq O(\log t) - \sum_{\rho}\mathrm{Re}\,\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big). \tag{391}$$

The sum over ρ is non-negative, for if we write $\rho = \beta + i\gamma$ (thus $0 \le \beta \le 1$, $\gamma \in \mathbb{R}$) then

$$\mathrm{Re}\,\frac{1}{s-\rho} = \frac{\sigma-\beta}{|s-\rho|^2} \ge 0 \qquad \text{and} \qquad \mathrm{Re}\,\frac{1}{\rho} = \frac{\beta}{|\rho|^2} \ge 0.$$

Hence (for $1 \le \sigma \le 2$, $t \ge 2$)

$$-\mathrm{Re}\,\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)} \le O(\log t), \tag{392}$$

and furthermore, if we choose t to coincide with the imaginary part γ of one fixed zero $\rho = \beta + i\gamma$, and take just the one term $\mathrm{Re}\,\frac{1}{s-\rho}$ in the sum which corresponds to this zero:

$$-\mathrm{Re}\,\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} \le O(\log t) - \frac{1}{\sigma-\beta}.$$

Substituting our upper bounds in the inequality (389), we get (when $1 < \sigma \le 2$ and $t = \gamma$ for some zero $\rho = \beta + i\gamma$):

$$\frac{3}{\sigma-1} - \frac{4}{\sigma-\beta} + O(1) + O(\log t) \ge 0,$$

i.e. there is an absolute constant $A > 0$ such that

$$\frac{4}{\sigma-\beta} \le \frac{3}{\sigma-1} + A\log t. \tag{393}$$

Note that this is also true, trivially, for $\sigma \ge 2$. From the last inequality it follows that (by solving for $1-\beta$):

$$\sigma - \beta \ge \frac{4}{\frac{3}{\sigma-1} + A\log t} \Longrightarrow 1-\beta \ge \frac{4}{\frac{3}{\sigma-1} + A\log t} - (\sigma-1) = \frac{1 - A(\sigma-1)\log t}{\frac{3}{\sigma-1} + A\log t}.$$

We now make the choice $\sigma - 1 = \frac{1}{2A\log t}$, in order to obtain a positive numerator in the last expression. This yields

$$1-\beta \gg \frac{1}{\log t}. \tag{394}$$

Finally, since $\zeta(\overline{s}) = \overline{\zeta(s)}$ and since $\zeta(s)$ has no zero arbitrarily near $\sigma = 1$ with $|t| \le 2$, we can also say that for *every* $t \in \mathbb{R}$, if there exists any zero with $\rho = \beta + it$ then (possibly with a smaller implied constant than in (394)):

$$1-\beta \gg \frac{1}{\log(|t|+2)}. \tag{395}$$

This completes the proof. □

We remark that the above zero-free region can be improved. The sharpest known zero-free region today, which was obtained independently by Vinogradov and Korobov in 1958,

essentially says that the power of log in Theorem 11.1 can be reduced from 1 to $\frac{2}{3}$: There is an absolute constant $c > 0$ such that $\zeta(s)$ does not have any zeros in the region

$$t \geq 3, \qquad \sigma \geq 1 - \frac{c}{(\log t)^{\frac{2}{3}}(\log\log t)^{\frac{1}{3}}}. \tag{396}$$

See §11.3 around (427) for some further comments.

11.2. **Zero-free Regions for $L(s,\chi)$.** It is easy to extend the previous results to the zeros of $L(s,\chi)$ when χ is a *fixed* character. But for many purposes it is important to also have estimates that are explicit with respect to q. This raises some difficult problems, and the results so far known are better for complex characters than for real characters.

The key to proving zero free regions is the following generalization of (389):

Lemma 11.2. *For any Dirichlet character χ modulo q and any $s = \sigma + it$ with $\sigma > 1$ we have*

$$3\Big(-\frac{L'(\sigma,\chi_0)}{L(\sigma,\chi_0)}\Big) + 4\Big(-\mathrm{Re}\,\frac{L'(\sigma+it,\chi)}{L(\sigma+it,\chi)}\Big) + \Big(-\mathrm{Re}\,\frac{L'(\sigma+2it,\chi^2)}{L(\sigma+2it,\chi^2)}\Big) \geq 0, \tag{397}$$

where χ_0 is the principal character modulo q.

Proof. Recall that logarithmic differentiation of the Euler product formula gives

$$-\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n=1}^{\infty} \chi(n)\Lambda(n)n^{-s} \qquad (\sigma > 1). \tag{398}$$

cf. (113) in Example 3.7. From this we see that the left hand side of (397) equals

$$\sum_{n=1}^{\infty} \Lambda(n)n^{-\sigma} \cdot \mathrm{Re}\,\Big(3\chi_0(n) + 4\chi(n)e^{-it\log n} + \chi(n)^2 e^{-2it\log n}\Big).$$

We claim that each term here is ≥ 0. If $(n,q) > 1$ then the nth term vanishes, since $\chi_0(n) = \chi(n) = 0$. Now assume $(n,q) = 1$. Then $\chi_0(n) = 1$, and $\chi(n) = e^{i\alpha}$ for some $\alpha \in \mathbb{R}$, so that the nth term equals $\Lambda(n)n^{-\sigma}\Big(3 + 4\cos\theta + \cos 2\theta\Big)$ with $\theta = \alpha - t\log n$; and this is ≥ 0 by (388). □

Note that when χ is a real character (and only then) we have $\chi^2 = \chi_0$, and this affects the argument when using (397) to get a zero-free region for $L(s,\chi)$.

When treating the last two terms in (397) we will use the following:

Lemma 11.3. *For any **primitive** Dirichlet character χ modulo $q \geq 3$ and any s with $1 < \sigma \leq 2$ we have*

$$-\mathrm{Re}\,\frac{L'(s,\chi)}{L(s,\chi)} < -\sum_{\rho} \mathrm{Re}\,\frac{1}{s-\rho} + O\Big(\log q + \log(|t|+2)\Big), \tag{399}$$

where the implied constant is absolute, and where ρ runs through all the non-trivial zeros of $L(s,\chi)$, or a (possibly empty) subset of these.

Proof. This is a simple consequence of Proposition 10.7 (which only applies when χ is primitive!). This proposition says:

$$\frac{L'(s,\chi)}{L(s,\chi)} = -\tfrac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})} + B(\chi) + \sum_\rho\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big),$$

(where $a=0$ if $\chi(-1)=1$ and $a=1$ if $\chi(-1)=-1$), and thus

$$-\mathrm{Re}\,\frac{L'(s,\chi)}{L(s,\chi)} = \tfrac{1}{2}\log\frac{q}{\pi} + \frac{1}{2}\mathrm{Re}\,\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})} - \mathrm{Re}\,B(\chi) - \mathrm{Re}\sum_\rho\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big).$$

Here recall that $\mathrm{Re}\,B(\chi) = -\sum_\rho \mathrm{Re}\,\frac{1}{\rho}$ (an absolutely convergent sum, all terms being non-negative), cf. (385). We also note that the Γ term above is $O\big(\log(|t|+2)\big)$, for all $1<\sigma\le 2$ and $t\in\mathbb{R}$, as follows from the asymptotic formula in Problem 8.7. Hence we get (399), with ρ running through all the non-trivial zeros of $L(s,\chi)$.

Finally note that all terms in the sum over ρ are ≥ 0 (thus giving a contribution $\le$ to the right hand side of (399)), for if $\rho=\beta+i\gamma$ is any zero then $\mathrm{Re}\,\frac{1}{s-\rho} = \frac{\sigma-\beta}{|s-\rho|^2} \ge 0$ since $\sigma>1$ and $\beta\le 1$. □

We now give results on zero-free regions. The following two theorems collect results obtained by Gronwall (1913) and Titchmarsh (1930,1933). We start with the case of χ *complex.*

Theorem 11.4. *There exists an absolute constant $c>0$ such that for every $q\in\mathbb{Z}^+$ and every* ***complex*** *Dirichlet character χ modulo q, $L(s,\chi)$ has no zero in the region*

$$\sigma \ge \begin{cases} 1-\frac{c}{\log q|t|} & \text{if } |t|\ge 1,\\ 1-\frac{c}{\log q} & \text{if } |t|\le 1.\end{cases} \tag{400}$$

(Of course, the assumption that χ is complex implies that $q\ge 3$, and that χ is non-principal.)

Proof. Without loss of generality we may (and will) assume $t\ge 0$, for the zeros of $L(s,\chi)$ with $t<0$ are the complex conjugates of the zeros of $L(s,\overline{\chi})$ with $t>0$, since $L(\overline{s},\chi) = \overline{L(s,\overline{\chi})}$.

Let us first assume that χ is *primitive.*

We will use Lemma 11.2. Concerning the first term in (397) we note that

$$-\frac{L'(\sigma,\chi_0)}{L(\sigma,\chi_0)} = \sum_{n=1}^\infty \chi_0(n)\Lambda(n)n^{-\sigma} \le -\frac{\zeta'(\sigma)}{\zeta(\sigma)} < \frac{1}{\sigma-1}+O(1) \tag{401}$$

for $1 < \sigma \le 2$ (the implied constant being absolute).

We choose $t \ge 0$ to be the imaginary part of a non-trivial zero $\rho = \beta + i\gamma$ of $L(s,\chi)$. Then by Lemma 11.3, where in the sum we retain only the one term corresponding to our selected zero ρ, we have

$$-\mathrm{Re}\,\frac{L'(\sigma+ti,\chi)}{L(\sigma+ti,\chi)} < -\frac{1}{\sigma-\beta} + O\Big(\log q + \log(t+2)\Big) \qquad \text{for } 1 < \sigma \le 2. \tag{402}$$

Concerning the last term in (397) we cannot always apply Lemma 11.3 directly to $\chi^2 \pmod q$ since this character may not be primitive. Instead we let χ_1 be the unique primitive character modulo $q_1 = c(\chi^2)$ that induces χ^2 (thus $\chi_1 = \chi^2$ if χ^2 is primitive). Note that χ^2 is non-principal since χ is complex; hence $q_1 \ge 3$ and Lemma 11.3 applies to χ_1, giving (when we take the ρ-sum to be over the empty set)

$$-\mathrm{Re}\,\frac{L'(\sigma+2ti,\chi_1)}{L(\sigma+2ti,\chi_1)} < O\Big(\log q_1 + \log(t+2)\Big) \qquad \text{for } 1 < \sigma \le 2. \tag{403}$$

Also recall that $L(s,\chi^2)$ and $L(s,\chi_1)$ are related by $L(s,\chi^2) = L(s,\chi_1)\prod_{p|q}(1-\chi_1(p)p^{-s})$ (cf. Lemma 4.22); thus

$$\frac{L'(s,\chi^2)}{L(s,\chi^2)} = \frac{L'(s,\chi_1)}{L(s,\chi_1)} + \sum_{p|q}\frac{\chi_1(p)(\log p)p^{-s}}{1-\chi_1(p)p^{-s}}$$

so that

$$\Big|\frac{L'(s,\chi^2)}{L(s,\chi^2)} - \frac{L'(s,\chi_1)}{L(s,\chi_1)}\Big| \le \sum_{p|q}\frac{(\log p)p^{-\sigma}}{1-p^{-\sigma}} \le \sum_{p|q}\log p \le \log q \qquad (\sigma \ge 1). \tag{404}$$

Combining this with (403) we conclude

$$-\mathrm{Re}\,\frac{L'(\sigma+2ti,\chi^2)}{L(\sigma+2ti,\chi^2)} < O\Big(\log q + \log(t+2)\Big). \tag{405}$$

Using now Lemma 11.2 combined with (401), (402), (405) we get

$$\frac{4}{\sigma-\beta} \le \frac{3}{\sigma-1} + O\Big(\log q + \log(t+2)\Big).$$

Exactly as in the argument between (393) and (394) this is seen to imply

$$1-\beta \gg \frac{1}{\log q + \log(t+2)},$$

and this implies that (400) holds with an absolute constant $c > 0$.

It remains to treat the case when χ is not primitive. Then let χ_1 modulo q_1 be the primitive character which induces χ. Since χ is complex so is χ_1, and hence $q_1 \ge 3$ and the

above result applies to χ_1, i.e. $L(s,\chi_1)$ has no zero in the region

$$\sigma \geq \begin{cases} 1 - \frac{c}{\log q_1 |t|} & \text{if } |t| \geq 1, \\ 1 - \frac{c}{\log q_1} & \text{if } |t| \leq 1. \end{cases} \tag{406}$$

But recall that $L(s,\chi) = L(s,\chi_1) \prod_{p|q}(1-\chi_1(p)p^{-s})$; hence the only zeros of $L(s,\chi)$ additional to those of $L(s,\chi_1)$ are on $\sigma = 0$; i.e. also $L(s,\chi)$ is without any zero in the above region. (We here assume – as we may – that the absolute constant $c > 0$ has been fixed to be smaller than $\log 3$, so that (406) implies $\sigma > 0$.) Finally, since $q > q_1$, the region given by (406) *contains* the region given by (400); hence the theorem is proved. □

Note that the above theorem in particular applies for $t = 0$; thus it gives a new proof of the fact that $L(1,\chi) \neq 0$ for each complex character χ!

We next turn to the case of χ *real.*

Theorem 11.5. *There exists an absolute constant $c > 0$ such that for every $q \in \mathbb{Z}^+$ and every **real** nonprincipal Dirichlet character χ modulo q, $L(s,\chi)$ has **at most one** zero in the region*

$$\sigma \geq \begin{cases} 1 - \frac{c}{\log q|t|} & \textit{if } |t| \geq 1, \\ 1 - \frac{c}{\log q} & \textit{if } |t| \leq 1, \end{cases} \tag{407}$$

*and if $L(s,\chi)$ has such a zero then this has to be a (simple) **real** zero.*

Proof. We may assume $t \geq 0$. Let us first assume that χ is *primitive.* The bounds (401) and (402) remain true in the present situation, with the same proofs. However the discussion of $\frac{L'(\sigma+2ti,\chi^2)}{L(\sigma+2ti,\chi^2)}$ needs modification, since now χ^2 is the principal character modulo q, and hence the corresponding primitive character is the trivial character $\chi_1 \equiv 1$ (modulo $q_1 = 1$), i.e. $L(s,\chi_1) = \zeta(s)$, and Lemma 11.3 does not apply to this χ_1. For t large there is no problem; we have $-\mathrm{Re}\,\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)} \leq O(\log t)$ for $1 \leq \sigma \leq 2$, $t \geq 2$, as we saw in (392); however for t small the term $\frac{1}{s-1}$ from (390) may blow up. Now if we go through the argument between (390) and (392) again we find that, for *all* $t \geq 0$ and $1 < \sigma \leq 2$:

$$-\mathrm{Re}\,\frac{\zeta'(s)}{\zeta(s)} \leq \mathrm{Re}\,\frac{1}{s-1} + O\big(\log(t+2)\big)$$

Since we also have $\left|\frac{L'(s,\chi^2)}{L(s,\chi^2)} - \frac{\zeta'(s)}{\zeta(s)}\right| \leq \log q$ (cf. (404), which holds without change when χ^2 is principal), we conclude that, for all $1 < \sigma \leq 2$, $t \geq 0$,

$$-\mathrm{Re}\,\frac{L'(\sigma+2it,\chi^2)}{L(\sigma+2it,\chi^2)} \leq \mathrm{Re}\,\frac{1}{\sigma-1+2it} + O\Big(\log q + \log(t+2)\Big).$$

Using this together with Lemma 11.2, (401) and (402), we get

$$\frac{4}{\sigma-\beta} \le \frac{3}{\sigma-1} + \operatorname{Re}\frac{1}{\sigma-1+2it} + O\Big(\log q + \log(t+2)\Big),$$

where now $t = \gamma \ge 0$, the imaginary part of some non-trivial zero $\rho = \beta + i\gamma$ of $L(s,\chi)$. In other words, there is an absolute constant $A > 0$ such that

$$\frac{4}{\sigma-\beta} \le \frac{3}{\sigma-1} + \frac{\sigma-1}{(\sigma-1)^2+4t^2} + AL, \qquad \text{with } L := \log q + \log(t+2).$$

Note that (with an appropriate choice of A) this is also true, trivially, for $\sigma \ge 2$; i.e. the inequality holds for all $\sigma > 1$. It follows that

$$\sigma - \beta \ge \frac{4}{\frac{3}{\sigma-1} + \frac{\sigma-1}{(\sigma-1)^2+4t^2} + AL}$$

$$\implies 1-\beta \ge \frac{4}{\frac{3}{\sigma-1} + \frac{\sigma-1}{(\sigma-1)^2+4t^2} + AL} - (\sigma-1) = \frac{1 - \frac{(\sigma-1)^2}{(\sigma-1)^2+4t^2} - AL(\sigma-1)}{\frac{3}{\sigma-1} + \frac{\sigma-1}{(\sigma-1)^2+4t^2} + AL}. \tag{408}$$

We wish to choose $\sigma = 1 + \frac{\delta}{L}$ for some *constant* $\delta > 0$ with $A\delta < 1$. Then the numerator of the last expression equals $1 - \frac{\delta^2}{\delta^2+4L^2t^2} - A\delta$, and this is positive so long as we assume that $4L^2t^2$ is large enough; thus let us *assume* $t = \gamma \ge \frac{c'}{L}$ where c' is a positive constant. Then $1 - \frac{\delta^2}{\delta^2+4L^2t^2} - A\delta > 1 - \frac{\delta^2}{4c'^2} - A\delta$, and we see that for *any* choice of $c' > 0$ we can fix $\delta > 0$ so small that $1 - \frac{\delta^2}{4c'^2} - A\delta > 0$. Note that for any such choice of constants $c', \delta > 0$, and with $\sigma = 1 + \frac{\delta}{L}$, the last expression in (408) is $\gg \frac{1}{L} = \frac{1}{\log q + \log(t+2)}$ (with the implied constant depending on c' and δ). Hence, if we also notice $\frac{c'}{L} < \frac{c'}{\log q}$, we have proved the following: *For every $c' > 0$ there exists some $c > 0$ such that for every $q \ge 3$ and every real primitive Dirichlet character χ modulo q, $L(s,\chi)$ does not have any zero $\rho = \beta + i\gamma$ in the region*

$$\gamma \ge \frac{c'}{\log q} \quad \text{and} \quad \beta \ge 1 - \frac{c}{\log q + \log(t+2)}. \tag{409}$$

Hence to complete the proof of Theorem 11.5 it now only remains to discuss the zeros of $L(s,\chi)$ with $0 \le \gamma < \frac{c'}{\log q}$. For this we will not use Lemma 11.2; instead we consider $\frac{L'(\sigma,\chi)}{L(\sigma,\chi)}$ for $\sigma > 1$. On the one hand we have the crude lower bound (for any $1 < \sigma \le 2$)

$$-\frac{L'(\sigma,\chi)}{L(\sigma,\chi)} = \sum_{n=1}^{\infty} \chi(n)\Lambda(n)n^{-\sigma} \ge -\sum_{n=1}^{\infty} \Lambda(n)n^{-\sigma} = \frac{\zeta'(\sigma)}{\zeta(\sigma)} > -\frac{1}{\sigma-1} + O(1).$$

On the other hand, if $\rho = \beta + i\gamma$ is any zero of $L(s,\chi)$ with $\gamma > 0$ then also $\overline{\rho} = \beta - i\gamma$ is a zero of $L(s,\chi)$ (since χ is real), and hence Lemma 11.3 (where in the sum we only retain

our fixed ρ and $\overline{\rho}$) gives

$$-\frac{L'(\sigma,\chi)}{L(\sigma,\chi)} < -\mathrm{Re}\left(\frac{1}{\sigma-\rho}+\frac{1}{\sigma-\overline{\rho}}\right)+O\big(\log q\big) = -\frac{2(\sigma-\beta)}{(\sigma-\beta)^2+\gamma^2}+O\big(\log q\big). \tag{410}$$

Combining these two inequalities we get

$$\frac{2(\sigma-\beta)}{(\sigma-\beta)^2+\gamma^2} < \frac{1}{\sigma-1}+A\log q,$$

where $A>0$ is an absolute constant. Note that (with an appropriate choice of A) this is also true, trivially, for $\sigma\geq 2$; i.e. the inequality holds for all $\sigma>1$. *Now assume* $0<\gamma<\frac{1}{2}(\sigma-\beta)$. Then

$$\begin{aligned}\frac{8}{5(\sigma-\beta)} < \frac{2(\sigma-\beta)}{(\sigma-\beta)^2+\gamma^2} &< \frac{1}{\sigma-1}+A\log q\\ \Longrightarrow \sigma-\beta &> \frac{8}{\frac{5}{\sigma-1}+5A\log q}\\ \Longrightarrow 1-\beta &> \frac{8}{\frac{5}{\sigma-1}+5A\log q}-(\sigma-1)=\frac{3-5A(\sigma-1)\log q}{\frac{5}{\sigma-1}+5A\log q},\end{aligned}$$

and hence if we take $\sigma=1+\frac{1}{5A\log q}$ then we get

$$1-\beta>\frac{1}{15A\log q}.$$

Recall that this was proved under the assumption that $0<\gamma<\frac{1}{2}(\sigma-\beta)$; in particular it holds if $0<\gamma<\frac{1}{2}(\sigma-1)=\frac{1}{10A\log q}$. Let us also note that a very similar argument applies to prove that if ρ_1,ρ_2 are any *two real* zeros (or a real double zero) of $L(s,\chi)$, then $\min(\rho_1,\rho_2)<1-(15A\log q)^{-1}$. [Details: If ρ_1,ρ_2 are two real zeros (or a double zero) of $L(s,\chi)$ with $0<\rho_1\leq\rho_2<1$, then Lemma 11.3 gives, for any $\sigma>1$:

$$-\frac{L'(\sigma,\chi)}{L(\sigma,\chi)} < -\mathrm{Re}\left(\frac{1}{\sigma-\rho_1}+\frac{1}{\sigma-\rho_2}\right)+O\big(\log q\big) \leq -\frac{2}{\sigma-\rho_1}+O(\log q),$$

which is the same as (410) with $\beta=\rho_1$ and $\gamma=0$. The rest of the proof is as before.]

Hence we have proved: *There exist some constants $c'>0$ and $c>0$ such that for every $q\in\mathbb{Z}^+$ and every real primitive Dirichlet character χ modulo q, $L(s,\chi)$ has at most one zero in the region*

$$0\leq\gamma<\frac{c'}{\log q}\quad\text{and}\quad\beta\geq 1-\frac{c}{\log q}, \tag{411}$$

and if $L(s,\chi)$ has such a zero then this has to be a simple ***real*** *zero.*

Combining our results around (409) and (411) we obtain the statement of the theorem,[21] except that χ is restricted to be primitive. Using this it is now easy to extend also to

[21]Here, of course, we first choose $c'>0$ so that (411) holds; then apply (409) for *this* c'.

the case when χ is a non-primitive, non-principal real character, exactly as in the last paragraph of the proof of Theorem 11.4. □

Remark 11.1. In connection with Definition 9.3 we may now note that *if χ is a primitive character modulo $q \geq 3$ then the non-trivial zeros of $L(s,\chi)$ are exactly those zeros which lie in the* ***open*** *strip $\{0 < \sigma < 1\}$*. Indeed, Theorem 11.4 and Theorem 11.5 together with the fact that $L(\sigma,\chi) \neq 0$ for $\sigma \geq 1$ imply that $L(s,\chi) \neq 0$ for all s on the line $\sigma = 1$. Hence by Corollary 9.7 (especially the symmetry about the line $\sigma = \frac{1}{2}$) it also follows that $L(s,\chi)$ does not have any zero on the line $\sigma = 0$, except for the trivial zero at $s = 0$ if $a = 0$.

We will next discuss results which say that if an "exceptional" zero as in Theorem 11.5 occurs, then at least it occurs only very seldom, i.e. only for very few values of q. The following result is due to Landau 1918:

Theorem 11.6. *There exists an absolute constant $c > 0$ such that for any two distinct real primitive characters χ_1, χ_2 to the moduli $q_1, q_2 \geq 3$ respectively; if the corresponding L-functions have real zeros β_1, β_2, then*

$$\min(\beta_1, \beta_2) < 1 - \frac{c}{\log q_1 q_2}.$$

(Note that the possibility $q_1 = q_2$ is not excluded in the theorem.)

We first prove a uniqueness lemma for primitive characters which ought to have been made clear in §4.6.

Lemma 11.7. *If χ_1 and χ_2 are primitive Dirichlet characters modulo q_1 and q_2, respectively, and if there is some Dirichlet character χ which is induced by both χ_1 and χ_2, then $\chi_1 = \chi_2$ and $q_1 = q_2 = c(\chi)$.*

(In other words: A Dirichlet character cannot be induced by two distinct primitive Dirichlet characters.)

Proof. Note that for given χ, Lemma 4.21 gives the uniqueness of a character $\chi_1 \in X_{q_1}$ inducing χ *under the assumption that* $q_1 = c(\chi)$. Hence it now suffices to prove $q_1 = q_2 = c(\chi)$. In fact it suffices to prove $q_1 = c(\chi)$ since $q_2 = c(\chi)$ will then follow analogously.

Since χ_1 induces χ, we have $q_1 \mid q$ and $\chi(n) = \chi_1(n)$ for all integers n with $(n,q) = 1$. Hence q_1 is a period of $[\chi(n)$ restricted by $(n,q) = 1]$ and thus $c(\chi) \mid q_1$ (by Lemma 4.20). Next note that if n, n' are any integers with $(n, q_1) = (n', q_1) = 1$ and $n \equiv n' \pmod{c(\chi)}$, then there are some $t, t' \in \mathbb{Z}$ such that $(n + tq_1, q) = 1$ and $(n' + t'q_1, q) = 1$ (as in the proof of Lemma 4.21), and then

$$\chi_1(n) = \chi_1(n + tq_1) = \chi(n + tq_1) = \chi(n' + t'q_1) = \chi_1(n' + t'q_1) = \chi_1(n'),$$

where the middle equality holds since $n + tq_1 \equiv n' + t'q_1 \pmod{c(\chi)}$. This shows that $c(\chi)$ is a period of $[\chi_1(n)$ restricted by $(n, q_1) = 1]$. Hence since χ_1 is primitive we have $q_1 \leq c(\chi)$, and this combined with $c(\chi) \mid q_1$ gives $q_1 = c(\chi)$. □

Proof of Theorem 11.6. Note that $\chi_1\chi_2$ is a Dirichlet character modulo q_1q_2. If $\chi_1\chi_2$ is principal then $\chi_1(n)\chi_2(n) = 1$ whenever $(n, q_1q_2) = 1$, and since $\chi_j(n) = \pm 1$ this implies $\chi_1(n) = \chi_2(n)$ whenever $(n, q_1q_2) = 1$, i.e. χ_1 and χ_2 induce the same Dirichlet character modulo q_1q_2. This is impossible by Lemma 11.7, since χ_1 and χ_2 are distinct primitive characters. Hence $\chi_1\chi_2$ *is nonprincipal.*

Now if χ' modulo q' is the unique primitive character which induces $\chi_1\chi_2$ then

$$\Big|\frac{L'(s,\chi_1\chi_2)}{L(s,\chi_1\chi_2)} - \frac{L'(s,\chi')}{L(s,\chi')}\Big| \le \log q_1q_1 \qquad (\sigma \ge 1),$$

exactly as in (404). Also, by Lemma 11.3,

$$-\frac{L'(\sigma,\chi')}{L(\sigma,\chi')} < O\big(\log q'\big), \qquad \forall \sigma \in [1,2].$$

Combining these two we get, since $q' = c(\chi_1\chi_2) \le q_1q_2$:

$$-\frac{L'(\sigma,\chi_1\chi_2)}{L(\sigma,\chi_1\chi_2)} \le O\big(\log q_1q_2\big) + O\big(\log q'\big) \le O\big(\log q_1q_2\big), \qquad \forall \sigma \in [1,2].$$

Lemma 11.3 also gives

$$-\frac{L'(\sigma,\chi_j)}{L(\sigma,\chi_j)} < -\frac{1}{\sigma-\beta_j} + O\big(\log q_j\big), \qquad \forall \sigma \in [1,2].$$

Now consider the expression

$$-\frac{\zeta'(\sigma)}{\zeta(\sigma)} - \frac{L'(\sigma,\chi_1)}{L(\sigma,\chi_1)} - \frac{L'(\sigma,\chi_2)}{L(\sigma,\chi_2)} - \frac{L'(\sigma,\chi_1\chi_2)}{L(\sigma,\chi_1\chi_2)} = \sum_{n=1}^{\infty} \Lambda(n)\big(1+\chi_1(n)\big)\big(1+\chi_2(n)\big)n^{-\sigma} \ge 0.$$

On substituting the previous upper bounds, and also $-\frac{\zeta'(\sigma)}{\zeta(\sigma)} < \frac{1}{\sigma-1} + O(1)$, we get

$$0 < \frac{1}{\sigma-1} - \frac{1}{\sigma-\beta_1} + O\big(\log q_1\big) - \frac{1}{\sigma-\beta_2} + O\big(\log q_2\big) + O\big(\log q_1q_2\big),$$

i.e.

$$\frac{1}{\sigma-\beta_1} + \frac{1}{\sigma-\beta_2} < \frac{1}{\sigma-1} + O\big(\log q_1q_2\big).$$

This implies

$$\frac{2}{\sigma - \min(\beta_1,\beta_2)} < \frac{1}{\sigma-1} + A\log q_1q_2,$$

where A is an absolute constant. Exactly as in the argument between (393) and (394) this is seen to imply

$$1 - \min(\beta_1,\beta_2) \gg \frac{1}{\log q_1q_2}.$$

□

Corollary 11.8. *There exists an absolute constant $c > 0$ such that for any $q \geq 3$, there is at most **one** real nonprincipal character $\chi \in X_q$ for which $L(s,\chi)$ has a real zero β with $\beta \geq 1 - \frac{c}{\log q}$.*

Proof. Let c be as in Theorem 11.6. We may assume $c < 1$. Now assume that χ_1 and χ_2 are two nonprincipal Dirichlet characters to the same modulus $q \geq 3$, and assume that the corresponding L-functions have real zeros β_1, β_2, respectively. Let χ_j' modulo q_j' be the primitive Dirichlet character which induces χ_j ($j = 1, 2$). Recall that all the zeros of $L(s,\chi_j)$ additional to those of $L(s,\chi_j')$ have real part equal to 0, since $L(s,\chi_j) = L(s,\chi_j')\prod_{p|q}(1-\chi_j'(p)p^{-s})$. Hence if $\beta_1, \beta_2 > 0$ then β_j is a zero of $L(s,\chi_j')$ for $j = 1, 2$, and Theorem 11.6 gives $\min(\beta_1,\beta_2) < 1 - \frac{c}{\log q_1' q_2'} \leq 1 - \frac{c}{2\log q}$. This is also true, trivially, if either β_1 or β_2 is ≤ 0 (since $c < 1$ and $\log q > 1$). This concludes the proof. □

Corollary 11.9. *There is an absolute constant $c > 0$ such that if $q_1 < q_2 < q_3 < \dots$ is the (possible) sequence of positive integers q with the property that there is a real primitive $\chi \pmod q$ for which $L(s,\chi)$ has a real zero β satisfying $\beta > 1 - \frac{c}{\log q}$, then $q_{j+1} > q_j^2$ for all $j = 1, 2, \dots$.*

Proof. Let us write c' for the absolute constant in Theorem 11.6. Now if $q_1 < q_2 < \dots$ is the sequence as above, so that for each j there is some real zero $\beta_j > 1 - \frac{c}{\log q_j}$ of $L(s,\chi)$ for some $\chi \in X_{q_j}$, then Theorem 11.6 gives, for all $j \geq 1$,

$$1 - \frac{c'}{\log q_j q_{j+1}} > \min(\beta_j, \beta_{j+1}) > \min\Big(1 - \frac{c}{\log q_j}, 1 - \frac{c}{\log q_{j+1}}\Big) = 1 - \frac{c}{\log q_j};$$

hence

$$\log q_j q_{j+1} > \frac{c'}{c}\log q_j,$$

and hence if we made the choice of c as $c = \frac{1}{3}c'$ then $q_j q_{j+1} > q_j^3$, which gives the result. □

Corollary 11.10. *There is an absolute constant $c > 0$ such that the following holds. For any $z \geq 3$, there exists at most one real primitive χ to a modulus $q \leq z$ for which $L(s,\chi)$ has a real zero β satisfying $\beta > 1 - \frac{c}{\log z}$.*

Proof. Again write c' for the absolute constant in Theorem 11.6. Let $z \geq 3$ and assume that, contrary to the claim, there exist two distinct real primitive characters $\chi_1 \in X_{q_1}$, $\chi_2 \in X_{q_2}$, $q_1, q_2 \leq z$ such that $L(s,\chi_j)$ has a real zero β_j satisfying $\beta_j > 1 - \frac{c}{\log z}$ for both $j = 1, 2$. Then Theorem 11.6 gives $1 - \frac{c}{\log z} < \min(\beta_1,\beta_2) < 1 - \frac{c'}{\log q_1 q_2} \leq 1 - \frac{c'}{2\log z}$, and this is a contradiction if we have made the choice $c = \frac{1}{2}c'$. □

The only obvious general upper bound for a real zero of an L-function corresponding to a real primitive χ is that which can be derived from the class-number formula relating $L(1,\chi)$ and a class number $h(d)$, and using $h(d) \geq 1$:

Proposition 11.11. *There is an absolute constant $c > 0$ such that for any real zero β of an L-function corresponding to a real nonprincipal Dirichlet character χ modulo q, we have*

$$\beta < 1 - \frac{c}{q^{\frac{1}{2}}(\log q)^{1+a}}, \tag{412}$$

where, as usual, $a = 0$ if $\chi(-1) = 1$ and $a = 1$ if $\chi(-1) = -1$.

Proof. We first assume that χ is a primitive character modulo q ($q \geq 3$). Then $\chi = \left(\frac{d}{\cdot}\right)$ for some fundamental discriminant $d = \pm q$, by Theorem 4.35, and by Dirichlet's class number formula, Theorem 5.4, we have

$$L(1,\chi) = \begin{cases} \frac{2\pi}{w\sqrt{|d|}} h(d) & \text{if } d < 0, \\ \frac{\log \varepsilon_d}{\sqrt{d}} h(d) & \text{if } d > 0, \end{cases} \qquad (w \in \{2,4,6\}).$$

This immediately implies

$$L(1,\chi) \gg |d|^{-\frac{1}{2}} \begin{Bmatrix} \log|d| & \text{if } d > 0 \\ 1 & \text{if } d < 0 \end{Bmatrix},$$

since $h(d) \geq 1$, and if $d > 0$ then $\varepsilon_d = \frac{1}{2}(x + y\sqrt{d}) \geq \frac{1}{2}(1+\sqrt{d})$ so that $\log \varepsilon_d \geq \log \frac{1}{2}(1+\sqrt{d}) \gg \log d$ (recall that if $d > 0$ then $d \geq 5$, since d is a fundamental discriminant). Recall that $d > 0$ holds if and only if $\chi(-1) = 1$ (Lemma 4.37), i.e. if and only if $a = 0$. Hence the above lower bound can be expressed as

$$L(1,\chi) \gg \frac{(\log q)^{1-a}}{q^{\frac{1}{2}}}. \tag{413}$$

Now let σ be any real number in the interval $1 - \frac{1}{\log q} \leq \sigma < 1$. Then by the mean value theorem we have

$$L(1,\chi) - L(\sigma,\chi) = (1-\sigma)L'(\xi,\chi)$$

for some $\xi \in (\sigma, 1)$, and since $L'(\xi,\chi) = O\big(\log^2 q\big)$ by Problem 10.1(c), we conclude that

$$L(\sigma,\chi) \geq L(1,\chi) - A(1-\sigma)\log^2 q,$$

where $A > 0$ is an absolute constant. Combining this with (413) we see that there is an absolute constant $c > 0$ such that

$$1 - \frac{c}{q^{\frac{1}{2}}(\log q)^{1+a}} \leq \sigma \leq 1 \Longrightarrow L(\sigma,\chi) > 0.$$

This proves the proposition. □

11.3. ***Alternative method.** There is an alternative method, due to Landau (1924), of obtaining zero-free regions for L-functions, which does not use an infinite product formula for the L-function, and thus in which the analytic character of the L-function for $\sigma \le 0$ need not be known.

We will here explain this method, borrowing from Titchmarsh [70, Ch. III].

To get started we prove the following; the *Borel-Carathéodory Theorem:*

Theorem 11.12. *Suppose that $f(z)$ is analytic in the open disc $|z - z_0| < R$ and has the Taylor expansion*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n. \tag{414}$$

Furthermore suppose that

$$\operatorname{Re} f(z) \le U \qquad \text{for all } z \text{ with } |z - z_0| < R. \tag{415}$$

Then

$$|c_n| \le \frac{2(U - \operatorname{Re} c_0)}{R^n} \qquad \forall n \ge 1, \tag{416}$$

and, in $|z - z_0| \le r < R$ we have

$$\big|f(z) - f(z_0)\big| \le \frac{2r}{R - r}\big(U - \operatorname{Re} f(z_0)\big), \tag{417}$$

$$\Big|\frac{f^{(m)}(z)}{m!}\Big| \le \frac{2R}{(R - r)^{m+1}}\big(U - \operatorname{Re} f(z_0)\big), \qquad \forall m \ge 1. \tag{418}$$

(Note that below we spend some extra work to obtain the precise constants "2" in (416), (417), (418). This constant is in fact the best possible, as may be seen by considering the function $f(z) = z/(1 + z)$, for which $\operatorname{Re} f(z) < \frac{1}{2}$ throughout $|z| < 1$. However, for the application in the present section these precise constants are not important.)

Proof. In fact we have already gone through most of the steps needed for the proof of this theorem, when we proved Lemma 8.2. We give a quick review ($f(z)$ now takes the role of $g(z)$ in the proof of Lemma 8.2): If we write $c_n = a_n + ib_n$ ($a_n, b_n \in \mathbb{R}$) then from the Taylor expansion (414) we get

$$\operatorname{Re} f(re^{i\theta}) = \sum_{n=0}^{\infty} a_n r^n \cos n\theta - \sum_{n=0}^{\infty} b_n r^n \sin n\theta, \qquad (0 \le r < R,\ \theta \in \mathbb{R}),$$

where the series are uniformly absolutely convergent with respect to θ for fixed r. Hence by basic Fourier analysis we get, for $n \geq 0$,

$$(1+\delta_{n0})\pi a_n r^n = \int_0^{2\pi} \big(\mathrm{Re}\, f(re^{i\theta})\big) \cos n\theta \, d\theta;$$
$$-(1-\delta_{n0})\pi b_n r^n = \int_0^{2\pi} \big(\mathrm{Re}\, f(re^{i\theta})\big) \sin n\theta \, d\theta.$$

Hence for any $n \geq 1$ we have

$$\left.\begin{matrix}\pi|a_n|r^n \\ \pi|b_n|r^n\end{matrix}\right\} \leq \int_0^{2\pi} \big|\mathrm{Re}\, f(re^{i\theta})\big| \, d\theta = -2\pi a_0 + \int_0^{2\pi} \Big(\big|\mathrm{Re}\, f(re^{i\theta})\big| + \mathrm{Re}\, f(re^{i\theta})\Big) \, d\theta$$
$$\leq -2\pi \mathrm{Re}\, c_0 + 4\pi U^+,$$

where $U^+ := \max(0, U)$, and thus

$$|c_n| = \sqrt{a_n^2 + b_n^2} \leq \frac{2\sqrt{2}(2U^+ - \mathrm{Re}\, c_0)}{r^n}.$$

We may improve the constants slightly as follows. If we set (for $n \geq 1$ fixed) $\omega = \arg(c_n)$ then $|c_n| = c_n e^{-i\omega} = \mathrm{Re}\big(c_n e^{-i\omega}\big) = a_n \cos\omega + b_n \sin\omega$ and the above formulas for a_n, b_n imply

$$\pi|c_n|r^n = \pi r^n (a_n \cos\omega + b_n \sin\omega) = \int_0^{2\pi} \big(\mathrm{Re}\, f(re^{i\theta})\big)\big(\cos\omega \cos n\theta - \sin\omega \sin n\theta\big) \, d\theta$$
$$= \int_0^{2\pi} \big(\mathrm{Re}\, f(re^{i\theta})\big) \cos(\omega + n\theta) \, d\theta,$$

and hence as above

$$\pi|c_n|r^n \leq -2\pi \mathrm{Re}\, c_0 + 4\pi U^+.$$

Finally we may apply this last inequality to the function $f_1(z) := f(z) - \alpha$ where α is any real constant; note that $f_1(z)$ satisfies the bound (415) with $U - \alpha$ in place of U, and f_1 has the same Taylor coefficients c_n as f except that c_0 is replaced with $c_0 - \alpha$; hence we obtain

$$\pi|c_n|r^n \leq -2\pi \mathrm{Re}\,(c_0 - \alpha) + 4\pi(U-\alpha)^+ = 2\pi\big(2(U-\alpha)^+ + \alpha - \mathrm{Re}\, c_0\big).$$

To get the best bound possible here we take $\alpha = U$, and thus conclude

$$\pi|c_n|r^n \leq 2\pi\big(U - \mathrm{Re}\, c_0\big).$$

This is true for all $r \in [0, R)$, and letting $r \to R^-$ we obtain (416). Now (417) follows from (416) as follows: Let us write $\beta_0 = 2(U - \mathrm{Re}\, c_0)$ for short. Then

$$\big|f(z) - f(z_0)\big| = \Big|\sum_{n=1}^{\infty} c_n (z - z_0)^n\Big| \leq \sum_{n=1}^{\infty} |c_n| r^n \leq \beta_0 \sum_{n=1}^{\infty} (r/R)^n = \frac{2r\beta_0}{R - r},$$

i.e. (417) holds. Similarly (418) follows from (416) as follows:

$$\begin{aligned}\big|f^{(m)}(z)\big| &= \Big|\sum_{n=m}^{\infty} n(n-1)\cdots(n-m+1)c_n(z-z_0)^{n-m}\Big| \\ &\le \beta_0 \sum_{n=m}^{\infty} n(n-1)\cdots(n-m+1)\frac{r^{n-m}}{R^n} \\ &= \beta_0\Big(\frac{d}{dr}\Big)^m \sum_{n=0}^{\infty}(r/R)^n = \beta_0\Big(\frac{d}{dr}\Big)^m \frac{R}{R-r} = \beta_0\frac{R\cdot m!}{(R-r)^{m+1}}.\end{aligned}$$

□

Landau's method to obtain zero-free regions depends on the following two lemmas.

Lemma 11.13. *If $f(s)$ is analytic in the disc $|s-s_0|\le r$ and $f(s_0)\neq 0$, then*

$$\Big|\frac{f'(s)}{f(s)} - \sum_{\rho}\frac{1}{s-\rho}\Big| \ll \frac{\log M}{r} \qquad \text{for all } s \text{ in the disc } |s-s_0|\le \tfrac14 r,$$

where the implied constant is absolute, ρ runs through the zeros of $f(s)$ such that $|\rho-s_0|\le \frac12 r$, and

$$M := \sup\Big\{\Big|\frac{f(s)}{f(s_0)}\Big| \;:\; s\in\mathbb{C},\ |s-s_0|\le r\Big\}.$$

Proof. The function $g(s) = f(s)\prod_\rho (s-\rho)^{-1}$ is analytic for $|s-s_0|\le r$, and not zero for $|s-s_0|\le \frac12 r$. If s lies on the circle $|s-s_0|=r$ then each ρ satisfies $|s-\rho|\ge \frac12 r \ge |s_0-\rho|$, and hence

$$\Big|\frac{g(s)}{g(s_0)}\Big| = \Big|\frac{f(s)}{f(s_0)}\prod_\rho \frac{s_0-\rho}{s-\rho}\Big| \le \Big|\frac{f(s)}{f(s_0)}\Big| \le M.$$

This inequality therefore holds for s inside the circle also, by the maximum principle. Now consider the function

$$h(s) = \log\frac{g(s)}{g(s_0)},$$

where the branch of the logarithm is chosen so that $h(s_0)=0$. This function $h(s)$ is analytic for $|s-s_0|\le \frac12 r$, and in this disc satisfies

$$h(s_0)=0, \qquad \operatorname{Re} h(s)\le \log M.$$

Hence by the Borel-Carathéodory Theorem (cf. (417) in Theorem 11.12), for any s with $|s-s_0|\le \frac14 r$ we have

$$\big|h'(s)\big| \le \frac{2\cdot\frac12 r}{(\frac12 r-\frac14 r)^2}\log M = \frac{16\log M}{r}.$$

This gives the result stated. □

Lemma 11.14. *Suppose that $f(s)$ is as in the previous lemma, and also that $f(s)$ has no zeros in the right-hand half of the circle $|s-s_0|\le r$. Then*

$$-\mathrm{Re}\,\frac{f'(s_0)}{f(s_0)} \le O\Big(\frac{\log M}{r}\Big).$$

If furthermore $f(s)$ has a zero ρ_0 between $s_0-\frac{1}{2}r$ and s_0, then

$$-\mathrm{Re}\,\frac{f'(s_0)}{f(s_0)} \le O\Big(\frac{\log M}{r}\Big) - \frac{1}{s_0-\rho_0}.$$

The implied constants in both big-O's are absolute.

Proof. Lemma 11.13 gives

$$-\mathrm{Re}\,\frac{f'(s_0)}{f(s_0)} \le O\Big(\frac{\log M}{r}\Big) - \sum_\rho \mathrm{Re}\,\frac{1}{s_0-\rho},$$

where ρ runs through the zeros of $f(s)$ such that $|\rho-s_0|\le\frac{1}{2}r$. By assumption we have $\mathrm{Re}\,(s_0-\rho)\ge 0$ for each ρ and thus $\mathrm{Re}\,\big(\frac{1}{s_0-\rho}\big)\ge 0$. Now both claims follow at once. □

We can now prove the following general theorem, which we will later apply with special forms of the functions $\theta(t)$ and $\phi(t)$.

Theorem 11.15. *Assume that*

$$\zeta(s) = O\big(\psi(t)\big) \qquad \text{for all } s \text{ with } t\ge 1 \text{ and } 1-\theta(t)\le\sigma\le 2,$$

where $\psi:[1,\infty)\to[e,\infty)$ is an increasing function and $\theta:[1,\infty)\to(0,1]$ is a decreasing function satisfying

$$\log\big(\theta(t)^{-1}\big) \ll \log\psi(t), \qquad \forall t\ge 1. \tag{419}$$

Then there is a constant $c>0$ such that $\zeta(s)$ has no zeros in the region

$$t\ge 0, \qquad \sigma \ge 1 - c\frac{\theta(2t+1)}{\log\psi(2t+1)}. \tag{420}$$

Proof. Let $\beta+i\gamma$ be a non-trivial zero of $\zeta(s)$ with $\gamma\ge 0$. Then from Remark 10.1 we know that $\gamma>6$. Take σ_0 arbitrary with $1<\sigma_0\le 2$ and set $s_0=\sigma_0+i\gamma$ and $s_0'=\sigma_0+2i\gamma$; then recall that we have the inequality (cf. (389)):

$$3\Big(-\frac{\zeta'(\sigma_0)}{\zeta(\sigma_0)}\Big) + 4\Big(-\mathrm{Re}\,\frac{\zeta'(s_0)}{\zeta(s_0)}\Big) + \Big(-\mathrm{Re}\,\frac{\zeta'(s_0')}{\zeta(s_0')}\Big) \ge 0 \qquad \forall\sigma_0>1. \tag{421}$$

We have for all (cf. Problem 2.1(a))

$$\zeta(\sigma_0+it)^{-1} = \sum_{n=1}^\infty \mu(n)n^{-\sigma_0-it} = O\Big(\sum_{n=1}^\infty n^{-\sigma_0}\Big) = O\big(\zeta(\sigma_0)\big) = O\Big(\frac{1}{\sigma_0-1}\Big).$$

We set $r = \theta(2\gamma+1) \le 1$. Then by the assumptions, since θ is decreasing and ψ is increasing, all s with $1 - r \le \sigma \le 2$ and $1 \le t \le 2\gamma + 1$ satisfy $\zeta(s) = O\big(\psi(t)\big) = O\big(\psi(2\gamma+1)\big)$. This bound trivially also holds when $2 \le \sigma \le 3$, since then $\zeta(s) = O(1)$, while $\psi(t) \ge e$. Using the last two bounds it follows that for all s with $|s - s_0| \le r$ and all s' with $|s' - s_0'| \le r$ we have

$$\Big|\frac{\zeta(s)}{\zeta(s_0)}\Big| = O\Big(\frac{\psi(2\gamma+1)}{\sigma_0 - 1}\Big) \qquad \text{and} \qquad \Big|\frac{\zeta(s')}{\zeta(s_0')}\Big| = O\Big(\frac{\psi(2\gamma+1)}{\sigma_0 - 1}\Big).$$

Applying Lemma 11.14 to these two discs we obtain (since $\frac{\psi(2\gamma+1)}{\sigma_0-1} \ge e$)

$$-\mathrm{Re}\,\frac{\zeta'(s_0')}{\zeta(s_0')} \le O\left(\frac{1}{\theta(2\gamma+1)}\log\Big(\frac{\psi(2\gamma+1)}{\sigma_0-1}\Big)\right) \tag{422}$$

and, if $\beta > \sigma_0 - \frac{1}{2}r$,

$$-\mathrm{Re}\,\frac{\zeta'(s_0)}{\zeta(s_0)} \le O\left(\frac{1}{\theta(2\gamma+1)}\log\Big(\frac{\psi(2\gamma+1)}{\sigma_0-1}\Big)\right) - \frac{1}{\sigma_0 - \beta}. \tag{423}$$

Also recall that

$$-\frac{\zeta'(\sigma_0)}{\zeta(\sigma_0)} < \frac{1}{\sigma_0 - 1} + O(1). \tag{424}$$

Using the last three inequalities, (422), (423), (424), together with (421), we obtain

$$\frac{3}{\sigma_0 - 1} + O(1) + O\left(\frac{1}{\theta(2\gamma+1)}\log\Big(\frac{\psi(2\gamma+1)}{\sigma_0-1}\Big)\right) - \frac{4}{\sigma_0 - \beta} \ge 0,$$

i.e. there is a constant A (i.e. a number independent of γ), which we choose to satisfy $A \ge 1$, such that

$$\frac{4}{\sigma_0 - \beta} \le \frac{3}{\sigma_0 - 1} + \frac{A}{\theta(2\gamma+1)}\log\Big(\frac{\psi(2\gamma+1)}{\sigma_0-1}\Big).$$

This implies (we here solve for $1-\beta$, and use the abbreviations $\psi := \psi(2\gamma+1)$, $\theta := \theta(2\gamma+1)$, $x := \sigma_0 - 1$)

$$\sigma_0 - \beta \ge \frac{4}{\frac{3}{x} + \frac{A}{\theta}\log\big(\frac{\psi}{x}\big)} \Longrightarrow 1 - \beta \ge \frac{4}{\frac{3}{x} + \frac{A}{\theta}\log\big(\frac{\psi}{x}\big)} - x = \frac{1 - \frac{Ax}{\theta}\log\big(\frac{\psi}{x}\big)}{\frac{3}{x} + \frac{A}{\theta}\log\big(\frac{\psi}{x}\big)}. \tag{425}$$

Let us fix a constant $0 < a < \frac{1}{3}$ (i.e. a number independent of γ) so small that

$$Aa\big(2 + |\log a|\big) < \tfrac{1}{4} \qquad \text{and} \qquad Aa\Big(\sup_{t\ge 1}\frac{\log(\theta(t)^{-1})}{\log\psi(t)}\Big) < \tfrac{1}{4}. \tag{426}$$

(Note that the supremum in the last inequality is $< \infty$, because of our assumption (419).) We now choose $x = a\frac{\theta}{\log\psi}$. Note that $0 < x < \frac{1}{3}$, so that $\sigma_0 = 1 + x$ lies in the interval $1 < \sigma_0 < \frac{4}{3} < 2$ as it should. We also have

$$\frac{Ax}{\theta}\log\big(\frac{\psi}{x}\big) = Aa\frac{\log\big(\psi/x\big)}{\log\psi} = Aa\Big(1 + \frac{\log\log\psi}{\log\psi} + \frac{\log(a^{-1})}{\log\psi}\Big) + Aa\frac{\log(\theta^{-1})}{\log\psi}$$
$$\le Aa\big(2 + |\log a|\big) + \tfrac{1}{4} < \tfrac{1}{2},$$

by (426). Hence the numerator in the last expression in (425) is $> \frac{1}{2}$, and the denominator is

$$\frac{3}{x} + \frac{A}{\theta}\log\big(\frac{\psi}{x}\big) < \frac{3}{x} + \frac{1}{2x} = \frac{7}{2}x.$$

Hence (425) implies

$$1 - \beta \ge \frac{1}{7}x = \frac{a}{7}\frac{\theta}{\log\psi} = \frac{a}{7}\frac{\theta(2\gamma+1)}{\log\psi(2\gamma+1)},$$

i.e. we have proved the desired bound (420). Recall that this was derived under the assumption that $\beta > \sigma_0 - \frac{1}{2}r$; but in the other case we have

$$\beta \le \sigma_0 - \tfrac{1}{2}r = 1 + x - \tfrac{1}{2}\theta = 1 + a\frac{\theta}{\log\psi} - \tfrac{1}{2}\theta < 1 + \big(\tfrac{1}{3} - \tfrac{1}{2}\big)\theta \le 1 - \frac{1}{6}\frac{\theta}{\log\psi},$$

i.e. (420) holds in this case as well. □

In particular we can take $\theta(t) = \frac{1}{2}$, $\psi(t) = t + 2$ in Theorem 11.15 (this is ok by (281) in Proposition 7.3); then the conclusion is that there is a constant $c > 0$ such that $\zeta(s)$ has no zeros in the region $t \ge 0$, $\sigma \ge 1 - \frac{c}{\log(t+2)}$, i.e. we have a new proof of Theorem 11.1.

As another example we state the sharpest known zero-free region today, which was obtained independently by Vinogradov and Korobov in 1958: There is an absolute constant $c > 0$ such that $\zeta(s)$ does not have any zeros in the region

$$t \ge 3, \qquad \sigma \ge 1 - \frac{c}{(\log t)^{\frac{2}{3}}(\log\log t)^{\frac{1}{3}}}. \tag{427}$$

This can be deduced from the following bound:

$$\zeta(s) = O\Big(t^{100\max(0,1-\sigma)^{\frac{3}{2}}}\log^{\frac{2}{3}} t\Big) \qquad \text{for all } s \text{ with } \frac{1}{2} \le \sigma \le 2,\ t \ge 2, \tag{428}$$

with an absolute implied constant. This bound follows from the (independent) work of Vinogradov and Korobov (1958) on bounding exponential sums of the form $\sum_{a<n\le b} n^{-it} =$

$\sum_{a<n\le b} e^{-it\log n}$ and similar sums; cf. Richert [57].[22] It follows from (428) that we may take

$$\theta(t) = \Big(\frac{\log\log(t+10)}{100\log t}\Big)^{\frac{2}{3}}; \qquad \psi(t) = 10\log^2(t+10)$$

in Theorem 11.15, and this gives the zero-free region (427).

Finally let us remark that it is known that $\zeta(1+it)$ is unbounded as $t\to\infty$ (cf., e.g. Titchmarsh [70, Thm. 8.9(A)]), and hence it is *impossible* to obtain from Theorem 11.15 a full vertical strip $\sigma \ge 1-\varepsilon$ as a zero-free region (since the assumptions of the theorem necessarily imply $\psi(t)\to\infty$ as $t\to\infty$).

11.4. **Problems.**

* *Problem* 11.1. Can you prove Theorem 11.15 using the infinite product expansion of $\zeta(s)$ instead of Lemma 11.13, Lemma 11.14?

[22] Richert [57] only formulates this result for $\frac{1}{2} \le \sigma \le 1$, but it is a simple task to extend it further to also cover $1 \le \sigma \le 2$.

12. The numbers $N(T)$ and $N(T,\chi)$

12.1. The number $N(T)$.

Definition 12.1. For each $T > 0$, we let $N(T)$ be the number of zeros of $\zeta(s)$ in the rectangle $0 < \sigma < 1$, $0 < t < T$.

Recall that Riemann in his memoir made the following conjecture:

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T) \qquad \text{as } T \to \infty. \tag{429}$$

This asymptotic relation was proved by von Mangoldt (1905), and we will give a proof in this section.

Definition 12.2. For $T > 0$, if T is not the ordinate (viz. the imaginary part) of a zero of $\zeta(s)$, we set

$$S(T) = \frac{1}{\pi}\Delta_L \arg\zeta(s), \tag{430}$$

where L denotes the line from 2 to $2+iT$ and then to $\frac{1}{2}+iT$. If T is the ordinate of a zero then we define $S(T) = \lim_{T'\to T^-} S(T')$.[23]

We remark that $S(T)$ may equivalently be defined as

$$S(T) = \frac{1}{\pi}\arg\zeta(\frac{1}{2}+iT),$$

provided that the argument is defined by continuous variation along L, or, equivalently, by continuous horizontal movement from $+\infty+iT$ to $\frac{1}{2}+iT$, *starting with the value* 0, and *"going below" any zero on the line* $t = T$. [The two definitions are equivalent because[24] $\big|\zeta(s)-1\big| \le \sum_{n=2}^{\infty} n^{-\sigma} = \frac{1}{6}\pi^2 - 1 < 1$ for all s with $\sigma \ge 2$, and also $\big|\zeta(s)-1\big| \le \zeta(\sigma)-1 \to 0$ as $\sigma \to \infty$.]

Theorem 12.1. *We have*

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + \frac{7}{8} + S(T) + O(T^{-1}), \qquad \forall T \ge 1. \tag{431}$$

Proof. Let $T \ge 1$ be given. We may assume that T does not coincide with the ordinate of a zero of $\zeta(s)$, since the remaining case will then follow by taking a limit $T' \to T^-$ (using the fact that both $N(T)$ and $S(T)$ are left continuous by definition).

[23]Thus we define $S(T)$ so as to be left continuous. Note that this differs from e.g. Titchmarsh [70] who defines $S(T)$ to be right continuous. Correspondingly Titchmarsh also defines $N(T)$ to be right continuous, whereas our $N(T)$ is left continuous.

[24]Here we use the fact that $\zeta(2) = \frac{\pi^2}{6}$, cf. Problem 9.4(d). However all we need here is that $\zeta(2) < 2$, and this can be proved more easily by an integral estimate: $\zeta(2) = 1 + \frac{1}{4} + \sum_{n=3}^{\infty} n^{-2} \le 1 + \frac{1}{4} + \int_2^{\infty} x^{-2}\,dx = \frac{7}{4}$.

Recall that the zeros of $\zeta(s)$ in the strip $0 < \sigma < 1$ are identical in position and order of multiplicity with the zeros of $\xi(s)$, and that $\xi(s)$ does not have any zeros outside this strip. Cf. Corollary 9.3, Remark 9.2. Also recall that $\xi(s)$ does not have any zeros for $s \in \mathbb{R}$ (cf. Remark 10.1). Hence by the argument principle we have

$$N(T) = \frac{1}{2\pi}\Delta_R \arg \xi(s),$$

where R is the rectangle in the s plane with vertices at

$$2, \quad 2+iT, \quad -1+iT, \quad -1,$$

described in the positive sense. Note that for s real $\xi(s)$ is real and never zero, and $\xi(s) > 0$ for large real s; hence $\xi(s) > 0$ for all real s. Hence there is no change in $\arg \xi(s)$ as s moves along the base of the rectangle R. Further, the change of $\arg \xi(s)$ as s moves from $\frac{1}{2}+iT$ to $-1+iT$ and then to -1 is equal to the change as s moves from 2 to $2+iT$ and then to $\frac{1}{2}+iT$, since

$$\xi(\sigma+it) = \xi(1-\sigma-it) = \overline{\xi(1-\sigma+it)}.$$

Hence

$$N(T) = \frac{1}{\pi}\Delta_L \arg \xi(s),$$

where L denotes the line from 2 to $2+iT$ and then to $\frac{1}{2}+iT$ (just as in the definition of $S(T)$, Definition 12.2).

Now recall that, by definition,

$$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = (s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s+1)\zeta(s),$$

and hence

$$N(T) = \frac{1}{\pi}\Big(\Delta_L \arg(s-1) + \Delta_L \arg(\pi^{-\frac{1}{2}s}) + \Delta_L \arg(\Gamma(\tfrac{1}{2}s+1)) + \Delta_L \arg(\zeta(s))\Big).$$

We have

$$\Delta_L \arg(s-1) = \arg(iT - \tfrac{1}{2}) = \tfrac{1}{2}\pi + O(T^{-1})$$

and

$$\Delta_L \arg(\pi^{-\frac{1}{2}s}) = \Delta_L(-\tfrac{1}{2}t\log\pi) = -\tfrac{1}{2}T\log\pi.$$

Next to compute $\Delta_L \arg\Gamma(\frac{1}{2}s+1)$ we recall that we have defined $\log\Gamma(z)$ as an analytic function in $z \in \mathbb{C}\setminus(-\infty, 0]$ (cf. Definition 8.3), and the definition immediately implies $\log\Gamma(z) \in \mathbb{R}$ for all $z > 0$. Hence

$$\Delta_L \arg\Gamma(\tfrac{1}{2}s+1) = \text{Im } \log\Gamma(\tfrac{5}{4}+\tfrac{1}{2}iT).$$

Applying here Stirling's formula (Theorem 8.17) in the version proved in Problem 8.6 (with $\alpha = \frac{5}{4}$) we obtain

$$\begin{aligned}\Delta_L \arg \Gamma(\tfrac{1}{2}s+1) &= \operatorname{Im} \log \Gamma(\tfrac{5}{4}+\tfrac{1}{2}iT) \\ &= \operatorname{Im}\Big[(\tfrac{3}{4}+\tfrac{1}{2}iT)\log\big(\tfrac{1}{2}iT\big) - \tfrac{1}{2}iT + \log\sqrt{2\pi} + O\big(T^{-1}\big)\Big] \\ &= \operatorname{Im}\Big[(\tfrac{3}{4}+\tfrac{1}{2}iT)(\log(\tfrac{1}{2}T)+\tfrac{1}{2}\pi i)\Big] - \tfrac{1}{2}T + O\big(T^{-1}\big) \\ &= \tfrac{1}{2}T\log(\tfrac{1}{2}T) - \tfrac{1}{2}T + \tfrac{3}{8}\pi + O\big(T^{-1}\big).\end{aligned}$$

Finally we have by definition

$$\Delta_L \arg \zeta(s) = \pi S(T).$$

Collecting these facts we obtain

$$\begin{aligned}N(T) = \frac{1}{\pi}\Delta_L \arg \xi(s) &= \tfrac{1}{2} - \tfrac{1}{2\pi}T\log\pi + \tfrac{1}{2\pi}T\log(\tfrac{1}{2}T) - \tfrac{1}{2\pi}T + \tfrac{3}{8} + O(T^{-1}) + S(T) \\ &= \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + \frac{7}{8} + S(T) + O(T^{-1}).\end{aligned}$$

□

Now the asymptotic relation conjectured by Riemann, (429), follows from Theorem 12.1 together with the following bound.

Theorem 12.2.

$$S(T) = O(\log T) \qquad \text{as } T \to \infty. \tag{432}$$

We will give two proofs of this result. The first proof is due to Backlund (1914,1918), and basically only uses the fact that we have a bound on the growth of $\zeta(s)$ in the critical strip, cf. Proposition 7.3.

First proof of Theorem 12.2. (We follow Ingham [34, p. 69].) We may assume from start that $T \geq 10$. We may also assume that T is not the ordinate of a zero of $\zeta(s)$, since the remaining case will then follow by taking a limit $T' \to T^-$. Let m be the number (necessarily finite, as will appear) of distinct points s' on L (excluding end-points) at which $\operatorname{Re}\zeta(s') = 0$. Then

$$S(T) = \frac{1}{\pi}\Delta_L \arg\zeta(s) \leq m+1, \tag{433}$$

for, when s describes one of the $m+1$ pieces into which L is divided by the points s', $\arg\zeta(s)$ cannot vary by more than π since $\operatorname{Re}\zeta(s)$ does not change sign. Now no point s' can lie on the line segment between 2 and $2+iT$, since, as we have seen, $\big|\operatorname{Re}\zeta(s)-1\big| < \big|\zeta(s)-1\big| \leq \frac{1}{6}\pi^2 - 1 < 1$ when $\sigma = 2$. Thus m is the number of distinct points σ of the

interval $\frac{1}{2} < \sigma < 2$ at which $\operatorname{Re}\zeta(\sigma + iT) = 0$, and this is the number of distinct zeros of the function

$$g(s) = \tfrac{1}{2}\big(\zeta(s+iT) + \zeta(s-iT)\big)$$

on the segment $\frac{1}{2} < s < 2$ of the real axis. (This is so because for real s we have $g(s) = \frac{1}{2}\big(\zeta(s+iT) + \zeta(s-iT)\big) = \frac{1}{2}\big(\zeta(s+iT) + \overline{\zeta(s+iT)}\big) = \operatorname{Re}\zeta(s+iT)$.)

Since $g(s)$ is analytic, except at $s = 1 \pm iT$, and not identically zero (since $g(2) > 0$ as seen above), the number m is finite, and we obtain an upper bound for m by using Jensen's formula (Proposition 8.3) for the disc with center $s = 2$ and some radius $R \in [1.6, 1.9]$ chosen so that $g(s)$ does not have any zeros on the boundary $|s-2| = R$. [25] Note that g is analytic in a neighbourhood of our disc since $T \geq 10$. Hence, if $s_1, \dots, s_n$ are all the zeros of $g(s)$ in our disc $|s-2| < R$:

$$\sum_{j=1}^{n} \log\frac{R}{|s_j|} = \frac{1}{2\pi}\int_0^{2\pi} \log\big|g(2 + Re^{i\theta})\big|\, d\theta - \log\big|g(2)\big|.$$

Note that $\log\frac{R}{|s_j|} \geq 0$ for each s_j, and if $\frac{1}{2} < s_j < 2$ then $\log\frac{R}{|s_j|} > \log\frac{16}{15} > 0$; hence the above gives

$$m < \big(\log\tfrac{16}{15}\big)^{-1}\Big(\frac{1}{2\pi}\int_0^{2\pi} \log\big|g(2 + Re^{i\theta})\big|\, d\theta - \log\big|g(2)\big|\Big). \tag{434}$$

But by (281) in Proposition 7.3 (with $\delta = 0.1$) we have

$$\big|g(s)\big| \leq \tfrac{1}{2}\big|\zeta(s+iT)\big| + \tfrac{1}{2}\big|\zeta(s-iT)\big| \ll (2T)^{1-\delta} \ll T$$

for all s with $|s-2| = R$ (the implied constants being absolute). Using this together with $\big|g(2) - 1\big| \leq \frac{1}{6}\pi^2 - 1$ in (434), we get

$$m \ll \log T.$$

This completes the proof, in view of (433). □

We will next give a proof of Theorem 12.2 based on studying the infinite product expansion of $\zeta(s)$ (following Davenport's book).

Lemma 12.3. *If $\rho = \beta + i\gamma$ runs through all the nontrivial zeros of $\zeta(s)$ then*

$$\sum_{\rho} \frac{1}{1 + (T-\gamma)^2} = O(\log T) \qquad \text{as } T \to \infty.$$

Proof. We start with the inequality (391), which states that

$$-\operatorname{Re}\frac{\zeta'(s)}{\zeta(s)} < O(\log t) - \sum_{\rho} \operatorname{Re}\Big(\frac{1}{s-\rho} + \frac{1}{\rho}\Big) \tag{435}$$

[25] Using Remark 8.1 we can simply fix $R = \frac{7}{4}$, say.

for $1 \le \sigma \le 2$ and $t \ge 2$. In this formula we take $s = 2 + iT$, where we keep $T \ge 10$, say. For such s we have $\left|\frac{\zeta'(s)}{\zeta(s)}\right| = \left|-\sum_{n=1}^{\infty} \Lambda(n) n^{-s}\right| \le \sum_{n=1}^{\infty} \Lambda(n) n^{-2} = -\frac{\zeta'(2)}{\zeta(2)} = O(1)$, and hence

$$\sum_{\rho} \operatorname{Re}\Big(\frac{1}{s-\rho} + \operatorname{Re}\frac{1}{\rho}\Big) < O\big(\log T\big).$$

As we have seen earlier we have $\operatorname{Re}\frac{1}{s-\rho} \ge 0$ and $\operatorname{Re}\frac{1}{\rho} \ge 0$ for all zeros ρ, and since

$$\operatorname{Re}\frac{1}{s-\rho} = \frac{2-\beta}{(2-\beta)^2 + (T-\gamma)^2} \ge \frac{1}{4 + (T-\gamma)^2} \ge \frac{1}{4} \cdot \frac{1}{1 + (T-\gamma)^2}$$

we obtain the assertion of the lemma. □

Two immediate corollaries of the above lemma are:

Corollary 12.4. *The number of zeros of $\zeta(s)$ with $T - 1 < \gamma < T + 1$ is $O(\log T)$, as $T \to \infty$.*

Corollary 12.5.

$$\sum_{\substack{\rho \\ \gamma \notin (T-1, T+1)}} \frac{1}{(T-\gamma)^2} = O(\log T) \qquad \text{as } T \to \infty.$$

We can also deduce the following:

Corollary 12.6. *For any s with $t \ge 2$ and $-1 \le \sigma \le 2$ and s not coinciding with a zero of ζ, we have*

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{\substack{\rho \\ |t-\gamma|<1}} \frac{1}{s-\rho} + O(\log t).$$

Proof. By Proposition 10.3, applied at s and at $2 + it$ and subtracted,

$$\begin{aligned} &\frac{\zeta'(s)}{\zeta(s)} - \frac{\zeta'(2+it)}{\zeta(2+it)} \\ &= -\frac{1}{s-1} + \frac{1}{1+it} - \frac{1}{2}\frac{\Gamma'(1+\frac{1}{2}s)}{\Gamma(1+\frac{1}{2}s)} + \frac{1}{2}\frac{\Gamma'(2+\frac{1}{2}it)}{\Gamma(2+\frac{1}{2}it)} + \sum_{\rho}\Big(\frac{1}{s-\rho} - \frac{1}{2+it-\rho}\Big). \end{aligned}$$

The first two terms in the right hand side are $O(t^{-1})$, and by Stirling's formula (cf. Problem 8.7) we have

$$\begin{aligned} -\frac{1}{2}\frac{\Gamma'(1+\frac{1}{2}s)}{\Gamma(1+\frac{1}{2}s)} + \frac{1}{2}\frac{\Gamma'(2+\frac{1}{2}it)}{\Gamma(2+\frac{1}{2}it)} &= -\frac{1}{2}\log\Big(\frac{1+\frac{1}{2}s}{2+\frac{1}{2}it}\Big) + O(t^{-1}) \\ &= -\frac{1}{2}\log\Big(1 - \frac{1-\frac{1}{2}\sigma}{2+\frac{1}{2}it}\Big) + O(t^{-1}) = O(t^{-1}). \end{aligned}$$

Furthermore (as also pointed out in the proof of Lemma 12.3) $\frac{\zeta'(2+it)}{\zeta(2+it)} = O(1)$. Hence

$$\frac{\zeta'(s)}{\zeta(s)} = O(1) + \sum_\rho \Big(\frac{1}{s-\rho} - \frac{1}{2+it-\rho} \Big).$$

For the terms with $|\gamma - t| \geq 1$, we have

$$\Big| \frac{1}{s-\rho} - \frac{1}{2+it-\rho} \Big| = \frac{2-\sigma}{|(s-\rho)(2+it-\rho)|} \leq \frac{3}{|\gamma-t|^2},$$

and the sum of these is $O(\log t)$ by Corollary 12.5. As for the terms with $|\gamma - t| < 1$, we have $|2+it-\rho| \geq 1$, and the number of terms is $O(\log t)$ by Corollary 12.4. This gives the result claimed. □

Second proof of Theorem 12.2. As before we may assume that $T \geq 10$ and that T is not the ordinate of a zero of $\zeta(s)$. From the definition of $S(T)$ we have

$$\begin{aligned} \pi S(T) = -\int_{\frac12+iT}^{\infty+iT} \operatorname{Im} \frac{\zeta'(s)}{\zeta(s)}\, ds &= O(1) - \int_{\frac12+iT}^{2+iT} \operatorname{Im} \frac{\zeta'(s)}{\zeta(s)}\, ds \\ &= O(\log T) - \sum_{\substack{\rho \\ |T-\gamma|<1}} \int_{\frac12+iT}^{2+iT} \operatorname{Im} \frac{1}{s-\rho}\, ds, \end{aligned} \tag{436}$$

where in the last step we used Corollary 12.6. Now (since $\frac{1}{s-\rho}$ equals the logarithmic derivative of $s-\rho$) we have for each ρ:

$$\int_{\frac12+iT}^{2+iT} \operatorname{Im} \frac{1}{s-\rho}\, ds = \Delta_{L'} \arg(s-\rho), \tag{437}$$

where L' is the line segment from $\frac12 + iT$ to $2+iT$. This last number has absolute value $\leq \pi$. The number of terms in the sum in (436) is $O(\log T)$, by Corollary 12.4. Hence we conclude

$$\pi S(T) = O(\log T).$$

□

We conclude by some remarks. Note that if the ordinates $\gamma > 0$ of the zeros of $\zeta(s)$ are enumerated in increasing order $\gamma_1 \leq \gamma_2 \leq \ldots$, then, as a simple consequence of Theorem 12.1 and Theorem 12.2,

$$\gamma_n \sim \frac{2\pi n}{\log n} \qquad \text{as } \ n \to \infty. \tag{438}$$

(Cf. Problem 12.1.) The following result also holds:

$$\lim_{n\to\infty} \big(\gamma_{n+1} - \gamma_n \big) = 0. \tag{439}$$

This was proved by Littlewood in 1924; note that is it *not* a consequence of (438) or of Theorems 12.1, 12.2.

We also remark that Littlewood (1924) proved that

$$\int_0^T S(t)\,dt = O(\log T) \qquad \text{as } T \to \infty, \tag{440}$$

and this indicates a high degree of cancellation among the values of the function $S(T)$, and in particular shows that it is appropriate to retain the term $\frac{7}{8}$ in Theorem 12.1.

12.2. **The Number $N(T,\chi)$.**

Definition 12.3. For each Dirichlet character χ and each $T > 0$, we let $N(T,\chi)$ be the number of zeros of $L(s,\chi)$ in the rectangle $0 < \sigma < 1$, $|t| < T$.

(Note that it is no longer appropriate to consider only the upper half-plane, since the zeros are not in general symmetrically placed with respect to the real axis.)

Theorem 12.7. *For any primitive Dirichlet character χ modulo $q \geq 3$ and any $T \geq 2$ we have*

$$\tfrac{1}{2}N(T,\chi) = \frac{T}{2\pi}\log\frac{qT}{2\pi} - \frac{T}{2\pi} + O(\log T + \log q), \tag{441}$$

where the implied constant is absolute.

Proof. The proof is on the same lines as the proofs of Theorem 12.1 and Theorem 12.2, but it is convenient now to consider the variation in $\arg\xi(s,\chi)$ as s describes the rectangle R with vertices

$$\tfrac{5}{2} - iT, \quad \tfrac{5}{2} + iT, \quad -\tfrac{3}{2} + iT, \quad -\tfrac{3}{2} - iT,$$

so as to avoid the possible zero at $s = -1$. As before we may assume that T is not the ordinate of a zero of $L(s,\chi)$, since the remaining case will follow by taking a limit $T' \to T^-$. Our rectangle R includes exactly one trivial zero of $L(s,\chi)$, at either $s = 0$ or $s = 1$ (cf. Corollary 9.7(ii)), and therefore

$$2\pi\big(N(T,\chi) + 1\big) = \Delta_R \arg\xi(s,\chi).$$

The contribution of the left half of the contour is equal to that of the right half, since by the functional equation (Theorem 9.6)

$$\xi(s,\chi) = \frac{\tau(\chi)}{i^a q^{\frac{1}{2}}}\xi(1-s,\overline{\chi}) = \frac{\tau(\chi)}{i^a q^{\frac{1}{2}}}\overline{\xi(1-\overline{s},\chi)},$$

and thus

$$\arg\xi(\sigma+it,\chi) = c + \arg\overline{\xi(1-\sigma+it,\chi)},$$

where c is a constant independent of $s = \sigma + it$.

Recall that by definition $\xi(s,\chi) = (q/\pi)^{\frac12(s+a)}\Gamma\big(\frac12(s+a)\big)L(s,\chi)$. Hence (if we denote by R' the right half of R)

$$\pi\big(N(T,\chi)+1\big) = \Delta_{R'}\arg(q/\pi)^{\frac12(s+a)} + \Delta_{R'}\arg\Gamma\big(\tfrac12(s+a)\big) + \Delta_{R'}\arg L(s,\chi).$$

Here

$$\Delta_{R'}\arg(q/\pi)^{\frac12(s+a)} = T\log(q/\pi)$$

and, by Problem 8.6 (with $\alpha = \frac14 + \frac12 a$)

$$\begin{aligned}\Delta_{R'}\arg\Gamma\big(\tfrac12(s+a)\big) &= \operatorname{Im}\,\log\Gamma\big(\tfrac12(\tfrac12+iT+a)\big) - \operatorname{Im}\,\log\Gamma\big(\tfrac12(\tfrac12-iT+a)\big)\\ &= 2\operatorname{Im}\,\log\Gamma\big(\tfrac14+\tfrac12 a+\tfrac12 iT\big)\\ &= 2\mathrm{Im}\,\Big(\big(-\tfrac14+\tfrac12 a+\tfrac12 iT\big)\log\big(\tfrac12 iT\big)\Big) - T + O(T^{-1})\\ &= T\log\tfrac12 T - T + O(1).\end{aligned}$$

Combining these formulas we obtain

$$\tfrac12 N(T,\chi) = \frac{T}{2\pi}\log\frac{qT}{2\pi} - \frac{T}{2\pi} + \frac{1}{2\pi}\Delta_{R'}\arg L(s,\chi) + O(1),$$

and hence to complete the proof of the theorem it now suffices to prove that

$$\Delta_{R'}\arg L(s,\chi) = O\big(\log T + \log q\big). \tag{442}$$

This will be done by extending Lemma 12.3 and Corollaries 12.4–12.6 to the present situation.

Lemma 12.8. *If $\rho = \beta + i\gamma$ runs through all the nontrivial zeros of $L(s,\chi)$, then*

$$\sum_\rho \frac{1}{1+(T-\gamma)^2} = O\big(\log q + \log(|T|+2)\big), \qquad \forall T\in\mathbb{R},$$

where the implied constant is absolute.

Proof. The proof is completely similar to the proof of Lemma 12.3: We start with the inequality in Lemma 11.3, which says that for any s with $1<\sigma\le 2$ we have

$$-\operatorname{Re}\frac{L'(s,\chi)}{L(s,\chi)} < -\sum_\rho \operatorname{Re}\frac{1}{s-\rho} + O\Big(\log q + \log(|t|+2)\Big),$$

with an absolute implied constant. In this formula we take $s = 2+iT$; for such s we have $\big|\frac{L'(s,\chi)}{L(s,\chi)}\big| = \big|-\sum_{n=1}^\infty \Lambda(n)\chi(n)n^{-s}\big| \le \sum_{n=1}^\infty \Lambda(n)n^{-2} = -\frac{\zeta'(2)}{\zeta(2)} = O(1)$, and hence

$$\sum_\rho \operatorname{Re}\frac{1}{s-\rho} < O\big(\log q + \log(|T|+2)\big).$$

Now since

$$\operatorname{Re}\frac{1}{s-\rho}=\frac{2-\beta}{(2-\beta)^2+(T-\gamma)^2}\geq\frac{1}{4+(T-\gamma)^2}\geq\frac{1}{4}\cdot\frac{1}{1+(T-\gamma)^2}$$

we obtain the assertion of the lemma. □

The following two corollaries follow immediately from the lemma:

Corollary 12.9. *For any $T\in\mathbb{R}$, the number of zeros of $L(s,\chi)$ with $T-1<\gamma<T+1$ is $O\big(\log q+\log(|T|+2)\big)$.*

Corollary 12.10.

$$\sum_{\substack{\rho\\ \gamma\notin(T-1,T+1)}}\frac{1}{(T-\gamma)^2}=O\big(\log q+\log(|T|+2)\big),\qquad\forall T\in\mathbb{R}.$$

We can also deduce the following:

Corollary 12.11. *For any s satisfying $-1\leq\sigma\leq 2$, $|t|\geq 2$ and $L(s,\chi)\neq 0$, we have*

$$\frac{L'(s,\chi)}{L(s,\chi)}=\sum_{\substack{\rho\\ |t-\gamma|<1}}\frac{1}{s-\rho}+O\big(\log q+\log|t|\big).$$

Proof. The proof is quite similar to the proof of Corollary 12.6. By Proposition 10.7, applied at s and at $2+it$ and subtracted,

$$\frac{L'(s,\chi)}{L(s,\chi)}-\frac{L'(2+it,\chi)}{L(2+it,\chi)}=-\frac{1}{2}\frac{\Gamma'(\frac{a}{2}+\frac{s}{2})}{\Gamma(\frac{a}{2}+\frac{s}{2})}+\frac{1}{2}\frac{\Gamma'(1+\frac{a}{2}+\frac{1}{2}it)}{\Gamma(1+\frac{a}{2}+\frac{1}{2}it)}+\sum_{\rho}\Big(\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\Big).$$

By Stirling's formula (cf. Problem 8.7) we have (since $|t|\geq 2$)

$$\begin{aligned}-\frac{1}{2}\frac{\Gamma'(\frac{a}{2}+\frac{s}{2})}{\Gamma(\frac{a}{2}+\frac{s}{2})}+\frac{1}{2}\frac{\Gamma'(1+\frac{a}{2}+\frac{1}{2}it)}{\Gamma(1+\frac{a}{2}+\frac{1}{2}it)}&=-\frac{1}{2}\log\Big(\frac{\frac{a}{2}+\frac{s}{2}}{1+\frac{a}{2}+\frac{1}{2}it}\Big)+O(|t|^{-1})\\&=-\frac{1}{2}\log\Big(1-\frac{1-\frac{1}{2}\sigma}{1+\frac{a}{2}+\frac{1}{2}it}\Big)+O(|t|^{-1})=O(|t|^{-1}).\end{aligned}$$

Furthermore, as pointed out in the proof of Lemma 12.8, $\frac{L'(2+it,\chi)}{L(2+it,\chi)}=O(1)$. Hence

$$\frac{L'(s,\chi)}{L(s,\chi)}=O(1)+\sum_{\rho}\Big(\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\Big).$$

For the terms with $|\gamma-t|\geq 1$, we have

$$\Big|\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\Big|=\frac{2-\sigma}{|(s-\rho)(2+it-\rho)|}\leq\frac{3}{|\gamma-t|^2},$$

and the sum of these is $O(\log q + \log|t|)$ by Corollary 12.10. As for the terms with $|\gamma - t| < 1$, we have $|2 + it - \rho| \geq 1$, and the number of terms is $O(\log q + \log|t|)$ by Corollary 12.9. This gives the result claimed. □

Finally we can now give the proof of the bound (442), viz.

$$\Delta_{R'} \arg L(s,\chi) = O\big(\log T + \log q\big),$$

where we recall that $T \geq 2$ and R' is the sequence of line segments going from $\frac{1}{2} - iT$ to $\frac{5}{2} - iT$, then to $\frac{5}{2} + iT$ and finally to $\frac{1}{2} + iT$.

For all s with $\sigma \geq 2$ we have $\big|L(s,\chi) - 1\big| \leq \sum_{n=2}^{\infty} n^{-\sigma} \leq \sum_{n=2}^{\infty} n^{-2} \leq \frac{3}{4}$ (cf. footnote 24); hence

$$\begin{aligned}\Delta_{R'} \arg L(s,\chi) &= O(1) + \int_{\frac{1}{2}-iT}^{2-iT} \operatorname{Im} \frac{L'(s,\chi)}{L(s,\chi)}\, ds - \int_{\frac{1}{2}+iT}^{2+iT} \operatorname{Im} \frac{L'(s,\chi)}{L(s,\chi)}\, ds \\ &= O\big(\log q + \log T\big) + \sum_{\substack{\rho \\ |-T-\gamma|<1}} \int_{\frac{1}{2}-iT}^{2-iT} \operatorname{Im} \frac{1}{s-\rho}\, ds - \sum_{\substack{\rho \\ |T-\gamma|<1}} \int_{\frac{1}{2}+iT}^{2+iT} \operatorname{Im} \frac{1}{s-\rho}\, ds,\end{aligned}$$

where in the last step we used Corollary 12.11. Each term in the last two sums has absolute value $\leq \pi$ (cf. (437)), and the number of terms in each sum is $O(\log q + \log T)$, by Corollary 12.9. Hence we conclude that (442) holds, and this completes the proof of Theorem 12.7. □

Finally let us note several consequences of Theorem 12.7. First of all a fact about the number of zeros in certain fixed bounded intervals, as $q \to \infty$:

Corollary 12.12. *Given any constant $C > 0$ there exists some $T_0 > 0$ such that for every $q \geq 3$ and every primitive Dirichlet character χ modulo q we have*

$$N(T_0, \chi) > C \log q.$$

Proof. This is a direct consequence of Theorem 12.7 (cf. Problem 12.3 below). □

In particular, using the above corollary for any fixed $C > 0$ shows that the estimate

$$\sum_{\rho} \frac{1}{1+\gamma^2} = O(\log q)$$

(which comes from Lemma 12.8 with $T = 0$) is essentially the best possible.

We next give some extensions of the $N(T,\chi)$-asymptotics to the case of imprimitive characters.

Corollary 12.13. *For every Dirichlet character χ and every $T \geq 2$ we have*

$$\tfrac{1}{2}N(T,\chi) = \frac{T}{2\pi}\log\frac{c(\chi)T}{2\pi} - \frac{T}{2\pi} + O(\log T + \log c(\chi)), \tag{443}$$

where the implied constant is absolute.

Proof. Let $\chi_1 \pmod{q_1}$ be the primitive character which induces χ. Then $q_1 = c(\chi)$. (Cf. Lemma 4.21 and also Lemma 11.7.) Recall that $L(s,\chi)$ and $L(s,\chi_1)$ have exactly the same zeros in the strip $0 < \sigma < 1$, because of the formula $L(s,\chi) = L(s,\chi_1)\prod_{p|q}(1-\chi_1(p)p^{-s})$; hence $N(T,\chi) = N(T,\chi_1)$ for all $T > 0$.

Now the corollary follows from Theorem 12.7 applied to χ_1, if $q_1 \geq 3$. In the remaining case we have $q_1 = 1$ and χ_1 is the trivial character, thus $L(s,\chi_1) = \zeta(s)$ and $N(T,\chi_1) = 2N(T)$ (cf. Definition 12.1), and the corollary follows from Theorem 12.1, Theorem 12.2. □

As we have noted in §11.2 the formula $L(s,\chi) = L(s,\chi_1)\prod_{p|q}(1-\chi_1(p)p^{-s})$ implies that the zeros of $L(s,\chi)$ are exactly those of $L(s,\chi_1)$ together with the zeros of $\prod_{p|q}(1-\chi_1(p)p^{-s})$, and the latter all lie on the imaginary axis, $\sigma = 0$. For purposes of also counting these zeros we define the following counting function:

Definition 12.4. For each Dirichlet character χ and each $T > 0$, we let $N_R(T,\chi)$ be the number of zeros of $L(s,\chi)$ in the rectangle $-\frac{3}{2} < \sigma < \frac{5}{2}$, $|t| < T$.

(Thus if χ is primitive and nontrivial then $N_R(T,\chi) = N(T,\chi) + 1$, cf. Corollary 9.7.)

Corollary 12.14. *For all Dirichlet characters χ modulo q and all $T \geq 2$ we have*

$$N_R(T,\chi) = \frac{T}{\pi}\log\frac{T}{2\pi} + O(T\log(q+1)), \tag{444}$$

where the implied constant is absolute.

Proof. Let $\chi_1 \bmod q_1$ be the primitive character which induces χ. Then by the above discussion we have

$$N_R(T,\chi) = N_R(T,\chi_1) + N_0, \tag{445}$$

where N_0 are the number of zeros of $\prod_{p|q}(1-\chi_1(p)p^{-s})$ with $|t| < T$. For each $p \mid q$ with $\chi_1(p) \neq 0$ (viz. $p \nmid q_1$) we have

$$1-\chi_1(p)p^{-s} = 0 \iff e^{-(\log p)s} = \overline{\chi_1(p)} \iff s = \frac{i}{\log p}\big(\arg\chi_1(p) + 2\pi k\big) \quad (k \in \mathbb{Z}),$$

since $|\chi_1(p)| = 1$. Note also that each of these zeros is a *simple* zero of $1-\chi_1(p)p^{-s}$, since $\frac{d}{ds}(1-\chi_1(p)p^{-s}) = \chi_1(p)p^{-s}\log p \neq 0$ for all s. Hence the number of zeros of $1-\chi_1(p)p^{-s}$ with $|t| < T$ is $\leq O(T\log p + 1) = O(T\log p)$, and thus

$$N_0 \leq \sum_{p|q} O(T\log p) = O(T\log q). \tag{446}$$

(This is true also for $q = 1$, since then $N_0 = 0$.) We also have $N_R(T, \chi_1) = N(T, \chi_1) + 1$ (if $q_1 \geq 3$) or $N_R(T, \chi_1) = N(T, \chi_1) = 2N(T)$ (if $\chi_1 = 1$), and hence by Theorem 12.7 or Theorems 12.1–12.2 we have

$$N_R(T, \chi_1) = \frac{T}{\pi} \log \frac{q_1 T}{2\pi} - \frac{T}{\pi} + O\big(\log T + \log q_1\big) = \frac{T}{\pi} \log \frac{T}{2\pi} + O\big(T \log(q_1 + 1)\big)$$

$$= \frac{T}{\pi} \log \frac{T}{2\pi} + O\big(T \log(q + 1)\big). \tag{447}$$

The corollary follows from (445), (446), (447). □

12.3. **Problems.**

Problem 12.1. Prove (438). [Hint. Compare Problem 7.1.]

Problem 12.2. Give an alternative proof of the bound (442), by imitating the first proof of Theorem 12.2, on p. 190.

Problem 12.3. (a). Prove Corollary 12.12.
(b). Prove the following generalization of Corollary 12.12: Given any constant $C > 0$ there exists some $T_0 > 0$ such that for every $q \geq 1$, every primitive Dirichlet character χ modulo q, and every $T \geq 0$ we have

$$N(T + T_0, \chi) - N(T, \chi) > C(\log q + \log(T + 2)).$$

13. The explicit formula for $\psi(x)$

(Davenport Chapter 17.)

Recall that $\Lambda(n) = \begin{cases} \log p & \text{if } n = p^r \\ 0 & \text{otherwise.} \end{cases}$.

Definition 13.1. We set

$$\psi(x) = \sum_{n \le x} \Lambda(n). \tag{448}$$

This is sometimes called Tchebychev's (auxiliary) ψ-function.

In this lecture we will prove von Mangoldt's formula which expresses $\psi(x)$ as a sum over the zeros of Riemann's $\zeta(s)$. This formula was stated in Riemann's memoir (1860) and proved by von Mangoldt in 1895.

The function $\psi(x)$ has discontinuities at the points where x is a prime power, and in order that von Mangoldt's formula remain valid at these points, it is necessary to modify the definition by taking the mean of the values on the left and on the right:

Definition 13.2. We set

$$\psi_0(x) = \frac{1}{2}\big(\lim_{t \to x^-} \psi(t) + \lim_{t \to x^+} \psi(t)\big). \tag{449}$$

In other words, $\psi_0(x) = \psi(x)$ when x is not a prime power, and $\psi_0(x) = \psi(x) - \frac{1}{2}\Lambda(x)$ when x is a prime power.

Theorem 13.1. *For any $x > 1$ we have*

$$\psi_0(x) = x - \sum_\rho \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1 - x^{-2}), \tag{450}$$

where ρ runs through all the nontrivial zeros of $\zeta(s)$, and the sum is to be understood in the symmetric sense as

$$\lim_{T \to \infty} \sum_{|\gamma| < T} \frac{x^\rho}{\rho} \qquad (\rho = \beta + i\gamma). \tag{451}$$

We will give the proof for $x \ge 2$, and leave the case $1 < x < 2$ as an exercise; cf. Problem 13.1 below.

Remark 13.1. In (450) we may note that $\frac{\zeta'(0)}{\zeta(0)} = \log 2\pi$; this follows from Proposition 10.3 and (374) in Proposition 10.4.

The basic idea of the proof of Theorem 13.1 is to use the discontinuous integral

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{y^s}{s}\,ds=\begin{cases}0 & \text{if } 0<y<1,\\ \frac{1}{2} & \text{if } y=1,\\ 1 & \text{if } y>1\end{cases} \tag{452}$$

(true for any $c>0$; see Lemma 13.3 below for a more precise statement), to pick out the terms in a Dirichlet series with $n\le x$, by taking $y=x/n$. Since

$$\sum_{n=1}^{\infty}\Lambda(n)n^{-s}=-\frac{\zeta'(s)}{\zeta(s)} \tag{453}$$

for $\sigma>1$, we may hope to obtain

$$\psi_0(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{x^s}{s}\,ds, \tag{454}$$

for $c>1$, at least with some appropriate interpretation of the integral (see Lemma 13.5 below for a precise statement which gives the above identity upon taking $T\to\infty$). Finally we try to move the contour in (454) to the left, all the way to $c\to-\infty$. Clearly, on a *formal* level, this gives (450), since the right hand side in (450) is just the sum of *all* residues of $-\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}$ (in particular the residues coming from the trivial zeros of $\zeta(s)$ at $s=-2,-4,-6,\dots$ add up to $-\sum_{m=1}^{\infty}\frac{x^{-2m}}{-2m}=-\log(1-x^{-2})$).

When carrying out the details of the argument outlined above it is best to work throughout with an integral over a *finite* interval; $\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{x^s}{s}\,ds$ (with T large), and bounding all errors uniformly with respect to x and T. This eventually leads to the following theorem, which immediately implies Theorem 13.1, but is more precise and more useful:

Definition 13.3. For $x>0$ we let $\langle x\rangle$ be the distance from x to the nearest prime power, other than x itself in case x is a prime power.

Theorem 13.2. *For any $x\ge 2$ and $T\ge 2$ we have*

$$\psi_0(x)=x-\sum_{|\gamma|<T}\frac{x^\rho}{\rho}-\frac{\zeta'(0)}{\zeta(0)}-\frac{1}{2}\log(1-x^{-2})+R(x,T), \tag{455}$$

where

$$|R(x,T)|\ll\frac{x\log^2(xT)}{T}+(\log x)\min\Big(1,\frac{x}{T\langle x\rangle}\Big), \tag{456}$$

where the implied constant is absolute.

We will give the proof in the next several pages, proving several lemmas as we need them along the way.

Lemma 13.3. *Set*

$$I(y,T)=\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s}{s}\,ds \qquad \text{and} \qquad \delta(y)=\begin{cases}0 & \text{if } 0<y<1,\\ \frac{1}{2} & \text{if } y=1,\\ 1 & \text{if } y>1.\end{cases} \tag{457}$$

Then, for any $y>0$, $c>0$, $T>0$,

$$\big|I(y,T)-\delta(y)\big|<\begin{cases}y^c\min\Big(1,\frac{1}{\pi T|\log y|}\Big) & \text{if } y\neq 1\\ \frac{c}{\pi T} & \text{if } y=1.\end{cases} \tag{458}$$

Proof. Suppose first $y>1$. By the residue theorem we have

$$\frac{1}{2\pi i}\int_C\frac{y^s}{s}\,ds=\mathrm{Res}_{s=0}\frac{y^s}{s}=1,$$

where C is the rectangle with vertices $c-Ti$, $c+Ti$, $-X+Ti$, $-X-Ti$, described in the positive sense, with arbitrary $T>0$, $X>0$. Let $X\to\infty$, keeping T fixed. Then the integral along the side $(-X+Ti,-X-Ti)$ tends to zero, since the path is of fixed length $2T$ and $|y^s/s|\le y^{-X}/X<1/X$ for all s on the path (since $y>1$, $X>0$). Hence

$$\frac{1}{2\pi i}\int_{c-Ti}^{c+Ti}\frac{y^s}{s}\,ds=1-\frac{1}{2\pi i}\int_{-\infty-Ti}^{c-Ti}\frac{y^s}{s}\,ds+\frac{1}{2\pi i}\int_{-\infty+Ti}^{c+Ti}\frac{y^s}{s}\,ds=1-J_1+J_2,$$

say. Here J_1 and J_2 are absolutely convergent and

$$\big|J_1\big|,\ \big|J_2\big|<\frac{1}{2\pi}\int_{-\infty}^{c}\frac{y^\sigma}{T}\,d\sigma=\frac{y^c}{2\pi T\log y}.$$

This proves the bound $\big|I(y,T)-\delta(y)\big|<\frac{y^c}{\pi T|\log y|}$. It remains to also prove $\big|I(y,T)-\delta(y)\big|<y^c$, and this is most easily proved by replacing the vertical path of integration by the arc Γ of the circle with center 0 which goes from $c-iT$ to $c+iT$ and which lies to the left of $\sigma=c$. By the residue theorem we have

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s}{s}\,ds=1+\frac{1}{2\pi i}\int_\Gamma\frac{y^s}{s}\,ds,$$

and here, since the circle of which Γ is part has radius $R=\sqrt{c^2+T^2}$,

$$\Big|\frac{1}{2\pi i}\int_\Gamma\frac{y^s}{s}\,ds\Big|<\frac{1}{2\pi}\frac{y^c}{R}2\pi R=y^c,$$

and this completes the proof of (458) in the case $y>1$.

The case $y<1$ is entirely similar (cf. also the proof of Lemma 7.8), we take both the rectangle C and the circle arc Γ to lie on the *right* of $\sigma=c$, and there is no residue inside the paths of integration.

The remaining case $y = 1$ is easily treated by direct computation. With $s = c + it$ we have

$$\frac{1}{2\pi i}\int_{c-Ti}^{c+Ti}\frac{ds}{s} = \frac{1}{2\pi i}\int_{-T}^{T}\frac{c-it}{c^2+t^2}\,i\,dt = \frac{1}{\pi}\int_0^T \frac{c}{c^2+t^2}\,dt \qquad \{t = cu\}$$
$$= \frac{1}{\pi}\int_0^{T/c}\frac{1}{1+u^2}\,du = \frac{1}{2} - \frac{1}{\pi}\int_{T/c}^{\infty}\frac{1}{1+u^2}\,du,$$

and the last integral is positive and $< \frac{1}{\pi}\int_{T/c}^{\infty}\frac{du}{u^2} = \frac{c}{\pi T}$. □

Now define

$$J(x,T) := \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{x^s}{s}\,ds. \tag{459}$$

Lemma 13.4. *For any $x > 0$, $c > 1$ and $T > 0$ we have*

$$\big|J(x,T) - \psi_0(x)\big| < \sum_{\substack{n=1\\(n\neq x)}}^{\infty}\Lambda(n)(x/n)^c\min\Big(1, \frac{1}{T|\log x/n|}\Big) + \frac{c}{T}\Lambda(x), \tag{460}$$

where we understand $\Lambda(x) := 0$ unless x is a prime power.

Proof. Since $-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty}\Lambda(n)n^{-s}$ for $\sigma > 1$ we have

$$J(x,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\sum_{n=1}^{\infty}\Lambda(n)n^{-s}\frac{x^s}{s}\,ds = \sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{(x/n)^s}{s}\,ds,$$

where the change of order is permitted since the sum $\sum_{n=1}^{\infty}\Lambda(n)n^{-s}$ is known to be uniformly convergent for all s on the finite path of integration. Applying now Lemma 13.3 (where we sacrifice some factors π in the denominators) we get

$$\Big|J(x,T) - \sum_{n=1}^{\infty}\Lambda(n)\delta(x/n)\Big| \le \sum_{\substack{n=1\\(n\neq x)}}^{\infty}\Lambda(n)\Big|\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{(x/n)^s}{s}\,ds - \delta(x/n)\Big|$$
$$< \sum_{\substack{n=1\\(n\neq x)}}^{\infty}\Lambda(n)(x/n)^c\min\Big(1,\frac{1}{T|\log x/n|}\Big) + \frac{c}{T}\Lambda(x).$$

This completes the proof, since $\sum_{n=1}^{\infty}\Lambda(n)\delta(x/n) = \psi_0(x)$. □

Next by estimating the series on the right of (460) we will prove:

Lemma 13.5. *For any $x > 2$, $1 < c \le 3$ and $T > 0$ we have*

$$\big|J(x,T) - \psi_0(x)\big| \ll \frac{x^c}{T(c-1)} + \frac{x\log^2 x}{T} + (\log x)\min\Big(1,\frac{x}{T\langle x\rangle}\Big). \tag{461}$$

Here and below all implied constants are absolute.

Proof. We first consider all terms in (460) for which $n \le \frac{3}{4}x$ or $n \ge \frac{5}{4}x$. For these, $|\log x/n|$ has a positive lower bounds, and so their contribution to the sum is, if we assume from now on $1 < c \le 3$,

$$\ll \frac{x^c}{T}\sum_{n=1}^{\infty}\Lambda(n)n^{-c} = \frac{x^c}{T}\Big(-\frac{\zeta'(c)}{\zeta(c)}\Big) \ll \frac{x^c}{T(c-1)}. \tag{462}$$

Consider next the terms for which $\frac{3}{4}x < n < x$. Let x_1 be the largest prime power less than x; we can suppose that $\frac{3}{4}x < x_1 < x$, since otherwise the terms under consideration vanish. For the term $n = x_1$, we have[26]

$$\log\frac{x}{n} = -\log\Big(1 - \frac{x-x_1}{x}\Big) \ge \frac{x-x_1}{x},$$

and therefore the contribution of this term to (460) is

$$\ll \Lambda(x_1)\min\Big(1, \frac{x}{T(x-x_1)}\Big) \ll (\log x)\min\Big(1, \frac{x}{T(x-x_1)}\Big).$$

For the other terms, we can put $n = x_1 - \nu$, where $0 < \nu < \frac{1}{4}x$, and then

$$\log\frac{x}{n} > \log\frac{x_1}{n} = -\log\Big(1 - \frac{\nu}{x_1}\Big) \ge \frac{\nu}{x_1}.$$

Hence the contribution of these terms to (460) is

$$\ll \sum_{0<\nu<\frac{1}{4}x} \Lambda(x_1-\nu)\cdot\frac{x_1}{T\nu} \ll \frac{x_1\log x}{T}\sum_{0<\nu<\frac{1}{4}x}\frac{1}{\nu} \ll \frac{x\log^2 x}{T}.$$

The terms with $x < n < \frac{5}{4}x$ are dealt with similarly, except that x_1 is replaced by x_2, the least prime power greater than x. [Details: We can suppose that $x < x_2 < \frac{5}{4}x$, since otherwise the terms under consideration vanish. For the term $n = x_2$, we have

$$\Big|\log\frac{x}{n}\Big| = -\log\Big(1 - \frac{x_2-x}{x_2}\Big) \ge \frac{x_2-x}{x_2},$$

and therefore the contribution of this term is

$$\ll \Lambda(x_2)\min\Big(1, \frac{x_2}{T(x_2-x)}\Big) \ll (\log x)\min\Big(1, \frac{x}{T(x_2-x)}\Big).$$

For the other terms, we can put $n = x_2 + \nu$, where $0 < \nu < \frac{1}{4}x$, and then

$$\Big|\log\frac{x}{n}\Big| = -\log\frac{x}{n} > -\log\frac{x_2}{n} = -\log\Big(1 - \frac{\nu}{x_2+\nu}\Big) \ge \frac{\nu}{x_2+\nu} \gg \frac{\nu}{x}.$$

[26]Here and several times below we use the inequality $-\log(1-t) \ge t$ for all $0 \le t < 1$, which follows from the Taylor series $-\log(1-t) = \sum_{n=1}^{\infty} n^{-1}t^n$.

Hence the contribution of these terms is

$$\ll \sum_{0<\nu<\frac{1}{4}x} \Lambda(x_2+\nu)\cdot\frac{x}{T\nu} \ll \frac{x\log x}{T}\sum_{0<\nu<\frac{1}{4}x}\frac{1}{\nu} \ll \frac{x\log^2 x}{T}. \quad]$$

The lemma follows by collecting the above estimates, and noticing that the last term $\frac{c}{T}\Lambda(x)$ in (460) is $\ll \frac{x\log^2 x}{T}$. □

We next optimize the choice of c:

Corollary 13.6. *For any $x > 2$ and $T > 0$, if we take $c = 1 + (\log x)^{-1}$ in the integral $J(x,T)$, then*

$$\big|J(x,T) - \psi_0(x)\big| \ll \frac{x\log^2 x}{T} + (\log x)\min\Big(1, \frac{x}{T\langle x\rangle}\Big). \tag{463}$$

Proof. Direct from Lemma 13.5, since $c = 1 + (\log x)^{-1}$ gives $1 < c < 3$ and $\frac{x^c}{T(c-1)} = \frac{ex\log x}{T} \ll \frac{x\log^2 x}{T}$. □

The next step is to replace the vertical line of integration in $J(x,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds$ by the other three sides of the rectangle with vertices at

$$c - iT, \quad c + iT, \quad -U + iT, \quad -U - iT,$$

where U is a large odd integer, and where we keep $c = 1 + (\log x)^{-1}$ as in the corollary. Thus the left vertical side passes halfway between two of the trivial zeros of $\zeta(s)$. The sum of the residues of the integrand at its poles inside the rectangle is (since $-\frac{\zeta'(s)}{\zeta(s)}$ has residue -1 at $s = 1$, residue 1 at each simple zero ρ of ζ; residue 2 at each double zero ρ of ζ, etc.)

$$x - \sum_{|\gamma|<T}\frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \sum_{0<2m<U}\frac{x^{-2m}}{-2m}.$$

Hence, so long as there is no zero $\rho = \beta + i\gamma$ with $\gamma = T$, we obtain

$$J(x,T) = \frac{1}{2\pi i}\int_L\Big(-\frac{\zeta'(s)}{\zeta(s)}\Big)\frac{x^s}{s}\,ds + \Big\{x - \sum_{|\gamma|<T}\frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \sum_{0<2m<U}\frac{x^{-2m}}{-2m}\Big\}, \tag{464}$$

where L is the broken line going from $c - iT$ to $-U - iT$, then to $-U + iT$, then to $c + iT$.

In order to have good bounds on the integrand along the horizontal sides of L we wish to stay as far as possible away from the zeros of $\zeta(s)$. We achieve this as follows: Start with an arbitrary $T \geq 2$, say. Then by Corollary 12.4 there are at most $O(\log T)$ zeros $\rho = \beta + i\gamma$ with $T \leq \gamma < T + 1$. Hence by replacing T by an appropriately chosen $T' \in [T, T+1]$ we can ensure that

$$|\gamma - T| \gg (\log T)^{-1} \tag{465}$$

for all the zeros $\rho = \beta + i\gamma$. Note that the new T still satisfies $T \geq 2$.

We recall further Corollary 12.6 which implies that

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{\substack{\rho \\ |\gamma - T| < 1}} \frac{1}{s-\rho} + O(\log T)$$

for all $s = \sigma + iT$ with $-1 \leq \sigma \leq 2$. With our present choice of T satisfying (465), each term $\frac{1}{s-\rho}$ is $O(\log T)$, and the number of terms is also $O(\log T)$; hence

$$\frac{\zeta'(\sigma+iT)}{\zeta(\sigma+iT)} = O\big((\log T)^2\big) \qquad \text{for} \quad -1 \leq \sigma \leq 2.$$

The same bound also holds for $\frac{\zeta'(\sigma-iT)}{\zeta(\sigma-iT)}$, and hence the contribution from $L \cap \{-1 \leq \sigma \leq 2\}$ to $\int_L \big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds$ is, since $c = 1 + (\log x)^{-1}$,

$$\ll \int_{-1}^{c} (\log T)^2 \frac{x^\sigma}{T}\,d\sigma = \frac{(\log T)^2}{T}\Big[\frac{x^\sigma}{\log x}\Big]_{\sigma=-1}^{\sigma=c} = \frac{x^c(\log T)^2}{T\log x} \ll \frac{x(\log T)^2}{T\log x}. \tag{466}$$

In order to bound the integrand in $\int_L \big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds$ for the remaining part of L, i.e. that part of L which lies in the half plane $\{\sigma \leq -1\}$, we first prove:

Lemma 13.7. *Let $D \subset \mathbb{C}$ be the half-plane $\{\sigma \leq -1\}$ minus a disc of radius $\frac{1}{2}$ around each of the points $s = -2, -4, -6, \ldots$. Then*

$$\Big|\frac{\zeta'(s)}{\zeta(s)}\Big| \ll \log(2|s|), \qquad \forall s \in D,$$

where the implied constant is absolute.

Proof. This is an easy consequence of the functional equation. It is easiest to use the functional equation in its unsymmetric form, cf. Problem 9.1(a):

$$\zeta(1-s) = 2(2\pi)^{-s}\cos\big(\tfrac{\pi}{2}s\big)\Gamma(s)\zeta(s).$$

Taking the logarithmic derivative of this relation we obtain

$$-\frac{\zeta'(1-s)}{\zeta(1-s)} = -\log(2\pi) - \tfrac{1}{2}\pi\tan\big(\tfrac{1}{2}\pi s\big) + \frac{\Gamma'(s)}{\Gamma(s)} + \frac{\zeta'(s)}{\zeta(s)}.$$

Thus, replacing s with $1-s$:

$$-\frac{\zeta'(s)}{\zeta(s)} = -\log(2\pi) - \tfrac{1}{2}\pi\cot\big(\tfrac{1}{2}\pi s\big) + \frac{\Gamma'(1-s)}{\Gamma(1-s)} + \frac{\zeta'(1-s)}{\zeta(1-s)}. \tag{467}$$

This is an identity between meromorphic functions in the whole complex plane. Assuming now $s \in D$, the complex number $1-s$ has real part ≥ 2, and thus $\frac{\zeta'(1-s)}{\zeta(1-s)} = O(1)$. We also

have, by Problem 8.7,

$$\frac{\Gamma'(1-s)}{\Gamma(1-s)} = \log(1-s) + O\big(|1-s|^{-1}\big) = \log|1-s| + O(1) = O\big(\log(2|s|)\big).$$

Finally we have $\cot(\frac{1}{2}\pi s) = O(1)$ for $s \in D$. [Proof: Since $f(s) = \cot(\frac{1}{2}\pi s)$ satisfies $f(\overline{s}) = \overline{f(s)}$, it suffices to prove $f(s) = O(1)$ for $s \in D$ with $t \geq 0$. Now note that

$$f(s) = i\frac{e^{\frac{1}{2}\pi is} + e^{-\frac{1}{2}\pi is}}{e^{\frac{1}{2}\pi is} - e^{-\frac{1}{2}\pi is}} = -i\frac{1+e^{\pi is}}{1-e^{\pi is}}.$$

Here $\operatorname{Re} \pi is = -\pi t \leq 0$ so that $\big|e^{\pi is}\big| \leq 1$. Note also that if $e^{\pi is}$ is close to 1 then $\pi\sigma = \arg(e^{\pi is})$ is close to $2\pi k$ for some $k \in \mathbb{Z}$, and $\pi t = -\log\big|e^{\pi is}\big|$ is close to 0, so that s must lie in the disc of radius $\frac{1}{2}$ about the point $2k$, contradicting the condition $s \in D$. Hence $|1 - e^{\pi is}| \gg 1$ holds for all $s \in D$. Since also $|1 + e^{\pi is}| \leq 1 + |e^{\pi is}| \leq 2$ we conclude that $f(s) = O(1)$ for $s \in D$.] These bounds together with (467) give the lemma. □

It follows from the above lemma that for s on the horizontal parts of $L \cap \{\sigma \leq -1\}$ we have

$$\Big|\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\Big| \ll \frac{\log(2|s|)}{|s|}x^\sigma \ll \frac{\log T}{T}x^\sigma,$$

since $T \geq 2$ and $(\log u)/u$ is a decreasing function for u large (in fact for all $u > e$). Hence the contribution from the horizontal parts of $L \cap \{\sigma \leq -1\}$ to $\int_L \big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds$ is

$$\ll \frac{\log T}{T}\int_{-U}^{-1} x^\sigma\,d\sigma = \frac{\log T}{T}\Big[\frac{x^\sigma}{\log x}\Big]_{\sigma=-U}^{\sigma=-1} \ll \frac{\log T}{Tx\log x}. \tag{468}$$

Similarly, on the vertical part of L we have

$$\Big|\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\Big| \ll \frac{\log U}{U}x^\sigma,$$

and thus the contribution from the vertical part of L to $\int_L \big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds$ is

$$\ll \frac{\log U}{U}\int_{-T}^{T} x^{-U}\,dt \ll \frac{T\log U}{Ux^U}. \tag{469}$$

Combining (466), (468), (469) we have now proved

$$\int_L \big(-\frac{\zeta'(s)}{\zeta(s)}\big)\frac{x^s}{s}\,ds = O\Big(\frac{x\log^2 T}{T\log x} + \frac{\log T}{Tx\log x} + \frac{T\log U}{Ux^U}\Big).$$

Note here that the second error term is subsumed by the first. Using the above bound in (464), and then letting $U \to \infty$ (which makes the error term $\frac{T\log U}{Ux^U}$ tend to 0) we get

$$J(x,T) = x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \sum_{m=1}^{\infty} \frac{x^{-2m}}{-2m} + O\Big(\frac{x\log^2 T}{T\log x}\Big)$$

$$= x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log\big(1-x^{-2}\big) + O\Big(\frac{x\log^2 T}{T\log x}\Big). \tag{470}$$

Finally using this together with Corollary 13.6 we obtain the formula (455)–(456) stated in Theorem 13.2 (since the sum of the two error terms $\frac{x\log^2 T}{T\log x}$ and $\frac{x\log^2 x}{T}$ is $\ll \frac{x\log^2 T + x\log^2 x}{T} \ll \frac{x\log^2(xT)}{T}$).

Note that we have proved this formula subject to a restriction on T, cf. (465), but this can now be removed: The effect of varying T by a bounded amount is to change $\sum_{|\gamma|<T}\frac{x^\rho}{\rho}$ in (455) by $O(\log T)$ terms, and each term is $O(x/T)$; hence the change of the sum is $O(\frac{x\log T}{T})$, and this is covered by the estimate on the right of (456). □□□

13.1. **The prime number theorem – again.** (Davenport Chapter 18.)

Using the explicit zero free region for $\zeta(s)$ which we proved in Theorem 11.1 we will now prove the prime number theorem with an explicit error term. We could do this by going through the same steps as in §7, now being able to be more explicit about the choice of path L in §7.4. However now that we have proved the explicit formula for $\psi_0(x)$, we are able to reach our goal in a much quicker way:

Theorem 13.8. *There is an absolute constant $c > 0$ such that*

$$\psi(x) = x + O\big(xe^{-c\sqrt{\log x}}\big) \qquad \text{as } x \to \infty. \tag{471}$$

Proof. It clearly suffices to prove the claim for *integers* $x \to \infty$. Thus from now on we assume that $x \geq 2$ is an integer. Recall from Theorem 13.2 that, for any $T \geq 2$,

$$\psi_0(x) = x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1-x^{-2}) + R(x,T), \tag{472}$$

where

$$|R(x,T)| \ll \frac{x\log^2(xT)}{T} + (\log x)\min\Big(1, \frac{x}{T\langle x\rangle}\Big) \ll \frac{x\log^2(xT)}{T}, \tag{473}$$

where the implied constants are absolute, and where the last step follows from $\langle x\rangle \geq 1$, which is true since x is an integer. It follows from Theorem 11.1 that there is some absolute constant $c_1 > 0$ such that $\beta < 1 - \frac{c_1}{\log T}$ holds for each zero $\rho = \beta + i\gamma$ of $\zeta(s)$ with $|\gamma| \leq T$.

Hence for each ρ in the sum in (472) we have

$$\left|x^\rho\right| = x^\beta < xe^{-c_1\frac{\log x}{\log T}},$$

and thus

$$\Big|\sum_{|\gamma|<T}\frac{x^\rho}{\rho}\Big| < xe^{-c_1\frac{\log x}{\log T}}\sum_{|\gamma|<T}\frac{1}{|\gamma|} = 2xe^{-c_1\frac{\log x}{\log T}}\sum_{2<\gamma<T}\frac{1}{\gamma}.$$

(Recall that $|\gamma| > 6 > 0$ for all ρ.) Here by partial summation, if $N(T)$ denotes the number of zeros in the critical strip with ordinates between 0 and T (as in Definition 12.1):

$$\sum_{2<\gamma<T}\frac{1}{\gamma} = \int_2^T t^{-1}\,dN(t) = \frac{N(T)}{T} + \int_2^T t^{-2}N(t)\,dt \ll \log T + \int_2^T \frac{\log T}{t}\,dt \ll \log^2 T,$$

where we used $N(t) \ll t\log t$ for $t \geq 2$, which follows from Theorems 12.1, 12.2. Hence

$$\Big|\sum_{|\gamma|<T}\frac{x^\rho}{\rho}\Big| \ll x(\log T)^2 e^{-c_1\frac{\log x}{\log T}}.$$

Using this together with (472) and (473) we get

$$\psi(x) = \psi_0(x) + O(\log x) = x + O\Big(\log x + x(\log T)^2 e^{-c_1\frac{\log x}{\log T}} + \frac{x\log^2(xT)}{T}\Big). \tag{474}$$

This is true for all integers $x \geq 2$ and all $T \geq 2$. Let us now keep $x \geq 3$ and choose $T = e^{\sqrt{\log x}}$ (thus $T > e > 2$); then $\log T = \sqrt{\log x}$, and we thus obtain

$$\psi(x) = x + O\Big(\log x + x(\log x)e^{-c_1\sqrt{\log x}} + \frac{x\big(\log x + \sqrt{\log x}\big)^2}{e^{\sqrt{\log x}}}\Big). \tag{475}$$

But note that for any fixed constant $\delta > 0$ we have $\log x \ll \log^2 x \ll e^{\delta\sqrt{\log x}}$ as $x \to \infty$ (cf. footnote 16). Hence if we fix c to be an arbitrary constant satisfying $0 < c < \min(c_1, 1)$, we get

$$\psi(x) = x + O\Big(xe^{-c\sqrt{\log x}}\Big) \qquad \text{as } x \to \infty. \tag{476}$$

This completes the proof. □

Remark 13.2. We give some motivation for the choice of T in (474): Assume that x is large. It is easily seen that the term "$\log x$" is subsumed by at least one of the other error terms in (474); and since $x(\log T)^2 e^{-c_1\frac{\log x}{\log T}}$ is an increasing function of T and $\frac{x\log^2(xT)}{T}$ is a decreasing function of T (at least when $T \geq 10$), the optimal choice of T is to make $x(\log T)^2 e^{-c_1\frac{\log x}{\log T}}$ and $\frac{x\log^2(xT)}{T}$ *roughly equal*. Thus, taking the logarithm, we wish to choose T so that $\log x + 2\log\log T - c_1\frac{\log x}{\log T} \approx \log x + 2\log\log(xT) - \log T$. Here the two "$\log x$" cancel, and it seems reasonable as a first try to assume that the log log-terms are negligible. Hence we are led to choose T so that $c_1\frac{\log x}{\log T} = \log T$, viz. $\log T = \sqrt{c_1\log x}$ and $T = e^{\sqrt{c_1\log x}}$. And indeed, this choice of T gives the end result $\psi(x) = x + O\Big(xe^{-c\sqrt{\log x}}\Big)$, *where c is any positive*

constant satisfying $c < \sqrt{c_1}$. [27] It is now also easy to see that this is *essentially* the best possible bound that can be deduced from (474), for if $T \geq e^{\sqrt{c_1 \log x}}$ then $x(\log T)^2 e^{-c_1 \frac{\log x}{\log T}} > xe^{-c_1 \frac{\log x}{\sqrt{c_1 \log x}}} = xe^{-\sqrt{c_1 \log x}}$ and if $T \leq e^{\sqrt{c_1 \log x}}$ then $\frac{x \log^2(xT)}{T} > xe^{-\sqrt{c_1 \log x}}$, i.e. the error term in (474) is *always* $\gg xe^{-\sqrt{c_1 \log x}}$.

Theorem 13.9. *There is an absolute constant* $c > 0$ *such that*

$$\pi(x) = \operatorname{Li} x + O\Big(xe^{-c\sqrt{\log x}}\Big) \qquad \text{as } x \to \infty. \tag{477}$$

This is the form of the prime number theorem proved by de la Vallée Poussin 1899.

Proof. This follows as a simple consequence of Theorem 13.8 using partial integration. (Cf. Problem 3.6 and Remark 6.1.) First of all recall from Remark 6.2 that $\psi(x) - O\big(\sqrt{x}\big) \leq \vartheta(x) \leq \psi(x)$ as $x \to \infty$; hence Theorem 13.8 implies (with the same constant $c > 0$)

$$\vartheta(x) = x + O\Big(xe^{-c\sqrt{\log x}}\Big) \qquad \text{as } x \to \infty. \tag{478}$$

Next, since $\vartheta(x) := \sum_{p \leq x} \log p$ and $\pi(x) := \sum_{p \leq x} 1$ we have

$$\pi(x) = \sum_{p \leq x} 1 = \int_{2-}^{x} \frac{1}{\log y}\, d\vartheta(y) = \frac{\vartheta(x)}{\log x} + \int_2^x \frac{1}{y \log^2 y}\, \vartheta(y)\, dy. \tag{479}$$

Using now (478), which must in fact hold for all $x \geq 2$ if we take the implied constant sufficiently large, we obtain:

$$= \frac{x}{\log x} + O\Big(\frac{xe^{-c\sqrt{\log x}}}{\log x}\Big) + \int_2^x \frac{1}{\log^2 y}\, dy + O\Big(\int_2^x \frac{e^{-c\sqrt{\log y}}}{\log^2 y}\, dy\Big) \tag{480}$$

Here by integration by parts ("backwards"; cf. the first line of the proof of Lemma 6.1):

$$\frac{x}{\log x} + \int_2^x \frac{1}{\log^2 y}\, dy = \int_2^x \frac{dy}{\log y} + \frac{2}{\log 2} = \operatorname{Li} x + O(1). \tag{481}$$

Furthermore, regarding the last error term in (480), for all $y \in [\sqrt{x}, x]$ we have $\sqrt{\log y} \geq \sqrt{\frac{1}{2} \log x} > \frac{1}{2}\sqrt{\log x}$, and hence if $x \geq 4$ then

$$\int_2^x \frac{e^{-c\sqrt{\log y}}}{\log^2 y}\, dy \ll \int_2^{\sqrt{x}} dy + \int_{\sqrt{x}}^x e^{-\frac{1}{2}c\sqrt{\log x}}\, dy \ll xe^{-\frac{1}{2}c\sqrt{\log x}}.$$

Hence we conclude

$$\pi(x) = \operatorname{Li} x + O\Big(xe^{-\frac{1}{2}c\sqrt{\log x}}\Big).$$

This proves (477), with $\frac{1}{2}c$ in place of c. □

[27] This is better than the restriction we obtained in the proof of Theorem 13.8, "$c < \min(c_1, 1)$", but this is of course important only if one is interested in giving an explicit value for c (which would also require that we first give an explicit value for c_1).

Remark 13.3. By estimating $\int_2^x \frac{e^{-c\sqrt{\log y}}}{\log^2 y}\,dy$ a bit more carefully one sees that one can actually take the constant c in (477) to be the *same* as in Theorem 13.8 (as opposed to "$\frac{1}{2}c$" which we obtained in the above proof). See Problem 13.2 below.

Remark 13.4. Note that the deduction in Davenport's book of Theorem 13.9 from Theorem 477 is slightly different from what we did above, since he uses $\pi_1(x) = \sum_{n\le x} \frac{\Lambda(n)}{\log n}$ as an intermediate function between $\psi(x)$ and $\pi(x)$, whereas we used $\vartheta(x)$ as an intermediate function.

Remark 13.5. Theorem 13.9 was improved to

$$\pi(x) = \operatorname{Li} x + O\Big(x \exp\Big(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}}\Big)\Big) \qquad \text{as } x \to \infty, \tag{482}$$

independently by Vinogradov and Korobov in 1958. (Cf. Problem 13.3 below.)

Finally we point out how much better error term we would obtain in the prime number theorem if we assume the Riemann Hypothesis, or a vertical strip as a zero-free region.

Proposition 13.10. *If there is an absolute constant $\frac{1}{2} \le \Theta < 1$ such that $\beta \le \Theta$ holds for all zeros $\rho = \beta + i\gamma$ of $\zeta(s)$, then*

$$\psi(x) = x + O(x^\Theta \log^2 x) \quad \text{and} \quad \pi(x) = \operatorname{Li} x + O(x^\Theta \log x) \qquad \text{as } x \to \infty. \tag{483}$$

In particular if the Riemann Hypothesis holds, viz. $\Theta = \frac{1}{2}$, then

$$\psi(x) = x + O(x^{\frac{1}{2}} \log^2 x) \quad \text{and} \quad \pi(x) = \operatorname{Li} x + O(x^{\frac{1}{2}} \log x) \qquad \text{as } x \to \infty.$$

Proof. Mimicking the proof of Theorem 13.8 but using the fact that we now have $\left|x^\rho\right| = x^\beta \le x^\Theta$ for all ρ in the sum in (472) we get

$$\psi(x) = x + O\Big(1 + x^\Theta (\log T)^2 + \frac{x\log^2(xT)}{T}\Big),$$

for all integers $x \ge 2$ and all $T \ge 2$. Here for large x we take $T = x^{1-\Theta}$; this gives the first relation in (483). (As in the proof of Theorem 13.8 we first obtain this relation only for all large *integers* x, but this implies that the relation actually holds for all large *real* x.)

Next mimicking the proof of Theorem 13.9 but using $\psi(x) = x + O(x^\Theta \log^2 x)$ in place of Theorem 13.8 we obtain

$$\pi(x) = \operatorname{Li} x + O\left(\frac{x^\Theta \log^2 x}{\log x} + \int_2^x \frac{y^\Theta \log^2 y}{y \log^2 y}\,dy\right) = \operatorname{Li} x + O\left(x^\Theta \log x + \int_2^x y^{\Theta-1}\,dy\right),$$

and this leads to the second relation in (483). □

Let us also note a converse to the above result:

Proposition 13.11. *If $\psi(x) = x + O(x^\alpha)$ as $x \to \infty$, for some fixed $\alpha < 1$, then $\beta \le \alpha$ for all zeros $\rho = \beta + i\gamma$ of $\zeta(s)$.*

Proof. Recall from (116) that

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^\infty \psi(x)x^{-s-1}\,dx \qquad (\sigma > 1).$$

Now define $R(x)$ by $\psi(x) = x + R(x)$; then

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_1^\infty x^{-s}\,dx + s\int_1^\infty R(x)x^{-s-1}\,dx = \frac{s}{s-1} + s\int_1^\infty R(x)x^{-s-1}\,dx.$$

We proved this for $\sigma > 1$, but if $\psi(x) = x+O(x^\alpha)$ as $x \to \infty$ then $R(x) = O(x^\alpha)$ as $x \to \infty$ and hence the integral $\int_1^\infty R(x)x^{-s-1}\,dx$ represents an analytic function in the half-plane $\{\sigma > \alpha\}$, so that the above relation gives a meromorphic continuation of $\frac{\zeta'(s)}{\zeta(s)}$ to $\{\sigma > \alpha\}$ with no poles except a simple pole at $s = 1$. This implies that $\zeta(s)$ does not have any zero s with $\sigma > \alpha$. □

13.2. **Problems.**

Problem 13.1. Prove Theorem 13.1 in the case $1 < x < 2$.

Problem 13.2. Let c be an arbitrary positive constant. Prove that

$$\int_2^x \frac{e^{-c\sqrt{\log y}}}{\log^2 y}\,dy \ll \frac{xe^{-c\sqrt{\log x}}}{\log^2 x} \qquad \text{as } x \to \infty,$$

and conclude from this that *if* $\psi(x) = x + O\big(xe^{-c\sqrt{\log x}}\big)$ as $x \to \infty$, *then* $\pi(x) = \operatorname{Li} x + O\Big(\frac{xe^{-c\sqrt{\log x}}}{\log x}\Big)$ as $x \to \infty$.

Problem 13.3. By using the zero-free region (427) proved independently by Vinogradov and Korobov in 1958, prove that there is an absolute constant $c > 0$ such that

$$\pi(x) = \operatorname{Li} x + O\Big(x\exp\Big(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}}\Big)\Big) \qquad \text{as } x \to \infty. \tag{484}$$

Problem 13.4. Prove that for every $x \geq 1$,

$$\psi_1(x) = \frac{x^2}{2} - \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - x\frac{\zeta'(0)}{\zeta(0)} + \frac{\zeta'(-1)}{\zeta(-1)} - \sum_{r=1}^\infty \frac{x^{1-2r}}{2r(2r-1)},$$

where in the first sum, ρ runs over all nontrivial zeros of $\zeta(s)$.
[Hint. This can be proved by mimicking the proof of the explicit formula for $\psi(x)$ in the present section. However things can be *very much* simplified in this case of $\psi_1(x)$!]

14. The explicit formula for $\psi(x,\chi)$

Definition 14.1. For any Dirichlet character χ we set

$$\psi(x,\chi)=\sum_{n\le x}\chi(n)\Lambda(n). \tag{485}$$

We also set

$$\psi_0(x,\chi)=\frac{1}{2}\big(\lim_{t\to x^-}\psi(t,\chi)+\lim_{t\to x^+}\psi(t,\chi)\big). \tag{486}$$

Thus $\psi_0(x,\chi)=\psi(x,\chi)$ when x is not a prime power, and $\psi_0(x,\chi)=\psi(x,\chi)-\frac{1}{2}\chi(x)\Lambda(x)$ when x is a prime power.

The sums $\psi(x,\chi)$ play much the same part in the prime number theorem for arithmetic progressions as that played by $\psi(x)$ in the prime number theorem itself, but now there is an aggregate of $\phi(q)$ such sums, one for each charater, instead of a single sum.

To start with we prove the analog of Theorem 13.2. As usual we write $a=0$ if $\chi(1)=1$ and $a=1$ if $\chi(1)=-1$. Let us also *define $b(\chi)$ to be the 0th coefficient in the Laurent expansion of $\frac{L'(s,\chi)}{L(s,\chi)}$ at $s=0$*. Thus $b(\chi)=\frac{L'(0,\chi)}{L(0,\chi)}$ if $\chi(-1)=-1$, while

$$\text{if } \chi(-1)=1: \qquad \frac{L'(s,\chi)}{L(s,\chi)}=\frac{1}{s}+b(\chi)+O(s) \qquad \text{as } s\to 0. \tag{487}$$

(Note that $b(\chi)$ can be expressed in terms of $B(\chi)$ using Proposition 10.7.)

Theorem 14.1. *For any $q\ge 3$, any primitive Dirichlet character χ modulo q, and any $x\ge 2$, $T\ge 2$, we have*

$$\psi_0(x,\chi)=-\sum_{|\gamma|<T}\frac{x^\rho}{\rho}-(1-a)\log x-b(\chi)+\sum_{m=1}^{\infty}\frac{x^{a-2m}}{2m-a}+R(x,T), \tag{488}$$

where the sum is taken over all the nontrivial zeros $\rho=\beta+i\gamma$ of $L(s,\chi)$ with $|\gamma|<T$, and

$$|R(x,T)|\ll\frac{x}{T}\log^2(qxT)+(\log x)\min\Big(1,\frac{x}{T\langle x\rangle}\Big), \tag{489}$$

where the implied constant is absolute.

The proof is a direct mimic of the proof of Theorem 13.2, and as in that proof we prove some lemmas as we need them along the way (note, though, that some of the lemmas in the proof of Theorem 13.2 can be used without any modification).

We let χ be a fixed primitive Dirichlet character modulo $q\ge 3$. Let us define

$$J(x,T):=\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\Big(-\frac{L'(s,\chi)}{L(s,\chi)}\Big)\frac{x^s}{s}\,ds. \tag{490}$$

Lemma 14.2. *For any $x > 0$, $c > 1$ and $T > 0$ we have*

$$\big|J(x,T) - \psi_0(x,\chi)\big| < \sum_{\substack{n=1\\(n\neq x)}}^{\infty} \Lambda(n)(x/n)^c \min\Big(1, \frac{1}{T|\log x/n|}\Big) + \frac{c}{T}\Lambda(x), \tag{491}$$

where we understand $\Lambda(x) := 0$ unless x is a prime power.

Proof. Since $-\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n=1}^{\infty} \chi(n)\Lambda(n)n^{-s}$ for $\sigma > 1$ we have

$$J(x,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \sum_{n=1}^{\infty} \chi(n)\Lambda(n)n^{-s}\,\frac{x^s}{s}\,ds = \sum_{n=1}^{\infty} \chi(n)\Lambda(n)\frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{(x/n)^s}{s}\,ds,$$

where the change of order is permitted since the sum $\sum_{n=1}^{\infty} \chi(n)\Lambda(n)n^{-s}$ is known to be uniformly convergent for all s on the finite path of integration. Applying now Lemma 13.3 (where we sacrifice some factors π in the denominators) we get

$$\begin{aligned}\Big|J(x,T) - \sum_{n=1}^{\infty}\chi(n)\Lambda(n)\delta(x/n)\Big| &\le \sum_{\substack{n=1\\(n\neq x)}}^{\infty} \Lambda(n)\Big|\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{(x/n)^s}{s}\,ds - \delta(x/n)\Big| \\ &< \sum_{\substack{n=1\\(n\neq x)}}^{\infty} \Lambda(n)(x/n)^c \min\Big(1, \frac{1}{T|\log x/n|}\Big) + \frac{c}{T}\Lambda(x).\end{aligned}$$

This completes the proof, since $\sum_{n=1}^{\infty}\chi(n)\Lambda(n)\delta(x/n) = \psi_0(x,\chi)$. □

Corollary 14.3. *For any $x > 2$ and $T > 0$, if we take $c = 1 + (\log x)^{-1}$ in the integral $J(x,T)$, then*

$$\big|J(x,T) - \psi_0(x,\chi)\big| \ll \frac{x\log^2 x}{T} + (\log x)\min\Big(1, \frac{x}{T\langle x\rangle}\Big). \tag{492}$$

Proof. This follows from Lemma 14.2 exactly as in the proof of Lemma 13.5 together with Corollary 13.6, since the right hand side in (491) is identical with the right hand side in (460). □

We next replace the vertical line of integration in $J(x,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds$ by the other three sides of the rectangle with vertices at

$$c - iT, \quad c + iT, \quad -U + iT, \quad -U - iT,$$

where U is a large integer $\equiv 1 + a \pmod 2$, and where we keep $c = 1 + (\log x)^{-1}$ as in the corollary. Thus the left vertical side passes halfway between two of the trivial zeros of $L(s,\chi)$. The sum of the residues of the integrand at its poles inside the rectangle is (since

$L(s,\chi)$ has no poles, and $\frac{L'(s,\chi)}{L(s,\chi)}$ has residue 1 at each simple zero ρ of $L(s,\chi)$; residue 2 at each double zero ρ of $L(s,\chi)$, etc.)

$$-\sum_{|\gamma|<T}\frac{x^\rho}{\rho}-\mathrm{Res}_{s=0}\Big(\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\Big)-\sum_{0<2m-a<U}\frac{x^{-(2m-a)}}{-(2m-a)}.$$

If $\chi(-1)=-1$ then $L(0,\chi)\neq 0$ and hence $\mathrm{Res}_{s=0}\Big(\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\Big)=\frac{L'(0,\chi)}{L(0,\chi)}=b(\chi)$. However, let us now assume $\chi(-1)=1$. Then the function $\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}$ has a *double* pole at $s=0$. To get hold of the residue we use (487) and $\frac{d}{ds}x^s=(\log x)x^s$ to see that the Laurent expansion of $\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}$ at $s=0$ is

$$\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}=\Big(\frac{1}{s}+b(\chi)+O(s)\Big)\cdot\frac{1}{s}\cdot\Big(1+(\log x)s+O(s^2)\Big)=\frac{1}{s^2}+\big(\log x+b(\chi)\big)\frac{1}{s}+O(1).$$

Thus

$$\mathrm{Res}_{s=0}\Big(\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\Big)=\log x+b(\chi).$$

Hence in general, so long as there is no zero $\rho=\beta+i\gamma$ with $\gamma=T$, we obtain

(493)
$$J(x,T)=\frac{1}{2\pi i}\int_\Gamma\Big(-\frac{L'(s,\chi)}{L(s,\chi)}\Big)\frac{x^s}{s}\,ds-\Big\{\sum_{|\gamma|<T}\frac{x^\rho}{\rho}+(1-a)\log x+b(\chi)+\sum_{0<2m-a<U}\frac{x^{-(2m-a)}}{-(2m-a)}\Big\},$$

where Γ is the broken line going from $c-iT$ to $-U-iT$, then to $-U+iT$, then to $c+iT$.

In order to have good bounds on the integrand along the horizontal sides of L we wish to stay as far as possible away from the zeros of $L(s,\chi)$. We achieve this as follows: Start with an arbitrary $T\geq 2$, say. Then by Corollary 12.9 there are at most $O(\log(qT))$ zeros $\rho=\beta+i\gamma$ with $T\leq\gamma<T+1$. Hence by replacing T by an appropriately chosen $T'\in[T,T+1]$ we can ensure that

$$|\gamma-T|\gg\frac{1}{\log(qT)} \tag{494}$$

for all the zeros $\rho=\beta+i\gamma$. Note that the new T still satisfies $T\geq 2$.

We recall further Corollary 12.11 which implies that

$$\frac{L'(s,\chi)}{L(s,\chi)}=\sum_{\substack{\rho\\|\gamma-T|<1}}\frac{1}{s-\rho}+O\big(\log(qT)\big)$$

for all $s=\sigma+iT$ with $-1\le\sigma\le 2$. With our present choice of T satisfying (494), each term $\frac{1}{s-\rho}$ is $O(\log(qT))$, and the number of terms is also $O(\log(qT))$; hence

$$\frac{L'(\sigma+iT,\chi)}{L(\sigma+iT,\chi)}=O\big(\log^2(qT)\big)\qquad\text{for}\quad -1\le\sigma\le 2.$$

The same bound also holds for $\frac{L'(\sigma-iT,\chi)}{L(\sigma-iT,\chi)}$, and hence the contribution from $\Gamma\cap\{-1\le\sigma\le 2\}$ to $\int_\Gamma\big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds$ is, since $c=1+(\log x)^{-1}$,

$$\ll\int_{-1}^{c}\log^2(qT)\frac{x^\sigma}{T}\,d\sigma=\frac{\log^2(qT)}{T}\Big[\frac{x^\sigma}{\log x}\Big]_{\sigma=-1}^{\sigma=c}=\frac{x^c\log^2(qT)}{T\log x}\ll\frac{x\log^2(qT)}{T\log x}. \tag{495}$$

In order to bound the integrand in $\int_\Gamma\big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds$ for the remaining part of Γ, i.e. that part of Γ which lies in the half plane $\{\sigma\le -1\}$, we first prove:

Lemma 14.4. *Let $D\subset\mathbb{C}$ be the half-plane $\{\sigma\le -1\}$ minus a disc of radius $\frac12$ around each of the points $s=-a,-a-2,-a-4,\dots$. Then*

$$\Big|\frac{L'(s,\chi)}{L(s,\chi)}\Big|\ll\log(q|s|),\qquad\forall s\in D,$$

where the implied constant is absolute.

Proof. We use the functional equation in its unsymmetric form, cf. Problem 9.1(b) (applied with $\overline{\chi}$ in place of χ):

$$L(1-s,\chi)=\varepsilon(\chi)2^{1-s}\pi^{-s}q^{s-\frac12}\cos\big(\tfrac12\pi(s-a)\big)\Gamma(s)L(s,\overline{\chi}),$$

where $\varepsilon(\chi)=\frac{i^a q^{\frac12}}{\tau(\overline{\chi})}$ (thus $\big|\varepsilon(\chi)\big|=1$). Taking the logarithmic derivative of this relation we obtain

$$-\frac{L'(1-s,\chi)}{L(1-s,\chi)}=\log q-\log(2\pi)-\tfrac12\pi\tan\big(\tfrac12\pi(s-a)\big)+\frac{\Gamma'(s)}{\Gamma(s)}+\frac{L'(s,\overline{\chi})}{L(s,\overline{\chi})}.$$

Thus, replacing s with $1-s$:

$$-\frac{L'(s,\chi)}{L(s,\chi)}=\log q-\log(2\pi)-\tfrac12\pi\cot\big(\tfrac12\pi(s+a)\big)+\frac{\Gamma'(1-s)}{\Gamma(1-s)}+\frac{L'(1-s,\overline{\chi})}{L(1-s,\overline{\chi})}. \tag{496}$$

Assuming now $s\in D$, the complex number $1-s$ has real part ≥ 2, and thus $\frac{L'(1-s,\chi)}{L(1-s,\chi)}=O(1)$. We also have $\frac{\Gamma'(1-s)}{\Gamma(1-s)}=O\big(\log(2|s|)\big)$ and $\cot(\frac12\pi(s+a))=O(1)$ for all $s\in D$, exactly as in the proof of Lemma 13.7. These bounds together with (496) give the lemma. □

It follows from the above lemma that for s on the horizontal parts of $\Gamma\cap\{\sigma\le -1\}$ we have

$$\Big|\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\Big|\ll\frac{\log(q|s|)}{|s|}x^\sigma\ll\frac{\log(qT)}{T}x^\sigma,$$

since $T \geq 2$ and $(\log u)/u$ is a decreasing function for u large (in fact for all $u > e$). Hence the contribution from the horizontal parts of $\Gamma \cap \{\sigma \leq -1\}$ to $\int_\Gamma \big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds$ is

$$\ll \frac{\log(qT)}{T}\int_{-U}^{-1} x^\sigma\,d\sigma = \frac{\log(qT)}{T}\Big[\frac{x^\sigma}{\log x}\Big]_{\sigma=-U}^{\sigma=-1} \ll \frac{\log(qT)}{Tx\log x}. \tag{497}$$

Similarly, on the vertical part of Γ we have

$$\Big|\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\Big| \ll \frac{\log(qU)}{U}x^\sigma,$$

and thus the contribution from the vertical part of L to $\int_\Gamma \big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds$ is

$$\ll \frac{\log(qU)}{U}\int_{-T}^{T} x^{-U}\,dt \ll \frac{T\log(qU)}{Ux^U}. \tag{498}$$

Combining (495), (497), (498) we have now proved

$$\int_\Gamma \big(-\frac{L'(s,\chi)}{L(s,\chi)}\big)\frac{x^s}{s}\,ds = O\Big(\frac{x\log^2(qT)}{T\log x} + \frac{\log(qT)}{Tx\log x} + \frac{T\log(qU)}{Ux^U}\Big).$$

Note here that the second error term is subsumed by the first. Using the above bound in (493), and then letting $U \to \infty$ (which makes the error term $\frac{T\log(qU)}{Ux^U}$ tend to 0) we get

$$J(x,T) = -\sum_{|\gamma|<T}\frac{x^\rho}{\rho} - (1-a)\log x - b(\chi) + \sum_{m=1}^{\infty}\frac{x^{a-2m}}{2m-a} + O\Big(\frac{x\log^2(qT)}{T\log x}\Big). \tag{499}$$

Finally using this together with Corollary 14.3 we obtain the formula (488)–(489) stated in Theorem 14.1 (since the sum of the two error terms $\frac{x\log^2(qT)}{T\log x}$ and $\frac{x\log^2 x}{T}$ is $\ll \frac{x\log^2(qT)+x\log^2 x}{T} \ll \frac{x\log^2(qxT)}{T}$).

Note that we have proved this formula subject to a restriction on T, cf. (494), but this can now be removed: The effect of varying T by a bounded amount is to change $\sum_{|\gamma|<T}\frac{x^\rho}{\rho}$ in (488) by $O(\log(qT))$ terms, and each term is $O(x/T)$; hence the change of the sum is $O(\frac{x\log(qT)}{T})$, and this is covered by the estimate on the right of (489). This completes the proof of Theorem 14.1. □□□

Corollary 14.5. *For any $q \geq 3$, any primitive Dirichlet character χ modulo q, and any $x \geq 2$, we have*

$$\psi_0(x,\chi) = -\sum_\rho\frac{x^\rho}{\rho} - (1-a)\log x - b(\chi) + \sum_{m=1}^{\infty}\frac{x^{a-2m}}{2m-a}, \tag{500}$$

where the sum is taken over all the nontrivial zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$, and should be understood in the symmetric sense as

$$\lim_{T\to\infty} \sum_{|\gamma|<T} \frac{x^\rho}{\rho}. \tag{501}$$

From the point of view of application to the distribution of primes in arithmetic progressions with a variable modulus, Theorem 14.1 (and even more Corollary 14.5) is of little use as it stands. It contains the unknown $b(\chi)$, and it contains terms x^ρ/ρ for which ρ may be very near either 1 or 0, and finally we also need a formula for *non-primitive* characters χ. Regarding the second point, recall that Theorem 11.4 and Theorem 11.5 state that there is an absolute constant $c > 0$ such that $L(s,\chi)$ has at most one zero within a distance $c/\log q$ of $s = 1$ (and so also at most one zero within a distance $c/\log q$ of $s = 0$), and this one zero is itself real and can only occur when χ is a real character. It is important to have this zero visible explicitly in the formula for $\psi(x,\chi)$.

In order to have a precise statement, we make the following definition.

Definition 14.2. Let us fix once and for all a numerical constant $0 < c < \frac{1}{4}$ such that both Theorem 11.4 and Theorem 11.5 hold with this constant c. Now for any nonprincipal Dirichlet character χ modulo q we call a zero $\rho = \beta + i\gamma$ of $L(s,\chi)$ *exceptional* if it satisfies $|\gamma| < 1$ and $\beta > 1 - \frac{c}{\log q}$. By Theorem 11.4 and Theorem 11.5 an exceptional zero can only occur if χ is real, and if it occurs then it is *real* and *unique* (for given χ); we will denote it by β_1.

Note that all exceptional zeros satisfy $\beta_1 > \frac{3}{4}$, since $\beta_1 > 1 - \frac{c}{\log q}$ with $c < \frac{1}{4}$ and $q \geq 3$. Note also that if $L(s,\chi)$ has an exceptional zero β_1 then also $1 - \beta_1$ is a zero of $L(s,\chi)$ (cf. Corollary 9.7(i)).

Theorem 14.6. *If χ is a nonprincipal character to the modulus q, and $2 \leq T \leq x$, then*

$$\psi(x,\chi) = -\begin{Bmatrix} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{Bmatrix} - {\sum_{|\gamma|<T}}' \frac{x^\rho}{\rho} + R(x,T), \tag{502}$$

where $\sum'$ denotes summation over all the zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ with $0 < \beta < 1$,[28] *excluding the zeros β_1 and $1 - \beta_1$ if they exist, and where*

$$|R(x,T)| \ll xT^{-1}\log^2(qx) + x^{\frac{1}{4}}\log x, \tag{503}$$

the implied constant being absolute.

(Although the error bound in (503) is much worse than the one in Theorem 14.1, the above formula is still more convenient in many situations.)

[28] If χ is primitive then these zeros are exactly the *nontrivial zeros* (cf. Remark 11.1). However if χ is not primitive then we have not defined the concept of "nontrivial zeros".

Proof. We will first *assume that χ is primitive*, modulo $q \geq 3$.

We may assume that x is an integer, since the effect on $\psi(x,\chi)$ of replacing x by the nearest integer is $O(\log x)$, which is covered by the bound in (503).

Given now any $2 \leq T \leq x$ with $x \in \mathbb{Z}$ we have $\langle x \rangle \geq 1$ and hence by Theorem 14.1:

$$\psi_0(x,\chi) = -\sum_{|\gamma|<T} \frac{x^\rho}{\rho} - (1-a)\log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{a-2m}}{2m-a} + O\Big(\frac{x}{T}\log^2(qxT)\Big),$$

where the sum is taken over all the nontrivial zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ with $|\gamma| < T$. Using also $\psi(x,\chi) = \psi_0(x,\chi) + O(\log x)$, $T \leq x$ and $\sum_{m=1}^{\infty} \frac{x^{a-2m}}{2m-a} \leq \sum_{k=1}^{\infty} x^{-k} = \frac{1}{x-1}$ we get

$$\psi(x,\chi) = -\sum_{|\gamma|<T} \frac{x^\rho}{\rho} - b(\chi) + O\big(xT^{-1}\log^2(qx)\big). \tag{504}$$

Next to get control on $b(\chi)$ we recall that by Proposition 10.7 we have (an identity of meromorphic functions on all $\mathbb{C}$):

$$\frac{L'(s,\chi)}{L(s,\chi)} = -\frac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})} + B(\chi) + \sum_\rho \Big(\frac{1}{s-\rho} + \frac{1}{\rho}\Big),$$

where ρ runs through all the nontrivial zeros of $L(s,\chi)$ (in particular all ρ are $\neq 0$). If we take $s = 2$ in this formula we obtain

$$\frac{L'(2,\chi)}{L(2,\chi)} = -\frac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(1+\frac{a}{2})}{\Gamma(1+\frac{a}{2})} + B(\chi) + \sum_\rho \Big(\frac{1}{2-\rho} + \frac{1}{\rho}\Big).$$

Subtracting the last two relations and using $\frac{L'(2,\chi)}{L(2,\chi)} = O(1)$ and $\frac{\Gamma'(1+\frac{a}{2})}{\Gamma(1+\frac{a}{2})} = O(1)$ we get

$$\frac{L'(s,\chi)}{L(s,\chi)} = O(1) - \frac{1}{2}\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})} + \sum_\rho \Big(\frac{1}{s-\rho} - \frac{1}{2-\rho}\Big).$$

Now if $a = 1$ then $\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})}$ is analytic at $s = 0$, whereas if $a = 0$ then $\frac{\Gamma'(\frac{s}{2}+\frac{a}{2})}{\Gamma(\frac{s}{2}+\frac{a}{2})}$ has the Laurent expansion $-2s^{-1} + C_0 + O(s)$ as $s \to 0$. Hence since $b(\chi)$ by definition is the 0th coefficient in the Laurent expansion of $\frac{L'(s,\chi)}{L(s,\chi)}$ at $s = 0$ we conclude

$$b(\chi) = O(1) - \sum_\rho \Big(\frac{1}{\rho} + \frac{1}{2-\rho}\Big).$$

Note that for the terms in this series with $|\gamma| \geq 1$ we have

$$\sum_{\substack{\rho \\ |\gamma|\geq 1}} \Big|\frac{1}{\rho} + \frac{1}{2-\rho}\Big| = 2\sum_{|\gamma|\geq 1} \frac{1}{|\rho(2-\rho)|} \ll \sum_{|\gamma|\geq 1} \frac{1}{\gamma^2} = O(\log q),$$

where in the last step we used Lemma 12.8 with $T = 0$. Similarly, using the fact that $|2-\rho| \gg 1 \gg 1+\gamma^2$ when $|\gamma| < 1$ we get

$$\sum_{\substack{\rho \\ |\gamma|<1}} \Big|\frac{1}{2-\rho}\Big| \ll \sum_{\substack{\rho \\ |\gamma|<1}} \frac{1}{1+\gamma^2} = O(\log q),$$

again by Lemma 12.8 with $T = 0$. Hence we obtain

$$b(\chi) = O(\log q) - \sum_{\substack{\rho \\ |\gamma|<1}} \frac{1}{\rho}.$$

We can therefore rewrite (504) as

$$\psi(x,\chi) = -\sum_{|\gamma|<T} \frac{x^\rho}{\rho} + \sum_{\substack{\rho \\ |\gamma|<1}} \frac{1}{\rho} + O\big(xT^{-1}\log^2(qx)\big). \tag{505}$$

If $L(s,\chi)$ does *not* have any exceptional zero then $\beta \le 1 - \frac{c}{\log q}$ holds for all ρ with $|\gamma| < 1$ (cf. Definition 14.2) and hence since the zeros are placed symmetrically about the line $\sigma = \frac12$ we also have $\beta \ge \frac{c}{\log q}$ and thus $\rho^{-1} = O(\log q)$ for all ρ. Also the number of ρ's with $|\gamma| < 1$ is $O(\log q)$ by Theorem 12.7 with $T = 2$, and hence $\sum_{|\gamma|<1} \rho^{-1} = O(\log^2 q)$, so that (502)–(503) hold (for recall that the nontrivial zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ are exactly the same as the zeros of $L(s,\chi)$ which satisfy $0 < \beta < 1$; cf. Remark 11.1).

Next assume that $L(s,\chi)$ *does* have an exceptional zero β_1. We use $\sum'$ to denote summation over the nontrivial zeros ρ *excluding* the zeros β_1 and $1-\beta_1$ if they exist (cf. the statement of the proposition). We then get from (505):

$$\psi(x,\chi) = -{\sum_{|\gamma|<T}}' \frac{x^\rho}{\rho} + {\sum_{\substack{\rho \\ |\gamma|<1}}}' \frac{1}{\rho} - \frac{x^{\beta_1}-1}{\beta_1} - \frac{x^{1-\beta_1}-1}{1-\beta_1} + O\big(xT^{-1}\log^2(qx)\big).$$

Just as before the second sum is $O(\log^2 q)$ and can hence be absorbed in the error term. We can also omit the term β_1^{-1}, which is $O(1)$. Finally, by the mean value theorem, $\frac{x^{1-\beta_1}-1}{1-\beta_1} = x^\sigma \log x$ for some σ between 0 and $1-\beta_1$, and $x^\sigma \log x < x^{\frac14}\log x$ since $\sigma < 1-\beta_1 < \frac14$. Hence we again obtain (502)–(503).

Hence the theorem is proved in the case of χ primitive modulo $q \ge 3$. Finally we extend the proof to the case of a general nonprincipal character χ modulo q. In this case we let χ_1 modulo $q_1 = c(\chi)$ be the corresponding primitive character. Then $q_1 \ge 3$ since χ is nonprincipal, and thus (502)–(503) hold for χ_1, q_1, viz.

$$\psi(x,\chi_1) = -\left\{\begin{matrix} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{matrix}\right\} - {\sum_{|\gamma|<T}}' \frac{x^\rho}{\rho} + O\Big(xT^{-1}\log^2(q_1x) + x^{\frac14}\log x\Big),$$

where $\sum'$ denotes summation over all the zeros $\rho = \beta + i\gamma$ of $L(s,\chi_1)$ with $0 < \beta < 1$, excluding the zeros β_1 and $1-\beta_1$ if they exist. Recall that $L(s,\chi)$ *and* $L(s,\chi_1)$ *have exactly the same zeros in the strip* $0 < \beta < 1$, since $L(s,\chi) = L(s,\chi_1)\prod_{p|q}(1-\chi_1(p)p^{-s})$. Furthermore if $L(s,\chi)$ has an exceptional zero β_1 then $\beta_1 > 1 - \frac{c}{\log q} \geq 1 - \frac{c}{\log q_1}$, and hence β_1 must *also* be the (unique) exceptional zero of $L(s,\chi_1)$. It follows that the sum

$$-\begin{cases} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{cases} - \sum_{|\gamma|<T}{}' \frac{x^\rho}{\rho}$$

is exactly the same for $L(s,\chi)$ *as for* $L(s,\chi_1)$, except when $L(s,\chi_1)$ has an exceptional zero β_1 but $L(s,\chi)$ does not have any exceptional zero. In this latter case the two sums differ exactly by $\frac{x^{1-\beta_1}}{1-\beta_1}$ (since this term appears in the $\sum'$-sum for $L(s,\chi)$ but not for $L(s,\chi_1)$), and using $1 - \frac{c}{\log q_1} < \beta_1 \leq 1 - \frac{c}{\log q}$ we can bound this difference as follows:

$$\frac{x^{1-\beta_1}}{1-\beta_1} = O(\log q) + \frac{x^{1-\beta_1}-1}{1-\beta_1} = O\big(\log q + x^{\frac14}\log x\big).$$

Furthermore note that

$$\begin{aligned} \big|\psi(x,\chi) - \psi(x,\chi_1)\big| \leq \sum_{\substack{n\leq x \\ (n,q)>1}} \Lambda(n) &= \sum_{p|q}\sum_{\substack{v\geq 1 \\ p^v\leq x}} \log p = O(\log x)\sum_{p|q}\log p \\ &= O\Big((\log x)(\log q)\Big) = O\Big(xT^{-1}\log^2(qx) + x^{\frac14}\log x\Big), \end{aligned} \tag{506}$$

where we used $T \leq x$ in the last step. It follows from these observations that (502)–(503) hold for our χ, q. □

15. The prime number theorem for Arithmetic Progressions (I)

(Davenport Chapter 20.)

Definition 15.1. Given $q \geq 1$ and a with $(a,q)=1$, we write $\pi(x;q,a)$ for the number of prime numbers which are $\leq x$ and congruent to a modulo q, viz.

$$\pi(x;q,a) = \#\big\{p \,:\, p \text{ is a prime number } \leq x \text{ and } p \equiv a \,(\mathrm{mod}\ q)\big\}.$$

Recall that $\psi(x,\chi) = \sum_{n\leq x} \chi(n)\Lambda(n)$. We now also define another ψ-function which is more directly related to the counting which we are interested in.

Definition 15.2. We set

$$\psi(x;q,a) = \sum_{\substack{n\leq x \\ n\equiv a \,(\mathrm{mod}\ q)}} \Lambda(n). \tag{507}$$

The relationship between $\psi(x;q,a)$ and $\psi(x,\chi)$ follows immediately from Lemma 1.5; we have, if $(a,q)=1$,

$$\psi(x;q,a) = \frac{1}{\phi(q)} \sum_{\chi\in X_q} \overline{\chi}(a)\psi(x,\chi). \tag{508}$$

Recall the definition of an exceptional zero; Definition 14.2. We will now assume that the constant c in Definition 14.2 is so small that also Corollary 11.8 holds with this c; thus for each $q\in\mathbb{Z}^+$ there is at most *one* $\chi\in X_q$ for which $L(s,\chi)$ has an exceptional zero.

Theorem 15.1. *There exists an absolute constant $c_1>0$ such that for all $x\geq 2$ and all integers q,a satisfying $q\geq 1$ and $(a,q)=1$, we have*

$$\psi(x;q,a) = \frac{x}{\phi(q)} - \frac{\overline{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1} + O\Big(xe^{-c_1\sqrt{\log x}}\Big), \tag{509}$$

where the implied constant is absolute. In (509)*, χ_1 denotes the single real character modulo q, if it exists, for which $L(s,\chi_1)$ has an exceptional zero β_1; if such χ_1,β_1 do not exist then the corresponding term in* (509) *should be omitted.*

Proof. The cases $q=1$ and $q=2$ follow from Theorem 13.8; hence from now on we assume $q\geq 3$. Let us first note some trivial bounds on the terms involved in (509): We have

$$\psi(x;q,a) \leq (\log x) \sum_{\substack{n\leq x \\ n\equiv a \,(\mathrm{mod}\ q)}} 1 < (\log x)\Big(\frac{x}{q}+1\Big). \tag{510}$$

Using this together with $\phi(q) \gg \frac{q}{\log q}$ (cf. Problem 15.1) and $\beta_1 > \frac{3}{4}$ (as remarked just below Definition 14.2) we see that, for all $x \geq 2$, $q \geq 3$:

$$\psi(x;q,a) - \frac{x}{\phi(q)} + \frac{\overline{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1} = O\Big(\frac{x}{q}\log(qx) + \log x\Big), \tag{511}$$

where the implied constant is absolute. Hence (509) is true in a *trivial* way whenever $\frac{x}{q}\log(qx) + \log x \ll xe^{-c_1\sqrt{\log x}}$, i.e. whenever

$$\frac{x}{q}\log(qx) \ll xe^{-c_1\sqrt{\log x}}. \tag{512}$$

In particular it follows that *if c_2 is any fixed number with $c_2 > c_1$, then (509) is true in a trivial way whenever $q \geq e^{c_2\sqrt{\log x}}$.* [Proof: For every $X > 2$ we have $\sup_{q\geq 1,\, x\in[2,X]} \frac{x}{q}\log(qx) < \infty$; hence it suffices to check that (512) holds for "x large"; more specifically we may assume that $x \geq 2$ is so large that $e^{c_2\sqrt{\log x}} > e$. Now note that for x fixed, $\frac{x}{q}\log(qx)$ is a decreasing function of (real) q for $q > e$; hence it now suffices to prove that (512) holds for $q = e^{c_2\sqrt{\log x}}$. This is verified by a direct computation: For any $x \geq 2$ and $q = e^{c_2\sqrt{\log x}}$ we have

$$\frac{x}{q}\log(qx) = xe^{-c_2\sqrt{\log x}}\big(c_2\sqrt{\log x} + \log x\big) \ll xe^{-c_1\sqrt{\log x}} \qquad (\text{since } c_2 > c_1),$$

as desired.]

We now turn to proving (509) also in the *nontrivial* case, when (512) does not hold. We will use (508), which we here recall:

$$\psi(x;q,a) = \frac{1}{\phi(q)}\sum_{\chi\in X_q} \overline{\chi}(a)\psi(x,\chi). \tag{513}$$

The contribution of the principal character $\chi_0 \in X_q$ to the sum on the right provides the main term. To estimate this main term we first note that

$$\big|\psi(x,\chi_0) - \psi(x)\big| = \sum_{\substack{n\leq x\\(n,q)>1}} \Lambda(n) \ll (\log q)(\log x), \tag{514}$$

where the last step follows as in (506). By Theorem 13.8, there is an absolute constant $c_3 > 0$ such that

$$\psi(x) = x + O\big(xe^{-c_3\sqrt{\log x}}\big), \qquad \forall x \geq 2.$$

Hence

$$\psi(x,\chi_0) = x + O\big(xe^{-c_3\sqrt{\log x}} + (\log q)(\log x)\big), \qquad \forall x \geq 2,$$

and using this in (513) we get

$$\psi(x;q,a) = \frac{x}{\phi(q)} + \frac{1}{\phi(q)} \sum_{\substack{\chi \in X_q \\ \chi \neq \chi_0}} \overline{\chi}(a)\psi(x,\chi) + O\Big(\frac{1}{\phi(q)}\big\{xe^{-c_3\sqrt{\log x}} + (\log q)(\log x)\big\}\Big). \tag{515}$$

Next, for every $\chi \in X_q$ with $\chi \neq \chi_0$ and any $T \in [2,x]$ we have, by Theorem 14.6,

$$\psi(x,\chi) = -\begin{Bmatrix} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{Bmatrix} - {\sum_{|\gamma|<T}}' \frac{x^\rho}{\rho} + O\Big(xT^{-1}\log^2(qx) + x^{\frac{1}{4}}\log x\Big), \tag{516}$$

where $\sum'$ denotes summation over all the zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ with $0 < \beta < 1$ excluding the zeros β_1 and $1-\beta_1$ if they exist. By Theorem 11.4 and Theorem 11.5, every $\rho = \beta + i\gamma$ occurring in the $\sum'$-sum satisfies

$$\beta < 1 - \frac{c}{\log(qT)}$$

(where $c > 0$ is the constant which we have fixed in Definition 14.2). Hence

$$\big|x^\rho\big| = x^\beta < xe^{-c\frac{\log x}{\log(qT)}}$$

and thus

$${\sum_{|\gamma|<T}}' \frac{x^\rho}{\rho} = O\Big(xe^{-c\frac{\log x}{\log(qT)}} {\sum_{|\gamma|<T}}' \frac{1}{|\rho|}\Big).$$

Here the last sum can be estimated using the counting function $N(T,\chi)$ (cf. Definition 12.3) and the fact that $N(t,\chi) \ll t\log(qt)$ for all $t \geq 1$ (cf. Corollary 12.13):

$$\begin{aligned} {\sum_{1\leq|\gamma|<T}}' \frac{1}{|\rho|} &= \int_1^T t^{-1}\,dN(t,\chi) \leq T^{-1}N(T,\chi) + \int_1^T t^{-2}N(t,\chi)\,dt \\ &\ll \log(qT) + \int_1^T t^{-1}\log(qt)\,dt \leq \Big(1 + \int_1^T t^{-1}\,dt\Big)\log(qT) \ll (\log T)\log(qT), \end{aligned}$$

and since each ρ with $|\gamma| < 1$ satisfies $\beta > \frac{c}{\log q}$ by Theorem 11.4 and Theorem 11.5 and the $[\beta \leftrightarrow 1-\beta]$-symmetry, we also have

$${\sum_{|\gamma|<1}}' \frac{1}{|\rho|} \ll N(1,\chi)\cdot\log q \ll \log^2 q,$$

so that

$${\sum_{|\gamma|<T}}' \frac{1}{|\rho|} \ll (\log T)\log(qT) + \log^2 q \leq \big(\log T + \log q\big)\big(\log T + \log q\big) = \log^2(qT). \tag{517}$$

Using these bounds in (516) we get

(518)
$$\psi(x,\chi) = -\begin{Bmatrix} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{Bmatrix} + O\Big(xe^{-c\frac{\log x}{\log(qT)}}\log^2(qT) + xT^{-1}\log^2(qx) + x^{\frac14}\log x\Big).$$

Let us now assume $q \le e^{\sqrt{\log x}}$. We may also assume $x \ge 3$. We choose $T = e^{\sqrt{\log x}}$ (note $2 \le T \le x$), and then find that the error term in (518) is

$$\ll xe^{-\frac{c}{2}\frac{\log x}{\sqrt{\log x}}}\big(2\sqrt{\log x}\big)^2 + x(\log x)^2e^{-\sqrt{\log x}} + x^{\frac14}\log x \ll xe^{-c_4\sqrt{\log x}},$$

where $c_4 > 0$ is any fixed constant satisfying $c_4 < \min(\frac{c}{2}, 1)$. Hence:

$$\psi(x,\chi) = -\begin{Bmatrix} \frac{x^{\beta_1}}{\beta_1} & \text{if } \beta_1 \text{ exists} \\ 0 & \text{otherwise} \end{Bmatrix} + O\Big(xe^{-c_4\sqrt{\log x}}\Big). \tag{519}$$

Using this for each $\chi \in X_q$, $\chi \ne \chi_0$ in (515) we get

$$\psi(x;q,a) = \frac{x}{\phi(q)} - \frac{\overline{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1} + O\Big(\frac{1}{\phi(q)}\big\{xe^{-c_3\sqrt{\log x}} + (\log q)(\log x)\big\} + xe^{-c_4\sqrt{\log x}}\Big),$$

where the interpretation of the χ_1, β_1-term is exactly as in (509). It follows from this that, under our present assumption $q \le e^{\sqrt{\log x}}$, (509) holds with $c_1 = \min(c_3, c_4)$! On the other hand if $q \ge e^{\sqrt{\log x}}$, then by using $c_1 \le c_4 < 1$ and our remark just below (512), we see that (509) holds in a trivial way. □

Remark 15.1. The choice of T in (518) can be motivated in very much the same way as in Remark 13.2. Thus: We want to choose T so as to make $xe^{-c\frac{\log x}{\log(qT)}}\log^2(qT)$ and $xT^{-1}\log^2(qx)$ roughly equal. Taking the logarithm we see that this means that we wish to make $-c\frac{\log x}{\log q + \log T} + 2\log\log(qT) \approx -\log T + 2\log\log(qx)$, and if the log log-terms are negligible this gives $(\log T)(\log T + \log q) \approx c\log x$. If we don't wish to be specific about the constant factor in the exponent in the final bound (viz. c_1 in (509)) then all we have to do is to choose T so that $(\log T)(\log T + \log q)$ is both $\ll \log x$ and $\gg \log x$; i.e. both $(\log T)^2$ and $(\log T)(\log q)$ should be $\ll \log x$ and one of them should also be $\gg \log x$. Since we are assuming $\log q \le \sqrt{\log x}$ this is seen to be fulfilled by choosing $\log T \ll\gg \sqrt{\log x}$. Also, just as in Remark 13.2 (and using $xe^{-c\frac{\log x}{\log(qT)}}\log^2(qT) \ge xe^{-c\frac{\log x}{\log T}}\log^2 T$) we see that *no matter how we choose* T, *the error term in* (518) *is always* $\gg xe^{-\sqrt{c\log x}}$.

Remark 15.2. Note that the restriction "$q \le \exp\big[C(\log x)^{\frac12}\big]$" made in Davenport's book (p. 122 (7)) is *not* necessary for the statement of Theorem 15.1 (which is a modified version of Davenport's statement near p. 123 (9)). However, as we saw in our discussion near (512), for any fixed $C > c_1$, the statement of Theorem 15.1 is *trivial* (in fact *worse* than the trivial bound) outside the range $q \le \exp\big[C(\log x)^{\frac12}\big]$. On the other hand, for any $C' < c_1$ the statement of Theorem 15.1 is *nontrivial* when $q \le e^{C'\sqrt{\log x}}$, in the sense that the main term

$\frac{x}{\phi(q)}$ in (509) is asymptotically larger than the error term $O\big(xe^{-c_1\sqrt{\log x}}\big)$; viz. $xe^{-c_1\sqrt{\log x}}\Big/\frac{x}{\phi(q)}$ tends to 0 as $x\to\infty$, uniformly with respect to q in the range $1\le q\le e^{C'\sqrt{\log x}}$.

[Proof: $xe^{-c_1\sqrt{\log x}}\Big/\frac{x}{\phi(q)} = \phi(q)e^{-c_1\sqrt{\log x}} \le qe^{-c_1\sqrt{\log x}} \le e^{(C'-c_1)\sqrt{\log x}} \to 0$.]

It is in the possible term containing β_1 that one of the main difficulties in the theory of the distribution of primes in arithmetic progressions shows itself. The only universal upper bound that we have for β_1 is Proposition 11.11 which states[29] $\beta_1 < 1 - \frac{c_5}{q^{\frac12}(\log q)^2}$ for some absolute constant $c_5>0$. This leads to:

Corollary 15.2. *For any fixed constant δ with $0<\delta<1$, we have*

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O\Big(xe^{-c_1\sqrt{\log x}}\Big) \tag{520}$$

for all $x\ge 2$ and all integers q,a with $(a,q)=1$ and $1\le q\le (\log x)^{1-\delta}$. Here $c_1>0$ is the absolute constant from Theorem 15.1, and the implied constant in (520) *depends only on δ.*

Proof. As before we may assume $q\ge 3$. It follows from Theorem 15.1 combined with Proposition 11.11 (viz. $\beta_1 < 1-\frac{c_5}{q^{\frac12}(\log q)^2}$) that

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O\bigg(\frac{x}{\phi(q)}\exp\Big(-c_5\frac{\log x}{q^{\frac12}\log^2 q}\Big) + x\exp\Big(-c_1\sqrt{\log x}\Big)\bigg). \tag{521}$$

Using now $3\le q\le (\log x)^{1-\delta}$ it follows that there are some positive constants A, A' (which only depend on δ, c_1, c_5) such that

$$c_5\frac{\log x}{q^{\frac12}\log^2 q} \ge c_5\frac{\log x}{A(\log x)^{\frac12-\frac13\delta}} = \frac{c_5}{A}(\log x)^{\frac12+\frac13\delta} \ge c_1\sqrt{\log x} - A', \qquad \forall x\ge 2.$$

Hence, using also $\phi(q)\ge 1$,

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O\Big(xe^{-c_1\sqrt{\log x}}\big(e^{A'}+1\big)\Big) = \frac{x}{\phi(q)} + O\Big(xe^{-c_1\sqrt{\log x}}\Big).$$

□

Remark 15.3. Note that the severe restriction $q\le(\log x)^{1-\delta}$ is essentially *necessary* if we want the error term in Corollary 15.2 to be of the form $O\big(xe^{-C\sqrt{\log x}}\big)$ with some constant $C>0$. Indeed, if $q\ge\log x$ then the first error term in (521) is

$$\frac{x}{\phi(q)}\exp\Big(-c_5\frac{\log x}{q^{\frac12}\log^2 q}\Big) \ge x\exp\Big(-\log q - c_5\frac{\sqrt{\log x}}{\log^2(\log x)}\Big),$$

[29]For simplicity we sacrifice one factor $\log q$ in the case $a=0$.

and since $\frac{\sqrt{\log x}}{\log^2(\log x)}\Big/\sqrt{\log x} \to 0$ as $x \to \infty$, this error term is *not* $O\big(xe^{-C\sqrt{\log x}}\big)$ for any $C > 0$, unless $\log q \gg \sqrt{\log x}$. Thus: For any choice of $q = q(x)$ such that $q \geq \log x$ and $\log q = o\Big(\sqrt{\log x}\Big)$ the error term in (521) is *worse* than $O\big(xe^{-C\sqrt{\log x}}\big)$ as $x \to \infty$, no matter how small we fix the constant $C > 0$ to be.

However, even though we get an error term which is worse than $O\big(xe^{-C\sqrt{\log x}}\big)$, the formula (521) is still *nontrivial* for larger values of q. In fact so long as we keep $3 \leq q \leq (\log x)^{2-\delta}$, the error term in (521) is asymptotically smaller than the main term, $\frac{x}{\phi(q)}$. Furthermore, also for *much* larger values of q the formula (521) implies the bound $\psi(x; q, a) \ll \frac{x}{\phi(q)}$, which is not trivial. Cf. Problem 15.2 below.

Next we prove (after Page, 1935) that the asymptotic result of Corollary 15.2 holds for a much wider range of q, so long as we allow a fairly small (unknown!) set of exceptions:

Corollary 15.3. *There exists an absolute constant $c_6 > 0$ and a function $q_1 : \mathbb{R}_{\geq 2} \to \mathbb{Z}_{\geq 3}$ which satisfies $q_1(x) \log^4 q_1(x) \gg \log x$ as $x \to \infty$, such that for all $x \geq 2$ and all integers q, a satisfying $q \geq 1$, $q_1(x) \nmid q$ and $(a, q) = 1$, we have*

$$\psi(x; q, a) = \frac{x}{\phi(q)} + O\Big(xe^{-c_6\sqrt{\log x}}\Big), \tag{522}$$

where the implied constant is absolute.

Proof. Let us write $c_6 > 0$ for the absolute constant "c" in Corollary 11.10; by possibly decreasing c_6 we may assume $c_6 \leq c_1$ where $c_1 > 0$ is as in Theorem 15.1. We start by defining the function $q_1 : \mathbb{R}_{\geq 2} \to \mathbb{Z}_{\geq 3}$. For any $x \geq 2$, by Corollary 11.10 applied with $z = \max(e^{\sqrt{\log x}}, 3)$ there is at most one real primitive Dirichlet character χ to a modulus $q_1 \leq e^{\sqrt{\log x}}$ for which $L(s, \chi)$ has a real zero β satisfying $\beta > 1 - \frac{c_6}{\sqrt{\log x}}$. We set $q_1(x) := q_1$ if such a χ exists, and otherwise $q_1(x) := e^{\sqrt{\log x}}$.

Note that if $q_1 = q_1(x) < e^{\sqrt{\log x}}$ and β is the corresponding real zero then $\frac{c_6}{\sqrt{\log x}} > 1 - \beta \gg q_1^{-\frac{1}{2}}(\log q_1)^{-2}$ by Proposition 11.11; hence $q_1(x)(\log q_1(x))^4 \gg \log x$, and this is obviously also true for those x where $q_1(x) = e^{\sqrt{\log x}}$.

Now consider an arbitrary $x \geq 2$ and arbitrary integers q, a with $q \geq 1$, $q_1(x) \nmid q$ and $(a, q) = 1$. If $q \in \{1, 2\}$ or $q \geq e^{\sqrt{\log x}}$ then (522) holds trivially, by the argument in the beginning of the proof of Theorem 15.1 (since $c_6 \leq c_1 < 1$). Hence we may now assume $3 \leq q < e^{\sqrt{\log x}}$. By Theorem 15.1 we have

$$\psi(x; q, a) = \frac{x}{\phi(q)} - \frac{\overline{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1} + O\Big(xe^{-c_1\sqrt{\log x}}\Big), \tag{523}$$

where χ_1, β_1 are as described below (509). If χ_1, β_1 do not exist then by convention the corresponding term in (523) should be omitted and hence (522) holds (since $c_6 \leq c_1$). Now

assume that χ_1, β_1 exist. Let χ_1' be the unique primitive character modulo $q_1' := c(\chi_1)$ which induces χ_1. Since β_1 is a zero of $L(s,\chi_1)$ and $L(s,\chi_1) = L(s,\chi_1')\prod_{p|q}(1-\chi_1(p)p^{-s})$ and $\beta_1 > \frac{3}{4} > 0$, β_1 must also be a zero of $L(s,\chi_1')$. On the other hand $q_1' \le q < e^{\sqrt{\log x}}$ and $q_1' \neq q_1(x)$ (since $q_1' \mid q$, $q_1(x) \nmid q$); hence the definition of $q_1(x)$ implies that $\beta_1 \le 1 - \frac{c_6}{\sqrt{\log x}}$. Hence

$$\Big|\frac{\overline{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1}\Big| < \frac{4}{3}\frac{x}{\phi(q)}e^{-c_6\sqrt{\log x}},$$

and this together with (523) implies (522). □

Finally, let us see what we can deduce when assuming the *Generalized Riemann Hypothesis, GRH*, that is, the hypothesis that for each Dirichlet character χ, every zero of $L(s,\chi)$ which lies in the open critical strip $\{0 < \sigma < 1\}$ in fact lies on the line $\sigma = \frac{1}{2}$. (Note that $\zeta(s)$ is included here, by taking $\chi \equiv 1$.)

Theorem 15.4. *If GRH holds, then for all $x \ge 2$ and all integers q, a with $q \ge 1$, $(a,q) = 1$, we have*

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O\big(x^{\frac{1}{2}}\log^2 x\big), \tag{524}$$

where the implied constant is absolute.

Proof. Since the Riemann Hypothesis holds by assumption, we have by Proposition 13.10

$$\psi(x) = x + O(x^{\frac{1}{2}}\log^2 x), \qquad \forall x \ge 2. \tag{525}$$

Thus (524) holds if $q = 1$ or 2. Note also that if $q \ge x^{\frac{1}{2}}$ and $x \ge 2$ then $\psi(x;q,a) \ll x^{\frac{1}{2}}\log x$ (cf. (510)) and $\frac{x}{\phi(q)} \ll \frac{x}{q/\log q} \ll x^{\frac{1}{2}}\log x$ (where we used the fact that $q/\log q$ is an increasing function of q for $q \ge e$), and this implies that (524) holds trivially. Hence from now on we may assume $3 \le q < x^{\frac{1}{2}}$.

Let χ_0 be the principal character modulo q. Then (525), (514) and $q < x^{\frac{1}{2}}$ imply

$$\psi(x,\chi_0) = x + O(x^{\frac{1}{2}}\log^2 x). \tag{526}$$

Now let χ be an arbitrary nonprincipal character modulo q. Then $L(s,\chi)$ does not have any exceptional zero because we are assuming GRH, and hence by Theorem 14.6,

$$\psi(x,\chi) = -\sum_{|\gamma|<T}\frac{x^\rho}{\rho} + O\Big(xT^{-1}\log^2(qx) + x^{\frac{1}{4}}\log x\Big), \qquad \forall T \in [2,x],$$

where the sum is taken over all the zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ with $0 < \beta < 1$ and $|\gamma| < T$. But GRH implies that $\beta = \frac{1}{2}$ for each such ρ; hence

$$\psi(x,\chi) = O\Big(x^{\frac{1}{2}}\sum_{|\gamma|<T}|\rho|^{-1} + xT^{-1}\log^2(qx) + x^{\frac{1}{4}}\log x\Big).$$

We proved in (517) that $\sum_{|\gamma|<T} |\rho|^{-1} \ll \log^2(qT)$. Hence, taking $T = x^{\frac{1}{2}}$, and using also $3 \le q < x^{\frac{1}{2}}$, we get

$$\big|\psi(x,\chi)\big| \ll x^{\frac{1}{2}} \log^2 x.$$

Using this for each nonprincipal character χ mod q, and also using (526), we obtain (524) via the formula (508). □

Remark 15.4. Note that even with the powerful hypothesis of GRH, we do not get any useful result if $q \ge x^{\frac{1}{2}}$. In fact as was seen in the above proof, the formula (524) is "worse than trivial" whenever $q \ge x^{\frac{1}{2}}$, since then both $\psi(x;q,a)$ and $\frac{x}{\phi(q)}$ are $O\big(x^{\frac{1}{2}} \log x\big)$.

15.1. **Consequences for $\pi(x;q,a)$.** To go from asymptotic information about $\psi(x;q,a)$ to asymptotic information about $\pi(x;q,a)$ is an exercise in partial integration, and is completely similar to the proof of Theorem 13.9. As an intermediate step we use $\vartheta(x;q,a)$, the natural generalization of $\vartheta(x)$:

Definition 15.3. We set

$$\vartheta(x;q,a) := \sum_{\substack{p \le x \\ p \equiv a \,(\mathrm{mod}\ q)}} \log p.$$

Thus $\vartheta(x;1,0) = \vartheta(x)$, just as $\psi(x;1,0) = \psi(x)$ and $\pi(x;1,0) = \pi(x)$.

We first note that $\psi(x;q,a)$ and $\vartheta(x;q,a)$ are asymptotically quite close.

Lemma 15.5. *For all $x \ge 2$ and all integers q, a with $q \ge 1$ we have*

$$0 \le \psi(x;q,a) - \vartheta(x;q,a) \le \psi(x) - \vartheta(x) \ll \sqrt{x},$$

where the implied constant is absolute.

Proof. The first and second inequalities follow from

$$\psi(x;q,a) - \vartheta(x;q,a) = \sum_{\substack{n \le x \\ n \equiv a \,(\mathrm{mod}\ q) \\ n \text{ not a prime}}} \Lambda(n).$$

The last bound was proved in Remark 6.2. □

Theorem 15.6. *Let $c_1 > 0$ be the absolute constant in Theorem 15.1. For all $x \ge 2$ and all integers q, a satisfying $q \ge 1$ and $(a,q) = 1$, we have*

$$\pi(x;q,a) = \frac{1}{\phi(q)} \operatorname{Li} x - \frac{\overline{\chi}_1(a)}{\phi(q)} \operatorname{Li}(x^{\beta_1}) + O\Big(x e^{-c_1\sqrt{\log x}}\Big), \tag{527}$$

where the implied constant is absolute. In (527)*, χ_1 denotes the single real character modulo q, if it exists, for which $L(s,\chi_1)$ has an exceptional zero β_1; if such χ_1, β_1 do not exist then the corresponding term in* (527) *should be omitted.*

Proof. Let x, q, a be arbitrary with $x \geq 2$, $q, a \in \mathbb{Z}$, $q \geq 1$, $(q,a) = 1$. From Theorem 15.1 and Lemma 15.5 we immediately obtain

$$\vartheta(x;q,a) = \frac{x}{\phi(q)} - \frac{\overline{\chi}_1(a) x^{\beta_1}}{\phi(q)\beta_1} + O\Big(xe^{-c_1\sqrt{\log x}}\Big), \tag{528}$$

where the implied constant is absolute. Here and in the remainder of the proof we use our convention that each term containing β_1, χ_1 should be omitted if there is no exceptional zero for any Dirichlet character mod q.

Next we note

$$\pi(x;q,a) = \sum_{\substack{p \leq x \\ p \equiv a \,(\mathrm{mod}\ q)}} 1 = \int_{2-} \frac{1}{\log y}\, d\vartheta(y;q,a) = \frac{\vartheta(x;q,a)}{\log x} + \int_2^x \frac{\vartheta(y;q,a)}{y \log^2 y}\, dy. \tag{529}$$

Using here (528) we get

$$\begin{aligned} &= \frac{1}{\phi(q)} \frac{x}{\log x} - \frac{\overline{\chi}_1(a)}{\phi(q)\beta_1} \frac{x^{\beta_1}}{\log x} + O\Big(\frac{x}{\log x} e^{-c_1\sqrt{\log x}}\Big) + \frac{1}{\phi(q)} \int_2^x \frac{1}{\log^2 y}\, dy \\ &\qquad - \frac{\overline{\chi}_1(a)}{\phi(q)\beta_1} \int_2^x \frac{y^{\beta_1}}{y \log^2 y}\, dy + O\Big(\int_2^x \frac{e^{-c_1\sqrt{\log y}}}{\log^2 y}\, dy \Big). \end{aligned} \tag{530}$$

Here the first and the fourth term together give $\frac{1}{\phi(q)}\big(\operatorname{Li} x + O(1)\big)$, just as in (481). Next, the second and the fifth term together give, by substituting $t = y^{\beta_1}$ in the integral and then using (481) again, with x^{β_1} in place of x,

$$-\frac{\overline{\chi}_1(a)}{\phi(q)} \Big(\frac{x^{\beta_1}}{\log(x^{\beta_1})} + \int_{2^{\beta_1}}^{x^{\beta_1}} \frac{1}{\log^2 t}\, dt \Big) = -\frac{\overline{\chi}_1(a)}{\phi(q)} \Big(\operatorname{Li}(x^{\beta_1}) + O(1) \Big).$$

Finally the two error terms in (530) are both $O\big(xe^{-c_1\sqrt{\log x}}\big)$ (cf. Problem 13.2 regarding the second of the error term). This completes the proof. □

From Theorem 15.6 one may derive analogues of Corollaries 15.2, 15.3 for $\pi(x;q,a)$. We do not pursue this; instead we conclude by stating what can be said about $\pi(x;q,a)$ modulo GRH:

Theorem 15.7. *If GRH holds, then for all $x \geq 2$ and all integers q, a with $q \geq 1$, $(a,q) = 1$, we have*

$$\pi(x;q,a) = \frac{1}{\phi(q)} \operatorname{Li} x + O\big(x^{\frac{1}{2}} \log x\big), \tag{531}$$

where the implied constant is absolute.

We leave the proof as an exercise (see Problem 15.4).

15.2. **Problems.**

Problem 15.1. (a). Prove that $\phi(q) \gg \frac{q}{\log q}$ for all $q \geq 2$.
(b). Prove the stronger fact that $\phi(q) \gg \frac{q}{\log\log q}$ for all $q \geq 3$.

Problem 15.2. In this problem we prove and make more precise the claims in Remark 15.3.
(a). Prove that if we keep $3 \leq q \leq (\log x)^{2-\delta}$ for some fixed constant $\delta > 0$, then the error term in (521) is asymptotically smaller than the main term, i.e.

$$\frac{\frac{x}{\phi(q)} \exp\Big(-c_5 \frac{\log x}{q^{\frac{1}{2}} \log^2 q}\Big) + x \exp\Big(-c_1\sqrt{\log x}\Big)}{\frac{x}{\phi(q)}} \to 0 \qquad \text{as } x \to \infty,$$

uniformly with respect to q in the range $3 \leq q \leq (\log x)^{2-\delta}$.
(b). Prove that in the range $(\log x)^2 \leq q \leq e^{c_1\sqrt{\log x}}$, the formula (521) is equivalent with the bound

$$\psi(x; q, a) = O\Big(\frac{x}{\phi(q)}\Big).$$

Note that this is not a trivial fact!

[Remark regarding the bound in (b): Although this is non-trivial, it can be proved without using Dirichlet L-functions, and in a wider range. In fact, the *Brun-Titchmarsh Theorem* (which was proved by Titchmarsh 1930 [68] using Brun's sieve; cf. also [48, Ch. 3 (Thm. 9)]), says that for all q in the (much larger!) range $1 \leq q \leq x^{1-\varepsilon}$ (where $\varepsilon > 0$ is any fixed constant), $\pi(x; q, a) \ll_\varepsilon \frac{x}{\phi(q)\log x}$. Hence $\psi(x; q, a) \leq \pi(x; q, a) \log x \ll_\varepsilon \frac{x}{\phi(q)}$.]

Problem 15.3. To make the condition on $q_1(x)$ in Corollary 15.3 a tiny bit more explicit, prove the following: A function $q_1 : \mathbb{R}_{\geq 2} \to \mathbb{Z}_{\geq 3}$ satisfies $q_1(x) \log^4 q_1(x) \gg \log x$ as $x \to \infty$ if and only if $q_1(x) \gg \frac{\log x}{(\log\log x)^4}$ as $x \to \infty$.

Problem 15.4. Prove Theorem 15.7.

Problem 15.5. (Difficult!?!) Let $Q : \mathbb{R}^+ \to \mathbb{R}^+$ be an increasing function satisfying $Q(x) \ll x^{1-\varepsilon}$ as $x \to \infty$, for some fixed $\varepsilon > 0$. Prove that the following two statements (A) and (B) are equivalent:
(A). "$\psi(x; q, a) \sim \frac{x}{\phi(q)}$ as $x \to \infty$, uniformly with respect to all integers q with $1 \leq q \leq Q(x)$ and all integers a with $(a, q) = 1$."
(B). "$\pi(x; q, a) \sim \frac{\mathrm{Li}\, x}{\phi(q)}$ as $x \to \infty$, uniformly with respect to all integers q with $1 \leq q \leq Q(x)$ and all integers a with $(a, q) = 1$."

Remark 15.5. We here list the various ranges, in increasing order, for which (A) and (B) hold, either unconditionally or modulo (successively more powerful) conjectures:
• (A) and (B) are known to hold when $Q(x) = (\log x)^{2-\delta}$, for any fixed $\delta > 0$; cf. (521) and Problem 15.2(a).
•The Siegel-Walfisz Theorem 16.6 which we will prove below shows that (A) and (B) hold when $Q(x) = (\log x)^A$, for any fixed $A > 0$ (but these results are *non-effective* when $A > 2$).
•*If there are no Siegel zeros*, then (A) and (B) hold when $Q(x) = e^{C\sqrt{\log x}}$ for any fixed $C \in (0, c_1)$; cf. Theorem 15.1 and Remark 15.2.
•Under GRH, (A) and (B) hold when $Q(x) = x^{\frac{1}{2}}(\log x)^{-2-\varepsilon}$, for any fixed $\varepsilon > 0$; cf. Theorem 15.4.
•It has been proposed that (A) and (B) should hold even in the much larger range $Q(x) = x^{1-\varepsilon}$, for any fixed $\varepsilon > 0$! (Cf., e.g., [25, §IV.2.].)

16. Siegel's Theorem

(Davenport Chapter 21.)

Siegel's Theorem (Siegel 1935, [65]), in the first of its two forms, reads:

Theorem 16.1. *Let $\varepsilon > 0$ be given. Then for every $q \geq 3$ and every real primitive character χ modulo q,*

$$L(1,\chi) \gg q^{-\varepsilon}, \tag{532}$$

where the implied constant only depends on ε.

It follows that for every fundamental discriminant d we have

$$h(d) \gg |d|^{\frac{1}{2}-\varepsilon} \qquad \text{if } d<0, \tag{533}$$

and

$$h(d)\log\varepsilon_d \gg d^{\frac{1}{2}-\varepsilon} \qquad \text{if } d>0. \tag{534}$$

Cf. Theorem 4.35 and Theorem 5.4. Again, the implied constant only depends on ε.

In its second form, Siegel's Theorem reads:

Theorem 16.2. *For any $\varepsilon > 0$ there exists a positive number $c(\varepsilon)$ such that, for every $q \in \mathbb{Z}^+$ and every real Dirichlet character χ modulo q,*

$$s > 1 - c(\varepsilon) q^{-\varepsilon} \Longrightarrow L(s,\chi) \neq 0. \tag{535}$$

Proof of the implication Theorem 16.1 $\Rightarrow$ Theorem 16.2. Let $\varepsilon > 0$ be given. Then by Theorem 16.1 applied with $\frac{1}{2}\varepsilon$ in place of ε, there is a constant $C_1 > 0$ which only depends on ε such that

$$L(1,\chi) \geq C_1 q^{-\frac{1}{2}\varepsilon}$$

for every $q \geq 3$ and every real primitive Dirichlet character χ modulo q. Furthermore there is an absolute constant $C_2 > 0$ such that

$$\big|L'(s,\chi)\big| \leq C_2 \log^2 q, \qquad \forall s \in \big[1-\tfrac{1}{\log q}, 1\big],$$

by Problem 10.1(c). Hence

$$L(s,\chi) > C_1 q^{-\frac{1}{2}\varepsilon} - C_2(1-s)\log^2 q, \qquad \forall s \in \big[1-\tfrac{1}{\log q}, 1\big],$$

and hence

$$s > 1 - \frac{C_1}{C_2}\frac{q^{-\frac{1}{2}\varepsilon}}{\log^2 q} \Longrightarrow L(s,\chi) > 0.$$

Hence with a suitable choice of $0 < c(\varepsilon) < 1$, (535) holds for all $q \geq 3$ and all real primitive characters χ modulo q.

Finally if χ is an arbitrary Dirichlet character, then if χ_1 modulo q_1 is the corresponding primitive character, we know that $L(s,\chi)$ and $L(s,\chi_1)$ have exactly the same zeros in the half plane $\{\sigma > 0\}$, and hence $L(s,\chi) \neq 0$ for all $s > 1 - c(\varepsilon)q_1^{-\varepsilon}$ (this is true also when $q_1 = 1$, i.e. $L(s,\chi_1) = \zeta(s)$; cf. Remark 10.1). But $q_1 \leq q$; hence $q^{-\varepsilon} \leq q_1^{-\varepsilon}$ and thus $L(s,\chi) \neq 0$ for all $s > 1 - c(\varepsilon)q^{-\varepsilon}$. □

Remark 16.1. One can also show by a fairly direct analysis that Theorem 16.2 ⇒ Theorem 16.1. See Problem 16.1 below.

It follows from Theorem 16.2 that any real zero β of $L(s,\chi)$, for real nonprincipal χ, satisfies

$$\beta \leq 1 - c(\varepsilon)q^{-\varepsilon}. \tag{536}$$

This is a much superior estimate to any we have had hitherto. It has, however, the disadvantage of being ***noneffective***, in the sense that it is not possible, with existing knowledge, to assign a numerical value to $c(\varepsilon)$ for a particular value of ε (for example for $\varepsilon = \frac{1}{4}$). Similarly Theorem 16.1 is also noneffective, meaning that we do not know of a way, for a particular value of ε, to assign a numerical value to the implied constant in (532). Also the class number bounds from below, (533) and (534), are noneffective.

Proof of Siegel's Theorem 16.1. Suppose that χ_1 is a real primitive character modulo $q_1 \geq 3$, and χ_2 is a real primitive character modulo $q_2 \geq 3$, where $q_1 \neq q_2$. As we noted in the proof of Theorem 11.6, $\chi_1\chi_2$ is a nonprincipal character modulo q_1q_2. Set

$$F(s) := \zeta(s)L(s,\chi_1)L(s,\chi_2)L(s,\chi_1\chi_2). \tag{537}$$

Then $F(s)$ is analytic in the whole complex plane except for a simple pole at $s = 1$, and its residue at this pole is

$$\lambda = L(1,\chi_1)L(1,\chi_2)L(1,\chi_1\chi_2). \tag{538}$$

A key step in the proof is to bound $F(s)$ from below for $s < 1$ near 1 in a way which shows that *if λ is small then $F(s) > 0$ for some $s < 1$ near 1:*

Lemma 16.3. *There is an absolute constant $C > 0$ such that for any $\frac{7}{8} \leq s < 1$ which satisfies $C\lambda < (1-s)(q_1q_2)^{-4(1-s)}$, we have $F(s) > 0$.*

Proof. Using the Euler products for $\zeta(s)$ and each $L(s,\chi)$ we get

$$F(s) = \prod_p \Big((1-p^{-s})(1-\chi_1(p)p^{-s})(1-\chi_2(p)p^{-s})(1-\chi_1(p)\chi_2(p)p^{-s}) \Big)^{-1}$$

$$= \prod_p \begin{cases} (1-p^{-s})^{-4} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{1\} \\ (1-p^{-s})^{-1} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0\} \\ (1-p^{-2s})^{-2} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{-1\} \text{ or } \{-1,1\} \\ (1-p^{-s})^{-2} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0,1\} \\ (1-p^{-2s})^{-1} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0,-1\} \end{cases}. \tag{539}$$

It follows from this that $F(s)$ is given by an absolutely convergent Dirichlet series for $\sigma > 1$:

$$F(s) = \sum_{n=1}^{\infty} a_n n^{-s}, \tag{540}$$

where $a_1 = 1$ and $a_n \geq 0$ for all n. [Detailed proof: We define a_n ($n \geq 1$) to be the unique multiplicative sequence corresponding to the product (539), i.e. set $a_1 = 1$ and for each prime p define a_{p^r} ($r = 1, 2, 3, \ldots$) by the power series relations (for $|z| < 1$)

$$a_1 + a_p z + a_{p^2} z^2 + a_{p^3} z^3 + \ldots = \begin{cases} (1-z)^{-4} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{1\} \\ (1-z)^{-1} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0\} \\ (1-z^2)^{-2} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{-1\} \text{ or } \{-1,1\} \\ (1-z)^{-2} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0,1\} \\ (1-z^2)^{-1} & \text{if } \{\chi_1(p),\chi_2(p)\} = \{0,-1\}, \end{cases}$$

and finally for composite $n \geq 1$ define a_n (uniquely) by multiplicativity. Then $a_{p^r} \geq 0$ for all primes p and all $r \geq 1$ (since $(1-z)^{-1} = 1+z+z^2+\ldots$ and $(1-z^2)^{-1} = 1+z^2+z^4+\ldots$ have all Taylor coefficients ≥ 0), and hence $a_n \geq 0$ for all n. Now since (539) was obtained by multiplying together four convergent Euler products, we know that the infinite product (539) converges for each $s \in \mathbb{C}$ with $\sigma > 1$. Applying this fact for $s = \sigma > 1$ and using Lemma 2.8 (with $f(n) := a_n n^{-\sigma}$) we obtain $\sum_{n=1}^{\infty} a_n n^{-\sigma} < \infty$ for all $\sigma > 1$; hence the Dirichlet series in the right hand side of (540) has abscissa of absolute convergence $\sigma_a \leq 1$. Finally the equality in (540) follows from (539) and Proposition 2.7 (with $f(n) = a_n n^{-s}$).]

It now follows just as in the argument on p. 17(top) that $F(s)$ has a power series expansion

$$F(s) = \sum_{m=0}^{\infty} b_m (2-s)^m \qquad \big(\text{when } \ |s-2| < 1\big),$$

with $b_0 \geq 1$ and $b_m \geq 0$ for all m. Since $F(s)$ is analytic in the whole complex plane except for a simple pole at $s = 1$ with residue λ, the function $F(s) - \frac{\lambda}{s-1}$ is entire. Note also

$$\frac{\lambda}{s-1} = \frac{\lambda}{1-(2-s)} = \lambda \sum_{m=0}^{\infty} (2-s)^m \qquad \big(\text{when } |s-2| < 1\big);$$

hence the function $F(s) - \frac{\lambda}{s-1}$ has the power series expansion

$$F(s) - \frac{\lambda}{s-1} = \sum_{m=0}^{\infty} (b_m - \lambda)(2-s)^m, \qquad \forall s \in \mathbb{C}. \tag{541}$$

(The power series must have radius of convergence $= \infty$ since the function in the left hand side is entire.)

On the circumference of the circle $|s-2| = \frac{3}{2}$, the function $\zeta(s)$ is bounded, and by Problem 10.1(a) the L-functions satisfy

$$L(s, \chi_1) = O\big(\sqrt{q_1}\big), \qquad L(s, \chi_2) = O\big(\sqrt{q_2}\big), \qquad L(s, \chi_1\chi_2) = O\big(\sqrt{q_1 q_2}\big). \tag{542}$$

Here and in all "big O's" in the rest of the proof, the implied constant is absolute. Thus

$$F(s) = O\big(q_1 q_2\big) \qquad \text{for all } s \text{ with } |s-2| = \tfrac{3}{2}.$$

Similarly from (538) and Problem 10.1(a) (applied in a crude way) we get $\lambda = O\big(q_1 q_2\big)$, and hence $\frac{\lambda}{s-1} = O\big(q_1 q_2\big)$ for all s with $|s-2| = \frac{3}{2}$. Hence

$$F(s) - \frac{\lambda}{s-1} = O\big(q_1 q_2\big) \qquad \text{for all } s \text{ with } |s-2| = \tfrac{3}{2}.$$

Hence by Cauchy's inequalities for the coefficients of a power series, applied to the function (541), we have

$$\big|b_m - \lambda\big| = O\big(q_1 q_2 (\tfrac{2}{3})^m\big).$$

Hence for any real $\frac{7}{8} \leq s < 1$ and any $M \in \mathbb{Z}^+$ we have

$$\sum_{m=M}^{\infty} \big|b_m - \lambda\big|(2-s)^m = \sum_{m=M}^{\infty} O\big(q_1 q_2 (\tfrac{2}{3})^m (2-s)^m\big) = O\big(q_1 q_2\big) \sum_{m=M}^{\infty} (\tfrac{3}{4})^m = O\big(q_1 q_2 (\tfrac{3}{4})^M\big),$$

and thus (using also $b_1 \geq 1$ and $b_m \geq 0$ for all m)

$$\begin{aligned} F(s) - \frac{\lambda}{s-1} &= \sum_{m=0}^{M-1} (b_m - \lambda)(2-s)^m + \sum_{m=M}^{\infty} (b_m - \lambda)(2-s)^m \\ &\geq 1 - \lambda \sum_{m=0}^{M-1} (2-s)^m - O\big(q_1 q_2 (\tfrac{3}{4})^M\big) \\ &\geq 1 - \lambda \frac{(2-s)^M - 1}{1-s} - A q_1 q_2 (\tfrac{3}{4})^M, \end{aligned} \tag{543}$$

where A is an absolute constant which we can take to be ≥ 1. Now choose M as the smallest positive integer for which $Aq_1q_2(\frac{3}{4})^M < \frac{1}{2}$. Then in fact $\frac{3}{8} \leq Aq_1q_2(\frac{3}{4})^M < \frac{1}{2}$ and hence

$$\log \tfrac{3}{8} \leq \log(Aq_1q_2) + M\log\big(\tfrac{3}{4}\big) < \log \tfrac{1}{2};$$

and thus

$$M \leq \frac{\log(\frac{8}{3}Aq_1q_2)}{\log(\frac{4}{3})} \leq O(1) + 4\log(q_1q_2)$$

and

$$(2-s)^M = e^{M\log(1+1-s)} \leq e^{M(1-s)} \leq e^{O(1)+4(1-s)\log(q_1q_2)} \leq O\big((q_1q_2)^{4(1-s)}\big).$$

Using this together with $Aq_1q_2(\frac{3}{4})^M < \frac{1}{2}$ in (543) we get

$$F(s) > \tfrac{1}{2} - O\Big(\frac{\lambda}{1-s}(q_1q_2)^{4(1-s)}\Big). \tag{544}$$

Here the implied constant is absolute, and hence we obtain the statement of the lemma. □

We now complete the proof of Theorem 16.1 using the above lemma. We distinguish two cases, and the distinction depends on the given positive number $\varepsilon > 0$. If there is a real primitive character for which $L(s,\chi)$ has a real zero between $1-\frac{1}{8}\varepsilon$ and 1, we choose χ_1 to be such a character and β_1 to be the zero in question. Then $F(\beta_1) = 0$, independently of what χ_2 may be. Otherwise, if there is no primitive character for which $L(s,\chi)$ has some real zero between $1-\frac{1}{8}\varepsilon$ and 1, then we instead choose χ_1 to be any real primitive character and β_1 to be any number satisfying $1-\frac{1}{8}\varepsilon < \beta_1 < 1$. In this case $F(\beta_1) < 0$, independently of what χ_2 may be, since $\zeta(s) < 0$ when $0 < s < 1$,[30] and the three L-functions in (537) are positive when $s = 1$ and do not vanish for $\beta_1 \leq s \leq 1$.

Hence in either of the two cases we have $F(\beta_1) \leq 0$, no matter what χ_2 may be, and hence by Lemma 16.3,

$$\lambda \gg (1-\beta_1)(q_1q_2)^{-4(1-\beta_1)},$$

where the implied constant is absolute. From now on we keep χ_1 and β_1 fixed. Let χ_2 be any real primitive character to a modulus $q_2 > q_1$. Then by (538) and Problem 10.1(b),

$$\lambda \ll (\log q_1)(\log(q_1q_2))L(1,\chi_2),$$

again with the implied constant being absolute. Hence

$$L(1,\chi_2) \gg q_2^{-4(1-\beta_1)}(\log q_2)^{-1}$$

where *the implied constant depends on* q_1, β_1 (and thus ultimately on ε), but not on q_2, χ_2. Since $4(1-\beta_1) < \frac{1}{2}\varepsilon$ the last inequality implies $L(1,\chi_2) \gg q_2^{-\varepsilon}$. Recall that we have proved

[30]Proof: $\zeta(s) \in \mathbb{R}$ for all $s \in \mathbb{R}$, $s \neq 1$, and $\zeta(s) < 0$ for all $s < 1$ sufficiently near 1, since $\zeta(s)$ has a simple pole at $s = 1$ with residue 1. Hence it suffices to prove that $\zeta(s) \neq 0$ for $0 < s < 1$, and this follows from Remark 10.1.

this for all $q_2 > q_1$ and all real primitive characters χ_2 modulo q_2. By possibly increasing the implied constant we may now allow q_2 to be arbitrary ≥ 3; i.e. we have proved (532). □

Remark 16.2. Regarding the *noneffectiveness* pointed out earlier: Note carefully that by the very nature of the above proof one cannot give an explicit value for the implied constant in (532). This is because, in the "first case" above, the lower bound on $L(1,\chi)$ is given *in terms of a supposed counterexample to GRH*; a large real zero β_1 to some $L(s,\chi)$. So if GRH is true but remains unproved, then the above proof cannot be exploited to give an explicit lower bound on $L(1,\chi)$.

Remark 16.3. The discussion of the "second case" in the above proof can be simplified by referring to Problem 16.1 below. In fact, if there is no primitive character for which $L(s,\chi)$ has some real zero between $1-\frac{1}{8}\varepsilon$ and 1, then by Problem 16.1 we have $L(1,\chi) \gg_\varepsilon \frac{1}{\log q}$ for all $q \geq 3$ and all primitive characters χ mod q, and this is a bound which is quite a bit stronger than Theorem 16.1!

Remark 16.4. Note that the bound which we obtain in (544) has a better constant in the exponent than Davenport's p. 129(8) simply because we used better bounds on the L-functions, cf. (542); but this has no effect on the quality of the final result, Theorem 16.1.

Remark 16.5. The choice of M in (543) can be motivated as follows: We would really like to choose that M which makes $1-\lambda\frac{(2-s)^M-1}{1-s} - Aq_1q_2(\frac{3}{4})^M$ as large as possible. However in the end we are only interested to know whether there exists some M for which that expression is *positive* (cf. the statement of Lemma 16.3). Then $\lambda\frac{(2-s)^M-1}{1-s} < 1$ and $Aq_1q_2(\frac{3}{4})^M < 1$; and in the opposite direction we have that *if* (*) $[\lambda\frac{(2-s)^M-1}{1-s} < \frac{1}{2}$ and $Aq_1q_2(\frac{3}{4})^M < \frac{1}{2}]$ then the desired positivity holds. Hence since we anyway allow unknown implied constants in our bounds, we may just as well ask if there is some M which satisfies (*). Now since $\lambda\frac{(2-s)^M-1}{1-s}$ is an increasing function of M, and $Aq_1q_2(\frac{3}{4})^M$ is a decreasing function of M, it is clear that if (*) holds for *some* M, then (*) holds for M as chosen in the proof text!

Remark 16.6. Another brief and simple proof of Siegel's Theorem was given by Goldfeld [22]. It is a quite useful exercise in relation to this course to work through the details of that proof!

16.1. *** Some history.** [I here more or less copy Davenport pp. 127-8 verbatim, mainly for myself to learn and collect these classical references.]

Siegel's theorem was the culmination of a series of discoveries by several mathematicians. The problem of proving that $h(d) \to \infty$ as $d \to -\infty$, or even of proving that $h(d) \geq 2$ if $-d$ is sufficiently large, was propounded by Gauss, but no progress toward its solution was made until much later. Hecke proved that if the inequality $\beta < 1 - \frac{c_2}{\log q}$ holds for the real zeros of L-functions formed with real primitive characters, then $h(d) > c_3|d|^{\frac{1}{2}}/\log|d|$. (We will prove a result in Problem 16.1 below which when combined with Dirichlet's class number formula immediately implies Hecke's result.) In particular this conclusion would follow from the GRH.

In 1933 Deuring [16] proved the unexpected result that the *falsity* of the classical Riemann hypothesis for $\zeta(s)$ implies that $h(d) \geq 2$ if $-d$ is sufficiently large, and shortly afterward Mordell proved that this assumption also implies that $h(d) \to \infty$ as $d \to -\infty$. Their work was based on a study of the behavior, as $d \to -\infty$, of

$$\sum_{Q} \sum_{(x,y)\in\mathbb{Z}^2\setminus\{\mathbf{0}\}} Q(x,y)^{-s},$$

where Q runs through a representative set of forms of discriminant d.

In 1934, Heilbronn [32] took a further important step forward. He proved that the falsity of the *GRH* implies that $h(d) \to \infty$ as $d \to -\infty$. Together with the result of Hecke, this gave an unconditional proof that $h(d) \to \infty$ as $d \to -\infty$, and so solved Gauss' problem.

Also in 1934, Heilbronn and Linfoot proved that there are at most ten negative discriminants d for which $h(d) = 1$. As nine such d were known,

$$-3, -4, -7, -8, -11, -19, -43, -67, -163, \tag{545}$$

the question was whether there is a tenth such discriminant. If there were, then the L-function $L(s, \chi_d)$ would have a real zero β larger that $\frac{1}{2}$. In 1966, Baker [3] and Stark [66] proved independently that there is no such tenth discriminant. Baker noted that his fundamental theorem in transcendence theory provides a solution of this class number problem in view of earlier work of Gelfond and Linnik. Stark was inspired by a paper of Heegner [31] in which elliptic modular functions were used to show that there is no tenth discriminant with class number 1. It was long thought that Heegner's argument was incomplete, partly because it seemed to depend on an unproved conjecture of Weber. However, in retrospect it has now been found that Heegner's proof is essentially correct; the obscure details have been clarified by Deuring [17] and Stark [67].

In 1976, Goldfeld [23] showed that an effective lower bound for the class number of imaginary quadratic fields could be constructed, if there exists an elliptic curve E defined over the field of rational numbers, whose Mordell-Weil rank is 3, and whose associated L-series has a zero of order 3 at $s = 1$. The existence of such a curve was established by Gross and Zagier [26], which enabled Oesterlé [55, 54] to show that

$$h(d) \geq \frac{1}{55}(\log|d|) \prod_{\substack{p|d \\ p<|d|}} \Big(1 - \frac{2\sqrt{p}}{p+1}\Big) \tag{546}$$

for all quadratic discriminants $d < 0$ (viz., $d \equiv 0$ or $1 \pmod 4$ and $\sqrt{d} \notin \mathbb{Z}$).

We remark that for *positive* d, Gauss conjectured that $h(d) = 1$ infinitely often, and this has still not been contradicted or justified. (Cf. [9, p. 151].)

16.2. **The prime number theorem for Arithmetic Progressions (II).** (Davenport Chapter 22.)

Using Siegel's Theorem 16.2 together with Theorem 15.1 we now obtain the *Siegel-Walfisz Theorem:*

Theorem 16.4. *For any fixed constant $N > 0$ we have*

$$\psi(x; q, a) = \frac{x}{\phi(q)} + O\Big(xe^{-c_1\sqrt{\log x}}\Big) \tag{547}$$

for all $x \geq 2$ and all integers q, a with $(a,q) = 1$ and $1 \leq q \leq (\log x)^N$. Here $c_1 > 0$ is the absolute constant from Theorem 15.1, and the implied constant in (547) *depends only on N, but is* ***non-effective****.*

Proof. We may assume $q \geq 3$. It follows from Theorem 15.1 combined with Theorem 16.2 that

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O\Big(\frac{x}{\phi(q)} \exp\Big(-c(\varepsilon)\frac{\log x}{q^\varepsilon}\Big) + x\exp\Big(-c_1\sqrt{\log x}\Big)\Big), \tag{548}$$

where the implied constant is absolute. If we take $\varepsilon = \frac{1}{3N}$ and keep $q \in [1, (\log x)^N]$ then $\frac{\log x}{q^\varepsilon} \geq \frac{\log x}{(\log x)^{\frac{1}{3}}} = (\log x)^{\frac{2}{3}}$. Hence the first error term in (548) is subsumed by the second error term, so long as we take a sufficiently large implied constant. Hence we get (547). Note that the required implied constant depends on $c(\varepsilon)$, and hence is non-effective. □

Remark 16.7. I prefer the statement of Theorem 16.4 (which is the same as e.g. [48, Ch. 11.3 (Cor. 19)]) over Davenport's p. 133(4), since Theorem 16.4 is *stronger* due to the constant in the exponent being independent of N: Theorem 16.4 almost directly implies Davenport's p. 133(4) with an *absolute* implied constant,[31] but (as far as I can see) to prove the opposite implication is just as difficult as proving Theorem 16.4 itself. Note that the only difference between the proof of Theorem 16.4 and the proof of Davenport's p. 133(4) is that in the former we choose ε to be *strictly smaller* than $\frac{1}{2N}$.

We also give the analogous result for the more basic function $\psi(x,\chi)$, which will be useful to refer to a few times in later sections. (We leave the proof as an exercise; see Problem 16.2 below.)

Theorem 16.5. *For any fixed constant $N > 0$ the following bound holds for all $x \geq 2$, all integers q with $1 \leq q \leq (\log x)^N$, and all nonprincipal Dirichlet characters χ modulo q:*

$$\psi(x,\chi) = O\Big(xe^{-c_1\sqrt{\log x}}\Big). \tag{549}$$

Here $c_1 > 0$ is an absolute constant, and the implied constant depends only on N, but is ***non-effective****.*

We also give the corresponding result for $\pi(x;q,a)$, which is called by the same name as Theorem 16.4, the *Siegel-Walfisz Theorem:*

[31] Proof: Assume Theorem 16.4 and let $N > 0$ be given. Then there is a constant $C_1 > 0$ which depends on N such that $\big|\psi(x;q,a) - \frac{x}{\phi(q)}\big| \leq C_1 x e^{-c_1\sqrt{\log x}}$ whenever $x \geq 2$, $(a,q) = 1$ and $1 \leq q \leq (\log x)^N$. We also have $\big|\psi(x;q,a) - \frac{x}{\phi(q)}\big| \leq \psi(x,1,0) + x = \psi(x) + x \leq C_2 x$ for some *absolute* constant $C_2 > 0$, by Theorem 7.10 or (more elementarily) by Tchebychev's bound (262). Now by fixing the positive constant $C(N)$ sufficiently small we have $\min\big(C_1 e^{-c_1\sqrt{\log x}}, C_2\big) \leq 2C_2 e^{-C(N)\sqrt{\log x}}$ for all $x \geq 2$. It follows that $\big|\psi(x;q,a) - \frac{x}{\phi(q)}\big| \leq 2C_2 x e^{-C(N)\sqrt{\log x}}$ whenever $x \geq 2$, $(a,q) = 1$ and $1 \leq q \leq (\log x)^N$, and we are done.

Theorem 16.6. *For any fixed constant $N > 0$ we have*

$$\pi(x;q,a) = \frac{1}{\phi(q)} \operatorname{Li} x + O\Big(xe^{-c_1\sqrt{\log x}}\Big) \tag{550}$$

for all $x \geq 2$ and all integers q, a with $(a,q) = 1$ and $1 \leq q \leq (\log x)^N$. Here $c_1 > 0$ is the absolute constant from Theorem 15.1, and the implied constant in (550) *depends only on N, but is* ***non-effective****.*

Proof. We may assume $q \geq 3$. It follows from Theorem 15.6 combined with ($\operatorname{Li}(x^{\beta_1}) \ll x^{\beta_1}$ and) Theorem 16.2 that

$$\pi(x;q,a) = \frac{1}{\phi(q)} \operatorname{Li} x + O\bigg(\frac{x}{\phi(q)} \exp\Big(-c(\varepsilon)\frac{\log x}{q^\varepsilon}\Big) + x\exp\Big(-c_1\sqrt{\log x}\Big)\bigg), \tag{551}$$

where the implied constant is absolute. From here the proof is the same as for Theorem 16.4. □

16.3. **Goal for the remainder of the course: Good bounds on average.** One of the main goal for the remainder of this course is to prove the following simple and far-reaching result of Bombieri:

Theorem 16.7. *For any positive constant A, there exists a positive constant B such that for all $x \geq 2$ we have*

$$\sum_{q \leq x^{\frac{1}{2}}(\log x)^{-B}} \max_{(a,q)=1} \max_{y \leq x} \Big|\psi(y;q,a) - \frac{y}{\phi(q)}\Big| \ll x(\log x)^{-A}, \tag{552}$$

where the implied constant only depends on A.

(In fact we will prove that one can take $B = A + 5$ in the above statement; thus trivially also every $B > A + 5$ works.)

Note that, up to factors of $\log x$, *Theorem 16.7 gives what GRH gives on average!* Namely: If GRH holds then by Theorem 15.4 we have

$$\max_{(a,q)=1} \max_{y \leq x} \Big|\psi(y;q,a) - \frac{y}{\phi(q)}\Big| \ll x^{\frac{1}{2}} \log^2 x,$$

uniformly over all x and q, and adding this over q with $1 \leq q \leq x^{\frac{1}{2}}(\log x)^{-B}$ we get

$$\sum_{q \leq x^{\frac{1}{2}}(\log x)^{-B}} \max_{(a,q)=1} \max_{y \leq x} \Big|\psi(y;q,a) - \frac{y}{\phi(q)}\Big| \ll x(\log x)^{2-B},$$

i.e. (552) holds with $B = A + 2$. We also point out that for x large and for *"almost"* all q in the range $x^{\frac{1}{2}}(\log x)^{-B} \ll q \leq x^{\frac{1}{2}}(\log x)^{-B}$, Theorem 16.7 implies an asymptotic formula for $\psi(x;q,a)$ of essentially the same quality as what GRH gives (viz., up to log-factors)! Cf. Problems 16.3 and 16.4 below for precise statements.

16.4. **Problems.**

Problem 16.1. Prove that for every nonprincipal real character χ modulo q, if β_1 denotes the largest real zero of $L(s,\chi)$ (so that $\beta_1 \geq -1$ by Corollary 9.7(ii)), then

$$L(1,\chi) \gg \min\Big(1-\beta_1, \frac{1}{\log q}\Big), \tag{553}$$

where the implied constant is absolute. (In particular this result shows that Theorem 16.2 implies Theorem 16.1, i.e. the opposite of the direction we proved on p. 232.)

Problem 16.2. Prove Theorem 16.5.

Problem 16.3. Prove that if GRH holds, then for each $A > 0$ the following holds for all $x \geq 2$ and all integers q, a with $1 \leq q \leq x^{\frac{1}{2}}(\log)^{-A-2}$ and $(a,q) = 1$:

$$\psi(x;q,a) = \frac{x}{\phi(q)}\Big(1 + O\big((\log x)^{-A}\big)\Big), \tag{554}$$

where the implied constant is absolute. Also prove that for any q with $x^{\frac{1}{2}}(\log)^{-A-2} \ll q \leq x$, the above bound implies the bound in Theorem 15.4 except for an extra factor $\log\log x$. (Thus in the range $x^{\frac{1}{2}}(\log x)^{-A-2} \ll q \leq x^{\frac{1}{2}}(\log x)^{-A-2}$ the bound (554) is essentially the best we can prove using GRH.)

Problem 16.4. Prove the following consequence of Theorem 16.7: Let $A, B > 0$ be as in Theorem 16.7, and take any constant C with $0 < C < A$. Then for every $x \geq 2$, setting $Q = x^{\frac{1}{2}}(\log x)^{-B}$ there is a subset $\mathcal{S} \subset \{1, 2, \dots, \lfloor Q \rfloor\}$ with $\#\mathcal{S} \leq Q(\log x)^{-C}$ such that for every q, a with $1 \leq q \leq Q$, $q \notin \mathcal{S}$ and $(a,q) = 1$ we have

$$\psi(x;q,a) = \frac{x}{\phi(q)}\Big(1 + O\big((\log x)^{C-A}\big)\Big), \tag{555}$$

where the implied constant depends only on A.

Remark 16.8. Since $\#\mathcal{S}/Q \to 0$ as $x \to \infty$, we may in particular say that "(555) holds for almost all $q \leq Q$ as $x \to \infty$"; hence also "$\psi(x;q,a) \sim \frac{x}{\phi(q)}$ holds for almost all $q \leq x^{\frac{1}{2}}(\log x)^{-B}$ as $x \to \infty$".

Remark 16.9. In particular for any $K > 0$ we know that we can take $A = 2K+1$ and $B = 2K+6$ in Theorem 16.7; if we also take $C = K+1$ then we see that if $Q = x^{\frac{1}{2}}(\log x)^{-2K-6}$, then

$$\psi(x;q,a) = \frac{x}{\phi(q)}\Big(1 + O\big((\log x)^{-K}\big)\Big),$$

for all $q \leq Q$ with the possible exception of at most $Q(\log x)^{-K-1}$ values of q. This is a slightly stronger statement than what Davenport states on p. 162 (lines 6–10), with $B \leftrightarrow K$.

17. The Polya-Vinogradov Inequality

(Davenport Chapter 23.)

The following is the *Polya-Vinogradov Inequality:*

Theorem 17.1. *If χ is a nonprincipal character modulo q and $M \in \mathbb{Z}$, $N \in \mathbb{Z}_{\geq 0}$, then*

$$\Big| \sum_{n=M+1}^{M+N} \chi(n) \Big| \leq 2q^{\frac{1}{2}} \log q. \tag{556}$$

Proof. The following elementary argument is due to Schur. First assume that χ is primitive (and nonprincipal; thus $q \geq 3$). Then by Lemma 9.4 we have

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) e\Big(\frac{an}{q}\Big),$$

where $\big|\tau(\overline{\chi})\big| = \sqrt{q}$ by Lemma 9.5. Hence

$$\sum_{n=M+1}^{M+N} \chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) \sum_{n=M+1}^{M+N} e\Big(\frac{an}{q}\Big) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q-1} \overline{\chi}(a) \sum_{n=M+1}^{M+N} e\Big(\frac{an}{q}\Big).$$

(In the last step we used $\overline{\chi}(q) = 0$; this is true since χ is nonprincipal.) The inner sum is a finite geometric series which equals

$$= \frac{e\big(\frac{a(M+N+1)}{q}\big) - e\big(\frac{a(M+1)}{q}\big)}{e\big(\frac{a}{q}\big) - 1}.$$

Hence

$$\Big| \sum_{n=M+1}^{M+N} \chi(n) \Big| \leq q^{-\frac{1}{2}} \sum_{a=1}^{q-1} \frac{2}{\big|e\big(\frac{a}{q}\big) - 1\big|} = q^{-\frac{1}{2}} \sum_{a=1}^{q-1} \frac{1}{\big|\sin\big(\pi a/q\big)\big|} = q^{-\frac{1}{2}} \sum_{a=1}^{q-1} \frac{1}{\sin\big(\pi a/q\big)}. \tag{557}$$

To bound the last sum in a convenient way we note that the function $f(x) = \frac{1}{\sin(\pi x)}$ is convex in $0 < x < 1$, since $f'(x) = -\pi(\sin \pi x)^{-2} \cos(\pi x)$ is an increasing function of x in this interval. Hence

$$f(\alpha) \leq \frac{1}{\delta} \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} f(x)\, dx \tag{558}$$

for all α, δ satisfying $0 < \alpha - \frac{1}{2}\delta < \alpha + \frac{1}{2}\delta < 1$. [Proof: Since f is convex we have $f(\alpha) \leq \frac{f(\alpha+h)+f(\alpha-h)}{2}$ for all $h \in [0, \frac{1}{2}\delta]$. Integrating this inequality over $h \in [0, \frac{1}{2}\delta]$ we obtain $\frac{1}{2}\delta f(\alpha) \leq \frac{1}{2} \int_0^{\frac{1}{2}\delta} \big(f(\alpha + h) + f(\alpha - h)\big)\, dh = \frac{1}{2} \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} f(x)\, dx$; hence (558).]

Using (558) with $\delta = \frac{1}{q}$ and $\alpha = \frac{1}{q}, \frac{2}{q}, \dots, \frac{q-1}{q}$ we get, from (557),

$$\Big|\sum_{n=M+1}^{M+N} \chi(n)\Big| \le q^{-\frac{1}{2}} \cdot q \int_{\frac{1}{2q}}^{\frac{2q-1}{2q}} \frac{1}{\sin(\pi x)}\,dx = 2q^{\frac{1}{2}} \int_{\frac{1}{2q}}^{\frac{1}{2}} \frac{1}{\sin(\pi x)}\,dx$$

Now $\sin(\pi x) > 2x$ for $0 < x < \frac{1}{2}$, so that the above is

$$< 2q^{\frac{1}{2}} \int_{\frac{1}{2q}}^{\frac{1}{2}} \frac{dx}{2x} = q^{\frac{1}{2}} \log q.$$

Hence we have proved

$$\Big|\sum_{n=M+1}^{M+N} \chi(n)\Big| < q^{\frac{1}{2}} \log q \tag{559}$$

for primitive χ.

Suppose now that $\chi \bmod q$ is not primitive, but still nonprincipal. Let $\chi_1 \bmod q_1$ be the corresponding primitive character. Then $q_1 \mid q$, and we write $q = q_1 r$. Hence

$$\sum_{n=M+1}^{M+N} \chi(n) = \sum_{\substack{n=M+1\\(n,r)=1}}^{M+N} \chi_1(n).$$

Now $\sum_{d|m} \mu(d) = 1$ or 0 according as $m = 1$ or $m > 1$ (cf. (740) in the solution to Problem 3.7), so that the above is

$$= \sum_{n=M+1}^{M+N} \chi_1(n) \sum_{d|(n,r)} \mu(d) = \sum_{d|r} \mu(d) \sum_{\substack{n=M+1\\d|n}}^{M+N} \chi_1(n) = \sum_{d|r} \mu(d)\chi_1(d) \sum_{\frac{M+1}{d} \le m \le \frac{M+N}{d}} \chi_1(m).$$

In view of (559), the inner sum has absolute value $< q_1^{\frac{1}{2}} \log q_1$, so that

$$\Big|\sum_{n=M+1}^{M+N} \chi(n)\Big| < q_1^{\frac{1}{2}} (\log q_1) \sum_{d|r} \big|\mu(d)\big| \le q_1^{\frac{1}{2}} (\log q_1) \sum_{d|r} 1 \le q_1^{\frac{1}{2}} (\log q_1) \cdot 2 \sum_{\substack{d|r\\d\le\sqrt{r}}} 1$$

$$\le q_1^{\frac{1}{2}} (\log q_1) \cdot 2\sqrt{r} = 2q^{\frac{1}{2}} \log q_1 \le 2q^{\frac{1}{2}} \log q, \tag{560}$$

where we used the fact that $d \mapsto \frac{r}{d}$ gives a bijection between the set of divisors $d \mid r$ with $d > \sqrt{r}$ and the set of divisors $d \mid r$ with $d < \sqrt{r}$. This completes the proof. □

Remark 17.1. In (560) we proved the bound $\sum_{d|r} |\mu(d)| \le 2\sqrt{r}$, but of course one can give much sharper bounds as $r \to \infty$. Note that $\sum_{d|r} |\mu(d)| = 2^{\omega(r)}$, where where $\omega(r)$ is the number of distinct primes in the prime factorization of r. We have the bound $\omega(r) \ll \frac{\log r}{\log\log r}$ for all $r \ge 3$ (cf. Problem 17.1 below); hence $2^{\omega(r)} \le \exp\big(C\frac{\log r}{\log\log r}\big) = r^{\frac{C}{\log\log r}}$ where $C > 0$

is some absolute constant, and in particular for any fixed $\varepsilon > 0$ we have $2^{\omega(r)} \ll r^{\varepsilon}$ $(\forall r \geq 1)$ where the implied constant depends on ε. For our purpose in (560) it is most natural to use this fact with the fairly large choice $\varepsilon = \frac{1}{2}$; and as seen in (560) we then get a good numerical value ("2") for the implied constant.

Remark 17.2. If we ask for a bound on $\left|\sum_{n=M+1}^{M+N} \chi(n)\right|$ which only depends on q then Theorem 17.1 is close to best possible, and modulo GRH the optimal bound is in fact $\sqrt{q}\log\log q$. More precisely: Schur has proved that

$$\max_N \Big|\sum_{n\leq N} \chi(n)\Big| > \frac{1}{2\pi}\sqrt{q}$$

for all primitive $\chi \bmod q$, and Paley (1932) has proved that

$$\max_N \Big|\sum_{n\leq N} \left(\frac{d}{n}\right)\Big| > \tfrac{1}{7}\sqrt{d}\log\log d$$

for infinitely many quadratic discriminants $d > 0$. In the opposite direction Montgomery and Vaughan [46] have shown that, *assuming the GRH*,

$$\Big|\sum_{n=M+1}^{M+N} \chi(n)\Big| \ll \sqrt{q}\log\log q$$

for all nonprincipal characters $\chi \bmod q$.

Remark 17.3. However, if we allow the bound to also depend on the length of the sum, N, then there are useful, better bounds, obtained by Burgess [8] (cf. also [36, Ch. 12]). For example Burgess proved that

$$\Big|\sum_{n=M+1}^{M+N} \chi(n)\Big| \ll_{\varepsilon} N^{\frac{1}{2}} q^{\frac{3}{16}+\varepsilon}$$

for any nonprincipal $\chi \bmod q$.

17.1. **Problems.**

Problem 17.1. Let $\omega(q)$ be the number of distinct primes in the prime factorization of q.
(a). Prove that $\omega(q) \ll \frac{\log q}{\log\log q}$ for all $q \geq 3$.
(b). Prove that the bound in (a) is the best possible, i.e. $\limsup_{q\to\infty} \frac{\omega(q)\log\log q}{\log q} > 0$.

Problem 17.2. (= Problem 18.5(a), home assignment.) Let $d(n)$ be the number of divisors of n. Prove that for any $\varepsilon > 0$ there is some constant $C(\varepsilon) > 0$ such that $d(n) \leq C(\varepsilon)n^{\varepsilon}$ for all $n \geq 1$.

18. Further prime number sums

(Davenport Chapter 24.)

In 1937 Vinogradov (see [74, Ch. IX]) introduced a method for estimating sums $\sum_{p\le N} f(p)$ in which f is oscillatory but not multiplicative. His starting point was the following. Let $P = \prod_{p\le N^{\frac12}} p$. Then for n in the range $1 \le n \le N$ we have $(n,P) = 1$ if and only if $n = 1$ or n is a prime number in the interval $N^{\frac12} < n \le N$. [This fact is what the sieve of Eratosthenes is based on.] Hence

$$f(1) + \sum_{N^{\frac12}<p\le N} f(p) = \sum_{\substack{n\le N\\(n,P)=1}} f(n) = \sum_{\substack{t|P\\t\le N}} \mu(t) \sum_{r\le N/t} f(rt). \tag{561}$$

We stress that in "$\sum_{\substack{n\le N\\(n,P)=1}}$" and "$\sum_{r\le N/t}$" it is implicitly understood that the summation variable (n and r, respectively) runs through all *positive integers* satisfying the stated condition; this convention about all summation variables being *positive integers* is used several times below (and has been used in previous sections).

[Proof of the last identity in (561): (This can be seen as an application of the inclusion-exclusion principle.) Given n with $1 \le n \le N$, the total factor of "$f(n)$" in the last double sum is: $\sum_{\substack{t|P\\t|n\\t\le N}} \mu(t) = \sum_{t|(n,P)} \mu(t)$, which equals 1 if $(n,P) = 1$, otherwise 0 (cf. (740) in the solution of Problem 3.7); hence the last double sum equals $\sum_{\substack{n\le N\\(n,P)=1}} f(n)$, as desired.]

Thus we are led to bound sums of the kind $\sum_{r\le N/t} f(rt)$. We need to show that these sums are small. However, we cannot hope to get much cancellation when t is nearly as large as N, for then the sum contains few terms. Therefore Vinogradov rearranged the terms arising from $t \mid P$, $\delta N \le t \le N$, but this entailed great complications. Later Vaughan [72] found a new version of Vinogradov's method in which the details are much simpler; this is the method which we will present here.

The sum which we will actually treat is $\sum_{n\le N} f(n)\Lambda(n)$, and not $\sum_{p\le N} f(p)$! However, it is not difficult to carry over bounds on one of these sums to bounds on the other; cf. Problem 18.1 below for an example.

Vaughan's method leads to the following fundamental bound on $\sum_{n\le N} f(n)\Lambda(n)$.

Proposition 18.1. *For any function $f : \mathbb{Z}^+ \to \mathbb{C}$ and any real numbers $N, U, V \geq 2$ with $N \geq UV$, we have*

$$\Big|\sum_{n\leq N} f(n)\Lambda(n)\Big| \ll \sum_{n\leq U}\big|f(n)\big|\Lambda(n) + (\log UV)\sum_{t\leq UV}\Big|\sum_{r\leq N/t} f(rt)\Big|$$

$$+ (\log N)\sum_{d\leq V}\max_{1\leq w\leq N/d}\Big|\sum_{1\leq h\leq w} f(dh)\Big| + N^{\frac{1}{2}}(\log N)^3 \max_{U\leq M\leq N/V}\Delta(f,M,N,V), \tag{562}$$

where $\Delta(f, M, N, V)$ denotes any non-negative real number satisfying

$$\Big|\sum_{M<m\leq 2M} b_m \sum_{V<k\leq N/m} c_k f(mk)\Big| \leq \Delta(f,M,N,V)\Big(\sum_{M<m\leq 2M}|b_m|^2\Big)^{\frac{1}{2}}\Big(\sum_{k\leq N/M}|c_k|^2\Big)^{\frac{1}{2}} \tag{563}$$

for all complex numbers b_m, c_k.

Proof. For given $U, V \geq 2$ we let

$$F(s) = \sum_{m\leq U}\Lambda(m)m^{-s}, \qquad G(s) = \sum_{d\leq V}\mu(d)d^{-s}, \tag{564}$$

and note the identity (for $\sigma > 1$)

$$-\frac{\zeta'(s)}{\zeta(s)} = F(s) - \zeta(s)F(s)G(s) - \zeta'(s)G(s) + \Big(-\frac{\zeta'(s)}{\zeta(s)} - F(s)\Big)\cdot(1-\zeta(s)G(s)). \tag{565}$$

Calculating the Dirichlet series coefficients of the four functions on the right-hand side, we get

$$\Lambda(n) = a_1(n) + a_2(n) + a_3(n) + a_4(n), \tag{566}$$

where

$$a_1(n) = \begin{cases}\Lambda(n) & \text{if } n \leq U\\ 0 & \text{if } n > U;\end{cases} \tag{567}$$

$$a_2(n) = -\sum_{\substack{mdr=n\\ m\leq U\\ d\leq V}}\Lambda(m)\mu(d); \tag{568}$$

$$a_3(n) = \sum_{\substack{hd=n\\ d\leq V}}\mu(d)\log h; \tag{569}$$

$$a_4(n) = -\sum_{\substack{mk=n\\ m>U\\ k>1}}\Lambda(m)\Big(\sum_{\substack{d|k\\ d\leq V}}\mu(d)\Big). \tag{570}$$

[Proof of (567)–(570): (567) is clear by definition. For the other three we use the following general formula for multiplication of Dirichlet series, in the half plane where both series

are absolutely convergent:

$$\Big(\sum_{n=1}^{\infty}\alpha_n n^{-s}\Big)\Big(\sum_{n=1}^{\infty}\beta_n n^{-s}\Big)=\sum_{n=1}^{\infty}\delta_n n^{-s},\qquad \text{where } \delta_n=\sum_{md=n}\alpha_m\beta_d. \tag{571}$$

(To see this formula one first notices that the left hand side equals $\sum_{m=1}^{\infty}\sum_{d=1}^{\infty}\alpha_m\beta_d(md)^{-s}$, and then for each $n\ge 1$ collect all terms with $md=n$.) Using (571) and $\zeta'(s)=-\sum_{n=1}^{\infty}(\log n)n^{-s}$ we get (569). Also using (571) and $-\frac{\zeta'(s)}{\zeta(s)}-F(s)=\sum_{m>U}\Lambda(m)m^{-s}$ and $1-\zeta(s)G(s)=1-\sum_{k=1}^{\infty}\Big(\sum_{\substack{dm=k\\ d\le V}}\mu(d)\Big)k^{-s}=-\sum_{k=2}^{\infty}\Big(\sum_{\substack{d|k\\ d\le V}}\mu(d)\Big)k^{-s}$ we get (570). Finally using (571) we also have the general formula for a triple product;

$$\Big(\sum_{n=1}^{\infty}\alpha_n n^{-s}\Big)\Big(\sum_{n=1}^{\infty}\beta_n n^{-s}\Big)\Big(\sum_{n=1}^{\infty}\gamma_n n^{-s}\Big)=\sum_{n=1}^{\infty}\delta_n n^{-s},\qquad \text{where } \delta_n=\sum_{mdr=n}\alpha_m\beta_d\gamma_r,$$

and using this we immediately get (568).]

We multiply throughout by $f(n)$ and sum; then

$$\sum_{n\le N}f(n)\Lambda(n)=S_1+S_2+S_3+S_4, \tag{572}$$

where

$$S_i=\sum_{n\le N}f(n)a_i(n). \tag{573}$$

Note that the first term in the right hand side of (562) is just the trivial bound on $\big|S_1\big|$.

We write S_2 in the form

$$\begin{aligned}S_2=-\sum_{n\le N}\sum_{\substack{mdr=n\\ m\le U\\ d\le V}}f(n)\Lambda(m)\mu(d)&=-\sum_{\substack{mdr\le N\\ m\le U\\ d\le V}}f(mdr)\Lambda(m)\mu(d)\\ &=-\sum_{\substack{m\le U\\ d\le V}}\Lambda(m)\mu(d)\sum_{r\le N/(md)}f(rmd)\\ &=-\sum_{t\le UV}\Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big)\sum_{r\le N/t}f(rt). \qquad (574)\end{aligned}$$

Again we have a linear combination of the sums $\sum_{r\le N/t}f(rt)$ but now we can control the range of t by ensuring that UV is substantially smaller than N. Using

$$\Big|\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big|\le\sum_{m|t}\Lambda(m)\le\log t\le\log UV$$

for $t \le UV$ we get

$$\big|S_2\big| \le (\log UV) \sum_{t\le UV} \Big| \sum_{r\le N/t} f(rt)\Big|. \tag{575}$$

We next treat S_3:

$$S_3 = \sum_{n\le N} f(n) \sum_{\substack{hd=n\\ d\le V}} \mu(d)\log h = \sum_{\substack{hd\le N\\ d\le V}} f(hd)\mu(d)\log h = \sum_{d\le V}\mu(d) \sum_{h\le N/d} f(hd)\log h. \tag{576}$$

We again wish to express this in terms of sums $\sum_{h\le x} f(hd)$, which we can hope to bound. Let us introduce the (temporary) notation $F_d(w) := \sum_{h\le w} f(hd)$. Then by integration by parts:

$$\sum_{h\le N/d} f(hd)\log h = \int_{1-}^{N/d} (\log w)\, dF_d(w) = \Big[(\log w)F_d(w)\Big]_{w=1-}^{w=N/d} - \int_1^{N/d} \frac{F_d(w)}{w}\, dw$$

and hence

$$\begin{aligned}\Big|\sum_{h\le N/d} f(hd)\log h\Big| &\le \log\Big(\frac{N}{d}\Big)\Big|F_d\Big(\frac{N}{d}\Big)\Big| + \Big(\sup_{1\le w\le N/d}\big|F_d(w)\big|\Big)\int_1^{N/d}\frac{dw}{w}\\ &\le 2(\log N)\max_{1\le w\le N/d}\Big|\sum_{1\le h\le w} f(hd)\Big|.\end{aligned}$$

(Here we only have to let w run through the *integers* in the interval $[1, N/d]$, since $F_d(w) = \sum_{1\le h\le w} f(hd)$ only depends on the integer part of w; hence there are only finitely many choices of w (for given N, d) and thus we can certainly write "max" in place of "sup".) From this we conclude that

$$\big|S_3\big| \ll (\log N)\sum_{d\le V}\max_{1\le w\le N/d}\Big|\sum_{1\le h\le w} f(dh)\Big|. \tag{577}$$

Remark 18.1. Note that (577) says exactly the same thing as the bound in Davenport's book, p. 140(3);

$$\big|S_3\big| \ll (\log N)\sum_{d\le V}\max_{1\le w\le N/d}\Big|\sum_{w\le h\le N/d} f(dh)\Big|.$$

[Proof: Note that for each integer w with $1 \le w \le N/d$ we have

$$\begin{aligned}\Big|\sum_{1\le h\le w} f(dh)\Big| = \Big|\sum_{1\le h\le N/d} f(dh) - \sum_{w<h\le N/d} f(dh)\Big| &\le \Big|\sum_{1\le h\le N/d} f(dh)\Big| + \Big|\sum_{w+1\le h\le N/d} f(dh)\Big|\\ &\le 2\max_{w'}\Big|\sum_{w'\le h\le N/d} f(dh)\Big|.\end{aligned}$$

Hence

$$\max_{w}\Big|\sum_{1\le h\le w} f(dh)\Big| \le 2\max_{w'}\Big|\sum_{w'\le h\le N/d} f(dh)\Big|.$$

Similarly

$$\max_{w'}\Big|\sum_{w'\le h\le N/d} f(dh)\Big| \le 2\max_{w}\Big|\sum_{1\le h\le w} f(dh)\Big|,$$

and this completes the proof.]

Finally we treat S_4, which is the most complicated sum:

$$S_4 = \sum_{n\le N} f(n)a_4(n) = -\sum_{n\le N} f(n)\sum_{\substack{mk=n\\ m>U\\ k>1}} \Lambda(m)\Big(\sum_{\substack{d|k\\ d\le V}}\mu(d)\Big).$$

Note here that $\sum_{\substack{d|k\\ d\le V}}\mu(d) = 0$ whenever $1 < k \le V$ (compare (740) in the solution of Problem 3.7); hence

(578)
$$S_4 = -\sum_{n\le N} f(n)\sum_{\substack{mk=n\\ m>U\\ k>V}} \Lambda(m)\Big(\sum_{\substack{d|k\\ d\le V}}\mu(d)\Big) = -\sum_{U<m\le N/V}\Lambda(m)\sum_{V<k\le N/m}\Big(\sum_{\substack{d|k\\ d\le V}}\mu(d)\Big)f(mk).$$

We apply dyadic decomposition in the m-variable:

$$S_4 = \sum_{\substack{M\in\{2^0U,2^1U,2^2U,\dots\}\\ M<N/V}}\Big\{\sum_{M<m\le\min(N/V,2M)}\Lambda(m)\sum_{V<k\le N/m}\Big(\sum_{\substack{d|k\\ d\le V}}\mu(d)\Big)f(mk)\Big\}, \tag{579}$$

and for each M we view the expression within the brackets as a bilinear form, $\langle (b_m),(c_k)\rangle \mapsto \sum_{m,k} f(mk)b_mc_k$. Now as in the statement of the proposition (cf. (563)) we assume that $\Delta = \Delta(f,M,N,V) \ge 0$ is such that

$$\Big|\sum_{M<m\le 2M} b_m \sum_{V<k\le N/m} c_k f(mk)\Big| \le \Delta\Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{k\le N/M}|c_k|^2\Big)^{\frac12} \tag{580}$$

for any complex numbers b_m, c_k. Then from (579), by applying (580) with

$$b_m = \begin{cases}\Lambda(m) & \text{if } m\le N/V\\ 0 & \text{else}\end{cases} \qquad\text{and}\qquad c_k = \sum_{\substack{d|k\\ d\le V}}\mu(d),$$

and using $\big|c_k\big| \le \sum_{d|k} 1 = d(k)$ where $d(k)$ is the divisor function defined in Problem 3.9(a), we get

$$\big|S_4\big| \ll (\log N)\max_{U\le M\le N/V}\Delta(f,M,N,V)\cdot\Big(\sum_{M<m\le 2M}\Lambda(m)^2\Big)^{\frac12}\Big(\sum_{k\le N/M} d(k)^2\Big)^{\frac12}.$$

Here the sum over m is bounded by noting that

$$\sum_{m\le z}\Lambda(m)^2 \le (\log z)\sum_{m\le z}\Lambda(m) \ll z\log z, \qquad \forall z\ge 2. \tag{581}$$

For the sum over k we have

$$\sum_{k\le z} d(k)^2 \ll z(\log 2z)^3, \qquad \forall z\ge 1, \tag{582}$$

cf. Problems 18.4, 18.5 below (where we also show that this bound is best possible). Combining these estimates, we see that

$$\big|S_4\big| \ll N^{\frac12}(\log N)^3 \max_{U\le M\le N/V}\Delta(f,M,N,V). \tag{583}$$

Combining (572) and (573) with (575), (577) and (583), we obtain (562). □

Remark 18.2. In some situations sharper estimates can be obtained by treating S_2 more carefully: Write

$$S_2 = \sum_{t\le UV} = \sum_{t\le U} + \sum_{U<t\le UV} = S_2' + S_2''.$$

Then treat S_2' as we did S_2, and S_2'' as we did S_4. This method will be used in the proof of Bombieri's Theorem, in §21.

Let us now discuss what is needed for the bound in Proposition 18.1 to be non-trivial. To be specific, let us from now on suppose that $|f(n)|\le 1$ for all n. The *trivial* bound on $\sum_{n\le N} f(n)\Lambda(n)$ then is

$$\Big|\sum_{n\le N} f(n)\Lambda(n)\Big| \le \sum_{n\le N}\Lambda(n) \ll N,$$

and we would like the bound in Proposition 18.1 to be asymptotically better than this, viz. to be $= o(N)$ as $N\to\infty$! The first term in (562) is unproblematic so long as $U = o(N)$, since $\sum_{n\le U}\big|f(n)\big|\Lambda(n) \le \sum_{n\le U}\Lambda(n) \ll U$. Next, the *trivial* estimate on the second term in (562) is

$$(\log UV)\sum_{t\le UV}\Big|\sum_{r\le N/t} f(rt)\Big| \le (\log UV)\sum_{t\le UV}\frac{N}{t} \ll N(\log UV)^2;$$

hence we do not require much cancellation in the sums $\sum_{t\le N/t} f(rt)$ to show that second term in (562) is $o(N)$. Similarly, the trivial estimate on the third term in (562) is

$$(\log N)\sum_{d\le V}\max_{1\le w\le N/d}\Big|\sum_{1\le h\le w} f(dh)\Big| \le (\log N)\sum_{d\le V}\frac{N}{d} \ll N(\log N)(\log V),$$

and again we see that we do not require much cancellation in the sums $\sum_h f(dh)$ to get $o(N)$. Finally regarding the last term in (562) we point out that a *trivial* choice of

$\Delta(f, M, N, V)$ (when $|f(n)| \le 1$ for all n) is $\Delta(f, M, N, V) = (2N)^{\frac12}$. Indeed; this is seen to be ok by Cauchy's inequality: For all choices of complex numbers b_m, c_k we have

$$\begin{aligned}\Big|\sum_{M<m\le 2M} b_m \sum_{V<k\le N/m} c_k f(mk)\Big| &\le \Big(\sum_{M<m\le 2M}|b_m|\Big)\cdot\Big(\sum_{V<k\le N/M}|c_k|\Big)\\ &\le (2M)^{\frac12}\Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\cdot\Big(\frac{N}{M}\Big)^{\frac12}\Big(\sum_{V<k\le N/M}|c_k|^2\Big)^{\frac12},\\ &= (2N)^{\frac12}\Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{V<k\le N/M}|c_k|^2\Big)^{\frac12},\end{aligned}$$

since the number of integers m with $M < m \le 2M$ is $\le 2M$. Using this choice of $\Delta(f, M, N, V)$ the last term in (562) is

$$N^{\frac12}(\log N)^3 \max_{U\le M\le N/V}\Delta(f, M, N, V) \ll N(\log N)^3;$$

thus we need only a slightly sharper bound for Δ to get $o(N)$.

Remark 18.3. However, let us note that if f is totally multiplicative and unimodular (viz. $\big|f(n)\big| = 1$ for all n) then we *cannot* improve on the bound $\Delta(f, M, N, V) \ll N^{\frac12}$. Indeed, if we choose $b_m = \overline{f(m)}$, $c_k = \overline{f(k)}$ we get, assuming $M < \frac{N}{3V}$, say,

$$\begin{aligned}\sum_{M<m\le 2M} b_m \sum_{V<k\le N/m} c_k f(mk) &= \sum_{M<m\le 2M}\overline{f(m)}\sum_{V<k\le N/m}\overline{f(k)}f(m)f(k)\\ &= \sum_{M<m\le 2M}\big|f(m)\big|^2\sum_{V<k\le N/m}\big|f(k)\big|^2 = \sum_{M<m\le 2M}\sum_{V<k\le N/m}1\\ &\ge \sum_{M<m\le 2M}\sum_{\frac{N}{3M}<k\le\frac{N}{2M}}1 \gg M\cdot\frac{N}{M} = N,\end{aligned}$$

while

$$\Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{V<k\le N/M}|c_k|^2\Big)^{\frac12} \ll M^{\frac12}\Big(\frac{N}{M}\Big)^{\frac12} = N^{\frac12};$$

so that we are forced to take $\Delta(f, M, N, V) \gg N^{\frac12}$. For this reason the principal applications of the method described in this section involve functions f which are *not* multiplicative.

For most functions f we are not able to determine the optimal choice of $\Delta(f, M, N, V)$. However, in the following proposition we give a very useful approach to finding a "good" Δ.

Proposition 18.2. *If $|f(n)| \le 1$ for all n, then for any real numbers $N, U, V \ge 2$ with $N \ge UV$, we have*

$$\Big|\sum_{n\le N} f(n)\Lambda(n)\Big| \ll U + (\log N)\sum_{t\le UV}\max_{w}\Big|\sum_{w\le r\le N/t} f(rt)\Big| + N^{\frac12}(\log N)^3 \max_{U\le M\le N/V}\max_{V\le j\le N/M}\Big(\sum_{V<k\le N/M}\Big|\sum_{\substack{M<m\le 2M\\ m\le N/k\\ m\le N/j}} f(mj)\overline{f(mk)}\Big|\Big)^{\frac12}. \tag{584}$$

Proof. By Cauchy's inequality we always have

$$\begin{aligned}\Big|\sum_{M<m\le 2M} b_m \sum_{V<k\le N/m} c_k f(mk)\Big| &\le \Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{M<m\le 2M}\Big|\sum_{V<k\le N/m} c_k f(mk)\Big|^2\Big)^{\frac12}\\ &= \Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{V<j\le N/M} c_j \sum_{V<k\le N/M}\overline{c_k}\sum_{\substack{M<m\le 2M\\ m\le N/j\\ m\le N/k}} f(mj)\overline{f(mk)}\Big)^{\frac12}\\ &\ll \Big(\sum_{M<m\le 2M}|b_m|^2\Big)^{\frac12}\Big(\sum_{V<j\le N/M}|c_j|^2\sum_{V<k\le N/M}\Big|\sum_{\substack{M<m\le 2M\\ m\le N/j\\ m\le N/k}} f(mj)\overline{f(mk)}\Big|\Big)^{\frac12},\end{aligned}$$

where in the last step we used $\big|c_j\overline{c_k}\big| \le \frac12|c_j|^2 + \frac12|c_k|^2$ and the symmetry between j and k. It follows that we can take

$$\Delta(f,M,N,V) \ll \max_{V\le j\le N/M}\Big(\sum_{V<k\le N/M}\Big|\sum_{\substack{M<m\le 2M\\ m\le N/k\\ m\le N/j}} f(mj)\overline{f(mk)}\Big|\Big)^{\frac12}.$$

Hence the fourth term in (562) is covered by the second line of (584). Next we note that both the second and the third term in (562) is subsumed by the second term in (584) (recall Remark 18.1 regarding the third term in (562)).

Finally note that, using $|f(n)| \le 1$, the first term in (562) is $\ll \sum_{n\le U}\Lambda(n) \ll U$. (This is the only place where we need the assumption $|f(n)| \le 1$!) This completes the proof of (562). □

18.1. **Example: An exponential sum formed with primes.** (Chapter 25 in Davenport.)

Vinogradov first used his method to estimate the important sum

$$S(\alpha) = \sum_{n\le N}\Lambda(n)e(n\alpha). \tag{585}$$

Bounds on this sum are of fundamental importance in the proof of Vinogradov's 3-primes theorem, cf. the next section. We will use our general estimates of the previous section to bound $S(\alpha)$. It turns out that the result depends on rational approximations of α:

Proposition 18.3. *If $\alpha \in \mathbb{R}$ and*

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{q^2} \qquad (q \in \mathbb{Z}^+,\ a \in \mathbb{Z},\ (a,q)=1), \tag{586}$$

then

$$\left|S(\alpha)\right| \ll (Nq^{-\frac{1}{2}} + N^{\frac{4}{5}} + N^{\frac{1}{2}}q^{\frac{1}{2}})(\log N)^4, \tag{587}$$

where the implied constant is absolute.

Remark 18.4. For any given real number $\alpha \in \mathbb{R}$ there does exist a rational approximation $\frac{a}{q}$ satisfying (586), and if α is irrational then there actually exists an infinite sequence of distinct such rational approximations. This follows from the following well-known theorem by Dirichlet on Diophantine approximation.

Lemma 18.4. *For every $\alpha \in \mathbb{R}$ and every real $Q \geq 1$, there is a rational number $\frac{a}{q}$ such that*

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qQ}; \qquad 1 \leq q \leq Q; \qquad (a,q) = 1.$$

Proof. For each $j = 0, 1, 2, \ldots, \lfloor Q \rfloor$ we pick $\beta_j \in [0,1)$ so that $j\alpha \equiv \beta_j \pmod 1$. Let $\beta'_0, \beta'_1, \ldots, \beta'_{\lfloor Q \rfloor}$ be a permutation of $\beta_0, \beta_1, \ldots, \beta_{\lfloor Q \rfloor}$ such that $0 = \beta'_0 \leq \beta'_1 \leq \ldots \leq \beta'_{\lfloor Q \rfloor} < 1$; then since $\sum_{j=1}^{\lfloor Q \rfloor}(\beta'_j - \beta'_{j-1}) + (1 - \beta'_{\lfloor Q \rfloor}) = 1$, at least one of the differences $\beta'_j - \beta'_{j-1}$ ($j \in \{1, 2, \ldots, \lfloor Q \rfloor\}$) and $1 - \beta'_{\lfloor Q \rfloor}$ must be $\leq \frac{1}{\lfloor Q \rfloor + 1} < \frac{1}{Q}$. Hence either there are $0 \leq j < k \leq \lfloor Q \rfloor$ such that $|\beta_j - \beta_k| < \frac{1}{Q}$ or else there is some $0 \leq j \leq \lfloor Q \rfloor$ such that $|1 - \beta_j| < \frac{1}{Q}$; in fact in the second case we must have $j \geq 1$ since $\beta_0 = 0$. In the first case it follows that $\big\|(k-j)\alpha\big\| < \frac{1}{Q}$ (where $\|\beta\|$ denotes the distance from β to the nearest integer), and in the second case it follows that $\big\|j\alpha\big\| < \frac{1}{Q}$; hence there is some q with $1 \leq q \leq \lfloor Q \rfloor$ such that $\big\|q\alpha\big\| < \frac{1}{Q}$. This means that there is an integer a such that $\big|q\alpha - a\big| < \frac{1}{Q}$, viz. $\left|\alpha - \frac{a}{q}\right| < \frac{1}{qQ}$. Finally if $d = (a,q) > 1$ then we can replace $\langle a, q\rangle$ with $\langle a/d, q/d\rangle$ and then $\left|\alpha - \frac{a}{q}\right| < \frac{1}{qQ}$ continues to hold while also $(a,q) = 1$. □

Remark 18.5. Returning to Proposition 18.3, we note that the *trivial* bound on $S(\alpha)$ is $\big|S(\alpha)\big| \leq \sum_{n \leq N} \Lambda(n) \ll N$. Hence Proposition 18.3 gives a *power saving* versus the trivial bound for any α for which we can find a rational approximation $\frac{a}{q}$ to α satisfying (586) and $N^{\varepsilon} \leq q \leq N^{1-\varepsilon}$. (Here ε is some fixed small positive constant.)

Remark 18.6. * (External reading.) To further appreciate the error term in Proposition 18.3 we mention a basic concept from diophantine analysis: An irrational number $\alpha \in \mathbb{R}$ is said

to be of *(diophantine) type* κ if there is a constant $C > 0$ such that

$$\Big|\alpha - \frac{a}{q}\Big| > \frac{C}{q^\kappa}, \qquad \text{for all } \ a \in \mathbb{Z},\ q \in \mathbb{Z}^+. \tag{588}$$

The smallest possible value of κ is $\kappa = 2$ (by Lemma 18.4), and in fact, for any given $\kappa > 2$, the set of α's of type κ is of full Lebesgue measure in $\mathbb{R}$. Cf. Problem 18.2 below. [Also, by the deep Thue-Siegel-Roth theorem, if α is irrational and *algebraic*, then α is of type κ for any $\kappa > 2$.] Now if α is fixed and of diophantine type κ, then one can prove that Proposition 18.3 implies $\big|S(\alpha)\big| \ll N^\tau$ as $N \to \infty$, where the exponent τ is any fixed number with

$$\tau > \max\Big(\frac{4}{5}, \frac{2\kappa - 1}{2\kappa}\Big). \tag{589}$$

(Cf. Problem 18.3 below.) In particular, for (Lebesgue-)almost every $\alpha \in \mathbb{R}$ we have $\big|S(\alpha)\big| \ll N^{\frac{4}{5}+\varepsilon}$ as $N \to \infty$, for any $\varepsilon > 0$; and by Thue-Siegel-Roth this holds in particular when α is irrational and algebraic (e.g. when $\alpha = \sqrt[3]{2}$). We also see that we get a power saving versus the trivial bound for *any* α which has a diophantine type. (The complement set, viz. the set of irrational α's which *don't* have a diophantine type are called the *Liouville numbers*; this set is quite "small" in the sense that it has "Hausdorff dimension 0" – although the Liouville numbers are still uncountably many, and form a dense set in $\mathbb{R}$.)

Proof of Proposition 18.3. The bound is trivial if $N < 10$; hence from now on we assume $N \geq 10$. By Proposition 18.2 we have, for any $U, V \geq 2$ with $UV \leq N$:

$$\begin{aligned} \big|S(\alpha)\big| = \Big|\sum_{n\leq N} \Lambda(n) e(n\alpha)\Big| &\ll U + (\log N) \sum_{t \leq UV} \max_w \Big| \sum_{w \leq r \leq N/t} e(rt\alpha)\Big| \\ &+ N^{\frac12} (\log N)^3 \max_{U \leq M \leq N/V} \max_{V \leq j \leq N/M} \Big(\sum_{V < k \leq N/M} \Big| \sum_{\substack{M < m \leq 2M \\ m \leq N/k \\ m \leq N/j}} e(m(j-k)\alpha)\Big|\Big)^{\frac12}. \end{aligned} \tag{590}$$

We note the following general bound, for any $\beta \in \mathbb{R}$ and any integers $N_1 < N_2$,

$$\Big|\sum_{n=N_1}^{N_2} e(n\beta)\Big| = \Big|\frac{e((N_2+1)\beta) - e(N_1\beta)}{e(\beta) - 1}\Big| \ll \min\Big(N_2 - N_1, \frac{1}{\|\beta\|}\Big), \tag{591}$$

where $\|\beta\|$ denotes the distance from β to the nearest integer. Here the first bound $N_2 - N_1$ follows simply since the number of terms in $\sum_{n=N_1}^{N_2}$ is $N_2 - N_1 + 1 \ll N_2 - N_1$, and the second bound, $\frac{1}{\|\beta\|}$, follows since $\big|e((N_2+1)\beta) - e(N_1\beta)\big| \leq 2$ and $\big|e(\beta) - 1\big| \gg \|\beta\|$. Using (591) we can bound the second term in (590) as follows:

$$(\log N) \sum_{t \leq UV} \max_w \Big| \sum_{w \leq r \leq N/t} e(rt\alpha)\Big| \ll (\log N) \sum_{t \leq UV} \min\Big(\frac{N}{t}, \frac{1}{\|t\alpha\|}\Big) \tag{592}$$

We give a lemma on how to bound this type of sum:

Lemma 18.5. *If $\alpha \in \mathbb{R}$ and*

$$\Big|\alpha - \frac{a}{q}\Big| \le \frac{1}{q^2} \qquad (q \in \mathbb{Z}^+,\ a \in \mathbb{Z},\ (a,q)=1), \tag{593}$$

then for any $N, T \ge 1$ we have

$$\sum_{t \le T} \min\Big(\frac{N}{t}, \frac{1}{\|t\alpha\|}\Big) \ll \Big(\frac{N}{q} + T + q\Big)\log(2qT). \tag{594}$$

We postpone the proof lemma. Using the lemma we get from (592):

$$\begin{aligned}(\log N)\sum_{t \le UV} \max_{w} \Big|\sum_{w \le r \le N/t} e(rt\alpha)\Big| &\ll (\log N)\Big(\frac{N}{q} + UV + q\Big)\log(2qUV) \\ &\le \Big(\frac{N}{q} + UV + q\Big)\big(\log 2qN\big)^2. \end{aligned} \tag{595}$$

Next, the *last* term in (590) is, using (591),

$$\ll N^{\frac12}(\log N)^3 \max_{U \le M \le N/V} \max_{V \le j \le N/M} \Big(\sum_{V < k \le N/M} \min\Big(M, \frac{1}{\|(k-j)\alpha\|}\Big)\Big)^{\frac12}.$$

Here note that for each integer j in the interval $V \le j \le N/M$ we have

$$\begin{aligned}\sum_{V < k \le N/M} \min\Big(M, \frac{1}{\|(k-j)\alpha\|}\Big) &\le \sum_{-N/M \le m \le N/M} \min\Big(M, \frac{1}{\|m\alpha\|}\Big) \\ &\le M + 2\sum_{1 \le m \le N/M} \min\Big(\frac{N}{m}, \frac{1}{\|m\alpha\|}\Big).\end{aligned}$$

Hence the last term in (590) is in fact

$$\begin{aligned}&\ll N^{\frac12}(\log N)^3 \max_{U \le M \le N/V} \Big(M + \sum_{1 \le m \le N/M} \min\Big(\frac{N}{m}, \frac{1}{\|m\alpha\|}\Big)\Big)^{\frac12} \\ &\ll N^{\frac12}(\log N)^3 \max_{U \le M \le N/V} \Big(M + \frac{N}{q} + \frac{N}{M} + q\Big)^{\frac12}\big(\log qN\big)^{\frac12} \\ &\ll \big(NV^{-\frac12} + Nq^{-\frac12} + NU^{-\frac12} + N^{\frac12}q^{\frac12}\big)\big(\log qN\big)^4,\end{aligned}$$

where we again used Lemma 18.5. Using this bound together with (595) we get from (590):

$$\big|S(\alpha)\big| \ll \big(UV + q + NV^{-\frac12} + Nq^{-\frac12} + NU^{-\frac12} + N^{\frac12}q^{\frac12}\big)\big(\log qN\big)^4$$

We now make the choice $U = V = N^{\frac25}$, and obtain

$$\big|S(\alpha)\big| \ll \big(N^{\frac45} + q + Nq^{-\frac12} + N^{\frac12}q^{\frac12}\big)\big(\log qN\big)^4.$$

If $q \le N$ then this implies the desired bound, (587). On the other hand if $q > N$ then (587) is trivial, since then $\big|S(\alpha)\big| \le \sum_{n\le N}\Lambda(n) \ll N \ll N^{\frac12}q^{\frac12}$. This completes the proof of the proposition. □

It remains to prove Lemma 18.5.

Proof of Lemma 18.5. Write $t = hq+r$ with $1 \le r \le q$ and put $\beta = \alpha - \frac{a}{q}$ (thus $|\beta| \le q^{-2}$). Then $t\alpha = (hq+r)(\beta+\frac{a}{q}) = ha + (ra/q + hq\beta + r\beta)$ with $h \in \mathbb{Z}$ and thus

$$\sum_{t\le T} \min\Big(\frac{N}{t}, \frac{1}{\|t\alpha\|}\Big) \le \sum_{0\le h\le T/q}\sum_{r=1}^{q} \min\Big(\frac{N}{hq+r}, \frac{1}{\|ra/q+hq\beta+r\beta\|}\Big).$$

We consider first those terms for which $h = 0, 1 \le r \le \frac12 q$. For these terms we have $|r\beta| \le \frac{1}{2q}$, so that the contribution of these terms is

(596)
$$\ll \sum_{1\le r\le q/2} \frac{1}{\Big\|\frac{ra}{q}\Big\| - \frac{1}{2q}} \le \sum_{\substack{m\in(\mathbb{Z}/q\mathbb{Z})\\ m\not\equiv 0 \,(\mathrm{mod}\ q)}} \frac{1}{\Big\|\frac{m}{q}\Big\| - \frac{1}{2q}} \le 2\sum_{1\le m\le q/2} \frac{1}{\frac{m}{q} - \frac{1}{2q}} \ll q\sum_{n=1}^{q-1}\frac{1}{n} \ll q\log(2q).$$

For all remaining terms we have $hq + r \gg (h+1)q$, and thus the contribution of these terms is

$$\ll \sum_{0\le h\le T/q}\sum_{r=1}^{q} \min\Big(\frac{N}{(h+1)q}, \frac{1}{\|ra/q+hq\beta+r\beta\|}\Big).$$

Now note that for fixed h, and for any given interval $I \subset \mathbb{R}$ of length q^{-1}, there are at most 4 values of r, $1 \le r \le q$, for which

$$\frac{ra}{q} + hq\beta + r\beta \in I + \mathbb{Z}. \tag{597}$$

(Notation: $A + B = \{a + b \,:\, a \in A, b \in B\}$, for any $A, B \subset \mathbb{R}$.) [Proof of the claim: Let J be the interval I translated by $-hq\beta$ and made q^{-1} longer at each end (in other words the interval is expanded by a factor of 3, from its central point). Then since $|r\beta| \le q^{-1}$ for all r with $1 \le r \le q$, (597) can only hold if $\frac{ra}{q} \in J + \mathbb{Z}$, viz. $ra \in qJ + q\mathbb{Z}$. Here qJ is an interval of length 3 and therefore there are at most 4 congruence classes $b \in \mathbb{Z}/q\mathbb{Z}$ for which $b \in qJ + q\mathbb{Z}$. Hence since $(a, q) = 1$, there are at most 4 integers r with $1 \le r \le q$ such that $ra \in qJ + q\mathbb{Z}$.]

We use this fact for the q intervals $I_j = [jq^{-1}, (j+1)q^{-1}]$, $j = 0, 1, 2, \dots, q-1$, and note that when $ra/q + hq\beta + r\beta$ belongs to $I_j + \mathbb{Z}$ we have $\big\|ra/q + hq\beta + r\beta\big\| \ge \min(\frac{j}{q}, 1 - \frac{j+1}{q})$,

and thus

$$\min\Big(\frac{N}{(h+1)q}, \frac{1}{\|ra/q+hq\beta+r\beta\|}\Big) \le \begin{cases} \frac{N}{(h+1)q} & \text{if } j=0 \text{ or } q-1 \\ \frac{q}{\min(j,q-j-1)} & \text{else.} \end{cases}$$

Hence

$$\sum_{0\le h\le T/q}\sum_{r=1}^{q}\min\Big(\frac{N}{(h+1)q}, \frac{1}{\|ra/q+hq\beta+r\beta\|}\Big) \le 8\sum_{0\le h\le T/q}\Big(\frac{N}{(h+1)q}+\sum_{j=1}^{\lceil q/2\rceil}\frac{q}{j}\Big)$$

$$\ll \frac{N}{q}\log(2T)+\Big(\frac{T}{q}+1\Big)q\log(2q) \tag{598}$$

Adding the bounds in (596) and (598) we obtain the bound claimed in the lemma. □

18.2. *** Equidistribution of $p\alpha$ mod 1.** [I didn't get time to write this section yet, but I will hopefully mention it briefly in class.]

18.3. **Problems.**

Problem 18.1. Let $f:\mathbb{Z}^+\to\mathbb{C}$ be a given function satisfying $|f(n)|\le 1$ for all n, and set $S(X)=\sum_{n\le X}f(n)\Lambda(n)$ and $T(X)=\sum_{p\le X}f(p)$. Assume that

$$S(X)=O\big(X^a\big) \qquad \text{as } X\to\infty,$$

where a is some fixed constant, $\frac12<a<1$. Then prove that

$$T(X)=O\big(X^a\big) \qquad \text{as } X\to\infty.$$

* *Problem* 18.2. Recall the definition of α having diophantine type κ; see (588). Prove that for any given $\kappa>2$, the set of α's of type κ has full Lebesgue measure in $\mathbb{R}$.

* *Problem* 18.3. Let α be an irrational number of diophantine type κ, let $S(\alpha)$ be as in (585), and let τ be any fixed number with

$$\tau>\max\Big(\frac45,\frac{2\kappa-1}{2\kappa}\Big). \tag{599}$$

Then prove (using Proposition 18.3) that $\big|S(\alpha)\big|\ll N^\tau$ as $N\to\infty$.

Problem 18.4. Carry out the details of the proof of (582), $\sum_{k\le z}d(k)^2\ll z(\log 2z)^3$ for all $z\ge1$, given in Davenport's book.

Problem 18.5. (a). Prove that $d(n) \ll_\varepsilon n^\varepsilon$, for any fixed $\varepsilon > 0$ and all $n \geq 1$.
[Hint. If n has the prime factorization $n = \prod_{j=1}^r p_j^{\alpha_j}$ then $d(n) = \prod_{j=1}^r (\alpha_j + 1)$.]
(b). Prove that $\sum_{n=1}^\infty \frac{d(n)^2}{n^s} = \frac{\zeta(s)^4}{\zeta(2s)}$ when $\sigma > 1$.
(c). Prove that there exist real constants c_3, c_2, c_1, c_0 and some constant $a < 1$, such that

$$\sum_{k \leq x} d(k)^2 = c_3 x(\log x)^3 + c_2 x(\log x)^2 + c_1 x(\log x) + c_0 x + O\big(x^a\big), \qquad \text{as } x \to \infty. \tag{600}$$

Also prove that $c_3 = \pi^{-2}$, and hence $\sum_{k \leq x} d(k)^2 \sim \pi^{-2} x(\log x)^3$ as $x \to \infty$.
[Hint. One approach is to work with the integral $\frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{\zeta(s)^4}{\zeta(2s)} \frac{x^s}{s}\, ds$ using the technique of §13; however we need not move the vertical line of integration all the way to $-\infty$; it suffices to move it from $\sigma = c > 1$ to $\sigma =$ some appropriately chosen number between $\frac{1}{2}$ and 1.]
(d). *This part is only for fun, for those who are interested; it gives no credit on the home assignment.* Use a computer to find numerical approximations to the constants c_3, c_2, c_1, c_0, and compute some values of the three functions $x \mapsto \sum_{k \leq x} d(k)^2$, $x \mapsto c_3 x(\log x)^3$ and $x \mapsto c_3 x(\log x)^3 + c_2 x(\log x)^2 + c_1 x(\log x) + c_0 x$.

19. Sums of three primes

(Davenport Chapter 26)

Hardy and Littlewood showed ([29]), assuming GRH, that every sufficiently large odd number is a sum of three odd primes. In their argument, the hypothesis was required to provide estimates corresponding to our estimates of $S(\alpha)$ in (585). In 1937 Vinogradov [73] used his new estimates to treat sums of three primes unconditionally:

Theorem 19.1. *(Vinogradov, 1937.) There exists some $X > 0$ such that every integer $n > X$ can be expressed as a sum of three odd primes.*

The proof of Theorem 19.1 which we will give in the present section gives the existence of X as a *non-effective* constant! In this respect it is worse than the original proof by Vinogradov in [73], which gives an effective constant. In fact, according to [43, p. 321], by working out numerical bounds on the implied constants in Vinogradov's proof, one can show that every odd integer $n > 3^{3^{15}}$ is the sum of three odd primes. (To get a feeling for the size of this number we may note that $10^{6846168} < 3^{3^{15}} < 10^{6846169}$.)

It is still not known whether *every* odd integer $n \geq 9$ is the sum of three odd primes; the best known lower bound today is due to Liu and Wang (2002, [44]; also see the survey in [43]): Every odd integer $n > e^{3100}$ is the sum of three odd primes. (We note $10^{1346} < e^{3100} < 10^{1347}$.)

To prove Theorem 19.1, instead of considering the number of representations of n as a sum of three primes, we deal with the related quantity

$$r(n) = \sum_{\substack{k_1,k_2,k_3 \\ k_1+k_2+k_3=n}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3). \tag{601}$$

(Here and below we will continue to use the convention that all summation variables by default run through all *positive integers* satisfying the given conditions, except variables named "p", which instead run through all *primes* satisfying the given conditions.) Thus $r(n)$ is a weighted counting of the number of representations of n as a sum of three prime powers. We obtain a kind of *generating function* for $r(n)$ as follows: Setting

$$S(\alpha) = \sum_{k \leq N} \Lambda(k) e(k\alpha) \tag{602}$$

as in §18.1, we see that

$$S(\alpha)^3 = \sum_{n \leq 3N} r'(n) e(n\alpha), \tag{603}$$

where $r'(n)$ is defined in the same way as $r(n)$ but with the further restriction that all the k_i are $\le N$, viz.

$$r'(n) = \sum_{\substack{k_1,k_2,k_3 \le N \\ k_1+k_2+k_3=n}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3), \tag{604}$$

Thus $r'(n) = r(n)$ for $n \le N$. As $S(\alpha)^3$ is a trigonometric polynomial, we can calculate $r(N)$ by the Fourier coefficient formula

$$r(N) = \int_0^1 S(\alpha)^3 e(-N\alpha)\, d\alpha. \tag{605}$$

We shall find that the integrand is large when α is near a rational number with a small denominator; by estimating the contributions made by these peaks, we will prove the following:

Theorem 19.2. *For any fixed $A > 0$ we have*

$$r(N) = \frac{1}{2}\mathfrak{S}(N)N^2 + O(N^2(\log N)^{-A}), \tag{606}$$

for all integers $N \ge 2$, where the implied constant only depends on A, and where

$$\mathfrak{S}(N) = \Big(\prod_{p|N}\Big(1 - \frac{1}{(p-1)^2}\Big)\Big)\Big(\prod_{p\nmid N}\Big(1 + \frac{1}{(p-1)^3}\Big)\Big). \tag{607}$$

Remark 19.1. If N is even then $\mathfrak{S}(N) = 0$, so that (606) says $r(N) = O(N^2(\log N)^{-A})$; but in fact one can see in a direct way from the definition that $r(N) = O(N(\log N)^4)$ (cf. Problem 19.1). Hence for N even the result of Theorem 19.2 is quite weak!

For N odd we have (using Proposition 2.6)

$$\mathfrak{S}(N) > \prod_{p\ge 3}\Big(1 - \frac{1}{(p-1)^2}\Big) > 0,$$

i.e. $\mathfrak{S}(N)$ is bounded from below by a positive constant which is independent of N. In particular Theorem 19.2 implies that $r(N) \gg N^2$ for all sufficiently large odd N.

Theorem 19.1 follows as a fairly easy consequence of Theorem 19.2. In fact the contribution made to $r(N)$ by proper prime powers can be seen to be $\ll N^{\frac{3}{2}}(\log N)^2$ (this is carefully proved in (609) below) and hence Theorem 19.2 implies the following more precise result:

Theorem 19.3. *There exist some positive constants X and c such that every odd integer $N > X$ can be expressed as a sum of three odd primes in $> cN^2(\log N)^{-3}$ ways.*

Proof of [Theorem 19.2 ⇒ Theorem 19.3]. Recall that we have defined

$$r(n) = \sum_{\substack{k_1,k_2,k_3\\ k_1+k_2+k_3=n}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3).$$

Hence, writing P for the set of prime numbers,

$$\sum_{\substack{p_1,p_2,p_3\\ p_1+p_2+p_3=n}} (\log p_1)(\log p_2)(\log p_3) = r(n) - \sum_{\substack{k_1,k_2,k_3\\ k_1+k_2+k_3=n\\ \{k_1,k_2,k_3\}\not\subset P}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3), \tag{608}$$

and here

$$\begin{aligned}\sum_{\substack{k_1,k_2,k_3\\ k_1+k_2+k_3=n\\ \{k_1,k_2,k_3\}\not\subset P}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3) &\le 3 \sum_{\substack{k_1,k_2,k_3\\ k_1+k_2+k_3=n\\ k_1\notin P}} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3)\\ &= 3 \sum_{2\le r\le \log_2 n} \sum_{p\le n^{1/r}} (\log p) \sum_{\substack{k_2,k_3\\ k_2+k_3=n-p^r}} \Lambda(k_2)\Lambda(k_3),\end{aligned}$$

where in the last step we substituted $k_1 = p^r$. We may continue, overestimating the inner sum in a trivial way:

$$\begin{aligned}&\ll \sum_{2\le r\le \log_2 n} \sum_{p\le n^{1/r}} (\log p) \sum_{\substack{k_2,k_3\\ k_2+k_3=n-p^r}} (\log n)^2\\ &= \sum_{2\le r\le \log_2 n} \sum_{p\le n^{1/r}} (\log p)(n-p^r-1)(\log n)^2\\ &\le n(\log n)^2 \sum_{2\le r\le \log_2 n} \sum_{p\le n^{1/r}} \log p = n(\log n)^2 \sum_{2\le r\le \log_2 n} \vartheta(n^{1/r})\\ &\ll n(\log n)^2 \sum_{2\le r\le \log_2 n} n^{1/r} \le n(\log n)^2\big(n^{\frac12} + (\log_2 n)n^{\frac13}\big) \ll n^{\frac32}(\log n)^2. \end{aligned} \tag{609}$$

Using this bound on the last term in (608) together with the result from Theorem 19.2 that $r(n) \gg n^2$ for all sufficiently large odd $n \ge 3$ (cf. Remark 19.1) we see that (608) implies that

$$\sum_{\substack{p_1,p_2,p_3\\ p_1+p_2+p_3=n}} (\log p_1)(\log p_2)(\log p_3) \gg n^2$$

for all sufficiently large odd integers n. Here $\log p_j \le \log n$ holds for all p_j appearing in the sum; hence

$$\sum_{\substack{p_1,p_2,p_3\\ p_1+p_2+p_3=n}} (\log n)^3 \gg n^2,$$

viz.

$$\#\big\{\langle p_1,p_2,p_3\rangle \in P^3 \,:\, p_1+p_2+p_3=n\big\} \gg n^2(\log n)^{-3}$$

for all sufficiently large odd n. Finally note that for n odd, the only way for some of the primes in $p_1+p_2+p_3=n$ to be *even* (i.e. $=2$) is if $n-4$ is a prime and then two of p_1,p_2,p_3 equal 2 and the third equals $n-4$. Hence the number of triples $\langle p_1,p_2,p_3\rangle$ of *odd* primes with $p_1+p_2+p_3=n$ is $\gg n^2(\log n)^{-3}-3 \gg n^2(\log n)^{-3}$, for n sufficiently large. □

Remark 19.2. One may argue that when counting the number of ways to express n as a sum of three primes, what we really should count is the number of "genuinly different" representations, i.e. count the number of *equivalence classes* into which the set

$$\big\{\langle p_1,p_2,p_3\rangle \in P^3 \,:\, p_1+p_2+p_3=n\big\}$$

is partitioned if we consider $\langle p_1,p_2,p_3\rangle$ and $\langle p_1',p_2',p_3'\rangle$ to be equivalent whenever $\langle p_1',p_2',p_3'\rangle = \langle p_{\tau(1)},p_{\tau(2)},p_{\tau(3)}\rangle$ for some permutation τ of $\{1,2,3\}$. However since each equivalence class contains at most six triples $\langle p_1,p_2,p_3\rangle$, Theorem 19.3 remains true also with this interpretation.

We now embark on the proof of the main result:

Proof of Theorem 19.2. We divide the range of integration in (605) into subintervals for detailed treatment. Let $P=(\log N)^B$, $Q=N(\log N)^{-B}$, where $B>0$ will be specified later in terms of A. We assume that N is so large that

$$\frac{1}{P^2} > \frac{2}{Q}.$$

For $1\le q\le P$, $1\le a\le q$, $(a,q)=1$, we set

$$\mathfrak{M}(q,a) = \Big[\frac{a}{q}-\frac{1}{Q},\frac{a}{q}+\frac{1}{Q}\Big]. \tag{610}$$

These sets $\mathfrak{M}(q,a)$ are called the "major arcs", and we think of each $\mathfrak{M}(q,a)$ as an interval, or "arc", on the "circle" $\mathbb{R}/\mathbb{Z}$ (the real numbers modulo 1), which we will identify with $[0,1)$; thus $\mathfrak{M}(1,1)$ can be thought of as the "arc" $[0,Q^{-1}]\cup[1-Q^{-1},1)$. We note that any two major arcs $\mathfrak{M}(q,a)$ and $\mathfrak{M}(q',a')$ with $\frac{a}{q}\neq\frac{a'}{q'}$ are disjoint, since

$$\Big|\frac{a}{q}-\frac{a'}{q'}\Big| \ge \frac{1}{qq'} \ge \frac{1}{P^2} > \frac{2}{Q}.$$

We also let $\mathfrak{M}$ be the union of all major arcs $\mathfrak{M}(q,a)$, and let $\mathfrak{m}$ be the complement in $[0,1)$ of $\mathfrak{M}$. Note that $\mathfrak{m}$ is also a finite union of arcs on $\mathbb{R}/\mathbb{Z}$; we call these the "minor arcs".

We now estimate the contribution of the major arcs to the integral (605). To this end we first determine the size of $S(\alpha)$ for $\alpha\in\mathfrak{M}(q,a)$. We write $\alpha=\frac{a}{q}+\beta$ so that $|\beta|\le Q^{-1}$

and

$$S(\alpha) = \sum_{k \le N} \Lambda(k) e\Big(\frac{ka}{q}\Big) e(k\beta).$$

Here the factor $e(k\beta)$ oscillates quite slowly, so it should be possible to control $S(\alpha)$ if we can control the sum $\sum_k \Lambda(k) e\Big(\frac{ka}{q}\Big)$ over a variable range; and since the function $k \mapsto e\Big(\frac{ka}{q}\Big)$ is periodic modulo q we wish to write it as a linear combination of Dirichlet characters modulo q (this can only be done if we restrict to k's with $(k,q) = 1$, since otherwise $\chi(k) = 0$ for all $\chi \in X_q$), so as to express things in terms of our well-studied functions $\psi(x,\chi) = \sum_{k \le x} \chi(k)\Lambda(k)$. Using Lemma 1.5 we see that for all $k \in \mathbb{Z}$,

$$\begin{aligned}
\begin{Bmatrix} e(ka/q) & \text{if } (k,q) = 1 \\ 0 & \text{if } (k,q) > 1 \end{Bmatrix} &= \sum_{m \in (\mathbb{Z}/q\mathbb{Z})^\times} e\Big(\frac{ma}{q}\Big) \begin{Bmatrix} 1 & \text{if } m \equiv k \pmod q \\ 0 & \text{if } m \not\equiv k \pmod q \end{Bmatrix} \\
= \frac{1}{\phi(q)} \sum_{m \in (\mathbb{Z}/q\mathbb{Z})^\times} e\Big(\frac{ma}{q}\Big) \sum_{\chi \in X_q} \overline{\chi(m)}\chi(k) &= \frac{1}{\phi(q)} \sum_{\chi \in X_q} \Big(\sum_{m \in (\mathbb{Z}/q\mathbb{Z})^\times} e\Big(\frac{ma}{q}\Big)\overline{\chi(m)}\Big)\chi(k) \\
&= \frac{1}{\phi(q)} \sum_{\chi \in X_q} \tau(\overline{\chi})\chi(a)\chi(k)
\end{aligned}$$

(cf. (346) concerning the last step). This is our desired linear combination formula. Using it we get

$$\begin{aligned}
S(\alpha) &= \sum_{\substack{k \le N \\ (k,q)=1}} \Lambda(k) e\Big(\frac{ka}{q}\Big) e(k\beta) + \sum_{\substack{k \le N \\ (k,q)>1}} \Lambda(k) e\Big(\frac{ka}{q}\Big) e(k\beta) \\
&= \frac{1}{\phi(q)} \sum_{\chi \in X_q} \tau(\overline{\chi})\chi(a) \sum_{k \le N} \Lambda(k)\chi(k) e(k\beta) + O\Big(\sum_{p \mid q} \sum_{r \le \log_p N} \Lambda(p^r) \Big) \\
&= \frac{1}{\phi(q)} \sum_{\chi \in X_q} \tau(\overline{\chi})\chi(a) \int_1^N e(x\beta)\, d\psi(x,\chi) + O\Big(\sum_{p \mid q} \log N \Big) \\
&= \frac{1}{\phi(q)} \sum_{\chi \in X_q} \tau(\overline{\chi})\chi(a) \Big(e(N\beta)\psi(N,\chi) - 2\pi i\beta \int_1^N e(x\beta)\psi(x,\chi)\, dx \Big) + O\Big((\log N)^2 \Big),
\end{aligned} \tag{611}$$

where in the last step we used $\sum_{p \mid q} 1 = \omega(q) \ll \log q \le \log P \le \log N$ (cf. Problem 17.1); the implied constant is absolute. In order to bound $\psi(x,\chi)$, we note that by Theorem 16.5 with $2B$ in place of "N", $\psi(x,\chi) = O\Big(x e^{-c_1\sqrt{\log x}} \Big)$ holds for all nonprincipal $\chi \in X_q$ and all $x \ge 2$ with $(\log x)^{2B} \ge q$. Here $c_1 > 0$ is an absolute constant, and the implied constant depends only on B but is *noneffective!* It follows that for all x with $\exp\big(q^{\frac{1}{2B}}\big) \le x \le N$ we

have

$$\psi(x,\chi) = O\Big(Ne^{-c_1\sqrt{\log N}}\Big). \tag{612}$$

Furthermore for all x with $x \le \exp\big(q^{\frac{1}{2B}}\big)$ we have

$$\big|\psi(x,\chi)\big| \le \psi(x) = O\big(x\big) = O\Big(\exp\big(q^{\frac{1}{2B}}\big)\Big) = O\Big(\exp\big(P^{\frac{1}{2B}}\big)\Big) = O\Big(e^{\sqrt{\log N}}\Big).$$

Thus (612) in fact holds for all $1 \le x \le N$, and hence

$$e(N\beta)\psi(N,\chi) - 2\pi i\beta\int_1^N e(x\beta)\psi(x,\chi)\,dx = O\Big((1+|\beta|N)Ne^{-c_1\sqrt{\log N}}\Big), \tag{613}$$

for each nonprincipal $\chi \in X_q$.

To treat the principal character $\chi_0 \in X_q$ we define $R(x)$ through

$$\psi(x,\chi_0) = \lfloor x\rfloor + R(x).$$

Now for all $1 \le x \le N$ we have

$$\psi(x,\chi_0) = \psi(x) - \sum_{p|q}\sum_{r\le \log_p x}\Lambda(p^r) = \psi(x) - O\Big(\sum_{p|q}\log N\Big) = \psi(x) - O\Big(\log^2 N\Big)$$

in the same way as in (611), and hence by Theorem 13.8,

$$R(x) = \psi(x) - \lfloor x\rfloor - O\big(\log^2 N\big) = O\Big(xe^{-c_2\sqrt{\log x}}\Big) + O\big(\log^2 N\big) = O\Big(Ne^{-c_2\sqrt{\log N}}\Big)$$

for all $1 \le x \le N$, where $c_2 > 0$ is an absolute constant. Hence

$$\begin{aligned} &e(N\beta)\psi(N,\chi_0) - 2\pi i\beta\int_1^N e(x\beta)\psi(x,\chi_0)\,dx \\ &= \Big(e(N\beta)\lfloor N\rfloor - 2\pi i\beta\int_1^N e(x\beta)\lfloor x\rfloor\,dx\Big) + \Big(e(N\beta)R(N) - 2\pi i\beta\int_1^N e(x\beta)R(x)\,dx\Big) \\ &= \int_{1-}^N e(x\beta)d\lfloor x\rfloor + O\Big((1+|\beta|N)Ne^{-c_2\sqrt{\log N}}\Big), \end{aligned} \tag{614}$$

where the last step follows by "backwards integration by parts". We give the integral (the "main term") in the last line the name $T(\beta)$, and note that it equals

$$T(\beta) := \int_{1-}^N e(x\beta)d\lfloor x\rfloor = \sum_{k=1}^N e(k\beta).$$

Using (613) and (614) in (611), together with the fact that $\tau(\chi_0) = \mu(q)$ and $\big|\tau(\chi)\big| \le \sqrt{q}$ for any $\chi \pmod q$ (both these facts follow from Problem 9.2) we now get:

$$S(\alpha) = \frac{\mu(q)}{\phi(q)}T(\beta) + O\Big(q^{\frac12}(1+|\beta|N)Ne^{-c_3\sqrt{\log N}}\Big) \qquad (c_3 = \min(c_2,c_1)).$$

But $|\beta|N \le Q^{-1}N = (\log N)^B$ and $q^{\frac12} \le P^{\frac12} = (\log N)^{\frac12 B}$ and thus

$$S(\alpha) = \frac{\mu(q)}{\phi(q)} T(\beta) + O\Big(N e^{-c_4\sqrt{\log N}}\Big)$$

where $c_4 > 0$ is any fixed constant strictly less than c_3. Consequently (using also $\big|\frac{\mu(q)}{\phi(q)}T(\beta)\big| \le \big|T(\beta)\big| \le N$)

$$S(\alpha)^3 = \frac{\mu(q)}{\phi(q)^3} T(\beta)^3 + O\Big(N^3 e^{-c_4\sqrt{\log N}}\Big).$$

This holds for all $\alpha \in \mathfrak{M}(q,a)$, and hence the contribution of the arc $\mathfrak{M}(q,a)$ to the integral (605) is (again using $Q^{-1}N = (\log N)^B$)

$$\frac{\mu(q)}{\phi(q)^3} e\Big(-\frac{aN}{q}\Big) \int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\, d\beta + O\Big(N^2 e^{-c_5\sqrt{\log N}}\Big),$$

where $c_5 > 0$ is any fixed constant strictly less than c_4. Summing over the various major arcs, we see that

$$\int_{\mathfrak{M}} S(\alpha)^3 e(-N\alpha)\, d\alpha = \sum_{q\le P} \frac{\mu(q)}{\phi(q)^3} c_q(N) \int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\, d\beta + O\Big(N^2 e^{-c_6\sqrt{\log N}}\Big), \tag{615}$$

where $c_6 > 0$ is any fixed constant strictly less than c_5, and where $c_q(n)$ is Ramanujan's sum,

$$c_q(n) = \sum_{\substack{a=1\\(a,q)=1}}^{q} e\Big(\frac{an}{q}\Big),$$

which we introduced in Problem 9.3.

We now estimate the integral and the sum occurring on the right hand side of (615). The sum $T(\beta)$ only depends on $\beta \bmod 1$ and is a geometric series with sum

$$\frac{e((N+1)\beta) - e(\beta)}{e(\beta) - 1} = O\Big(\min(N, \|\beta\|^{-1})\Big).$$

Hence

$$\int_{1/Q}^{1-1/Q} |T(\beta)|^3\, d\beta = O\Big(\int_{1/Q}^{1/2} \beta^{-3}\, d\beta\Big) = O\big(Q^2\big) = O\big(N^2 (\log N)^{-2B}\big),$$

so that

$$\int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\, d\beta = \int_0^1 T(\beta)^3 e(-N\beta)\, d\beta + O\big(N^2(\log N)^{-2B}\big).$$

The integral on the right equals

$$\int_0^1 \Big(\sum_{1\le k_1,k_2,k_3\le N} e((k_1+k_2+k_3)\beta)\Big) e(-N\beta)\,d\beta = \sum_{\substack{1\le k_1,k_2,k_3\le N\\ k_1+k_2+k_3=N}} 1,$$

viz. the number of ways of writing N in the form $N=k_1+k_2+k_3$ with positive integers k_1,k_2,k_3, and this is

$$= \tfrac{1}{2}(N-1)(N-2) = \tfrac{1}{2}N^2 + O(N).$$

Hence

$$\int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\,d\beta = \tfrac{1}{2}N^2 + O\big(N^2(\log N)^{-2B}\big).$$

Using this in (615) we get

$$\begin{aligned}\int_{\mathfrak{M}} S(\alpha)^3 e(-N\alpha)\,d\alpha &= \sum_{q\le P} \frac{\mu(q)}{\phi(q)^3} c_q(N)\Big(\tfrac{1}{2}N^2 + O\big(N^2(\log N)^{-2B}\big)\Big) + O\Big(N^2 e^{-c_6\sqrt{\log N}}\Big)\\ &= \frac{1}{2}N^2 \sum_{q\le P} \frac{\mu(q)}{\phi(q)^3} c_q(N) + O\Big(N^2(\log N)^{-2B} \sum_{q\le P} \frac{|c_q(N)|}{\phi(q)^3}\Big) + O\Big(N^2 e^{-c_6\sqrt{\log N}}\Big).\end{aligned}$$

Using the trivial estimate $|c_q(N)| \le \phi(q)$ together with the fact that $\phi(q) \gg q/\log q$ for all $q \ge 2$ (cf. Problem 15.1) we see that

$$\sum_{q\le P} \frac{|c_q(N)|}{\phi(q)^3} \le \sum_{q=1}^{\infty} \frac{1}{\phi(q)^2} = O(1),$$

so that the above gives

$$\int_{\mathfrak{M}} S(\alpha)^3 e(-N\alpha)\,d\alpha = \frac{1}{2}N^2 \sum_{q\le P} \frac{\mu(q)}{\phi(q)^3} c_q(N) + O\Big(N^2(\log N)^{-2B}\Big).$$

Next, again using $|c_q(N)| \le \phi(q)$ and $\phi(q) \gg q/\log q$ ($\forall q \ge 2$), thus $\phi(q) \gg_\varepsilon q^{1-\varepsilon}$ ($\forall q \ge 1$) for any fixed $0 < \varepsilon < \frac{1}{2}$, we get

$$\begin{aligned}\Big|\sum_{q>P} \frac{\mu(q)}{\phi(q)^3} c_q(N)\Big| \le \sum_{q>P} \frac{1}{\phi(q)^2} &\ll_\varepsilon \sum_{q>P} q^{-2+2\varepsilon} \le \int_{P-1}^{\infty} x^{-2+2\varepsilon}\,dx\\ &\ll_\varepsilon (P-1)^{-1+2\varepsilon} \ll P^{-1+2\varepsilon} = (\log N)^{-B+2\varepsilon B},\end{aligned}$$

and thus if we choose $\varepsilon = \frac{1}{2B}$ we have

$$\Big|\sum_{q>P} \frac{\mu(q)}{\phi(q)^3} c_q(N)\Big| \ll (\log N)^{-B+1}$$

(where the implied constant depends only on B). Hence

$$\sum_{q\le P}\frac{\mu(q)}{\phi(q)^3}c_q(N)=\sum_{q=1}^{\infty}\frac{\mu(q)}{\phi(q)^3}c_q(N)+O\big((\log N)^{-B+1}\big).$$

Here the infinite sum can be factored as an Euler product if we recall from Problem 9.3 that

$$c_q(N)=\frac{\phi(q)}{\phi(\frac{q}{(q,N)})}\mu\Big(\frac{q}{(q,N)}\Big)$$

and (thus) $c_q(N)$ is multiplicative as a function of q. Hence, by Proposition 2.7 and since $\mu(p^\alpha)=0$ for all $\alpha\ge 2$,

$$\begin{aligned}\sum_{q=1}^{\infty}\frac{\mu(q)}{\phi(q)^3}c_q(N)&=\prod_p\Big(1+\frac{\mu(p)}{\phi(p)^3}c_p(N)\Big)=\prod_p\Big(1-\frac{c_p(N)}{(p-1)^3}\Big)\\&=\prod_{p\mid N}\Big(1-\frac{1}{(p-1)^2}\Big)\prod_{p\nmid N}\Big(1+\frac{1}{(p-1)^3}\Big)=\mathfrak{S}(N),\end{aligned}$$

cf. our definition (607). Hence we conclude:

$$\int_{\mathfrak{M}}S(\alpha)^3e(-N\alpha)\,d\alpha=\frac12\mathfrak{S}(N)N^2+O\Big(N^2(\log N)^{-B+1}\Big). \tag{616}$$

To complete the argument we must show that the minor arcs contribute a smaller amount. We note that

$$\begin{aligned}\Big|\int_{\mathfrak{m}}S(\alpha)^3e(-N\alpha)\,d\alpha\Big|&\le\int_{\mathfrak{m}}\big|S(\alpha)\big|^3\,d\alpha\le\Big(\sup_{\alpha\in\mathfrak{m}}\big|S(\alpha)\big|\Big)\int_{\mathfrak{m}}\big|S(\alpha)\big|^2\,d\alpha\\&\le\Big(\sup_{\alpha\in\mathfrak{m}}\big|S(\alpha)\big|\Big)\int_0^1\big|S(\alpha)\big|^2\,d\alpha\end{aligned} \tag{617}$$

and here the last integral is

$$\begin{aligned}\int_0^1\big|S(\alpha)\big|^2\,d\alpha&=\int_0^1\sum_{k_1\le N}\sum_{k_2\le N}\Lambda(k_1)\Lambda(k_2)e\big((k_1-k_2)\alpha\big)\,d\alpha\\&=\sum_{k_1\le N}\Lambda(k_1)\sum_{k_2\le N}\Lambda(k_2)\int_0^1e\big((k_1-k_2)\alpha\big)\,d\alpha\\&=\sum_{k\le N}\Lambda(k)^2\le(\log N)\sum_{k\le N}\Lambda(k)\ll N\log N.\end{aligned} \tag{618}$$

(Here the first two lines were simply a proof of Parseval's formula, in the present situation.) Finally for every $\alpha\in\mathfrak{m}$ Dirichlet's theorem on Diophantine approximation (Lemma 18.4) says that there is a rational number a/q such that $\Big|\alpha-\frac{a}{q}\Big|\le\frac{1}{qQ}$, $1\le q\le Q$ and $(a,q)=1$.

If $q \le P$ then $\alpha \in \mathfrak{M}(a,q)$; hence in the present case we must have $P < q \le Q$. Hence by Proposition 18.3 we have

$$\begin{aligned}\big|S(\alpha)\big| \ll (Nq^{-\frac12} + N^{\frac45} + N^{\frac12}q^{\frac12})(\log N)^4 \ll \big(NP^{-\frac12} + N^{\frac45} + N^{\frac12}Q^{\frac12}\big)(\log N)^4 \\ \ll N(\log N)^{-(B/2)+4}.\end{aligned}$$

Hence

$$\sup_{\alpha\in\mathfrak{m}}\big|S(\alpha)\big| \ll N(\log N)^{-(B/2)+4}$$

and hence from (617) and (618) we conclude that

$$\Big|\int_{\mathfrak{m}} S(\alpha)^3 e(-N\alpha)\,d\alpha\Big| \ll N^2(\log N)^{-(B/2)+5}.$$

This together with (616) gives the desired result, on taking $B = 2A + 10$. □

19.1. **Problems.**

Problem 19.1. Prove that for all even positive integers N we have $r(N) \ll N(\log N)^4$. [Hint. For each nonvanishing term in (601), at least one k_i must be a power of 2.]

Problem 19.2. Prove (using Theorem 19.2) that if $t(N)$ is the number of ways to write N as a sum of three odd primes, i.e.

$$t(N) = \#\big\{\langle p_1,p_2,p_3\rangle \,:\, p_1+p_2+p_3 = N\big\},$$

then

$$t(N) \sim \frac{\mathfrak{S}(N)N^2}{2(\log N)^3} \qquad \text{as } N\to\infty. \tag{619}$$

[Hint is available upon request (`astrombe@math.uu.se`) — I don't want to spoil things for those who first want to work without a hint.]

20. The Large Sieve

(Davenport chapter 27.)

The large sieve was first proposed by Linnik in a short but important paper of 1941, [42]. In a subsequent series of papers, Rényi developed the method by adopting a probabilistic attitude. His estimates were not optimal, and in 1965 Roth substantially modified Rényi's approach to obtain an essentially optimal result [58]. Bombieri further refined the large sieve, and used it to describe the distribution of primes in arithmetic progressions [5], see the next section.

Rényi's approach to the large sieve concerns an extension of Bessel's inequality. We recall that Bessel's inequality asserts that if $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_R$ are orthonormal members of an inner product space V over the complex numbers, and if $\boldsymbol{\xi} \in V$, then

$$\sum_{r=1}^{R} |(\boldsymbol{\xi}, \boldsymbol{\phi}_r)|^2 \le \|\boldsymbol{\xi}\|^2. \tag{620}$$

In number theory we frequently encounter vectors which are not quite orthonormal. Thus, with possible applications in mind, we seek an inequality

$$\sum_{r=1}^{R} |(\boldsymbol{\xi}, \boldsymbol{\phi}_r)|^2 \le A\|\boldsymbol{\xi}\|^2. \tag{621}$$

valid for all $\boldsymbol{\xi}$, where A depends on $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_R$; we hope to find that A is near 1 when the $\boldsymbol{\phi}_r$ are in some sense nearly orthonormal. Boas has characterized the constant A for which (621) holds [4]:

Proposition 20.1. *If $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_R$ are arbitrary vectors in an inner product space V, and $A \ge 0$, then the bound*

$$\sum_{r=1}^{R} |(\boldsymbol{\xi}, \boldsymbol{\phi}_r)|^2 \le A\|\boldsymbol{\xi}\|^2, \qquad \forall \boldsymbol{\xi} \in V, \tag{622}$$

holds if and only if

$$\sum_{r=1}^{R}\sum_{s=1}^{R} u_r\overline{u}_s(\boldsymbol{\phi}_r, \boldsymbol{\phi}_s) \le A\sum_{r=1}^{R} |u_r|^2, \qquad \forall u_1, \ldots, u_R \in \mathbb{C}. \tag{623}$$

(Note that the double sum in the left hand side of (623) is always *real*, since $\overline{u_r\overline{u}_s(\boldsymbol{\phi}_r, \boldsymbol{\phi}_s)} = u_s\overline{u}_r(\boldsymbol{\phi}_s, \boldsymbol{\phi}_r)$ for all r, s.)

Proof. Clearly (622)⇔(623) holds when $A = 0$, both statements being equivalent with $\boldsymbol{\phi}_1 = \ldots = \boldsymbol{\phi}_R = \mathbf{0}$ in this case. Hence we may now assume $A > 0$.

First assume that (623) holds. Take $\boldsymbol{\xi} \in V$ arbitrary. Then for any $u_1, \ldots, u_R \in \mathbb{C}$ we have

$$\begin{aligned} 0 \le \Big\| \boldsymbol{\xi} - \sum_{r=1}^{R} u_r \boldsymbol{\phi}_r \Big\|^2 &= \Big(\boldsymbol{\xi} - \sum_{r=1}^{R} u_r \boldsymbol{\phi}_r, \boldsymbol{\xi} - \sum_{r=1}^{R} u_r \boldsymbol{\phi}_r \Big) \\ &= \|\boldsymbol{\xi}\|^2 - \sum_{r=1}^{R} u_r \overline{(\boldsymbol{\xi}, \boldsymbol{\phi}_r)} - \sum_{r=1}^{R} \overline{u}_r (\boldsymbol{\xi}, \boldsymbol{\phi}_r) + \sum_{r=1}^{R} \sum_{s=1}^{R} u_r \overline{u}_s (\boldsymbol{\phi}_r, \boldsymbol{\phi}_s), \\ &= \|\boldsymbol{\xi}\|^2 - 2\mathrm{Re} \sum_{r=1}^{R} \overline{u}_r (\boldsymbol{\xi}, \boldsymbol{\phi}_r) + \sum_{r=1}^{R} \sum_{s=1}^{R} u_r \overline{u}_s (\boldsymbol{\phi}_r, \boldsymbol{\phi}_s), \end{aligned}$$

and by (623) this implies

$$0 \le \|\boldsymbol{\xi}\|^2 - 2\mathrm{Re} \sum_{r=1}^{R} \overline{u}_r (\boldsymbol{\xi}, \boldsymbol{\phi}_r) + A \sum_{r=1}^{R} |u_r|^2.$$

We here take $u_r = (\boldsymbol{\xi}, \boldsymbol{\phi}_r)/A$; then the above inequality simplifies to read

$$0 \le \|\boldsymbol{\xi}\|^2 - \frac{1}{A} \sum_{r=1}^{R} \big| (\boldsymbol{\xi}, \boldsymbol{\phi}_r) \big|^2,$$

i.e. (622) holds.

Conversely, now assume that (622) holds. Let $u_1, \ldots, u_R \in \mathbb{C}$ be arbitrary and set $\boldsymbol{\xi} = \sum_{s=1}^{R} u_s \boldsymbol{\phi}_s$. Then

$$\begin{aligned} \big\| \boldsymbol{\xi} \big\|^2 = \Big(\boldsymbol{\xi}, \sum_{s=1}^{R} u_s \boldsymbol{\phi}_s \Big) = \sum_{s=1}^{R} \overline{u}_s (\boldsymbol{\xi}, \boldsymbol{\phi}_s) &\le \Big(\sum_{s=1}^{R} |u_s|^2 \Big)^{\frac{1}{2}} \Big(\sum_{s=1}^{R} \big| (\boldsymbol{\xi}, \boldsymbol{\phi}_s) \big|^2 \Big)^{\frac{1}{2}} \\ &\le A^{\frac{1}{2}} \big\| \boldsymbol{\xi} \big\| \Big(\sum_{s=1}^{R} |u_s|^2 \Big)^{\frac{1}{2}}, \end{aligned}$$

where we used (622) in the last step. If $\boldsymbol{\xi} \neq \mathbf{0}$ then we divide both sides of the above inequality by $\|\boldsymbol{\xi}\|$ and square; this gives (623). Note that (623) also holds in the remaining case $\boldsymbol{\xi} = \mathbf{0}$, trivially. □

Using the above lemma we will now prove:

Theorem 20.2. *Let $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_R$ be arbitrary vectors in an inner product space V over the complex numbers, and set*

$$A = \max_r \sum_{s=1}^{R} \big| (\boldsymbol{\phi}_r, \boldsymbol{\phi}_s) \big|. \tag{624}$$

Then

$$\sum_{r=1}^{R}\big|(\boldsymbol{\xi},\boldsymbol{\phi}_r)\big|^2 \le A\big\|\boldsymbol{\xi}\big\|^2, \qquad \forall \boldsymbol{\xi}\in V. \tag{625}$$

Proof. Given $u_1,\dots,u_R\in\mathbb{C}$ we have $|u_r\overline{u}_s|\le \frac12|u_r|^2+\frac12|u_s|^2$ for all r,s, and hence

$$\sum_{r=1}^{R}\sum_{s=1}^{R} u_r\overline{u}_s(\boldsymbol{\phi}_r,\boldsymbol{\phi}_s) \le \sum_{r=1}^{R}\sum_{s=1}^{R}\big(\tfrac12|u_r|^2+\tfrac12|u_s|^2\big)\big|(\boldsymbol{\phi}_r,\boldsymbol{\phi}_s)\big| = \sum_{r=1}^{R}|u_r|^2\sum_{s=1}^{R}\big|(\boldsymbol{\phi}_r,\boldsymbol{\phi}_s)\big|$$
$$\le \Big(\max_r\sum_{s=1}^{R}\big|(\boldsymbol{\phi}_r,\boldsymbol{\phi}_s)\big|\Big)\sum_{r=1}^{R}|u_r|^2.$$

Hence (623) holds with A as in (624). Hence by Proposition 20.1, (625) holds. □

Following Davenport and Halberstam ([14], 1966), we consider the large sieve to be an inequality of the following kind:

Definition 20.1. Given $N\in\mathbb{Z}^+$ and $\delta>0$, we say that *the number* $\Delta=\Delta(N,\delta)>0$ *satisfies the large sieve inequality for* N,δ, if the following holds: For any $M\in\mathbb{Z}$ and any sequence of complex numbers $\{a_n\}$, if we set

$$S(\alpha)=\sum_{n=M+1}^{M+N} a_n e(n\alpha),$$

then for any $R\in\mathbb{Z}^+$ and any real numbers $\alpha_1,\dots,\alpha_R$, satisfying $\|\alpha_r-\alpha_s\|\ge\delta$ for[32] all $r\ne s$, we have

$$\sum_{r=1}^{R}|S(\alpha_r)|^2\le\Delta\sum_{n=M+1}^{M+N}|a_n|^2. \tag{626}$$

The value of M is in fact irrelevant in the above definition, since for any K we can put

$$T(\alpha)=\sum_{n=K+1}^{K+N} a_{M-K+n}e(n\alpha)=e((K-M)\alpha)S(\alpha),$$

and then T has frequencies in the range $K+1\le n\le K+N$ and $\big|T(\alpha)\big|=\big|S(\alpha)\big|$ for all $\alpha\in\mathbb{R}$. We also note the following:

Proposition 20.3. *Given* N,δ, *if* Δ *satisfies the large sieve inequality for* N,δ, *then*

$$\Delta\ge\max(N,\delta^{-1}-1).$$

[32]Recall from p. 254 that $\|\beta\|$ denotes the distance from β to the nearest integer; thus $\|\alpha_r-\alpha_s\|\ge\delta$ means that α_r (mod 1) and α_s (mod 1) have distance $\ge\delta$ on the "circle" $\mathbb{R}/\mathbb{Z}$.

Proof. Taking $M = 0$, $a_n = 1$ $(\forall n)$, $R = 1$ and $\alpha_1 = 0$ in Definition 20.1 we see that Δ must satisfy $\big|S(0)\big|^2 \le \Delta N$, viz. $N^2 \le \Delta N$, viz. $\Delta \ge N$.

It remains to prove $\Delta \ge \delta^{-1} - 1$, and here we may clearly assume $\delta < 1$ since the inequality is otherwise trivial. Note that for *any* choice of $\alpha_1, \dots, \alpha_R \in \mathbb{R}$, we have

$$\int_0^1 \sum_{r=1}^{R} \big|S(\alpha_r + \beta)\big|^2 \, d\beta = R \int_0^1 |S(\beta)|^2 \, d\beta = R \sum_{n=M+1}^{M+N} |a_n|^2,$$

and hence there is some $\beta \in \mathbb{R}$ for which

$$\sum_{r=1}^{R} \big|S(\alpha_r + \beta)\big|^2 \ge R \sum_{n=M+1}^{M+N} |a_n|^2.$$

Note that the points $\{\alpha_r + \beta\}_{r=1}^R$ are separated by at least $\delta \,(\mathrm{mod}\ 1)$ if and only if the points $\{\alpha_r\}_{r=1}^R$ are so separated. Hence we conclude that $\Delta \ge R$ holds for every $R \in \mathbb{Z}^+$ such that there *exists* a set of R points separated by at least $\delta \,(\mathrm{mod}\ 1)$. But such a set of points exists if and only if $R\delta \le 1$, and so the largest choice of R is $R = \lfloor \delta^{-1} \rfloor$. Hence $\Delta \ge \lfloor \delta^{-1} \rfloor \ge \delta^{-1} - 1$. □

The above proposition shows that the following theorem is essentially the best possible.

Theorem 20.4. *The large sieve inequality holds with* $\Delta = N + 3\delta^{-1}$.

Proof. If $R = 1$ then

$$\sum_{r=1}^{R} \big|S(\alpha_r)\big|^2 = \big|S(\alpha_1)\big|^2 = \Big| \sum_{n=M+1}^{M+N} a_n e(n\alpha_1) \Big|^2 \le N \sum_{n=M+1}^{M+N} |a_n|^2,$$

by Cauchy's inequality, i.e. (626) holds even with $\Delta = N$. Hence from now on we may assume $R \ge 2$, and thus $\delta \le \frac{1}{2}$. Using also our remark that the exact value of M is irrelevant we see that it suffices to show that

$$\sum_{r=1}^{R} \Big| \sum_{k=-K}^{K} a_k e(k\alpha_r) \Big|^2 \le (2K + 3\delta^{-1}) \sum_{k=-K}^{K} \big|a_k\big|^2, \tag{627}$$

for any $\delta \le \frac{1}{2}$ and $K \ge 0$. (For note that the number of terms in $\sum_{k=-K}^{K}$ is $N = 2K + 1$, and thus (627) actually implies that the large sieve inequality holds with $\Delta = N - 1 + 3\delta^{-1}$ if N is odd, but $\Delta = N + 3\delta^{-1}$ if N is even, e.g. by taking $a_{-K} = 0$ in (627).)

We will now apply Theorem 20.2 for the inner product space

$$V = \ell^2 = \Big\{ \boldsymbol{x} = (x_k)_{k \in \mathbb{Z}} \,:\, x_k \in \mathbb{C},\ \sum_{k \in \mathbb{Z}} |x_k|^2 < \infty \Big\},$$

with the standard inner product $(\boldsymbol{x},\boldsymbol{y}) = \sum_{k\in\mathbb{Z}} x_k\overline{y}_k$. Given some numbers $b_k \geq 0$ ($k\in\mathbb{Z}$) such that $\sum_k b_k < \infty$, and $b_k > 0$ when $-K \leq k \leq K$, we set

$$\boldsymbol{\xi} = (\xi_k)_{k\in\mathbb{Z}} \in V; \qquad \xi_k = \begin{cases} a_k b_k^{-\frac{1}{2}} & \text{if } -K \leq k \leq K \\ 0 & \text{else,} \end{cases}$$

and, for each $r = 1,\dots,R$,

$$\boldsymbol{\phi}_r = (\phi_{r,k})_{k\in\mathbb{Z}} \in V; \qquad \phi_{r,k} = b_k^{\frac{1}{2}} e(-k\alpha_r)$$

Then Theorem 20.2 says that

$$\sum_{r=1}^{R}\Big|\sum_{k=-K}^{K} a_k e(k\alpha_r)\Big|^2 \leq A\sum_{k=-K}^{K}|a_k|^2 b_k^{-1}$$

where

$$A = \max_r \sum_{s=1}^{R}\big|(\boldsymbol{\phi}_r,\boldsymbol{\phi}_s)\big| = \max_r \sum_{s=1}^{R}\Big|\sum_{k=-\infty}^{\infty} b_k e\big(k(\alpha_s-\alpha_r)\big)\Big| = \max_r \sum_{s=1}^{R}\Big|B(\alpha_s-\alpha_r)\Big|$$

where

$$B(\alpha) := \sum_{k=-\infty}^{\infty} b_k e(k\alpha).$$

Hence in order to prove (627) and thus completing the proof of Theorem 20.4, it now suffices to choose the b_k's in such a way that $b_k \geq 1$ when $-K \leq k \leq K$ (and $b_k \geq 0$, $\forall k\in\mathbb{Z}$ and $\sum_k b_k < \infty$ as before) and such that

$$\sum_{s=1}^{R}\Big|B(\alpha_s-\alpha_r)\Big| \leq 2K + 3\delta^{-1} \tag{628}$$

for all $r \in \{1,\dots,R\}$.

We may first note that if we were to take $b_k = 1$ for $-K \leq k \leq K$ and $b_k = 0$ otherwise, we would obtain the inferior estimate

$$\sum_{s=1}^{R}\Big|B(\alpha_s-\alpha_r)\Big| \leq 2K + O\big(\delta^{-1}\log(\delta^{-1})\big). \tag{629}$$

(Cf. Problem 20.1 below.) To obtain a sharper estimate we take "smoother" b_k, namely

$$b_k = \begin{cases} 1 & \text{if } |k| \leq K, \\ 1-(|k|-K)/L & \text{if } K \leq |k| \leq K+L, \\ 0 & \text{if } |k| \geq K+L, \end{cases}$$

where L is a positive number to be chosen later. To write $B(\alpha)$ in closed form we first note the identity (for $\alpha \notin \mathbb{Z}$).

$$\sum_{|j|\le J}(J-|j|)e(j\alpha) = \sum_{|j|\le J} e(j\alpha)\cdot \#\big\{j_1,j_2\in\{1,\dots,J\} \,:\, j_1-j_2=j\big\}$$
$$= \sum_{j_1=1}^{J}\sum_{j_2=1}^{J} e((j_1-j_2)\alpha) = \Big|\sum_{j=1}^{J} e(j\alpha)\Big|^2$$
$$= \Big|\frac{e((J+1)\alpha)-e(\alpha)}{e(\alpha)-1}\Big|^2 = \Big|\frac{e(\frac12 J\alpha)-e(-\frac12 J\alpha)}{e(\frac12\alpha)-e(-\frac12\alpha)}\Big|^2 = \Big(\frac{\sin\pi J\alpha}{\sin\pi\alpha}\Big)^2.$$

We apply this identity firstly with $J=K+L$ and secondly with $J=K$, and then subtract and divide by L. This gives:

$$B(\alpha) = \frac{\sin^2\big(\pi(K+L)\alpha\big)-\sin^2(\pi K\alpha)}{L\sin^2(\pi\alpha)} \qquad (\alpha\notin\mathbb{Z}).$$

Hence by letting $\alpha\to 0$ we get $B(0)=2K+L$ (as is of course also clear directly from $B(0)=\sum b_k$). The above formula also implies

$$\big|B(\alpha)\big| \le \frac{1}{L\sin^2(\pi\alpha)} \le \frac{1}{4L\|\alpha\|^2}, \qquad \forall\alpha\in\mathbb{R}\setminus\mathbb{Z}.$$

(In the last step we used $\sin^2(\pi\alpha)\ge 4\big\|\alpha\big\|^2$; to prove this it suffices to consider $0\le\alpha\le\frac12$, since both sides of the inequality are even and periodic modulo 1 as functions of α; and for $0\le\alpha\le\frac12$ the inequality follows from $\sin(\pi\alpha)\ge 2\alpha$, which is clear since $\sin(\pi\alpha)$ is concave for $0\le\alpha\le\frac12$.)

Fix an arbitrary $r\in\{1,2,\dots,R\}$. Now since $B(\alpha_s-\alpha_r)$ only depends on $\alpha_s \bmod 1$, we may as well assume that all points $\alpha_1,\alpha_2,\dots,\alpha_R$ lie in $(\alpha_r-\frac12,\alpha_r+\frac12]$. We *order* these points as $\alpha'_{1-S}\le\alpha'_{2-S}\le\ldots\le\alpha'_{R-S}$ where S is adjusted so that $\alpha'_0=\alpha_r$. Then since our points are separated by at least δ, we have $\big\|\alpha'_s-\alpha_r\big\|\ge|s|\delta$ for all s, and hence

$$\sum_{s=1}^{R}\Big|B(\alpha_s-\alpha_r)\Big| = B(0)+\sum_{\substack{1-S\le s\le R-S\\ s\neq 0}}\Big|B(\alpha'_s-\alpha_r)\Big| \le 2K+L+2\sum_{s=1}^{\infty}\frac{1}{4Ls^2\delta^2}$$
$$= 2K+L+\frac{\zeta(2)}{2L\delta^2} \le 2K+L+\frac{1}{L\delta^2},$$

where we used the fact that $\zeta(2)<2$, cf. footnote 24 on p. 188. Now set $L=\lceil\delta^{-1}\rceil$; then the above is

$$\le 2K+\delta^{-1}+1+\delta^{-1}\le 2K+3\delta^{-1},$$

since $\delta\le\frac12$. Thus we have (628), and the proof is complete. □

Remark 20.1. Note the very simple and beautiful proof by Gallagher [21] given on pp. 156-7 in Davenport's book, of the fact that the large sieve inequality holds with $\Delta = \pi N + \delta^{-1}$, which is sharper than Theorem 20.4 when $N\delta$ is small.

Atle Selberg has proved that the large sieve inequality holds with $\Delta = N + \delta^{-1} - 1$, which is sharper than both Theorem 20.4 and Gallagher's bound. (Cf. the survey article paper [47] by Montgomery, as well as Vaaler [71], and Bombieri [6].)

We now give some applications of the large sieve. Our starting point will be the following special case of the large sieve inequality.

Proposition 20.5. *For any $Q, N \in \mathbb{Z}^+$, $M \in \mathbb{Z}$ and any complex numbers a_n, we have*

$$\sum_{q\le Q}\sum_{\substack{1\le a\le q\\(a,q)=1}}\Big|S\Big(\frac{a}{q}\Big)\Big|^2 \le (N+3Q^2)\sum_{n=M+1}^{M+N}|a_n|^2. \tag{630}$$

Proof. This follows by applying the large sieve inequality (Theorem 20.4) with the points α_r being the so called *Farey fractions* with denominator $\le Q$, i.e. all the points in the set

$$\mathcal{F}_Q = \Big\{\frac{a}{q} \,:\, 1\le a\le q\le Q,\ (a,q)=1\Big\}.$$

If $\frac{a}{q}$ and $\frac{a'}{q'}$ are two distinct such fractions, then

$$\Big\|\frac{a}{q}-\frac{a'}{q'}\Big\| = \Big\|\frac{aq'-a'q}{qq'}\Big\| \ge \frac{1}{qq'} \ge \frac{1}{Q^2},$$

and hence we can apply the large sieve with $\delta = Q^{-2}$. This gives the stated inequality. □

We now use the above result to formulate the large sieve in the manner of Rényi.

Theorem 20.6. *Let $\mathcal{N}$ be a set of Z integers in the interval $M+1 \le n \le M+N$, and let $Z(q,h)$ denote the number of these integers which are congruent to $h \pmod q$. Let $\mathcal{P}$ be a set of P prime numbers p, with $p \le Q$ for all $p \in \mathcal{P}$. Let $0 < \tau < 1$, and suppose that $Z(p,h) = 0$ for at least τp values of $h \pmod p$, for all $p \in \mathcal{P}$. Then*

$$Z \le \frac{N+3Q^2}{\tau P}. \tag{631}$$

Proof. We use Proposition 20.5 with a_n specialized to be the characteristic function of the set $\mathcal{N}$, so that $S(\alpha) = \sum_{n\in\mathcal{N}} e(n\alpha)$. Then for any $q \ge 1$ we have

$$\sum_{a=1}^{q}\Big|S\Big(\frac{a}{q}\Big)\Big|^2 = \sum_{a=1}^{q}\sum_{m\in\mathcal{N}}\sum_{n\in\mathcal{N}} e\Big(\frac{a(m-n)}{q}\Big) = \sum_{m\in\mathcal{N}}\sum_{n\in\mathcal{N}}\sum_{a=1}^{q} e\Big(\frac{a(m-n)}{q}\Big)$$

The innermost sum is q if $m \equiv n \pmod q$, but 0 if $m \not\equiv n \pmod q$; hence we get

$$= q \sum_{m\in\mathcal{N}} \sum_{\substack{n\in\mathcal{N} \\ n\equiv m \,(\mathrm{mod}\ q)}} 1 = q \sum_{h=1}^{q} \#\big\{ \langle m,n\rangle \in \mathcal{N} \,:\, m \equiv n \equiv h \,(\mathrm{mod}\ q)\big\} = q \sum_{h=1}^{q} Z(q,h)^2.$$

We can use this to get an upper bound on the *variance*[33] of $Z(q,h)$;

$$V(q) := \frac{1}{q} \sum_{h=1}^{q} \Big(Z(q,h) - \frac{Z}{q} \Big)^2, \tag{632}$$

averaged over q running through prime numbers. Indeed, for any $q \in \mathbb{Z}^+$ we have

$$q^2 V(q) = q \sum_{h=1}^{q} \Big(Z(q,h) - \frac{Z}{q} \Big)^2 = q \sum_{h=1}^{q} Z(q,h)^2 - 2Z \sum_{h=1}^{q} Z(q,h) + Z^2$$
$$= q \sum_{h=1}^{q} Z(q,h)^2 - Z^2,$$

and combining the above relations we get

$$q^2 V(q) = \sum_{a=1}^{q} \Big| S\Big(\frac{a}{q}\Big)\Big|^2 - Z^2 = \sum_{a=1}^{q-1} \Big| S\Big(\frac{a}{q}\Big)\Big|^2, \tag{633}$$

since $S(0) = Z$. Let us here specialize q to be a *prime*; $q = p$. Then

$$\sum_{a=1}^{p-1} \Big| S\Big(\frac{a}{p}\Big)\Big|^2 = \sum_{\substack{1\le a\le p \\ (a,p)=1}} \Big| S\Big(\frac{a}{p}\Big)\Big|^2,$$

and hence (633) together with Proposition 20.5 implies

$$\sum_{p\le Q} p^2 V(p) \le (N + 3Q^2) Z. \tag{634}$$

This inequality (which is of interest in itself) was proved for an *arbitrary* subset $\mathcal{N} \subset \{M+1, M+2, \dots, M+N\}$. Now assume that $\mathcal{N}$ satisfies the assumption made in the theorem; then for each $p \in \mathcal{P}$ there are at least τp values of $h \,(\mathrm{mod}\ p)$ such that $\big(Z(p,h) - \frac{Z}{p}\big)^2 = Z^2/p^2$, and hence $V(p) \ge \tau Z^2/p^2$, by (632). Hence (634) implies

$$\tau P Z^2 \le (N + 3Q^2) Z,$$

which completes the proof of Theorem 20.6. □

[33] To be precise, $V(q)$ is the variance of $Z(q,h)$ for given q if h is taken to be a (discrete) random variable with uniform distribution in $\mathbb{Z}/q\mathbb{Z}$. To see this we need only note that the expected value of $Z(q,h)$ equals $\frac{1}{q}\sum_{h=1}^{q} Z(q,h) = \frac{Z}{q}$. Note that our $V(q)$ differs from Davenport's $V(q)$ by a factor of $\frac{1}{q}$, which we have inserted to make $V(q)$ truely be the variance of $Z(q,h)$.

Remark 20.2. To appreciate the strength of the bound in Theorem 20.6, suppose that $\mathcal{N}$ is the set of squares in the interval $1 \le n \le N$ with N large. Set $Q = N^{\frac{1}{2}}$, and let $\mathcal{P}$ be the set of odd primes $p \le N^{\frac{1}{2}}$. Then $Z(p,h) = 0$ for quadratic nonresidues $h \pmod p$, so that $Z(p,h) = 0$ for at least $\frac{1}{2}(p-1)$ values of h. Hence $\tau = \frac{1}{3}$ and $P \sim 2N^{\frac{1}{2}}/\log N$, and we obtain the bound $Z \ll N^{\frac{1}{2}} \log N$, which is not far from the truth, $Z \sim N^{\frac{1}{2}}$.

Remark 20.3. One can prove that for a *fixed* set $\mathcal{P}$ of P primes, and a *fixed* $0 < \tau < 1$, if Z_N is the largest possible cardinality of a set $\mathcal{N}$ of integers in the interval $1 \le n \le N$ such that $Z(p,h) = 0$ for at least τp values of $h \pmod p$, for all $p \in \mathcal{P}$, then

$$\lim_{N\to\infty} \frac{Z_N}{N} = \prod_{p\in\mathcal{P}} \frac{\lfloor (1-\tau)p \rfloor}{p} \le (1-\tau)^P \tag{635}$$

(cf. Problem 20.2 below). On the other hand Theorem 20.6 implies the bound

$$\limsup_{N\to\infty} \frac{Z_N}{N} \le \frac{1}{\tau P}.$$

This is often a lot weaker than (635), but in the special case when $\tau \approx \frac{1}{P}$ it *agrees* with (635), up to a constant factor. However, a key feature of Theorem 20.6 is of course that it gives a very useful bound in cases when N is *not* extremely much larger than P!

[Details: The fact that $\frac{1}{\tau P} > (1-\tau)^P$ for all $P \ge 1$ and $0 < \tau < 1$ is easily checked e.g. by differentiation: For given $P \ge 1$ we have to prove that the function $f(\tau) = \tau(1-\tau)^P$ is always $< \frac{1}{P}$ for $0 < \tau < 1$. Now $f'(\tau) = (1-\tau)^{P-1}\big(1-(P+1)\tau\big)$; hence $f(\tau)$ is increasing for $0 \le \tau \le \frac{1}{P+1}$ and decreasing for $\frac{1}{P+1} \le \tau \le 1$, and for all $0 \le \tau \le 1$ we have $f(\tau) \le f(\frac{1}{P+1}) = \frac{1}{P+1}\big(1-\frac{1}{P+1}\big)^P < \frac{1}{P}$. Note also that if $\tau = \frac{1}{P+1}$ then $(1-\tau)^P \sim e^{-1}$ as $P \to \infty$, and the same holds if $\tau = \frac{1}{P}$; in particular if $\tau = \frac{1}{P}$ we have $(1-\tau)^P \gg 1 = \frac{1}{\tau P}$ for all $P \ge 1$.]

Finally we give an application (due to Gallagher) of the large sieve to estimating averages of character sums. Let X_q^* be the set of all *primitive* characters $\chi \in X_q$.

Theorem 20.7. *For any $M \in \mathbb{Z}$, $Q, N \in \mathbb{Z}^+$ and any complex numbers a_n we have*

$$\sum_{1\le q\le Q} \frac{q}{\phi(q)} \sum_{\chi\in X_q^*} \Big| \sum_{n=M+1}^{M+N} a_n \chi(n) \Big|^2 \le (N+3Q^2) \sum_{n=M+1}^{M+N} |a_n|^2. \tag{636}$$

Proof. Recall from Lemma 9.4 and Lemma 9.5 that, if $\chi \in X_q^*$,

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{h=1}^{q} \overline{\chi}(h) e\Big(\frac{hn}{q}\Big)$$

for all $n \in \mathbb{Z}$, and that the Gauss sum $\tau(\overline{\chi})$ has absolute value $\big|\tau(\overline{\chi})\big| = \sqrt{q}$. On multiplying both sides by a_n and summing, we see that

$$\sum_{n=M+1}^{M+N} a_n \chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{h=1}^{q} \overline{\chi}(h) \sum_{n=M+1}^{M+N} a_n e\Big(\frac{hn}{q}\Big) = \frac{1}{\tau(\overline{\chi})} \sum_{h=1}^{q} \overline{\chi}(h) S\Big(\frac{h}{q}\Big),$$

where we write $S(\alpha) = \sum_{n=M+1}^{M+N} a_n e(n\alpha)$ as before. Hence, for any $q \geq 1$,

$$\begin{aligned}\sum_{\chi \in X_q^*} \Big| \sum_{n=M+1}^{M+N} a_n \chi(n)\Big|^2 &= \frac{1}{q} \sum_{\chi \in X_q^*} \Big|\sum_{h=1}^{q} \overline{\chi}(h) S\Big(\frac{h}{q}\Big)\Big|^2 \leq \frac{1}{q} \sum_{\chi \in X_q} \Big|\sum_{h=1}^{q} \overline{\chi}(h) S\Big(\frac{h}{q}\Big)\Big|^2 \\ &= \frac{1}{q} \sum_{\chi \in X_q} \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} \overline{\chi}(h_1)\chi(h_2) S\Big(\frac{h_1}{q}\Big)\overline{S\Big(\frac{h_2}{q}\Big)} \\ &= \frac{1}{q} \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} S\Big(\frac{h_1}{q}\Big)\overline{S\Big(\frac{h_2}{q}\Big)} \sum_{\chi \in X_q} \overline{\chi}(h_1)\chi(h_2),\end{aligned}$$

and using Lemma 1.5 we get

$$= \frac{\phi(q)}{q} \sum_{\substack{1 \leq h_1 \leq q \\ (h_1,q)=1}} \Big|S\Big(\frac{h_1}{q}\Big)\Big|^2.$$

Hence

$$\sum_{1 \leq q \leq Q} \frac{q}{\phi(q)} \sum_{\chi \in X_q^*} \Big| \sum_{n=M+1}^{M+N} a_n \chi(n)\Big|^2 \leq \sum_{1 \leq q \leq Q} \sum_{\substack{1 \leq h_1 \leq q \\ (h_1,q)=1}} \Big|S\Big(\frac{h_1}{q}\Big)\Big|^2 \leq (N + 3Q^2) \sum_{n=M+1}^{M+N} |a_n|^2,$$

where in the last step we used the large sieve inequality in the form of Proposition 20.5. □

20.1. **Problems.**

Problem 20.1. Prove that if $b_k = 1$ for $-K \leq k \leq K$ and $b_k = 0$ otherwise, then the bound (629) holds.

Problem 20.2. Let $\mathcal{P}$ be a fixed set of P primes and let $0 < \tau < 1$ be a fixed number. For each $N \in \mathbb{Z}^+$ we let Z_N denote the largest possible cardinality of a set $\mathcal{N}$ of integers in the interval $1 \leq n \leq N$ such that $Z(p,h) = 0$ for at least τp values of $h \pmod p$, for all $p \in \mathcal{P}$. Prove that

$$Z_N \sim N \prod_{p \in \mathcal{P}} \frac{\lfloor (1-\tau)p \rfloor}{p} \qquad \text{as } N \to \infty. \tag{637}$$

21. Bombieri's Theorem

(Davenport Chapter 28.)

Definition 21.1. For $q \in \mathbb{Z}^+$, $a \in \mathbb{Z}$ with $(a,q)=1$ and $x \geq 1$ we put

$$E(x;q,a) = \psi(x;q,a) - \frac{x}{\phi(q)}, \tag{638}$$

$$E(x,q) = \max\big\{|E(x;q,a)| \,:\, a \in \mathbb{Z},\ (a,q)=1\big\}, \tag{639}$$

and

$$E^*(x,q) = \max_{y \leq x} E(y,q). \tag{640}$$

The following theorem by Bombieri says that $E^*(x,q)$ is significantly smaller than $x/\phi(q)$ for most $q \leq x^{\frac{1}{2}}(\log x)^{-A}$.

Theorem 21.1. *Let $A > 0$ be fixed. Then for all $x \geq 2$ and all Q in the range $x^{\frac{1}{2}}(\log x)^{-A} \leq Q \leq x^{\frac{1}{2}}$ we have*

$$\sum_{1 \leq q \leq Q} E^*(x,q) \ll x^{\frac{1}{2}} Q (\log x)^5, \tag{641}$$

where the implied constant only depends on A, but is ***non-effective****.*

(The dependence on A and the non-effectiveness of the implied constant only arises at the very end of the proof when we apply Siegel's Theorem; cf. (670) below.)

Remark 21.1. Note that Theorem 21.1 implies Theorem 16.7. (Proof: For any given $A > 0$ we may apply Theorem 21.1 with $Q = x^{\frac{1}{2}}(\log x)^{-(A+5)}$ and $A+5$ in place of A, to deduce that $\sum_{1 \leq q \leq x^{\frac{1}{2}}(\log x)^{-(A+5)}} E^*(x,q) \ll x(\log x)^{-A}$, viz.

$$\sum_{1 \leq q \leq x^{\frac{1}{2}}(\log x)^{-(A+5)}} \max_{y \leq x} \max_{(a,q)=1} \Big|\psi(y;q,a) - \frac{y}{\phi(q)}\Big| \ll x(\log x)^{-A}.$$

Hence the statement of Theorem 16.7 holds with $B = A+5$.)

Remark 21.2. To assess the strength of the bound in Theorem 21.1 we note that for each $y \geq 1$ there are at most $\frac{y}{q}+1$ integers $1 \leq n \leq y$, $n \equiv a \pmod q$, and hence, given any large x we have for all $q, y \leq x$ and all a:

$$\psi(y;q,a) \ll \frac{x \log x}{q} \qquad \text{and} \qquad \frac{y}{\phi(q)} \ll \frac{x \log(q+1)}{q} \ll \frac{x \log x}{q}$$

(we used $\phi(q) \gg \frac{q}{\log(q+1)}$, cf. Problem 15.1). Thus

$$E^*(x,q) \ll \frac{x \log x}{q}, \qquad \forall q \leq x.$$

Consequently we have the trivial bound

$$\sum_{1\le q\le Q} E^*(x,q) \ll \sum_{1\le q\le Q} \frac{x\log x}{q} \ll x(\log x)^2 \tag{642}$$

for all $Q \le x$. (Note that we don't get any better bound by this trivial approach even if we take Q to be significantly smaller than x; for instance if $Q \approx x^{\frac{1}{1000}}$ we still obtain the *same* bound, $\sum_{1\le q\le Q} E^*(x,q) \ll x(\log x)^2$.) From this we see that the result of Theorem 21.1 is *trivial* for $Q \gg x^{\frac12}(\log x)^{-3}$. We also see that the result of Theorem 21.1 never represents more than a *log power saving* versus the trivial bound (642), since the bound in the right hand side of (641) is always $\ge x(\log x)^{5-A}$ where A is the fixed constant in Theorem 21.1. However, *this log power saving is a very important one:* As we saw in Problem 16.4, Theorem 21.1 implies that the asymptotic relation $\psi(x;q,a) \sim \frac{x}{\phi(q)}$ holds for x large and for "almost" all q-values in a range which is essentially of the same quality as what GRH gives! (Namely: In the range $q \le x^{\frac12}(\log x)^{-B}$, where B is any fixed constant > 5.)

We now start on the proof of Theorem 21.1, which falls into several parts. (Note that we have structured the proof in a slightly different order than in Davenport's book.) The main work will be spent on proving the following bound. Recall that we denote by X_q^* the set of all *primitive* characters $\chi \in X_q$; cf. p. 277.

Proposition 21.2. *We have, for all $x \ge 2$ and $Q \ge 1$,*

$$\sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi\in X_q^*} \max_{y\le x} \big|\psi(y,\chi)\big| \ll (x + x^{\frac56}Q + x^{\frac12}Q^2)(\log Qx)^4, \tag{643}$$

where the implied constant is absolute.

Theorem 21.1 is in fact a (comparatively) direct and easy consequence of Proposition 21.2 combined with our estimates from the prime number theorem for arithmetic progressions, in particular Siegel's Theorem. This deduction, Prop. 21.2 ⇒ Theorem 21.1, is carried out on pp. 288–292.

Remark 21.3. The bound (643) can also be expressed as follows:

$$\sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi\in X_q^*} \max_{y\le x} \big|\psi(y,\chi)\big| \ll \left\{ \begin{array}{ll} x & \text{if } Q \le x^{\frac16} \\ x^{\frac56}Q & \text{if } x^{\frac16} \le Q \le x^{\frac13} \\ x^{\frac12}Q^2 & \text{if } x^{\frac13} \le Q \end{array} \right\} (\log Qx)^4.$$

In order to prove Proposition 21.2 we will use our method of estimating $\sum_{n\le N} f(n)\Lambda(n)$ from §18. In Remark 18.3 we observed that this method fails if f is multiplicative; in particular we are not able to bound $\big|\psi(x,\chi)\big|$ by this method. Nevertheless we can use the method to bound an *average* of $\big|\psi(x,\chi)\big|$ over various χ, by using the large sieve. More precisely, we will use the large sieve to deduce the following estimate:

Proposition 21.3. *For any real $Q, M, N \geq 1$ and any complex numbers $\{a_m\}$ and $\{b_n\}$, we have*

$$\sum_{q\leq Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{u\in\mathbb{R}}\Big|\sum_{m\leq M}\sum_{\substack{n\leq N\\ mn\leq u}}a_m b_n\chi(mn)\Big| \ll (M+Q^2)^{\frac12}(N+Q^2)^{\frac12}\Big(\sum_{m\leq M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\leq N}|b_n|^2\Big)^{\frac12}\log(2MN). \tag{644}$$

Proof. We use the large sieve in the form of the inequality (cf. Theorem 20.7)

$$\sum_{q\leq Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\Big|\sum_{m\leq M}a_m\chi(m)\Big|^2 \ll (M+Q^2)\sum_{m\leq M}|a_m|^2, \tag{645}$$

and similarly for $\{b_n\}_{n\leq N}$. Using also Cauchy's inequality we get

$$\begin{aligned}
&\sum_{q\leq Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\Big|\sum_{m\leq M}\sum_{n\leq N}a_m b_n\chi(mn)\Big|\\
&=\sum_{q\leq Q}\sum_{\chi\in X_q^*}\Big|\Big(\sqrt{\frac{q}{\phi(q)}}\sum_{m\leq M}a_m\chi(m)\Big)\Big(\sqrt{\frac{q}{\phi(q)}}\sum_{n\leq N}b_n\chi(n)\Big)\Big|\\
&=\Big(\sum_{q\leq Q}\sum_{\chi\in X_q^*}\frac{q}{\phi(q)}\Big|\sum_{m\leq M}a_m\chi(m)\Big|^2\Big)^{\frac12}\Big(\sum_{q\leq Q}\sum_{\chi\in X_q^*}\frac{q}{\phi(q)}\Big|\sum_{n\leq N}b_n\chi(n)\Big|^2\Big)^{\frac12}\\
&\ll (M+Q^2)^{\frac12}(N+Q^2)^{\frac12}\Big(\sum_{m\leq M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\leq N}|b_n|^2\Big)^{\frac12}.
\end{aligned} \tag{646}$$

To introduce the condition $mn\leq u$ in the inner summation we will use the following formula:

Lemma 21.4. *For any $T>0, \beta>0, \alpha\in\mathbb{R}$ with $\beta\neq|\alpha|$ we have*

$$\int_{-T}^{T}e^{it\alpha}\frac{\sin t\beta}{\pi t}\,dt=\left\{\begin{matrix}1 & \text{if } |\alpha|<\beta\\ 0 & \text{if } |\alpha|>\beta\end{matrix}\right\}+O\Big(\frac{1}{T\big|\beta-|\alpha|\big|}\Big), \tag{647}$$

where the implied constant is absolute.

Proof. By Lemma 13.3 we have, for all $y>0$ with $y\neq1$, and all $c>0$, $T>0$:

$$\Big|\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s}{s}\,ds-\delta(y)\Big|<\frac{y^c}{\pi T|\log y|},$$

where $\delta(y)=0$ if $0<y<1$ while $\delta(y)=1$ if $y>1$. Here we may let $c\to0^+$ to conclude

$$\Big|\frac{1}{2\pi i}\int_{-iT}^{iT}\frac{y^s}{s}\,ds-\delta(y)\Big|\leq\frac{1}{\pi T|\log y|}.$$

In the integral we substitute $s = it$; this gives

$$\frac{1}{2i}\int_{-T}^{T} \frac{y^{it}}{\pi t}\, dt = \delta(y) + O\Big(\frac{1}{T|\log y|}\Big), \qquad \forall T > 0,\ y \in \mathbb{R}^+ \setminus \{1\}.$$

We apply this formula for $y = e^{\alpha+\beta}$ and $y = e^{\alpha-\beta}$, and subtract; since $\frac{1}{2i}\big(e^{(\alpha+\beta)it} - e^{(\alpha-\beta)it}\big) = e^{it\alpha}\sin t\beta$ this gives:

$$\int_{-T}^{T} e^{it\alpha}\frac{\sin t\beta}{\pi t}\, dt = \delta(e^{\alpha+\beta}) - \delta(e^{\alpha-\beta}) + O\Big(\frac{1}{T|\alpha+\beta|} + \frac{1}{T|\alpha-\beta|}\Big), \qquad \forall T > 0.$$

Here if $\beta > |\alpha|$ then $\delta(e^{\alpha+\beta}) = 1$ and $\delta(e^{\alpha-\beta}) = 0$, while if $0 < \beta < |\alpha|$ then $\delta(e^{\alpha+\beta}) = \delta(e^{\alpha-\beta})$. We also have $\frac{1}{T|\alpha+\beta|} \le \frac{1}{T|\beta-|\alpha||}$ and $\frac{1}{T|\alpha-\beta|} \le \frac{1}{T|\beta-|\alpha||}$, since $\beta > 0$. Hence we get (647). □

Applying the above lemma with $\beta = \log u$ ($u > 1$) and $\alpha = -\log(mn)$ ($m, n \in \mathbb{Z}^+$) we obtain

$$\int_{-T}^{T} m^{-it}n^{-it}\frac{\sin(t\log u)}{\pi t}\, dt = \begin{Bmatrix} 1 & \text{if } mn < u \\ 0 & \text{if } mn > u \end{Bmatrix} + O\Big(T^{-1}\Big|\log\frac{mn}{u}\Big|^{-1}\Big),$$

so long as $mn \neq u$. Multiplying this formula with $a_m b_n \chi(mn)$ and adding over all $\langle m, n\rangle \in \{1, \dots, \lfloor M\rfloor\} \times \{1, \dots, \lfloor N\rfloor\}$ we get, for any $u > 1$ which is not an integer, and any $T > 0$:

$$\sum_{m\le M}\sum_{\substack{n\le N\\ mn\le u}} a_m b_n \chi(mn) = \int_{-T}^{T} A(t,\chi)B(t,\chi)\frac{\sin(t\log u)}{\pi t}\, dt + O\bigg(T^{-1}\sum_{m\le M}\sum_{n\le N}|a_m b_n|\Big|\log\frac{mn}{u}\Big|^{-1}\bigg), \tag{648}$$

where

$$A(t,\chi) = \sum_{m\le M} a_m \chi(m) m^{-it}, \qquad B(t,\chi) = \sum_{n\le N} b_n \chi(n) n^{-it}.$$

When proving (644) it is no loss of generality to assume that u only runs through the numbers of the form $u = k + \frac{1}{2}$, where k is an integer, $1 \le k \le MN$. (This is because the inner sum $\sum_{m\le M}\sum_{\substack{n\le N\\ mn\le u}} a_m b_n \chi(mn)$ in (644) only depends on the integer part of u, equals 0 for $u < 1$, and is constant for all $u \ge MN$.) For all such $u = k + \frac{1}{2}$, and all $\langle m, n\rangle \in \{1, \dots, \lfloor M\rfloor\} \times \{1, \dots, \lfloor N\rfloor\}$, using the fact that $\big|\log\alpha\big| \gg \min(1, |\alpha - 1|)$ for all $\alpha > 0$, we have:

$$\Big|\log\frac{mn}{u}\Big| \gg \min\Big(1, \Big|\frac{mn}{u} - 1\Big|\Big) \ge \min\Big(1, \Big|\frac{u \pm \frac{1}{2}}{u} - 1\Big|\Big) = \min\Big(1, \Big|\frac{1}{2u}\Big|\Big) \gg \frac{1}{u} \gg \frac{1}{MN}.$$

We also have

$$\big|\sin(t\log u)\big| \le \min(1, |t\log u|) \le \min\big(1, |t|\log(2MN)\big).$$

Hence from (648) we conclude

$$\begin{aligned}
&\max_u \Big|\sum_{m\le M}\sum_{\substack{n\le N\\ mn\le u}} a_m b_n \chi(mn)\Big| \\
&\ll \int_{-T}^{T} \big|A(t,\chi)B(t,\chi)\big| \min\Big(\frac{1}{|t|}, \log(2MN)\Big)\, dt + \frac{MN}{T}\sum_{m\le M}\sum_{n\le N}|a_m b_n|. \\
&\ll \int_{-T}^{T} \big|A(t,\chi)B(t,\chi)\big| \min\Big(\frac{1}{|t|}, \log(2MN)\Big)\, dt + \frac{M^{\frac32}N^{\frac32}}{T}\Big(\sum_{m\le M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\le N}|b_n|^2\Big)^{\frac12},
\end{aligned}$$

where in the last step we used Cauchy's inequality. Multiplying the above bound with $\frac{q}{\phi(q)}$ and adding over all $\langle q,\chi\rangle$ with $1\le q\le Q$, $\chi\in X_q^*$, we get

$$\begin{aligned}
&\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_u \Big|\sum_{m\le M}\sum_{\substack{n\le N\\ mn\le u}} a_m b_n \chi(mn)\Big| \\
&\ll \int_{-T}^{T}\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\big|A(t,\chi)B(t,\chi)\big| \min\Big(\frac{1}{|t|}, \log(2MN)\Big)\, dt \\
&\qquad + \frac{M^{\frac32}N^{\frac32}}{T}\Big(\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}1\Big)\Big(\sum_{m\le M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\le N}|b_n|^2\Big)^{\frac12}. \qquad (649)
\end{aligned}$$

Here by (646) (applied with $a_m m^{-it}$, $b_n n^{-it}$ in place of a_m, b_n) we have

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\big|A(t,\chi)B(t,\chi)\big| \ll (M+Q^2)^{\frac12}(N+Q^2)^{\frac12}\Big(\sum_{m\le M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\le N}|b_n|^2\Big)^{\frac12}.$$

(The implied constant is absolute, just like all other implied constants on the last few pages. In particular the last bound is uniform with respect to $t\in\mathbb{R}$.) Using this and $\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}1 \le \sum_{q\le Q} q \ll Q^2$ we see that (649) is

$$\begin{aligned}
&\ll (M+Q^2)^{\frac12}(N+Q^2)^{\frac12}\Big(\sum_{m\le M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\le N}|b_n|^2\Big)^{\frac12}\int_{-T}^{T}\min\Big(\frac{1}{|t|},\log(2MN)\Big)\,dt \\
&\qquad + \frac{M^{\frac32}N^{\frac32}}{T}Q^2\Big(\sum_{m\le M}|a_m|^2\Big)^{\frac12}\Big(\sum_{n\le N}|b_n|^2\Big)^{\frac12}.
\end{aligned}$$

Taking here $T=(MN)^{\frac32}$ we obtain (644). □

Proof of Proposition 21.2. If $Q^2>x$ then (644) implies (643) on taking $M=1$, $a_1=1$, $b_n=\Lambda(n)$, $N=x$.

We now assume $Q^2 \le x$, and prove (643) using the identity of §18 (cf. (572), (573), (574), (576), (577), (578), where we take y in place of "N")[34]: For any given $U, V, y \ge 1$ we have

$$\psi(y,\chi) = \sum_{j=1}^{4} S_j(y,\chi,U,V), \tag{650}$$

where

$$S_1(y,\chi,U,V) = \sum_{n\le \min(U,y)} \Lambda(n)\chi(n) = O(U); \tag{651}$$

$$S_2(y,\chi,U,V) = -\sum_{t\le UV} \Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}} \Lambda(m)\mu(d)\Big) \sum_{r\le y/t} \chi(rt); \tag{652}$$

$$S_3(y,\chi,U,V) = \sum_{d\le V} \mu(d) \sum_{h\le y/d} \chi(hd)\log h = O\Big((\log y)\sum_{d\le V} \max_{1\le w\le y/d} \Big|\sum_{h\le w} \chi(h)\Big|\Big); \tag{653}$$

$$S_4(y,\chi,U,V) = -\sum_{U<m\le y/V} \Lambda(m) \sum_{V<k\le y/m} \Big(\sum_{\substack{d|k\\ d\le V}} \mu(d)\Big)\chi(mk). \tag{654}$$

In the following discussion, from here until relation (664), we consider x, Q, U, V as given, and satisfying $x \ge 2$, $1 \le Q \le \sqrt{x}$, $U, V \ge 1$, $UV \le x$; and we work with *absolute* implied constants in all "big O" and "$\ll$" bounds. We will later choose U and V as functions of Q and x.

By (650) we have

$$\sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi\in X_q^*} \max_{y\le x} \big|\psi(y,\chi)\big| \le \sum_{j=1}^{4} \sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi\in X_q^*} \max_{y\le x} \big|S_j(y,\chi,U,V)\big|. \tag{655}$$

To treat the contribution from S_4 (viz., $j=4$) in (655) we will apply dyadic decomposition to the m-variable in (654). We first note that for any given real number $M \in [2,x]$, if we put

$$a_m = \begin{cases} \Lambda(m) & \text{if } \max(U,M) < m \le \min(\frac{x}{V}, 2M) \\ 0 & \text{else;} \end{cases} \qquad b_k = \begin{cases} \sum_{\substack{d|k\\ d\le V}} \mu(d) & \text{if } k > V \\ 0 & \text{else;} \end{cases}$$

[34]Note that the assumptions $U, V \ge 2$, $UV \le N$ made in Proposition 18.1 are not needed for the quoted identities to hold. Note however that the sum (654) is empty (i.e. $S_4 = 0$) if $UV \ge y$.

then

$$\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\Big|\sum_{\substack{U<m\le y/V\\ M<m\le 2M}}\Lambda(m)\sum_{V<k\le y/m}\Big(\sum_{\substack{d\mid k\\ d\le V}}\mu(d)\Big)\chi(mk)\Big|$$

$$=\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\Big|\sum_{1\le m\le 2M}\sum_{\substack{1\le k\le x/M\\ mk\le y}}a_m b_k\chi(mk)\Big|. \tag{656}$$

[Proof: For fixed q,χ in the above sum one checks by inspection that if the pair $\langle m,k\rangle$ occurs in the first sum, then it also occurs in the second sum and gives the same contribution there. On the other hand if the pair $\langle m,k\rangle$ occurs and gives a nonzero contribution in the second sum then $a_m\neq 0$ forces $U<m$ and $M<m\le 2M$, and $b_k\neq 0$ forces $V<k$, and furthermore $mk\le y$ so that $k\le y/m$ and $m\le y/k<y/V$; thus the pair $\langle m,k\rangle$ also occurs in the first sum and gives the same contribution there.]

Now if we replace "$\max_{y\le x}$" by "$\max_{y\in\mathbb{R}}$" in the last line of (656) (thus making the total sum *larger*), then this expression is exactly of the same form as the left hand side of (644) (with $M\leftarrow 2M$ and $N\leftarrow x/M$); hence Proposition 21.3 gives that (656) is

$$\ll (Q^2+M)^{\frac12}\Big(Q^2+\frac{x}{M}\Big)^{\frac12}\Big(\sum_{M<m\le 2M}\Lambda(m)^2\Big)^{\frac12}\Big(\sum_{V<k\le x/M}d(k)^2\Big)^{\frac12}\log x$$

$$\ll \big(Q+M^{\frac12}\big)\big(Q+x^{\frac12}M^{-\frac12}\big)\big(M\log M\big)^{\frac12}\Big(\frac{x}{M}\log^3 x\Big)^{\frac12}\log x$$

$$\ll \big(Q^2x^{\frac12}+QxM^{-\frac12}+Qx^{\frac12}M^{\frac12}+x\big)(\log x)^3$$

(cf. (581) and (582) and recall $2\le M\le x$). Adding the above bound over $M=2^\mu U$ for all integers $\mu\ge 0$ with $2^\mu U\le x/V$ we obtain

$$\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|S_4(y,\chi,U,V)\big|$$

$$\ll \sum_{\substack{\mu\ge 0\\ (2^\mu U\le x/V)}}\big(Q^2x^{\frac12}+Qx(2^\mu U)^{-\frac12}+Qx^{\frac12}(2^\mu U)^{\frac12}+x\big)(\log x)^3$$

$$\ll \big(Q^2x^{\frac12}+QxU^{-\frac12}+QxV^{-\frac12}+x\big)(\log x)^4. \tag{657}$$

To treat S_2 we consider two ranges of t in (652), namely $t\le U$ and $t>U$. That is, we write

$$S_2(y,\chi,U,V)=S_2'(y,\chi,U,V)+S_2''(y,\chi,U,V),$$

where

$$S_2'(y,\chi,U,V) = -\sum_{t\le U}\Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big)\sum_{r\le y/t}\chi(rt)$$

and

$$S_2''(y,\chi,U,V) = -\sum_{U<t\le UV}\Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big)\sum_{r\le y/t}\chi(rt).$$

(Note that $S_2''=0$ if $y\le U$.) We treat S_2'' exactly as we did S_4. Namely: We first note that for any given real number $T\in[2,x]$, if we put

$$a_t = \begin{cases} \sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d) & \text{if } \ \max(U,T)<t\le \min(UV,2T)\\ 0 & \text{else}\end{cases}$$

and $b_r=1$ $(\forall r)$, then

$$\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\Big|\sum_{\substack{U<t\le UV\\ T<t\le 2T}}\Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big)\sum_{r\le y/t}\chi(rt)\Big|$$

$$= \sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\Big|\sum_{1\le t\le 2T}\sum_{\substack{1\le r\le x/T\\ tr\le y}}a_t b_r\chi(rt)\Big|. \tag{658}$$

Now if we replace "$\max_{y\le x}$" by "$\max_{y\in\mathbb{R}}$" in the last line of (658) (thus making the total sum *larger*), then this expression is exactly of the same form as the left hand side of (644) (with $M\leftarrow 2T$ and $N\leftarrow x/T$); hence Proposition 21.3 gives that (658) is

$$\ll (Q^2+T)^{\frac12}\Big(Q^2+\frac{x}{T}\Big)^{\frac12}\Big(\sum_{T<t\le 2T}\Big(\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big)^2\Big)^{\frac12}\Big(\frac{x}{T}\Big)^{\frac12}\log x \tag{659}$$

Here note that for any $t\in\mathbb{Z}^+$ we have

$$\Big|\sum_{\substack{md=t\\ m\le U\\ d\le V}}\Lambda(m)\mu(d)\Big| \le \sum_{m|t}\Lambda(m) = \sum_{p|t}\sum_{r=1}^{\mathrm{ord}_p(t)}\Lambda(p^r) = \sum_{p|t}\mathrm{ord}_p(t)\log p = \log t. \tag{660}$$

and hence (659) is

$$\begin{aligned}&\ll \big(Q+T^{\frac12}\big)\big(Q+x^{\frac12}T^{-\frac12}\big)\big(T\log^2 T\big)^{\frac12}\Big(\frac{x}{T}\Big)^{\frac12}\log x\\&\ll \big(Q^2x^{\frac12}+QxT^{-\frac12}+Qx^{\frac12}T^{\frac12}+x\big)(\log x)^2.\end{aligned}$$

Adding the above bound over $T=2^\tau U$ for all integers $\tau\ge 0$ with $2^\tau U\le UV$ we obtain

$$\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|S_2''(y,\chi,U,V)\big|\ll \big(Q^2x^{\frac12}+QxU^{-\frac12}+Qx^{\frac12}U^{\frac12}V^{\frac12}+x\big)(\log x)^3. \tag{661}$$

We next treat S_2': Using (660) we have

$$\big|S_2'(y,\chi,U,V)\big|\ll\sum_{t\le U}(\log t)\Big|\sum_{r\le y/t}\chi(t)\chi(r)\Big|\le(\log U)\sum_{t\le U}\Big|\sum_{r\le y/t}\chi(r)\Big|,$$

and by the Pólya-Vinogradov inequality, Theorem 17.1, we get, if $\chi\in X_q^*$ with $q>1$,

$$\big|S_2'(y,\chi,U,V)\big|\ll(\log U)Uq^{\frac12}\log q\ll q^{\frac12}U(\log qU)^2,$$

since such a χ must be nonprincipal. On the other hand for $q=1$ (viz. $\chi=$ the trivial character) we have the bound

$$\big|S_2'(y,\chi,U,V)\big|\ll(\log U)\sum_{t\le U}\frac{y}{t}\ll y(\log U)^2.$$

Combining these estimates we find that

$$\begin{aligned}\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|S_2'(y,\chi,U,V)\big|&\ll x(\log U)^2+\sum_{2\le q\le Q}q^{\frac32}U(\log qU)^2\\&\ll x(\log U)^2+Q^{\frac52}U(\log QU)^2\ll\big(x+Q^{\frac52}U\big)(\log x)^2\end{aligned} \tag{662}$$

(since we are assuming $Q^2\le x$ and $U\le x$).

We treat S_3 as we did S_2': From (653) and the Pólya-Vinogradov inequality (Theorem 17.1) we have, if $\chi\in X_q^*$ with $q>1$,

$$\big|S_3(y,\chi,U,V)\big|\ll(\log y)\sum_{d\le V}q^{\frac12}\log q\le q^{\frac12}V(\log qx)^2,$$

while if $q=1$,

$$\big|S_3(y,\chi,U,V)\big|\ll(\log y)\sum_{d\le V}\frac{y}{d}=y(\log y)(\log V)\le y(\log yV)^2.$$

Hence

$$\begin{aligned}\sum_{1\le q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|S_3(y,\chi,U,V)\big|&\ll x(\log Vx)^2+\sum_{2\le q\le Q}q^{\frac32}V(\log qx)^2\\&\ll\big(x+Q^{\frac52}V\big)(\log x)^2.\end{aligned} \tag{663}$$

Using the estimates (651), (657), (661), (662) and (663) in (655) we get

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi(y,\chi)\big| \ll \big(Q^2x^{\frac12}+x+QxU^{-\frac12}+QxV^{-\frac12}+U^{\frac12}V^{\frac12}Qx^{\frac12} + Q^{\frac52}U+Q^{\frac52}V\big)(\log x)^4. \tag{664}$$

If $x^{\frac13}\le Q\le x^{\frac12}$ then we take $U=V=x^{\frac23}Q^{-1}$, and get

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi(y,\chi)\big| \ll \big(Q^2x^{\frac12}+x+Q^{\frac32}x^{\frac23}+x^{\frac76}\big)(\log x)^4 \ll Q^2x^{\frac12}(\log x)^4$$

(we used $x^{\frac13}\le Q$ in the last step to see that all of x, $x^{\frac76}$, $Q^{\frac32}x^{\frac23}$ are $\le Q^2x^{\frac12}$). In the other case, $1\le Q<x^{\frac13}$, we take $U=V=x^{\frac13}$, and get

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi(y,\chi)\big| \ll \big(Q^2x^{\frac12}+x+Qx^{\frac56}+Q^{\frac52}x^{\frac13}\big)(\log x)^4 \ll \big(Q^2x^{\frac12}+x+Qx^{\frac56}\big)(\log x)^4$$

(we used $Q<x^{\frac13}$ in the last step to get $Q^{\frac52}x^{\frac13}<Qx^{\frac56}$). Hence in both cases (643) holds, and we are done. □

Remark 21.4. The choice of U,V in (664) can be motivated as follows: For any given number $Y\in[1,x]$, if we vary $U,V\in[1,Y]$ in such a way that that $UV=Y$, then (664) is minimized, up to at worst a factor 2, when $U=V=\sqrt{Y}$. [Proof: For fixed $Q\ge1$, $x\ge2$ we set $f(U)=QxU^{-\frac12}+Q^{\frac52}U$, so that our task is to minimize the function $f(U)+f(Y/U)$ over $U\in[1,Y]$. By symmetry ($U\leftrightarrow Y/U$) we may assume $1\le U\le\sqrt{Y}$. By differentiation we find that $f(U)$ is decreasing for $U<U_0$ and increasing for $U>U_0$, where $U_0:=(x/2)^{\frac23}Q^{-1}$. Hence if $\sqrt{Y}\le U_0$ then $f(U)\ge f(\sqrt{Y})$, while if $\sqrt{Y}\ge U_0$ then $f(Y/U)\ge f(\sqrt{Y})$. Hence in both cases we have

$$f(U)+f(Y/U)\ge f(\sqrt{Y}),$$

which proves the claim.] It follows from this that we may keep $U=V$ when minimizing (664); thus our task is to choose $U\in[1,x]$ so as to minimize

$$QxU^{-\frac12}+UQx^{\frac12}+Q^{\frac52}U = QxU^{-\frac12}+\big(Qx^{\frac12}+Q^{\frac52}\big)U.$$

If $Q<x^{\frac13}$ then this expression is $\ll\gg QxU^{-\frac12}+Qx^{\frac12}U$ and in the usual way we see that this is minimized (up to an absolute constant) by making both terms equal, viz. by taking $U=x^{\frac13}$. On the other hand if $Q\ge x^{\frac13}$ then the above expression is $\ll\gg QxU^{-\frac12}+Q^{\frac52}U$, which is minimized (up to an absolute constant) by taking $U=x^{\frac23}Q^{-1}$.

Finally we now use Proposition 21.2 to prove Theorem 21.1:

Proof of Theorem 21.1. Recall that (by Lemma 1.5)

$$\psi(y;q,a)=\frac{1}{\phi(q)}\sum_{\chi\in X_q}\overline{\chi}(a)\psi(y,\chi).$$

From $\psi(y,\chi_0)$ we wish to subtract the main term y; hence we put

$$\psi'(y,\chi) := \begin{cases} \psi(y,\chi) & \text{if } \chi \neq \chi_0, \\ \psi(y,\chi_0) - y & \text{if } \chi = \chi_0. \end{cases}$$

Then

$$\psi(y;q,a) - \frac{y}{\phi(q)} = \frac{1}{\phi(q)} \sum_{\chi \in X_q} \overline{\chi}(a)\psi'(y,\chi),$$

and hence, for all $y, q \geq 1$ and all a with $(a,q) = 1$:

$$\big|E(y;q,a)\big| \leq \frac{1}{\phi(q)} \sum_{\chi \in X_q} \big|\psi'(y,\chi)\big|.$$

Since the right hand side is independent of a we conclude that also

$$E(y;q) \leq \frac{1}{\phi(q)} \sum_{\chi \in X_q} \big|\psi'(y,\chi)\big|.$$

For any $\chi \in X_q$ we let χ_1 denote the corresponding primitive character. Then $\psi'(y,\chi)$ and $\psi'(y,\chi_1)$ are nearly equal, since

$$\begin{aligned} \psi'(y,\chi_1) - \psi'(y,\chi) &= \sum_{p|q} \sum_{1 \leq k \leq \log_p y} \chi_1(p^k) \log p = O\Big(\sum_{p|q} \Big\lfloor \frac{\log y}{\log p} \Big\rfloor \log p\Big) \\ &= O\Big((\log y) \sum_{p|q} \log p\Big) = O\Big((\log y)(\log q)\Big) = O\Big((\log qy)^2\Big). \end{aligned}$$

Hence

$$E(y;q) \ll (\log qy)^2 + \frac{1}{\phi(q)} \sum_{\chi \in X_q} \big|\psi'(y,\chi_1)\big|$$

and thus for all $x, q \geq 1$,

$$E^*(x;q) \ll (\log qx)^2 + \frac{1}{\phi(q)} \sum_{\chi \in X_q} \max_{y \leq x} \big|\psi'(y,\chi_1)\big|.$$

Hence we get the following bound on the left hand side of (641), for any $x, Q \geq 1$:

$$\begin{aligned} \sum_{1 \leq q \leq Q} E^*(x,q) &\ll \sum_{1 \leq q \leq Q} \Big((\log qx)^2 + \frac{1}{\phi(q)} \sum_{\chi \in X_q} \max_{y \leq x} \big|\psi'(y,\chi_1)\big|\Big) \\ &\leq Q(\log Qx)^2 + \sum_{1 \leq q \leq Q} \sum_{\chi \in X_q} \frac{1}{\phi(q)} \max_{y \leq x} \big|\psi'(y,\chi_1)\big|. \end{aligned}$$

Here in the last double sum, χ_1 visits every primitive character which has modulus $q_1 \leq Q$, and no other primitive characters. Moreover, any fixed primitive character χ_1 modulo $q_1 \leq Q$ appears in the above double sum exactly for those χ's which are induced by χ_1, i.e.

for exactly one $\chi \in X_q$ for each q_1-multiple $q \in \{q_1, 2q_1, 3q_1, 4q_1, \ldots\}$ with $q \le Q$, and for no other χ's. Hence we get, writing $q = kq_1$:

$$\sum_{1\le q\le Q} E^*(x,q) \ll Q(\log Qx)^2 + \sum_{q_1\le Q}\sum_{\chi_1\in X^*_{q_1}} \max_{y\le x}|\psi'(y,\chi_1)|\Big(\sum_{1\le k\le Q/q_1}\frac{1}{\phi(kq_1)}\Big). \tag{665}$$

Here we note the general inequality

$$\phi(kq_1) = kq_1\prod_{p|kq_1}\Big(1-\frac{1}{p}\Big) \ge kq_1\prod_{p|k}\Big(1-\frac{1}{p}\Big)\prod_{p|q_1}\Big(1-\frac{1}{p}\Big) = \phi(k)\phi(q_1),$$

which implies (for any $q_1 \in \mathbb{Z}^+$ and $z \in \mathbb{R}$, $z \ge 1$)

$$\sum_{1\le k\le z}\frac{1}{\phi(kq_1)} \le \frac{1}{\phi(q_1)}\sum_{1\le k\le z}\frac{1}{\phi(k)}.$$

Next, to bound $\sum_{1\le k\le z}\frac{1}{\phi(k)}$ we use the fact that $\phi(k)$ is multiplicative (and positive) to see that

$$\begin{aligned}\sum_{1\le k\le z}\frac{1}{\phi(k)} &\le \prod_{p\le z}\Big(1+\frac{1}{\phi(p)}+\frac{1}{\phi(p^2)}+\frac{1}{\phi(p^3)}+\ldots\Big)\\ &= \prod_{p\le z}\Big(1+\frac{1}{p-1}+\frac{1}{p(p-1)}+\frac{1}{p^2(p-1)}+\ldots\Big) = \prod_{p\le z}\Big(1+\frac{1}{(p-1)}\sum_{r=0}^{\infty}p^{-r}\Big)\\ &= \prod_{p\le z}\Big(1+\frac{1}{(p-1)(1-p^{-1})}\Big) = \prod_{p\le z}\Big(1+\frac{p}{(p-1)^2}\Big),\end{aligned}$$

and thus, taking the logarithm:

$$\begin{aligned}\log\sum_{1\le k\le z}\frac{1}{\phi(k)} &\le \sum_{p\le z}\log\Big(1+\frac{p}{(p-1)^2}\Big) \le \sum_{p\le z}\frac{p}{(p-1)^2} = \sum_{p\le z}\Big(\frac{1}{p}+\frac{2p-1}{p(p-1)^2}\Big)\\ &= \sum_{p\le z}\Big(p^{-1}+O\big(p^{-2}\big)\Big) = \Big(\sum_{p\le z}p^{-1}\Big)+O(1) = \log\log(z+1)+O(1),\end{aligned}$$

by Proposition 6.5. Exponentiating back we get

$$\sum_{1\le k\le z}\frac{1}{\phi(k)} \ll \log(z+1).$$

Hence, if we from now on assume $Q \le x$, we get that for all $q_1 \le Q$,

$$\sum_{1\le k\le Q/q_1}\frac{1}{\phi(kq_1)} \le \frac{1}{\phi(q_1)}\sum_{1\le k\le Q/q_1}\frac{1}{\phi(k)} \ll \frac{1}{\phi(q_1)}\log\Big(\frac{Q}{q_1}+1\Big) \ll \frac{\log x}{\phi(q_1)}.$$

Using this in (665), we conclude, writing now q, χ in place of q_1, χ_1,

$$\sum_{1\le q\le Q} E^*(x,q) \ll Q(\log Qx)^2 + (\log x)\sum_{q\le Q}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi'(y,\chi)\big|.$$

Hence we see that in order to prove Theorem 21.1 it suffices to prove that for all $x \ge 2$ and all Q with $x^{\frac12}(\log x)^{-A} \le Q \le x^{\frac12}$ we have

$$\sum_{q\le Q}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi'(y,\chi)\big| \ll x^{\frac12}Q(\log x)^4. \tag{666}$$

We now bring the bound (643) into use. First note that for any $U \ge 1$, by taking $Q = 2U$ in (643); then sacrificing all terms with $q \le U$ in the left hand side and dividing by U (using $q/U > 1$ for $q > U$), we get

$$\sum_{U<q\le 2U}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi(y,\chi)\big| \ll \Big(\frac{x}{U} + x^{\frac56} + x^{\frac12}U\Big)(\log Ux)^4. \tag{667}$$

Hence, for any given Q_1 with $1 \le Q_1 \le Q$, by summing the above over $U = 2^k$ where k runs through all integers with $\frac12 Q_1 < 2^k \le 2Q$, we get (also using the fact that $\psi'(y,\chi) = \psi(y,\chi)$ for every nontrivial primitive character χ)

$$\begin{aligned}\sum_{Q_1<q\le Q}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi'(y,\chi)\big| &\le \sum_{\substack{k\\ \frac12 Q_1<2^k\le 2Q}}\ \sum_{2^k<q\le 2^{k+1}}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi(y,\chi)\big|\\ &\ll \sum_{\log_2(\frac12 Q_1)<k\le\log_2(2Q)}\Big(\frac{x}{2^k} + x^{\frac56} + x^{\frac12}2^k\Big)(\log 2Qx)^4\\ &\ll \Big(\frac{x}{Q_1} + x^{\frac56}\log Q + x^{\frac12}Q\Big)(\log Qx)^4. \end{aligned}\tag{668}$$

Let us now assume $x \ge 3$ (without loss of generality) and take $Q_1 = (\log x)^A$. Using $x^{\frac12}(\log x)^{-A} \le Q \le x^{\frac12}$ we then see that the bound in (668) is $\ll x^{\frac12}Q(\log x)^4$, and thus we have

$$\sum_{(\log x)^A<q\le Q}\frac{1}{\phi(q)}\sum_{\chi\in X_q^*}\max_{y\le x}\big|\psi'(y,\chi)\big| \ll x^{\frac12}Q(\log x)^4. \tag{669}$$

Finally to treat q with $q \le (\log x)^A$, note that by Theorem 16.5 applied with $N = A+1$, for all y with $\exp\big(q^{\frac{1}{A+1}}\big) \le y \le x$ and any nonprincipal Dirichlet character $\chi \in X_q$ we have

$$\psi(y,\chi) = O\big(ye^{-c\sqrt{\log y}}\big) = O\big(xe^{-c\sqrt{\log x}}\big). \tag{670}$$

But for y with $1 \le y < \exp\big(q^{\frac{1}{A+1}}\big)$ we have the even stronger bound

$$\big|\psi(y,\chi)\big| \le \psi(y) \ll y < \exp\big(q^{\frac{1}{A+1}}\big) \le \exp\big((\log x)^{\frac{A}{A+1}}\big).$$

Hence in fact

$$\max_{y\le x}\big|\psi(y,\chi)\big| \ll xe^{-c\sqrt{\log x}}.$$

Adding this over all q with $2 \le q \le (\log x)^A$ we get, since $\psi(y,\chi) = \psi'(y,\chi)$ when χ is nontrivial and primitive,

$$\sum_{2\le q\le(\log x)^A} \frac{1}{\phi(q)} \sum_{\chi\in X_q^*} \max_{y\le x}\big|\psi'(y,\chi)\big| \ll (\log x)^A \cdot xe^{-c\sqrt{\log x}} \ll x(\log x)^{-A} \ll x^{\frac12}Q. \tag{671}$$

Finally when χ is the trivial character we have, for all y with $1 \le y \le x$,

$$\big|\psi'(y,\chi)\big| = \big|\psi(y) - y\big| \ll ye^{-c\sqrt{\log y}} \ll xe^{-c\sqrt{\log x}} \ll x^{\frac12}Q.$$

Adding this together with (669) and (671) we obtain (666). Hence Theorem 21.1 is proved. □

References

1. L. V. Ahlfors, *Complex analysis*, McGraw-Hill, 1966.
2. Eric Bach, *Explicit bounds for primality testing and related problems*, Math. Comp. **55** (1990), no. 191, 355–380.
3. A. Baker, *Linear forms in the logarithms of algebraic numbers*, Mathematika **13** (1966), 204–216, ref from Montgomery and Vaughans book, p.394.
4. R. P. Boas, Jr., *A general moment problem*, Amer. J. Math. **63** (1941), 361–370.
5. E. Bombieri, *On the large sieve*, Mathematika **12** (1965), 201–225.
6. Enrico Bombieri, *Le grand crible dans la théorie analytique des nombres*, Astérisque (1987), no. 18, 103.
7. Johannes Buchmann and Ulrich Vollmer, *Binary quadratic forms*, Algorithms and Computation in Mathematics, vol. 20, Springer, Berlin, 2007, An algorithmic approach.
8. D. A. Burgess, *On character sums and L-series. II*, Proc. London Math. Soc. (3) **13** (1963), 524–536.
9. H. Cohn, *Advanced number theory*, Dover, 1962.
10. Alina Carmen Cojocaru and M. Ram Murty, *An introduction to sieve methods and their applications*, London Mathematical Society Student Texts, vol. 66, Cambridge University Press, Cambridge, 2006.
11. Brian Conrey, *Problems and progress in l-functions and rmt*, talk at Bordeaux, 2006.
12. J. B. Conrey, *More than two fifths of the zeros of the Riemann zeta function are on the critical line*, J. Reine Angew. Math. **399** (1989), 1–26.
13. John H. Conway, *The sensual (quadratic) form*, Carus Mathematical Monographs, vol. 26, Mathematical Association of America, Washington, DC, 1997, With the assistance of Francis Y. C. Fung.
14. H. Davenport and H. Halberstam, *The values of a trigonometrical polynomial at well spaced points*, Mathematika **13** (1966), 91–96.
15. Harold Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Revised and with a preface by Hugh L. Montgomery.
16. Max Deuring, *Imaginäre quadratische Zahlkörper mit der Klassenzahl 1*, Math. Z. **37** (1933), no. 1, 405–415.
17. ———, *Imaginäre quadratische Zahlkörper mit der Klassenzahl Eins*, Invent. Math. **5** (1968), 169–179.
18. H. M. Edwards, *Riemann's zeta function*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974, Pure and Applied Mathematics, Vol. 58.
19. David Farmer, *Number theory background, i*, Tech. report, 2006, talk.
20. John B. Fraleigh, *First course in abstract algebra, a*, Addison Wesley; 7 edition (November 16, 2002), 2002.
21. P. X. Gallagher, *The large sieve*, Mathematika **14** (1967), 14–20.
22. Dorian M. Goldfeld, *A simple proof of Siegel's theorem*, Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 1055.
23. ———, *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **3** (1976), no. 4, 624–663.
24. S. M. Gonek, *Moments ii*, talk, 2006.
25. Timothy Gowers, *The princeton companion to mathematics*, Princeton University Press, 2008, (Gowers is EDITOR).
26. Benedict H. Gross and Don B. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.
27. H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974, London Mathematical Society Monographs, No. 4.
28. G. H. Hardy, *Sur les zéros de la fonction $\zeta(s)$ de riemann*, Comptes Rendus Mathematique **158** (1914), 1012–1014, ref from Titchmarsh' book.
29. G. H. Hardy and J. E. Littlewood, *Some problems of partitio numerorum; iii: On the expression of a number as a sum of primes*, Acta Mathematica **44** (1922), 1–70.

30. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Clarendon Press, Oxford, 1938.
31. Kurt Heegner, *Diophantische Analysis und Modulfunktionen*, Math. Z. **56** (1952), 227–253.
32. Hans Heilbronn, *On the class-number in imaginary quadratic fields*, Quarterly J. of Math. **5** (1934), 150–160.
33. Christopher Hooley, *On the Barban-Davenport-Halberstam theorem. I*, J. Reine Angew. Math. **274/275** (1975), 206–223, Collection of articles dedicated to Helmut Hasse on his seventy-fifth birthday, III.
34. A. E. Ingham, *The distribution of prime numbers*, Cambridge Mathematical Library, 1932.
35. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer-Verlag, 1990.
36. Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004.
37. Y. Katznelson, *An introduction to harmonic analysis*, Dover, 1976.
38. E. Kowalski, *The large sieve and its applications*, Cambridge University Press, Cambridge, 2008.
39. E. Landau, *Vorlesungen Über Zahlentheorie i, ii, iii*, S. Hirzel, 1927.
40. S. Lang, *Algebraic Number Theory*, Springer-Verlag, 1986.
41. Norman Levinson, *More than one third of zeros of Riemann's zeta-function are on $\sigma = 1/2$*, Advances in Math. **13** (1974), 383–436.
42. U. V. Linnik, *"The large sieve."*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **30** (1941), 292–294.
43. Ming-Chit Liu and Tianze Wang, *On linear ternary equations with prime variables—Baker's constant and Vinogradov's bound*, A panorama of number theory or the view from Baker's garden (Zürich, 1999), Cambridge Univ. Press, Cambridge, 2002, pp. 311–324.
44. ———, *On the Vinogradov bound in the three primes Goldbach conjecture*, Acta Arith. **105** (2002), no. 2, 133–175.
45. H. L. Montgomery, *Primes in arithmetic progressions*, Michigan Math. J. **17** (1970), 33–39.
46. H. L. Montgomery and R. C. Vaughan, *Exponential sums with multiplicative coefficients*, Invent. Math. **43** (1977), no. 1, 69–82.
47. Hugh L. Montgomery, *The analytic principle of the large sieve*, Bull. Amer. Math. Soc. **84** (1978), no. 4, 547–567.
48. Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, Cambridge, 2007.
49. M. Ram Murty, *Problems in analytic number theory*, second ed., Graduate Texts in Mathematics, vol. 206, Springer, New York, 2008, Readings in Mathematics.
50. T. Nagell, *Introduction to number theory*, Wiley & Sons, 1951.
51. Melvyn B. Nathanson, *Elementary methods in number theory*, Graduate Texts in Mathematics, vol. 195, Springer-Verlag, New York, 2000.
52. J. Neukirch, *Algebraic number theory (oevers. av tyska boken)*, Springer-Verlag, 1999.
53. Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An introduction to the theory of numbers*, fifth ed., John Wiley & Sons Inc., New York, 1991.
54. J. Oesterlé, *Le problème de Gauss sur le nombre de classes*, Enseign. Math. (2) **34** (1988), no. 1-2, 43–67.
55. Joseph Oesterlé, *Nombres de classes des corps quadratiques imaginaires*, Astérisque (1985), no. 121-122, 309–323, Seminar Bourbaki, Vol. 1983/84.
56. H. A. Priestley, *Introduction to complex analysis*, second ed., Oxford University Press, Oxford, 2003.
57. H.-E. Richert, *Zur Abschätzung der Riemannschen Zetafunktion in der Nähe der Vertikalen $\sigma = 1$*, Math. Ann. **169** (1967), 97–101.
58. K. F. Roth, *On the large sieves of Linnik and Rényi*, Mathematika **12** (1965), 1–9.
59. W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, 1976.
60. ———, *Real and complex analysis*, McGraw-Hill, 1987.

61. ———, *Fourier analysis on groups*, John Wiley & Sons, 1990.
62. Atle Selberg, *On the zeros of Riemann's zeta-function*, Skr. Norske Vid. Akad. Oslo I. **1942** (1942), no. 10, 59, according to Titchmarsh's book, THIS is where he proves positive proportion of zeros lie on the line!
63. J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.
64. Jean-Pierre Serre, *Linear representations of finite groups*, Springer-Verlag, New York, 1977, Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.
65. Carl Ludwig Siegel, *über die classenzahl quadratischer zahlkörper*, Acta Arithmetica **1** (1935), 83–86.
66. H. M. Stark, *A complete determination of the complex quadratic fields of class-number one*, Michigan Math. J. **14** (1967), 1–27.
67. ———, *On the "gap" in a theorem of Heegner*, J. Number Theory **1** (1969), 16–27.
68. E. C. Titchmarsh, *A divisor problem*, Rend. Circ. Mat. Palermo **54** (1930), 414–429.
69. E. C. Titchmarsh, *The theory of functions (2nd ed.)*, Oxford University Press, 1939.
70. ———, *The theory of the riemann zeta-function (2nd ed.)*, Oxford University Press, 1986.
71. Jeffrey D. Vaaler, *Some extremal functions in Fourier analysis*, Bull. Amer. Math. Soc. (N.S.) **12** (1985), no. 2, 183–216.
72. Robert-C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), no. 16, A981–A983.
73. I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, Comptes Rendues (Doklady) de l'Academy des Sciences de l'USSR **15** (1937), 191–294, Ref avskriven från jDgEhRdZ97.
74. ———, *The method of trigonometrical sums in the theory of numbers*, Dover Publications Inc., Mineola, NY, 2004, Translated from the Russian, revised and annotated by K. F. Roth and Anne Davenport, Reprint of the 1954 translation.