Analytic Number Theory, undergrad, Courant Institute, Spring 2017
http://www.math.nyu.edu/faculty/goodman/teaching/NumberTheory2017/index.html

Section 3, Dirichlet's theorem

version 1.1, February 16, 2017

1 Introduction.

Dirichlet's theorem was mentioned already. An *arithmetic progression* is a sequence of the form $x_k = a + kn$. The theorem is that if a and n are relatively prime, then infinitely many of the numbers x_k are prime numbers. The sequence with $a = 5$ and $n = 12$ seems to have lots of primes:

$$\underline{5},\ \underline{17},\ \underline{29},\ \underline{41},\ \underline{53},\ 65,\ 77,\ \underline{89},\ \underline{101},\ \underline{113},\ 125, \underline{137}, \underline{149}, 161, \underline{173}, 185,\ \cdots\ .$$

If a and n are not relatively prime, then there are not infinitely many primes.

Dirichlet's theorem is important because if its intrinsic interest, and because of the ideas that go into its proof. Here is a summary, using terminology and notation that will be explained in this section. Some algebra allows us to engineer new Euler product identities that relate simple Dirichlet series to sums involving only primes equal to a mod n. The algebra has two parts. The *Discrete Fourier transform* is a useful basis of the vector space $\mathbb{C}^n$. These basis vectors may be thought of as *characters* of the *group* $G = C_{n_1} \times \cdots \times C_{n_m}$. Here, C_n is the *cyclic group* of order n, and $C_{n_1} \times \cdots$ is a *cartesian product* of cyclic groups. The other piece of algebra involves the *multiplicative* group, G_n, of integers mod n that are relatively prime to n. *Dirichlet* characters, mod n, are the characters of G_n. A function of an integer x is periodic with period n if

$$\chi(x + n) = \chi(x) \tag{1}$$

for all x. A function is *multiplicative*[1] if

$$\chi(xy) = \chi(x)\chi(y)\ \ , \quad \text{for any integers } x \text{ and } y. \tag{2}$$

Dirichlet characters are periodic and multiplicative. We construct them by showing that G_n is *isomorphic* to a cartesian product of cyclic groups.

The Euler product formulas in this section apply to Dirichlet series called L *functions*. There is a Dirichlet L function for any character:

$$L_\chi(s) = \sum_1^\infty \chi(x)x^{-s}\ . \tag{3}$$

[1] A function with this property is often called *completely multiplicative*. A function is multiplicative if $\chi(xy) = \chi(x)\chi(y)$ when x and y are relatively prime. A Dirichlet series with multiplicative coefficients has an Euler product representation even if the coefficients are not completely multiplicative. The Dirichlet characters happen to be completely multiplicative.

The Euler product representation is

$$L_\chi(s) = \prod_p \left(1 - \chi(p)p^{-s}\right)^{-1} \ . \tag{4}$$

The formal derivation, and the proof, are the same as for the Riemann zeta function, if you use the character property (2). Taking the log of both sides of (4) turns the Euler product into an infinite sum over just primes[2] (using $\log(1-z) = -z + O(|z|^2)$, which is true for complex z with $|z| < \frac{1}{2}$)

$$\log(L_\chi(s)) = -\sum_p \log\bigl(1 - \chi(p)p^{-s}\bigr) = \sum_p \chi(p)p^{-s} + O\bigl(p^{-2s}\bigr) \ .$$

As with the zeta function, we will study the behavior of $L_\chi(s)$ for $s > 1$ in the limit $s \downarrow 1$. On the other hand, if we substitute the series for $\log(L)$ and the definition of g_a, we get

$$g_a(s) = \sum_p \left(\sum_\chi c_\chi \chi(p) \right) p^{-s} + \sum_p O\bigl(p^{-2s}\bigr) \ .$$

We will show in Sections 2 and 5 that if a is relatively prime to n it is possible to choose coefficients c_χ (depending on a) so that for any integer k

$$\sum_\chi c_\chi \chi(x) = \begin{cases} 1 & \text{if } x = a \bmod n \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This implies that

$$g_a(s) = \sum_{(p,n)=1} p^{-s} + \sum_p O\bigl(p^{-2s}\bigr) \ .$$

We will also see that $c_0 \neq 0$, which implies that $g_a(s) \to \infty$ as $s \downarrow 1$. If there were only finitely many primes p equal to $a \bmod n$, then the right side would be bounded as $s \downarrow 1$. This is the contradiction that proves that there infinitely many primes $p \equiv a \bmod n$.

Review of complex numbers

Here are some basic facts about complex numbers. If course, i is the imaginary unit, which satisfies $i^2 = -1$. If $z = x + iy$ is a complex number, then $\overline{z} = x - iy$ is the *complex conjugate*. Important properties are $\overline{z}z = x^2 + y^2 = |z|^2$ and $\overline{e^{i\theta}} = \overline{\cos(\theta) + i\sin(\theta)} = \cos(\theta) - i\sin(\theta) = e^{-i\theta}$. This implies the famous formula $e^{i\pi} = -1$, which leads to something that is used many times below: $e^{2\pi ij} = 1$ if j is an integer. Also, $\left|e^{i\theta}\right| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = 1$.

[2] We will later switch to the derivative of $\log(L(s))$, which is $L'(s)/L(s)$, but the log is is enough to make the point here.

2 Linear Algebra and the DFT

This and the next few sections are about constructing characters that are multiplicative and may be used to represent general functions as in (5). This section constructs functions like χ but which are periodic and additive rather than periodic and multiplicative:

$$w(a+n) = w(a) \ , \quad w(a+b) = w(a)w(b) \ . \tag{6}$$

Section 4 then explains how to turn multiplication mod n into addition mod n. The log/exponential pair doesn't apply because the log of an integer is not an integer, and a periodic function of $\log(x)$ is not a periodic function of x.

The functions $w(a)$ are *discrete Fourier modes*. A general periodic function with period n may be expressed as a sum of Fourier modes. This representation is called the *discrete Fourier transform*, or *DFT*. *Harmonic analysis* is in large part about the many uses of Fourier representations in mathematics. In this course we will use Fourier series and the Fourier integral in addition to the DFT. The DFT, used here in the proof of Dirichlet's theorem on primes, is the first taste of harmonic analysis in number theory.

A *cycle* of length n, also called the *cyclic group*[3] of order n, is written C_n. It may be thought of as the first n non-negative inteters:

$$C_n = \{0, 1, \ldots, n-1\} \ .$$

It also may be described as the set of *equivalence classes* of the integers mod n, as we will soon do. A complex valued function on C_n, which we write $f(a)$ for $a \in C_n$, is determined by the n complex numbers $f(a)$, for $a \in C_n$. The same function may be thought of as a periodic function of the integer variable a. If $f(a+n) = f(a)$ for all a, then f is determined by its values on $a = 0, 1, \ldots, n-1$. For example, the values $f(n)$ and $f(-2n)$ are the same as $f(0)$. It is helpful, sometimes, to think of f as a column vector in $\mathbb{C}^n$. The operation of addition in C_n is defined as ordinary addition mod n. If $a \in C_n$ and $b \in C_n$, then $c = a + b \in C_n$ is the integer in C_n that is equal to $a + b$ mod n. We will describe addition and modular arithmetic in more detail soon.

The *discrete Fourier modes* are the functions $w_j(a)$ defined by

$$w_j(a) = e^{2\pi i j a/n} \ . \tag{7}$$

As long as j is an integer, w_a satisfies the two properties (6). To verify that w is periodic if j is an integer:

$$w_j(a+n) = e^{2\pi i j(a+n)/n} = e^{2\pi i j a/n} e^{2\pi i j} = e^{2\pi i j a/n} = w_j(a) \ .$$

To verify that w is additive/multiplicative,

$$w_j(a+b) = e^{2\pi i j(a+b)/n} = e^{2\pi i j a/n} e^{2\pi i j b/n} = w_j(a) w_j(b) \ .$$

[3]The term *group* is defined in Section 4

The function $w_j(a)$ is also a periodic function of the mode number, j:

$$w_{j+n}(a) = w_j(a) \ .$$

Therefore, we have a "complete set" of Fourier modes if we take the n functions $w_0, \ldots, w_{n-1}$. These discrete Fourier modes form a basis for $\mathbb{C}^n$ (two proofs below). This means that if f is any periodic function, then there are coefficients c_j so that

$$f(a) = \sum_0^{n-1} c_j w_j(a) \ . \tag{8}$$

Both proofs are based on the fact that the values $w_j(a)$ are powers of

$$z = e^{2\pi i/n} \ , \ \ w_j(a) = z^{ja} = \left(z^j\right)^a = (z^a)^j \ .$$

One of the proofs uses linear algebra and orthogonality (brief review included). The inner products can be calculated because the sums involved are geometric series with powers of z. The other proof (which comes from the book of Apostol) uses the fact that the right side of (8) involves a polynomial in z:

$$p(t) = \sum_0^{n-1} c_j t^j \ .$$

The representation formula (8) may be written

$$f(a) = p(z^a) \ .$$

Polynomial interpolation is the problem of finding a polynomial p of degree $n-1$ that takes given values at n distinct given points. These two proofs lead to two different paths for generalizing the DFT.

Linear algebra, proofs ommitted

The main object in linear algebra is a vector space *over* a *field* [4] of *scalars*. For now, the field of scalars will be the complex numbers $\mathbb{C}$. A vector space, $\mathbb{V}$, is a collection of *vectors*. To be a vector space, it must be possible to add vectors. If $f \in \mathbb{V}$ and $g \in \mathbb{V}$, then $f + g \in \mathbb{V}$. We use f and g to denote vectors to remind ourselves that our vectors often are functions. Vector addition must be commutative and associative: $f+g = g+f$ and $(f+g)+h = f+(g+h)$. There is a zero vector $0 \in \mathbb{V}$ so that $0 + f = f$ for all $f \in \mathbb{V}$. There is an operation of

[4] *Field* is a concept in abstract algebra. Common examples are $\mathbb{R}$ and $\mathbb{C}$. Other examples are important in number theory, as we will see. If x and y are scalars, the sum $x + y$ and product xy are defined. Addition and multiplication are commutative and associative. The distributive property holds: $x(y + z) = (xy) + (xz)$. There is unique multiplicative unit, written 1, and a unique additive unit, written 0. Every scalar x has a unique additive inverse, written $-x$. Every scalar x has a unique multiplicative inverse, written x^{-1}. Because of the distributive and associative properties, $-x = (-1)x$. That means that you get the additive inverse of x by multiplying the additive inverse of the multiplicative unit by x.

scalar multiplication. If $x \in \mathbb{C}$ and $f \in \mathbb{V}$, then $xf \in \mathbb{V}$. Scalar multiplication is distributive and associative in every possible way: $(xy)f = x(yf)$, $(x+y)f = xf + yf$, $x(f+g) = xf + gf$.

A *basis* for $\mathbb{V}$ is a set of vectors $g_1, \ldots, g_n$ so that for any $f \in \mathbb{V}$ there is a unique set of scalars (complex numbers in the examples here) $c_1, \ldots, c_n$ so that

$$f = \sum_1^n c_j g_j \ .$$

The sum on the right is a *linear combination* of the vectors g_j. A set of vectors $h_1, \ldots, h_m$ is *linearly independent* if

$$\sum_1^m c_j h_j = 0 \implies c_j = 0 \text{ for all } j.$$

If the h_j are linearly independent then $m \leq n$. The vectors in a basis are linearly independent. If $m = n$ and the h_j are linearly independent, then the h_j form a basis. As a consequence, if $\mathbb{V}$ has a basis with n vectors, then any other basis also has n vectors. This n is the *dimension* of $\mathbb{V}$. If $m > n$ and any $f \in \mathbb{V}$ is a linear combination of vectors h_j, then the coefficients c_j are not unique and the h_j are *overcomplete*. A vector space that does not have a finite basis[5] is called *infinite dimensional*. There are infinite dimensional vector spaces later in this course.

The vector space $\mathbb{C}^n$ is the set of n component column vectors:

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} \ .$$

We write f_a for the $a-$th component of f. If we think of f as a periodic function of a with period n, then we write $f(a)$ instead of f_a. If we have a collection of functions, we may call the $j-$th function f_j, which has values $f_j(a)$. This kind of conflict of notation is confusing for beginners and old-times alike. There seems to be no sensible way to avoid it.

The vector space $\mathbb{V} = \mathbb{C}^n$ has a special structure called a *Hermitian* structure, or an *inner product* (they mean the same thing). The inner product of vectors f and g is

$$\langle f, g \rangle = \sum_{a=1}^n \overline{f}_a g_a \ .$$

[5]It is redundant to say *finite* here, because the definition calls for n vectors. The word is included for emphasis, and because there are other definitions of basis in which infinite bases are allowed.

This has the properties ($c \in \mathbb{C}$ is any complex number, $\overline{c}$ is its complex conjugate.)

$$\langle f, g\rangle = \overline{\langle g, f\rangle} \qquad \text{(not quite commutative)}$$
$$\langle f, g + ch\rangle = \langle f, g\rangle + c\langle f, h\rangle \qquad \text{(linear in the second vector)}$$
$$\langle f + ch, g\rangle = \langle f, g\rangle + \overline{c}\langle h, g\rangle \qquad \text{(anti-linear in the second vector)}$$
$$\langle f, f\rangle = \sum_1^n |f_a|^2 = \|f\|^2 > 0 \text{ if } f \neq 0 \qquad \text{(norm from inner product)}$$
$$\text{if } \langle f, g\rangle = 0 \text{ , then } \quad \|f + g\|^2 = \|f\|^2 + \|g\|^2 \qquad \text{(pythagorean theorem)}$$

The following formula is a consequence of the properties above: If u_j are orthogonal vectors, $\langle u_j, u_k\rangle = 0$ for $j \neq k$, and f is expressed in terms of them

$$f = \sum c_j u_j \ , \tag{9}$$

then the "size" of f is determined by the coefficients

$$\|f\|^2 = \sum |c_j|^2 \|u_j\| \ . \tag{10}$$

More generally, if g is another vector/function expanded in the u vectors

$$g = \sum d_j u_j \ ,$$

then

$$\langle f, g\rangle = \sum \overline{c}_j d_j \|u_j\|^2 \ .$$

The coefficients in the representation formula, also called the *expansion* formula, (9) are given by

$$c_j = \frac{1}{\|u_j\|^2}\langle u_j f\rangle \ . \tag{11}$$

The discrete Fourier modes (7), for $j = 0, 1, \ldots, n-1$, are orthogonal to each other.

$$\langle w_j, w_j\rangle = \|w_j\|^2 = n \ , \tag{12}$$
$$\langle w_j, w_k\rangle = 0 \ , \qquad \text{if } j \neq k \tag{13}$$

The first formula depends on the fact that $\left|e^{i\theta}\right| = 1$ for real θ:

$$\langle w_j, w_j\rangle = \sum_1^n \left|e^{2\pi i j a/n}\right| = \sum_1^n 1 = n \ .$$

The orthogonality relation uses geometric series, as promised. We define $z = e^{2\pi i(k-j)/n}$ in the calculation.

$$\begin{aligned}
\langle w_j, w_k \rangle &= \sum_{a=1}^{n} \overline{w_j(a)}\, w_k(a) \\
&= \sum_{1}^{n} e^{-2\pi ija/n} e^{2\pi ika/n} \\
&= \sum_{1}^{n} e^{2\pi i(k-j)a/n} \\
&= \sum_{1}^{n} z^a \\
&= \frac{z^{n+1} - z}{z-1} \\
&= z\frac{z^n - 1}{z-1} \; .
\end{aligned}$$

If j and k are in the range $\{0, 1, \ldots, n-1\}$, then $z^n = 1$ and $z \neq 1$ (if $j \neq k$). These n Fourier modes therefore form a basis for $\mathbb{C}^n$.

For future reference, here are the DFT formulas written specifically for the discrete Fourier modes. The formulas were not derived in this order, but it is convenient to list them this way. *Very iomportant*: there have to be factors $\frac{1}{n}$ somewhere in the DFT formulas. Different people put them in different places. You always have to check the DFT formulas of the specific thing you're reading to see where your author puts them. The *discrete Fourier transform* is the general coefficient formula (11) specialized for discrete Fourier modes

$$c_j = \frac{1}{n}\langle w_j, f \rangle = \frac{1}{n} \sum_{1}^{n} e^{-2\pi ija/n} f(a) \; . \tag{14}$$

The discrete *Fourier inversion formula* is the general expansion formula (9) specialized to discrete Fourier modes:

$$f(a) = \sum_{j=0}^{n-1} c_j e^{2\pi ija/n} \; . \tag{15}$$

The general version of the Pythagorean theorem (10) specializes to the discrete *Plancharel theorem*

$$\sum_{a=1}^{n} |f(a)|^2 = n \sum_{j=0}^{n-1} |c_j|^2 \; . \tag{16}$$

The sums above run over n distinct j or a values. It is natural to make a run from 1 to n if you think of $f(a)$ as component a of a vector in $\mathbb{C}^n$. But if you think of a as a possible value of an integer mod n, it might be more natural

to use the range from 0 to $n-1$. This doesn't matter, if we think of $f(a)$ as a periodic function of a. The formula (14) makes c_j a periodic function of j, in the sense that just applying the formula gives $c_{j+n} = c_j$. Therefore we can use 0 to $n-1$ or 1 to n for j also. It will be most convenient for us to take both a and j in the range 0 to $n-1$ from now on.

At this point I should put in a few nice examples where you can calculate the DFT sums explicitly. Unfortunately, I don't know any elementary examples. The DFT is a powerful theoretical tool, but not so much for paper and pencil calculation. This may be because there are not so many simple and natural periodic functions $f(a+n) = f(a)$? Time permitting, this class may be able to do some of the amazing calculations when you take the DFT of Dirichlet characters. These are called *Ramanujan sums* (for the principal character) and *Gauss sums* otherwise.[6]

To construct general characters, we need the DFT for periodic functions of more than one integer variable. For this, we develop discrete Fourier modes and Fourier representations for functions of more than integer variable. Suppose $f(a_1, a_2)$ is periodic with period n_1 in a_1 and periodic with period n_2 in a_2. This means that

$$f(a_1 + jn_1, a_2 + kn_2) = f(a_1, a_2) \ , \ \text{ if } j \text{ and } k \text{ are integers.} \tag{17}$$

An example of a function like this is

$$w_j(a_1, a_2) = e^{2\pi i(j_1 a_1/n_1 + j_2 a_2/n_2)} \ . \tag{18}$$

These functions, like the one variable, are additive/multiplicative in the sense that

$$w_j(a_1, a_2) w_j(b_1, b_2) = w_j(a_1 + b_1, a_2 + b_2) \ .$$

The addition $a_1 + b_1$ can be done mod n_1, and similarly for $a_2 + b_2$, because w_j is periodic with periods n_1 and n_2.

If f is doubly periodic in the sense that it satisfies (17) then f is determined by the values $f(a_1, a_2)$ for a_1 in the range $\{1, 2, \ldots, n_1\}$ and $a_2 \in \{1, 2, \ldots, n_2\}$. There are $n = n_1 n_2$ such pairs (a_1, a_2), so the space of all doubly periodic functions forms a vector space of dimension n. In this context, this is sometimes written $\mathbb{C}^{n_1 \times n_2}$. The number of distinct discrete Fourier modes (18) is also $n_1 n_2$. We may take, for example $j = (j_1, j_2)$, where $j_1 \in \{0, 1, \ldots, n_1 - 1\}$ and $j_2 \in \{0, 1, \ldots, n_2 - 1\}$. The DFT formula analogous to (14) is

$$c_j = \frac{1}{n_1 n_2} \sum_{a_1=0}^{n_1-1} \sum_{a_2=0}^{n_2-1} e^{-2\pi i(j_1 a_1/n_1 + j_2 a_2/n_2)} f(a_1, a_2) \ . \tag{19}$$

The discrete Fourier inversion formula for two variable functions is

$$f(a_1, a_2) = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} c_j e^{2\pi i(j_1 a_1/n_1 + j_2 a_2/n_2)} \tag{20}$$

[6] It's fair to ask how Ramanujan, who comes 100 years after Gauss gets his name on a subset of the calculations Gauss did. There is a reason.

The Plancharel identity is

$$\sum_{a_1=0}^{n_1-1}\sum_{a_2=0}^{n_2-1} |f(a_1,a_2)|^2 = n_1 n_2 \sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1} |c_j|^2 \ . \tag{21}$$

One way to prove these formulas is to compute that the discrete Fourier modes are orthogonal. The inner product sums are geometric series, but now in two integer variables. Another way is to use the one variable formulas twice. For example, we can use the one variable formula (14) in the a_1 variable with a_2 fixed to define quantities

$$\widetilde{c}_{j_1}(a_2) = \frac{1}{n_1}\sum_{a_1=0}^{n_1-1} e^{-2\pi i j_1 a_1/n_1} f(a_1,a_2) \ .$$

The one variable Fourier inversion formula gives

$$f(a_1,a_2) = \sum_{j_1=0}^{n_1-1} \widetilde{c}_{j_1} e^{2\pi i j_1 a_1/n_1} \ .$$

We can take the one variable DFT of $\widetilde{c}_j(a_2)$ in the a_2 variable, which gives

$$c_{j_1,j_2} = \frac{1}{n_2} e^{-2\pi i j_2 a_2/n_2}\, \widetilde{c}_{j_1}(a_2) \ .$$

If we substitute in the definition of $\widetilde{c}_{j_1}(a_2)$, the result is the same as (19). The two variable inversion formula and Plancharel formula are verified in the same way.

This extends to DFT formulas for functions of m variables. Let $a = (a_1,\ldots,a_m)$ and $j = (j_1,\ldots,j_m)$, which sometimes are called a *multi-indices*. Let $f(a) = f(a_1,\ldots,a_m)$ be a function of m integer variables that is periodic in variable r with period n_r. The discrete Fourier modes are

$$w_j(a) = e^{2\pi i(j_1 a_1/n_1 + \cdots + j_m a_m/n_m)} \ .$$

These are periodic and orthogonal as before.

3 Modular multiplication, $\phi(n)$ and G_n

Dirichlet characters are constructed using the properties of multiplication mod n. The characters are periodic (1), which means that $\chi(x)$ is actually a function of the *residue class* $\overline{x}$, defined below. The characters are multiplicative (2), so they depend on the properties of multiplication mod n. While addition mod n is simple, multiplication is more subtle.

The *residue class* of x mod n is the set of integers equal to x mod n. It is written $\overline{x}$. The formal definition is

$$y \in \overline{x} \ \text{ if } \ y = x + kn \ \text{ for some integer } \ k \ . \tag{22}$$

For example, the residue class of 12 mod 5 is

$$\overline{12} = \{\ldots, -13, -8, -3, 2, 7, 12, 17, 22, \ldots, \} \ .$$

Here are some properties of residue classes.

1. $x \in \overline{x}$, take $k = 0$ in (22).

2. If $y \in \overline{x}$, then $x \in \overline{y}$. If $y = x = kn$, then $x = y + (-k)n$.

3. If $y \in \overline{x}$ then $\overline{y} = \overline{x}$. Since $\overline{x}$ and $\overline{y}$ are sets, saying $\overline{x} = \overline{y}$ is the same as saying that $z \in \overline{x}$ if and only if $z \in \overline{y}$. Two sets that have the same elements are the same set. We show that $z \in \overline{y}$ and $y \in \overline{x}$ implies that $z \in \overline{x}$. The reverse argument is the same. If $z \in \overline{y}$, then $z = y + jn$ for some integer j. If $y \in \overline{x}$, then $y = x + kn$ for some integer k. This implies that $z = (k + kn) + jn = x + (j + k)n$, which implies that $z \in \overline{x}$.

A number $y \in \overline{x}$ is a *representative* of the residue class. Each residue class has exactly one representative among the numbers $\{0, 1, \ldots, n-1\}$. For example, $2 \in \overline{12}$ mod 5. Therefore, there are n residue classes mod n.

One way to define operations on residue classes is to do the operation on representatives and then show that the residue class of the result is independent of the representatives chosen. As an example, here is a way to define addition of residue classes. We want $\overline{x_1} + \overline{x_2}$ to be a residue class. To find it, choose representatives $y_1 \in \overline{x_1}$ and $y_2 \in \overline{x_2}$. Then

$$\overline{x_1} + \overline{x_2} = \overline{y_1 + y_2} \ .$$

The left side is the definition of the + operation on the right. It is one of the n residue classes mod n. The + operation on residue classes is "well defined" because the result is independent of the representatives $y_1 \in \overline{x_1}$ and $y_2 \in \overline{x}_2$. This is the same as saying that the residue class of $y_1 + y_2$ is the same as the residue class of $x_1 + x_2$. To show this, if $y_1 = x_1 + k_1 n$ and $y_2 = x_2 + k_2 n$, then

$$y_1 + y_2 = x_1 + x_2 + (k_1 + k_2)n \ .$$

This shows that $y_1 + y_2$ and $x_1 + x_2$ are in the same residue class.

For example, we write $2 + 4 = 1$ mod 5. The residue class way to say this is $\overline{2} + \overline{4} = \overline{1}$ mod 5. If we take representatives $2 \in \overline{2}$ and $4 \in \overline{4}$, then we get $\overline{2} + \overline{4} = \overline{6} = \overline{1}$. We would get the same answer with representative $-1 \in \overline{4}$, because $\overline{2 + (-1)} = \overline{1}$.

The discrete Fourier modes (7) may be thought of as periodic functions of x or as functions of the residue class $\overline{x}$. The definition could be written

$$w_j(\overline{x}) = e^{2\pi i j y} \ ,$$

where y on the right is any representative of the residue class $\overline{x}$. The definition of $w_j(\overline{x})$ works because the result is independent of the representative. If $z \in \overline{x}$ is a different representative, then $e^{2\pi i j z/n} = e^{2\pi i j y/n}$. We will not distinguish

between $w_j(x)$ and $w_j(\overline{x})$, even though in some sense the functions are different. One is a function of an integer argument x while the other is a function of the residue class $\overline{x}$. The definition has the additive/multiplicative property that

$$w_j(\overline{x} + \overline{y}) = w_j(\overline{x}) w_j(\overline{y}) \ .$$

This differs from the original version (6) in that this one has modular addition while the original version has integer addition. The result is the same because w is periodic.

Dirichlet characters depend on modular multiplication, not modular addition. The definition is similar. You multiply $\overline{x_1}$ and $\overline{x_2}$ by multiplying representatives. The result is independent of the representative chosen, because if $y_1 = x_1 + k_1 n$ and $y_2 = x_2 + k_2 n$, then

$$y_1 y_2 = (x_1 + k_1 n)(x_2 + k_2 n) = x_1 x_2 + (k_1 + k_2 + k1 k_2 n) n \ .$$

That is, $y_1 y_2$ is equal to $x_1 x_2$ plus an integer multiple of n. For example, mod 5 we have $\overline{2} \cdot \overline{4} = \overline{8} = \overline{3}$. This could be written $2 \cdot 4 = 3 \bmod 5$. Both modular addition and modular multiplication are associative and abelian (commutative). The distributive property for modular addition and modular multiplication is

$$(\overline{x} + \overline{y}) + \overline{z} = \overline{x} + (\overline{y} + \overline{z}) \quad , \quad (\overline{x} \cdot \overline{y}) \cdot \overline{z} = \overline{x} \cdot (\overline{y} \cdot \overline{z}) \ .$$

The operations are abeilan

$$\overline{x} + \overline{y} = \overline{y} + \overline{x} \quad , \quad \overline{x} \cdot \overline{y} = \overline{y} \cdot \overline{x} \ .$$

The *identity element* for modular addition is $\overline{0}$, which has

$$\overline{0} + \overline{x} = \overline{x}$$

For modular multiplication it is $\overline{1}$.

Addition and multiplication differ with respect to inverses. For any $\overline{x}$ there is a $\overline{y}$ so that[7] $\overline{x} + \overline{y} = 0$. This $\overline{y}$ is the *additive inverse* of $\overline{x}$ mod n. For example, $\overline{3}$ is the additive inverse of $\overline{2}$ mod 5. Every residue class has an additive inverse. But not every residue class have a multiplicative inverse. For one thing, there is never a multiplicative inverse of 0. There is no x so that $\overline{x} \cdot \overline{0} = 1$.

We can go further. If x is not relatively prime[8] to n then x does not have a multiplicative inverse mod n. To see this, suppose p divides n and p divides x. Then p divides xy for any integer y. For that reason, p divides $n - xy - kn$ for any integer k. This says that $1 \notin \overline{xy}$, because $n - xy - kn \neq 1$ for any integer k. If $(x, n) = 1$ and $y \in \overline{x}$, then also $(y, n) = 1$. If y and n have p in common, then $x = y - kn$ also has p, so $p \leq (x, n)$. We will see (soon) that if $(x, n) = 1$ then x does have a multiplicative inverse mod n. The set of residue classes with $(x, n) = 1$ is written G_n. The *order* of G_n, which is the number of elements in

[7] We should write $\overline{0}$ instead of 0, but we choose instead to abuse notation.

[8] The gcd of x and n is written (x, n). The numbers x and n are relatively prime if and only if $(x, n) = 1$.

G_n, is written $|G_n|$. The *Euler totient* function is $\phi(n) = |G_n|$. Figure 3 lists G_n and $\phi(n)$ for n up to 17. The "Structure" column will be explained soon.

Theorem. $\overline{x}$ has a multiplicative inverse mod n if and only if $(x, n) = 1$.

Proof. The "only if" part is above. If $(x, n) > 1$ then x has no multiplicative inverse mod n. We give to variants on the "if" part. Both of them depend on a lemma, which is the heart of the matter.

Lemma. (*Cancellation Lemma*) If x, y, and z are all relatively prime to n, and $\overline{xz} = \overline{yz}$ mod n, then $\overline{x} = \overline{y}$ mod n.

Proof of the Lemma. Without loss of generality, we may assume x, y, and z are among the numbers $\{1, 2, \ldots, n-1\}$, and $x > y$. If they weren't, we would just take different representatives of the equivalence classes $\overline{x}$, $\overline{y}$ and $\overline{x}$, and possibly interchange x with y. If $\overline{xz} = \overline{yz}$ mod n, then

$$xz = yz + kn \ , \quad \text{for some integer } k.$$

The equation may be rewritten as

$$(x - y)z = kn \ .$$

If the left side is zero, then we showed $x = y$, which proves the lemma. Suppose the left side is not zero. Since $(z, n) = 1$, all the primes in n are in $x - y$; if p^a divides n, then p^a divides $x - y$. Putting all these prime powers together, we conclude that $|x - y| \geq n$. But we chose $0 < x < n$ and $0 < y < n$, and $y < x$, so $0 < x - y < n$. This contradiction proves the lemma.

A proof of the Theorem. The *pigeonhole principle* is the following.[9] Suppose A and B are two finite sets with[10] $|A| = |B| = m$. Suppose there is a function $f : A \to B$ with the property that is *one to one*,[11] which means that $f(a_1) \neq f(a_2)$ if $a_1 \neq a_2$. Then for every $b \in B$, there is exactly one $a \in A$ with $f(a) = b$. The *image* of f in B is the set of $b \in B$ so that $f(a) = b$ for some $a \in A$. It is written[12] $\mathrm{Im}(f) \subseteq B$. If f is one to one, then $|\mathrm{Im}(f)| = |A| = m$. If B is a finite set and $\mathrm{Im}(f) \subseteq B$ and $|\mathrm{Im}(f)| = |B|$, then $\mathrm{Im}(f) = B$. A function is *onto* if $\mathrm{Im}(f) = B$. This means that for every $b \in B$ there is an $a \in A$ with $f(a) = b$.

The name of this principle comes from the picture of pigeons who live in holes. No two pigeons fit in the same hole. If there are the same number of pigeons as holes, then every hole is taken. The pigeonhole principle is *nonconstructive*. You learn that there is some a, but you don't know what a is.

[9] Math majors should practice talking to each other this way. For example: "My date last night had the following agreeable characteristics."

[10] We write $|A|$ for the number of elements in A.

[11] Two to one would mean that there are two values $a_1 \neq a_2$ with $f(a_1) = f(a_2)$. Then f would "take" the two values a_1 and a_2 to one common value in B.

[12] You have to use context to decide whether Im means "image" or "imaginary part".

Now, suppose we have $z \in G_n$ and we want $x \in G_n$ with $xz = 1$. Here, we write $z \in G_n$ for a residue class, not an integer with $\overline{z} \in G_n$. We write $1 \in G_n$ for the residue class of $1 \in \mathbb{Z}$. We "find" x using the pideonhole principle. Define $f(x) = xz$. This maps the finite set G_n to itself, so G_n plays the role of both A and B in the pigeonhole principle. The lemma, which we said was the heart of the matter, shows that f is one to one. Therefore, there is an $x \in G_n$ with $f(x) = xz = 1$. This proves the theorem.

Another proof of the Theorem. Suppose we have $z \in G_n$ and we want $x \in G_n$ so that $xz = 1$. If $z = 1$, then z is its own multiplicative inverse, so suppose $z \neq 1$. Look at the sequence of powers z, z^2, and so on. It might happen that $z^a = 1$ for some a. Since $z \neq 1$, this implies that $a \geq 2$. Choose $x = z^{a-1}$. Then $xz = z^a = 1$, which shows that z^{a-1} is the multiplicative inverse of z. This argument depends on the fact that modular multiplication is associative (so $z^a z^b = z^c$ for any non-negative a and b with $a + b = c$) and commutative.

Can it happen that $z^a \neq 1$ for all positive a? No. Since G_n is a finite set, the powers z^a cannot all be distinct. Therefore, there must be $a < b$ (without loss of generality) with $z^a = z^b$. Write this as $z^a \cdot 1 = z^a z^{b-a}$. The Lemma, with x for z^a and z for z^{b-1} implies that $1 = z^{b-a}$. If $b - a = 1$ we learn $z = 1$, which is a contradiction. If $b > a - 1$, we learn that $z^{b-a-1}z = 1$, which produces the multiplicative inverse. This argument is slightly better than the first one, because this one is a bit more constructive. It tells you a little more about the multiplicative inverse, namely that it is a power of z.

Both proofs rely on the fact that G_n is finite. The theorem may not be true otherwise. For example, the set of positive integers seems to have a lot in common with G_n. The cancellation Lemma is true for positive integers. Yet 2 has no multiplicative inverse. The map $f(x) = 2x$ is one to one, but it is not onto. If B is an infinite set, it is possible that there is a $C \subset B$ (say, the even positive integers as a subset of all positive integers) where $C \neq B$ but they have the same "number" of elements.

To summarize, G_n is the set of residue classes mod n with $(x, n) = 1$. The definition makes sense because if $y \in \overline{x}$ and $(x, n) = 1$ then $(y, n) = 1$. The order of G_n is $\phi(n)$, the Euler phi function. Multiplication of residue classes is well defined, abelian and associative. There is a unique identity element $1 \in G_n$ with $1 \cdot x = x$ for all $x \in G_n$. Every $z \in G_n$ has a unique multiplicative inverse x. We write this as $x = z^{-1}$. You can check that the usual rules of exponents work. For example, the multiplicative inverse of z^2 is x^2, because $x^2 z^2 = xxzz = (xz)(xz) = 1$. Therefore $\left(z^{-1}\right)^2 = \left(z^2\right)^{-1}$. We write this as z^{-2}.

As an example, consider G_{10}. The inverse of $\overline{3}$ is $\overline{7}$, because $\overline{3} \cdot \overline{7} = \overline{21} = \overline{1}$, mod 10. We might write this informally as $3^{-1} = 7 \bmod 10$. The inverse of $\overline{9}$ is $\overline{9}$, because $9 \cdot 9 = 81 = 1 \bmod 10$. But we didn't have to calculate $9 \cdot 9$ to know that $9^{-1} = 9 \bmod 10$. We know from the general theorem that 9 has

n	G_n	structure	$\phi(n)$
2	1	C_1	1
3	1, 2	C_2	2
4	1, 3	C_2	2
5	1, 2, 3, 4	C_4	4
6	1, 5	C_2	2
7	1, 2, 3, 4, 5, 6	C_6	6
8	1, 3, 5, 7	$C_2 \times C_2$	4
9	1, 2, 4, 5, 7, 8	C_6	6
10	1, 3, 7, 9	C_4	4
11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	C_{10}	10
12	1, 5, 7, 11	$C_2 \times C_2$	4
13	1, 3, ... , 12	C_{12}	12
14	1, 3, 5, 9, 11, 13	C_6	6
15	1, 2, 4, 7, 8, 11, 13, 14	$C_4 \times C_2$	8
16	1, 3, 5, 7, 9, 11, 13, 15	$C_4 \times C_2$	8
17	1, 2, 3, ..., 14, 15, 16	C_{16}	16

Figure 1: The multiplicative groups mod k, their structures, and their orders

a multiplicative inverse, and the other possibilities, 3 and 7, were eliminated. Also, $9 = -1 \bmod 10$ and $(-1)^{-1} = -1 \bmod$ any n.

Consider multiplication mod 17 and look for 2^{-1}. We know it must be a power of 2 (from the second proof), so we consider the sequence $2, 4, 8, 16$. We stop at 16 because $16 = -1 \bmod 17$. Therefore $16^2 = 1 \bmod 17$. This means that $16^2 = \left(2^4\right)^2 = 2^8 = 1 \bmod 17$. Therefore, $2^{-1} = 2^7 = 128 \bmod 17$. We can compute the representative of $\overline{128}$ in $\{1, \ldots, 16\}$, which is the remainder in 128/17. Or we can compute powers of 2 mod 17:

$$2 \overset{\times 2}{\to} 4 \overset{\times 2}{\to} 8 \overset{\times 2}{\to} 16 \overset{\times 2}{\to} 32 = 15 \overset{\times 2}{\to} 30 = 13 \overset{\times 2}{\to} 26 = 9 \overset{\times 2}{\to} 18 = 1 \; .$$

Thus, $2 \cdot 9 = 18 = 1 \bmod 17$, so $2^{-1} = 9 \bmod 176$.

The Euler ϕ function is the order of G_n, which is the number of numbers $\{1, 2, \ldots, n-1\}$ relatively prime to n. Figure 3 lists some values. You will notice that $\phi(p) = p - 1$ if p is a prime. You can also verify the formula

$$n = \sum_{d|n} \phi(d) \; . \tag{23}$$

The sum is all numbers d that are divisors of n. For example, the divisors[13] of 7 are $\{1, 7\}$, so the equation (23) is

$$7 = \phi(7) + \phi(1) = 6 + 1 \; .$$

[13] I always forget whether to include the "endpoints" 1 and n in the divisor list. The definition is: d divides n if n/d is an integer. This includes $d = 1$ and $d = n$.

The divisors of 12 are $\{12, 6, 4, 3, 2, 1\}$. The equation (23) is (numbers taken from Figure 3)

$$12 = \phi(12) + \phi(6) + \phi(4) + \phi(3) + \phi(2) + \phi(1) = 4 + 2 + 2 + 2 + 1 + 1 \ .$$

This formula is important and easy to prove, when we're ready.

4 Structure of finite commutativitive groups

The discrete Fourier modes of Section 2 are similar to characters except that the the multiplicative property (2) of characters is replaced by the additive property (6) of discrete Fourier modes. However, these properties may be thought of as almost the same. Both C_n of Section 2 and G_n of Section 3 are examples of the abstract object called a *finite abelian group*. An *isomorphism* of groups (or of any two mathematical objects of the same *category*) is a one to one and onto mapping that "preserves" whatever structure defines the category, the group operation in this case. This section shows that G_n is isomorphic to a product of groups of the form C_k. The multi-variate discrete Fourier modes of the form (18), under such an isomorphism, become characters of G_n.

This section is a crash course on groups, mainly abelian (commutative) groups, mainly finite abelian groups. It goes just far enough to find the characters of G_n. We first define abstract groups and the notion of *isomorphism* between groups. We then describe an abstract version of modular arithmetic. The ideas are the same, only the names of the objects change. Residue classes become *cosets*, etc. We use the notions of *subgroup* and *quotient group* to create an abstract version of reducing mod n. The abstract notion of *product group* generalizes the notion of a function of more than one variable, as in (18). The main theorem is that any finite abelian group, in particular G_n, is isomorphic to a product group, where the *factors* are cyclic groups. When Figure 3 says, for example, that the *structure* of G_{15} is $C_4 \times C_2$, this means that G_{15} is isomorphic to $C_4 \times C_2$.

Let's get started. An abstract *group* is a collection of objects, together with an operation of *composition*. Examples are $\mathbb{R}$ with addition as the composition operation, or $\mathbb{C} - \{0\}$, which is all complex numbers except zero, with multiplication as the composition operation. Another example, which plays no roll in this section, has the set of $n \times n$ invertable matrices with matrix multiplication as the composition operation.

We denote an abstract group by G. The composition is a function that has two arguments that are elements of G and produces an element of G. For $x \in G$ and $y \in G$, the composition may be written "additively" as $a + b$, or "multiplicatively" as xy, or in some generic way such as $x * y$. This set and operation forms a group if the operation has the following properties

- (Associativity) If $x \in G$ and $y \in G$, and $z \in G$, then

$$x * (y * z) = (x * y) * z \ .$$

- (Identity) There is a unique *identity* element $e \in G$ so that $e * x = x$ and $x * e = x$ for all $x \in G$.
- (Invertibility), for any $x \in G$, there is a unique $y \in G$ so that $x * y = e$, and a unique $z \in G$ so that $z * x = e$. We say that y is the right inverse of x and z is the left inverse.

The group is *abelian*, or *commutative*, if it also satisfies

- (Commutativity) For all $x \in G$ and $y \in G$,

$$x * y = y * x \ .$$

All groups in this section are abelian. Non-abelian groups play a big role in other parts of number theory.

Section 2 is based on the fact that C_n is a group under addition. Section 3 is devoted to defining modular multiplication, which is associative on all of C_n. However, multiplication mod n has an inverse only for x with $(x, n) = 1$. This set is *closed* under multiplication, which means that if $(x, n) = 1$ and $(y, n) = 1$, then $(xy, n) = 1$. If neither x nor y has any primes in n, then xy doesn't either. Thus, G_n is an abelian group under multiplication.

As was said, an isomorphism is a structure-preserving identification of two groups. Suppose G and H are groups. We use $*$ for the composition in G or in H, even when these are different. If we must distinguish, we write $*_G$ and $*_H$. A *bijection* one to one and onto map. An *isomorphism* from G to H is a bijection $f\colon G \to H$ so that

$$f(x) * f(y) = f(x * y) \ . \tag{24}$$

A bijection has the property that for every $a \in H$, there is a unique $x \in G$ with $f(x) = a$. We write f^{-1} for the inverse map, which is also a bijection, so $x = f^{-1}(a)$. If $a = f(x)$ and $b = f(y)$, then (24) implies that f^{-1} is also a homomorphism, which means that $f^{-1}(a * b) = f^{-1}(a) * f^{-1}(b)$. We often talk about G and H being *isomorphic*, since the isomorphism goes both ways. An example is the isomorphism between $\mathbb{R}^+$ (the positive real numbers) under multiplication and $\mathbb{R}$ under addition. The isomorphism one way is log and the other way is the exponential.

Figure 3 suggests that the groups G_p for prime p are isomorphic to C_{p-1}. A proof of this is below. But for now, an example. A *generator* of a group G is an element g so that all of G consists of powers of g. That is, $G = \{e, g, g * g, g * g * g, \cdots, \}$. This looks simpler if we write the operation multiplicatively:

$$G = \left\{1, g, g^2, \cdots, g^a, \ldots, g^{m-1}\right\} \ .$$

Of course, m is the order of G. We say that G is *cyclic* if it has a generator in this sense. Most cyclic groups have more than one generator. To see whether 2 is a generator of G_5, we compute the powers

$$2 \stackrel{\times 2}{\to} 2^2 = 4 \stackrel{\times 2}{\to} 2^3 = 8 = 3 \stackrel{\times 2}{\to} 2^4 = 6 = 1 \ .$$

This shows that $g = 2$ is a generator for G_5, and that the corresponding *cycle* is $\{1, 2, 4, 3\}$. To see whether 3 is a generator, we calculate

$$3 \stackrel{\times 3}{\to} 3^2 = 9 = 4 \stackrel{\times 3}{\to} 3^3 = 12 = 2 \stackrel{\times 3}{\to} 3^4 = 6 = 1 \ .$$

This shows that 3 also generates G_5.

A group G is cyclic or order m if and only if G is isomorphic to the additive group C_m. The additive group C_m has elements $\{0, 1, \ldots, m-1\}$. The isomorphism takes $1 \in G$ to $0 \in C_m$ and $g \in G$ to $1 \in C_m$. After that, the isomorphism is determined by the group property (24). We see that $g^2 = g * 2 \stackrel{f}{\to} 1 * 1 = 1 + 1 = 2$, and generally $g^a \stackrel{f}{\to} a$. Recall that C_m is additive, which is why $1 * 1 = 1 + 1$ in that group. The mapping $f(g^a) = a$ is well defined because of the properties of exponent, and because g generates a cycle of order m. For example, $g^m = 1$ in G, and $m = 0$ in C_m, so the mapping $1 = g^m \to m = 0$ is consistent.

All the multiplicative groups G_p, for prime p, are cyclic, as we will see. But it isn't always easy to find a generator. For example, 2 is not a generator of G_7 because mod 7,

$$2 \stackrel{\times 2}{\to} 2^2 = 4 \stackrel{\times 2}{\to} 2^3 = 8 = 1 \ .$$

To be a generator of G_7, an element has to generate a cycle of length $|G_7| = 6$. Instead, 2 generates a cycle of length 3. To see whether 3 is a generator, we calculate the cycle 3 generates:

$$3 \stackrel{\times 3}{\to} 3 \cdot 3 = 9 = 2 \stackrel{\times 3}{\to} 2 \cdot 3 = 6 = 4 \stackrel{\times 3}{\to} 4 \cdot 3 = 12 = 5 \stackrel{\times 3}{\to} 5 \cdot 3 = 15 = 1 \ .$$

This is a cycle of length $|G_7| = 6$, so 3 is a generator of G_7.

Some of the groups G_n are not cyclic. The first one is G_8 with order 4. The *order* of an element $x \in G$ is the smallest $a > 0$ with $x^a = 1$. If G is cyclic, then there is at least one $x \in G$ whose order is the order of G. We know G_8 is not cyclic because G_8 has no elements of order 4. We just check the three elements that are not the identity, all calculations mod 8:

$$3 \stackrel{\times 3}{\to} 3 \cdot 3 = 1 \qquad \text{(order 2)}$$

$$5 \stackrel{\times 5}{\to} 5 \cdot 5 = 25 = 24 + 1 = 1 \qquad \text{(order 2)}$$

$$7 \stackrel{\times 7}{\to} 7 \cdot 7 = 49 = 48 + 1 = 1 \qquad \text{(order 2)}$$

Exercise 6 asks you to explore the cycle lengths in other G_n. The structure of G_8 involves a *cartesian product* of groups, which we now describe.

Suppose G and H are groups. The *cartesian product*, also called *direct product* or *cross product*, is the set of pairs (x, y) with $x \in G$ and $y \in H$. It is written $G \times H$. You do the group operation in the product by doing the appropriate operation in each *factor*

$$(x_1, y_1) * (x_2, y_2) = (x_1 * x_2, y_1 * y_2) \ .$$

For example, if G and H are both the additive integers $\mathbb{Z}$, then the product is $\mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$. This consists of all "integer lattice points" in the plane. You add two elements of $\mathbb{Z}^2$ using ordinary vector addition: you add the x coordinates and add the y coordinates. If G and H are finite, then the number of pairs is $|G \times H| = |G| \cdot |H|$.

This fact may be surprising at first. The product of C_m and C_n "is" (more properly, is isomorphic to) C_{mn} if $(m, n) = 1$. To see this, let g be a generator of G and h a generator of H. Then (writing multiplicatively)

$$(g, h)^a = (g^a, h^a) \ ,$$

because the operation is done separately in each factor. The element $(g, h)^a$ is the identity in $G \times H$ if and only if it is the identity in each factor. This means $(g^a, h^a) = (1, 1)$. Be careful here, in $(1, 1)$, the first 1 is the identity element of G and the second 1 is the identity in H. If the elements of G and H are residue classes modulo m and n respectively, then $\overline{1}$ mod m is not the same set as $\overline{1}$ mod n. You are welcome to write $(1_G, 1_H)$, if this makes it clearer for you. Now the punch line: If $(g^a, h^a) = (1, 1)$, then $g^a = 1$ and $h^a = 1$. Since g is order m, $g^a = 1$ implies that a is a multiple of m. Similarly, $h^a = 1$ implies that a is a multiple of n. If $(m, n) = 1$, then the smallest integer that is a multiple of m and a multiple of n is mn. This shows that the smallest a with $(g, h)^a = (1, 1)$ is $a = mn$. Therefore, the elements $(g, h)^a$ form a cycle of length $mn = |G \times H|$. This shows that (g, h) is a generator of $G \times H$, and that it is cyclic.

It is possible to take the cross product of more than two groups. The cross product operation is associative in the sense that

$$(G_1 \times G_2) \times G_3 = G_1 \left(\times G_2 \times G_3\right) \ .$$

Therefore, we write it without parentheses as $G_1 \times G_2 \times G_3$. The elements of $G_1 \times G_2 \times G_3$ are triples (x_1, x_2, x_3), with $x_i \in G_i$ and elementwise composition. You can check, for example, that $C_2 \times C_3 \times C_5$ is a isomorphic to C_{30}.

Theorem. Every finite abelian group is isomorphic to a product of cyclic groups.

Remarks. We are interested in the groups G_n specifically. It seems to be easier to give an abstract proof that applies to all finite abelian groups. Much of the reasoning in the proof applies only to abelian groups. Non-abelian groups are more complicated.

Proof. The proof is a kind of induction. We build an increasing family of *subgroups* $H_1 \subset H_2 \subset \cdots \subset G$ that are isomorphic to products of cyclic groups. We "build" H_{j+1} from H_j by adding a new generator. This makes H_{j+1} bigger than H_j. Since G is finite, this process must stop, and can only stop with $H_k = G$.

The construction uses the notions of subgroup (as we said), and quotient group. If G is a group and $H \subset G$, then H is a *subgroup* of G if H is *closed*

under the group operation and under inverses. That is, if $h_1 \in H$ and $h_2 \in H$, then $h_1 * h_2 \in H$ and if $h \in H$ then $h^{-1} \in H$. You can see that if H is a subgroup, then $e \in H$. For if $h \in H$, then $h^{-1} \in H$ and therefore $h * h^{-1} = e \in H$. The subgroup H is *trivial* if it consists only of e. The subgroup H is *proper* if it is not G, if there is some $x \in G$ with $x \notin H$. A subgroup is *non-trivial* if it is not trivial and proper.

For $G = \mathbb{Z}$, the non-trivial subgroups are $n\mathbb{Z}$, which is the set of integer multiples of n. Every non-trivial subgroup of $\mathbb{Z}$ has this form. If G is a product group $G = G_1 \times G_2$, then there is a subgroup isomorphic to G_1 which consists of all elements of the form $(x, 1)$, with $x \in G_1$ and 1 the identity in G_2. This subgroup is non-trivial if neither G_1 nor G_2 is the trivial group. If G is any finite group and $x \in G$, then the cycle generated by x is a subgroup of G that is isomorphic to to C_m with m being the order of x. The subgroup H is the set $H = \left\{1, x, x^2, \ldots, x^{m-1}\right\}$. This is closed under the group operation because $x^a * x^b = x^{a+b}$. It is closed under inverses because $x^{-1} = x^{m-1}$. This subgroup is proper if x is not a generator of G.

If G is an abelian group and $H \subset G$ is a subgroup, then there is a *quotient* group G/H whose elements are *cosets*. For any $x \in G$, the coset containing x will be written $\overline{x}$. The definition is that $y \in \overline{x}$ if there is an $h \in H$ with $y = hx$ (writing the group operation multiplicatively). Cosets are a generalization of residue classes. If $G = \mathbb{Z}$ and $H = n\mathbb{Z}$ (additive groups), then elements h have the form kn for integer k. The statement $y = x + h$ (writing the group operation additively for this example) is the same as $y = x + kn$, which is our definition of residue class.

You can check that general cosets for abelian groups have the properties of residue classes. There is an identity coset, which is $\overline{1} = H$. If $x \in \overline{y}$, then $y \in \overline{x}$. If $x \in \overline{y}$ and $y \in \overline{z}$, then $x \in \overline{z}$. For example, if $x \in \overline{y}$, then there is an $h \in H$ with $x = hy$. We multiply by h^{-1} and use the associativity property to get $h^{-1}x = y$, where, of course, $h^{-1} \in H$. Therefore $y \in \overline{x}$. If $x = h_1 y$ and $y = h_2 z$, then $x = h_1 h_2 z$, so $x \in \overline{z}$. If G is finite, then all cosets are the same size:[14] $|\overline{x}| = |H|$. We can think of each coset as a copy of H.[15] Distinct cosets are disjoint. If $\overline{x} \neq \overline{y}$, and $z \in \overline{x}$, then $z \notin \overline{y}$. Therefore, G consists of a collection of cosets of equal size. For that reason $|G| = \#\left\{cosets\right\} \cdot |H|$. The order of H is a divisor of the order of G. In particular, if $x \in G$, then the order of x divides $|G|$, because the cycle generated by x is a subgroup of G. If x generates a cycle of length k in G_n, then k divides $\phi(n)$. This is true for all the cycle lengths listed in Figure 3.

The cosets form a group if G is abelian. You "multiply" (or add or do whatever the group operation is) two cosets by multiplying representatives and taking the corresponding coset.

$$\overline{x} \cdot \overline{y} = \overline{x \cdot y} \; .$$

[14] There is a bijection between elements of H and elements of $\overline{x}$. Associate to each $h \in H$ the value $hx \in \overline{x}$. If $h_1 \neq h_2$ then $h_1, \neq h_2 x$ (the cancellation property). Every element of $\overline{x}$ has this from.

[15] It is possible that "co" in "coset" is for "copy".

We have to verify that this makes sense and leads to a group. Making sense means that the resulting coset is independent of the representatives chosen. Suppose $x' \in \overline{x}$, and $y' \in \overline{y}$, and $z = xy$, and $z' = x'y'$. We must verify that $z' \in \overline{z}$. Here goes. If $x' \in \overline{x}$, then there is an $h \in H$ with $x' = hx$. Similarly, there is a $k \in H$ with $y' = ky$. Then[16] $z' = (ha) \cdot (ky) = (hk) \cdot (xy) = (hk)z$. Since H closed under multiplication, $hk \in H$, which shows that $z' \in \overline{z}$.

You should check that modular addition of residue classes is the same as this general construction with $H = n\mathbb{Z}$ and $G = \mathbb{Z}$. The arguments given for that in Section 3 are exactly those given here, with addition as the group operation. The argument given there for modular multiplication is not the same, because it mixes the two operations addition and multiplication. In fact, the corresponding "theorem" is not true. Residue classes do not form a group under residue class multiplication, only the residue classes relatively prime to n. This is why $|G_n| < n$. You have to leave out some cosets.

You might worry that if we need to define Dirichlet characters $\chi(x)$ for all x, not just x with $(x, n) = 1$. We do, but the definition is $\chi(x) = 0$ if $(x, n) > 1$. The purpose of Dirichlet characters is to find primes equal to x mod n, but only if $(x, n) = 1$. The value of $\chi(x)$ when $(x, n) > 1$ is not relevant in the Euler product for $L(x, \chi)$.

Here are the examples of subgroup and quotient group we are interested in. If $G = G_1 \times G_2$ and $H = G_1$, then $G/H = G_2$. Suppose G is a cycle of length nm generated by g. Then the element $h = g^n$ generates a cycle of length m. The quotient group is (isomorphic to) a cycle of length n. Given the theorem we are proving, these must essentially be the only examples.

Here is a strategy for proving the theorem. Let H be a cyclic subgroup of G and let $K = G/H$ be the quotient group. If $G = H \times K$, and if K is a product of cyclic groups, then G is also such a product. The problem is that G need not be isomorphic to $H \times K$. For example, suppose $G = C_4$ generated by g with $g^4 = 1$, but $g^a \neq 1$ for $a = 1, 2, 3$. Let H be generated by g^2, so $H = \{1, g^2\}$. The cosets are $\overline{1} = \{1, g^2\}$, and $\overline{g} = \{g, g^3\}$. Thus, K is isomorphic to C_2, with elements $\{\overline{1}, \overline{g}\}$. This is cyclic of order 2 because, in K,

$$\overline{g}^{\,2} = \overline{g^2} = \overline{1}\,.$$

In this example,[17] $H = C_2$, $K = C_2$, but $G \neq C_2 \times C_2$.

Here, finally, is the proof of the theorem. We work by induction, building an increasing family of subgroups $H_j \subseteq G$ with each H_j being a product

$$H_j = C_{n_1} \times \cdots \times C_{n_j}\,.$$

The induction step is to show that if $H_j \neq G$, then it is possible to find another factor $C_{n_{j+1}}$ and a larger subgroup H_{j+1} in G. To get started, let n_1 be the length of the longest cycle in G and choose $h_1 \in G$ that has that cycle length.

[16] This uses the fact that the group is abelian. In non-abelian groups, multiplication of cosets may not be well defined.

[17] We write = between groups when they are isomorphic.

The step from $j = 1$ to $j = 2$ illustrates the trick to avoid the counter-example above. We have a cycle of length h_1 generated by h_1. Consider the quotient group $K_2 = G/H_1$. Let n_2 be the length of the longest cycle in K_2. Then[18] $n_2 \leq n_1$. This implies that $\overline{h_2}^{n_2} = \overline{1}$ in K_2, which means that $h_2^{n_2} = h_1^r$, for some $r \in \{0, 1, \ldots, n_1 - 1\}$. If $r = 0$, which means that $h_2^{n_2} = 1$, then we've succeeded. There is a map from $C_{n_1} \times C_{n_2}$ into H_2, which is the "span" of h_1 and h_2. To define the map, take $(a_1, a_2) \in C_{n_1} \times C_{n_2}$, and map it to

$$f(a_1, a_2) = h_1^{a_1} h_2^{a_2} \ .$$

Let $H_2 \subseteq G$ be the set of all elements of G of that form. You can check that f is one to one, onto H (by definition), and preserves the group operations in $C_{n_1} \times C_{n_2}$ and H, which makes f an isomorphism.

But what happens if $h_2^{n_2} \neq 1$ in G? That's what happened in the counter-example. The answer is that we replace h_2 with $\widetilde{h}_2 = h_2 h_1^\alpha$ and we show that there is an α so that $\widetilde{h}_2^{n_2} = 1$ in G.

4.1 Proof that G_p is cyclic

The group G_p has two choices. Either it can be a single cycle of size $p - 1$, or it can be a product of shorter cycles. It cannot be a product of shorter cycles because if it were, there would be too many distinct solutions to the equation $x^d = 1$. Recall that the order of an element a is the smallest n with $a^n = 1$.

5 Dirichlet characters

Let G be a finite abelian group. A *character* of G is complex valued function, $\chi(x)$, defined for all $x \in G$ with the property that

$$\chi(x * y) = \chi(x)\chi(y) \tag{25}$$

for any $x \in G$ and $y \in G$. If the group operation is written additively, this would be written $\chi(x+y) = \chi(x)\chi(y)$. If the operation is written multiplicatively, the character relation would be written $\chi(xy) = \chi(x)\chi(y)$. Of course, these are all equivalent. The additive one was done in Section 2. It's the DFT. In our notation, suppose g is a generator for the cyclic group C_n, then there are n characters of the form

$$\chi_j(g^k) = e^{2\pi i j k/n} \ .$$

You can see that every character has this form. If χ is any character, then $(\chi(g))^n = \chi(g^n) = 1$, so $\chi(g)$ is one of the numbers $e^{2\pi i j/n}$.

The more complicated case, where G is a product of cycles, is the same but with bookkeeping. Let m be the number of cycles, and n_r the order of cycle r.

[18] If $n_2 > n_1$, then there is an $h_2 \in G$ with $\overline{h_2}^a \neq \overline{1}$ in K_2 for $a < n_2$. That implies that $h_2^a \neq 1$ in G, so h_2 generates a cycle in G at least as long as $n_2 > n_1$, which contradicts n_1 being the longest cycle length in G.

Let g_r be a generator of cycle r. A typical element of G has the form

$$x = g_1^{k_1} \cdots g_m^{k_m} \ .$$

The numbers k_r are in the range $\{0, 1, \ldots, n_r - 1\}$. The order of G is $n = n_1 n_2 \cdots n_m$. We pick "mode numbers" j_r in the range $\{0, 1, \ldots, n_r - 1\}$. The number of choices is n. We write j for the collection of mode numbers: $j = (j_1, \ldots, j_r)$. In analysis, j might be called a "multi-index". The character corresponding to j is[19]

$$\chi_j(x) = e^{2\pi i j_1 k_1/n_1} \cdots e^{2\pi i j_m k_m/n_m} = e\left(\frac{j_1 k_1}{n_1} + \cdots + \frac{j_r k_r}{n_r}\right) \ .$$

As for the single cycle case, you can see that every character of G has this form. The number of characters is the order of G.

We re-state some of the results from Section 2 in the present notation. Suppose $f(x)$ is a complex valued function defined on the group G. Then f may be represented in terms of characters

$$f(x) = \sum_\chi c_\chi \chi(x) \ . \tag{26}$$

The sum is over all n characters of G. The coefficients are given by

$$c_\chi = \frac{1}{n} \sum_G \overline{\chi}(x) f(x) \ .$$

This may be written in a slightly different way. Since $\chi(x)$ is a complex number of magnitude 1, and since χ is a character of G, we have

$$\overline{\chi}(x) = \chi(x)^{-1} = \chi(x^{-1}) \ .$$

The formula for the coefficient may be written as

$$c_\chi = \frac{1}{n} \sum_G \chi(x^{-1}) f(x) \ . \tag{27}$$

The Dirichlet characters are the characters of G_n, the equivalence classes of numbers relatively prime to n. The order of G_n is $\phi(n)$. If a is an integer relatively prime to n, we write $\chi(a) = \chi(\overline{a})$. It is an "abuse of notation" to write $\chi(a)$ when a is an integer, but without abuses like this mathematics would be much harder to read. In the simple case where $n = p$ is prime and g is a generator mod p, there is a k with $a \equiv g^k \bmod p$. The Dirichlet character is

$$\chi_j(a) = e^{2\pi i j k/(p-1)} \ .$$

So far, we have not given a value to $\chi(0)$, or to $\chi(a)$ if a is not relatively prime to n. The convention (which we will see is quite convenient) is to set $\chi(a) = 0$

[19] In number theory, some people use the notation $e(t) = e^{2\pi i t}$.

if a is not relatively prime to n. The number of Dirichlet characters mod n is $\phi(n)$.

Dirichlet characters are used to represent functions $f(x)$ of an integer x that are periodic with period n and are non-zero only if x is relatively prime to n. Any such function can be written as a linear combination of Dirichlet characters:

$$f(x) = \sum_{\chi} c_\chi \chi(x) \ , \tag{28}$$

where the sum is over Dirichlet characters mod n, and

$$c_\chi = \frac{1}{\phi(n)} \sum_{(x,n)=1} \overline{\chi}(x) f(x) \ .$$

Recall that $(x,n) = \gcd(x,n) = 1$ if x is relatively prime to n. I did not use the form $\chi(x^{-1})$, because I don't know a simple formula for the multiplicative inverse of x mod n. If we decide $\chi(x) = 0$ if $(x,n) > 1$, then the coefficient formula may be written

$$c_\chi = \frac{1}{\phi(n)} \sum_{0}^{n-1} \overline{\chi}(x) f(x) \ . \tag{29}$$

The representation formula (26) then implies that $f(x) = 0$ if $(x,n) > 1$.

Dirichlet characters are useful for representing many specific functions f. But for Dirichlet's theorem on prime numbers in an arithmetic progression, we are interested in a specific function. Suppose $(a,n) = 1$ and we want to prove there are infinitely many primes equal to a mod n. The function is

$$f(x) = \begin{cases} 1 & \text{if } x \equiv a \bmod n \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

When we expand this f in terms of characters, the sum (29) reduce to

$$c_\chi = \frac{1}{\phi(n)} \overline{\chi}(a) \ . \tag{31}$$

We are interested in a that is relatively prime to n, written $(a,n) = 1$. For such an a, we know that $\chi(a) \neq 0$ for any character:

$$c_\chi \neq 0 \ . \tag{32}$$

This is because the characters all take values $e^{i\xi}$ where ξ is real if $(a,n) = 1$.

The point of this representation will become clear in Section 6. The first term in parentheses is the mean, which is the important term. The rest of the terms have mean value zero. Because of *cancellations* these terms contribute less. As a function of x, $\left|\overline{\chi}_j(a)\chi_j(x)\right|$ is the same size as 1 (exactly the same size). The cancellation comes from the fact that summing over x gives zero. If $j \neq 0$, then

$$\sum_{0}^{n-1} \chi_j(x) = 0 \ .$$

This is because of orthogonality:

$$\langle \chi_0, \chi_j \rangle = \sum \overline{\chi}_0(x)\chi_j(x) = \sum \chi_j(x) = 0 \; .$$

The summand $\chi(x)$ is zero if $(x, n) > 1$, so you get the same answer if you sum over all $0, 1, \ldots, n-1$, or just the relatively prime x. The formula was derived for summation over G_n, which is relatively prime x, but you don't have to distinguish. We will use this.

6 L functions and the Dirichlet theorem

It is almost straightforward to assemble the proof of Dirichlet's theorem on primes in an arithmetic progression. We will use the dominated convergence theorem to prove that if $s > 0$ the L function (3) has the product representation (4). We will calculate the logarithmic derivative, as we did for the Riemann zeta function, to get a sum involving primes in an arithmetic progression. The algebra of the DFT and Dirichlet characters comes in here.

The one new thing is the *non-vanishing* of non-principal Dirichlet series at $s = 1$:

$$L_\chi(1) \neq 0 \; . \tag{33}$$

We can use ideas we already have to prove that $L_\chi(s) \sim \frac{C}{s-1}$ as $s \downarrow 1$ for the principal character. It also is "routine" (uses known ideas in a known way) to show that the sum (3) converges when $s = 1$ for non-principal χ. But it does not converge absolutely. The Euler product for non-principal L_χ probably does not converge absolutely (we can verify the theorems we are proving later). So it seems possible that the product (4) converges to zero, as this Euler product does:

$$\frac{1}{\zeta(1)} = \prod_p \left(1 - p^{-1}\right) = \lim_{x \to \infty} \prod_{p \leq x} \left(1 - p^{-1}\right) = 0 \; .$$

First the routine stuff. If $s > 1$ then the product (4) converges absolutely. The bound that worked for the zeta function works here too. For $p \geq 2$ (as all primes are) a Taylor series (or mean value theorem) calculation gives the inequality

$$\left| 1 - \left(1 - \chi(p)p^{-s}\right)^{-1} \right| \leq C \left| \chi(p)p^{-s} \right| = C\, p^{-s} \; .$$

Recall that for $s > 1$ the prime sum is finite, because (for example) it is less than the sum over all positive integers

$$\sum_p p^{-s} \leq \sum_n n^{-s} = \zeta(s) < \infty \; .$$

Next, the product (4) is equal to the sum (3) if $s > 1$. This too may be done using our zeta function methods. The partial products are

$$P_x = \prod_{p \leq x} \left(1 - \chi(p)p^{-s}\right)^{-1} \; .$$

This is equal to the sum over integers whose prime factorization contains only $p \le x$ (as we saw when we were talking about zeta). Denote this set by F_x. We say $n \in F_x$ if $p|n \Rightarrow n \le x$. Then

$$\prod_{p \le x} \left(1 - \chi(p)p^{-s}\right)^{-1} = \sum_{n \in F_x} \chi(n) n^{-s} \ .$$

To use the dominated convergence theorem, we express the right side sum as

$$\sum_1^\infty a_n(x) \ ,$$

where

$$a_n(x) = \begin{cases} \chi(n)n^{-s} & \text{if } \ n \in F_x \\ 0 & \text{otherwise} \ . \end{cases}$$

Clearly, $a_n(x) \to \chi(n)n^{-s}$ as $x \to \infty$. In fact, $a_n(x) = \chi(n)n^{-s}$ if x is larger than the largest prime factor of n. Also, $|a_n(x)| \le n^{-s}$ for all x. Therefore $D_n = n^{-s}$ is a dominating series. The D_n sum is finite for $s > 1$ because it's the zeta sum. This proves that $P_x(s) \to L_\chi(s)$ as $x \to \infty$.

Next, the non-principal sums are finite even for $s = 1$. Any non-principal character has the property that

$$\sum_{j=kn+1}^{(k+1)n} \chi(j) = 0 \ . \tag{34}$$

This is true with $k = 0$ because the non-principal character is orthogonal to the principal character. It is true for $k > 0$ because a character is a periodic function. We use this to find *cancellation* in the Dirichlet series (3). We write the sum as a sum over periods

$$L_\chi(s) = \sum_{k=0}^\infty \left(\sum_{j=kn+1}^{(k+1)n} \chi(j) j^{-s} \right) = \sum_1^\infty S_k \ .$$

For convenience, we rewrite the period sum in the form

$$S_k = \sum_{j=1}^n \chi(j)(kn+j)^{-s} \ . \tag{35}$$

The cancellation is seen in the fact that the period sums are one power smaller than the individual terms. That is, we will prove

$$\left| \sum_{j=1}^n \chi(j)(kn+j)^{-s} \right| = |S_k| \le C k^{-s-1} \ . \tag{36}$$

This implies that the S_k sum converges absolutely for any $s > 0$. In particular, the S_k sum provides a definition of $L_\chi(s)$ that makes sense for any $s > 0$. For

$s > 1$ the two definitions (the S_k sum and the original Dirichlet sum) agree (because the Dirichlet sum converges absolutely and therefore doesn't depend on the order of the terms).

This is an example of *analytic continuation*. Analytic continuation refers to a definition (an infinite sum or product or integral) that makes sense for some parameter range. To *continue* the function, you find a different definition that agrees with your original definition where it makes sense. If the new definition makes sense in a wider parameter range, we say that we have *continued* the function. The geometric series is a simple example. The series

$$S(z) = \sum_0^\infty z^n$$

converges absolutely if $|z| < 1$. However, in that z range, we have the formula

$$S(z) = \sum_0^\infty z^n = \frac{1}{1-z} \ .$$

The formula $S(z) = 1/(1-z)$ gives a *continuation* of $S(z)$ to all z values except $z = 1$. In this spirit, the cancellation inequality (36) extends $L_\chi(s)$ from the original $s > 1$ to $s > 0$.

The approximate cancellation inequality (36) is a consequence of the exact cancellation (34) and the fact that the function k^{-s} is "flatter than it is small". The smallness is $k^{-s} \to 0$ as $k \to \infty$. The flatness is that the derivative (with respect to n) goes to zero faster:

$$\frac{d}{dk} k^{-s} \sim k^{-s-1} \ .$$

In fact, if $1 \leq j \leq n$ then

$$\left| k^{-s} - (k+j)^{-s} \right| \leq C k^{-s-1} \ . \tag{37}$$

This follows from the mean value theorem, which gives, in this case,

$$k^{-s} - (k+j)^{-s} = j(-s)\xi^{-s-1} \ , \quad \text{for some } \xi \text{ in the interval } k \leq \xi \leq k+j \ .$$

This implies the inequality (37) with $C = ns$. This is adequate for us here, because $s > 0$ is fixed and the period n is fixed while $k \to \infty$.

If $(kn+j)^{-s}$ were constant in the range $1 \leq j \leq n$, then the character cancellation formula (34) would imply that $S_k = 0$. Instead, $(kn+j)^{-s}$ is almost constant, in the sense of (37). This is the "flatness" mentioned before, and it gives the approximate cancellation

$$\begin{aligned} S_k &= \sum_1^n \chi(j)(kn+j)^{-s} \\ &= \sum_1^n \chi(j)(kn)^{-s} + \sum_1^n \chi(j) \left((kn+j)^{-s} - (kn)^{-s} \right) \end{aligned}$$

The first sum on the right is zero, so we have the bound

$$\begin{aligned} |S_k| &\le \sum_1^n \left|(kn+j)^{-s} - (kn)^{-s}\right| \\ &\le \sum_1^n C(kn)^{-s-1} \\ &\le n\, n^{-s-1} C k^{-s-1} \ . \end{aligned}$$

This proves the bound we need (36).

The last simple thing is the $L_\chi(s)$ for the principal character. In this case, we have

$$\sum_1^n \chi(j) = \phi(n) \ .$$

Therefore, the argument that gave (36) for the non-principal characters now gives (for $k > 0$)

$$S_k = \phi(n)(kn)^{-s} + O\big(\, k^{-s-1}\big) \ .$$

Therefore, for any $\sigma_0 > 0$, we can say that

$$\sum_{k=0}^\infty S_k = \frac{\phi(n)}{n^s} k^{-s} + O\big(\, k^{-s-1}\big) = \frac{\phi(n)}{n^s}\zeta(s) + O(1) \ .$$

If you want to know what happens for s near 1, you can set $s = 1$ in the denominator (justify this). The result is

$$L_\chi(s) = \frac{\phi(n)}{n}\frac{1}{s-1} + O(1) \ .$$

This is an analytic continuation of L_χ for the principal character to values $s > 0$ with $s \ne 1$.

The information about primes in an arithmetic progression comes from the a linear combination of logarithmic derivatives of the L functions of period n. Suppose $s > 1$. As for the zeta function, we have

$$\log(L_\chi(s)) = -\sum_p \log\left(1 - \chi(p)p^{-s}\right) \ .$$

Therefore (note: $1 - \chi(p)p^{-s} \neq 0$ for $s > 2$ and $p \geq 2$)

$$\begin{aligned}
\frac{L'_\chi(s)}{L_\chi(s)} &= \frac{d}{ds}\log(L_\chi(s)) \\
&= -\sum_p \frac{d}{ds}\log\left(1 - \chi(p)p^{-s}\right) \\
&= -\sum_p \frac{\frac{d}{ds}\left(1 - \chi(p)p^{-s}\right)}{1 - \chi(p)p^{-s}} \\
&= -\sum_p \frac{\chi(p)\log(p)p^{-s}}{1 - \chi(p)p^{-s}} \\
\frac{L'_\chi(s)}{L_\chi(s)} &= -\sum_{k=1}^{\infty} \chi(k)\Lambda(k)k^{-s}\ . \qquad (38)
\end{aligned}$$

The algebra of the last step is like the algebra that we used in the Euler product section to calculate ζ'/ζ (look it up). You use the geometric series formula

$$\frac{\chi(p)p^{-s}}{1 - \chi(p)p^{-s}} = \sum_{j=1}^{\infty} \chi(p)^j \left(p^j\right)^{-s}$$

The character is multiplicative, so $\chi(p)^j = \chi(p^j)$. If $k = p^j$, then $\log(p) = \Lambda(k)$. This accounts for all the non-zero terms on the right of (38), because $\Lambda(k) = 0$ if k is not a prime power.

7 Exercises

1. Define the $n \times n$ complex matrix M as follows. The discrete Fourier representation (8), written more explicitly as (15), defines a *map* that takes a vector $c \in \mathbb{C}^n$ to a vector $f \in \mathbb{C}^n$, which is written

$$c \mapsto f = Mc\ .$$

   (a) Identify the entries m_{jk}. It will be convenient to write m_{jk} as a power of $z = e^{2\pi i/n}$.

   (b) Calculate M^2, M^4 and M^*M. Here, M^* is the conjugate transpose of M. The jk entry of M^* is the complex conjugate of the kj entry of M. If M were real, then $M^* = M^t$ (M^t is the transpose), but our M is not real. Use these calculations to identify the determinant $\det(M)$ and the possible eigenvalues of M. A *Vandermonde* matrix is a matrix whose columns are powers of a given number:

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{pmatrix}$$

Show that M is a Vandermonde matrix and identify the numbers x_k in terms of z.

(c) The Vandermone formula is

$$\det(V) = \prod_{j<k} (x_j - x_k) \ , \tag{39}$$

(or maybe the negative of this?). Show that this formula correctly predicts that V is singular if $x_j = x_k$ for any pair with $j \neq k$.

(d) (*harder, do if you have time*) Prove (39). *Hint*: Start by seeing what happens when you use Gauss elimination to put zeros below the 1 in the first column. Then continue to the second column, putting the matrix in upper triangular form. There is a pattern you may spot.

(e) Use the Vandermonde formula (39) to find a different formula for $\det(M)$ (or possibly $\det(M^*M)$ or $\det(M^2)$). Use this to show

$$\prod_{1}^{n-1} \left(z^j - 1\right) = n \ .$$

(f) Check this formula explicitly for $n = 4$ by doing a bit of arithmetic with complex numbers.

2. The Plancharel identities such as (16) lead to many interesting specific formulas when you put in specific functions f.

(a) What formula do you get if you take $f(0) = 1$ and $f(a) = 0$ if $a \neq 0$ mod n?

(b) What formula do you get if you take n to be even and $f(a) = 1$ for $a = 0, 1, \ldots, n/2$, and $f(a) = 0$ for $a = n/2 + 1, \ldots, n - 1$?

(c) (*More challenging and time consuming, do as time permits*) Take the limit $n \to \infty$ and get a formula for a simple infinite sum. Try to evaluate the limits of the discrete Fourier coefficients with fixed k as $n \to \infty$:

$$b_k = \lim_{n\to\infty} c_k \ , \quad d_k = \lim_{n\to\infty} c_{n-k} \ .$$

The answers are simpler than the c_k formulas. You can relate the formula you get to Euler's formula

$$S_1 = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

using a trick:

$$S_1 = \sum_{k \text{ even}} \frac{1}{k^2} + \sum_{k \text{ odd}} \frac{1}{k^2} \ ,$$

and

$$S_1 = \sum_{k \text{ even}} \frac{1}{k^2} = \frac{1}{4} S_1 \ ,$$

(why?).

3. Show that if x is relatively prime to n and if $y \equiv x \bmod n$, then y is relatively prime to n. Hint: If y and n have a prime in common, then x and n have the same prime in common.

4. Show that if x, y, and z are all relatively prime to n, and if $xy \equiv xz \bmod n$, then $y \equiv z \bmod n$.

5. Here are three questions about generators of G_p. I do not know the answers to all of them, but I believe experts in number theory do know. Try to figure them out:

   (a) For which primes p does $\overline{2}$ generate G_p?

   (b) How many generators does G_p have?

   (c) Is there a formula or simple recipe for finding a generator of G_p that is better than trial and error?

6. Verify the cycle lengths and cycle structure and identify generators for G_{15} and G_{16}.

7. The *Chinese remainder theorem* states that if a and b are relatively prime, and if c and d are any integers, then there is an integer m so that $m \equiv c \bmod a$ and $m \equiv d \bmod b$.

8. Find an n so that G_n is not cyclic or isomorphic to a product of two cyclic groups.