Notes on Expanders and Property T

Richard Evan Schwartz

July 15, 2020

1 Overview

These notes are organized around the goal of explaining why the family of Cayley graphs of $SL_3(\boldsymbol{Z}/p)$, as p ranges over all primes, is an expander family. There is nothing original in the notes, but they gather together essentially all the material you need to understand this result and some related classics. After a background chapter, here is what the notes cover:

Expander Families: I give the two definitions of an expander family of graphs, one based on eigenvalues of the graph Laplacian and one based on the Cheeger constant. I then show that the two definitions are equivalent. One of the directions is much harder than the other, and relies on the "hard Cheeger Inequality", Equation 11 below. I learned the proof by reading pp 7-8 of the paper by Fan Chung called *Four Proofs of Cheeger's Inequality and Graph Partition Algorithms*. My exposition follows the proof given there, except that I change the notation and definitions a bit, only consider the d-regular case, and add explanations.

The Main Examples: I show that the family of Cayley graphs of $SL_3(\boldsymbol{Z}/p)$, relative to the generating set formed by the elementary matrices mod p, is a 12-regular expander family. The strategy is to use a bit of representation theory to reduce the expander property to a special case of Property T for $SL_3(\boldsymbol{Z})$, and then use Yehuda Shalom's argument to establish that special case. I learned Shalom's argument from his IHES paper, *Bounded Generation and Kazhdan's Property T*. The application to expanders only requires a certain finite dimensional version of Property T, and in this case we can replace all the functional analysis with linear algebra.

Bounded Generation: Though it is not necessary for the expander result, I include a proof of the bounded generation property of $SL_3(\boldsymbol{Z})$. This is a beautiful result, and closely related to the expander property. The proof I give of the bounded generation property of $SL_3(\boldsymbol{Z})$ is taken directly from pp 197-206 the book by B. Bekka, P. de la harpe, and A. Valette called *Kazhdan's Property T*. The main difference between what I write and what is written in the book is that I include detailed proof outlines to explain the strategy of the proof in a top-down way, and I streamline the notation.

Dirichlet's Theorem: One step of the bounded generation proof is not elementary: It requires Dirichlet's classic result about primes in arithmetic progressions. For the sake of completeness, and because I saw this as a great opportunity to really learn the proof, I include a proof of this result. My proof skimps a bit on some of the routine limiting arguments, but otherwise it is all there. I learned the proof from notes by Anthony Várilly entitled *Dirichlet's Theorem on Primes in Arithmetic Progressions*. Sometimes my exposition follows Várilly's notes and sometimes I reorganize and simplify things. One step of the proof uses the analytic continuation of the Riemann ζ-function, Lemma 6.11. I learned this proof from Harvard course notes of Noam Elkies.

So, why read these notes? I think that these notes are more accessible than any of the sources above. Also, as I said, I tried go "all the way to the bottom" and include essentially everything that you need to understand the theorems. These notes are a record of the odyssey I took in order to learn all this material. Any mistakes I made along the way are my fault, and certainly not the fault of the sources I used.

I think that you could read these notes provided that you have had a semester course in each of linear algebra, abstract algebra, and (for Dirichlet's Theorem) complex analysis. You should be able to tell from the background chapter about how much mathematics is being assumed.

This summer, Jasper Liu, an undergrad at Brown, has been these reading these notes. I want to thank him for pointing out numerous typos and other glitches.

2 Background

2.1 The Hermitian Form

Let $\boldsymbol{R}$ and $\boldsymbol{C}$ denote the real and complex numbers respectively. Usually we are interested in $\boldsymbol{R}$ but we gain perspective, and can prove things more easily, by going into $\boldsymbol{C}$. The following operation on $\boldsymbol{C}^n \times \boldsymbol{C}^n$ extends the usual dot product:

$$\langle V, W\rangle = V \cdot \overline{W} \tag{1}$$

Here $\overline{W}$ is the conjugate vector. If $W = (w_1, ..., w_n)$ then $\overline{W} = (\overline{w}_1, ..., \overline{w}_n)$. The gadget $\langle , \rangle$ is called the *standard Hermitian Form*. It satisfies the following properties:

- If V and W are real then $\langle V, W\rangle = V \cdot W$.
- $\langle V_1 + V_2, W\rangle = \langle V_1, W\rangle + \langle V_2, W\rangle$.
- $\langle V, W_1 + W_2\rangle = \langle V, W_1\rangle + \langle V, W_2\rangle$.
- $\langle \lambda V, W\rangle = \lambda\langle V, W\rangle$.
- $\langle V, \lambda W\rangle = \overline{\lambda}\langle V, W\rangle$.
- $\langle W, V\rangle = \overline{\langle V, W\rangle}$.

We also note that the properties imply that $\langle V, V\rangle \geq 0$ with equality iff $V = 0$. We write $\|V\| = \sqrt{\langle V, V\rangle}$. The quantity $\|V - W\|$ measures the Euclidean distance in $\boldsymbol{C}^n$ between V and W. There is one additional property we will use. Let M^t denote the transpose of a possibly non-square matrix and let M^* denote the entry-wise conjugate of M^t. Then, assuming the multiplications all make sense,

$$\langle M(V), W\rangle = \langle V, M^*(W)\rangle. \tag{2}$$

2.2 Unitary Transformations

Let $T : \boldsymbol{C}^n \to \boldsymbol{C}^n$ be some complex linear transformation. An *eigenvector* of T is some nonzero vector V such that $T(V) = \lambda V$ for some complex λ, called the *eigenvalue*. The eigenvalues of T are the roots of the characteristic polynomial $\det(T - \lambda I)$, where I is the identity matrix. There are always n complex roots, counting multiplicity.

T is *unitary* if $\langle T(V), T(W)\rangle = \langle V, W\rangle$ for all $V, W \in \boldsymbol{C}^n$. By Equation 2 this is the same as saying that $T^{-1} = T^*$ when T is treated as a matrix. All the eigenvalues of T are unit complex numbers.

Lemma 2.1 *A unitary T has an orthonormal basis of eigenvectors.*

Proof: Some unitary transformations have all unequal eigenvalues, and these eigenvalues vary algebraically. So, any unitary transformation can be perturbed an arbitrary small amount so that it has all unequal eigenvalues. By continuity, it suffices to prove our result in this case. If V and W are two eigenvectors corresponding to distinct eigenvalues λ and μ then

$$\lambda\langle V, W\rangle = \langle T(V), W\rangle = \langle V, T^*(W)\rangle = \langle V, T^{-1}(W)\rangle =^* \langle V, \overline{\mu} W\rangle = \mu\langle V, W\rangle.$$

The starred equality uses the fact that $\mu^{-1} = \overline{\mu}$. Thus, if we take unit vectors corresponding to all the eigenvalues, we get an orthonormal basis. ♠

Lemma 2.2 *Suppose that A and B are commuting unitary transformations, meaning that $AB = BA$. Then A and B have a simultaneous orthonormal basis of eigenvectors.*

Proof: Lemma 2.1 also works for a unitary transformation $T : X \to X$ where X is a complex subspace of $\boldsymbol{C}^n$. If (λ, V) eigenvalue-eigenvector pair for A, then $\lambda B(V) = B(\lambda V) = BA(V) = AB(V)$. Hence $B(V)$ is also a λ-eigenvector for A. Hence B preserves A-eigenspaces. So setting X equal to one of these A-eigenspaces, X has an orthonormal basis of B-eigenvectors. But this is also an orthonormal basis of A-eigenvectors. Now we take the union of these sub-bases, as X ranges over all A-eigenspaces. ♠

Lemma 2.3 *Suppose $\{T_i\}$ is a collection of n unitary transformations with the property that $\|T_i(V) - V\| < \epsilon_i$ for all i. Let T be the product $T_1, ...T_n$. We have $\|T(V) - V\| < \sum_{i=1}^n \epsilon_i$.*

Proof: By induction it suffices to consider the case $n = 2$. By the triangle inequality

$$\|T_1T_2(V) - V\| \leq \|T_1(T_2(V)) - T_1(V)\| + \|T_1(V) - V\| =^*$$
$$\|T_2(V) - V\| + \|T_1(V) - V\| < \epsilon_1 + \epsilon_2.$$

The starred equality comes from the fact that T_2 is an isometry of $\boldsymbol{C}^n$. ♠

2.3 Real Symmetric Matrices

An $n \times n$ matrix M is called *Hermitian* if $M = M^*$. When the entries of M are real, we have $M = M^t$ and we call M *real symmetric*. We state the results here for real symmetric matrices, but (if you care) you should be able to see how the first two results generalize to the Hermitian case.

Lemma 2.4 *M has real eigenvalues. If, additionally, $M = X^tX$, then all the eigenvalues are non-negative.*

Proof: If λ is an eigenvalue of M and $V \in \boldsymbol{C}^n$ is the corresponding eigenvector, then using Equation 2 and the fact that $M = M^t$, we have

$$\lambda\langle V, V\rangle = \langle M(V), V\rangle = \langle V, M(V)\rangle = \langle V, \lambda V\rangle = \overline{\lambda}\langle V, V\rangle.$$

Hence $\lambda = \overline{\lambda}$. This forces λ to be real. If $M = X^tX$ then

$$\lambda\langle V, V\rangle = \langle M(V), V\rangle = \langle X^tX(V), V\rangle = \langle X(V), X(V)\rangle \geq 0.$$

This completes the proof. ♠

Lemma 2.5 *Eigenvectors corresponding to distinct eigenvalues of M are orthogonal.*

Proof: If V_1 and V_2 correspond to distinct eigenvalues λ_1 and λ_2 then

$$\lambda_1\langle V_1, V_2\rangle = \langle M(V_1), V_2\rangle = \langle V_1, M(V_2)\rangle = \lambda_2\langle V_1, V_2\rangle.$$

Since at most one λ_i is zero, this forces $\langle V_1, V_2\rangle = 0$. ♠

Lemma 2.6 *Let $\boldsymbol{C}_0^n$ be the subspace of $\boldsymbol{C}^n$ consisting of vectors whose coordinates sum to 0. If $M(\boldsymbol{C}_0^n) = \boldsymbol{C}_0^n$ and all the pairs (λ, V) with $V \in \boldsymbol{C}_0^n$ have nonzero λ, then $\boldsymbol{C}_0^n$ has an orthonormal basis of real eigenvectors.*

Proof: The dimension of $\boldsymbol{C}_0^n$ is $n-1$, and we can identify it with a copy of $\boldsymbol{C}^{n-1}$. In general, $\boldsymbol{C}_0^n$ splits as a direct sum of eigenspaces, and within each one, we choose any orthonormal basis. We can replace each basis vector ξ by one of $i\xi$ or $(\xi + \overline{\xi})/\|\xi + \overline{\xi}|$. One of these is a unit real vector with the same eigenvalue. ♠

2.4 The Graph Laplacian

Let G be a connected graph. Let $\boldsymbol{V}$ denote the vector space of complex valued functions defined on the set of vertices of G. Let $\boldsymbol{V}_0$ denote the set of such functions which sum to 0. We usually identify $\boldsymbol{V}$ with $\boldsymbol{C}^n$ and $\boldsymbol{V}_0$ with $\boldsymbol{C}_0^n$. Here $n = |G|$, the number of vertices of G.

The graph Laplacian is a linear transformation $L : \boldsymbol{V} \to \boldsymbol{V}$ given by

$$Lf(v) = D_v f(v) - \sum_{w \sim v} f(w). \tag{3}$$

Here D_v is the degree of v and $w \sim v$ means that w is adjacent to v. The constant functions are eigenvectors of L having eigenvalue 0.

Lemma 2.7 *The only eigenfunctions corresponding to the value* 0 *are the constant functions.*

Proof: If $L(f) = 0$ it means that the value of f at each vertex is the average of the neighbors. But if f is not constant, then f has some maximum. This forces f to have the same value at all adjacent vertices, and so on. Hence f is constant. ♠

The matrix for L is the degree matrix minus the adjacency matrix. Therefore, L is symmetric. The constant functions are orthogonal to $\boldsymbol{V}_0$. Hence $\boldsymbol{V}_0$ has an orthonormal basis of real eigenvectors corresponding to nonzero real eigenvalues of L.

Lemma 2.8 *The nonzero eigenvalues of* L *are positive.*

Proof: Label the vertices of G with numbers from 1 to n. Label the edges of G with numbers 1 to N. The *incidence matrix* M is given by

- $M_{ij} = +1$ if vertex i is incident to edge j and the other vertex incident to j has a higher label.
- $M_{ij} = -1$ if vertex i is incident to edge j and the other vertex incident to j has a lower label.
- Otherwise $M_{ij} = 0$.

We have $L = M^t M$. From this factorization and Lemma 2.4, the $n-1$ nonzero eigenvalues of L are all positive. ♠

2.5 Groups and Graphs

Let Γ be a group. A *generating set* for Γ is a finite list $S = \{g_1, ..., g_d\}$ of elements such that every element of Γ is some product of these elements. We always take S to be *symmetric*, which means that $g \in S$ iff $g^{-1} \in S$.

Example: For any ring R, the group $SL_3(R)$ is the group of 3×3 determinant 1 matrices with elements in R. The group $SL_3(\boldsymbol{Z})$ is generated by the 12 elementary matrices. These are the matrices which differ from the identity matrix by a single nonzero entry which is either $+1$ or -1. The finite group $SL_3(\boldsymbol{Z}/p)$, is generated by the elementary matrices mod p. The homomorphism $\pi_p : SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p)$ reduces the matrix entries mod p.

The input to a Cayley graph is a pair (G, S) where G is a group and S is a symmetric generating set for G. The *Cayley graph* $G(\Gamma, S)$ is the graph whose vertices are elements of G, and whose edges join g to gh for each $h \in S$. This is a regular graph of degree $d = |S|$.

An *action* of a group Γ on a set Σ is a homomorphism from Γ into the group of permutations of Σ. In particular, there are two actions of Γ on the set of vertices of the Cayley graph $G = G(\Gamma, S)$: the *left action* and the *right action*. They come from left and right multiplication respectively. Given $h \in \Gamma$, we have

$$L_h(g) = hg, \qquad R_h(g) = gh^{-1}. \tag{4}$$

Note $L_h(gs) = hgs = L_h(g)s$. This means that L_h maps edges of G to edges G. The left action is a graph automorphism.

Notwithstanding this nice structure, we only care about the right action. First, the reason for the inverse in formula for the right action is that

$$R_{h_1h_2}(g) = g(h_1h_2)^{-1} = gh_2^{-1}h_1^{-1} = R_{h_1} \circ R_{h_2}(g).$$

The right action is not a graph automorphism, but it is nicer in some ways. It can be described as "follow the arrows". The edges of G are labeled by the generators. If h is a generator of Γ then R_h tells each vertex to move along the directed edge labeled by h. In general, if h is some word in the generators, then h defines a directed walk on the graph, and R_h tells each vertex to perform that directed walk. The significance of this is as follows: If we have some set Σ, then each generator of the group maps vertices of G near $\partial\Sigma$ to vertices of G near $\partial\Sigma$. Here $\partial\Sigma$ is the set of edges having exactly one endpoint in Σ.

2.6 The Right Regular Representation

Let U_n denote the group of unitary transformations of $\boldsymbol{C}^n$. Let Γ be a group. A homomorphism $\rho : \Gamma \to U_n$ is called a *unitary representation*. Here we give the main example of interest.

When Γ is a finite group, we identify $\boldsymbol{C}^n$ with the vector space of complex valued functions on the vertices of $G = G(\Gamma, S)$ which sum to 0. Given any $h \in \Gamma$ we define $\rho_h : \boldsymbol{C}^n \to \boldsymbol{C}^n$ as follows:

$$\rho_h f = f \circ R_h. \tag{5}$$

Tha value of $\rho_h f$ on v equals the value of f on $R_h(v)$. This gives us a homomorphism $\rho : G \to U_n$.

To see that $\rho(G) \subset U_n$, we need to see that ρ_h is a unitary transformation. This is true just because R_h is just permuting the vertices of G_p. Hence

$$\langle \rho_h(f), \rho_h(g) \rangle = \sum_{v \in V(G)} f(vh^{-1})\overline{g}(vh^{-1}) = \sum_{w \in V(G)} f(w)\overline{g}(w) = \langle f, g \rangle.$$

The representation $\rho : G \to U_n$ just constructed is called the *right regular representation*.

Starting with a unitary representation of a finite group Γ, as above, we can sometimes get unitary representations a larger (and possibly infinite) group $\widehat{\Gamma}$ as follows: If we have a homomorphism $\pi : \widehat{\Gamma} \to \Gamma$, then the composition $\rho \circ \pi$ is a unitary representation of $\widehat{\Gamma}$. We will produce unitary representations of $SL_3(\boldsymbol{Z})$ this way, composing the map $\pi_p : SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p)$ with the right regular representation of $SL_3(\boldsymbol{Z}/p)$.

2.7 Kazhdan's Property T

Let $\rho : SL_3(\boldsymbol{Z}) \to U_n$ be a unitary representation. We say that ρ is *ϵ-small* if there is some unit vector Θ such that $\|g(\Theta) - \Theta\| < \epsilon$ for each elementary matrix g. We are talking here about 12 conditions, one per elementary matrix. Relatedly, we say that ρ has a *fixed unit vector* if there is some unit vector which is fixed by $\rho(g)$ for all $g \in SL_3(\boldsymbol{Z})$. We will prove the following result:

Theorem 2.9 *There is some constant $\epsilon_0 > 0$, independent of ρ and n, with the following property. If ρ is ϵ_0-small then ρ has a fixed unit vector.*

Remark: Theorem 2.9 is a special case of what is known as *Kazhdan's Property T* for $SL_3(\boldsymbol{Z})$. The statement is the same, but $\boldsymbol{C}^n$ is allowed to be an arbitrary Hilbert space. In this case, the standard Hermitian form on $\boldsymbol{C}^n$ is replaced by a related gadget on the Hilbert space. The proof in the general case is almost the same, except for one step which requires functional analysis in place of linear algebra. See the remark at the end of §4.4.

One step in the proof of Theorem 2.9 is difficult, the bounded generation property for $SL_3(\boldsymbol{Z})$. Now we describe an easier result which bypasses this difficulty. Let Γ_p denote the kernel of the map

$$\pi_p : SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p).$$

The group Γ_p is the subgroup of matrices in $SL_3(\boldsymbol{Z})$ congruent to the identity mod p.

Theorem 2.10 *There is some constant $\epsilon_0 > 0$, independent of ρ and n and p, with the following property. If ρ is the identity on Γ_p for some p, and ρ is ϵ_0-small, then ρ has a fixed unit vector.*

Theore 2.10 is a special case of Theorem 2.9, and it is all that we need for the expander result.

3 Expander Families

3.1 Basic Definitions

There are 2 definitions of an expander family. We give them in turn.

Eigenvalue Definition: Given the graph G, we let $\rho(G)$ equal the lowest positive eigenvalue associated to the graph Laplacian L of G. Let $|G|$ denote the number of vertices of G. Let $d(G)$ denote the max degree of G. An *expander family* is an infinite sequence $G_1, G_2, ...$ of graphs satisfying the following properties:

1. $|G_n| \to \infty$ with n.
2. There is some D such that $d(G_n) < D$ for all n.
3. There is some $\epsilon > 0$ such that $\rho(G_n) > \epsilon$ for all n.

The expander family is *d-regular* if all the graphs are d-regular for the same d. That is, all the vertices have the same degree.

Cheeger Constant Definition: Let $V(G)$ denote the set of vertices of a graph G. Given $\Sigma \subset V(G)$, let $\partial\Sigma$ denote the set of edges connecting vertices in Σ to vertices not in Σ. The *Cheeger constant* $h(G)$ is defined as follows:

$$h(G) = \min_{\Sigma \subset V(G)} \frac{|\partial\Sigma|}{|\Sigma|}. \tag{6}$$

The minimum is taken over all subsets having at most half the vertices of G. (Without this restriction the definition is useless.) The Cheeger constant is small when there is a "bottleneck", a subset Σ that is relatively large but only having a small number of edges pointing out of it. We say that a *Cheeger expander family* is a sequence of graphs $\{G_n\}$ having the first two properties of an eigenvalue expander family and also having the property that $h(G_n) > h_0$ for some $h_0 > 0$ and all n.

For our examples, the Cheeger constant definition turns out to be much easier to verify. However, the two definitions are equivalent. One goal of this chapter is to prove

Theorem 3.1 *A family $\{G_n\}$ of graphs is a Cheeger expander family if and only if it is an eigenvalue expander family.*

3.2 The Main Examples

Recall that $SL_3(\boldsymbol{Z}/p)$ is generated by the set S of 12 elementary matrices mod p. Let

$$G_p = G(SL_3(\boldsymbol{Z}/p), S) \tag{7}$$

be the associated Cayley graph. This is a 12-regular graph for any prime p. The number of vertices is $p^3(p^3-1)(p^2-1)$. We omit the proof because this fact is not used anywhere in our main argument.

Theorem 3.2 (Main) *The family $\{G_p\}$ is an expander family.*

Proof: We show that $\{G_p\}$ is a Cheeger expander family. Let $\Sigma_p \subset V(G_p)$ be a vertex set which realizes the Cheeger constant h_p of G_p. Let $f_p \in \boldsymbol{R}_0^n$ be the unit vector which is positive constant on Σ and negative constant on its complement. Here $n = |V(G_p)|$. Consider the representation $\widehat{\rho} = \rho \circ \pi$, where

$$\pi_p : SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p), \qquad \rho_p : SL_3(\boldsymbol{Z}/p) \to \boldsymbol{C}^n$$

are respectively the reduction map and the right regular representation.

Let $g \in S$ be an elementary matrix, Let $T_g = \widehat{\rho}_p(g)$. The right action of R_g moves a vertex in or out of Σ only if this vertex is incident to an edge of $\partial\Sigma$. There are at most $2|\partial\Sigma|$ such incident vertices. Hence $T_p(f_p) = f_p$ except on at most $2|\partial\Sigma|$ vertices. When $T_p(f_p)$ and f_p disagree, the absolute value $|T_p(f_p) - f_p|$ is at most twice the value of f_p on Σ. Hence

$$\|T_g(f_p) - f_p\| \leq \sqrt{8h_p}\|f_p\|_\Sigma < \sqrt{8h_p}. \tag{8}$$

Here $\|f_p\|_\Sigma < 1$ is the norm of the function which equals f_p on Σ and is 0 elsewhere. If $\{G_p\}$ is not a Cheeger expander family, we can choose p so that the expression in Equation 8 is smaller than ϵ_0 for all elementary matrices g. Here ϵ_0 is the constant from Theorem 2.10. But then $\widehat{\rho}_p$ has a fixed unit vector by Theorem 2.10.

A fixed unit vector for $\widehat{\rho}_p$ corresponds to a constant function on $V(G_p)$. But the sum of the values of this function is 0 and hence the constant would be 0. This is impossible for a unit vector, contradiction. ♠

Remark: Essentially the same proof works for the family $SL_n(\boldsymbol{Z}/p)$ for any $n > 3$. The family $SL_2(\boldsymbol{Z}/p)$ turns out to be an expander family of 4-regular graphs, the most beautiful one of all. However, the proof is much harder.

3.3 The Rayleigh Quotient

The rest of this chapter is devoted to proving Theorem 3.1. If you are satisfied with the statement that our examples form a Cheeger constant expander family, you need not read the rest of the chapter.

In this section we introduce a useful technical tool called the Rayleigh quotient. One could view this object as the bridge between the Cheeger constant and the lowest positive eigenvalue of the Laplacian.

Let L be the graph Laplacian, as above. Define for the Rayleigh quotient for any function $f : V(G) \to \boldsymbol{R}$:

$$\rho(f) = \frac{|\langle Lf, f\rangle|}{\langle f, f\rangle}. \tag{9}$$

We define the *Rayleigh Quotient of* G as

$$\rho(L) = \inf_{f \in \boldsymbol{R}_0^n - \{0\}} \rho(f). \tag{10}$$

This inf is taken over all nonzero real vectors whose coordinates sum to 0.

Lemma 3.3 *$\rho(L)$ equals the lowest positive eigenvalue of L.*

Proof: We can normalize so that $\|f\| = 1$. Let $\xi_1, ..., \xi_{n-1}$ be an orthonormal basis of real eigenvectors of $\boldsymbol{C}_0^n$. We have $f = \sum a_i \xi_i$ for some real constants $a = (a_1, ..., a_{n-1}) \in \boldsymbol{R}^{n-1}$. We compute

$$F(a) := |\langle Lf, f\rangle| = \sum a_i^2 \lambda_i.$$

We are trying to minimize F subject to the constraint $G(a) = \sum a_i^2 = 1$. Using Lagrange multipliers, we see that ∇F and ∇G are parallel at a min. This forces all the terms in a to be 0 except for the ones corresponding to a single eigenvalue. The min occurs when the special eigenvalue is the lowest one. ♠

Referring to the eigenvalue expander condition, we can summarize the previous lemma by saying that $\rho(G) = \rho(L)$.

3.4 Proof of the Equivalence

In this section we prove Theorem 3.1. We start with the easy direction.

Lemma 3.4 *If $\{G_n\}$ is an eigenvalue expander family then $\{G_n\}$ is a Cheeger expander family.*

Proof: Suppose that this is false. Suppose G is one of our graphs, and we have $h(G) = \epsilon$ for some small ϵ. Let Σ be a corresponding set which realizes this bound. Let $\Sigma' = V(G) - \Sigma$. We have $|\Sigma| \leq |\Sigma'|$. We define $f : V(G) \to \boldsymbol{R}$ so that $f \equiv |\Sigma'|$ on Σ and $f \equiv -|\Sigma|$ on Σ'. These properties imply that $f \in \boldsymbol{V}_0$. Note that $\langle f, f\rangle \geq |\Sigma||\Sigma'|^2$. Also, $L(f)$ is 0 except on vertices incident to edges of ∂E. The value of $L(f)$ on these vertices is at most $2d|\Sigma'|$, where d is the max degree. But then

$$|\langle Lf, f\rangle| \leq 4d^2|\Sigma'|^2|\partial\Sigma|,$$

because we only get nonzero contributions to the sum for the vertices incident to ∂S. Combining these estimates we have

$$\rho(G) \leq \frac{4d^2|\Sigma'|^2||\partial\Sigma|}{|\Sigma||\Sigma'|^2} = \frac{4d^2|\partial\Sigma|}{|\Sigma|} = 4d^2 h(G) = 4d^2\epsilon.$$

This makes $\rho(G)$ small as well, a contradiction. ♠

Now we prove the converse of Lemma 3.4. This is the hard direction. The proof below is taken from Fan Chung's paper, though I change the notation somewhat and add explanatory detail. Also, our definition of the Cheeger constant differs by a factor of d from the one in Chung's paper, and this accounts for a difference in the final inequality, Equation 11. These constants are not important for the overall result.

We will prove through a series of lemmas that

$$\text{lowest positive eigenvalue} = \rho(G) \geq \frac{h(G)^2}{2d}. \tag{11}$$

This is known as one of the Cheeger inequalities – the "hard one". Theorem 3.1 follows immediately from Lemma 3.4 and from this inequality. For ease of exposition will assume that G has an even number of vertices (as it does in the main examples). The odd case requires only a minor tweak.

We have $\rho(G) = \rho(g)$, where g is an eigenvector corresponding to the lowest positive eigenvalue of L. Perturbing g by a value that is as small as we like, we can arrange that g never takes the same value twice. We retain the inequality $\rho(G) > \rho(g) - \epsilon$ and we can make ϵ arbitrarily small. Let m be a value such that $g > m$ on half the vertices and $g < m$ on the other half. Let $\gamma = g - m$.

Lemma 3.5 $\rho(g) \geq \rho(\gamma)$.

Proof: Note that $L(\gamma) = L(g)$ and $\langle g, m\rangle = 0$ because $g \in \boldsymbol{R}_0^{2N}$. From this we see that

$$\langle L(\gamma), \gamma\rangle = \langle L(g), g\rangle, \qquad \langle \gamma, \gamma\rangle = \langle g, g\rangle + m^2.$$

Our inequality follows immediately from the preceding equation. ♠

For the rest of the proof, it is convenient to work with a non-negative function. We define a new function γ_+ so that $\gamma = \gamma_+$ whenever $\gamma > 0$ and otherwise $\gamma_+ = 0$. We define γ_- so that $\gamma = \gamma_+ - \gamma_-$. Replacing γ by $-\gamma$ if needed we can arrange that $\rho(\gamma_+) \leq \rho(\gamma_-)$.

Lemma 3.6 $\rho(\gamma) \geq \rho(\gamma_+)$.

Proof: We compute

$$\rho(\gamma) = \frac{\langle L(\gamma_+) - L(\gamma_-), \gamma_+ - \gamma_-\rangle}{\langle \gamma_+ - \gamma_-, \gamma_+ - \gamma_-\rangle} = \frac{\langle L(\gamma_+), \gamma_+\rangle + \langle L(\gamma_-), \gamma_-\rangle}{\langle \gamma_+, \gamma_+\rangle + \langle \gamma_-, \gamma_-\rangle} + X,$$

$$X = -\frac{\langle L(\gamma_+), \gamma_-\rangle + \langle L(\gamma_-), \gamma_+\rangle}{\langle \gamma_+, \gamma_+\rangle + \langle \gamma_-, \gamma_-\rangle}.$$

The term X is positive because $L(\gamma_\pm)$ is negative whenever $\gamma_\mp$ is positive. From this calculation, we have

$$\rho(\gamma) \geq \frac{\langle L(\gamma_+), \gamma_+\rangle + \langle L(\gamma_-), \gamma_-\rangle}{\langle \gamma_+, \gamma_+\rangle + \langle \gamma_-, \gamma_-\rangle} \geq \min \rho(\gamma_+), \rho(\gamma_-) = \rho(\gamma_+).$$

The last inequality comes from the fact that the Farey sum $(a+c)/(b+d)$ of two positive rationals a/b and c/d is at least as large as the min of the two rationals. ♠

At this point we set $f = \gamma_+$. To finish the proof, it suffices to show that $\sqrt{\rho(f)} \geq h(G)/\sqrt{2d}$. This is what we will do.

Lemma 3.7 *We have*

$$\sqrt{\rho(f)} \geq \frac{\sum_{u\sim v} |f^2(u) - f^2(v)|}{\sqrt{2d}A}. \tag{12}$$

Proof: Define

$$A = \sum_v f^2(v), \qquad B = \sum_{u \sim v} f(u)f(v). \tag{13}$$

The first sum is the sum over all vertices and the second sum is the sum over all the edges. We have

$$\langle f, f\rangle = A, \qquad \langle L(f), f\rangle = \sum_v \sum_{u\sim v} f(v)(f(v) - f(u)) = dA - 2B \geq 0.$$

To explain the last inequality we observe that

$$dA \pm 2B = \sum_{u\sim v} (f(u) \pm f(v))^2 = \|V_\pm\|^2 \geq 0. \tag{14}$$

Here $V_\pm$ is the vector obtained by stringing out the terms $f(u) \pm f(v)$ for each edge $u \sim v$.

Note that

$$A(dA + 2B) = A \times \Big(2dA - (dA - 2B)\Big) \leq 2dA^2.$$

Combining these equations, and using $dA - 2B \geq 0$, we have

$$\rho(f) = \frac{dA - 2B}{A} = \frac{(dA-2B)(dA+2B)}{A(dA+2B)} \geq \frac{(dA-2B)(dA+2B)}{2dA^2}. \tag{15}$$

Now we go back to Equation 14. Since we are only summing over each edge once in Equation 14, we can always order so that $f(u) \geq f(v)$. This makes $V_\pm$ a non-negative vector. By the Cauchy-Schwarz inequality,

$$(dA - 2B)(dA + 2B) = \|V_+\|^2\|V_-\|^2 \geq (V_+ \cdot V_-)^2 =$$

$$\Big(\sum_{u\sim v}(f(u) - f(v))(f(u) + f(v))\Big)^2 = \Big(\sum_{u\sim v} f^2(u) - f^2(v)\Big)^2. \tag{16}$$

Combining Equations 15 and 16 we get Equation 12. ♠

Now we come to the interesting part of the proof. We order the vertices of G as $v_1, ..., v_{2N}$ in such a way that $f(v_i) \geq f(v_{i+1})$ for all i. For the first half of the vertices, the order is uniquely determined and for the second half the order is essentially arbitrary. Let Σ_i denote the first i vertices of G.

Lemma 3.8 *We have*

$$\sum_{u\sim v} f^2(u) - f^2(v) = \sum_{i=1}^{N} (f^2(v_i) - f^2(v_{i+1}))|\partial\Sigma_i|. \tag{17}$$

Proof: Again, we are ordering the edges so that $f(u) > f(v)$ for all $u \sim v$. Consider an edge $v_i \sim v_j$. We order the vertices so that $i < j$. This edge belongs to the sets $\partial\Sigma_i, ..., \partial\Sigma_{j-1}$ and at the same time

$$f^2(v_i) - f^2(v_j) = (f^2(v_i) - f^2(v_{i+1})) + \cdots + (f^2(v_{j-1}) - f^2(v_j)).$$

This works because f is decreasing with i. In case $j > N$, the sum terminates at N because all the other terms are 0. Equation 17 then follows from counting the same terms in two different ways. ♠

Since $|\Sigma_i| = i$ for $i = 1, ..., N$, the definition of the Cheeger constant tells us that

$$|\partial\Sigma_i| \geq h(G) \times i, \qquad i = 1, ..., N. \tag{18}$$

Combining this with Equation 17, we get

$$\sum_{u\sim v} |f^2(u) - f^2(v)| \geq h(G) \sum_{i=1}^{N} (f^2(v_i) - f^2(v_{i+1})) \times i = h(G)A. \tag{19}$$

The last equality comes from a telescoping sum and uses the fact that $f(v_i)$ is decreasing with i and satisfies $f(v_{N+1}) = 0$. Equations 12 and 19 together immediately imply Equation 11. This completes the proof of Theorem 3.1.

4 Property T

We have proved the Main Theorem modulo Theorem 2.10. In this chapter we will prove both Theorem 2.9 and Theorem 2.10. Before we start the proof, we mention an abuse of notation we indulge in. We have a representation $\rho : SL_3(\boldsymbol{Z}) \to U_n$. Given $g \in SL_3(\boldsymbol{Z})$ and $V \in \boldsymbol{C}^n$ we will often write $g(V)$ when what we really mean is $\rho(g)(V)$.

4.1 Bounded Generation

Now only do the elementary matrices generate $SL_3(\boldsymbol{Z})$ but something much stronger is true.

Theorem 4.1 (Bounded Generation) *Every element $M \in SL_3(\boldsymbol{Z})$ can be written like this:*

$$M = g_1^{n_1} ... g_{57}^{n_{57}} \tag{20}$$

for elementary matrices $g_1, ..., g_{57} \in S$ and integers $n_1, ..., n_{57}$.

Here $n_1, ..., n_{57}$ are allowed to be arbitrarily large, and some may be 0. Thanks to some inefficiency on my part, the bound of 57 is a bit worse than the bound of 48 from the proof in the book by Bekka, de la Harpe, and Valette. Any finite bound suffices for the proof of Theorem 2.9. The bounded generation property is false for $SL_2(\boldsymbol{Z})$.

Now we describe a much weaker property that is sufficient for Theorem 2.10. This property also works for $SL_2(\boldsymbol{Z})$, though $SL_2(\boldsymbol{Z})$ does not have Property T. Recall that Γ_p is the kernel of the map $SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p)$.

Lemma 4.2 (Weak Bounded Generation) *For any prime p the following is true. Every element $M \in SL_3(\boldsymbol{Z})$ can be written like this:*

$$M = M_p g_1^{n_1} ... g_9^{n_9}, \tag{21}$$

where $M_p \in \Gamma_p$ and $g_1, ..., g_9 \in S$ and $n_1, ..., n_9 \in \boldsymbol{Z}$.

Proof: Given that the map $SL_3(\boldsymbol{Z}) \to SL_3(\boldsymbol{Z}/p)$ is a homomorphism, it suffices to show that every element of $SL_3(\boldsymbol{Z}/p)$ can be written as the product of 9 powers of elementary matrices. Multiplying on the left or the right by a power of an elementary matrix performs an operation whereby one adds a multiple of one row/column to another row/column. So we just have to show that we can reduce $M \in SL_3(\boldsymbol{Z})$ to the identity matrix using at most 9 operations. This is just Gaussian elimination; it works because $\boldsymbol{Z}/p$ is a field! After at most 3 column operations and then at most 2 row operations we arrange that the bottom row of M is $(0,0,1)$ and the right column of M is $(0,0,1)^t$. After at most 2 more column operations and at most 1 more row operation we arrange that the middle row of M is $(0,1,0)$ and the middle column is $(0,1,0)^t$. Since the determinant is 1, the top left entry is 1. ♠

4.2 The Reduction Argument

In this section we deduce Theorems 2.9 and 2.10 from the relevant bounded generation property and from Theorem 4.3 below. Let $A_\pm, B_\pm, C_\pm, D_\pm$ respectively be the elementary matrices

$$\begin{bmatrix} 1 & 0 & \pm 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \pm 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & \pm 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ \pm 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

These matrices generate a group Γ which is written as $SL_2(\boldsymbol{Z}) \ltimes \boldsymbol{Z}^2$. The matrices $A_\pm$ and $B_\pm$ generate a copy of $\boldsymbol{Z}^2$ and the matrices $C_\pm$ and $D_\pm$ generate a copy of $SL_2(\boldsymbol{Z})$. The group $\boldsymbol{Z}^2$ is normal in Γ. By construction $A_\pm$ and $B_\pm$ are contained in the normal subgroup, namely $\boldsymbol{Z}^2$, of Γ but $C_\pm$ and $D_\pm$ are not. We say that $A_\pm$ and $B_\pm$ are *normally placed*. All in all 4 elementary matrices are normally placed in a copy of $SL_2(\boldsymbol{Z}) \rtimes \boldsymbol{Z}^2$. The remaining 8 elementary matrices can be normally placed, 4 at a time, inside groups Γ' and Γ'' which are different copies of $SL_2(\boldsymbol{Z}) \rtimes \boldsymbol{Z}^2$. We state the next result for Γ but it also works for Γ' and Γ''.

Theorem 4.3 *There is some $\epsilon_1 > 0$ with the following property: Suppose we have a unitary representation $\rho : \Gamma \to U_n$ and a unit vector $\Psi \in \boldsymbol{C}^n$ such that each elementary matrix in Γ moves Ψ less than ϵ_1. Then some unit vector in $\boldsymbol{C}^n$ is fixed by every element of $\boldsymbol{Z}^2$.*

Remark: Shalom proves this result for any unitary representation w.r.t. an arbitrary Hilbert space, and he establishes that one can take $\epsilon_1 = 1/10$.

Let $\rho : SL_3(\boldsymbol{Z}) \to U_n$, be as in Theorem 2.9. We can restrict ρ to Γ.

Lemma 4.4 *Let $\epsilon_2 \in (0,1)$. Suppose $V \in \boldsymbol{C}^n$ is a unit vector such that $\|g(V)-V\| < \epsilon_1\epsilon_2/2$ for each generator g of Γ. Then $\|h(V)-V\| \leq \epsilon_2$ for all $h \in \boldsymbol{Z}^2$.*

Proof: Write $\boldsymbol{C}^n = \boldsymbol{C}^n_1 \oplus \boldsymbol{C}^n_2$, where $\boldsymbol{C}^n_1$ is the set of $\boldsymbol{Z}^2$ invariant vectors and $\boldsymbol{C}^n_2$ is the orthogonal complement. The splitting is ρ-invariant because $\boldsymbol{Z}^2$ is normal in Γ. This means that Γ also acts by unitary transformations on $\boldsymbol{C}^n_2$. Write $V = V_1 + V_2$ where $V_j \in \boldsymbol{C}^n_j$. Given any $h \in \boldsymbol{Z}^2$ we have $h(V_1) = V_1$ and $\langle h(V_i), V_j\rangle = 0$ for $i \neq j$. Hence

$$\|h(V)-V\| = \|h(V_2)-V_2\| \leq 2\|V_2\|. \tag{22}$$

Since there are no $\boldsymbol{Z}^2$-invariant vectors, Theorem 4.3 (and scaling) implies that there is some $g \in S$ such that $\|g(V_2)-V_2\| \geq \|V_2\|\epsilon_1$. But then

$$\epsilon_1\|V_2\| \leq \|g(V_2)-V_2\| \leq \|g(V)-V\| \leq (\epsilon_2\epsilon_2)/2. \tag{23}$$

This gives $\|V_2\| \leq \epsilon_2/2$. But then Equation 22 finishes the proof. ♠

Corollary 4.5 *Let ϵ_0 and Θ be as in Theorem 2.9. If $\epsilon_0 < \epsilon_1\epsilon_2/2$ then $\|M(\Theta)-\Theta\| < 57\epsilon_2$ for all $M \in SL_3(\boldsymbol{Z})$.*

Proof: Since every elementary matrix is normally placed inside one of $\Gamma, \Gamma', \Gamma''$, the previous result gives $\|g^n(\Theta)-\Theta\| < \epsilon_2$ for every elementary g and every $n \in \boldsymbol{Z}$. By Lemma 2.3, every product of 57 powers of elementary matrices moves Θ at most $57\epsilon_2$. ♠

Proof of Theorem 2.9: We choose $\epsilon_2 < 1/57$ and then set $\epsilon_0 = \epsilon_1\epsilon_2/2$. The $SL_3(\boldsymbol{Z})$-invariant set

$$Q = \bigcup_{M \in SL_3(\boldsymbol{Z})} M(\Theta)$$

is less than 1 unit from the unit vector Θ and hence has a nonzero center of mass W'. Since Q is $SL_3(\boldsymbol{Z})$-invariant, its center of mass W' is fixed by every element of $SL_3(\boldsymbol{Z})$. But then so is the unit vector $W = W'/\|W'\|$. ♠

Proof of Theorem 2.10: Corollary 4.5, with 9 in place of 57, works for a representation of the form in Theorem 2.10 because the element M_p in the Weak Bounded Generation Lemma fixes every vector, including Θ. The rest of the proof is the same. ♠

4.3 A Measure Theoretic Result

Now we turn to the proof of Theorem 4.3. The lemma in this section (suitably generalized) appears in Shalom's paper, and he attributes it to Marc Burger.

Let A be a finite list of points in $\boldsymbol{R}^2 - \{0\}$, such that each point is given a positive weight. We assume that the sum of all the weights is 1. For each subset $X \subset \boldsymbol{R}^2 - \{0\}$ we define $m(X)$ to be the number of points of A in X, counted with their weights. If $\{X_i\}$ is a finite collection of disjoint subsets of $\boldsymbol{R}^2 - \{0\}$ then $m(\bigcup X_i) = \sum m(X_i)$.

Lemma 4.6 *Let S denote the collection of 4 elementary 2×2 matrices. There is $X \subset \boldsymbol{R}^2 - \{0\}$ and $g \in S$ with $|m(gX) - m(X)| \geq 1/4$.*

Proof: The statement and conclusion of the lemma are invariant with respect to rotation by 90 degrees about the origin. So, without loss of generality we may assume that $m(X) \geq 1/2$, where X is the set of points (x, y) with $y/x \in [0, \infty)$. (This is the union of the $(++)$ and $(--)$ quadrants with the Y-axis removed.) Let $X = X_0 \cup X_1$ where X_0 is the union of points (x, y) where $y/x \in [0, 1)$ and X_1 is the complementary set. One of these sets has at most half the mass of the other. So, we have an index $i \in \{0, 1\}$ such that $m(X) - m(X_i) \geq 1/4$. But we can find some $g \in S$ such that $gX \subset X_i$. ♠

4.4 A Special Measure

Let $\rho : \boldsymbol{Z}^2 \to U_n$ be unitary. Let $A = \rho(1, 0)$ and $B = \rho(0, 1)$. Let $\{\xi\}$ be an orthonormal basis of simultaneous eigenvectors for A and B, as guaranteed by Lemma 2.2.

We think of the torus $\boldsymbol{T}$ as pairs of unit complex numbers. Given a unit vector $V \in \boldsymbol{C}^n$, we write $V = \sum_{i=1}^n a_i \xi_i$. We then introduce the weighted sum of points in $\boldsymbol{T}$ according to the following rule: The point (λ, μ) is weighted by the squared-norm of the component of V in the joint (λ, μ) eigenspace. That is, this point gets weight

$$|a_{i_1}|^2 + ... + |a_{i_k}|^2,$$

where $\xi_{i_1}, ..., \xi_{i_k}$ are the eigenvectors having A-eigenvalue λ and B-eigenvalue μ. Let ν_V be the corresponding measure.

We say that ν_V is *ϵ-concentrated* if $\nu_V(D) > 1 - \epsilon$, where D is the disk of radius ϵ about $(1,1)$. Now we set $V = \Psi$, the vector that is almost fixed by the generators in the statement of Theorem 4.3.

Lemma 4.7 *For any $\epsilon_2 > 0$, the measure ν_Ψ is ϵ_2-concentrated if ϵ_1 is small enough.*

Proof: The basis coefficients for Ψ and $A(\Psi)$ respectively are $\{a_i\}$ and $\{\lambda_i a_i\}$. We compute

$$\|A(\Psi) - \Psi\|^2 = \langle A(\Psi), A(\Psi)\rangle + \langle \Psi, \Psi\rangle - \langle A(\Psi), \Psi\rangle - \langle A, A(\Psi)\rangle =$$

$$1 + 1 - \sum_{i=1}^n |a_i|^2 \lambda_i - \sum_{i=1}^n |a_i|^2 \overline{\lambda}_i = 2 - \sum_{i=1}^n 2|a_i|^2 Re(\lambda_i).$$

A similar calculation works for B. Hence

$$\sum_{i=1}^n |a_i|^2 Re(\lambda_i) > 1 - \epsilon_1^2/2, \qquad \sum_{i=1}^n |a_i|^2 Re(\nu_i) > 1 - \epsilon_1^2/2. \tag{24}$$

The only unit complex numbers having real part near 1 are near 1. Hence, Equation 24 establishes our result. ♠

Remark: The proof of Property T for $SL_3(\boldsymbol{Z})$ in the infinite dimensional case is almost the same as in the finite dimensional case except that the measure ν_p comes from what is called a *projection valued measure* associated to ρ. This is basically constructed out of the spectral theorem for unitary operators on Hilbert space. This is a much more advanced approach.

4.5 The Group Action

So far we have been working entirely $\boldsymbol{Z}^2$, but now we remember that ρ is actually a unitary representation of the larger group $\Gamma = SL_2(\boldsymbol{Z}) \ltimes \boldsymbol{Z}^2$. Consider the generator $C = C_+$ of Γ. We have

$$CAC^{-1} = A, \qquad CBC^{-1} = AB. \tag{25}$$

Recall that we have our orthonormal basis of eigenvectors $\xi_1, ..., \xi_n$. For ease of exposition we consider the case when all eigenvalues are distinct. The general case can be treated either as a limiting case, or else by the trick or just listing out an eigenvalue k times if it appears with multiplicity k.

Note that $C(\xi_k)$ is an eigenvector for $CAC^{-1} = A$ and $CBC^{-1} = AB$. The corresponding eigenvectors are λ_k and $\lambda_k\mu_k$. But that means that we can write

$$C(\Psi) = \sum_{i=1}^{n} a_i C(\xi_i) = \sum_{i=1}^{n} a'_i \xi_i. \tag{26}$$

So, $\nu_{C(\Psi)}$ has two descriptions:

1. It gives weight a_k^2 to the point $(\lambda_k, \lambda_k^{-1}\mu_k)$.
2. It gives weight $(a'_k)^2$ to the point (λ_k, μ_k).

Item 1 needs some explanation. As we have already mentioned, $C(\xi_k)$ is an eigenvector for the transformation $A = CAC^{-1}$ with eigenvalue λ_k and an eigenvector for the transformation $AB = CBC^{-1}$ with eigenvalue $\lambda_k\mu_k$. Therefore $C(\zeta_k)$ is an eigenvector for B with eigenvalue $\lambda_k^{-1}\mu_k$.

Item 2 just says that $\nu_{C(\Psi)}$ is supported on exactly the same points as is ν_Ψ. The important point is that ν_Ψ and $\nu_{C(\Psi)}$ are supported on the same set of points. All that changes are the weights given to these points. Item 1 gives one description of how the weights are assigned and Item 2 gives a second description.

4.6 Changing Coordinates

Now we are ready to change coordinates and interpret Items 1 and 2 above in terms of the flat square torus.

We identify the torus $\boldsymbol{T}$ with the unit square $[-1/2, 1/2]^2$ with its sides glued together. The correspondence is

$$(x, y) \sim (\exp(2\pi i x), \exp(2\pi i y)). \tag{27}$$

We define $C^*(x, y) = (x, x + y)$. We have

$$(x, y) \sim (\lambda_k, \mu_k) \implies C^*(x, y) \sim (\lambda_k, \lambda_k^{-1}\mu_k).$$

Item 1 then gives us the following: If $X = \{(\lambda, \mu)\}$ is one of the points in our weighted collection, then

$$\nu_{C(\Psi)}(X) = \nu_\Psi(C^*(X)). \tag{28}$$

When X is some other point not in our weighted collection, both sides of Equation 28 are zero. Since Equation 28 holds for every point, it holds for all $X \subset \boldsymbol{T}$. The same argument applies to the other generators of $SL_2(\boldsymbol{Z})$, giving:

$$\nu_{C_\pm(\Psi)}(X) = \nu_\Psi(C^*_\pm(X)), \qquad \nu_{D_\pm(\Psi)}(X) = \nu_\Psi(D^*_\pm(X)), \tag{29}$$

provided that these maps are appropriately defined. More precisely:

- $C^*_+(x,y) = (x, x+y)$.
- $C^*_-(x,y) = (x, -x+y)$.
- $D^*_+(x,y) = (x+y, y)$.
- $D^*_-(x,y) = (x-y, y)$.

Lemma 4.8 *For any $\epsilon_2 > 0$ and any $g \in \{C_\pm, D_\pm\}$, we can arrange that $|\nu_\Psi(X) - \nu_{g\Psi}(X)| < \epsilon_2$ for all $X \subset \boldsymbol{T}$, provided that ϵ_1 is sufficiently small.*

Proof: Let Ψ and Ψ' respectively denote the list of $\{a_i\}$ and $\{a'_i\}$ corresponding to points in X. What we are saying is that we can choose ϵ_1 small enough to guarantee that $\|\Psi\|^2$ and $\|\Psi'\|^2$ are within ϵ_2 of each other. We treat the case when $\|\Psi\| \geq \|\Psi'\|$. The other case has a similar treatment. So, we want to show that $\|\Psi'\| > \|\Psi\| - \epsilon_2$. We note that

$$\|\Psi - \Psi'\|^2 \leq \sum |a_i - a'_i|^2 = \|\Psi - C(\Psi)\|^2 < \epsilon_1^2.$$

Hence $\|\Psi - \Psi'\| < \epsilon_1$. By the triangle inequality $\|\Psi'\| > \|\Psi\| - \epsilon_1$. Hence

$$\|\Psi'\|^2 > \|\Psi\|^2 - 2\epsilon_1\|\Psi\| - \epsilon_1^2 > \|\Psi\|^2 - 2\epsilon_1 - \epsilon_1^2.$$

This gives us what we want as soon as $2\epsilon_1 + \epsilon_1^2 < \epsilon_2$. ♠

Proof of Theorem 4.3: Combining Equation 29 with Lemma 4.8, we can pick our favorite $\epsilon_2 > 0$ and then choose ϵ_1 small enough so that

$$|\nu_\Psi(X) - \nu_\Psi(g^*(X))| < \epsilon_2, \tag{30}$$

for each generator $g \in \{C_\pm, D_\pm\}$ and all $X \subset \boldsymbol{T}$. Again, these are the maps from Lemma 4.6. When $\epsilon_2 < 1/4$ this looks a lot like a contradiction to

Lemma 4.6, but we are not quite there yet: The domain for ν_Ψ is $\boldsymbol{T}$ and the domain for the measure m in Lemma 4.6 is $\boldsymbol{R}^2 - \{0\}$. We now fix this.

Most of the mass of ν_Ψ is concentrated very near the origin, and moreover $C_\pm$ and $D_\pm$ both map the open square $(-1/4, 1/4)^2$ into $(-1/2, 1/2)^2$, a set which injectively sits inside $\boldsymbol{T}$. So, we slightly modify ν_Ψ by omitting the points outside the square $(-1/4, 1/4)^2$ and rescaling so that the new measure has total mass 1. Let m be the modified measure. We re-interpret m as being a measure on $\boldsymbol{R}^2$. The no-fixed-vector assumption says that actually m is a measure on $\boldsymbol{R}^2 - \{0\}$ with total mass 1.

We are only dropping off a small amount of mass, so we can still arrange

$$|m(X) - m(g^*(X))| < 1/4, \tag{31}$$

for all $X \subset \boldsymbol{R}^2 - \{0\}$ and all $g \in \{C_\pm, D_\pm\}$. Now we have a direct contradiction to Lemma 4.6. This contradiction finishes the proof. $\spadesuit$

5 The Bounded Generation Property

The proof of the Main Theorem is done, and also the proof of Theorem 2.10 is done. So, the rest of these notes are only relevant for the reader who is interested in Theorem 2.9 and/or the bounded generation property of $SL_3(\boldsymbol{Z})$. As I mentioned above, this proof is taken directly from pp 197-206 of the book by Bekka, de la Harpe, and Valette, though I organize things a bit differently and use streamlined notation. Perhaps all this will make things easier for the reader.

5.1 Background

The proof uses three results from number theory.

- **Fermat's Little Theorem:** If p is prime and does not divide a, then $a^{p-1} \equiv 1 \bmod p$. *Proof*: The multiplicative group $(\boldsymbol{Z}/p)^*$ has order $p-1$ and the order of the subgroup $\langle a \rangle$ divides the order of the group.

- **The Chinese Remainder Theorem:** If n_1, n_2 are relatively prime integers and a_1, a_2 are any integers, then there is some A having the property that $A \equiv a_i \bmod n_i$ for $i = 1, 2$. *Proof*: The map
$$\phi : \boldsymbol{Z}/(n_1 n_2) \to \boldsymbol{Z}/n_1 \oplus Z/n_2$$
is injective because if n_i divides some $m \in \boldsymbol{Z}/A$ for $i = 1, 2$ then $n_1 n_2$ divides m as well. Hence, by the pidgeonhole principle, ϕ is also surjective.

- **Dirichlet's Theorem:** If a is relatively prime to n then there is some prime p such that $p \equiv a \bmod n$. (In fact, there are infinitely many.) There is no elementary way to say why this is true, but I will give more or less the full proof in the next chapter.

The proof also uses the following linear algebra result.

Lemma 5.1 *Let $A \in SL_2(\boldsymbol{Z})$ and let s be any positive integer. There are integers f and g such that $A^s = fI + gA$.*

Proof: The general case follows from induction and the case $s = 2$. The case $s = 2$ follows from the Cayley-Hamilton Theorem, but one can also see this directly by calculating that $A^2 = -I + \text{trace}(A)A$. The trace is the sum of the diagonal entries. ♠

5.2 Proof Outline

Multiplying a 3×3 matrix M by a power of an elementary matrix on one side or the other adds an integer multiple of a row/column to some other row/column, and all such *operations* arise this way. We write $M \to_k N$ if there is a sequence of at most k operations which converts M to N, and we shorten $\to_1$ to $\to$. Note the symmetry: $M \to_k N$ then $M^t \to_k N^t$ and $M^{-1} \to_k N^{-1}$ and $N \to_k M$. Given an arbitrary $M \in SL_3(\boldsymbol{Z})$ we produce $M_1, M_2 \in SL_3(\boldsymbol{Z})$ such that $M \to_7 M_1 \to_2 M_2 \to_{48} I$. Hence $M \to_{57} I$.

Step 1: We have

$$M \to_7 M_1 = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad ad - bc = 1.$$

Replacing M_1 by one of M_1^t, M_1^{-1}, $\widetilde{M}_1$ if needed, we arrange $b \equiv 3 \bmod 4$. Here we have set $\widetilde{M}_1 = (M_1^t)^{-1}$.

Step 2: There are primes p and q such that $P = (p-1)/2$ and $Q = (q-1)/2$ are relatively prime integers, and

$$M_1 \to_2 M_2 = \begin{bmatrix} u & p & 0 \\ q & v & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

for some $u, v \in \boldsymbol{Z}$. This step uses Dirichlet's theorem. Note that there are integers k and ℓ such that $kP + \ell Q = 1$. Hence $M_2 = M_2^{kP} M_2^{\ell Q}$.

Step 3: $M_3 \to_4 I$ and $M_4 \to_4 I$, where

$$M_3 = \begin{bmatrix} u^{kP} & p & 0 \\ * & * & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad M_4 = \begin{bmatrix} v^{\ell Q} & -q & 0 \\ * & * & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

for appropriately chosen starred entries. This uses Fermat's little theorem.

Step 4: For any starred entries we have

$$M_2^{kP} \to_{20} M_3, \qquad \widetilde{M_2^{\ell Q}} \to_{20} M_4.$$

Combining Steps 3 and 4, and symmetry, $M_2^{kP} \to_{24} I$ and $M_2^{\ell Q} \to_{24} I$. Since $M_2 = M_2^{kP} M_2^{\ell Q}$ we have $M_2 \to_{48} I$.

5.3 Step 1

Write

$$M_0 = \begin{bmatrix} * & * & x' \\ * & * & y' \\ x & y & z \end{bmatrix},$$

If $x = y = x' = y' = 0$ then $z = \pm 1$. If $z = 1$ the $M_0 = M_1$. If $z = -1$ then $M_0 \to_2 M_1$. Otherwise, replacing M_0 with M_0^t if needed we can assume that not both x, y are 0.

We claim that there is some t such that one of one of the two pairs $(x, y + tz)$ or $(x + tz, y)$ are relatively prime. Assuming this claim, one operation reduces to the case when x, y are relatively prime. Then there are integers u, v such that $ux + vy = 1$. We add $(1 - z)u$ times the first column and $(1-z)v$ times the second column to the third column. This gives us $z = 1$ after 2 operations. But now 4 more operations make $x' = y' = x = y = 0$. Hence $M_0 \to_7 I$.

Now we prove our claim. If $x = 0$ then y, z are we take $t = 0$. A similar argument works if $y = 0$. So, assume that both x, y are nonzero. Let $\delta = \text{GCD}(x, y)$. If $\delta = 1$ we are done. Otherwise, write $y = \delta y'$. By the C.R.T. we can choose t so that we have $t \equiv 1 \bmod \delta$ and $t \equiv 0 \bmod y'$. By construction $x + tz \equiv 1 \bmod \delta$ and hence is relatively prime to δ. Likewise $x + tz$ is relatively prime to y' because x is relatively prime to y'. Hence $x + tz$ is relatively prime to y.

5.4 Step 2

Let a, b, c, d be as in the matrix M_1 above. Note that (b, d) are relatively prime. Since b is odd, $(b, 4d)$ are also relatively prime. By Dirichlet's Theorem there is some prime p such that $p = b + 4\mu d$ for some integer μ. Adding $4d\mu$ times the middle row of M to the top row, we get

$$M_1 \to M_1' = \begin{bmatrix} u & p & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad ud - pc = 1, \qquad p \equiv 3 \bmod 4.$$

Note that $u \neq 0$ because otherwise p would divide $\det(M_1')$. To get $M_1' \to M_2$ we just need a pair (η, q) where $q = \eta u + c$ is prime and P, Q are relatively prime. Here $P = (p - 1)/2$ and $Q = (q - 1)/2$.

Let $\delta = \mathrm{GCD}(u, p-1)$ and write $p - 1 = \delta p'$. By the C.R.T. we can choose t such that $t \equiv c \bmod u$ and $t \equiv -1 \bmod p'$.

Lemma 5.2 *The number t is relatively prime to $u(p-1)$, and $t-1$ is relatively prime to P.*

Proof: Since u and c are relatively prime and $t \equiv c \bmod u$, we see that t and u are relatively prime. By construction t and p' are relatively prime. We just need to show that t and δ are relatively prime. We have

$$1 \equiv \det(M_1') \equiv -pc \equiv (1-p)c - c \equiv -c \equiv -t \bmod \delta.$$

The last congruence comes from the fact that δ divides u and u divides $c - t$. Our calculation shows that $t \equiv -1 \bmod \delta$. Hence t and δ are relatively prime. This proves that t is relatively prime to $u(p-1)$.

Our argument shows that $t \equiv -1 \bmod r$ for all primes r dividing $p-1$. This means that $t - 1 \equiv -2 \bmod r$ for all primes r dividing $p-1$. Hence, $t-1$ is not divisible by any prime dividing $p-1$ except perhaps 2. Since $p \equiv 3 \bmod 4$, the expression $P = (p-1)/2$ is an odd integer which, from what we have just seen, must be relatively prime to $t-1$. $\spadesuit$

By Dirichlet's Theorem, there is a positive prime q such that

$$q \equiv t \bmod u(p-1).$$

Since $q-1$ and $t-1$ are congruent mod P, we see that P and $2Q$ are relatively prime. Hence P and Q are relatively prime. We have $t = mu + c$ for some integer m and $q = m'u(p-1) + t$ for some other integer m'. Setting $\eta = m + m'u(p-1)$, we see that $q = \eta u + c$.

5.5 Step 3

The proofs for M_3 and M_4 are the same, so we treat M_3. Note that u is not divisible by p because $\det(M_2) = 1$. By Fermat's theorem $u^{2P} \equiv 1 \bmod p$. Therefore, we have $u^P \equiv \pm 1 \bmod p$. We treat the two cases at the same time by a careful use of notation. There is some α such that $u^P \mp \alpha p = \pm 1$. We make the undetermined entries of M_3 equal to α and ± 1 and note that

$$M_3 = \begin{bmatrix} u^{kP} & p & 0 \\ \underline{\alpha} & \underline{\pm 1} & \underline{0} \\ 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} \pm 1 & 0 & 0 \\ \alpha & \pm 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \to_k I, \qquad k = 2 \mp 1.$$

We have underlined the (active) row involved in the operation.

5.6 Step 4

Given any 2×2 matrix A, let $\widehat{A}$ denote the 3×3 matrix

$$\widehat{A} = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Here $\mathbf{0}$ stands for a row or column of 2 zeros. The following more general result implies Step 4: Given any $s \in \boldsymbol{N}$ and any $x, y \in \boldsymbol{Z}$ so that $a^s y - bx = 1$ we have

$$\widehat{A^s} \to_{20} \widehat{B}, \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad B = \begin{bmatrix} a^s & b \\ x & y \end{bmatrix}. \tag{32}$$

As a warmup, let's consider the case $s = 1$ first. Since $ad - bc = ay - bx = 1$ we have $a(d - y) = b(c - x)$. Since a and b relatively prime, this is only possible if $x = c + ak$ for $y = d + bk$ for some k. Hence $\widehat{A} \to \widehat{B}$. Now we outline the general case.

Proof Outline: Let f, g be such that $A^s = fI + gA$, as in Lemma 5.1. We introduce the matrices

$$C = \begin{bmatrix} f + ag & bg & 0 \\ * & * & 1 \\ 1 & * & * \end{bmatrix}, \quad J = \begin{bmatrix} * & 0 & * \\ 0 & g & f \\ 1 & * & * \end{bmatrix} \quad S = \begin{bmatrix} 1 & 0 & -b \\ 0 & g & f + ga \\ 1 & * & * \end{bmatrix}. \tag{33}$$

The starred entries are unspecified but they must be such that the matrix lies in $SL_3(\boldsymbol{Z})$. When we fill in the starred entries of C in some way we call it an *C-filling*. Likewise for J and S. We establish the following results:

1. For any C-filling we have $\widehat{A^s} \to_6 C$.
2. If some S-filling gives $S \to_k I$, then $C \to_{k+1} \widehat{B}$ for some C-filling.
3. For any J-filling, we have some S-filling so that $S \to_5 J$.
4. For some J-filling we have $J \to_8 I$.

Combining Statements 3 and 4, we see that $S \to_{13} I$ for some S-filling. By Statement 2, there is some C-filling so that $C \to_{14} \widehat{A}$. By Statement 1, $\widehat{A^s} \to_6 C \to_{14} \widehat{B}$. ♠

Proof of Statement 1: For any C-filling, we have

$$C = \begin{bmatrix} f+ag & bg & 0 \\ \underline{*} & \underline{*} & \underline{1} \\ 1 & * & * \end{bmatrix} \to \begin{bmatrix} f+ag & bg & 0 \\ * & * & 1 \\ \underline{1} & \underline{*} & \underline{1} \end{bmatrix} \to \begin{bmatrix} f+ag & bg & \underline{0} \\ * & * & \underline{0} \\ 1 & * & \underline{1} \end{bmatrix} \to_2$$

$$\begin{bmatrix} \underline{f+ag} & \underline{bg} & \underline{0} \\ * & * & 0 \\ 0 & 0 & 1 \end{bmatrix} \to_1 \begin{bmatrix} f+ag & bg & 0 \\ cg & f+dg & 0 \\ 0 & 0 & 1 \end{bmatrix} = fI + gA = \widehat{A}^s.$$

The last arrow is the case $s = 1$ of the result, which we already established. ♠

Proof of Statement 2: Since A^s mod b is lower triangular, $(A^s)_{11} \equiv a^s$ mod b. At the same time $(A^s)_{11} = fa + g$. Hence $B_{11} = fa + g - bu$ for some u. Hence

$$\widehat{B} = \begin{bmatrix} fa+g-bu & \underline{b} & 0 \\ x & \underline{y} & 0 \\ 0 & \underline{0} & 1 \end{bmatrix} \to \begin{bmatrix} fa+g & b & 0 \\ x' & y & 0 \\ 0 & 0 & 1 \end{bmatrix} = C'.$$

Now we compute that

$$C'S = \begin{bmatrix} fa+g & b & 0 \\ x' & y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -b \\ 0 & g & f+ga \\ 1 & * & * \end{bmatrix} = C,$$

where C-filling depends on the S-filling. The fact that entry $(C)_{32} = 1$ comes comes the fact that $(f+ag)y - bx' = 1$, which comes $\det(C') = 1$. In summary, $C = C'S \to_k C' \to \widehat{B}$. ♠

Proof of Statement 3:

$$J = \begin{bmatrix} \underline{*} & 0 & * \\ \underline{0} & g & f \\ \underline{1} & * & * \end{bmatrix} \to \begin{bmatrix} * & 0 & * \\ 0 & g & f \\ \underline{1} & \underline{*} & \underline{1} \end{bmatrix} \to \begin{bmatrix} \underline{1} & * & * \\ \underline{0} & g & f \\ \underline{1} & * & 1 \end{bmatrix} \to$$

$$\begin{bmatrix} \underline{1} & 0 & * \\ \underline{0} & g & f \\ \underline{1} & * & 1 \end{bmatrix} \to \begin{bmatrix} 1 & \underline{0} & -b \\ 0 & \underline{g} & f \\ 1 & \underline{*} & * \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -b \\ 0 & g & f+ga \\ 1 & * & * \end{bmatrix} = S.$$

More precisely, for any J-filling, there is some S-filling so that $S \to_5 J$. ♠

Proof of Statement 4: Now we show that $J \to_8 I$ for a suitable J-filling. Note that $g^2 \equiv 1 \bmod f$. To see this congruence, note that

$$1 \equiv \det(fI + gA) \equiv \deg(gA) \equiv g^2 \det(A) \equiv g^2 \bmod f.$$

Since f divides $g^2 - 1 = (g+1)(g-1)$, there are integers $f_\pm, g_\pm$ such that $f = f_+ f_-$ and $g \pm 1 = f_\pm g_\pm$. Let

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & g & f_- \\ 0 & g_- & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} g_+ & 0 & 1 \\ -f_- & 1 & 0 \\ g & 0 & f_+ \end{bmatrix}.$$

We compute that $GH = J$ for some J-filling. Using $g - f_- g_- = 1$, we have

$$G = \begin{bmatrix} 1 & 0 & \underline{0} \\ 0 & g & \underline{f_-} \\ 0 & g_- & \underline{1} \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & f_- \\ \underline{0} & \underline{0} & \underline{1} \end{bmatrix} \to I,$$

Hence $G \to_2 I$. Using $g - f_+ g_+ = -1$ in the underlined arrow, we have

$$H = \begin{bmatrix} g_+ & \underline{0} & 1 \\ -f_- & \underline{1} & 0 \\ g & \underline{0} & f_+ \end{bmatrix} \to \begin{bmatrix} g_+ & 0 & \underline{1} \\ 0 & 1 & \underline{0} \\ g & 0 & \underline{f_+} \end{bmatrix} \underline{\to} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & f_+ \end{bmatrix} \to_4 I.$$

Hence $H \to_6 I$. Combining these two results, we have $J \to_8 I$. ♠

5.7 Discussion

What role does Dirichlet's theorem really play in all this? We could try skipping Step 2 and just use M_1 in place of M_2. The same argument gives

$$M_1^{k\phi(b)/2} \to_{24} I, \qquad M_1^{\ell\phi(c)/2} \to_{24} I.$$

for any $k, \ell \in \boldsymbol{Z}$. Here ϕ is the Euler ϕ function. Letting δ be the GCD of $\phi(b)/2$ and $\phi(c)/2$, we get $M_1^\delta \to_{48} I$. If b, c are not prime, then δ will probably be quite large, and in all cases δ will be even. So, Dirichlet's theorem lets us modify M_1 to guarantee that $\delta = 1$ for a "nearby" matrix.

I wonder if there is a completely different strategy for proving the bounded generation property that does not require Dirichlet's Theorem. To use an analogy, one common way to solve Rubik's cube is to go layer-by-layer, using complicated moves to get the last layer. But then there are completely different methods which ignore the layer structure.

6 Dirichlet's Theorem

6.1 A Warmup

Let us start by giving the classic analytic proof that there are infinitely many prime numbers. Starting from the identity

$$\frac{1}{1-\frac{1}{p}} = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots. \tag{34}$$

one sees that every term of the form $1/k$, with $k = 1, 2, 3, ...$ appears exactly once in the product of such expressions, taken over primes:

$$\sum_{k=1}^{\infty} \frac{1}{k} = \prod_p \frac{1}{1-\frac{1}{p}}. \tag{35}$$

This is the prototypical *Euler product*. If there were only finitely many primes, the right hand side would be finite. But the left hand side is infinite. This contradiction shows that there are infinitely many primes.

Some readers may feel queasy dealing with divergent series. So, for any $s > 1$ we note that we have an identity

$$\zeta(s) := \sum_{k=1}^{\infty} \frac{1}{k^s} = \prod_p \frac{1}{1-\frac{1}{p^s}}. \tag{36}$$

Both sides converge. Letting $s \to 1$, the left hand side (the Riemann ζ function) blows up. We mean that $\lim_{s\to 1} |\zeta(s)| = \infty$. If there were only finitely many primes, the right hand side could not also blow up. This way of doing things avoids dealing with divergent series.

Taking Logs: One beautiful thing about the Euler product proof is that one can squeeze a lot more information out of it. For $x \in [0, 1/2]$ we have

$$\log\left(\frac{1}{1-x}\right) \leq 2x.$$

We mean to apply this inequality to $x = 1/2, 1/3, 1/5, ...$ Taking logs of both sides of Equation 35, we get

$$\infty = \sum_p \log\left(\frac{1}{1-\frac{1}{p}}\right) \leq 2\sum_p \frac{1}{p}.$$

This shows that the sums of the reciprocals of the primes diverges, a much stronger statement than the fact that there are infinitely many.

6.2 Proof Outline

We fix n, and a relatively prime to n. Let $(\boldsymbol{Z}/n)^*$ denote the multiplicative group of residue classes mod n that are relatively prime to n. The order of $(\boldsymbol{Z}/n)^*$ is $\phi(n)$, where ϕ is the Euler ϕ function.

Dirichlet Characters: A *Dirichlet character* is a unitary representation

$$\chi : (\boldsymbol{Z}/n)^* \to U_1. \tag{37}$$

Each element $T \in U_1$ has the form $T(V) = \lambda V$ for some unit complex λ. So, more simply, a Dirichlet character is a homomorphism from $(\boldsymbol{Z}/n)^*$ into S^1, the group of unit complex numbers.

The image of a Dirichlet character is always a root of unity. The *principal Dirichlet character* is the character χ_0 whose value is identically 1. The rest we call *non-principal*. Below we establish the *orthogonality relation:*

$$\sum_{\chi \in G} \chi(p)\overline{\chi}(a) = 0 \tag{38}$$

when $p, a \in (\boldsymbol{Z}/n)^*$ are unequal members. Here G is the set of all Dirichlet characters. When $p = a$ the sum equals $\phi(n)$.

Dirichlet L-Series: Given a Dirichlet character χ, we extend χ to $\boldsymbol{N}$, the set of natural numbers, by the following rule: $\chi(m) = 0$ if $(m, n) \neq 1$ and $\chi(m) = \chi([m])$ when $(m, n) = 1$. Here (m, n) is the GCD of m and n, and $[m] \in (\boldsymbol{Z}/n)^*$ is the residue class of m mod n. First of all, Equation 38 extends as well, and says that the sum equals 0 when $p \not\equiv a$ mod n and otherwise equals $\phi(n)$. Second of all, we define, for $s \geq 1$ real:

$$L(\chi, s) = \sum_{k=1}^{\infty} \frac{\chi(s)}{n^s}. \tag{39}$$

Regardless of χ, this series is majorized by a convergent geometric series for $s > 1$, so $L(\chi, s)$ is defined for all $s > 1$. Here are three facts we need:

1. For χ_0 principal, $L(\chi_0, s)$ blows up as $s \to 1$. This is easy.

2. For χ non-principal, $L(\chi, 1)$ exists and is bounded. This is easy.

3. For χ non-principal, $L(\chi, 1) \neq 0$. This is harder.

Logs of Euler Products: The functions $L(\chi, s)$ are not necessarily real valued when $\chi \neq \chi_0$, so we have to argue that there is a well-defined way to take the log of this function. Assume this for now. Using an argument similar to the Euler product expansion in the previous section, one sees that there is some function $g(\chi, s)$ such that

$$\log L(\chi, s) = \sum_p \frac{\chi(p)}{p^s} + g(\chi, s). \tag{40}$$

This equation holds for all $s > 1$. The function $g(\chi, s)$ is continuous and defined for all $s \geq 1$.

Summing over Characters: Let $S(a, n)$ denote the set of primes congruent to a mod n. We want to see that $S(a, n)$ is an infinite set. Take $s > 1$ and consider the expression obtained by summing over all characters:

$$F_a(s) = \sum_\chi \log L(\chi, s) \times \overline{\chi}(a). \tag{41}$$

Given the 3 properties of the L-series mentioned above, all the terms of $F_a(s)$ stay bounded as $s \to 1$ except the term coming from the principle character χ_0. Hence $F_a(s)$ blows up as $s \to 1$.

From Equation 40 we have $F_a(s) = G_a(s) + H_a(s)$, where

$$G_a(s) = \sum_\chi g(\chi, s)\overline{\chi(a)}, \qquad H_a(s) = \sum_\chi \sum_p \frac{\chi(p)\overline{\chi}(a)}{p^s}. \tag{42}$$

G_a is continuous and $G_a(1)$ is finite. Hence $H_a(s)$ blows up as $s \to 1$. Interchanging the order of summation and using (the extended version of) Equation 38, we have

$$H_a(s) = \sum_p \Big(\sum_\chi \frac{\chi(p)\overline{\chi}(a)}{p^s}\Big) = \sum_{p \in S(a,n)} \frac{\phi(n)}{p_s}. \tag{43}$$

All the other terms vanish! Since $H_a(s)$ blows up as $s \to 1$ there must be infinitely many primes in $S(a, n)$. This completes the proof.

The rest of the notes are devoted to filling in the details of this outline.

6.3 Finite Abelian Groups

A number of the form p^k is called *primary* when p is prime. A finite group is called *primary* when its order is primary. A group is called *cyclic* if it is isomorphic to $\boldsymbol{Z}/N$ for some N. The result we prove in this section is part of the classification of finite abelian groups:

Theorem 6.1 *Every finite abelian group is the product of cyclic primary groups.*

This is Theorem 2.14.1 in *Topics in Algebra, 2nd Ed.* by I. N. Herstein, though the full proof there is spread out through several sections of the book. For the sake of completeness we give a proof here. Our proof is a bit short on details, and this may not the best place to learn it for the first time. Let G be a finite abelian group of order $|G|$.

Lemma 6.2 *Suppose p is prime and divides $|G|$. Then G has an element of order p.*

Proof: For $\boldsymbol{Z}/N$ the element N/p does the job. So, assume G is not cyclic. Pick any proper subgroup H of G and consider the quotient map $\phi : G \to G/H$. If p divides $|H|$ then, by induction on the order, H has an element of order p. Hence, so does G. Otherwise p divides $|G/H|$ and by induction $|G/H|$ has some element $\phi(g)$ of order p. We $g^p = a \in H$. Let $o(a)$ be the order of a. Let $y = g^{o(a)}$. Since p does not divide $o(a)$, the element y is not the identity. On the other hand, $y^p = a^{o(a)}$ is the identity. Hence y is the desired element. ♠

A *Sylow p-subgroup* of G is a primary subgroup whose order is the largest power of p that divides $|G|$.

Lemma 6.3 *G has a Sylow p-subgroup for each prime p dividing $|G|$.*

Proof: Let p^α be the highest power of p dividing $|G|$. By the preceding result, G has a subgroup H of order p. Consider $\phi : G \to G/H$. By induction G/H has a subgroup S of order $p^{\beta-1}$. But then $\phi^{-s}(S)$ is the desired subgroup of G having order p^β. ♠

Lemma 6.4 *The Sylow p-subgroup of G is unipue when G is abelian. Any Sylow subgroup intersects the product of the others only in the identity.*

Proof: Let p^α be the highest power of p dividing $|G|$. Suppose H_1 and H_2 are unequal Sylow p-subgroups. We have the formula

$$|H_1 H_2| = \frac{|H_1||H_2|}{|H_1 \cap H_2|}.$$

Here $H_1 H_2$ is the set of products of the form $h_1 h_2$ with $h_1 \in H_1$ and $h_2 \in H_2$. This set is in fact a group. If H_1 and H_2 are unequal Sylow p-subgroups, then $H_1 H_2$ is a subgroup whose order is divisible by $p^{\alpha+1}$, a contradiction.

Suppose $g \in H_1 \cap H_2 ... H_k$. Then $o(g)$ divides a power of p_1 and a power of $p_2 ... p_k$. Hence $o(g) = 1$, making g the identity. ♠

Corollary 6.5 *Every finite abelian group is the product of primary groups.*

Proof: Let $H_1, ..., H_k$ be the Sylow subgroups of G. The product mapping $\psi(h_1, ..., h_k) = h_1 ... h_k$ gives a homomorphism from $H_1 \oplus ... \oplus H_k$ to G. Both groups have the same order, and ψ is injective by the previous result. Hence ψ is an isomorphism. ♠

Lemma 6.6 *Every primary abelian group G is the product of cyclic groups.*

Proof: Let G be a counter-example of lowest order. So, G is not cyclic and G does not factor into smaller groups. Let p be the relevant prime and let $g \in G$ be an element of maximal order. Let $H = \langle g \rangle$, the subgroup generated by g. Consider $\phi : G \to G/H$. If G/H is not cyclic then by induction $G = S_1 \oplus S_1$. But then $G = \phi^{-1}(S_1) \oplus \phi^{-1}(S_2)$, a contradiction.

Hence G/H is cyclic. Let $p^m = |G/H|$. Let $a \in G$ be an element such that $\phi(a)$ generates G/H. If a^{p^m} is the identity we are done, because then G splits as $H \oplus \langle a \rangle$. Otherwise we have $a^{p^m} = g^{p^n}$ for some n.

If $|G/H| > |H|$ then a is an element of G having larger order than g, which is a contradiction. Hence $|G| \geq p^{2m}$. The element a has order at least $p^m \times (p^{2m}/p^n)$, which forces $n \geq m$. Let us say $n = m + k$. Then $a^{p^m} = (g^k)^{p^m}$. Set $b = ag^{-k}$. By construction $\phi(b)$ generates G/H and b^{ϕ^m} is the identity. Hence $G = H \oplus \langle b \rangle$. ♠

6.4 Properties of Characters

A kth root of unity ω is *primitive* if $k > 0$ is the smallest power such that $\omega^k = 1$.

Lemma 6.7 *Suppose that $n_1, ..., n_k$ are relatively prime and $\omega_1, ..., \omega_k$ are such that ω_j is a primitive (n_j)th root of unity for $j = 1, ..., k$. Then the product $\omega = \omega_1 ... \omega_k$ is a primitive $(n_1 ... n_k)$th root of unity.*

Proof: By induction, it suffices to prove the case $n = 2$. If $\omega_1\omega_2$ is not a primitive $(n_1 n_2)$th root of unity, then after reordering we can find some $d > 1$ such that d divides n_2 and $(\omega_1\omega_2)^{n_1 n_2/d} = 1$. But then $\omega_3^{n_2/d} = 1$, where $\omega_3 = \omega_2^{n_1}$ is a primitive (n_2)th root of unity. This is a contradiction. ♠

The group $(\boldsymbol{Z}/n)^*$ is a finite abelian group and so is the set of Dirichlet characters. The character group law is $(\chi_1\chi_2)(c) = \chi_1(c)\chi_2(c)$.

Lemma 6.8 *Let f be the order of p in $(\boldsymbol{Z}/n)^*$. Let $g = \phi(n)/f$. Then there are exactly g Dirichlet characters which map p to any given fth root of unity.*

Proof: By Theorem 6.1, we have $(\boldsymbol{Z}/n)^* = C_1 \oplus ... \oplus C_\ell$, where C_i is cyclic and primary. We write $p = (p_1, ..., p_\ell)$ with $p_i \in C_i$. Let a_i be a generator of C_i. We specify a character χ uniquely by $\chi(a_1), ..., \chi(a_\ell)$, and $\chi(a_j)$ can be any $|C_j|$th root of unity. Hence there are $\phi(n)$ characters.

Let J_p denote the group of fth roots of unity and let J_p' denote the set of images of p under the maps given by the characters. Since the characters form a group, J_p' is a subgroup of J_p, and moreover, each member of J_p' is the image of p under $\phi(n)/|J_p'|$ characters. To finish the proof, we just need to show that J_p' contains a primitive fth root of unity, for then $J_p' = J_p$.

For each maximal primary number q^α dividing f, there is an index j such that the order of p_j is q^α. We define a character χ by setting $\chi(a_h) = 1$ if $h \neq j$ and $\chi(a_j) = \exp(2\pi i j/|C_j|)$. By construction

$$\chi(p) = \chi\Big(a_j^{\frac{|C_j|}{q^\alpha}}\Big) = \exp(2\pi i/q^\alpha)$$

is a primitive (q^α)th root of unity. But then J_p contains the product over all such roots of unity corresponding to maximal primary powers dividing f. By the previous lemma, this product is a primitive fth root of unity. ♠

6.5 The Orthogonality Relation

In this section we prove Equation 38 and a related result.

Lemma 6.9 *Let $k \in (\boldsymbol{Z}/n)^*$. If $k \neq 1$ then $\sum_\chi \chi(k) = 0$. The sum is taken over all Dirichlet characters.*

Proof: By Lemma 6.8, there is a Dirichlet character ψ such that $\psi(k) \neq 1$. Let Σ be the sum in question. We have

$$\Sigma = \sum_\chi \chi(k) = \sum_\chi (\psi\chi)(k) = \psi(k) \sum_\chi \chi(k) = \psi(k)\Sigma.$$

Since $\psi(k) \neq 1$ we have $\Sigma = 0$. ♠

Call a pair (a, p) *good* if Equation 38 holds for (a, p). Since $\chi(1) = 1$ for all characters, the previous lemma says that $(1, k)$ is good as long as $k \neq 1$. Setting $k = a^{-1}p$, we see that $(1, a^{-1}p)$ is good. Now we compute that

$$\sum_\chi \chi(a)\overline{\chi}(p) = \sum_\chi \chi(a)\chi(a^{-1})\overline{\chi}(a^{-1})\overline{\chi}(p) = \sum_\chi \chi(1)\overline{\chi}(a^{-1}p) = 0.$$

The starred equality uses the fact that $\chi(a^{-1})\overline{\chi}(a^{-1}) = 1$ for all choices. This establishes the orthogonality relation.

We close this section with a related result which will useful in the next section.

Lemma 6.10 *Let χ be a Dirichlet character. If $\chi \neq \chi_0$, then $\sum_h \chi(h) = 0$. This sum is taken over all $h \in (\boldsymbol{Z}/n)^*$.*

Proof: This has the same kind of proof as Lemma 6.9. There is some $k \in (\boldsymbol{Z}/n)^*$ such that $\chi(k) \neq 1$. But then

$$\Sigma = \sum_h \chi(h) = \sum_h \chi(hk) = \chi(k) \sum_h \chi(h) = \chi(k)\Sigma.$$

This forces $\sum_h \chi(h) = 0$. ♠

Lemma 6.10 is false for χ_0. In this case, the sum is $\phi(n)$. Note that Lemma 6.9 and Lemma 6.10 make "dual" statements. In fact the group of characters on $(\boldsymbol{Z}/n)^*$ is known as the dual group of $(\boldsymbol{Z}/n)^*$.

6.6 Complex Analytic Functions

Given an open set $U \subset \boldsymbol{C}$, a function $f : U \to \boldsymbol{C}$ is *complex analytic* if f equals a convergent power series in a neighborhood of each $s_0 \in U$:

$$f(s) = \sum_{k=0}^{\infty} a_k (s - s_0)^k. \tag{44}$$

More usually, a complex analytic function is defined as one which has a complex derivative, and then the above property is derived as a consequence. We mention 4 properties of complex analytic functions:

Algebraic Property: If f and g are complex analytic in U then so are $f+g$ and fg. To see this, just expand out the series.

Cancellation Property: If $f(s_0) = 0$ then $g(s) = f(s)/(s - s_0)$ extends to be a complex analytic function in U. One can see this readily from the series definition, because we would have $a_0 = 0$ in Equation 44.

Taylor Series Property: The function f is infinitely differentiable, and the derivative of f is just the term-by-term differentiation of the series. Hence, the series in Equation 44 is just the Taylor series expansion of f at s_0. If $\Delta \subset U$ is any disk centered at s_0 then Equation 44 is valid throughout U. This last statement basically boils down to the series convergence test.

Convergence Property: Suppose $\{f_n\}$ is a sequence of complex analytic functions on U which converges uniformly on closed disks. Then $\lim_{n\to\infty} f_n$ is again a complex analytic function on U. By *uniform convergence* we mean that for each $\epsilon > 0$ and each closed disk $\Delta \subset U$ there is some N such that $\sup_{s\in\Delta} |f_m(s) - f_n(s)| < \epsilon$ when $m, n > N$.

Now we mention the important special case for us. For any positive integer n and $s = x + iy$ we interpret

$$n^s = n^x(\exp(i \log y)).$$

The exponential function $\exp(z)$ is complex analytic in the entire complex plane, and hence so are the functions $s \to n^s$ and $s \to 1/n^s$.

6.7 Convergence Properties

Let H_x denote the open halfplane consisting of s with $\mathrm{Re}(s) > x$.

We first deal with the principal character χ_0. We have

$$L(\chi_0, s) = \sum_{(k,n)=1} \frac{1}{k^s}.$$

Each summand appears once in the product of all terms in Equation 34 where p is prime and relatively prime to n. Hence

$$L(\chi_0, s) = \prod_{(p,n)=1} \frac{1}{1-\frac{1}{p^s}} = \Big(\prod_{p|n} 1 - \frac{1}{p^s}\Big) \times \zeta(s). \tag{45}$$

The first factor on the right hand side, a product over the primes dividing n, is a finite product. Moreover, $\zeta(s)$ forms a convergent geometric series for $s > 1$. By the Convergence Property for complex analytic functions, $L(\chi_0, s)$ is complex analytic in H_1 and blows up as $s \to 1$.

Now suppose that χ is some other character which is not the principal character. We will show that $L(\chi, s)$ is finite for all $s > 0$. By Lemma 6.10,

$$\sum_{k=1}^{n} \chi(tn+k) = 0, \qquad t = 0, 1, 2, 3, ... \tag{46}$$

We write

$$L(\chi, s) = \sum_{t=0}^{\infty} A_t, \qquad A_t = \sum_{k=1}^{n} \frac{\chi(tn+k)}{(tn+k)^s}. \tag{47}$$

By Equation 46,

$$A_t = \sum_{k=1}^{n} \chi(tn+k) \times \Big(\frac{1}{(tn)^s} - \frac{1}{(tn+k)^s}\Big).$$

Hence

$$|A_t| < \frac{n}{n^s} \times \Big|\frac{1}{t^s} - \frac{1}{(t+1)^s}\Big| < sn^{s-1} \times \frac{1}{t^{s+1}}.$$

The last inequality comes from the fact that $\frac{d}{dt} t^{-s} = -st^{-1-s}$, and this function is decreasing in absolute value as t increases. Since $\sum 1/t^{s+1}$ converges for all $s > 0$, so does the series for $L(\chi, s)$. By the Convergence Property for complex analytic functions, $L(\chi, s)$ is complex analytic in H_0.

We have yet to show that $L(\chi, 1) \neq 0$.

6.8 Logs of Euler Products

The proof here imitates what we did in the warm-up section on the Riemann ζ function. Just like Equation 34 we have

$$\frac{1}{1-\frac{\chi(p)}{p^s}} = 1 + \frac{\chi(p)}{p^s} + \frac{\chi^2(p)}{p^{2s}} + \frac{\chi^3(p)}{p^{3s}} + \cdots . \tag{48}$$

Using the multiplicative nature of χ and routine limiting arguments, we get

$$L(\chi, s) = \prod_p \frac{1}{1-\frac{\chi(p)}{p^s}}, \tag{49}$$

for all $s > 1$. Since $L(\chi, s)$ is not necessarily real valued, we have to specify how we take the logs. Integrating the equation

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i,$$

we get

$$\log\left(\frac{1}{1-x}\right) = \sum_{k=1}^{\infty} \frac{x^k}{k}. \tag{50}$$

This converges for any complex x with $|x| > 1$. In particular, using the fact that the log of the product is the sum of the logs and making suitable limiting arguments, we have

$$\log L(\chi, s) = \sum_p \frac{\chi(p)}{p^s} + g(\chi, s), \qquad g(\chi, s) = \sum_p \sum_{k=2}^{\infty} \frac{\chi(p^k)}{kp^{ks}}. \tag{51}$$

To show that $g(\chi, s)$ is well defined and bounded, note that for all $s \geq 1$:

$$\Big|\sum_{k=2}^{\infty} \frac{\chi(p^k)}{kp^{ks}}\Big| \leq \sum_{k=2}^{\infty} \frac{1}{p^k} = \frac{1}{p^2} \times \frac{1}{1-\frac{1}{p}} \leq \frac{2}{p^2}.$$

In the last equation we used $p \geq 2$. So when we sum over p we get a convergent series for all $s \geq 1$. In fact $g(\chi, s)$ is complex analytic in H_0, but we don't need to know this.

6.9 The Nonvanishing Property: Outline

We will assume that $L(\chi, 1) = 0$ for some non-principle Dirichlet character χ and derive a contradiction.

Step 1: We show that the product

$$F(s) = \prod_{\chi} L(\chi, s), \tag{52}$$

taken over all Dirichlet characters, is complex analytic in H_0.

Step 2: We show that in H_1 we have

$$F(s) = \sum_{j=1}^{\infty} \frac{a_j}{n^s}, \qquad a_j \geq 0. \tag{53}$$

Step 3: We produce another series

$$G(s) = \sum_{j=1}^{\infty} \frac{b_j}{n^s}, \qquad 0 \leq b_j \leq a_j \tag{54}$$

which diverges at $s = 1/\phi(n)$.

Step 4: We show that Equation 53 actually holds on H_0. By the comparison test, G converges at $s = 1/\phi(n)$. This contradiction finishes the proof.

Remark: Steps 3 and 4 almost seem unnecessary, but not quite. If we had two nonprincipal characters χ_1, χ_2 such that $L(\chi_j, 1) = 0$ for $j = 1, 2$ then $F(1) = 0$. Step 2 then would say that F is identically 0. So, Steps 1 and 2 already say that the bad event $L(\chi, 1) = 0$ can happen at most once. Note also that the non-real characters come in pairs. If χ is a character then so is the conjugate $\overline{\chi}$. Moreover, $L(\chi, 1) = 0$ iff $L(\overline{\chi}, 1) = 0$. Thus, for proving Dirichlet's Theorem, the only case we have to worry about is when χ is real. In the special case when n is prime, there is only non-real non-principle Dirichlet character, namely the Legendre symbol: $\chi(k) = \pm 1$ with the positive sign taken if and only if the polynomial $x^2 = k$ has a root mod n. This is one connection between Dirichlet's Theorem and matters related to e.g. quadratic reciprocity.

6.10 Nonvanishing Property: Details

Proof of Step 1: Define

$$Z(s) = \zeta(s) - \frac{1}{1-s}. \tag{55}$$

Lemma 6.11 *$Z(s)$ is complex analytic in H_0.*

Proof: We have

$$Z(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} - \int_{t=1}^{\infty} \frac{dt}{t^s} = \sum_{n=1}^{\infty} Z_n(s), \qquad Z_n(s) = \int_n^{n+1} \Big(\frac{1}{n^s} - \frac{1}{t^s}\Big) dt. \tag{56}$$

We also have

$$|Z_n(s)| \leq \max_{u \in [n,n+1]} \left|\frac{d}{du}\frac{1}{u^s}\right| \leq \frac{s}{n^{x+1}}.$$

Thus, the sum defining $Z(s)$ converges uniformly on closed disks contained in H_0. So, by the Convergence Property for complex analytic functions, the limit of such a sum is also complex analytic in H_0. ♠

If follows immediately from Equation 45 that

$$L(\chi_0, s) = A(s) + B(s)\frac{1}{1-s},$$

Where A and B are complex analytic in H_0. We have already seen that $L(\chi, s)$ is complex analytic in H_0 when χ is not principle.

Suppose that $L(\chi_1, 1) = 0$. By the Cancellation Property and the Algebraic Property, the function

$$C(s) = L(\chi_0, s)L(\chi_1, s) = A(s)L(\chi_1, s) + B(s)\frac{L(\chi_1, s)}{1-s}$$

is complex analytic in H_0. Finally,

$$F(s) = C(s) \times \prod_{\chi \neq \chi_0, \chi_1} L(\chi, s),$$

is the product of two functions which are complex analytic in H_0. Hence, $F(s)$ is complex analytic in H_0.

Proof of Step 2: Let p be a prime that does not divide n. Let $f(p), g(p), J_p$ be as in Lemma 6.8. Using the Euler product expansion (and closely following Várilly's notes) we have

$$F(s) = \prod_{\chi} \prod_{(p,n)=1} \frac{1}{1 - \frac{\chi(p)}{p^s}} = \prod_{(p,n)=1} \frac{1}{A(p,s)}, \qquad A(p,s) = \prod_{\chi} \Big(1 - \frac{\chi(p)}{p^s}\Big). \tag{57}$$

From the Lemma 6.8, we have

$$A(p,s) = B(p,s)^{g(p)}, \qquad B(p,s) = \prod_{\omega \in J_p} \Big(1 - \frac{\omega}{p^s}\Big) =$$

$$p^{-f(p)s} \prod_{\omega \in J_p} (p^s - \omega) =^* p^{-f(p)s}(p^{f(p)s} - 1) = 1 - \frac{1}{p^{f(p)s}}. \tag{58}$$

The starred equality has the following justification: the polynomial $x^M - 1$ factors as $(x - \omega_1)...(x - \omega_M)$, and we then set $x = p^s$ and $M = f(p)$.

Equations 57 and 58 combine to prove

$$F(s) = \prod_{(p,n)=1} \Big(\frac{1}{1 - \frac{1}{p^{sf(p)}}}\Big)^{g(p)}. \tag{59}$$

We can now expand this out just as in Equation 35, and we see that we get Equation 53 for some non-negative sequence $\{a_j\}$.

Proof of Step 3: Let f, g, ϕ be positive integers with $fg = \phi$. Compare the series:

$$\frac{1}{1 - \gamma^{\phi}}, \qquad \Big(\frac{1}{1 - \gamma^f}\Big)^g.$$

Both consist of positive terms, and every term of the series on the left also appears as a term in the series on the right. With this in mind, we define

$$G(s) = \prod_{(p,n)=1} \frac{1}{1 - \frac{1}{p^{s\phi(n)}}} = \sum_{j=1}^{\infty} \frac{b_j}{n^s} = L(\chi_0, s\phi(n)). \tag{60}$$

We have $F(s) \geq G(s)$ in the strong sense that $0 \leq b_j \leq a_j$ for all j. The last equality in Equation 60 shows the series diverges at $s = 1/\phi(n)$.

Proof of Step 4: We simplify the argument in Várilly's notes. The key to the proof is the following identity, which derives from the Taylor series expansion for e^z:

$$\sum_{k=0}^{\infty} \frac{1}{k!}(\log j)^k = j \qquad \forall j = 1, 2, 3, ... \tag{61}$$

Every point in H_0 has the form $s-1$ for some $s \in H_1$. Since F is complex analytic in H_0, by restriction F is also complex analytic in a disk Δ of radius $1+\epsilon$ about any point $s \in H_1$, as long as $\epsilon > 0$ is small enough. By the Taylor series property for F, applied to $s - 1 \in \Delta$, we have:

$$F(s-1) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{d^k F}{ds^k}(s). \tag{62}$$

Using Equation 53, we compute

$$\frac{d^k F}{ds^k}(s) = \sum_{j=1}^{\infty} (-1)^k \frac{a_j (\log j)^k}{j^s}. \tag{63}$$

Combining these equations, we have a convergent double sum, which we rearrange as:

$$F(s-1) = \sum_{j=1}^{\infty} \frac{a_j}{j^s} \Big(\sum_{k=0}^{\infty} \frac{1}{k!}(\log j)^k \Big) =^{*} \sum_{j=1}^{\infty} \frac{a_j}{j^s} \times j = \sum_{j=1}^{\infty} \frac{a_j}{j^{s-1}}. \tag{64}$$

The starred equality is Equation 61. This final expression is exactly the series from Equation 53 evaluated at $s - 1$. Hence Equation 53 holds not just in H_1 but also in H_0. This completes the proof.