Random Walks on Finite Groups

Hyojeong Son
Advisor: Robert Hough

August, 2019

Abstract

The main goal of this journal is to study random walks. The first part of this journal reports the crucial statements of group representation and character theory, Fourier analysis, and Markov chain, which are preliminary when studying random walks. The second part of the journal studies random walks by proving upper bound Lemma and studying the spectrum of a graph. Lastly, we present research done on sandpile dynamics on tiling graphs. In this journal, we focus on sandpile dynamics on the triangular lattice. Readers who are interested in sandpile dynamics on various tiling graphs may refer to the research journal[**HS19**].

In chapter one, we start by defining a representation. Representation theory is useful when studying a group because it often simplifies problems on groups into known problems in linear algebra. We prove every representation of a finite group G on a complex vector space is completely reducible. Later, we will see the set of irreducible characters form an orthonormal basis for the space of class functions on G. We note that regular representation contains all the irreducible representations; this is a key when proving Schur's second orthogonality relations which shows that columns of the character table are orthogonal, and thus, the character table is invertible. At the end of this section, we construct the character table for S_3 as an application of chapter one.

Having studied relevant statements in representation theory, we introduce the Fourier transform in Chapter two. We begin by proving a finite group G is isomorphic to its dual. Then we present the definition of Fourier transform and show how to invert it. We highlight the convolution identity which leads to Corollary 2.2.4; this corollary shows that Fourier transform is a ring isomorphism from $\mathbb{C}^G$ to $\mathbb{C}^{\widehat{G}}$.

In chapter three, we move to topics in the Markov chain. In the first section, we review the basics of probability theory such as linearity of expectation

and Markov inequality, the properties that are needed in a later section. In the remaining sections, we aim to prove the existence and uniqueness of the stationary distribution of a finite Markov chain. We assume that readers have no prior knowledge on the Markov chain, and hence, we start by studying Markov property, which sometimes referred to as memoryless property. Several definitions related to the Markov chain, including period, aperiodic, and irreducible have been introduced to see the behavior of the chain in depth. In the next section, we see the difference between recurrent and transient states and prove any state of finite Markov chain is either recurrent or transient (Corollary 3.3.4). Finally, in the last section, we prove the existence and uniqueness of the stationary distribution; we prove it by constructing a stationary distribution.

Chapter four consists of examples of random walks and a collection of remarkable results on the random walk. Random walks have full applications in various fields of studies such as economy, computer science, chemistry, and physics. A simple example of a random walk is a random walk on $\mathbb{Z}$. At each time step, a random walker move one step left or right from the current position with equal probability. Questions like, what is the expected position at time t can be answered by studying the walk. We conclude the chapter by giving a proof of upper bound Lemma by Diaconis and Shahshahani [**D88**], and studying a convolution operator.

Chapter five leads us to topics in the abelian sandpile model. In a 1987 paper by Bak, Tang, and Wisenfield, the sandpile model was first introduced. This model was represented as an example of a dynamic system with a self-organized criticality; it was one of the milestone discoveries in statistical physics in the 20th century. Since then, the model has been widely studied in physics and mathematics; there are more than 1800 returns in Google Scholar search under "abelian sandpile."

Let $G = \{V, E\}$ be a graph with a set of vertices and a set of edges. In the abelian sandpile model, chips are distributed on each vertex $v \in V$. We call a vertex v is stable if it has a fewer number of chips than it is degree; otherwise, we call the vertex v is unstable. Assume that we have one unstable vertex. We topple chips from one unstable vertex by sending out one chip to each neighboring vertex while setting one vertex as a sink, where passed chips to the sink are removed. Observe that toppling may cause other stable vertices to become unstable. We repeat toppling until we have no unstable vertex in the graph G; sandpile without unstable vertex is called a stable sandpile. We conclude chapter five by showing the existence and uniqueness of the stabilization of a sandpile.

In chapter six, we investigate sandpile dynamics on the triangular lattice with periodic and open boundary conditions. We define spectral parameters and state the theorem that gives the spectral parameter of the triangular tiling. We conclude the chapter six by giving optimization problems to determine spectral gap and spectral factors. For periodic tilings, there is no difference in the asymptotic mixing time between periodic and open boundary conditions. However, we discovered that for $D4$ lattice in dimension 4, there is a choice of a boundary with the open boundary mixing controlled by the 3-dimensional boundary. Readers who are interested in further information other than triangular lattice may want to take a look at the research journal by the author of this journal and Professor Robert Hough [**HS19**].

Acknowledgement

I would like to express my sincere gratitude to Professor Robert Hough for his supervision and continuous support throughout my undergraduate studies. I appreciate him for introducing such an exciting field of mathematics. It was a fantastic opportunity for me to participate in mathematical research under his guidance. You have set an example of excellence as a researcher, mentor, and role model.

I would like to thank the Stony Brook Summer Math Foundation and NSF Grants DMS-1712682 and DMS-1802336 for financial support.

Contents

1

Group Representation and Character Theory

When analyzing a group, studying its action on vector space reveals valuable information, and it is often easier than studying a group itself. Representation is a great tool that breaks down problems regarding abstract groups into linear algebra. In this section, we prove Maschke's theorem, which tells us that every finite group is a direct sum of irreducible representations. However, we need to note that this theorem does not guarantee that we have unique decomposition into irreducibles. We will give a proof of uniqueness in the following section.

1.1 Maschke's theorem

Definition 1.1.1. (Representation). A representation φ of group G is a homomorphism from G to $\mathrm{GL}(V)$ where V is a vector space.

Remark. The symbol $\deg(\varphi)$ denotes the dimension of V. Note that V is called a representation space of φ.

We say a representation φ of a group G is trivial when $\varphi(g) = 1$ for all g in G.

Definition 1.1.2. (G-invariant Subspace). Given a representation φ of a group G on a vector space V, G-invariant subspace W is a subspace of V if for all $w \in W$, $\varphi_g w \in W$ for all $g \in G$.

Remark. We call $\varphi|W : G \to GL(V)$ is a subrepresentation of φ when W is a G-invariant subspace of V.

Note that when a group is trivial, G-invariant subspaces are equivalent to subspaces of V. It is because $\varphi(g) = \mathcal{I}$ for all and unique element $g \in G$. Thus if W is a subspace of V, then $\varphi_g w = \mathcal{I} w \in W$ which implies W is a G-invariant subspace. Hence, when a group is trivial, W being a subspace of V automatically means that W is a G-invariant subspace.

Definition 1.1.3. (Equivalence). Given two representations φ and ψ of a group G on vector spaces V and W respectively, two representations are equivalent when there exists an isomorphism T from V to W such that $\psi_g = T\varphi_g T^{-1}$ for all $g \in G$.

There is a more relaxed version of definition than equivalence:

Definition 1.1.4. (Morphism). Given two representations φ and ψ of a finite group G on complex vector spaces V and W respectively, a morphism from φ to ψ is a linear map T from V to W such that below diagram commutes for all $g \in G$.

$$\begin{tikzcd} V \arrow[r, "\varphi_g"] \arrow[d, "T"'] & V \arrow[d, "T"] \\ W \arrow[r, "\psi_g"'] & W \end{tikzcd}$$

Note that a linear map T does not need to be an isomorphism. A set of morphisms are denoted as $\mathrm{Hom}_G(\varphi, \psi)$. We note that $\mathrm{Hom}_G(\varphi, \psi)$ is a subspace of $\mathrm{Hom}(V, W)$.

We may ask ourselves whether $\ker(T)$ and $\mathrm{Im}(T)$ are G-invariant subspaces of V and W, similar to other cases when we have homomorphism. Assume throughout Lemma 1.1.1 and 1.1.2 that two representations φ and ψ of a group G are given and T is a morphism between those representation spaces V and W respectively.

Lemma 1.1.1. $\ker(T)$ *is a G-invariant subspace of V.*

Proof. We show that given $v \in \ker(T)$, $\varphi_g(v)$ is also in $\ker(T)$. Since $T \in \mathrm{Hom}_G(\varphi, \psi)$,

$$T\varphi_g(v) = \psi_g T(v). \tag{1.1}$$

Hence, $\psi_g T(v)$ also equals to 0 because $T(v) = 0$. From (1.1), $T\varphi_g(v) = 0$. This concludes that $\ker(T)$ is a G-invariant subspace. $\square$

Lemma 1.1.2. $\mathrm{Im}(T)$ *is a* G*-invariant subspace of* W.

Proof. Given $w \in \mathrm{Im}(T)$, we show $\psi_g(w)$ is also an image of T. Suppose $T(v) = w$. Then $\psi_g(w)$ equals to $\psi_g T(v)$. Applying the fact that T is a morphism, we get

$$\psi_g T(v) = T\varphi_g(v).$$

Hence, $\psi_g T(v)$ is an image of T. Thus, $\psi_g(w) \in \mathrm{Im}(T)$. This concludes that $\mathrm{Im}(T)$ is a G-invariant subspace of W. $\square$

Definition 1.1.5. (Irreducible Representation). Given a representation φ of a group G on a vector space V, a group G is called irreducible if V and $\{0\}$ are only G-invariant subspaces of V.

Definition 1.1.6. (Decomposable Representation). Given a representation φ of a group G on a vector space V, φ is decomposable if V can be decomposed into two nonzero G-invariant subspaces.

Definition 1.1.7. (Completely Reducible). A representation φ of a group G on a vector space V is completely reducible if φ can be decomposed into direct sums of irreducible representations.

Observe that decomposition of decomposable representation into two G-invariant subspaces is not unique. Suppose the set $\{V_1, V_2, \cdots, V_n\}$ is a complete set of G-invariant subspaces of V with $V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$. Recalling that direct sums of subspaces is a subspace, a representation space V can be decomposed into two G-invariant subspaces in various ways. For example, first, V is a direct sum of V_1 and $V_2 \oplus V_3 \oplus \cdots \oplus V_n$. Alternately, a direct sum of $V_1 \oplus V_2$ and $V_3 \oplus V_4 \oplus \cdots \oplus V_n$ is V. However, the decomposition of completely reducible representation into direct sums of irreducible representations is unique up to isomorphism. We prove the uniqueness in the later chapter.

Irreducibility, decomposability, and completely reducibility are shared properties among equivalent representations; this is useful because when $\varphi \sim \psi$, if ψ is irreducible, then it automatically implies φ is also irreducible.

Throughout Lemma 1.1.3, 1.1.4, 1.1.5 and 1.1.6, we assume that φ and ψ are two representations of a group G on vector spaces V and W respectively. Also assume that φ is equivalent to ψ, and let T be an isomorphism such that below diagram commutes.

$$\begin{tikzcd} V \arrow[r, "\varphi_g"] \arrow[d, "T"'] & V \arrow[d, "T"] \\ W \arrow[r, "\psi_g"'] & W \end{tikzcd}$$

Lemma 1.1.3. *If ψ is an irreducible representation, then φ is also irreducible.*

Proof. Let ψ be an irreducible representation. Assume, for the sake of contradiction, φ is not irreducible. Then there exists a nonzero proper G-invariant subspace V'. Then since φ is equivalent to ψ, for $v' \in V'$

$$\varphi_g v' = T^{-1}\psi_g T v'. \tag{1.2}$$

Because of our assumption that V' is a G-invariant space, $\varphi_g v'$ is in V'. Thus by (1.2), we have

$$T^{-1}\psi_g T v' \in V'. \tag{1.3}$$

Applying T on the both sides of (1.3) gives $\psi_g T v' \in T(V')$. Since the choice of v' was arbitrary, $\psi_g T v' \in T(V')$ implies $T(V')$ is a G-invariant subspace of W. However, we do not have G-invariant subspace of W because we assumed that ψ is an irreducible representation. Thus, we have reached a contradiction, and hence our assumption was wrong. Therefore, φ is irreducible, and this completes the proof. $\square$

Lemma 1.1.4. *If ψ is a decomposable representation, then φ is also decomposable.*

Proof. Let $W = W_1 \oplus W_2$ where W_1 and W_2 are G-invariant subspaces of W. Suppose $T^{-1}(W_i) = V_i$ for $i = 1, 2$. We show $V = V_1 \oplus V_2$ where V_1 and V_2 are G-invariant subspaces.

We first show V_1 and V_2 are G-invariant subspaces. We know $\varphi_g(v_1)$ equals to $T^{-1}\psi_g T(v_1)$ and $T(v_1) \in W_1$. Recalling that W_1 is a G-invariant subspace gives us that $\varphi_g(v_1) = T^{-1}\psi_g T(v_1) \in T^{-1}(W_1) = V_1$. Thus, V_1 is a G-invariant subspace. A dual argument verifies V_2 is a G-invariant subspace.

We now show $V_1 \cap V_2 = \{0\}$. Suppose $v \in V_1 \cap V_2$. Then applying T on both sides yields $T(v) \in T(V_1) \cap T(V_2)$ and this equals to $T(v) \in W_1 \cap W_2$ by the definition. Since $W_1 \cap W_2 = \{0\}$ and T is an isomorphism, we arrive at v equals to 0.

Lastly, we prove $V = V_1 + V_2$. Since φ is equivalent to ψ, $\varphi_g(v) = T^{-1}\psi_g T(v)$. Observe that $T(v) \in W$. Thus, $T(v) = w_1 + w_2$ for some $w_1 \in W_1$ and $w_2 \in W_2$. Hence, $\varphi_g(v) = T^{-1}\psi_g(w_1 + w_2)$. Recalling that W_1 and W_2 are G-invariant subspace yields $\varphi_g(v) = T^{-1}\psi_g(w_1 + w_2) \in T^{-1}(W_1 + W_2) \in T^{-1}(W_1) + T^{-1}(W_2) = V_1 + V_2$. Since V_1 and V_2 are G-invariant subspaces, $\varphi_g(v) \in V_1 + V_2$ implies that $v \in V_1 + V_2$, and this completes the proof. $\square$

Lemma 1.1.5. *If ψ is a completely reducible representation, then φ is also completely reducible.*

Proof. Let ψ be a completely reducible representation. Then a representation space W of ψ can be decomposed into the direct sums of G-invariant subspaces of W:

$$W = W_1 \oplus W_2 \oplus \cdots \oplus W_n \tag{1.4}$$

where W_i is a nonzero G-invariant subspace and $\psi|_{W_i}$ is a irreducible representation for all i. Suppose $T^{-1}(W_i) = V_i$ for all i. We claim that $V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$ where V_i is nonzero G-invariant subspace and $\varphi|_{V_i}$ is a irreducible representation for all i.

We first show V is a direct sum of V_i's. In order to do so, we need to show that $V_i \cap \sum_{i \neq j} V_j = \{0\}$ for all $1 \leq i, j \leq n$. Suppose $v \in V_i \cap \sum_{i \neq j} V_j$. Applying T on the both sides yields

$$T(v) \in T(V_i) \cap T(\sum_{i \neq j} V_j).$$

Since T is an isomorphism, we get

$$T(v) \in T(V_i) \cap \Big(T(V_1) + T(V_2) + \cdots + T(V_{i-1}) + T(V_{i+1}) + \cdots + T(V_n)\Big).$$

Equivalently,

$$T(v) \in W_i \cap \Big(W_1 + W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_n\Big).$$

However, we know $W_i \cap \sum_{i \neq j} W_j = \{0\}$ for all $1 \leq i, j \leq n$ because W is a direct sums of W_i's. Thus,

$$T(v) = 0$$

and this implies $v = 0$ because T is an isomorphism. The choice of i was arbitrary, so we conclude that $V_i \cap \sum_{i \neq j} V_j = \{0\}$ for all $1 \leq i, j \leq n$.

Now let us to prove that each V_i is a G-invariant subspace. Since φ and ψ are equivalent to each other, for $v_i \in V_i$

$$\varphi_g v_i = T^{-1} \psi_g T v_i.$$

Observe that $T(v_i) \in W_i$. Since W_i is a G-invariant subspace, $\psi_g T(v_i) \in W_i$. Recalling that $T^{-1}(W_i) = V_i$ yields $T^{-1}\psi_g T(v_i) \in V_i$. Thus, V_i is G-invariant subspace. Lastly, we show $\varphi|_{V_i}$ is a irreducible representation. Observe that $\varphi|_{V_i}$ is equivalent to irreducible representation $\psi|_{W_i}$. From our previous Lemma 1.1.3, we get $\varphi|_{V_i}$ is also an irreducible representation. This completes the proof. □

Definition 1.1.8. (Unitary Representation). Let φ be a representation of a group G. Suppose a representation space V of φ is equipped with an inner product $\langle \cdot, \cdot \rangle$. A representation φ is unitary if

$$\langle \varphi_g(v), \varphi_g(w) \rangle = \langle v, w \rangle$$

for all $g \in G$ and $v, w \in V$.

In the next proposition, we prove that every representation of a finite group is equivalent to unitary representation. This proposition comes in handy when combined with lemmas that we established previously. Proving a unitary representation of a finite group is decomposable shows that every representation of a finite group is also decomposable from Lemma 1.1.3 and Lemma 1.1.4. We will see indeed this is true, and it is a key idea when proving Maschke's theorem.

Proposition 1.1.1. *Every representation of a finite group is equivalent to unitary representation.*

Proof. Let φ be a representation of a group $G = \{g_1, g_2, \ldots, g_n\}$ on a vector space V. We create a bijective map T from V to $\mathbb{C}_{|V|}$. Let $\langle\cdot,\cdot\rangle_1$ be an arbitrary inner product on $\mathbb{C}_{|V|}$. Define a new representation ψ in a following way:

$$\psi := T\varphi_g T^{-1}$$

Note that we have an equivalence relation between φ and ψ.

We define a new inner product $\langle\cdot,\cdot\rangle_2$ on W: for $v, w \in W$,

$$\langle v, w\rangle_2 := \sum_{i=1}^{n} \langle \psi_{g_i}(v), \psi_{g_i}(w)\rangle_1. \tag{1.5}$$

We first prove $\langle v, w\rangle_2$ is an inner product on W. Given an arbitrary element g_i in G, suppose $\psi_{g_i} v = v_i$ and $\psi_{g_i} w = w_i$. We show the inner product $\langle\cdot,\cdot\rangle_2$ satisfies the inner product axioms for all vectors $v, w, u \in W$ and for all scalars $c_1, c_2 \in \mathbb{C}$. We first see conjugate symmetry holds:

$$\begin{aligned}\langle v, w\rangle_2 &= \sum_{i=1}^{n} \langle \psi_{g_i}(v), \psi_{g_i}(w)\rangle_1 \\ &= \sum_{i=n}^{n} \langle v_i, w_i\rangle_1 \\ &= \langle v_1, w_1\rangle_1 + \langle v_2, w_2\rangle_1 + \cdots + \langle v_n, w_n\rangle_1.\end{aligned}$$

Since $\langle\cdot,\cdot\rangle_1$ is an inner product, we know $\langle v_i, w_i\rangle_1$ equals to $\overline{\langle w_i, v_i\rangle}_1$. Hence

$$\begin{aligned}\langle v, w\rangle_2 &= \overline{\langle w_1, v_1\rangle}_1 + \overline{\langle w_2, v_2\rangle}_1 + \cdots + \overline{\langle w_n, v_n\rangle}_1 \\ &= \sum_{i=n}^{n} \overline{\langle w_i, v_i\rangle}_1 \\ &= \overline{\langle w, v\rangle}_2.\end{aligned}$$

We next check positive definiteness of the inner product; by (1.5) and recalling that $\langle \cdot, \cdot \rangle_1$ is an inner product, we get

$$\begin{aligned}
\langle v, v \rangle_2 &= \sum_{i=1}^{n} \langle \psi_{g_i} v, \psi_{g_i} v \rangle_1 \\
&= \sum_{i=1}^{n} \langle v_i, v_i \rangle_1 \\
&\geq 0,
\end{aligned}$$

as desired.

We also observe that $\langle v, v \rangle_2$ equals to 0 if and only if $v = 0$. When $v = 0$, it is obvious that $\langle v, v \rangle_2 = 0$. It is easy to see the opposite direction also works. Suppose $\langle v, v \rangle_2 = \sum_{i=1}^{n} \langle \psi_{g_i} v, \psi_{g_i} v \rangle_1 = 0$. This implies $\langle \psi_e v, \psi_e v \rangle_1$, where e is an element identity of G, must equal to 0. Since $\varphi_2 v = v$, we get $\langle v, v \rangle_1 = 0$ and this happens if and only if $v = 0$.

We finally show the linearity of the inner product:

$$\begin{aligned}
\langle c_1 v + c_2 u, w \rangle_2 &= \sum_{i=1}^{n} \langle \psi_{g_i}(c_1 v + c_2 u), \psi_{g_i}(w) \rangle_1 \\
&= \sum_{i=1}^{n} \langle \psi_{g_i}(c_1 v) + \psi_{g_i}(c_1 u), \psi_{g_i}(w) \rangle_1 \\
&= \sum_{i=1}^{n} \Big(c_1 \langle \psi_{g_i}(v), \psi_{g_i}(w) \rangle_1 + c_2 \langle \psi_{g_i}(u), \psi_{g_i}(w) \rangle_1 \Big) \\
&= c_1 \langle v, w \rangle_2 + c_2 \langle u, w \rangle_2.
\end{aligned}$$

Hence $\langle \cdot, \cdot \rangle_2$ is an inner product. It remains for us to show that ψ is a unitary representation. For an arbitrary element g_j in G,

$$\begin{aligned}
\langle \psi_{g_j} v, \psi_{g_j} w \rangle_2 &= \sum_{i=1}^{n} \langle \psi_{g_i}(\psi_{g_j} v), \psi_{g_i}(\psi_{g_j} w) \rangle_1 \\
&= \sum_{i=1}^{n} \langle \psi_{g_{ij}} v, \psi_{g_{ij}} w \rangle_1.
\end{aligned}$$

Since g_{ij} sums over all the elements of G as i ranges over 1 to n,

$$\langle \psi_{g_j} v, \psi_{g_j} w \rangle_2 = \langle v, w \rangle_2.$$

This shows that ψ is a unitary representation with the inner product $\langle \cdot, \cdot \rangle_2$. Therefore, φ is equivalent to a unitary representation ψ. $\square$

Lemma 1.1.6. *If φ is a unitary representation of a group G and φ is not irreducible, then φ is decomposable.*

Proof. Let W be a G-invariant subspace, and let V be a representation space. Let us denote the orthogonal complement of W as $W^{\perp}$. Then $V = W \oplus W^{\perp}$. It only remains for us to show that $W^{\perp}$ is a G-invariant subspace in order to prove φ is a decomposable representation. Since φ is a unitary representation, for $w \in W$, $w' \in W^{\perp}$,

$$\begin{aligned}\langle w, \varphi_g(w') \rangle &= \langle \varphi_{g^{-1}}(w), \varphi_{g^{-1}g}(w') \rangle \\ &= \langle \varphi_{g^{-1}}(w), \varphi_e(w') \rangle \\ &= \langle \varphi_{g^{-1}}(w), w' \rangle.\end{aligned}$$

Because W is a G-invariant subspace and w is in W, $\varphi_{g^{-1}}(w)$ is in W. Hence $\langle \varphi_{g^{-1}}(w), w' \rangle = 0$, and therefore

$$\langle w, \varphi_g(w') \rangle = 0. \tag{1.6}$$

As a result, $\varphi_g(w') \in W^{\perp}$. This shows $W^{\perp}$ is also a G-invariant subspace, and this completes the proof. $\square$

We know from Proposition 1.1.1, that any representation φ of a finite group G is equivalent to a unitary representation ψ. There are two cases to consider: case 1 is when ψ is irreducible; case 2 is when ψ is not irreducible. In case 1, we are done. In case 2, applying Lemma 1.1.6 yields that ψ is decomposable. Since irreducibility and decomposability are shared properties among a class of equivalence representations, we conclude that a representation φ of a finite group is either irreducible or decomposable.

Now we are ready to prove Maschke's theorem. Maschke's theorem requires a representation space V to be a vector space over fields of characteristic zero; recall that field of $\mathbb{C}$ is characteristic zero.

Theorem 1.1.7. *(Maschke's theorem). Every representation of a finite group G on a complex vector space V is completely reducible.*

Proof. We use mathematical induction on the dimension of a degree of representation to prove Maschke's theorem. When the degree of representation is 1, the representation is necessarily irreducible, and hence, completely reducible. For the inductive step, assume that Maschke's theorem holds for any representation with a degree less than n. Now suppose the degree of φ is n. If φ is an irreducible representation of G, then we are done. So suppose φ is not an irreducible representation, then there exists a G-invariant subspace W of V. We want to prove the existence of a G-invariant subspace W' such that $V = W \oplus W'$. From our inductive hypothesis, we know that subrepresentations $\varphi|_W$ and $\varphi|_{W'}$ are completely reducible. Because their direct sums are completely reducible, we conclude that φ is completely reducible. □

Remark. Since we are working on complex vector spaces, we can place $|G|$ in the denominator when defining a map $\overline{T}$. If the characteristic of a field F divides the order of G, then Maschke's theorem does not hold.

Observe that all degree 1 representations are irreducible representation because it is not possible to have a proper nonzero subspace with a degree less than its representation space, which is 1.

For degree 2 or 3 representations, there is an easy way to check whether the representation is irreducible or not. The idea is simple; we use the fact that if a degree 2 representation is not irreducible, then there exists a one-dimensional G-invariant subspace. Suppose there exists a common eigenvector v, then $\mathbb{C}v$, a degree one subspace, forms a G-invariant subspace. Hence, the representation is not irreducible. For degree 3 representation φ of a group G, we may apply Maschke's theorem. Suppose there exists a common eigenvector v of φ_g for all $g \in G$. If φ is irreducible, then there exists either one-dimensional or two-dimensional G-invariant subspace. In case we have one-dimensional G-invariant subspace, we can apply the same logic we used in degree 2 representations. In case we have two-dimensional G-invariant subspace, from Maschke's theorem, we know every representation of a finite group is completely irreducible; hence there exists a complementary one-dimensional G-invariant subspace. Then again, we can apply the same logic again. As a result, we have following proposition:

Proposition 1.1.2. *Let φ be a degree 2 representation of a group G. Then φ is irreducible if and only if there exists no common eigenvector of φ_g for all $g \in G$.*

Proof. Let V be a representation space.
($\Rightarrow$): We give a proof of contraposition; if there exists a common eigenvector among φ_g for all $g \in G$, then φ is not irreducible. Let w be a common eigenvector of all φ_g for all $g \in G$. Hence $\varphi_g w = \lambda_g w$ for all $g \in G$, where $\lambda_g \in \mathbb{C}$; here the value of λ_g is dependent on the choice of g. Let W be a subspace of V formed by a basis $\{w\}$. It follows that a subspace W is a proper nonzero subspace because $\varphi_g w = \lambda_g w \in W \subseteq V$ for all $w \in W$. Hence we conclude that φ is not irreducible.

($\Leftarrow$): We also give a proof of contraposition to prove; if φ is not irreducible, then there exists a common eigenvector of φ_g for all $g \in G$. Assume φ is not an irreducible representation. Then there exists a proper subspace W of V such that $\varphi_g w \in W$ for all $w \in W$ and $g \in G$. Since W is a proper subspace of V, the G-invariant subspace W must be one-dimensional. Then it follows that $\{w\}$ forms a basis for W, and $\varphi_g w \in W$; this implies $\varphi_g w = \lambda_g w$ where $\lambda_g \in \mathbb{C}$. This shows that w is a common eigenvector of all φ_g, and this completes the proof. $\square$

In general, there is an easy way to check the irreducibility of representations with some help of character theory, which will be introduced in a later chapter.

1.2 Schur's Lemma

We are now ready to prove Schur's lemma which is an essential statement in representation theory. It tells us that given two irreducible representations and a linear transformation T between them, the linear transformation T must be invertible; otherwise T is trivial. In case we have identical representations, the map T between them is some scalar multiple of an identity map.

Lemma 1.2.1. *(Schur's Lemma). Let φ and ψ be two irreducible finite-dimensional representations of a group G on complex vector spaces V and W respectively, and let T be in $\mathrm{Hom}_G(\varphi, \psi)$. Then following holds;*
(1) If $\varphi \nsim \psi$, then there exists no nontrivial linear map T between V and W;

(2) If $\varphi = \psi$, then any linear map T between V and W is a some scalar multiple of the identity map.

Proof. (1): We give a proof by contraposition: if there exists a nontrivial linear transformation T from V to W, then φ is equivalent to ψ. We observe that T must be invertible with given conditions. Recall from Lemma 1.1.1 and 1.1.2, $T \in \mathrm{Hom}_G(\varphi, \psi)$ implies $\ker(T)$ and $\mathrm{Im}(T)$ are G-invariant subspaces. Since φ and ψ are irreducible representations, they have no proper subrepresentation. Hence, $\ker(T)$ is either 0 or V, and $\mathrm{Im}(T)$ is either 0 or W. If $\ker(T) = V$, then $T = 0$; however, we assumed $T \neq 0$. Hence $\ker(T) = 0$ and similarly we can verify $\mathrm{Im}(T) = W$. Therefore, T is an isomorphism. Then T is an invertible linear map such that $T \in \mathrm{Hom}_G(\varphi, \psi)$, and this implies that φ is equivalent to ψ.

(2): Suppose $\varphi = \psi$. Recall that the representation space V is a complex vector space. Due to the fundamental theorem of algebra, we can assume that there exists an eigenvalue λ of T. Note that $T - \lambda\mathcal{I}$, where $\mathcal{I}$ denotes the identity map, is not invertible according to the definition of an eigenvalue. Observe that the identity map $\mathcal{I}$ commutes with the action of the group. Since $\mathrm{Hom}_G(\varphi, \psi)$ is a subspace of $\mathrm{Hom}(V, W)$ and $T, \mathcal{I} \in \mathrm{Hom}_G(\varphi, \psi)$, their linear combination, $T - \lambda\mathcal{I}$ is in $\mathrm{Hom}_G(\varphi, \psi)$. However, from the previous part of the proof, we know that any nonzero linear map is invertible. Since $T - \lambda\mathcal{I}$ is not invertible, $T - \lambda\mathcal{I}$ must be equal to 0. Hence, $T = \lambda\mathcal{I}$, and this completes the proof. □

We prove a corollary followed after Schur's lemma. Before proving it, observe an immediate consequence of Schur's lemma. Suppose φ and ψ are two finite-dimensional irreducible representations of G on complex vector spaces V and W respectively, and φ is equivalent to ψ. Also, suppose there are two invertible maps T_1, T_2 in $\mathrm{Hom}_G(\varphi, \psi)$. Then $T_1 \circ T_2^{-1}$ is an invertible map from W to W and $T_1 \circ T_2^{-1} \in \mathrm{Hom}_G(\varphi, \psi)$. We can apply the second statement of Schur's lemma in this situation. After applying, we get $T_1 \circ T_2^{-1} = \lambda\mathcal{I}$ for some scalar λ. Multiplying T_2 to the both sides of the equation gives $T_1 = \lambda T_2$. Therefore, we conclude that $\dim \mathrm{Hom}_G(\varphi, \psi) = 1$.

By applying Schur's lemma to representations of an abelian group, we get interesting results as below:

Corollary 1.2.2. *The degree of any irreducible finite-dimensional representation of an abelian group on a complex vector space V is one.*

Proof. Let G be an abelian group. Then $\varphi_h\varphi_g = \varphi_g\varphi_h$ for all $g, h \in G$. Therefore, $\varphi_h \in \mathrm{Hom}_G(\varphi, \varphi)$. It follows from Schur's lemma that $\varphi_h = \lambda_h\mathcal{I}$ where λ_h is dependent on the choice of h. Then for $v \in V$ and $c \in \mathbb{C}$, $\varphi_h(cv) = \lambda_h v \in \mathbb{C}v$, and this implies $\mathbb{C}v$ is a G-invariant subspace. Since φ is irreducible representation, we conclude that the G-invariant subspace $\mathbb{C}v$ is V. Hence the representation space V is one-dimensional and this completes the proof. □

Remark. Given a representation of a finite abelian group, we know from the Maschke's theorem, the representation is completely reducible. Then from this Corollary, the degree of any irreducible representation of a finite abelian group is one. Hence, the matrix representation of a finite abelian group is diagonalizable.

1.3 Schur's Orthogonality Relations

Having proven Schur's lemma, now we are ready to prove Schur's orthogonality relations. When we talk about representation in this section, we refer to a matrix representation of it. Given a degree n representation φ, Schur's orthogonality relations tells us that $\{\varphi_{ij} \mid 1 \leq i, j \leq n\}$ forms an orthogonal set. The proof uses "averaging trick." Like previous sections of this journal, we work on a representation of a finite group on a complex vector space.

Proposition 1.3.1. *Let φ and ψ be two finite-dimensional irreducible representations of a group G on complex vector spaces V and W respectively, and let T be any linear map from V to W. Then following holds;*
(1) If $\varphi \nsim \psi$, then $\frac{1}{|G|}\sum_{g\in G}\psi_{g^{-1}}T\varphi_g = 0$;
(2) If $\varphi = \psi$, then $\frac{1}{|G|}\sum_{g\in G}\psi_{g^{-1}}T\varphi_g = \frac{\mathrm{tr}(T)}{\dim V}\mathcal{I}$.

Proof. Let us define a new map by using "averaging trick":

$$\overline{T} := \frac{1}{|G|}\sum_{h\in G}\psi_{h^{-1}}T\varphi_h.$$

We first show $\overline{T} \in \mathrm{Hom}_G(\varphi, \psi)$. From the definition of $\overline{T}$, for $g \in G$,

$$\begin{aligned}\overline{T}\varphi_g &= \frac{1}{|G|}\sum_{h\in G}\psi_{h^{-1}}T\varphi_h\varphi_g \\ &= \frac{1}{|G|}\sum_{h\in G}\psi_{h^{-1}}T\varphi_{hg}.\end{aligned}$$

After the change of variables $hg \to g'$, we are left with

$$\begin{aligned}\overline{T}\varphi_g &= \frac{1}{|G|}\sum_{g'\in G}\psi_{gg'^{-1}}T\varphi_{g'} \\ &= \psi_g\frac{1}{|G|}\sum_{g'\in G}\psi_{g'^{-1}}T\varphi_{g'} \\ &= \psi_g\overline{T}.\end{aligned}$$

Note that $\sum_{g'\in G}\psi_{g'^{-1}}T\varphi_{g'}$ is same as $\sum_{h\in G}\psi_h T\varphi_{h^{-1}}$ because both sum over all the finite elements of G. Hence $\overline{T}\varphi_g = \psi_g\overline{T}$ for all $g \in G$. Thus, $\overline{T} \in \mathrm{Hom}_G(\varphi, \psi)$. Now we apply Schur's lemma to finish the proof.

(1): Suppose $\varphi \not\sim \psi$. Then by Schur's lemma, we know there is no nontrivial map. Thus, we must have $\overline{T} = 0$.

(2): Suppose $\varphi = \psi$. Then $\overline{T} \in \mathrm{Hom}_G(\varphi, \varphi)$. After applying Schur's lemma, we see that

$$\overline{T} = \lambda\mathcal{I} \tag{1.7}$$

where $\lambda \in \mathbb{C}$. We can calculate the value of λ. The equation 1.7 implies

$$\lambda \dim V = \frac{1}{|G|}\sum_{h\in G}\mathrm{tr}(\varphi_{h^{-1}}T\varphi_h).$$

Since $\mathrm{tr}(AB) = \mathrm{tr}(BA)$,

$$\begin{aligned}\lambda \dim V &= \frac{1}{|G|}\sum_{h\in G}\mathrm{tr}(T\varphi_h\varphi_{h^{-1}}) \\ &= \frac{1}{|G|}\sum_{h\in G}\mathrm{tr}(T\varphi_{hh^{-1}}) \\ &= \frac{1}{|G|}\sum_{h\in G}\mathrm{tr}(T) \\ &= \frac{|G|}{|G|}\mathrm{tr}(T) = \mathrm{tr}(T).\end{aligned}$$

Upon dividing both sides by $\dim V$, we get $\lambda = \frac{\mathrm{tr}(T)}{\dim V}$. Thus, $\overline{T} = \frac{\mathrm{tr}(T)}{\dim V}\mathcal{I}$ and this completes the proof. □

Having proven the previous proposition, we are now ready give a proof of Schur's Orthogonality Relations.

Theorem 1.3.1. *(Schur's Orthogonality Relations). Let φ and ψ be irreducible representations of a finite group G on complex vector spaces V and W respectively. Then*
(1) If $\varphi \nsim \psi$, then $\langle \varphi_{ij}, \psi_{kl} \rangle = 0$;
(2) If $\varphi = \psi$, then $\langle \varphi_{ij}, \varphi_{kl} \rangle = \begin{cases} \frac{1}{\dim V} & \text{if } i = k \text{ and } k = l, \\ 0 & \text{others.} \end{cases}$

Proof. Let T be any linear map from V to W. In the proof of Proposition 1.3.1, we showed that $\overline{T} = \frac{1}{|G|}\sum_{h\in G}\psi_{h^{-1}}T\varphi_h \in \mathrm{Hom}_G(\varphi, \psi)$. There are two cases to consider: case 1 is when $\varphi \nsim \psi$; case 2 is when $\varphi = \psi$. In case 1, by the Schur's lemma $\overline{T} = 0$. Hence

$$\begin{aligned}(\overline{T})_{lj} &= \frac{1}{|G|}\sum_{h\in G}\sum_{k,i}(\psi_{h^{-1}})_{lk}T_{ki}(\varphi_h)_{ij} \\ &= \frac{1}{|G|}\sum_{k,i}T_{ki}\Big(\sum_{h\in G}(\psi_{h^{-1}})_{lk}(\varphi_h)_{ij}\Big) = 0.\end{aligned} \tag{1.8}$$

Recall that T is an arbitrary linear map from V to W. To make (1.8) always holds, we must have $\sum_{h\in G}(\psi_{h^{-1}})_{lk}(\varphi_h)_{ij}$ equal to 0. From Proposition 1.1.1, we know every representation of a finite group is unitary. Hence, we may assume φ and ψ are unitary representations, and therefore $(\psi_{h^{-1}})_{lk} = (\psi_h)^*_{lk} = \overline{(\psi_h)}_{kl}$. Thus,

$$\sum_{h\in G}(\varphi_h)_{ij}(\psi_{h^{-1}})_{lk} = \sum_{h\in G}(\varphi_h)_{ij}\overline{(\psi_h)}_{kl} = 0. \tag{1.9}$$

We now can calculate $\langle \varphi_{ij}, \psi_{kl} \rangle$:

$$\langle \varphi_{ij}, \psi_{kl} \rangle = \frac{1}{|G|}\sum_{g\in G}\varphi_{ij}(g)\overline{\psi_{kl}}(g)$$

The equation 1.9 implies

$$\langle \varphi_{ij}, \psi_{kl} \rangle = 0$$

as desired.

In case 2, when $\varphi = \psi$, by the Schur's lemma,

$$\overline{T} = \frac{1}{|G|}\sum_{h\in G}(\varphi_{h^{-1}}T\varphi_h) = \lambda\mathcal{I} \tag{1.10}$$

Recalling from Proposition 1.3.1 that $\lambda = \frac{\operatorname{tr}(T)}{\dim V}$ and substituting λ for $\frac{\operatorname{tr}(T)}{\dim V}$ in (1.10) yields

$$\frac{\operatorname{tr}(T)}{\dim V}\mathcal{I}_{lj} = \frac{1}{|G|}\sum_{h\in G}(\varphi_{h^{-1}})_{lk}T_{ki}(\varphi_h)_{ij}, \tag{1.11}$$

which is equivalently rewritten as

$$\frac{\sum_{ki} T_{ki}\delta_{ik}\delta_{jl}}{\dim V} = \frac{1}{|G|}\sum_{k,i} T_{ki}\sum_{h\in G}(\varphi_{h^{-1}})_{lk}(\varphi_h)_{ij} \tag{1.12}$$

where δ is a Kronecker delta function. Rearranging (1.13) gives

$$\sum_{ki} T_{ki}\left(\frac{\delta_{ik}\delta_{jl}}{\dim V} - \frac{1}{|G|}\sum_{h\in G}(\varphi_h)_{ij}(\varphi_{h^{-1}})_{lk}\right) = 0. \tag{1.13}$$

Recall that T is an arbitrary map from V to W. Thus, to make (1.13) always hold, we need

$$\frac{\delta_{ki}\delta_{lj}}{\dim V} = \frac{1}{|G|}\sum_{h\in G}(\varphi_h)_{ij}(\varphi_{h^{-1}})_{lk}. \tag{1.14}$$

Observe that we get $\frac{1}{|G|}\sum_{h\in G}(\varphi_h)_{ij}(\varphi_{h^{-1}})_{lk}$ equals to $\frac{1}{\dim V}$ when $k = i$ and $l = j$, and 0 otherwise. Again, by Proposition 1.1.1, we can assume φ is unitary and this implies that $(\varphi_{h^{-1}})_{lk} = (\varphi_h)^*_{lk} = \overline{(\varphi_h)_{kl}}$. Applying our observation to (1.14), we finally arrive at

$$\frac{1}{|G|}\sum_{h\in G}(\varphi_h)_{ij}\overline{(\varphi_h)_{kl}} = \begin{cases} \frac{1}{\dim V} & \text{if } i = k \text{ and } k = l, \\ 0 & \text{others.} \end{cases} \tag{1.15}$$

Note that $\langle \varphi_{ij}, \varphi_{kl}\rangle = \frac{1}{|G|}\sum_{g\in G}\varphi_{ij}(g)\varphi_{kl}(g)$, and this completes the proof. $\square$

1.4 Schur's First Orthogonality Relations

Definition 1.4.1. (Discrete Convolution). Given complex valued functions a and b, the discrete convolution of a and b is given by

$$a * b(x) = \sum_{y} a(xy^{-1})b(y).$$

Definition 1.4.2. ($\mathbb{C}^G$). Given a group G, $\mathbb{C}^G$ is a set of all functions from G to $\mathbb{C}$. Suppose $a, b \in \mathbb{C}^G$. The operation of addition is given by

$$(a+b)(g) = a(g) + b(g).$$

The scalar multiplication is given by

$$k(a(g)) = k \cdot a(g)$$

for k in $\mathbb{C}$. The operation of convolution as a multiplication is given by

$$a * b(g) = \sum_{y \in G} a(gy^{-1})b(y).$$

Lastly, the inner product is given by

$$\langle a, b \rangle = \frac{1}{|G|} \sum_{g \in G} a(g)\overline{b(g)}.$$

Theorem 1.4.1. *$\mathbb{C}^G$ is a ring with two binary operations $(+, *)$.*

Proof. Suppose $a, b, c \in \mathbb{C}^G$. We prove the multiplication is associative. From the definition of convolution, we get

$$\begin{aligned}(a * b) * c(g) &= \sum_{y \in G} (a * b)(gy^{-1})c(y) \\ &= \sum_{y \in G} \sum_{x \in G} a(gy^{-1}x^{-1})b(x)c(y).\end{aligned}$$

The change of variable $xy \to z$ leaves,

$$\begin{aligned}(a*b)*c(g) &= \sum_{y\in G}\sum_{z\in G} a(gz^{-1})b(zy^{-1})c(y)\\ &= \sum_{z\in G} a(gz^{-1})\sum_{y\in G} b(zy^{-1})c(y)\\ &= \sum_{z\in G} a(gz^{-1})(b*c)(z)\\ &= a*(b*c)(g).\end{aligned}$$

Hence the multiplication is associative. We skip the rest of the part of the proof because it is a straight forward to check $\mathbb{C}^G$ is a ring. $\square$

Remark. The Kronecker delta function δ_e is a multiplicative identity. To see this, we calculate $a*\delta_e(g)$: $a*\delta_e(g) = \sum_{y\in G} a(gy^{-1})\delta_e(y)$. Note that the Kronecker delta function $\delta_e(y)$ is not zero if and only if $y = e$. Then $\sum_{y\in G} a(gy^{-1})\delta_e(y) = a(ge^{-1})\cdot 1$. Therefore, $a*\delta_e(g)$ equals to $a(g)$. The same logic works to see $\delta_e * a(g) = a(g)$.

Definition 1.4.3. (Character of a Representation). Given a representation φ of a finite group G, the character χ of $\varphi(g)$ is

$$\chi_\varphi(g) = \operatorname{tr}(\varphi(g)).$$

Remark. By recalling that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, we see that $\chi_\varphi(hgh^{-1}) = \chi_\varphi(gh^{-1}h) = \chi_\varphi(g)$. Therefore, χ is constant on conjugacy class. Hence, χ is a class function.

Remark. When we talk about characters, we assume that representation is a matrix representation; in this way, we can calculate the trace of φ.

Definition 1.4.4. (Linear Character). A linear character is a homomorphism $\chi : G \to \mathbb{C}^\times$.

Remark. A linear character is a special kind of character of a representation; it is a character of a degree one representation. Observe that $\operatorname{tr}(\varphi(g)) = \varphi(g)$ when the degree of φ is one.

For the convenience of notation, a linear character is referred to as a character.

Let $(G, +)$ be a finite group with a binary operation $+$. Note that linear characters χ of G form a finite abelian group. For x, y in G,

$$\chi(x+y) = \chi(x) \cdot \chi(y) = \chi(y) \cdot \chi(x).$$

Since G is a finite group, for every $x \in G$, $x^n = 1$ for some n, where 1 is the identity element of G. Then $\chi(x^n) = \chi(x)^n = 1$. Thus, we can think of $\chi(x)$ as a n-th root of unity. In the case when $G = \mathbb{Z}/p\mathbb{Z}$, the characters are p-th root of unity.

$$\chi_n(x) = e^{2\pi i x n/p}.$$

Observe that the image of χ is the unit circle.

We also note that the product of the two characters is a character:

$$\begin{aligned}\chi_n(x) \cdot \chi_n(y) &= e^{2\pi i x n/p} \cdot e^{2\pi i y n/p} \\ &= e^{2\pi i (x+y) n/p} \\ &= \chi_n(x+y).\end{aligned}$$

Having made this observation, we can see the set of characters form an abelian group. Before giving a formal proof, we define a set of all characters of a group G.

Definition 1.4.5. (Dual Group). Let G be a finite abelian group. A dual group $\widehat{G}$ is a set of all irreducible characters of G with multiplication as an operator. The multiplication is given by

$$(\chi \cdot \eta)(g) = \chi(g)\eta(g)$$

where χ and η are characters of G and $g \in G$.

Remark. Recall, from Corollary 1.2.2, that the degree of an irreducible representation of a finite abelian group is one. Hence, an irreducible character of an irreducible representation of a finite abelian group is a linear character.

Lemma 1.4.2. *If G is a finite abelian group, then the dual group $\widehat{G}$ is an abelian group.*

Proof. We first prove $\widehat{G}$ is a group. The identity element is χ_1, a character of a trivial irreducible representation. Given an element $\chi \in \widehat{G}$, the inverse is

$\chi(g^{-1})$. The associativity law holds trivially. It remains for us to show $\widehat{G}$ is abelian. For $\chi, \psi \in \widehat{G}$ and $g_1, g_2 \in G$,

$$\begin{aligned}
\chi \cdot \psi(g_1 g_2) &= \chi(g_1 g_2)\psi(g_1 g_2) \\
&= \chi(g_1)\chi(g_2)\psi(g_1)\psi(g_2) \\
&= \chi(g_2)\chi(g_1)\psi(g_2)\psi(g_1) \\
&= \chi(g_2 g_1)\psi(g_2 g_1) \\
&= \chi \cdot \psi(g_2 g_1).
\end{aligned}$$

Thus $\widehat{G}$ is an abelian group. □

Observe that $Z(\mathbb{C}^G)$, the center of $\mathbb{C}^G$, is a subspace of $\mathbb{C}^G$. Let $a, b \in Z(\mathbb{C}^G)$. For $c \in \mathbb{C}^G$, $(a+b) * c(g)$ equals to $a * c(g) + b * c(g)$ which is equivalently rewritten as $c * a(g) + c * b(g)$ because a and b are in the center of $\mathbb{C}^G$. Hence, $(a+b) * c(g) = c * (a+b)(g)$, and this shows $Z(\mathbb{C}^G)$ is a subspace. A similar argument shows that the space of class functions of $\mathbb{C}^G$ is a subspace.

Previously, we showed that characters are class functions. We will see class functions are $Z(\mathbb{C}^G)$, and vice versa.

Theorem 1.4.3. *Given a function $a \in \mathbb{C}^G$, a is a class function if and only if a is in the center of $\mathbb{C}^G$*

Proof. $(\Rightarrow)$: Suppose $a \in \mathbb{C}^G$ is a class function. For $b \in \mathbb{C}^G$ and $x, y \in G$

$$b * a(x) = \sum_{y \in G} b(xy^{-1})a(y).$$

Since a is a class function, a is constant on the conjugate classes of G. Thus

$$b * a(x) = \sum_{y \in G} b(xy^{-1})a(xyx^{-1})$$

which is equivalently rewritten as,

$$b * a(x) = \sum_{y \in G} a(x(xy^{-1})^{-1})b(xy^{-1}).$$

Note that xy^{-1} sums over all the elements in the finite group G. Hence,

$$b * a(x) = a * b(x)$$

as desired.

($\Leftarrow$): Now suppose $a \in \mathbb{C}^G$ is in $Z(\mathbb{C}^G)$. Note that for $x, y \in G$,

$$\begin{aligned} a(yxy^{-1}) &= \sum_{z \in G} a(yz^{-1})\delta_{yx^{-1}}(z) \\ &= a * \delta_{yx^{-1}}(y). \end{aligned}$$

Because we assumed $a \in Z(\mathbb{C}^G)$, $a * \delta_{yx^{-1}}(y)$ equals to $\delta_{yx^{-1}} * a(y)$. Hence,

$$\begin{aligned} a(yxy^{-1}) &= \delta_{yx^{-1}} * a(y) \\ &= \sum_{z \in G} \delta_{yx^{-1}}(yz^{-1})a(z). \end{aligned}$$

The Kronecker delta function $\delta_{yx^{-1}}(yz^{-1})$ equals to 1 if and only if $yx^{-1} = yz^{-1}$ which happens if and only if when $z = x$. Therefore,

$$a(yxy^{-1}) = \sum_{z \in G} \delta_{yx^{-1}}(yz^{-1})a(z) = a(x)$$

as desired. □

Theorem 1.4.4. *Let $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_n$ be a complete set of distinct conjugacy classes of a finite group G. Then the set $\{\delta_{\mathcal{C}_i} | 1 \le i \le n\}$ forms a basis for the space of class function in $\mathbb{C}^G$.*

Proof. We first show $\{\delta_{\mathcal{C}_i} | 1 \le i \le n\}$ spans the space of class functions. Let a be a class function. Then

$$a(g) = \sum_i a(\mathcal{C}_i)\delta_{\mathcal{C}_i}(g).$$

Hence the $\delta_{\mathcal{C}_i}$ spans the space of class functions. Next we show the set $\{\delta_{\mathcal{C}_i} | 1 \le i \le n\}$ is orthogonal. Given two Kronecker delta functions $\delta_{\mathcal{C}_i}$ and $\delta_{\mathcal{C}_j}$ such that $i \neq j$, their inner product is

$$\langle \delta_{\mathcal{C}_i}, \delta_{\mathcal{C}_j} \rangle = \frac{1}{|G|} \sum_{g \in G} \delta_{\mathcal{C}_i}(g)\delta_{\mathcal{C}_j}(g).$$

However, $\delta_{\mathcal{C}_i}(g)\delta_{\mathcal{C}_j}(g)$ always equals to 0 when $i \neq j$. This establishes the orthogonality. Hence, $\{\delta_{\mathcal{C}_i} | 1 \leq i \leq n\}$ form a basis. $\square$

Remark. The dimension of the space of class functions of $\mathbb{C}^G$ equals to the class number of a group, the number of conjugacy classes in a group.

We showed the set $\{\delta_{\mathcal{C}_i} | 1 \leq i \leq n\}$ forms a basis for the space of class function in $\mathbb{C}^G$. In Theorem 1.4.3, we have shown the class function is the center of $\mathbb{C}^G$. Therefore, combined with Theorem 1.4.3, the set $\{\delta_{\mathcal{C}_i} | 1 \leq i \leq n\}$ also forms a basis for the center of $\mathbb{C}^G$.

Lemma 1.4.5. *Let G be a finite abelian group and let $\chi \in \widehat{G}$. Then*

$$\sum_{x \in G} \chi(x) = \begin{cases} |G| & \text{if } \chi = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $S(\chi) = \sum_{x \in G} \chi(x)$. The key idea of the proof is observing $S(\chi) = \chi(y)S(\chi)$ for all $h \in G$:

$$\begin{aligned} \chi(y) \cdot S(\chi) &= \chi(y) \sum_{x \in G} \chi(x) \\ &= \sum_{x \in G} \chi(xy). \end{aligned}$$

After the change of variables $z \to xy$, we get

$$\begin{aligned} \chi(y) \cdot S(\chi) &= \sum_{z \in G} \chi(z) \\ &= S(\chi). \end{aligned} \tag{1.16}$$

Hence,

$$S(\chi)\Big(1 - \chi(y)\Big) = 0.$$

If χ is trivial, then $\chi(y) = 1$ for all $y \in G$. Therefore $S(\chi) = \sum_{x \in G} \chi(x)$ equals to $|G|$. In case when χ is not trivial character, choose y with $\chi(y) \neq 1$. Then $S(\chi)$ must equal to 0, to make equation 1.16 always hold. This completes the proof. $\square$

Before we prove Schur's first orthogonality relations, we introduce an easy lemma.

Lemma 1.4.6. *Let φ and ψ be representations of a group G. If φ is equivalent to ψ, then χ_φ equals to χ_ψ.*

Proof. From the definition of a character,

$$\chi_\varphi(g) = \mathrm{tr}(\varphi(g)).$$

Since φ is equivalent to ψ, $\varphi(g) = T^{-1}\psi T(g)$ where T is an invertible map in $\mathrm{Hom}_G(\varphi, \psi)$. Then

$$\chi_\varphi(g) = \mathrm{tr}\,(T^{-1}\psi T(g)).$$

Recalling $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$ yields,

$$\chi_\varphi(g) = \mathrm{tr}\,(\psi T T^{-1}(g)) = \mathrm{tr}(\psi(g)) = \chi_\psi(g),$$

as desired. □

We now give the proof of Schur's first orthogonality relations. This theorem tells us that irreducible characters form an orthonormal set. In our later discussion, we will prove the set of irreducible characters form a basis for the space of class functions. Recall from the Theorem 1.4.3 that given a function in $\mathbb{C}^G$, the function is a class function if and only if the function is in the center of $\mathbb{C}^G$. Hence showing the set of irreducible characters form an basis for the space of class functions implies that the set also forms a basis for the center of $\mathbb{C}^G$.

Theorem 1.4.7. *(Schur's First Orthogonality Relations). Let φ and ψ be two irreducible representations of G. Then*

$$\langle \chi_\varphi, \chi_\psi \rangle = \begin{cases} 1 & if \varphi \sim \psi, \\ 0 & if \varphi \nsim \psi. \end{cases}$$

Proof. Let n and m be the degree of φ, ψ respectively. Then

$$\begin{aligned}\langle \chi_\varphi, \chi_\psi \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi_\varphi \overline{\chi_\psi} \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi_{ii}(g) \overline{\psi_{jj}(g)} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{|G|} \sum_{g \in G} \varphi_{ii}(g) \overline{\psi_{jj}(g)} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \langle \varphi_{ii}, \psi_{jj} \rangle. \end{aligned} \tag{1.17}$$

There are two cases consider: case 1 is when $\varphi \sim \psi$; case 2 is when $\varphi \nsim \psi$. For the case 1, first recall from Lemma 1.4.6 that when two representations are equivalent, their character is same. Hence,

$$\chi_\varphi = \chi_\psi.$$

Applying the Schur's orthogonality relations to the equation 1.17 yields

$$\langle \chi_\varphi, \chi_\psi \rangle = \langle \chi_\varphi, \chi_\varphi \rangle = \sum_{i=1}^{n} \langle \varphi_{ii}, \varphi_{ii} \rangle = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

For the case 2, again applying the Schur's orthogonality relations to the equation 1.17 yields

$$\langle \chi_\varphi, \chi_\psi \rangle = 0.$$

□

Remark. Since $\varphi \nsim \psi$ implies $\langle \chi_\varphi, \chi_\psi \rangle = 0$, the set of characters of distinct irreducible representations is orthogonal.

It remains for us to show that the set of irreducible characters spans the space of class functions on G.

Definition 1.4.6. (Regular Representation). The (left) regular representation R of a finite group G is a homomorphism from G to $GL(\mathbb{C}^G)$ such that following holds: for $c_h \in \mathbb{C}$,

$$R_g\Big(\sum_{h \in G} c_h h\Big) = \sum_{h \in G} c_h (gh) = \sum_{x \in G} c_{g^{-1}x} x.$$

Remark. Observe that the dimension of regular representation is $|G|$.

Theorem 1.4.8. *Let $\chi_1, \chi_2, \cdots, \chi_n$ be a complete set of irreducible characters of a finite group G. The irreducible characters form an orthonormal basis for the space of class functions on G.*

Proof. Since we showed the set $\{\chi_1, \chi_2, \cdots, \chi_n\}$ is orthonormal from the Schur's first orthogonality relations, it remains for us to show the set spans the space of class functions . Let W be a space constructed from a basis $\{\chi_1, \chi_2, \cdots, \chi_n\}$. Then the direct sum $W \oplus W^\perp$ is the space of class function on G. Showing $W^\perp = 0$ proves the set $\{\chi_1, \chi_2, \cdots, \chi_n\}$ spans space of class functions.

Let f be a class function on G such that f is orthogonal to all irreducible characters. Let φ be an irreducible representation of G with a representation space V. We define a map f_φ from V to V:

$$f_\varphi := \sum_{g\in G} f(g)\varphi_g.$$

We claim $f_g \in \mathrm{Hom}_G(\varphi, \varphi)$:

$$\begin{aligned}
\varphi_{y^{-1}} f_\varphi \varphi_y &= \varphi_{y^{-1}} \Big(\sum_{g\in G} f(g)\varphi_g\Big)\varphi_y \\
&= \sum_{g\in G} f(g)\varphi_{y^{-1}}\varphi_g\varphi_y \\
&= \sum_{g\in G} f(g)\varphi_{y^{-1}gy}.
\end{aligned}$$

Since f is a class function, $f(g)$ equals to $f(y^{-1}gy)$ for all $y \in G$. Hence

$$\varphi_{y^{-1}} f_\varphi \varphi_y = \sum_{g\in G} f(y^{-1}gy)\varphi_{y^{-1}gy}.$$

The variable $y^{-1}gy$ sums over all the elements in G as g iterates. Therefore,

$$\varphi_{y^{-1}} f_\varphi \varphi_y = f_\varphi$$

and this shows that $f_\varphi \in \mathrm{Hom}_G(\varphi, \varphi)$. Applying Schur's Lemma gives $f_\varphi = \lambda\mathcal{I}$. Then

$$\mathrm{tr}\,(f_\varphi) = \lambda \dim V. \tag{1.18}$$

Equivalently, the trace $\mathrm{tr}\,(f_\varphi)$ is

$$\begin{aligned}\mathrm{tr}\,(f_\varphi) &= \sum_{g\in G} f(g)\,\mathrm{tr}\,(\varphi_g) \\ &= \sum_{g\in G} f(g)\chi_\varphi(g).\end{aligned}$$

Substituting the above equation to equation 1.18 gives

$$\begin{aligned}\lambda &= \frac{1}{\dim V}\sum_{g\in G} f(g)\chi_\varphi(g) \\ &= \frac{|G|}{\dim V}\langle f, \overline{\chi_\varphi}\rangle.\end{aligned}$$

Recalling that the function f is orthogonal to all irreducible characters yields $\lambda = 0$ which proves $f_\varphi = 0$.

Now let ψ be a regular representation of the group G with representation space W. Let $\{e_g | g \in G\}$ be a basis for W. Then

$$f_\psi e_e = \sum_{g\in G} f(g)\psi_{(g)}e_e$$

where e is the identity element in G. Since ψ is the regular representation

$$f_\psi e_e = \sum_{g\in G} f(g)e_g.$$

Previously, we showed $f_\psi = 0$ where ψ is irreducible representation. Hence,

$$0 = \sum_{g\in G} f(g)e_g.$$

Recall that $\{e_g | g \in G\}$ form a basis for W. Thus we must have $f(g) = 0$ for all g in G. Hence any class function f such that the function f is orthogonal to all the set of irreducible characters is a zero function. Therefore, $W^\perp = 0$ and the set of irreducible characters span the space of class functions. This completes the proof. $\square$

Corollary 1.4.9. *The number of conjugacy classes of a finite group G equals to the number of equivalence classes of irreducible representations of G.*

Proof. Let $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_n$ be a complete set of conjugacy classes of G. From Theorem 1.4.4, we know the set $\{\delta_{\mathcal{C}_i} | 1 \leq i \leq n\}$ forms a basis for the space of class functions of G. Also from Theorem 1.4.8, we know the complete set of irreducible characters form a basis for the space of class functions of G. Since the numbers of elements in both bases are the same, the number of conjugacy classes equals the number of equivalence classes of irreducible representations. □

Corollary 1.4.10. *Decomposition of a representation φ of a finite group G into a direct sums of irreducibles is unique up to isomorphism. If φ_i is an irreducible representation, then the multiplicity of φ_i in φ equals to $\langle \chi_\varphi, \chi_{\varphi_i} \rangle$.*

Proof. Let

$$\varphi \sim \varphi_1^{\alpha_1} \oplus \varphi_2^{\alpha_2} \oplus \cdots \oplus \varphi_n^{\alpha_n}$$

where φ_i is an irreducible representation and α_i is a multiplicity of it for each i. Suppose there is another decomposition of φ into irreducibles such that

$$\varphi \sim \psi_1^{\beta_1} \oplus \psi_2^{\beta_2} \oplus \cdots \oplus \psi_n^{\beta_n}.$$

We construct the map from φ to φ; note this map is obviously nonzero. We apply Schur's Lemma; the map sends $\varphi_i^{\alpha_i}$ to $\psi_j^{\beta_j}$ and $\varphi_i \sim \psi_j$ where $\alpha_i = \beta_j$. Hence, the decomposition is unique up to isomorphism.

For the second part of the corollary, we use Schur's First Orthogonality Relations. Recall that $\langle \chi_{\varphi_i}, \chi_{\varphi_j} \rangle = 0$ when $\varphi_i \nsim \varphi_j$. Then

$$\begin{aligned}\langle \chi_\varphi, \chi_{\varphi_i} \rangle &= \alpha_1 \langle \chi_{\varphi_1}, \chi_{\varphi_i} \rangle + \alpha_2 \langle \chi_{\varphi_2}, \chi_{\varphi_i} \rangle + \cdots + \alpha_n \langle \chi_{\varphi_n}, \chi_{\varphi_i} \rangle \\ &= 0 + 0 + \cdots + \alpha_i \cdot 1 + 0 + \cdots + 0.\end{aligned}$$

This completes the proof. □

Remark. By the Corollary 1.4.10, we can decompose a representation φ of a finite group into distinct irreducible representations $\varphi \sim \varphi_1^{\alpha_1} \oplus \varphi_2^{\alpha_2} \oplus \cdots \oplus \varphi_n^{\alpha_n}$. Then $\langle \chi_\varphi, \chi_\varphi \rangle = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2$. We note that $\langle \chi_\varphi, \chi_\varphi \rangle$ is 1 if and only if only one of α_i equals to 1 and others are all 0; this happens if and only if when φ is an irreducible representation. Hence, we conclude that φ is irreducible if and only if $\langle \chi_\varphi, \chi_\varphi \rangle = 1$.

1.5 Schur's Second Orthogonality Relations

We briefly introduced a regular representation in the previous section. Regular representation is constructed from a multiplication table of a group. Studying regular representation gives us insight when decomposing representations into irreducibles. We will see that the regular representation contains all the irreducible representations and will also prove some fundamental statements related to it. At the end of this section, we give a proof of Schur's second orthogonality relations. These relations show that the columns of the character table are orthogonal, and therefore the character table is invertible. We will also see we can construct the character table even if we do not know all the irreducible characters.

Definition 1.5.1. ($\mathbb{C}G$). Given a group G, $\mathbb{C}G$ is a set of all linear combinations of G with coefficients in $\mathbb{C}$. Suppose $a = \sum_{g\in G} a_g g$ and $b = \sum_{g\in G} b_g g$. The operation of addition is given by

$$a + b = \sum_{g\in G} a_g g + \sum_{g\in G} b_g g = \sum_{g\in G} (a_g + b_g)g.$$

The scalar multiplication is given by

$$ka = \sum_{g\in G} k a_g g.$$

for k in $\mathbb{C}$. Lastly, the inner product is given by

$$\langle \sum_{g\in G} a_g g, \sum_{g\in G} b_g g\rangle = \sum_{g\in G} a_g \overline{b_g}.$$

Definition 1.5.2. (Regular Representation). The (left) regular representation R of a finite group G is a homomorphism from G to $GL(\mathbb{C}G)$ such that following holds: for $c_h \in \mathbb{C}$

$$R_g\Big(\sum_{h\in G} c_h h\Big) = \sum_{h\in G} c_h (gh).$$

As previously mentioned, the regular representation is constructed from the multiplication table of group. For example, let G be a symmetric group S_3. The multiplication table is given below.

$\cdot$	$()$	$(1,2)$	$(2,3)$	$(1,3)$	$(1,2,3)$	$(1,3,2)$
$()$	$()$	$(1,2)$	$(2,3)$	$(1,3)$	$(1,2,3)$	$(1,3,2)$
$(1,2)=(1,2)^{-1}$	$(1,2)$	$()$	$(1,2,3)$	$(1,3,2)$	$(2,3)$	$(1,3)$
$(2,3)=(2,3)^{-1}$	$(2,3)$	$(1,3,2)$	$()$	$(1,2,3)$	$(1,3)$	$(1,2)$
$(1,3)=(1,3)^{-1}$	$(1,3)$	$(1,2,3)$	$(1,3,2)$	$()$	$(1,2)$	$(2,3)$
$(1,3,2)=(1,2,3)^{-1}$	$(1,3,2)$	$(2,3)$	$(1,3)$	$(1,2)$	$()$	$(1,2,3)$
$(1,2,3)=(1,3,2)^{-1}$	$(1,2,3)$	$(1,3)$	$(1,2)$	$(2,3)$	$(1,3,2)$	$()$

Observe that the multiplication table above is constructed in a way that the identity element $()$ is located on the diagonal. Suppose the i-th element on the colunm is g. Then we place g^{-1} on the i-th row.

The associated regular representation R_g is $|G|\times|G|$ matrix with 1 where the element g presents in the multiplication table, and 0 otherwise. For example,

$$R_{(1,2,3)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Proposition 1.5.1. *Regular representation is a representation.*

Proof. Let G be a group with n elements, and let R be a regular representation of G such that g_i denotes the i-th element of the column. Then i-th element of the row is the inverse of g_i. We want to show

$$R_{g_i g_j} = R_{g_i} R_{g_j}.$$

Observe that $(R_{g_i g_j})_{xy} = 1$ if and only if $g_x^{-1} g_y = g_i g_j$. For $R_{g_i} R_{g_j}$,

$$(R_{g_i} R_{g_j})_{xy} = \sum_z (R_{g_i})_{xz} (R_{g_j})_{zy}.$$

Then $(R_{g_i} R_{g_j})_{xy} = 1$ if and only if $g_x^{-1} = g_i g_z^{-1}$ and $g_y = g_z g_j$. Hence,

$$g_x^{-1} g_y = g_i g_z^{-1} g_z g_j = g_i g_j$$

which implies $R_{g_ig_j} = R_{g_i}R_{g_j}$. This shows the regular representation is a representation. $\square$

Proposition 1.5.2. *Every regular representation of a finite group G is unitary.*

Proof. Let $\sum_{x\in G} a_x x, \sum_{x\in G} b_x x \in \mathbb{C}G$. Then for arbitrary element g in G,

$$\langle R_g \sum_{x\in G} a_x x, R_g \sum_{x\in G} b_x x\rangle = \langle \sum_{x\in G} a_x(gx), \sum_{x\in G} b_x(gx)\rangle.$$

The change of variables $gx \to h$ yields

$$\begin{aligned}\langle R_g \sum_{x\in G} a_x x, R_g \sum_{x\in G} b_x x\rangle &= \langle \sum_{h\in G} a_{g^{-1}h} h, \sum_{h\in G} b_{g^{-1}h} h\rangle \\ &= \sum_{h\in G} a_{g^{-1}h}\overline{b_{g^{-1}h}}.\end{aligned}$$

Again, the change of variables $g^{-1}h \to x$ leaves

$$\langle R_g \sum_{x\in G} a_x x, R_g \sum_{x\in G} b_x x\rangle = \langle \sum_{x\in G} a_x x, \sum_{x\in G} b_x x\rangle.$$

Hence, every regular representation of a finite group G is unitary. $\square$

Proposition 1.5.3. *The character χ_{reg} of regular representation R of a finite group G is given as follows; for $g \in G$*

$$\chi_{reg}(g) = \begin{cases} |G| & \textit{if } g{=}e, \\ 0 & \textit{otherwise}. \end{cases}$$

Proof. Let g_i denote the i-th element of the column of R. Then it follows that g_i^{-1} is the i-th element of the row. Observe that $(R_g)_{ii} = 1$ if and only if $g_i^{-1}g_i = g$ or equivalently, if and only if when $e = g$. Since R is $|G|{\times}|G|$ matrix, the trace of R_e equals to $|G|$. $\square$

Proposition 1.5.4. *The regular representation of a finite group G is not irreducible unless the group is trivial.*

Proof. Let χ_{reg} be a regular representation of a finite group G. The inner product $\langle \chi_{reg}, \chi_{reg}\rangle$ is

$$\langle \chi_{reg}, \chi_{reg}\rangle = \frac{1}{|G|}\sum_{g\in G} \chi_{reg}(g)\chi_{reg}(g).$$

From Proposition 1.5.3, $\chi_{reg(g)} = |G|$ when $g = e$; otherwise $\chi_{reg(g)} = 0$. Thus

$$\begin{aligned}\langle \chi_{reg}, \chi_{reg} \rangle &= \frac{1}{|G|} \chi_{reg}(e) \chi_{reg}(e) \\ &= \frac{1}{|G|} |G||G| = |G|.\end{aligned}$$

Recalling that a representation φ is irreducible if and only if $\langle \chi_\varphi, \chi_\varphi \rangle = 1$ shows the regular representation R is irreducible if and only if $|G| = 1$. $\square$

Theorem 1.5.1. *Let R be the regular representation of a finite group G such that*

$$R \sim \varphi_1^{\alpha_1} \oplus \varphi_2^{\alpha_2} \oplus \cdots \varphi_n^{\alpha_n}$$

where φ_i is irreducible representation for all i. Then the multiplicity α_i of φ_i equals to the degree of φ_i.

Proof. Recall from Schur's first orthogonality relations that $\langle \chi_{\varphi_i}, \chi_{\varphi_j} \rangle = 0$ when $\varphi_i \nsim \varphi_j$. Then

$$\begin{aligned}\langle \chi_R, \chi_{\varphi_i} \rangle &= \alpha_1 \langle \chi_{\varphi_1}, \chi_{\varphi_i} \rangle + \alpha_2 \langle \chi_{\varphi_2}, \chi_{\varphi_i} \rangle + \cdots + \alpha_n \langle \chi_{\varphi_n}, \chi_{\varphi_i} \rangle \\ &= 0 + 0 + \cdots + \alpha_i \cdot 1 + 0 + \cdots + 0.\end{aligned}$$

Observe that

$$\begin{aligned}\langle \chi_R, \chi_i \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi_R(g) \overline{\chi_i(g)} \\ &= \frac{1}{|G|} \chi_R(e) \overline{\chi_i}(e) \\ &= \frac{1}{|G|} |G| \deg \varphi_i \\ &= \deg \varphi^{(i)}.\end{aligned}$$

Hence we conclude that the multiplicity of φ_i equals to the degree of φ_i. $\square$

Remark. Recall, from Corollary 1.4.10, the decomposition of a representation R of a finite group G into a direct sum of irreducibles is unique up to isomorphism; this also applies to regular representation.

Theorem 1.5.2. *Let $\{\varphi_1, \varphi_2, \cdots, \varphi_n\}$ be a complete set of equivalence classes of irreducible representations of a finite group G. Then*

$$\sum_{i=1}^{k} (\deg \varphi_i)^2 = |G|.$$

Proof. Let R be a regular representation of a finite group G such that

$$R \sim \varphi_1^{\alpha_1} \oplus \varphi_2^{\alpha_2} \oplus \cdots \varphi_n^{\alpha_n}.$$

By Proposition 1.5.3,

$$\begin{aligned} |G| &= \chi_R(e) \\ &= \alpha_1 \chi_{\varphi_1}(e) + \alpha_2 \chi_{\varphi_2}(e) + \cdots + \alpha_n \chi_{\varphi_n}(e). \end{aligned}$$

where α_i is a multiplicity of φ_i. From Theorem 1.5.1, we know $\alpha_i = \deg \varphi_i$, and by the definition $\deg \varphi_i = \chi_{\varphi_i}(e)$. After substituting α_i's and $\chi_{\varphi_i}(e)$'s, we get

$$|G| = \deg {\varphi_1}^2 + \deg {\varphi_2}^2 + \cdots + \deg {\varphi_n}^2,$$

as desired. □

Corollary 1.5.3. *Regular representation of a group G contains all the irreducible representations.*

Proof. This is an immediate consequence of Theorem 1.5.1 and Theorem 1.5.2. □

Given a representation φ of a finite group G on a complex vector space V, we showed that if $\varphi \sim \varphi_1 \oplus \varphi_2 \oplus \cdots \oplus \varphi_n$ where φ_i is irreducible for all i, then $\{\chi_{\varphi_1}, \chi_{\varphi_2}, \cdots, \chi_{\varphi_n}\}$ forms a basis of the space of class functions on G. Recall that $\{\delta_C | C \in Cl(G)\}$ is also a basis of the space of class functions on G. The following theorem uses a relation between those two bases.

Theorem 1.5.4. *(Schur's Second Orthogonality Relations). Let φ be a representation of a finite group G such that $\varphi \sim \varphi_1 \oplus \cdots \oplus \varphi_n$ where φ_i is irreducible for all i. Then for g, h in G,*

$$\sum_i \chi_{\varphi_i}(g)\overline{\chi_{\varphi_i}(h)} = \begin{cases} \frac{|G|}{|C_G(h)|} & \textit{if } g,\ h \textit{ are conjugate,} \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. We proved that $\{\delta_{\mathcal{C}} | \mathcal{C} \in Cl(G)\}$ and $\{\chi_{\varphi_1}, \chi_{\varphi_2}, \cdots, \chi_{\varphi_n}\}$ are two bases for the space of class functions. Hence, we can rewrite $\delta_{C_G(h)}(g)$ in terms of irreducible characters, here $C_G(h)$ denotes the conjugacy class of h :

$$\begin{aligned}
\delta_{C_G(h)}(g) &= \sum_{i=1}^{n} \langle \delta_{C_G(h)}, \chi_{\varphi_i} \rangle \chi_{\varphi_i}(g) \\
&= \sum_{i=1}^{n} \frac{1}{|G|} \sum_{g' \in G} \delta_{C_G(h)}(g') \overline{\chi_{\varphi_i}(g')} \chi_{\varphi_i}(g) \\
&= \sum_{i=1}^{n} \frac{1}{|G|} \sum_{g' \in G} \delta_{C_G(h)}(g') \chi_{\varphi_i}(g'^{-1}) \chi_{\varphi_i}(g) \\
&= \sum_{i=1}^{n} \chi_{\varphi_i}(g) \frac{1}{|G|} \Big(\sum_{g' \in G} \delta_{C_G(h)}(g') \chi_{\varphi_i}(g'^{-1}) \Big) \\
&= \sum_{i=1}^{n} \chi_{\varphi_i}(g) \frac{1}{|G|} |C_G(h)| \chi_{\varphi_i}(h^{-1}) \\
&= \frac{|C_G(h)|}{|G|} \sum_{i=1}^{n} \chi_{\varphi_i}(g) \chi_{\varphi_i}(h^{-1}).
\end{aligned}$$

Hence

$$\sum_{i=1}^{n} \chi_{\varphi_i}(g) \chi_{\varphi_i}(h^{-1}) = \frac{|G|}{|C_G(h)|} \delta_{C_G(h)}(g) \tag{1.19}$$

In the case g, h are conjugate, $\delta_{C_G(h)}(g) = 1$. Then substituting 1 to $\delta_{C_G(h)}(g)$ in equation 1.19 yields

$$\sum_{i=1}^{n} \chi_{\varphi_i}(g) \chi_{\varphi_i}(h^{-1}) = \frac{|G|}{|C_G(h)|}.$$

In the case g, h are not conjugate, $\delta_{C_G(h)}(g) = 0$. Substituting 0 to $\delta_{C_G(h)}(g)$ in equation 1.19 yields

$$\sum_{i=1}^{n} \chi_{\varphi_i}(g) \chi_{\varphi_i}(h^{-1}) = 0.$$

This completes the proof. $\square$

1.6 Character Table for S_3

Character table is a two-dimensional table with $|Cl(G)|\times|Cl(G)|$ entries. The column of the character table corresponds to the conjugacy classes of G, and the row corresponds to the irreducible characters. In this section, we will construct the character table for S_3.

Recall that there are 6 elements in S_3:

$$S_3 = \{(1)\} \cup \{(1,2)(2,3)(3,1)\} \cup \{(1,2,3)(1,3,2)\}.$$

Observe that there are three conjugacy classes of S_3:

$$\{(1)\}, \{(1,2)(2,3)(3,1)\}, \{(1,2,3)(1,3,2)\}.$$

This implies there are exactly 3 irreducible representations of S_3; our goal is to calculate characters of those representations. Let φ_1, φ_2 and φ_3 denote three irreducible representations of S_3. Recall, from Theorem 1.5.2, the sum of degrees of all the irreducible representation of a finite group equals to the order of the group. Hence,

$$\deg {\varphi_1}^2 + \deg {\varphi_2}^2 + \deg {\varphi_3}^2 = |S_3| = 6.$$

All representations have a trivial representation. Let φ_1 be a trivial representation. Then $\varphi_1(g) = 1$ for all $g \in S_3$ and $\deg \varphi_1 = 1$. Substituting 1 for $\deg {\varphi_1}^2$ gives

$$1 + \deg {\varphi_2}^2 + \deg {\varphi_3}^2 = 6.$$

The only possible positive integer solution for above equation is $\deg \varphi_2 = 1$ and $\deg \varphi_3 = 2$. With given information, we can construct the character table as below:

	$\{(1)\}$	$\{(1,2)(2,3)(3,1)\}$	$\{(1,2,3)(1,3,2)\}$
χ_{φ_1}	1	1	1
χ_{φ_2}	1	x	y
χ_{φ_3}	2	z	w

To complete the character table recall following orthogonality relations. From Schur's first orthogonality relations (Theorem 1.4.7), the inner product $\langle \chi_{\varphi_i}, \chi_{\varphi_i} \rangle = 1$ if and only if $i \neq j$, and 0 otherwise. Therefore, the rows of the character table are orthogonal. Also, from Schur's second orthogonality relations (Theorem 1.5.4), we know the columns of the character table are orthogonal. Now we apply orthogonal relations to find x, y, z, and w.

Since the first and second rows are orthogonal:

$$1 + 3 \cdot x + 2 \cdot y = 0. \tag{1.20}$$

Applying Schur's First Orthogonality Relations to the second row gives

$$\begin{aligned} \langle \chi_{\varphi_2}, \chi_{\varphi_2} \rangle &= \frac{1}{6}(1 + 3x^2 + 2y^2) \\ &= 1. \end{aligned} \tag{1.21}$$

Observe that $(1,2)(1,2) = (1)$ and $(1,2,3)(1,2,3) = (1,3,2)$. Hence,

$$x^2 = 1, y^2 = y.$$

Substituting 1 for x and y for y^2 in euqations 1.20 and 1.21 leaves $x = -1$ and $y = 1$. Now the character table is

	$\{(1)\}$	$\{(1,2)(2,3)(3,1)\}$	$\{(1,2,3)(1,3,2)\}$
χ_{φ_1}	1	1	1
χ_{φ_2}	1	-1	1
χ_{φ_3}	2	z	w

We use the column orthogonality relations to find z and w. Since the first and second columns are orthogonal to each other,

$$1 \cdot 1 + 1 \cdot (-1) + 2 \cdot z = 0.$$

Also, the first and third columns are orthogonal. Therefore

$$1 \cdot 1 + 1 \cdot 1 + 2 \cdot w = 0.$$

Then $z = 0$ and $w = -1$ and this leaves

	$\{(1)\}$	$\{(1,2)(2,3)(3,1)\}$	$\{(1,2,3)(1,3,2)\}$
χ_{φ_1}	1	1	1
χ_{φ_2}	1	-1	1
χ_{φ_3}	2	0	-1

2

Fourier Transform on Finite Abelian Groups

The Fourier transform is widely used in signal processing and probability. In this journal we document some of the basic statements in Fourier transform that are necessary when studying the random walks. Before introducing the Fourier transform, we study the dual group $\widehat{G}$. We will see G, a finite abelian group, is isomorphic to its dual $\widehat{G}$. We will show $\widehat{G}$ forms an orthonormal basis for $\mathbb{C}^G$. As a result, we can we express a function $f : Z \to \mathbb{Z}/n\mathbb{Z}$ with linear combinations of elements of $\widehat{G}$. After establishing the Fourier transform and its inversion, we will give a well-known application of it, the Plancherel Theorem.

2.1 Dual Group $\widehat{G}$

Recall the definition of a (linear) character. A linear character is a homomorphism $\chi : G \to S^1$, where S^1 is a unit circle. Given a finite abelian group G, its dual group $\widehat{G}$ is a set of all characters with multiplication as a binary operation. In this section, we will show that G is isomorphic to $\widehat{G}$. By showing $G \cong \widehat{G}$ and recalling distinct characters form an orthonormal set, we conclude that $\widehat{G}$ form an orthonormal basis for $\mathbb{C}^G$.

Theorem 2.1.1. *Let G be a finite abelian group. Then $G \cong \widehat{G}$.*

Proof. Recall that a finite abelian group is isomorphic to direct products of cyclic groups. So suppose,

$$G \cong H_1 \times H_2 \times \cdots \times H_n.$$

where H_i is cyclic for all i. We claim $H_i \cong \widehat{H_i}$ for all i. Since H_i is cyclic, H_i is abelian. Recall that when H_i is abelian, the number of distinct irreducible characters equals to $|H_i|$. Therefore, $|H_i| = |\widehat{H_i}|$, and this implies H_i is isomorphic to $\widehat{H_i}$. Also, observe that $\widehat{H_i}$ is cyclic. Suppose h generates H_i and set $\chi(h^x) = \chi(h)^x = e^{2\pi i x/|H_i|}$ for $\chi \in \widehat{H_i}$. It follows that χ generates $\widehat{H_i}$. So far, we have proved

$$G \cong H_1 \times H_2 \times \cdots \times H_n \cong \widehat{H_1} \times \widehat{H_2} \times \cdots \times \widehat{H_n}. \tag{2.1}$$

It remains for us to show $\widehat{H_1} \times \widehat{H_2} \times \cdots \times \widehat{H_n} \cong \widehat{G}$. In order to show that, we claim $\widehat{A \times B} \cong \hat{A} \times \hat{B}$ for a finite abelian group A and B. Suppose $\chi \in \widehat{A \times B}$. Then let χ_A and χ_B be character restricted to A and B respectively; $\chi_A(a) = \chi(a, 1)$ and similarly $\chi_B = \chi(1, b)$. Define a map T as follows:

$$\begin{aligned} T : \widehat{A \times B} &\to \hat{A} \times \hat{B} \\ \chi &\mapsto (\chi_A, \chi_B) \end{aligned}$$

Then the map T is a homomorphism: for $\chi, \theta \in \widehat{A \times B}$

$$\begin{aligned} T(\chi\theta) &= ((\chi\theta)_A, (\chi\theta)_B) \\ &= (\chi_A\theta_A, \chi_B\theta_B) \\ &= (\chi_A, \chi_B)(\theta_A, \theta_B) \\ &= T(\chi) \cdot T(\theta). \end{aligned}$$

Observe that only trivial character is the kernel of the map T. This is because

$$\chi(a, b) = \chi_A(a)\chi_B(b) = \chi(a, 1)\chi(1, b) = 1$$

for all $a \in A$ and $b \in B$ if and only if χ is a trivial character. Hence, the map T is injective. Lastly, observe that for all $(\chi_A, \chi_B) \in \hat{A} \times \hat{B}$, there exists a character $\chi \in \widehat{A \times B}$, where $\chi = (\chi_A, \chi_B)$. Thus, T is surjective. This establishes the isomorphism between $\widehat{A \times B}$ and $\hat{A} \times \hat{B}$. Applying this fact to equation 2.1 yields

$$G \cong H_1 \times H_2 \times \cdots \times H_n \cong (H_1 \times \widehat{H_2 \times \cdots} \times H_n) \cong \widehat{G}$$

as desired. $\square$

Remark. Recall from Schur's First Orthogonality Relations that the set of distinct characters forms an orthonormal set. Theorem 2.1.1 implies $|G| = |\widehat{G}|$. Thus the set of characters form an orthonormal basis for $Z(\mathbb{C}^G)$; since we are working with a finite abelian group, $Z(\mathbb{C}^G) = \mathbb{C}^G$.

2.2 Fourier Analysis on Finite Abelian Groups

Having observed that $\widehat{G}$, a set of characters, forms an orthonormal basis for $\mathbb{C}^G$, where G is a finite abelian group, we now introduce the Fourier transform. The Fourier transform is a map from $\mathbb{C}^G$ to $\mathbb{C}^{\hat{G}}$. We will see the Fourier transform is a ring isomorphism that takes the convolution into multiplication; this only applies to a finite abelian group; further discussions are needed to establish the Fourier transform for the non-abelian group. At the end of this section, we will give a Plancherel Theorem as an application.

Definition 2.2.1. (Discrete Fourier Transform). Let f be a function from $\mathbb{Z}/n\mathbb{Z}$ to $\mathbb{C}$. The discrete Fourier transforms $\widehat{f}$ of a function f is given by

$$\begin{aligned}\widehat{f}([k]) &= |n|\langle f, \chi_k\rangle \\ &= \sum_{x=0}^{n-1} f([x])e^{-2\pi ikx/n} \\ &= \sum_{x=0}^{n-1} f([x])\big(\cos(2\pi kx/n) - i\sin(2\pi kx/n)\big).\end{aligned}$$

Definition 2.2.2. (Fourier Transform on Finite Abelian Group). Let G be a finite abelian group. The Fourier transform is a map $\mathcal{F} : \mathbb{C}^G \to \mathbb{C}^{\hat{G}}$ such that

$$\mathcal{F}(f) = \widehat{f}(\chi) = \sum_{g\in G} f(g)\overline{\chi(g)}.$$

Theorem 2.2.1. *Let G be a finite abelian group and $f \in \mathbb{C}^G$. Then the Fourier inversion is given by*

$$f = \frac{1}{|G|}\sum_{\chi\in\hat{G}} \widehat{f}(\chi)\chi.$$

Proof. Recall that $\widehat{G}$ forms a basis for $\mathbb{C}^G$. Then

$$\begin{aligned} f &= \sum_{\chi \in \widehat{G}} \langle f, \chi \rangle \chi \\ &= \sum_{\chi \in \widehat{G}} \frac{1}{|G|} \sum_{g \in G} f(g)\overline{\chi(g)}\chi. \end{aligned}$$

Since $\widehat{f}(\chi)$ is $\sum_{g \in G} f(g)\overline{\chi(g)}$, we get

$$f = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\chi$$

and this completes the proof. □

Definition 2.2.3. (Fourier Transform). Let f be an integrable function from $\mathbb{R}$ to $\mathbb{C}$. The Fourier transform of f is given by

$$\mathcal{F}\{f\}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x \xi} dx$$

where $\xi \in \mathbb{R}$.

Definition 2.2.4. (Inverse Fourier Transform). Let f be an integrable function from $\mathbb{R}$ to $\mathbb{C}$. Then

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\xi)e^{2\pi i x \xi} d\xi.$$

Lemma 2.2.2. *The inverse Fourier transform $\mathcal{F}^{-1}$ satisfies following equation:*

$$\mathcal{F}^{-1} = \mathcal{F}\mathcal{R} = \mathcal{R}\mathcal{F}$$

where $\mathcal{R}\{f\}(x) := f(-x)$.

Proof. From the definition of the Fourier transform and Fourier inversion,

$$\begin{aligned} \mathcal{F}\mathcal{R}\{\widehat{f}(\xi)\} &= \mathcal{F}\{\widehat{f}(-\xi)\} \\ &= \int_{-\infty}^{\infty} \widehat{f}(x)e^{2\pi i x \xi} dx \\ &= f(\xi) \\ &= \mathcal{F}^{-1}\{\widehat{f}(\xi)\}. \end{aligned}$$

Thus $\mathcal{F}^{-1} = \mathcal{F}\mathcal{R}$. It is obvious that $\mathcal{F}\mathcal{R} = \mathcal{R}\mathcal{F}$ from the definition of the Fourier transform and the flip operator $\mathcal{R}$. This completes the proof. □

Remark. The inverse Fourier transform is identical to Fourier transform except for the flip operator $\mathcal{R}$. Thus proving a statement for the inverse Fourier transforms automatically shows the statement is also valid for the Fourier transform.

Theorem 2.2.3. *(Convolution Identity). Functions $a, b \in \mathbb{C}^G$ satisfies following identities:*

$$\begin{aligned} &(1) \quad \mathcal{F}\{a * b\} = \mathcal{F}\{a\} \cdot \mathcal{F}\{b\} \\ &(2) \quad \mathcal{F}\{a\} * \mathcal{F}\{b\} = \mathcal{F}\{a \cdot b\}. \end{aligned}$$

Proof. (1): By the definition of the convolution and the Fourier transform,

$$\begin{aligned} \widehat{a * b}(\chi) &= \sum_{g \in G} a * b(g)\overline{\chi(g)} \\ &= \sum_{g \in G} \sum_{h \in G} a(gh^{-1})b(h)\overline{\chi(g)} \\ &= \sum_{h \in G} b(h) \sum_{g \in G} a(gh^{-1})\overline{\chi(g)}. \end{aligned}$$

After the change of variables $gh^{-1} \to g'$, we get

$$\begin{aligned} \widehat{a * b}(\chi) &= \sum_{h \in G} b(h) \sum_{g' \in G} a(g')\overline{\chi(g'h)} \\ &= \sum_{h \in G} b(h) \sum_{g'} a(g')\overline{\chi(g')\chi(h)} \\ &= \sum_{h \in G} b(h)\overline{\chi(h)} \sum_{g' \in G} a(g')\overline{\chi(g')} \\ &= \sum_{g' \in G} a(g')\overline{\chi(g')} \sum_{h \in G} b(h)\overline{\chi(h)} \\ &= \widehat{a(\chi)} \cdot \widehat{b(\chi)}, \end{aligned}$$

the desired result.

(2): The first part of the convolution identity $\mathcal{F}\{a * b\} = \mathcal{F}\{a\} \cdot \mathcal{F}\{b\}$ can be rewritten as

$$\mathcal{F}\big(\mathcal{F}^{-1}(\mathcal{F}\{a\}) * \mathcal{F}^{-1}(\mathcal{F}\{b\})\big) = \mathcal{F}\{a\} \cdot \mathcal{F}\{b\}.$$

Applying $\mathcal{F}^{-1}$ on both sides of above equation yields

$$\mathcal{F}^{-1}(\mathcal{F}\{a\}) * \mathcal{F}^{-1}(\mathcal{F}\{b\}) = \mathcal{F}^{-1}\big(\mathcal{F}\{a\} \cdot \mathcal{F}\{b\}\big). \tag{2.2}$$

From Lemma 2.2.2, we know proving the second part of the convolution identity works for the inverse Fourier transform also proves this identity works for the (forward) Fourier transform. Therefore, equation 2.2 can be rewritten as

$$\mathcal{F}\{a\} * \mathcal{F}\{b\} = \mathcal{F}\{a \cdot b\}$$

as desired. $\square$

Corollary 2.2.4. *Let G be a finite abelian group. Then $(\mathbb{C}^G, +, *)$ and $(\mathbb{C}^{\widehat{G}}, +, \cdot)$ are rings, and the Fourier transform is a ring isomorphism from $\mathbb{C}^G$ to $\mathbb{C}^{\widehat{G}}$.*

Proof. We first observe the Fourier transform $\mathcal{F} : \mathbb{C}^G \to \mathbb{C}^{\widehat{G}}$ is linear: for $c_1, c_2 \in \mathbb{C}$, $a, b \in \mathbb{C}^G$, and $\chi \in \widehat{G}$,

$$\begin{aligned}
\mathcal{F}(c_1 a + c_2 b) &= (\widehat{c_1 a + c_2 b})(\chi) \\
&= \sum_{g \in G} (c_1 a + c_2 b)(g)\overline{\chi}(g) \\
&= c_1 \sum_{g \in G} a(g)\overline{\chi}(g) + c_2 \sum_{g \in G} b(g)\overline{\chi}(g) \\
&= c_1 \widehat{a}(\chi) + c_2 \widehat{b}(\chi) \\
&= c_1 \mathcal{F}(a) + c_2 \mathcal{F}(b).
\end{aligned}$$

Hence the Fourier transform is linear. Also, from Theorem 2.2.1, we know the Fourier transform is invertible. Since invertible linear map is isomorphism, we conclude that the map $\mathcal{F}$ is an isomorphism. Lastly, Theorem 2.2.3 asserts $\mathcal{F}(a * b) = \mathcal{F}(a) \cdot \mathcal{F}(b)$. This concludes that the Fourier transform is a ring isomorphism. $\square$

We can detect whether $x \equiv 0 \mod n$ or not by looking at the value of $\sum_{k=0}^{n-1} e^{2\pi i k(x-y)/n}$. Observe that

$$\sum_{k=0}^{n-1} e^{2\pi i k(x-y)/n} = \begin{cases} n & \text{if } x \equiv y \mod n, \\ 0 & \text{if } x \not\equiv y \mod n. \end{cases} \tag{2.3}$$

Given a function $f(x)$ on $\mathbb{Z}/n\mathbb{Z}$, the Fourier transform $\widehat{f}(k)$ is

$$\widehat{f}(k) = \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x) e^{-2\pi ikx/n}.$$

Applying the Fourier inversion yields

$$\begin{aligned} f(0) &= \frac{1}{n} \sum_{k \in \mathbb{Z}/n\mathbb{Z}} \widehat{f}(k) \\ &= \frac{1}{n} \sum_{k \in \mathbb{Z}/n\mathbb{Z}} \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x) e^{-2\pi ikx/n} \\ &= \sum_{x \in \mathbb{Z}/n\mathbb{Z}} f(x) \Big(\frac{1}{n} \sum_{k \in \mathbb{Z}/n\mathbb{Z}} e^{-2\pi ikx/n} \Big). \end{aligned}$$

We know from Equation 2.3 that the value of $\frac{1}{n} \sum_{k \in \mathbb{Z}/n\mathbb{Z}} e^{-2\pi ikx/n}$ equals 1 if and only if $x = 0$, and 0 otherwise. Thus, the function $\frac{1}{n} \sum_{k \in \mathbb{Z}/n\mathbb{Z}} e^{-2\pi ikx/n}$ works as a tool that detects whether x equals to 0 or not.

Theorem 2.2.5. *(Parseval Identity). Let $f, g \in \mathbb{C}^G$ where G is a finite abelian group. Then*

$$\langle f, g \rangle = \frac{1}{|G|} \langle \widehat{f}, \widehat{g} \rangle.$$

Proof. We make substitution $f = \frac{1}{|G|} \sum_{x \in \widehat{G}} \widehat{f}(\chi)\chi$, $g = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{g}(\chi)\chi$ in the $\langle f, g \rangle$ to arrive at

$$\langle f, g \rangle = \Big\langle \frac{1}{|G|} \sum_{x \in \widehat{G}} \widehat{f}(\chi)\chi, \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{g}(\chi)\chi \Big\rangle.$$

which is equivalently rewritten as

$$\begin{aligned} \langle f, g \rangle &= \frac{1}{|G|^2} \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\widehat{g}(\chi) \\ &= \frac{1}{|G|} \Big(\frac{1}{|G|} \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\widehat{g}(\chi) \Big).. \end{aligned}$$

Since $|G|= |\widehat{G}|$ when G is finite abelian, we get

$$\begin{aligned}\langle f, g\rangle &= \frac{1}{|G|}\Big(\frac{1}{|\widehat{G}|}\sum_{\chi\in\widehat{G}} \widehat{f}(\chi)\widehat{g}(\chi)\Big)\\ &= \frac{1}{|G|}\langle \widehat{f}, \widehat{g}\rangle\end{aligned}$$

as desired. □

The immediate consequence of the Parseval identity is the Plancherel identity.

Corollary 2.2.6. *(Plancherel Identity). Let $f \in \mathbb{C}^G$ where G is a finite abelian group. Then*

$$||f||= ||\widehat{f}||.$$

Proof. This is the special case of Parseval's identity. □

3 Markov Chain

3.1 Probability Theory

Before introducing the Markov chain and random walk, the subject we are interested in, we review some of the basic concepts in probability theory.

Definition 3.1.1. (Sample Space). Given an experiment, the sample space Ω is a set of all possible outcomes of the experiment.

Definition 3.1.2. (Event). Let Ω be the sample space. The σ-algebra Σ in $\mathcal{P}(\Omega)$, the power set of Ω, is called an event.

Remark. Recall the definition of a σ-algebra. For $\Sigma \subseteq \mathcal{P}(\Omega)$, σ-algebra satisfies three properties: Ω is in Σ, Σ is closed under complement and countable unions.

Definition 3.1.3. (Axioms of Probability). Let Ω be the sample space of an experiment.
(1) The probability of an event E is a non-negative real number which satisfies following properties: for all $E \in S$

$$\mathbf{Prob}(E) \in \mathbb{R}, \quad 0 \leq \mathbf{Prob}(E) \leq 1.$$

(2) The probability of the sample space equals to 1:

$$\mathbf{Prob}(\Omega) = 1.$$

(3) For any countable sequence of disjoint events $E_1, E_2, \cdots$,

$$\mathbf{Prob}\big(\cup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} \mathbf{Prob}(E_i).$$

Definition 3.1.4. (Conditional Probability). For $F \in S$ with $\mathbf{Prob}(F) > 0$, the conditional probability of an event E given that event F has occurred is defined by

$$\mathbf{Prob}(E|F) = \frac{P(EF)}{P(F)}.$$

Definition 3.1.5. (Random Variable). A real-valued function on a sample space is called a random variable.

Definition 3.1.6. (Expectation). Let X be a random variable with probability mass function $f(x)$. The expected value of the random variable X is given by

$$E[X] = \int_{\mathbb{R}} x f(x) dx.$$

In case where X is a discrete random variable, the expectation of X is given by

$$E[X] = \sum_{i=1}^{\infty} x_i f(x_i).$$

We review some basic properties of expectation:

Proposition 3.1.1. *(Linearity of Expectation). Let $E[X]$ and $E[Y]$ be finite and k be a constant value. Then*

$$(1)\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$
$$(2)\mathbf{E}[kX] = k\mathbf{E}[X].$$

Proof. The proof is omitted. □

Proposition 3.1.2. *(The Law of Iterated Expectation). For random variables X and Y following holds;*

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]].$$

Proof. From the definition of the expectation, we get

$$\mathbf{E}[\mathbf{E}[X|Y]] = \int_{-\infty}^{\infty} \mathbf{E}[X|Y] f_Y(y) dy,$$

where f_Y is the probability density of Y. Again by the definition of the expectation, we get

$$\begin{aligned}\mathbf{E}[\mathbf{E}[X|Y]] &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx\right) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)dy\right)dx,\end{aligned}$$

where $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$. Then

$$\begin{aligned}\mathbf{E}[\mathbf{E}[X|Y]] &= \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\right)dx \\ &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \mathbf{E}[X]\end{aligned}$$

as we desired. □

Proposition 3.1.3. *For a non-negative discrete random variable X,*

$$\mathbf{E}[X] = \sum_{t\geq 1} \mathbf{Prob}(X \geq t).$$

Proof. Observe that

$$\begin{array}{lcccccccc} & \mathbf{Prob}(X \geq 1) & = & \mathbf{Prob}(X = 1) & + & \mathbf{Prob}(X = 2) & + & \mathbf{Prob}(X = 3) & + \cdots \\ & \mathbf{Prob}(X \geq 2) & = & & & \mathbf{Prob}(X = 2) & + & \mathbf{Prob}(X = 3) & + \cdots \\ & \mathbf{Prob}(X \geq 3) & = & & & & & \mathbf{Prob}(X = 3) & + \cdots \\ + & \vdots & & & & & & & \\ \hline \multicolumn{9}{l}{\quad \sum_{t\geq 1} \mathbf{Prob}(X \geq 1) = 1 \cdot \mathbf{Prob}(X = 1) + 2 \cdot \mathbf{Prob}(X = 2) + 3 \cdot \mathbf{Prob}(X = 3) + \cdots}\end{array}$$

Then

$$\sum_{t\geq 1} \mathbf{Prob}(X \geq 1) = \sum_{t=1}^{\infty} t \cdot \mathbf{Prob}(X = t) = \mathbf{E}[X]$$

as we wished. □

Definition 3.1.7. (Variance). The variance of a random variable X is given by

$$\mathbf{Var}X = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

Definition 3.1.8. (Indicator Random Variable). Let A be an event. The indicator random variable I_A is given by

$$I_A = \begin{cases} 1 & A \text{ occurs} \\ 0 & A \text{ does not occur} \end{cases}$$

Remark. Observe that the expectation $\mathbf{E}(I_A)$ is $\mathbf{E}(I_A) = 1 \cdot \mathbf{Prob}(A) + 0 \cdot \mathbf{Prob}(A) = \mathbf{Prob}(A)$.

Definition 3.1.9. (Uniform Distribution). The probability density function of uniform distribution $\mathbb{U}$ on the interval $[a, b]$ is given by

$$\mathbb{U} = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if} x < a \text{ or } x > b \end{cases}$$

where $b > a$

We introduce a classic inequality in probability theory, the Markov inequality:

Theorem 3.1.1. *(Markov Inequality). For a non-negative random variable X,*

$$\mathbf{Prob}(X > \lambda \cdot \mathbf{E}[X]) \leq \frac{1}{\lambda}$$

for all $\lambda \geq 1$.

Proof. Suppose X is finitely distributed with probability mass function $P(t) := \mathbf{Prob}(X = t)$. Then

$$\begin{aligned} \mathbf{E}[X] &= \sum_t t \cdot P(t) \\ &= \sum_{t \leq \lambda \cdot \mathbf{E}[X]} t \cdot P(t) + \sum_{t > \lambda \cdot \mathbf{E}[X]} t \cdot P(t) \\ &\geq \sum_{t > \lambda \cdot \mathbf{E}[X]} t \cdot P(t) \\ &\geq \lambda \cdot \mathbf{E}[X] \cdot \mathbf{Prob}(X > \lambda \cdot \mathbf{E}[X]). \end{aligned}$$

Hence $\mathbf{E}[X] \geq \Big(\lambda \cdot \mathbf{E}[X]\Big) \cdot \mathbf{Prob}(X > \lambda \mathbf{E}[X])$. Upon dividing both the left and the right hand sides by $\lambda \cdot \mathbf{E}[X]$ we have our desired result. $\square$

Corollary 3.1.2. *(Chebyshev's inequality). Let X be a random variable with $\mathbf{E}[X] = 0$ and $\mathbf{Var}[X] = d$. Then for any real number $\lambda > 0$,*

$$\mathbf{Prob}(X > \lambda\sqrt{d}) \leq \frac{1}{\lambda^2}.$$

Proof. The Chebyshev's inequality is a special case of the Markov inequality. Observe that

$$\mathbf{Prob}(X > \lambda\sqrt{d}) \leq \mathbf{Prob}(X^2 > \lambda^2 d).$$

Because $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \mathbf{E}[X^2] + 0 = d$, we have $\mathbf{E}[X^2] = d$. Applying the Markov inequality to $\mathbf{Prob}(X^2 > \lambda^2 d)$ yields

$$\begin{aligned}\mathbf{Prob}(X^2 > \lambda^2 d) &= \mathbf{Prob}(X^2 > \lambda^2 \cdot \mathbf{E}[X^2]) \\ &\leq \frac{\mathbf{E}[X^2]}{\lambda^2 d} \\ &= \frac{1}{\lambda^2}.\end{aligned}$$

Thus

$$\mathbf{Prob}(X > \lambda\sqrt{d}) \leq \mathbf{Prob}(X^2 > \lambda^2 d) \leq \frac{1}{\lambda^2}$$

as desired. □

3.2 Finite State Markov Chain

Definition 3.2.1. (Stochastic Matrix). Let P be a $n \times m$ matrix. The matrix P is stochastic if $\sum_{j=1}^{m} P_{ij} = 1$ for all i and j.

Lemma 3.2.1. *Let P be $n \times m$ matrix and Q be $m \times l$ matrix. If P and Q are stochastic, then their matrix multiplication PQ is also stochastic.*

Proof. From the definition,

$$[PQ]_{ij} = \sum_{k=1}^{m} P_{ik} Q_{kj}. \tag{3.1}$$

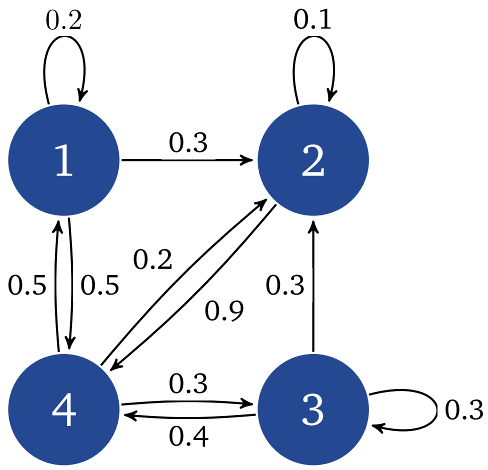


Fig. 3.1: Four States Markov Chain

The sum of i-th row of PQ is

$$\sum_{j=1}^{l}[PQ]_{ij} = \sum_{j=1}^{l}\sum_{k=1}^{m} P_{ik}Q_{kj}. \tag{3.2}$$

Since P and Q are stochastic, we know $\sum_{k=1}^{m} P_{ik} = 1 = \sum_{j=1}^{l} Q_{kj}$. Applying this to (3.2) yields

$$\begin{aligned}\sum_{j=1}^{l}[PQ]_{ij} &= \sum_{j=1}^{l}\sum_{k=1}^{m} P_{ik}Q_{kj} \\ &= \sum_{k=1}^{m} P_{ik} \sum_{j=1}^{l} Q_{kj} \\ &= \sum_{k=1}^{m} P_{ik} \cdot 1 \\ &= 1.\end{aligned}$$

Observe that the choice of i was arbitrary. This shows that PQ is stochastic. □

We now introduce Markov chain. A Markov chain is a sequence of random variables $(X_0, X_1, \cdots)$ with a Markov property; when a distribution of X_{n+1} only depends on X_n and independent on previous random variables $X_0, X_1, \cdots, X_{n-1}$ we say the random process satisfies Markov property.

A directed graph $G = (V, E)$ (Figure 3.1) is represented as an example of a four-state Markov chain; V is the set of vertices and E is the set of ordered edges. A set of vertices V represents states of the Markov chain and the directed edge from one state to the other state indicates a positive probabilities of moving from the one to the other.

Suppose there is an object moving according to Figure 3.1. For example, if the object is at state 2 currently, the probability of the object moving to the state 4 is 0.9, and the probability of the object remaining at the present position is 0.1.

Let $\mathcal{X}$ denote a state space of the Markov chain; in this case, $\mathcal{X} = \{1, 2, 3, 4\}$. Random variables $\{X_t\}_{t=1}^{\infty}$ is a function from $\mathbb{Z}_+ \cup \infty$ to $\mathcal{X}$. Let X_t denote the object's position at step t. The Markov property rules the behavior of random variables $\{X_t\}_{t=1}^{\infty}$. Suppose $X_t = x$ currently, and we want to know the probability of moving to state y in the next step. The Markov property tells us that the X_{t+1} only depends on the previous state X_t. Equivalently saying in symbols,

$$\mathbf{Prob}(X_{t+1} = x_{t+1} | X_1 = x_1, \cdots, X_t = x_t) = \mathbf{Prob}(X_{t+1} = x_{t+1} | X_t = x_t),$$

meaning that the previous states that the object have visited does not impact the probability distribution of next step; only present state does. Thus we get following equation:

$$\begin{aligned}&\mathbf{Prob}(X_0 = x_0, X_1 = x_1, \cdots, X_t = x_t)\\ &= \mathbf{Prob}(X_0 = x_0)\mathbf{Prob}(X_1 = x_1 | X_0 = x_0) \cdots \mathbf{Prob}(X_{n-1} = x_{n-1} | X_t = x_t)\end{aligned} \tag{3.3}$$

Definition 3.2.2. (Time-Homogeneous Markov Chain). A Markov chain is time-homogeneous when

$$\mathbf{Prob}(X_{t+1} = x | X_t = y) = \mathbf{Prob}(X_t = x | X_{t-1} = y).$$

Remark. We discuss only the time-homogeneous Markov chain in this journal. Throughout this section, we assume that the Markov chain is time-homogeneous.

Let us construct a transition matrix P of the Markov chain:

$$P(x, y) := \mathbf{Prob}(X_{t+1} = y | X_t = x).$$

Applying the above notation to (3.3) yields

$$\begin{aligned}\mathbf{Prob}(X_0 = x_0, X_1 = x_1, \cdots, X_t = x_t) \\ = P(x_0, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n).\end{aligned}$$

Since the set of random variables holds Markov property and is time-homogeneous, we know $|\mathcal{X}| \times |\mathcal{X}|$ matrix is sufficient as a transition matrix. Thus the transition matrix P is given as follows:

$$P = \begin{bmatrix} \mathbf{Prob}(X_{t+1} = 1 | X_t = 1) & \mathbf{Prob}(X_{t+1} = 2 | X_t = 1) & \mathbf{Prob}(X_{t+1} = 3 | X_t = 1) & \mathbf{Prob}(X_{t+1} = 4 | X_t = 1) \\ \mathbf{Prob}(X_{t+1} = 1 | X_t = 2) & \mathbf{Prob}(X_{t+1} = 2 | X_t = 2) & \mathbf{Prob}(X_{t+1} = 3 | X_t = 2) & \mathbf{Prob}(X_{t+1} = 4 | X_t = 2) \\ \mathbf{Prob}(X_{t+1} = 1 | X_t = 3) & \mathbf{Prob}(X_{t+1} = 2 | X_t = 3) & \mathbf{Prob}(X_{t+1} = 3 | X_t = 3) & \mathbf{Prob}(X_{t+1} = 4 | X_t = 3) \\ \mathbf{Prob}(X_{t+1} = 1 | X_t = 4) & \mathbf{Prob}(X_{t+1} = 2 | X_t = 4) & \mathbf{Prob}(X_{t+1} = 3 | X_t = 4) & \mathbf{Prob}(X_{t+1} = 4 | X_t = 4) \end{bmatrix}.$$

The associated transition matrix to the figure 3.1 is

$$P = \begin{bmatrix} 0.2 & 0.3 & 0 & 0.5 \\ 0 & 0.1 & 0 & 0.9 \\ 0 & 0.3 & 0.3 & 0.4 \\ 0.5 & 0.2 & 0.3 & 0 \end{bmatrix}.$$

Observe that $P(i, j)$ is the probability of moving from state i to state j in one step. Also note that the sum of each row equals to 1; this is because the value of X_{t+1} must take one from the state space. Since the row sum equals to 1, the transition matrix P is stochastic.

We may want to analyze the distribution of X_t in order to see how distribution changes as t increases. When analyzing, it is convenient to store probability distribution of X_t in $1 \times |\mathcal{X}|$ row vector μ_t as follows:

$$\mu_t = \begin{bmatrix} \mathbf{Prob}(X_t = 1) & \mathbf{Prob}(X_t = 2) & \mathbf{Prob}(X_t = 3) & \mathbf{Prob}(X_t = 4) \end{bmatrix}.$$

In case where $t = 0$, the distribution matrix μ_0 is called an initial distribution of the Markov chain.

Equipped with the transition matrix and the distribution matrix, the natural question would be how we get the next step's distribution. Observe that

$$\mu_{t+1}(i) = \sum_{j \in \mathcal{X}} \mu_t(i) P(i, j),$$

or equivalently,

$$\mu_{t+1} = \mu_t P.$$

Thus, multiplying the transition matrix P on the right side of the distribution of X_t gives the distribution of X_{t+1}. In general, we can write the previous equation as

$$\mu_t = \mu_0 P^t. \tag{3.4}$$

Note that, from Lemma 3.2.1, P^t is stochastic when P is stochastic. Also, note that,

$$P^t(i, j) = \delta_i P^t(j).$$

From (3.4), we know $\delta_i P^t(j)$ is the probability distribution $\mathbf{Prob}(X_{t+1} = j)$ with initial distribution δ_i. Therefore, $P^t(i, j)$ denotes the probability of state i of moving to state j in t-steps.

But what happens when we multiply transition matrix on the left side of the distribution of X_t? Suppose f is a function on $\mathcal{X}$. Then multiplying transition matrix P to the left side of f gives the expectation; for $x \in \mathcal{X}$, $Pf(x) = \sum_{y \in \mathcal{X}} f(x)P(x, y)$. Note that $P(x, y)$ can be thought as a probability of $X_1 = y$, given an initial distribution δx. Hence, $Pf(x) = \mathbf{E}_{\delta_x}(f(X_1))$.

We now give a formal definition of the Markov chain.

Definition 3.2.3. (Markov Chain). A sequence of random variables $X_0, X_1, \cdots$ that satisfies a Markov property is called a Markov Chain. Thus X_{n+1} is independent of $X_0, X_1, \cdots, X_{n-1}$ but conditionally on X_n. Equivalently saying, for $\mathbf{Prob}(X_1 = x_1, X_2 = x_2, \cdots, X_t = x_t) > 0$,

$$\begin{aligned} \mathbf{Prob}(X_{t+1} = x_{t+1} | X_1 = x_1, \cdots, X_t = x_t) &= \mathbf{Prob}(X_{t+1} = x_{t+1} | X_t = x_t) \\ &= P(x_t, x_{t+1}). \end{aligned}$$

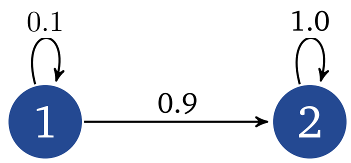
It is convention to write

$$\mathbf{Prob}(X_{t+1} = i | X_t = j, X_0 = k) = P_k(j, i).$$

Definition 3.2.4. (Accessibility). Let $\mathcal{X}$ be a finite state space of a Markov Chain. For $i, j \in \mathcal{X}$, we say j is accessible from i (written $i \to j$) if there exists a non-negative integer t such that

$$\mathbf{Prob}(X_t = j | X_0 = i) > 0.$$

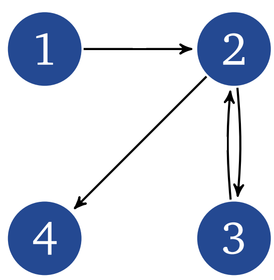
The accessibility relation is reflexive because for all $i \in \mathcal{X}$, $\mathbf{Prob}(X_0 = i | X_0 = i) = 1$. Also, this relation is transitive. For $i \to j$ and $j \to k$, it is easy to see $i \to k$. However, the accessibility relation is not symmetric. We can think of the below-directed graph as an example of a two-state Markov chain such that $1 \to 2$, but not $2 \to 1$.



Definition 3.2.5. (Communicate). Let $\mathcal{X}$ be a finite state space of a Markov chain. For $i, j \in \mathcal{X}$, we say i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.

Observe that the communicating relation is trivially symmetric; if $i \leftrightarrow j$, then $j \leftrightarrow i$. Recall that accessibility relation is reflexive and transitive; communicating relation inherits those properties. Thus communicating relation is an equivalence relation. Since communicating relation is an equivalence relation, communicating classes partitions a state space. Let us define

$$[i] := \{j \in \mathcal{X} | i \leftrightarrow j\}.$$



Then $[1], [2, 3], [4]$ partition the above Markov chain. If a Markov chain consists of one communicating class, we call the chain is irreducible;

Definition 3.2.6. (Irreducibility). We say a Markov chain is irreducible if it has only one communicating class. In other words, a Markov chain with state space $\mathcal{X}$ and transition matrix P is irreducible if for each $x, y \in \mathcal{X}$, there exists a non-negative integer t such that $P^t(x, y) > 0$.

If a Markov chain is irreducible, given any two states in state space $\mathcal{X}$, we can reach one state from the other state in some step; the number of steps depends on the choice of two states; this means we can reach all states given sufficient time, no matter what the initial distribution was.

Definition 3.2.7. (Period). The period of state i (written $d(i)$) is the greatest common divisor of a set $D(i) := \{t \geq 1 : P^t(i, i) > 0\}$.

Let $\mathcal{X}$ be a finite state space of an irreducible Markov chain. Then, for $i, j \in \mathcal{X}$, $i \to j$ and $j \to i$. Hence, $i \to i$. Therefore, we can assure that the set $\{t \geq 1 : P^t(i, i) > 0\}$ is not empty, and hence we can define the period in irreducible Markov chain without any problem.

Definition 3.2.8. (Aperiodic). A Markov chain is aperiodic if every state has period 1. Otherwise, we call the chain is periodic.

Lemma 3.2.2. *For an irreducible Markov chain, the period is constant on the state space.*

Proof. Let $\mathcal{X}$ be a finite state space of a Markov chain with transition matrix P and $i, j \in \mathcal{X}$. The periods of i and j are $\gcd D(i) = \gcd\{t \geq 1 : P^t(i, i) > 0\}$ and $\gcd D(j) = \gcd\{t \geq 1 : P^t(j, j) > 0\}$ respectively. First, we prove $\gcd D(i) \geq \gcd D(j)$.

Since P is irreducible, there exists $t_1, t_2 \in \mathbb{Z}^+$ such that

$$P^{t_1}(i, j) > 0, \quad P^{t_2}(j, i) > 0. \tag{3.5}$$

We observe that

$$P^{t_1+t_2}(j, j) = P^{t_2} \cdot P^{t_1}(j, j) = \sum_{k \in \mathcal{X}} P^{t_2}(j, k) P^{t_1}(k, j) \geq P^{t_2}(j, i) P^{t_1}(i, j).$$

Hence

$$P^{t_2+t_1}(j,j) \geq P^{t_2}(j,i)P^{t_1}(i,j) > 0, \tag{3.6}$$

and $t_1 + t_2 \in D(j)$. Suppose $t \in D(i)$. Because

$$P^{t+t_1+t_2}(j,j) \geq P^{t_2}(j,i)P^{t}(i,i)P^{t_1}(i,j) > 0,$$

we get $t + (t_1 + t_2) \in D(j)$. Since $t_1 + t_2, t + (t_1 + t_2) \in D(j)$,

$$\begin{aligned} t_1 + t_2 &= k_1 \cdot \gcd D(j) \\ t + t_1 + t_2 &= k_2 \cdot \gcd D(j) \end{aligned}$$

for some $k_1, k_2 \in \mathbb{Z}$. Then

$$t = (k_2 - k_1) \cdot \gcd D(j).$$

Since $k_2 - k_1 \in \mathbb{Z}$, we get $t \in D(j)$. Recall that t was an arbitrary element of $D(i)$ and greatest common divisor for $D(i)$ is $\gcd D(i)$ as we set at the beginning. Therefore $\gcd D(j) \leq \gcd D(j)$. A dual argument asserts $\gcd D(j) \geq \gcd D(j)$. Hence

$$\gcd D(j) = \gcd D(j)$$

as desired. □

Since the period of irreducible Markov chain is constant on the state space, showing the period of one state is aperiodic implies the irreducible Markov chain is aperiodic.

Lemma 3.2.3. *(Bézout's Identity). Let a and b be integers. Then there exists integers x and y such that*

$$ax + by = \gcd(a, b).$$

Proof. Let S be the set $S := \{ax + by \mid x, y \in \mathbb{Z}\}$. Observe the set S is not empty; for example, $\{-a, a, -b, b\} \subseteq S$. Since the set is not empty, we can apply the well-ordering principle. We claim that the smallest element $c = ax' + by'$ of the set S is $\gcd(a, b)$.

We first show c is a common divisor for a and b. From the division algorithm,

$$a = qc + r, \quad 0 \le r < c.$$

Then

$$\begin{aligned} r &= a - qc \\ &= a - q(ax' + by') \\ &= a(1 - qx') - b(qy'). \end{aligned}$$

Observe that r is a linear combination of a and b, and therefore $r \in S$. However, we set c to be the smallest element of S and $0 \le r < c$. Thus the only possible value that r can take is 0. Therefore we conclude c divides a. A dual argument shows c divides b.

It remains for us to show c is the greatest common divisor of a and b. Suppose d is any common divisor of a and b such that $a = de$ and $b = df$. Then

$$\begin{aligned} c &= ax' + by' \\ &= (de)x' + (df)y' \\ &= d(ex' + fy'). \end{aligned}$$

Hence $d \mid c$ and this implies $d \le c$. This concludes that c is a $\gcd(a, b)$. $\square$

Lemma 3.2.4. *Let S be any set of non-negative integers. Suppose $\gcd S = 1$ and the set S is closed under addition. Then there exists an integer x such that $y \ge x$ implies $y \in S$.*

Proof. From the Bézout's identity, we know there exits integers $a_1, a_2, \cdots, a_n$ such that

$$a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = 1 \tag{3.7}$$

where $s_i \in S$. There are two cases to consider: case 1 is when all the integer coefficients $a_i \ge 0$ for all i; case 2 is when we have negative integer coefficients.

In case 1, since S is closed under addition, $1 \in S$; this finishes the proof because for all non-negative integers x,

$$x = x \cdot 1 = x \cdot (a_1 s_1 + a_2 s_2 + \cdots + a_n s_n) \in S.$$

In case 2, suppose $a_1, a_2, \cdots, a_i \geq 0$ and $a_{i+1}, a_{i+2}, \cdots, a_n < 0$; otherwise, we can reorder the variables and relabel them. Since S is a set closed under addition, observe that

$$
\begin{aligned}
b_1 &:= a_1 s_1 + a_2 s_2 + \cdots + a_i s_i \in S \\
b_2 &:= -(a_{i+1} s_{i+1} + a_{i+2} s_{i+2} + \cdots + a_n s_n) \in S
\end{aligned}
$$

Also (3.7) can be rewritten as

$$
a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = b_1 - b_2 = 1. \tag{3.8}
$$

Now suppose $b_2 \geq 1$; when $b_2 = 0$, this case is same as the first case we discussed. We want to show there exists a positive integer y such that $x \geq y$ implies $x \in S$. In order to prove this, we need to show an arbitrary integer with some constraint can be written as a linear combination of elements of S with positive coefficients. Suppose $x \geq b_2^2$. Then by Euclidean algorithm,

$$
\begin{aligned}
x &= q b_2 + r, \quad 0 \leq r < b_2 \\
&= q b_2 + r \cdot 1.
\end{aligned} \tag{3.9}
$$

Substituting 1 with $b_1 - b_2$ by (3.8) yields

$$
x = q b_2 + r(b_1 - b_2) = r b_1 + (q - r) b_2.
$$

Showing $q - r$ is a positive coefficients proves that $x \in S$, and this completes the proof. Combining the equation and the inequality in (3.9) leaves

$$
b_2^2 \leq x = q b_2 + r < (q+1) b_2.
$$

Then $b_2^2 \leq (q+1) b_2$. Dividing both sides of inequality by b_2 and rewriting the inequality gives

$$
q \geq b_2 - 1.
$$

Subtracting r from the both sides of inequality yields

$$
q - r \geq b_2 - 1 - r. \tag{3.10}
$$

Recall that

$$
0 \leq r < b_2.
$$

This inequality implies

$$-1 - r \leq -1 < b_2 - 1 - r \tag{3.11}$$

Combining (3.10) with (3.11) results

$$q - r \geq b_2 - 1 - r > -1.$$

Hence $q - r \geq 0$ and x is a linear combination of b_1 and b_2 with positive coefficients. Thus x is an element of S and this completes the proof. □

Now we are ready to prove that in the case of finite Markov chain, the ergodicity is equivalent to aperiodicity plus the irreducibility.

Proposition 3.2.1. *For an aperiodic and irreducible Markov chain which has a finite state space $\mathcal{X}$ and transition matrix P, there exists $t_0 \in \mathbb{Z}^+$ such that $t \geq t_0$ implies $P^t(i,j) > 0$ for all $i, j \in \mathcal{X}$.*

Proof. Recall from Lemma 3.2.2, the period is constant on the state space of an irreducible Markov chain. Therefore $\gcd \mathcal{T}(i) = \gcd\{t \geq 1 \mid P^t(i,i) > 0\} = 1$ for all $i \in \mathcal{X}$. We first observe that $\mathcal{T}(i)$ is closed. Let $t_1, t_2 \in \mathcal{T}(i)$. Then $t_1 + t_2 \in \mathcal{T}(i)$; this is because

$$P^{t_1+t_2}(i,i) \geq P^{t_1}(i,i)P^{t_2}(i,i) > 0.$$

Hence, $\mathcal{T}(i)$ is closed under addition. Since the set $\mathcal{T}(i)$ is closed under addition and has greatest common divisor as a 1, it follows from Lemma 3.2.4 that there exists $t(i)$ such that $t_i \geq t(i)$ implies $t_i \in \mathcal{T}(i)$.

Since P is an irreducible Markov chain, there exits $t(ij)$ such that $P^{t(ij)}(i,j) > 0$; the value $t(ij)$ is dependent on the choice of i and j. Then

$$P^{t_i+t(ij)}(i,j) \geq P^{t_i}(i,i)P^{t(ij)}(i,j) > 0.$$

Letting $t(ij) := \max_{j\in\mathcal{X}}(t(ij))$. Then $t_{ij} \geq t(ij)$ implies $P^{t_i+t_{ij}}(i,j) > 0$. Finally, let the value of $t(i)$ be $\max_{i\in\mathcal{X}} t(i)$ where $t(i) \in \mathcal{T}(i)$. Then $t \geq t(i) + t(ij)$ implies

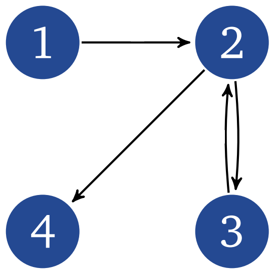
$$P^t(i,j) > 0 \text{ for all } i, j \in \mathcal{X}$$

as we desired. □

Definition 3.2.9. (Hitting Time). Let a sequence of random variables $(X_0, X_1, \cdots)$ be a Markov chain with finite state space $\mathcal{X}$ and transition matrix P. The hitting time for $x \in \mathcal{X}$ is

$$T_i := \min\{t \geq 0 | X_t = i\}.$$

Remark. We also define $T_i^+ := \min\{t \geq 1 | X_t = i\}$ and $T_i^{(n)}$ denotes n-th visit to the state i.



Suppose $X_0 = 1$. Then $P_1(T_2 < \infty) = 1$. However, in case when $X_0 = 4$, we get $P_1(T_2 < \infty) = 0$.

Definition 3.2.10. (Stopping Time). Let $\{X_0, X_1, \cdots\}$ be a Markov chain. Then the stopping time with respect to this Markov chain is a random variable $T : \mathcal{X} \to \mathbb{Z}_+ \cup \{0, \infty\}$ such that the event $\{T = n\}$ only depends on the previous $(X_0, X_1, \cdots, X_n)$ for all n.

The examples of stopping time are hitting time and n-th time of getting head when tossing a coin. However, the last exit time of a state i,

$$\max\{n \geq 0 | X_n = i\},$$

is not a hitting time because the last exit time depends on the future also.

We will see a Markov property plus stopping time implies a strong Markov property.

Theorem 3.2.5. *(Strong Markov Property). Let $(X_n | n \geq 0)$ be a Markov chain with a transition matrix P and T be a stopping time with respect to this time. Then*

$$\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | X_T = i_0, T < \infty) = P(i_0, i_1)P(i_1, i_2)\cdots$$

Proof. Let A be an event depend on $X_0, X_1, \cdots, X_T$. Then

$$\begin{aligned}
&\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots; A, T = n, X_T = i_0) \\
&= \frac{\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots; A, T = n, X_T = i_0)}{\mathbf{Prob}(A, T = n, X_T = i_0)} \cdot \mathbf{Prob}(A, T = n, X_T = i_0) \\
&= \mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | A, T = n, X_T = i_0)\mathbf{Prob}(A, T = n, X_T = i_0)
\end{aligned}$$

Since the event A depends on only $X_0, X_1, \cdots, X_n$ and independent on the future events after X_T,

$$\begin{aligned}
&\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots; A, T = n, X_T = i_0) \\
&= \mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | T = n, X_T = i_0)\mathbf{Prob}(A, T = n, X_T = i_0)
\end{aligned}$$

Because of Markov property, we get

$$\begin{aligned}
&\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots; A, T = n, X_T = i_0) \\
&= P(i_0, i_1)P(i_1, i_2)P(i_3, i_4)\cdots \mathbf{Prob}(A, T = n, X_T = i_0).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | T = n, X_T = i_0)\mathbf{Prob}(A, T = n, X_T = i_0) \\
&= P(i_0, i_1)P(i_1, i_2)P(i_3, i_4)\cdots \mathbf{Prob}(A, T = n, X_T = i_0)
\end{aligned}$$

By summing over for $n = 1, 2, 3, \cdots$, we get

$$\begin{aligned}
&\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | T < \infty, X_T = i_0)\mathbf{Prob}(A, T < \infty, X_T = i_0) \\
&= P(i_0, i_1)P(i_1, i_2)P(i_3, i_4)\cdots \mathbf{Prob}(A, T < \infty, X_T = i_0)
\end{aligned}$$

and upon dividing above equation by $\mathbf{Prob}(A, T < \infty, X_T = i_0)$, we get

$$\mathbf{Prob}(X_{T+1} = i_1, X_{T+2} = i_2, \cdots | T < \infty, X_T = i_0) = P(i_0, i_1)P(i_1, i_2)P(i_3, i_4)\cdots$$

as we desired. $\square$

Definition 3.2.11. (Regenerative Process). A regenerative process $\{X_t | t \geq 0\}$ is a random process such that there exist points $0 \leq T^1 < T^2 < T^3 < \cdots$ such that following two properties holds;

(1) $\{X_t | T^r \leq t < T^{r+1}\}$ has same distribution as $\{X_t | T^{r-1} \leq t < T^r\}$.
(2) $\{X_t | T^r \leq t < T^{r+1}\}$ is independent of $\{X_t | 0 \leq t < T^r\}$

We show the Markov chain with a stopping time T is a regenerative process.

Let T_i be a hitting time of a state i of a Markov chain:

$$T_i := T_i^{(1)} := \min\{t \geq 1 | X_t = i\} \quad \text{and} \quad T_i^{(0)} := 0$$

We recursively define n-th hitting time:

$$T_i^{(n)} := \min\{t \geq T_i^{(n-1)} + 1 | X_t = i\}.$$

Now, let us define n-th excursion:

$$\mathscr{X}_i^{(n)} := \{X_t | T_i^{(n)} \leq t < T_i^{(n+1)}\} \quad \text{for } n \in \mathbb{Z}_+$$

and

$$\mathscr{X}_i^{(0)} := \{X_t | 0 \leq t < T_i\}$$

Theorem 3.2.6. *Let $\{X_t | t \geq 0\}$ be a Markov chain and let $\{T_i^{(n)} | n = 0, 1, 2, \cdots\}$ be a set of n-th hitting time of state i. Then for $T_i^{(n)} < \infty$,*
(1) $\mathscr{X}_i^{(n)}$ is independent of $\mathscr{X}_i^{(0)}, \mathscr{X}_i^{(1)}, \cdots, \mathscr{X}_i^{(n-1)}$
(2) The distribution of $\mathscr{X}_i^{(n)}$ is same as $\mathscr{X}_i^{(n-1)}$ for all $n = 2, 3, 4, \cdots$.

Proof. By applying the strong Markov property (Theorem 3.2.5) at stopping time $T_i^{(n)}$, we get $\{X_{T_i^{(n)}+t} | t \geq 0\}$ is independent of $\{X_t | 0 \leq t < T_i^{(n)}\}$. Thus (1) holds.

Since Markov chain is assumed to be time-homogeneous, we get the same distribution for $\mathscr{X}_i^{(n)}$ for $n \in \mathbb{Z}_+$. □

3.3 Classification of States

Definition 3.3.1. (Recurrent). Let $\{X_t | t \geq 0\}$ be a Markov chain. Then state i is called recurrent if

$$\textbf{Prob}(T_i < \infty | X_0 = i) = 1.$$

Definition 3.3.2. (Transient). Let $\{X_t | t \geq 0\}$ be a Markov chain. Then state i is called transient if

$$\mathbf{Prob}(T_i < \infty | X_0 = i) < 1.$$

Let V_i denote the number of visits to a state i of a Markov chain.

$$V_i = \sum_{t=1}^{\infty} \mathbb{1}_{\{X_t = i\}} \quad \text{where } \mathbb{1}_{X_t = i} \begin{cases} 1 & \text{if } X_t = i, \\ 0 & \text{if } X_t \neq i. \end{cases}$$

Recall that $T_i^{(n)}$ is the n-th hitting time of the state i. Then it follows that $V_i = \max\{n \geq 0 | T_i^{(n)} < \infty\}$.

Lemma 3.3.1. $\mathbf{Prob}(V_i \geq t | X_0 = i) = \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t$ *for all* $t \in \mathbb{Z}_+ \cup \{0\}$.

Proof. When $t = 0$, it is trivially true from the definition. For inductive step, assume above statement holds for $t = n$. By applying the strong Markov property at $T_i^{(n)}$, we get

$$\mathbf{Prob}_i(V_i \geq n+1) = \mathbf{Prob}_i(V_i \geq n)\mathbf{Prob}_i(V_i \geq 1).$$

By inductive hypothesis, we know $\mathbf{Prob}_i(V_i \geq n) = \Big(\mathbf{Prob}_i(T_i < \infty)\}\Big)^n$. Thus,

$$\begin{aligned} \mathbf{Prob}_i(V_i \geq n+1) &= \Big(\mathbf{Prob}_i(T_i < \infty)\}\Big)^n \mathbf{Prob}_i(V_i \geq 1) \\ &= \Big(\mathbf{Prob}_i(T_i < \infty)\}\Big)^n \mathbf{Prob}_i(T_i < \infty) \\ &= \Big(\mathbf{Prob}_i(T_i < \infty)\}\Big)^{n+1}. \end{aligned}$$

as desired. □

Suppose state i is recurrent. Then a process starting from i will visit the state i with the probability 1. Again, by the Markov property, the process will revisit the state. Applying the Markov property recursively, we conclude that the state will be revisited infinitely many times.

Suppose now that a state i is transient. Then $\mathbf{Prob}(T_i < \infty | X_0 = i) < 1$ by definition. Then $\mathbf{Prob}(V_i < \infty) = 0$. Hence, the state i is revisited finitely many times during the process.

Lemma 3.3.2. *Suppose a state i of a Markov chain is recurrent. Then the followings are equivalent.*
(a) $\mathbf{Prob}(V_i = \infty | X_0 = i) = 1$
(b) $\mathbf{Prob}(T_i < \infty | X_0 = i) = 1$
(c) $\sum_{t=0}^{\infty} \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t = \infty$

Proof. From the previous discussion, we know (a) is equivalent to (b). Suppose now (b) holds. Observe that from Proposition 3.1.3 and Lemma 3.3.1,

$$\begin{aligned}
\mathbf{E}_i(V_i) &= \sum_{t=0}^{\infty} \mathbf{Prob}(V_i > r) \\
&= \sum_{t=0}^{\infty} \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t \\
&= \sum_{t=0}^{\infty} 1 \\
&= \infty.
\end{aligned}$$

Hence (b) implies (c).

Suppose (c) holds. Then

$$\begin{aligned}
E(V_i) &= \sum_{t=0}^{\infty} \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t \\
&= \frac{1}{1 - \mathbf{Prob}(T_i < \infty | X_0 = i)} \\
&= \infty.
\end{aligned}$$

Hence $\mathbf{Prob}(T_i < \infty | X_0 = i)$ equals to 1 as desired, and this completes the proof. □

Lemma 3.3.3. *Suppose a state i of a Markov chain is transient. Then the followings are equivalent.*
(a) $\mathbf{Prob}(V_i < \infty | X_0 = i) = 1$
(b) $\mathbf{Prob}(T_i < \infty | X_0 = i) < 1$
(c) $\sum_{t=0}^{\infty} \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t < \infty$

Proof. From the previous discussion, we know (a) is equivalent to (b). We show (b) is equivalent to (c). Suppose (b) holds. Then observe that from Proposition 3.1.3 and Lemma 3.3.1,

$$\begin{aligned}
\mathbf{E}_i(V_i) &= \sum_{t=0}^{\infty} \mathbf{Prob}(V_i > r) \\
&= \sum_{t=0}^{\infty} \Big(\mathbf{Prob}(T_i < \infty | X_0 = i)\Big)^t \\
&= \frac{1}{1 - \mathbf{Prob}(T_i < \infty | X_0 = i)} \\
&< \infty.
\end{aligned}$$

Hence (b) implies (c).

Suppose now (c) holds. Then

$$\begin{aligned}
E(V_i) &= \frac{1}{1 - \mathbf{Prob}(T_i < \infty | X_0 = i)} \\
&< \infty.
\end{aligned}$$

Hence $\mathbf{Prob}(T_i < \infty | X_0 = i)$ must be less than1, and this completes the proof. □

Corollary 3.3.4. *Any state of a Markov chain is either recurrent or transient.*

Proof. Observe that $\mathbf{Prob}(T_i < \infty | X_0 = i)$ is either less than 1 or equals to 1. Then from Lemma 3.3.2 and Lemma 3.3.3, all the states of Markov chain is either recurrent or transient. □

Corollary 3.3.5. *State i or a Markov chain with a transition matrix P is recurrent if $\sum_{n=1}^{\infty} P_{ii}^n = \infty$ and transient if $\sum_{n=1}^{\infty} P_{ii}^n < \infty$.*

Proof. From Lemma 3.3.2 and 3.3.3, we showed that state i is recurrent if $\mathbf{E}[V_i] = \infty$ and transient if $\mathbf{E}[V_i] < \infty$. Observe that

$$\begin{aligned}
\mathbf{E}[V_i] &= \mathbf{E}[\sum_{n=0}^{\infty} I_{\{X_n=i\}}|X_0 = i] \\
&= \sum_{n=0}^{\infty} \mathbf{E}[I_{\{X_n=i\}}|X_0 = i] \\
&= \sum_{n=0}^{\infty} \mathbf{Prob}(X_n = i|X_0 = i) \\
&= \sum_{n=0}^{\infty} P^n(i,i).
\end{aligned}$$

Hence state i is recurrent if $\sum_{n=1}^{\infty} P^n(i,i) = \infty$ and transient if $\sum_{n=1}^{\infty} P^n(i,i) < \infty$. □

Corollary 3.3.6. *If state i is a transient state of a Markov chain with a transition matrix P and a state space $\mathcal{X}$, then*

$$P^n(j,i) \to 0 \text{ as } n \to 0.$$

Proof. First observe that

$$\begin{aligned}
\mathbf{E}[V_i|X_0 = j] &= \mathbf{E}[\sum_{n=0}^{\infty} I_{\{X_n=i\}}|X_0 = j] \\
&= \sum_{n=0}^{\infty} \mathbf{Prob}(X_n = i|X_0 = j) \\
&= \sum_{n=0}^{\infty} P^n(j,i).
\end{aligned}$$

Since i is transient we have $\mathbf{E}[V_i|X_0 = j] < \infty$, and therefore in case $j \to i$, we must have $P^n(j,i) \to 0$ as $n \to \infty$ because the sum must be convergent; in case i is not accessible from j, we have $P^n(j,i) = 0$. □

Theorem 3.3.7. *Transient and recurrent are communicating class properties.*

Proof. Let a state i be transient and a state j communicate with j. Let P be a transition matrix of a Markov chain. Then there exists t_1 and t_2 such that

$$P^{t_1}(i,j) > 0, \; P^{t_2}(j,i) > 0.$$

Then
$$\sum_{t=0}^{\infty} P^t(j,j) \leq \frac{1}{P^{t_1}(i,j) \cdot P^{t_2}(j,i)} \sum_{t=0}^{\infty} P^{t_1+t_2+t}(i,i)$$
Since i is a transient state, therefore from Lemma 3.3.3, we know

$$\sum_{t=0}^{\infty} P^{t_1+t_2+t}(i,i) < \infty.$$

Therefore
$$\sum_{t=0}^{\infty} P^t(j,j) < \infty,$$

and therefore by Lemma 3.3.3, j is also transient. This proves that if one state of a communicating class is transient, all the states in the class are transient. Now suppose one of the states of communicating class is not transient. Then that state is recurrent by Corollary 3.3.4. By our previous observation that transient is a class property, if one of the states of the chain is recurrent, other states must be recurrent; otherwise, we see the contradiction. □

Even if $\mathbf{Prob}(T_i < \infty | X_0 = i) = 1$, observe that it does not guarantee $\mathbf{E}[T_i] < \infty$. We introduce definitions of positive recurrent and null recurrent.

Definition 3.3.3. (Positive Recurrent). A state i is positive recurrent if it is recurrent and $\mathbf{E}[T_i] < \infty$.

Definition 3.3.4. (Null Recurrent). A state i is positive recurrent if it is recurrent and $\mathbf{E}[T_i] = \infty$.

Positive and null recurrence are also class properties:

Lemma 3.3.8. *(Wald's Identity). Let $\{X_n\}_{n\geq 1}$ be independent and identically distributed (i.i.d.) random variables which have a finite expectation. Let T be a stopping time which is independent of $\{X_n\}_{n\geq 1}$ and which also has a finite expectation. Then*

$$\mathbf{E}[X_1 + X_2 + \cdots + X_T] = \mathbf{E}[X_1]\mathbf{E}[T].$$

Proof. Observe that
$$\sum_{t=1}^{T} X_t = \sum_{t=1}^{\infty} X_t I_{\{T\geq t\}},$$

where $I_{\{T\geq t\}}$ is a indicator random variable. Substituting $\sum_{t=1}^{T} X_t$ in $\mathbf{E}[\sum_{t=1}^{T} X_t]$ yields

$$\mathbf{E}[\sum_{t=1}^{T} X_t] = \mathbf{E}[\sum_{t=1}^{\infty} X_t I_{\{T\geq t\}}].$$

By the linearity of the expectation, we get

$$\mathbf{E}[\sum_{t=1}^{T} X_t] = \sum_{t=1}^{\infty} \mathbf{E}[X_t I_{\{T\geq t\}}].$$

By applying the law of iterated expectation (Theorem 3.1.2) we arrive at

$$\mathbf{E}[\sum_{t=1}^{T} X_t] = \sum_{t=1}^{\infty} \mathbf{E}[\mathbf{E}[X_t I_{\{T\geq t\}} | X_1, X_2, \cdots, X_{t-1}]].$$

Recall that T is a stopping time. Then by the strong Markov property, T is only dependent on $X_1, X_2, \cdots, X_{t-1}$. Thus $I_{\{T\geq t\}}$ is completely determined by $X_1, X_2, \cdots, X_{t-1}$, and therefore we can pull $I_{\{T\geq t\}}$ from conditional expectation:

$$\mathbf{E}[\sum_{t=1}^{T} X_t] = \sum_{t=1}^{\infty} \mathbf{E}[I_{\{T\geq t\}} \mathbf{E}[X_t | X_1, X_2, \cdots, X_{t-1}]].$$

Recall that X_n's are independent. Hence

$$\begin{aligned}\mathbf{E}[\sum_{t=1}^{T} X_t] &= \sum_{t=1}^{\infty} \mathbf{E}[I_{T\geq t} \mathbf{E}[X_1]] \\ &= \mathbf{E}[X_1] \sum_{t=1}^{\infty} \mathbf{E}[I_{\{T\geq t\}}]\end{aligned}$$

From the Definition 3.1.8, we observed that the expectation of the indicator random variable is $\mathbf{E}[I_{\{T\geq t\}}] = \mathbf{Prob}(T \geq t)$. Therefore

$$E[\sum_{t=1}^{T} X_t] = \mathbf{E}[X_1] \sum_{t=1}^{\infty} \mathbf{Prob}(T \geq t).$$

Applying the Proposition 3.1.3 yields

$$E[\sum_{t=1}^{T} X_t] = \mathbf{E}[X_1]\mathbf{E}[T],$$

as we wished. □

Theorem 3.3.9. *Positive and null recurrence are class properties.*

Proof. Let state i be null recurrent of a Markov chain with a transition matrix P, and let state i communicate with j. Since $i \leftrightarrow j$, there exists n satisfies $P^n(ij) > 0$; let n be a smallest positive integer satisfies the inequality. Now let A be a event such that

$$A = \{X_n = j \text{ and } T_i > n \text{ given } X_0 = i\}.$$

Because i is a null recurrent and $\mathbf{E}[T_i] \geq \mathbf{E}[T_i, A]$,

$$\begin{aligned}\infty > \mathbf{E}[T_i] &\geq \mathbf{E}[T_i, A] \\ &= \mathbf{E}[T_i|A]P(A) \\ &= (n + \mathbf{E}[T_{ij}])P(A)\end{aligned}$$

where T_{ij} denotes the first hitting time of state j given $X_0 = i$. In order to make above inequality holds, we must have $\mathbf{E}[T_{ij}] < \infty$; otherwise, we have $\mathbf{E}[T_i] > \infty$ which contradicts our assumption that i is null recurrent. It remains for us to show $\mathbf{E}[T_{ji}] < \infty$.

Let $\{Y_n\}_{n\geq 1}$ be i.i.d. random variables with each distribution same as T_i. Then n-th hitting time of state i is

$$T_i^{(n)} = Y_1 + Y_2 + \cdots + Y_n.$$

Let N be a number of time that the process visits state i until the process hits state j. Then N is a stopping time. To see $\mathbf{E}[N] < \infty$, let the probability p denotes the probability of the process visits state i before visiting state j. Suppose the process visits i, then by the strong Markov property, the future after the process visits i is independent of the past before the process hits the state i. Then we can observe that N is a geometric distribution with probability p, and therefore $\mathbf{E}[N] < \infty$. Observe that the following inequality holds;

$$T_{ji} \leq \sum_{n=1}^{N} Y_n.$$

Applying Wald's Identity (Lemma 3.3.8), yields

$$\mathbf{E}[T_{ji}] \leq \mathbf{E}[\sum_{n=1}^{N} Y_n] = \mathbf{E}[Y_1]\mathbf{E}[N] < \infty.$$

Hence $\mathbf{E}[T_{ji}] < \infty$ and

$$\mathbf{E}[T_j] \leq \mathbf{E}[T_{ji}] + \mathbf{E}[T_{ij}] < \infty$$

as we wished. Thus null recurrence is a class property. Observe that the recurrent state is either positive recurrent or null recurrent. If one state in the communicating class is positive recurrent, other states must also be positive recurrent; otherwise, since null recurrence is a class property, all the states should be null recurrent. □

Lemma 3.3.10. *An irreducible finite Markov chain with a transition probability P is recurrent.*

Proof. Since the chain is irreducible and finite, there must be at least one state i such that state i is visited infinitely many times; otherwise, if all the state are transient, we will eventually run out of state to visit as the chain moves. By the strong Markov property,

$$\begin{aligned} 0 &< P(X_n = i \text{ for infinitely many } n) \\ &= P(X_n = i \text{ for some } n) P_i(V_i = \infty). \end{aligned} \tag{3.12}$$

From Lemma 3.3.2, 3.3.3 and Corollary 3.3.4, we showed any state of Markov chain is either transient or recurrent. Hence $P_i(V_i = \infty)$ equals to either 1 or 0. To make the inequality 3.12 holds, we must have $P_i(V_i = \infty) = 1$. Therefore state i is recurrent. Since recurrent is a class property followed by Theorem 3.3.7, we conclude the chain is also recurrent. □

Proposition 3.3.1. *An irreducible finite Markov chain with a transition matrix P and a state space $\mathcal{X}$ is positive recurrent.*

Proof. Because positive recurrence is a class property by Theorem 3.3.9, showing at least one state of a Markov chain is positive recurrent proves the chain is positive recurrent. Recall that the transition matrix P is stochastic;

$$\sum_{j \in \mathcal{X}} P(i, j) = 1.$$

Lemma 3.2.1 implies when P is stochastic, P^n is also stochastic for all n. Hence

$$\lim_{n\to\infty} \sum_{j\in\mathcal{X}} P^n(i,j) = 1. \tag{3.13}$$

Since the sum is finite, we can put limits inside the finite sum;

$$\sum_{j\in\mathcal{X}} \lim_{n\to\infty} P^n(i,j) = 1$$

If all states are null recurrent, then $\lim_{n\to\infty} P^n(i,j) = 0$ for all j, and therefore $\sum_{j\in\mathcal{X}} \lim_{n\to\infty} P^n(i,j) = 0$. This contradicts the equation 3.13. Hence at least one state must be positive recurrent and this proves the chain is positive recurrent. □

3.4 Existence and Uniqueness of Stationary Distribution

Definition 3.4.1. (Stationary Distribution). Let π be a distribution of a Markov chain $\{X_n\}_{n\geq 0}$ with a transition matrix P and a state space $\mathcal{X}$. The distribution π is stationary if π satisfies the balance equation

$$\pi = \pi P$$

or equivalently,

$$\pi(i) = \sum_{j\in\mathcal{X}} \pi(j)P(j,i)$$

for all $j \in \mathcal{X}$

Remark. If $\mu_0 = \pi$, then $\mu_n = \pi$ for all n

Obviously, from the balance equation, we know the stationary distribution is a left eigenvector with an eigenvalue 1. One may ask whether all Markov chain has stationary distribution. However, we can think of counterexamples. Suppose a Markov chain is not irreducible; suppose a state space $\mathcal{X} = A \cup B$

where A and B has a stationary distribution π_a and π_b respectively. Given a transition matrix P as

$$P = \begin{bmatrix} P_A & 0 \\ 0 & P_B \end{bmatrix}$$

where P_A and P_B are transition matrix of A and B respectively. Then $\pi = \begin{bmatrix} q\pi_a & (1-q)\pi_b \end{bmatrix}$ where $0 \leq q \leq 1$ is an infinite set of stationary distributions of the chain. On the other hand, given a transition matrix P

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

any distribution π with $\sum_i \pi(i) = 1$ is a stationary distribution.

Now one may ask if stationary distribution exists, under what condition it is unique. We will see that the irreducible Markov chain has a unique stationary distribution if and only if it is positive recurrent. We will show the existence of unique stationary distribution by constructing it.

Definition 3.4.2. (Probability Flux). Given two subsets A and B of a state space of a Markov chain with a transition matrix P, the probability flux is given by

$$\text{flux}(A, B) = \sum_{i \in A} \sum_{j \in B} \pi(i) P(i, j).$$

Proposition 3.4.1. *A distribution π is stationary if and only if*

$$\text{flux}(A, A^\complement) = \text{flux}(A^\complement, A)$$

for all $A \subset \mathcal{X}$ and $\sum_{i \in \mathcal{X}} \pi(i) = 1$.

Proof. ($\Rightarrow$): Since the balance equation holds

$$\begin{aligned} \pi(i) &= \sum_{j \in \mathcal{X}} \pi(j) P(j, i) \\ &= \sum_{j \in A} \pi(j) P(j, i) + \sum_{j \in A^\complement} \pi(j) P(j, i) \end{aligned} \tag{3.14}$$

for some fixed subset A of the state space. Observe that $\pi(i) = \pi(i)\sum_{j\in\mathcal{X}} P(i,j)$ since P is stochastic. Thus

$$\pi(i)\sum_{j\in\mathcal{X}} P(i,j) = \sum_{j\in A}\pi(j)P(j,i) + \sum_{j\in A^{\complement}}\pi(j)P(j,i)$$

or equivalently,

$$\sum_{j\in A}\pi(i)P(i,j) + \sum_{j\in A^{\complement}}\pi(i)P(i,j) = \sum_{j\in A}\pi(j)P(j,i) + \sum_{j\in A^{\complement}}\pi(j)P(j,i).$$

Applying $\sum_{i\in A}$ on the both sides of above equation yields

$$\sum_{i\in A}\sum_{j\in A}\pi(i)P(i,j)+\sum_{i\in A}\sum_{j\in A^{\complement}}\pi(i)P(i,j) = \sum_{i\in A}\sum_{j\in A}\pi(j)P(j,i)+\sum_{i\in A}\sum_{j\in A^{\complement}}\pi(j)P(j,i). \tag{3.15}$$

Observe that $\sum_{i\in A}\sum_{j\in A}\pi(i)P(i,j)$ equals to $\sum_{i\in A}\sum_{j\in A}\pi(j)P(j,i)$. Hence 3.15 becomes

$$\sum_{i\in A}\sum_{j\in A^{\complement}}\pi(i)P(i,j) = \sum_{i\in A}\sum_{j\in A^{\complement}}\pi(j)P(j,i)$$

which can be rewritten as

$$\mathrm{flux}(A, A^{\complement}) = \mathrm{flux}(A^{\complement}, A)$$

as we wished.

($\Leftarrow$): Suppose $\mathrm{flux}(A, A^{\complement}) = \mathrm{flux}(A^{\complement}, A)$. By setting $A = \{i\}$, we get

$$\sum_{a\in\{i\}}\sum_{b\in\{i\}^{\complement}}\pi(a)P(a,b) = \sum_{a\in\{i\}^{\complement}}\sum_{b\in\{i\}}\pi(a)P(a,b) \tag{3.16}$$

The left side of the equation 3.16 equals to

$$\sum_{a\in\{i\}}\sum_{b\in\{i\}^{\complement}}\pi(a)P(a,b) = \sum_{b\in\{i\}^{\complement}}\pi(i)P(i,b).$$

The right side of the equation 3.16 equals to

$$\sum_{a\in\{i\}^{\complement}}\sum_{b\in\{i\}}\pi(a)P(a,b) = \sum_{a\in\{i\}^{\complement}}\pi(a)P(a,i).$$

Therefore,

$$\sum_{b\in\{i\}^\complement} \pi(i)P(i,b) = \sum_{a\in\{i\}^\complement} \pi(a)P(a,i). \tag{3.17}$$

Combining above equation 3.17 with $\sum_{i\in\mathcal{X}} \pi(i) = 1$ and naming variables differently, we get

$$\pi(i) = \sum_{j\in\mathcal{X}} \pi(j)P(j,i)$$

as desired. □

However, solving systems of equations may be cumbersome. We will construct a stationary distribution and give an exact formula at the end of this chapter.

Theorem 3.4.1. *An irreducible Markov chain $\{X_t\}_{t\geq 0}$ with a transition matrix P has a stationary distribution if and only if it is positive recurrent.*

Proof. ($\Rightarrow$): Recall that positive recurrence is a class property. We first show if an irreducible Markov chain is positive recurrent, then there exists a stationary distribution. We will show the existence of stationary distribution by constructing it. Let us define and recall some variables;

$$\begin{aligned}
&T_i^{(0)} = T_i^+ := \min\{t|t\geq 1, X_t = i\}\\
&T_i^{(n)} := \min\{t|t\geq T_i^{(n-1)}+1, X_t = i\};\\
&\mathscr{X}_i^{(0)} := \{X_t|0\leq t < T_i, X_0 = i\};\\
&\mathscr{X}_i^{(n)} := \{X_t|T_i^{(n)}\leq t < T_i^{(n+1)}\};\\
&N_j = \sum_{t=1}^{\infty} I_{\{X_t=j, T_i\geq t\}}.
\end{aligned}$$

Also recall from Theorem 3.2.6, given $T_i^{(n)}$, the n-th hitting time of state i, n-th excursion $\mathscr{X}_i^{(n)}$ is independent of $\mathscr{X}_i^{(0)}, \cdots, \mathscr{X}_i^{(n-1)}$. Furthermore, the distributions of excursions are same for all n. This implies that the proportion of the times that the chain spends at state j are same for all excursions. This also implies that the proportion of time that chain spends in state j for long run time is same as the proportion of the time spent in n-th excursion. To guarantee n-th excursion exists, we must have every state to be recurrent; otherwise $P_i(T_i < \infty) < 1$ and we are unable to define n-th excursion for sure.

Observe that our previous discussion implies

$$\mathbf{E}[T_i] = \sum_{j\in\mathcal{X}} \mathbf{E}[N_j|X_0 = i];$$

In words, the expected length of first excursion $\mathscr{X}_i^{(0)}$ equals the sum of expected number of visits to state during $\mathscr{X}_i^{(0)}$. To make sure the expected length of excursion to be finite, we must have every state of the chain to be positive recurrent; otherwise, we will have $\mathbf{E}[T_i] = \infty$.

Let us define a new variable $d(j)$;

$$\begin{aligned} d(j) &= \mathbf{E}[N_j|X_0 = i] \\ &= \sum_{t=1}^{\infty} \mathbf{E}[I_{\{X_t=j,T_i\geq t\}}|X_0 = i]. \end{aligned}$$

We will show $d(j)/\sum_{i\in\mathcal{X}} d(i)$ satisfies the balance equation, and therefore it is a stationary distribution. Note that $I_{\{X_t=j,T_i\geq t\}}$ is an indicator function, then

$$d(j) = \sum_{t=1}^{\infty} P_i(X_t = j, T_i \geq t) \tag{3.18}$$

Define an event

$$A_k := \{X_{t-1} = k\}$$

for all $k \in \mathcal{X}$. Then A_k's partition entire sample space. Applying this fact, equation 3.18 can be rewritten as

$$d(j) = \sum_{t=1}^{\infty} \sum_{k\in\mathcal{X}} P_i(X_t = j, X_{t-1} = k, T_i \geq t).$$

Observe that when $k = i$, $P_i(X_t = j, X_{t-1} = k, T_i \geq t) = 0$ because we cannot have $X_{t-1} = i$ at the same time $T_i \geq t$. Hence

$$d(j) = \sum_{t=1}^{\infty} \sum_{k\neq i} P_i(X_t = j, X_{t-1} = k, T_i \geq t). \tag{3.19}$$

For simplicity, let

$$a_{ij}(t) := \sum_{k\neq i} P_i(X_t = j, X_{t-1} = k, T_i \geq t).$$

Now observe that

$$a_{ij}(t) = \sum_{k \neq i} a_{ik}(t-1)P(k,j) \tag{3.20}$$

for $t = 2, 3, \cdots$. Let us rewrite equation (3.19) by using notation from (3.20);

$$\begin{aligned} d(j) &= \sum_{t=1}^{\infty} a_{ij}(t) \\ &= a_{ij}(1) + \sum_{t=2}^{\infty} a_{ij}(t) \end{aligned} \tag{3.21}$$

Here observe that $a_{ij}(1) = \sum_{k \neq i} P_i(X_1 = j, X_0 = k, T_i \geq 1)$ equals to $P(i,j)$. Substituting 3.21 with 3.20 yields

$$\begin{aligned} d(j) &= P(i,j) + \sum_{t=2}^{\infty} \sum_{k \neq i} a_{ik}(t-1)P(k,j) \\ &= P(i,j) + \sum_{k \neq i} \Big(\sum_{t=2}^{\infty} a_{ik}(t-1) \Big) P(k,j) \\ &= P(i,j) + \sum_{k \neq i} d(k)P(k,j) \end{aligned} \tag{3.22}$$

From the definition, $d(i) = \mathbf{E}[N_i | X_0 = i]$; it is a number of visits to state i before the chain visits state i for the first time after its first visit at time 0. Thus $d(i) = 1$. Applying this to (3.22) leaves

$$d(j) = \sum_{k \in \mathcal{X}} d(k)P(k,j).$$

Lastly, let us normalize $d(j)$;

$$\pi(j) := \frac{d(j)}{\sum_{i \in \mathcal{X}} d(j)}.$$

Then we see

$$\pi = \pi P$$

as we desired.

($\Leftarrow$): We now show if stationary distribution exists, then the chain is positive recurrent and furthermore the stationary distribution is unique. Suppose a stationary distribution exists for an irreducible Markov chain. Then there exists π such that satisfies the balance equation;

$$\pi = \pi P. \tag{3.23}$$

The balance equation (3.23) implies

$$\pi = \pi P^n$$

or equivalently,

$$\pi(i) = \sum_{j \in \mathcal{X}} \pi(j) P^n(j, i). \tag{3.24}$$

Applying limit to (3.24) leaves

$$\pi(i) = \lim_{n \to \infty} \sum_{j \in \mathcal{X}} \pi(j) P^n(j, i). \tag{3.25}$$

Since the sum is finite, (3.25) can be rewritten as

$$\pi(i) = \sum_{j \in \mathcal{X}} \lim_{n \to \infty} \pi(i) P^n(j, i).$$

By Corollary 3.3.6, if i is a transient state then, $\lim_{n\to\infty} P^n(j,i) = 0$. Suppose all states of chain are transient, then $\pi(i) = 0$ for all i. It contradicts the fact that $\sum_{i \in \mathcal{X}} \pi(i) = 1$. Thus our assumption that all state are transient is wrong and hence there must exists a recurrent state. Recall that recurrent is a class property; one state being recurrent implies all states must be recurrent.

We now prove the chain is positive recurrent, and show the uniqueness of stationary distribution. Suppose initial distribution $\mu_0 = \pi$, then $\mu_n = \pi$ for all n. We claim

$$\mathbf{E}[T_i | X_0 = i] \pi(i) = \mathbf{E}[T_i | X_0 = i] P(X_0 = i) = 1.$$

From Proposition 3.1.3, $\mathbf{E}[T_i | X_0 = i] P(X_0 = i)$ can be rewritten as

$$\begin{aligned} \mathbf{E}[T_i | X_0 = i] P(X_0 = i) &= \sum_{n=1}^{\infty} P(T_i \geq n | X_0 = i) P(X_0 = i) \\ &= \sum_{n=1}^{\infty} P(T_i \geq n, X_0 = i). \end{aligned}$$

From the basic set theory, we know given two sets A and B, $P(A \cap B) = P(A) - P(A \cap B^{\complement})$. Setting $A = \{X_1 \neq i, X_2 \neq i, \cdots, X_{n-1} \neq i\}$, $B = \{X_0 = i\}$ leaves

$$\sum_{n=1}^{\infty} \mathbf{Prob}(T_i \geq n, X_0 = i) = \mathbf{Prob}(T_i \geq 1, X_0 = i)$$
$$+ \sum_{n=2}^{\infty} \Big(\mathbf{Prob}(X_1 \neq i, X_2 \neq i, \cdots, X_{n-1} \neq i) - \mathbf{Prob}(X_0 \neq i, X_1 \neq i, \cdots, X_{n-1} \neq i) \Big)$$

Observe that $\mathbf{Prob}(T_i \geq 1, X_0 = i)$ equals to $\mathbf{Prob}(X_0 = i)$. Also recall that we assumed an initial distribution to be π, and therefore whole process is stationary; hence we can shift the index of $\mathbf{Prob}(X_1 \neq i, X_2 \neq i, \cdots, X_{n-1} \neq i)$ by one and still have same probability. Note that following holds;

$$\sum_{n=1}^{\infty} \mathbf{Prob}(T_i \geq n, X_0 = i) = \mathbf{Prob}(X_0 = i)$$
$$+ \sum_{n=2}^{\infty} \Big(\mathbf{Prob}(X_0 \neq i, X_1 \neq i, \cdots, X_{n-2} \neq i) - \mathbf{Prob}(X_0 \neq i, X_1 \neq i, \cdots, X_{n-1} \neq i) \Big) \tag{3.26}$$

For the simplicity of notation, let us define a new variable;

$$b_n := \mathbf{Prob}(X_0 \neq i, X_1 \neq i, \cdots, X_n \neq i). \tag{3.27}$$

Applying this notation (3.27) to (3.26) yields

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{Prob}(T_i \geq n, X_0 = i) &= \mathbf{Prob}(X_0 = i) + \sum_{n=2}^{\infty} (b_{n-2} - b_{n-1}) \\
&= \Big(\mathbf{Prob}(X_0 = i) + b_0 \Big) - \lim_{n \to \infty} b_n \\
&= \Big(\mathbf{Prob}(X_0 = i) + \mathbf{Prob}(X_0 \neq i) \Big) - \lim_{n \to \infty} b_n \\
&= 1 - \lim_{n \to \infty} b_n.
\end{aligned}$$

Observe that

$$\lim_{n \to \infty} b_n = \lim_{n \to \infty} \mathbf{Prob}(X_0 \neq i, \cdots, X_n \neq i),$$

the probability that recurrent state i will never be visited given an initial position at j. By the strong Markov property with hitting time $T_j^{(n)}$ and irreducibility of the chain, for each excursion, there exists a positive integer such that $P^n(j,i) > 0$. Hence $\lim_{n\to\infty} b_n = 0$. To sum up,

$$\mathbf{E}[T_i|X_0 = i]\pi(i) = 1,$$

and therefore

$$\pi(i) = \frac{1}{\mathbf{E}[T_i|X_0 = i]}.$$

Since mean time to return to state i is unique, π is also unique. To see the chain is positive recurrent, assume the chain is null recurrent. Then $\mathbf{E}[T_i|X_0 = i] = \infty$, and therefore $\pi(i) = 0$ for all $i \in \mathcal{X}$. But we may not have $\sum_{i\in\mathcal{X}} \pi(i) = 1$. Thus we have reached a contradiction and our original assumption that the chain is null recurrent is wrong, and the chain must be positive recurrent. This completes the proof. □

4

Random Walks on Finite Groups

4.1 Upper Bound Lemma

Having studied some of the essential statements of group representation, character theory, Fourier transform, and Markov chain, we are ready to introduce random walks on finite groups.

Definition 4.1.1. (Random Walk). A random walk driven by a probability measure μ on a group has a distribution at step n given by μ^{*n} where $\mu^{*1} = \mu$ and $\mu^{*n} = \mu * \mu^{*(n-1)}$.

We first verify μ^{*n} is a probability measure in the following proposition.

Proposition 4.1.1. *Let P and Q be probability distributions on a finite group G and suppose P and Q are independent. Then $P * Q$ is also a probability distribution on G.*

Proof. Since P and Q are probability distributions on a finite group G, following holds:

$$\sum_{g\in G} P(g) = 1, \sum_{g\in G} Q(g) = 1,$$
$$0 \leq P(g), Q(g) \leq 1 \quad \text{for all } g \in G.$$

From the definition of the convolution, for $g \in G$

$$P * Q(g) = \sum_{h\in G} P(gh^{-1})Q(h).$$

Observe that $P(gh^{-1})Q(h) \leq Q(h) \leq 1$ for $h \in G$ because $P(gh^{-1}) \leq 1$.

Also, we can easily verify that $\sum_{g \in G} P * Q(g) = 1$;

$$\begin{aligned}\sum_{g \in G} P * Q(g) &= \sum_{g \in G} \sum_{h \in G} P(gh^{-1})Q(h) \\ &= \sum_{h \in G} Q(h) \sum_{g \in G} P(gh^{-1}).\end{aligned}$$

Since $P(gh^{-1})$ sums over all the elements in G once, $\sum g \in GP(gh^{-1})$ equals to 1. Thus,

$$\begin{aligned}\sum_{g \in G} P * Q(g) &= \sum_{h \in G} Q(h) \cdot 1 \\ &= 1 \cdot 1 = 1.\end{aligned}$$

Hence $P * Q$ is a probability distribution on G. $\square$

We will introduce a property that convolution between two probability measures enjoys. Before stating the proposition, we recall the definition of support of a function.

By using convolution, we can find a probability distribution of a sum of two or more independent random variables. Let X and Y be two independent random variables on G such that $X \sim P$ and $Y \sim Q$. Let probability distribution $Z = XY$ be their sum, and let $Z \sim R$. To calculate $R(Z = g)$, first suppose $Y = h$ where $h \in G$. Then X must equal to gh^{-1} in order to make $Z = g$. Hence $R(Z = g) = \sum_{h \in G} P(gh^{-1})Q(h) = P * Q(g)$. This is an intuition behind the definition of the random walk; the n-th power of convolution of a probability measure μ on a finite group G is a probability distribution on n-th step.

Definition 4.1.2. (Support). Let P be a probability distribution on a group G. The support of P is given by $\operatorname{supp}(P) = \{g \in G | P(g) \neq 0\}$.

Proposition 4.1.2. *Let P and Q be probability distributions on a finite group G, and let P and Q be independent. Then* $\operatorname{supp}(P * Q) = \operatorname{supp}(P) \cdot \operatorname{supp}(Q)$.

Proof. From the definition of the support,

$$\operatorname{supp}(P * Q) = \{g \in G | \sum_{h \in G} P(gh^{-1})Q(h) \neq 0\}.$$

Then $g \in \operatorname{supp}(P * Q)$ if and only if there exist $h \in G$ such that $gh^{-1} \in \operatorname{supp}(P)$ and $h \in \operatorname{supp}(Q)$. Since $(gh^{-1}) \cdot h = g$, we see that $\operatorname{supp}(P * Q) = \operatorname{supp}(P) \cdot \operatorname{supp}(Q)$ as desired. $\square$

Definition 4.1.3. (Ergodic). A random walk driven by a probability measure μ is called ergodic if there exits a positive integer t such that

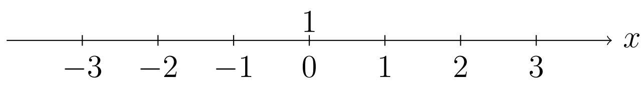
$$\operatorname{supp}(\mu^{*t}) = G.$$

Example 4.1.1. Let μ be a probability on $\mathbb{Z}/101\mathbb{Z}$ given by

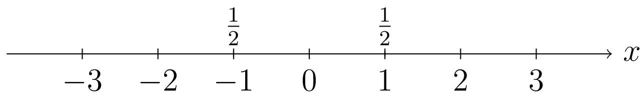
$$\mu(x) = \begin{cases} \frac{1}{2} & \text{if } x = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

This is called a simple random walk and the general cases of this walk will be revisited in the next example.

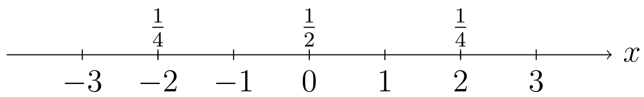
At step 0: μ is the probability distribution at step 0.



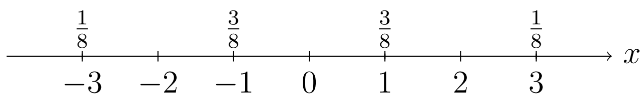
At step 1: μ is the probability distribution at step 1.



At step 2: $\mu * \mu$ is the probability distribution at step 2.



At step 3: μ^{*3} is the probability distribution at step 3.



The random walker's position at n-th steps has the distribution

$$\mu^{*n} = \mu * \mu^{*(n-1)}$$

on $\mathbb{Z}/101\mathbb{Z}$. It is beneficial to study the Fourier transform of $\widehat{\mu^{*n}}$. It follows from the definition of the Fourier transform that

$$\widehat{\mu}(n) = \sum_{x \in \mathbb{Z}/101\mathbb{Z}} \mu(x) e^{-2\pi i n x/101}.$$

Then

$$\begin{aligned}
\widehat{\mu * \mu}(n) &= \sum_x \mu * \mu(x) e^{-2\pi i n x/101} \\
&= \sum_x \sum_y \mu(x-y)\mu(y) e^{-2\pi i n x/101} \\
&= \sum_x \sum_y \mu(x-y) e^{-2\pi i n (x-y)/101} \mu(y) e^{-2\pi i n y/101} \\
&= \left(\widehat{\mu}(n)\right)^2
\end{aligned}$$

as expected; this can be also verified by Theorem 2.2.3. We see that after n steps $\widehat{\mu^{*n}}(\xi) = (\widehat{\mu}(\xi))^n$. Note that $|e^{2\pi i n \xi/101}| = 1$. Then

$$|\widehat{\mu}(\xi)| = \left| \sum_{x \in \mathbb{Z}/101\mathbb{Z}} \mu(x) e^{-2\pi i \xi x/101} \right| \leq \sum_{x \in \mathbb{Z}/101\mathbb{Z}} \mu(x) = 1.$$

Also, note that if μ is a uniform measure $\mathbb{U}$, then

$$\begin{aligned}
\widehat{\mathbb{U}}(\xi) &= \sum_{x \in \mathbb{Z}/101\mathbb{Z}} \mathbb{U}(x) e^{-2\pi i \xi x/101} \\
&= \frac{1}{101} \sum_{x \in \mathbb{Z}/101\mathbb{Z}} e^{-2\pi i \xi x/101} \\
&= \begin{cases} 1 & \text{if } \xi = 0, \\ 0 & \text{if } \xi \neq 0. \end{cases}
\end{aligned}$$

In this case, $\widehat{\mathbb{U}}(\xi)$ detects whether ξ equals 0 or not.

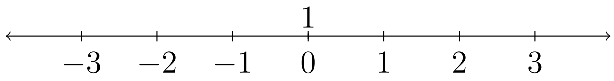


Fig. 4.1: Step 0 of a Simple Random Walk on $\mathbb{Z}$

More generally, given $\chi \in \hat{G}$, $\widehat{\mathbb{U}}(\chi)$ equals to δ_{χ_1}. This is because

$$\begin{aligned}
\widehat{\mathbb{U}}(\chi) &= \sum_{g\in G} \mathbb{U}(g)\overline{\chi(g)} \\
&= \sum_{g\in G} \frac{1}{|G|}\overline{\chi(g)} \\
&= \sum_{g\in G} \frac{1}{|G|}\chi_1(g)\overline{\chi(g)} \\
&= \langle \chi_1, \chi \rangle.
\end{aligned}$$

Finally by Schur's orthogonality relations (Theorem 1.4.7), we get

$$\widehat{\mathbb{U}}(\chi) = \delta_{\chi_1}.$$

Example 4.1.2. (Simple Random Walk on $\mathbb{Z}$)

Let the state space $\mathcal{X}$ be $\mathbb{Z}$. Let $\{X_n\}_{n\geq 1}$ be independent and identically distributed (i.i.d.) random variables where each random variable X_i is defined as follows:

$$X_i = \begin{cases} 1 & \text{with } \mathbf{Prob}(X_i = 1) = \frac{1}{2}, \\ -1 & \text{with } \mathbf{Prob}(X_i = 1) = \frac{1}{2}. \end{cases}$$

The simple random walk on $\mathbb{Z}$ is a random process defined as $W_0 = 0$ and

$$W_n := X_1 + X_2 + \cdots + X_n.$$

Here, note that at each step, the move is independent from the past;

$$W_{n+1} = W_n + X_{n+1}.$$

At each time, the random walker moves to either left or right by one step from the current position (Figure 4.1).

Observe that W_t is binomially distributed:

$$\mathbf{Prob}(W_{2t} = 2j) = \frac{1}{2^{2t}}\binom{2t}{t+j}. \tag{4.1}$$

To understand why (4.1) holds, think of selecting $t + j$ random variables among $2t$ independent random variables and assign the value 1 to each of them. For the remaining $2t - (t + j)$ random variables, we assign the value -1 to them. As a result, we get

$$W_{2t} = X_1 + X_2 + \cdots + X_{2t} = 1 \cdot (t + j) - 1 \cdot (2t - (t + j)) = 2j$$

as we desired.

Let us calculate the expectation of W_t. First, note that

$$\begin{aligned} \mathbf{E}[X_i] &= 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = 0 \\ \mathbf{E}[X_i^2] &= 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1. \end{aligned} \tag{4.2}$$

The linearity of expectation (Proposition 3.1.1) and (4.2) yields

$$\begin{aligned} \mathbf{E}[W_t] &= \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_t] = 0, \\ \mathbf{E}[W_t^2] &= \mathbf{E}[\Big(X_1 + \cdots + X_t\Big)^2] \\ &= \mathbf{E}[\sum_i^t X_i^2 + 2\sum_{i<j} X_i X_j] \\ &= \mathbf{E}[\sum_i^t X_i^2] + 2\sum_{i<j} \mathbf{E}[X_i]\mathbf{E}[X_j] \\ &= \mathbf{E}[\sum_i^t X_i^2] + 0 \\ &= t. \end{aligned}$$

Then the variance $\mathbf{Var}(W_t)$ is $\mathbf{E}[W_t^2] - \mathbf{E}[W_t]^2 = t$. Moreover, observe that if t increases, the probability approaches a normal distribution:

$$\frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

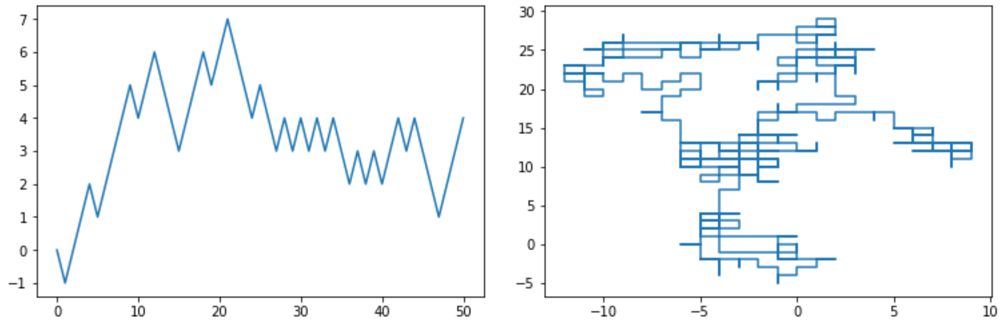


Fig. 4.2: Simple Random Walks on 1-Dimension (Left), Simple Random Walk on 2-Dimension (Right)

We give a generalized version of a simple random walk. There is a classic example of a random walk on n-dimension. When the dimension is 1, we can think of the random walk as a tossing a fair coin, similar to our previous discussion. When the dimension is 2, we can think of the random walk like a drunkard in New York City. The drunkard's initial location is X_0, and at each step, the drunkard has an equal probability of moving east, west, south, and north.

By using a Python, we can simulate random walks on n-dimensions. The left figure in Figure 4.2 illustrates the random walk on 1-dimension. The x-axis represents the number of steps, and the y-axis represents the position. The right figure illustrates the random walk on 2-dimension. The blue line represents the trajectory of a drunkard started from an initial position $(0, 0)$.

Example 4.1.3. (Simple Random Walk on $\mathbb{Z}_n$)

Let $\mathcal{X} = \mathbb{Z}_n = \{0, 1, 2, \cdots, n-1\}$ be a state space, and define a transition matrix P as follows:

$$P(j,k) = \begin{cases} \frac{1}{2} & \text{if } k \equiv j+1 \mod n, \\ \frac{1}{2} & \text{if } k \equiv j-1 \mod n, \\ 0 & \text{otherwise.} \end{cases}$$

A random walk on the n-cycle is a Markov chain with state space $\mathcal{X}$ and transition matrix P as defined above. We give two examples of random walk on n-cycle (Figure 4.3): 6-cycle and 7-cycle given that $X_0 = 0$.

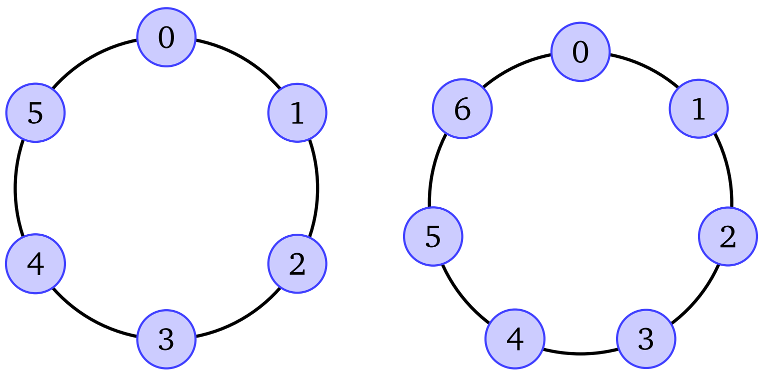


Fig. 4.3: Random Walk on 6-Cycle (Left) and 7-Cycle (Right)

r.v$\backslash\mathcal{X}$	0	1	2	3	4	5
X_1		$\checkmark$				$\checkmark$
X_2	$\checkmark$		$\checkmark$		$\checkmark$	
X_3		$\checkmark$		$\checkmark$		$\checkmark$
X_4	$\checkmark$		$\checkmark$		$\checkmark$	
X_5		$\checkmark$		$\checkmark$		$\checkmark$

r.v$\backslash\mathcal{X}$	0	1	2	3	4	5	6
X_1		$\checkmark$					$\checkmark$
X_2	$\checkmark$		$\checkmark$			$\checkmark$	
X_3		$\checkmark$		$\checkmark$	$\checkmark$		$\checkmark$
X_4	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	
X_5		$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
X_6	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

The left table describes 6-cycle, and the right table describes 7-cycle. The check-mark $\checkmark$ in (t, x)-th slot of the table indicates $P^t(0, x) > 0$; the state x is accessible from the state 0 in t-th step. We observe that both cycles are irreducible from the above tables; both cycles have one communicating class. Let us examine the period of each state of 6-cycle and 7-cycle.

We can partition the state $\mathcal{X}$ into two states; $\mathcal{X}_1 = \{0, 2, 4\}$, $\mathcal{X}_2 = \{1, 3, 5\}$. Notice that the period of any state is 2 in 6-cycle case. Hence 6-cycle is periodic.

The period of any state space in random walk on 7-cycle is 1. Hence, P is aperiodic. Then by Proposition 3.2.1, there exists $t(x)$ such that $P^t(x, y) > 0$ for all $x, y \in \mathcal{X}$ and $t > t(x)$. In this case, the $t(x)$ is 5.

We can think the random walk on the n-cycle as follows. We toss a fair coin. If the head is up, we move one step clockwise. Otherwise, we move one step counter-clockwise. Let Z be a random variable with uniform distribution on $\{-1, 1\}$. Then the transition matrix P of this Markov chain can be defined as

$$P(x, y) := \textbf{Prob}\{(x + Z) \mod n = y\}.$$

The stationary distribution of random walk on 6-cycle is

$$\begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \cdot \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}.$$

In general the stationary distribution of random walk on n-cycle is $\begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$.

Like we have seen, not every simple random walk on $\mathbb{Z}_n$ is aperiodic. We introduce a lazy random walk on n-cycle which is always aperiodic no matter what the value of n.

Definition 4.1.4. (Lazy Random Walk on n-cycle).

Let $\mathcal{X} = \mathbb{Z}_n = \{0, 1, 2, \cdots, n-1\}$ be a state space, and define a transition matrix P as follows:

$$P(j, k) = \begin{cases} \frac{1}{4} & \text{if } k \equiv j + 1 \mod n, \\ \frac{1}{2} & \text{if } k \equiv j \mod n, \\ \frac{1}{4} & \text{if } k \equiv j - 1 \mod n. \end{cases}$$

Note that $P(j, j) = \frac{1}{2} > 0$ for all $j \in \mathcal{X}$, and therefore every state has period 1. Hence every lazy random walk on n-cycle is aperiodic.

Example 4.1.4. (Simple Random Walk on a Finite Graph G).

Given a graph $G = (V, E)$ where V is a set of vertices, and E is a set of edges, a simple random walk on a finite graph G is a Markov chain with state space V and transition matrix P defined as follows; for $j, k \in V$

$$P(j,k) = \begin{cases} \frac{1}{\deg(j)} & \text{if } x \sim y \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the stationary distribution π of a simple random walk on a finite graph G is

$$(\frac{\deg(x_1)}{2|E|}, \frac{\deg(x_2)}{2|E|}, \ldots, \frac{\deg(x_n)}{2|E|})$$

for all $x_i \in V$.

The topics that we are interested in are how many steps are required to get close to the uniform distribution. We will use the total variance distance to measure the distance between probability measures.

Definition 4.1.5. (Total Variation Distance). Let μ and ν be probability measures on a measure space $(\mathscr{X}, \mathscr{B})$. The total variation distance between μ and ν is

$$||\mu - \nu||_{\mathrm{TV}(\mathscr{X})} = \sup_{A \in \mathscr{B}} |\mu(A) - \nu(A)|.$$

There is a nice identity relation between total variation distance and L^1 norm. We first recall L^1 norm.

Definition 4.1.6. (L^1 norm) Let f be a function in $\mathbb{C}^G$. The L^1 norm of the function f is defined as

$$||f||_1 = \sum_{g \in G} |f(g)|.$$

Proposition 4.1.3. *Let P and Q be probability measure on a finite group G. Then*

$$||P - Q||_{\mathrm{TV}} = \frac{1}{2}||P - Q||_1.$$

Proof. Let

$$A = \{g \in G | P(g) > Q(g)\}$$
$$B = \{g \in G | P(g) \leq Q(g)\}.$$

We claim $||P-Q||_{\text{TV}} = P(A)-Q(A) = Q(B)-P(B)$. If $P=Q$, then we are done. So suppose $P \neq Q$. Then it follows that there exists $g \in G$ such that $P(g) > Q(g)$. Now let

$$||P-Q||_{\text{TV}} = |P(C)-Q(C)| \tag{4.3}$$

for $C \subseteq G$. We claim if $||P-Q||_{\text{TV}} = |P(C)-Q(C)|$, then $||P-Q||_{\text{TV}} = |P(C^{c})-Q(C^{c})|$. A direct calculation shows it is true;

$$\begin{aligned}
|P(C^{c})-Q(C^{c})| &= |P(G\setminus C)-Q(G\setminus C)| \\
&= |1-P(C)-1+Q(C)| \\
&= |Q(C)-P(C)| \\
&= |P(C)-Q(C)| \\
&= ||P-Q||_{\text{TV}}.
\end{aligned}$$

Hence

$$||P-Q||_{\text{TV}} = |P(C^{c})-Q(C^{c})| = |P(C)-Q(C)|. \tag{4.4}$$

Suppose $P(g) > Q(g)$ for some $g \in G$ with $g \in C$; otherwise, from equation 4.4, we can assume $g \in C^{c}$ and get the same result. We want show $C \subseteq A$; since $|P(A)-Q(A)| > |P(C)-Q(C)| = ||P-Q||_{\text{TV}}$, showing $C \subseteq A$ proves $||P-Q||_{\text{TV}} = |P(A)-Q(A)|$ as required. First observe that $|P(C)-Q(C)| = P(C)-Q(C)$. Assume, for the sake of contradiction, $|P(C)-Q(C)| = Q(C)-P(C)$. Then

$$\begin{aligned}
|P(C\setminus\{g\})-Q(C\setminus\{g\})| &= Q(C\setminus\{g\})-P(C\setminus\{g\}) \\
&= \Big(Q(C)-P(C)\Big)+\Big(P(g)-Q(g)\Big).
\end{aligned}$$

From our assumption $P(g)-Q(g) > 0$. Therefore,

$$\begin{aligned}
|P(C\setminus\{g\})-Q(C\setminus\{g\})| &> Q(C)-P(C) \\
&= |P(C)-Q(C)| \\
&= ||P-Q||_{\text{TV}}.
\end{aligned} \tag{4.5}$$

However the inequality 4.5 contradicts equation 4.3 and the definition of the total variation distance. Thus, we have reached a contradiction and $||P-Q||_{\text{TV}} = |P(C)-Q(C)| = P(C)-Q(C)$. Therefore, $C \subseteq A$. Thus our assumption was wrong and $C \subseteq A$. It follows that $|P-Q|_{\text{TV}} = P(A) - Q(A)$. A similar argument asserts $|P-Q|_{\text{TV}} = Q(B) - P(B)$. Thus,

$$|P-Q|_{\text{TV}} = P(A) - Q(A) = Q(B) - P(B). \tag{4.6}$$

Applying equation 4.6 to $||P-Q||_{\text{TV}}$ gives

$$\begin{aligned}
||P-Q||_{\text{TV}} &= \frac{1}{2}\Big(P(A) - Q(A) + Q(B) - P(B)\Big) \\
&= \frac{1}{2}\Big(\sum_{g\in A}|P(g) - Q(g)| + \sum_{g\in B}|P(g) - Q(g)|\Big) \\
&= \frac{1}{2}\Big(\sum_{g\in G}|P(g) - Q(g)|\Big) \\
&= \frac{1}{2}||P-Q||_1,
\end{aligned}$$

the desired result. □

Note that total variation distance sums over all the subsets of the state space. By Proposition 4.1.3, we can get the total variation distance more easily.

We now introduce the upper bound Lemma introduced by Diaconis and Shahshahani. This Lemma provides a tool to determine the distance.

Lemma 4.1.1. *(Upper Bound Lemma). Let Q be a probability measure on a finite abelian group G and $\mathbb{U}$ be a uniform measure on G. Then*

$$||Q-\mathbb{U}||^2_{\text{TV}} \leq \frac{1}{4}\sum_{\chi\in\hat{G}\setminus\{\chi_1\}}|\hat{Q}(\chi)|^2.$$

Proof. By Proposition 4.1.3,

$$\begin{aligned}
||Q-\mathbb{U}||^2_{\text{TV}} &= \Big(\frac{1}{2}|Q-\mathbb{U}|_1\Big)^2 \\
&= \frac{1}{4}\Big(|Q-\mathbb{U}|_1\Big)^2 \\
&= \frac{1}{4}\Big(\sum_{g\in G}|Q(g) - \mathbb{U}(g)|\Big)^2 \\
&= \frac{1}{4}\Big(\sum_{g\in G}|Q(g) - \mathbb{U}(g)|\cdot\chi_1(g)\Big)^2 \\
&= \frac{1}{4}\Big(|G|\big\langle|Q-\mathbb{U}|, \chi_1\big\rangle\Big)^2.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality yields,

$$\begin{aligned}||Q-\mathbb{U}||_{\mathrm{TV}}^2 &\le \frac{1}{4}\Big(|G||Q-\mathbb{U}||\chi_1|\Big)^2\\ &= \frac{1}{4}|G|^2|Q-\mathbb{U}|^2\cdot 1.\end{aligned}$$

The Plancherel identity (Corollary 2.2.6) gives

$$\begin{aligned}||Q-\mathbb{U}||_{\mathrm{TV}}^2 &\le \frac{1}{4}|G||\widehat{Q}-\widehat{\mathbb{U}}|^2\\ &= \frac{1}{4}|G|\Big(\langle\widehat{Q},\widehat{Q}\rangle+\langle\widehat{\mathbb{U}},\widehat{\mathbb{U}}\rangle-2\langle\widehat{Q},\widehat{\mathbb{U}}\rangle\Big).\end{aligned} \tag{4.7}$$

Recall from our previous observation that $\widehat{\mathbb{U}}(\chi)=\delta_{\chi_1}$. Calculating inner products gives

$$\begin{aligned}\langle\widehat{Q},\widehat{Q}\rangle &= \frac{1}{|G|}\sum_{\chi\in\widehat{G}}\widehat{Q}(\chi)\overline{\widehat{Q}}(\chi),\\ \langle\widehat{\mathbb{U}},\widehat{\mathbb{U}}\rangle &= \frac{1}{|G|}\sum_{\chi\in\widehat{G}}\widehat{\mathbb{U}}(\chi)\overline{\widehat{\mathbb{U}}}(\chi)=\frac{1}{|G|}\sum_{\chi\in\widehat{G}}\delta_{\chi_1}\cdot\delta_{\chi_1}=\frac{1}{|G|},\\ \langle\widehat{Q},\widehat{\mathbb{U}}\rangle &= \frac{1}{|G|}\sum_{\chi\in\widehat{G}}\widehat{Q}(\chi)\overline{\widehat{\mathbb{U}}}(\chi)=\frac{1}{|G|}\sum_{\chi\in\widehat{G}}\widehat{Q}(\chi)\overline{\delta_{\chi_1}}(\chi)=\frac{1}{|G|}\widehat{Q}(\chi_1),\\ &= \langle Q,\chi_1\rangle=\frac{1}{|G|}\sum_{g\in G}Q(g)\chi_1(g)=\frac{1}{|G|}\sum_{g\in G}Q(g)\cdot 1=\frac{1}{|G|}.\end{aligned} \tag{4.8}$$

We can make substitutions in the equation 4.7 by using equations 4.8 to get the desired result;

$$\begin{aligned}||Q-\mathbb{U}||_{\mathrm{TV}}^2 &\le \frac{1}{4}|G|\Big(\langle\widehat{Q},\widehat{Q}\rangle+\langle\widehat{\mathbb{U}},\widehat{\mathbb{U}}\rangle-2\langle\widehat{Q},\widehat{\mathbb{U}}\rangle\Big)\\ &= \frac{1}{4}|G|\Big(\frac{1}{|G|}\sum_{\chi\in\widehat{G}}\widehat{Q}(\chi)\overline{\widehat{Q}}(\chi)+\frac{1}{|G|}-2\frac{1}{|G|}\Big)\end{aligned}$$

Observe that $\widehat{Q}(\chi_1)=\sum_{g\in G}Q(g)\chi_1(g)=\sum_{g\in G}Q(g)=1$. Then

$$\begin{aligned}||Q-\mathbb{U}||_{\mathrm{TV}}^2 &\le \frac{1}{4}|G|\Big(\frac{1}{|G|}\sum_{\chi\in\widehat{G}\setminus\{\chi_1\}}|\widehat{Q}(\chi)|^2+\frac{1}{|G|}+\frac{1}{|G|}-\frac{2}{|G|}\Big)\\ &= \frac{1}{4}\sum_{\chi\in\widehat{G}\setminus\{\chi_1\}}|\widehat{Q}(\chi)|^2.\end{aligned}$$

This completes the proof. □

Remark. We can apply upper bound lemma to random walk by setting $Q = P^{*k}$. Then we get $||P^{*k} - \mathbb{U}||^2_{\text{TV}} \leq \frac{1}{4}\left|\widehat{P}(\chi)\right|^{2k}$.

4.2 Spectrum of Graph

Definition 4.2.1. (Spectrum of the Graph). Let $G = (V, E)$ be a graph with V as the set of vertices and E as the set of edges. The spectrum of the graph is the set of eigenvalues of the adjacency matrix of graph G.

Remark. By the spectral theorem for symmetric matrices, we know an adjacency matrix of a graph G has real eigenvalues.

Recall that given a group G, $\mathbb{C}^G$ is a set of all functions from G to $\mathbb{C}$.

Definition 4.2.2. (Spectrum of the Random Walk). Let P be a probability measure on a finite group G, and let $T : \mathbb{C}^G \to \mathbb{C}^G$ be a convolution operator given by $T(h) = P * h$. The spectrum $\operatorname{spec}(P)$ of the random walk driven by a probability measure P is the set of all eigenvalues of the linear operator T.

Lemma 4.2.1. *Let P be a probability measure on a finite group G. Then*

$$\operatorname{spec}(P) \subseteq \{z \in \mathbb{C} \,|\, |z| \leq 1\}.$$

Proof. For $\widehat{P}(\chi) \in \operatorname{spec}(P)$

$$\begin{aligned}|\widehat{P}(\chi)| &= |G| \cdot |\langle P, \chi\rangle| \\ &= |\sum_{g\in G} P(g)\overline{\chi(g)}| \\ &\leq \sum_{g\in G} P(g)|\overline{\chi(g)}|.\end{aligned}$$

Since $|\overline{\chi(g)}| = 1$ for all $g \in G$,

$$|\widehat{P}(\chi)| \leq \sum_{g\in G} P(g) = 1.$$

□

Theorem 4.2.2. *Let G be a finite abelian group and f be an element of $\mathbb{C}^G$. If $T : \mathbb{C}^G \to \mathbb{C}^G$ is a convolution operator given by $T(h) = f * h$, then T is a diagonalizable linear operator.*

Proof. We first show T is a linear operator. Let $h_1, h_2 : G \to \mathbb{C}$. Then for $g_1 \in G$,

$$\begin{aligned} T(h_1 + h_2)(g_1) &= f * (h_1 + h_2)(g_1) \\ &= \sum_{g_2 \in G} f(g_1 g_2^{-1})(h_1 + h_2)(g_2) \\ &= \sum_{g_2 \in G} f(g_1 g_2^{-1})\Big(h_1(g_2) + h_2(g_2)\Big) \\ &= \sum_{g_2 \in G} \Big(f(g_1 g_2^{-1})h_1(g_2)\Big) + \sum_{g_2 \in G} \Big(f(g_1 g_2^{-1})h_2(g_2)\Big) \\ &= f * h_1(g_1) + f * h_2(g_1) \\ &= T(h_1)(g_1) + T(h_2)(g_1). \end{aligned}$$

Hence, T is a linear operator.

Recall from Schur's orthogonality relations (Theorem 1.4.7), irreducible characters form an orthonormal set. Hence showing T has irreducible characters as eigenvectors with eigenvalue $\widehat{f}(\chi)$ proves T is diagonalizable.

Let $\chi, \psi \in \widehat{G}$. From Theorem 2.2.3, we know

$$\widehat{f * \chi} = \widehat{f} \cdot \widehat{\chi}.$$

By Schur's orthogonality relations (Theorem 1.4.7), we know $\widehat{\chi}(\psi) = |G|\langle \chi, \psi \rangle = |G|$ if and only if $\chi = \psi$. Hence,

$$\widehat{f * \chi}(\psi) = \widehat{f}(\chi) \cdot |G| \delta_\chi. \tag{4.9}$$

Applying the inverse Fourier transforms to (4.9) gives

$$f * \chi = \widehat{f}(\chi) \cdot \chi$$

or equivalently,

$$T(\chi) = \widehat{f}(\chi)\chi.$$

Hence, χ is an eigenvector with the eigenvalue $\widehat{f}(\chi)$. Therefore we conclude that T is diagonalizable. $\square$

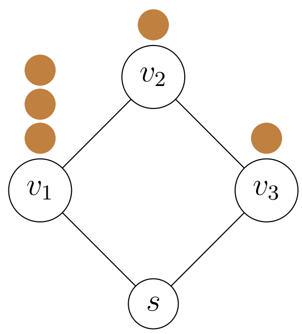
Applying theorem (Theorem 4.2.2) to random walks yields an immediate corollary.

Corollary 4.2.3. *Let G be a finite abelian group and let P be a probability measure on G. Then the set of all characters χ such that $\widehat{P}(\chi) = \lambda$ forms an orthonormal basis for the eigenspace of λ.*

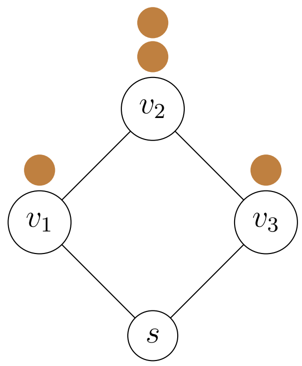
5 Abelian Sandpile Model

5.1 Abelian Sandpile Model and Laplacian Operator

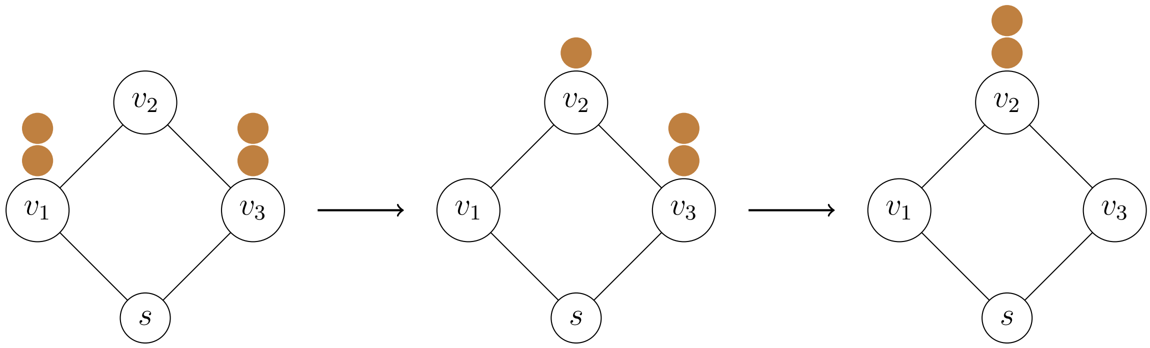
Let G be a graph with chips or grains of sand on its vertices as below; there are three grains of sand on v_1, one grain of sand on each of v_2 and v_3, and no sand on v_s.



We call a vertex is stable if it has fewer grains of sand than its degree; otherwise, we call a vertex unstable. Note that vertex v_1 is unstable. We can topple or fire an unstable vertex by giving one sand to each neighboring vertices. We designate one vertex as a sink, and any grains of sands that falls into the sink gets removed; the graph G has s as a sink vertex. After toppling the vertex v_1, we get the following configuration:

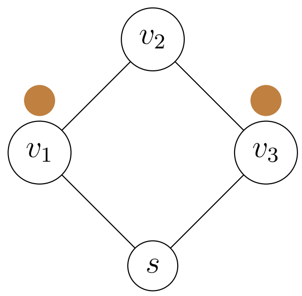


Observe that a sand that passed to the sink vertex s is removed. Also notice that toppling of v_1 made the other vertex v_2 be unstable. The below diagram shows the stabilization of a sandpile on G. At first, v_2 is toppled, and v_1 and v_3 are toppled in order.



Later, we will see that the order of toppling does not impact the final configuration of a sandpile, which is called an abelian property of the sandpiles.

Observer that v_2 is still unstable from the last configuration. Hence we topple v_2 again to get the stabilization.



We now have all the stable vertices. When all the vertices are stable, the configuration is called stable. There exists no more vertex to topple. Hence we stop the toppling.

Definition 5.1.1. (Abelian Sandpile Model). Let $G = (V, E, s)$ be a graph with a set of vertices V, a set of edges E, and $s \in V$ as a sink. The abelian sandpile model is a graph G with each vertex associated with a value that corresponds to the number of grains of sand on the vertex. Passed grains of sand to the sink s is removed.

Definition 5.1.2. (Stable). A vertex v is stable if it has less grains of sand than its degree.

Definition 5.1.3. (Unstable). A vertex v is unstable if it has equal or more grains of sand than its degree.

Definition 5.1.4. (Non-Sink Vertices). $\tilde{V}$ denotes a set of non-sink vertices of G:

$$\tilde{V} := V \backslash \{s\}$$

Definition 5.1.5. (Configuration). A configuration of G is given by

$$\text{Config}(G) := \mathbb{Z}\tilde{V} := \{\sum_{v \in \tilde{V}} c(v)v : c(v) \in \mathbb{Z} \text{ for all } v \in V\}$$

Remark. Since sandpiles have only nonnegative integer amounts of sand grains on each vertex, $c(v) \geq 0$ for all $v \in V$.

Definition 5.1.6. (Stable). A configuration c of G is stable if $c(v) < \deg(v)$ for all $v \in \tilde{V}$.

Definition 5.1.7. (Toppling). One unstable vertex can be toppled by sending out a grain of sand to each neighboring vertex from the unstable vertex.

As we observed previously, toppling might cause other vertices to be unstable and result in an avalanche of topplings.

Suppose a vertex v is unstable. Then toppling of vertices other than v would never make v stable; hence in order to make the vertex v to be stable, we need to topple v.

We denote illegal firing as sequences of vertex firings that might contain firing stable vertices; we call firing only unstable vertices a legal firing.

Given a configuration c, let us denote by c° the stabilization of c after sequences of toppling σ. In symbols,

$$c \xrightarrow{\sigma} c^\circ$$

Definition 5.1.8. (Global Sink). A vertex s designated as a global sink is globally accessible; meaning that there exists a path from each vertex to the sink.

Definition 5.1.9. (Adjacency Matrix). Given a graph $G = (V, E)$, the adjacency matrix A is $|V|\times|V|$ matrix defined by

$$A_{ij} = \begin{cases} 1 & \text{if } V_i \sim V_j, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 5.1.10. (Full Laplacian). Given a sandpile $G = (V, E, s)$, the full Laplacian Δ is defined by

$$\Delta = \deg(G) - A$$

where $\deg(G) = \operatorname{diag}\Big(\deg(v_1), \cdots, \deg(v_{|V|-1}), \deg(s)\Big)$ and A is the adjacency matrix of G.

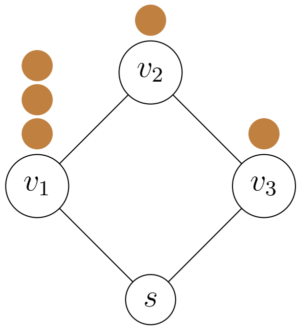
We note that the rows of full Laplacian matrix sum to zero. This is because the amount of grains lost from the unstable vertex equals to the amount of sand gained by its neighboring vertices.

Definition 5.1.11. (Reduced Laplacian). Given a sandpile $G = (V, E, s)$, the reduced Laplacian $\widetilde{\Delta}$ is defined by

$$\widetilde{\Delta} = \widetilde{\deg}(G) - \widetilde{A}$$

where $\widetilde{\deg}(G) = \operatorname{diag}\Big(\deg(v_1), \cdots, \deg(v_{|V|-1})\Big)$ and $\widetilde{A}$ is the reduced adjacency matrix where row and column corresponding to the sink are deleted.

Let us calculate the full Laplacian and the reduced Laplacian for the following sandpile.



The full Laplacian is given by

$$\Delta = \deg(G) - A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}.$$

The associated reduced Laplacian is

$$\widetilde{\Delta} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Observe that the configuration of the above sandpile is $4v_1 + v_2 + v_3$. As we seen before, toppling the vertex v_1 results in the configuration $v_1 + 2v_2 + v_3$. The corresponding calculation can be done by using the reduced Laplacian:

$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

5.2 Existence and Uniqueness of Stabilization

Lemma 5.2.1. *(Least Action Principle). Let c be a configuration on G. Let σ be a legal toppling sequence to stabilize c. then for any sequence of toppling τ that stabilizes c, $\sigma \leq \tau$.*

Proof. Let $v_1, v_2, \cdots, v_n$ be a legal toppling procedure according to σ. We will prove by induction on n. When $n = 0$, the statement is trivially true.

For induction step, suppose the statement is true for $n - 1 > 0$. Observe that since v_1 is unstable, the only way to make v_1 stable is to topple it at least once; as we observed before, toppling vertices other than v_1 would not make v_1 stable. Hence for any sequence of toppling τ that stabilizes c, $\tau(v_1) > 0$.

We now fire v_1 and get configuration c'. Then $v_2, v_3, \cdots, v_n$ is a legal sequence of toppling that results stabilization given the configuration c'. Define

$$\tau' := \tau - v_1.$$

Then τ' is a sequence of toppling that results stabilization of configuration c'. By inductive hypothesis we know

$$\sigma - v_1 \leq \tau'.$$

By adding v_1 on the both sides, we get

$$\sigma \leq \tau' + v_1 = \tau$$

as desired. □

Corollary 5.2.2. *Let c be a configuration. If $c \xrightarrow{\sigma} c_1$, $c \xrightarrow{\tau} c_2$, and c_1 and c_2 are stable, then $\sigma = \tau$ and $c_1 = c_2$.*

Lemma 5.2.3. *If a graph $G = (V, E)$ has a global sink, then every configuration has a stabilization.*

Proof. Let s denotes a global sink and N be the total number of chips in the configuration c. Given a vertex $v_0 \in V$, suppose $v_0, v_1, \cdots, v_n$ is a path from

v_0 to $v_n = s$. Then observe that v_{n-1} fire at most N times since every time v_{n-1} fires a chip falls into the sink $v_n = s$.

Note that v_{n-2} needs to fire at most $\deg(v_{n-1})$ times to cause v_{n-1} to fire so that a chip fall into a sink. Hence v_{n-2} fire at most $\deg(v_{n-1}) \cdot N$ times. With similar logic, we see that $v \in V$ fires at most $\deg(v_1) \cdot \deg(v_2) \cdots \deg(v_{n-1}) \cdot N$ times. It follows that the configuration c has only finitely many legal firing sequence, hence c has a stabilization. $\square$

Let c be a configuration. Define the chip addition operator E_v that adds a single chip at vertex v and then stabilizes:

$$E_v c = (c + 1_v)^\circ$$

Lemma 5.2.4. *For any graph with a global sink, the chip addition operators commute.*

Proof. Let c be a configuration on a graph G. Observe that whatever vertices unstable in $c + 1_v$ is also unstable in $c' = c + 1_v + 1_w$. We apply the firing sequence that stabilizes $c + 1_v$. Then we obtain the configuration $E_v c + 1_w$. After stabilization we get $E_w E_v c$ at the end.

By stabilizing $c + 1_w$ first, we get $E_v E_w c$. By the uniqueness of stabilization, we conclude that $E_w E_v c = E_v E_w c$ as desired. $\square$

Remark. This lemma justifies the term "abelian sandpile model."

6 Sandpile Dynamics on Tiling Graphs

6.1 Sandpile Dynamics

Sandpile dynamics on a finite connected graph $G = (V, E)$ may be described as follows. In the model, a node $s \in V$ is designated sink. Each non-sink vertex v is assigned a non-negative number $\sigma(v)$ of chips. If at some point $\sigma(v) \geq \deg(v)$ the vertex can topple, passing one chip to each neighbor; if a chip falls on the sink it is lost from the model.

Definition 6.1.1. (Sandpile). A sandpile on a graph G is a map $\sigma : G \to \mathbb{Z}_{\geq 0}$.

Definition 6.1.2. (Full Sandpile). The map $\sigma_{\text{full}} = \deg - 1$ is the full sandpile.

The set of stable sandpiles is indicated

$$\mathscr{S}(G) = \{\sigma : G \to \mathbb{Z}_{\geq 0} : \sigma \leq \sigma_{\text{full}}\}. \tag{6.1}$$

Definition 6.1.3. (Stable). A configuration σ is called stable if $\sigma(v) < \deg(v)$ for all $v \in V \setminus \{s\}$.

The dynamics in the model occur in discrete time steps, in which a chip is added to the model at a uniform random vertex, then all legal topplings occur until the model reaches a stable state.

Definition 6.1.4. (Recurrent). A sandpile c is recurrent if it is stable and for every configuration a, there exists a sandpile configuration b such that $c = (a + b)^\circ$.

Given a graph G, the set of recurrent sandpiles on the graph form an abelian group.

The set of recurrent states form the sandpile group and are indicated $\mathscr{G}(G)$. Its dual group is $\hat{\mathscr{G}}$.

Recall the definition of Laplacian operator given in previous chapter.

Definition 6.1.5. (Graph Laplacian). Denote Δ the graph Laplacian $\Delta f(v) = \deg(v)f(v) - \sum_{(v,w)\in E} f(w)$.

When G is a finite graph and a node s has been designated sink, the reduced Laplacian Δ' is obtained from Δ by removing the row and column corresponding to the sink.

Since the sandpile group of a graph with sink s is isomorphic to $\mathscr{G} = \mathbb{Z}^{V\setminus\{s\}}/\Delta'\mathbb{Z}^{V\setminus\{s\}}$ where Δ' is the reduced graph Laplacian obtained by omitting the row and column corresponding to the sink, the dual group is isomorphic to $\hat{\mathscr{G}} = (\Delta')^{-1}\mathbb{Z}^{V\setminus\{s\}}/\mathbb{Z}^{V\setminus\{s\}}$. Thus Δ' provides a natural mapping from $\hat{\mathscr{G}} \to \mathscr{G}$. A map in the reverse direction may be constructed via convolution with the graph Green's function.

Definition 6.1.6. (Harmonic Modulo 1). Given a function f on $\mathscr{T}$, say that f is harmonic modulo 1 if $\Delta f \equiv 0 \bmod 1$ and denote the set of such functions $\mathscr{H}(\mathscr{T})$.

Let
$$\mathscr{H}(G) = \{f : G \to \mathbb{R}, \Delta f \equiv 0 \bmod 1\}. \tag{6.2}$$

Since the random walk considered is a random walk on an abelian group, in terms of the mixing behavior there is no loss in assuming that the walk is started at the identity. Also, the transition kernel is diagonalized by the Fourier transform, that is, the characters, for $\xi \in \hat{\mathscr{G}}$, $\chi_\xi(g) = e^{2\pi i \xi(g)}$ are eigenfunctions for the transition kernel, and the eigenvalues are the Fourier coefficients
$$\hat{\mu}(\xi) = \frac{1}{|V|}\left(1 + \sum_{v\in V\setminus\{s\}} e(\xi_v)\right). \tag{6.3}$$

Definition 6.1.7. (Group Convolution). A random walk driven by a probability measure μ on a group has distribution at step n given by μ^{*n} where $\mu^{*1} = \mu$ and $\mu^{*n} = \mu * \mu^{*(n-1)}$ is the group convolution.

We recall some of the definitions.

Definition 6.1.8. (Total Variation Distance). The total variation distance between two probability measures μ and ν on a measure space $(\mathscr{X}, \mathscr{B})$ is

$$\|\mu - \nu\|_{\mathrm{TV}(\mathscr{X})} = \sup_{A \in \mathscr{B}} |\mu(A) - \nu(A)|. \tag{6.4}$$

Definition 6.1.9. (Total Variation Mixing Time). Given a measure μ driving sandpile dynamics on the group of recurrent sandpile states $\mathscr{G}(G)$ with uniform measure $\mathbb{U}_{\mathscr{G}}$, the total variation mixing time is

$$t_{\mathrm{mix}} = \min\left\{k : \left\|\mu^{*k} - \mathbb{U}_{\mathscr{G}(G)}\right\|_{\mathrm{TV}(\mathscr{G}(G))} < \frac{1}{e}\right\}. \tag{6.5}$$

Definition 6.1.10. (Cut-Off Phenomenon). Given a sequence of graphs G_n the sandpile dynamics is said to satisfy the cut-off phenomenon in total variation if, for each $\epsilon > 0$,

$$\left\|\mu^{*\lceil(1-\epsilon)t_{\mathrm{mix}}\rceil} - \mathbb{U}_{\mathscr{G}(G_n)}\right\|_{\mathrm{TV}(\mathscr{G}(G_n))} \to 1,$$
$$\left\|\mu^{*\lfloor(1+\epsilon)t_{\mathrm{mix}}\rfloor} - \mathbb{U}_{\mathscr{G}(G_n)}\right\|_{\mathrm{TV}(\mathscr{G}(G_n))} \to 0$$

as $n \to \infty$.

6.2 Sandpiles with Periodic and Open Boundary

Definition 6.2.1. (Periodic Space Tiling). let M be a non-singular $d \times d$ matrix, and let $\Lambda = M \cdot \mathbb{Z}^d < \mathbb{R}^d$ be a d-dimensional lattice. A (periodic) space tiling $\mathscr{T}$ is a connected graph embedded in $\mathbb{R}^d$ which is connected, is Λ-periodic, has finitely many vertices in a fundamental domain for $\mathbb{R}^d/\Lambda$, and has bounded degree.

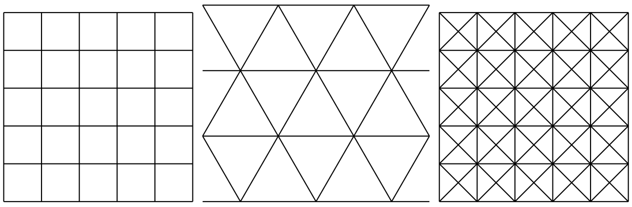


Fig. 6.1: The square, triangular and tetrakis square lattices are examples of tilings with reflecting families of lines such that the quotient by the reflection group is a bounded convex region of the plane.

Suppose without loss of generality that 0 is a vertex in $\mathscr{T}$. Given an integer $m \geq 1$, two types of graphs are considered.

(1) (Torus boundary condition) The graph $\mathbb{T}_m = \mathscr{T}/m\Lambda$ consists of m^d fundamental domains with opposite faces identified. By convention, 0 is designated sink.

(2) (Open boundary condition) In two dimensions, assume that there are vectors $v_1, ..., v_k$ in which $\mathscr{T}$ has translational symmetry, and lines $\ell_1, ..., \ell_k$, $\ell_i = \{x \in \mathbb{R}^2 : \langle x, v_i \rangle = 0\}$ such that $\mathscr{T}$ has reflection symmetry in the family of lines

$$\mathscr{F} = \{nv_i + \ell_i : 1 \leq i \leq k, n \in \mathbb{Z}\}. \tag{6.6}$$

In this case, let $\mathscr{R}$ be an open, connected, convex region cut out by some of the lines, and assume further that $\mathbb{R}^2$ is tiled by the reflections of $\mathscr{R}$ in the family of lines and that any sequence of reflections which maps $\mathscr{R}$ to itself is the identity map.

The results concerning sandpile dynamics are proved by studying the spectrum of the sandpile transition kernel.

6.3 Spectral Gap and Spectral Factors

In the case of a torus boundary condition, define the spectral parameter

$$\gamma = \inf \left\{ \sum_{x \in \mathscr{T}} 1 - \cos(2\pi\xi_x) : \xi \in \mathscr{H}(\mathscr{T}), \xi \not\equiv 0 \bmod 1 \right\}. \tag{6.7}$$

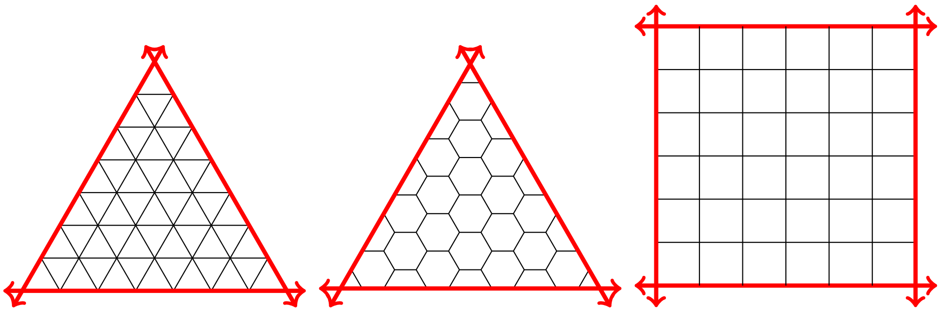


Fig. 6.2: The triangular, hex and square lattice configurations with open boundary condition.

In two dimensions, let $\mathscr{L}$ denote the set of lines which make up a segment of the boundary of $\mathscr{R}$ and let $\mathscr{C}$ be the set of pairs of lines from $\mathscr{L}$ which intersect at a corner of the boundary of $\mathscr{R}$. Write an affine line $a \in \mathscr{L}$ as $a = nv + \ell$ where $v \in \mathbb{R}^2$ and ℓ is the perpendicular line. A pair of affine lines $(a_1, a_2) \in \mathscr{C}$ have ℓ_1 and ℓ_2 that split $\mathscr{T}$ into four quadrants. Let $Q_{(a_1,a_2)}$ be the quadrant whose translate contains $\mathscr{R}$. Given $a \in \mathscr{L}$, let $\mathscr{H}_a(\mathscr{T})$ be those functions $\xi \in \mathscr{H}(\mathscr{T})$ which are anti-symmetric in ℓ, similarly given $(a_1, a_2) \in \mathscr{C}$, let $\mathscr{H}_{(a_1,a_2)}(\mathscr{T})$ be those functions in $\mathscr{H}(\mathscr{T})$ which are anti-symmetric in ℓ_1 and ℓ_2. Define spectral parameters

$$\begin{aligned}
\gamma_0 &= \inf_{\substack{\xi \in \mathscr{H}(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}} 1 - \cos(2\pi \xi_x) \\
\gamma_1 &= \frac{1}{2} \inf_{a \in \mathscr{L}} \inf_{\substack{\xi \in \mathscr{H}_a(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}} 1 - \cos(2\pi \xi_x) \\
\gamma_2 &= \inf_{(a_1,a_2) \in \mathscr{C}} \inf_{\substack{\xi \in \mathscr{H}_{(a_1,a_2)}(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in Q_{(a_1,a_2)}} 1 - \cos(2\pi \xi_x).
\end{aligned}$$

Let $\mathfrak{S}_S$ be the group generated by reflections in S, and let $\mathscr{H}_S(\mathscr{T})$ denote those harmonic modulo 1 functions which are anti-symmetric in $H_{j,0}$ for all $j \in S$, identified with functions on $\mathscr{T}/\mathfrak{S}_S$. Again, for $0 \leq i \leq d$ define the spectral parameters

$$\gamma_i = \inf_{\substack{S \subset \{1,2,\ldots,d\} \\ |S| = i}} \inf_{\substack{\xi \in \mathscr{H}_S(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}/\mathfrak{S}_S} 1 - \cos(2\pi \xi_x). \tag{6.8}$$

Thus $\gamma_0 = \gamma$. In dimension $d \geq 2$ define the jth spectral factor

$$\Gamma_j = \frac{d-j}{\gamma_j} \tag{6.9}$$

and $\Gamma = \max_j \Gamma_j$.

The following theorem determines the spectral of sandpile dynamics for plane and space tiling graphs.

Theorem 6.3.1. *Given a tiling $\mathscr{T}$, as $m \to \infty$, the spectral gap of the transition kernel of sandpile dynamics on $\mathbb{T}_m$ satisfies*

$$\mathrm{gap}_{\mathbb{T}_m} = (1+o(1))\frac{\gamma}{|\mathbb{T}_m|}. \tag{6.10}$$

If $\mathscr{T}$ has a family of reflection symmetries $\mathscr{F}$ and satisfies condition A, then the spectral gap of the transition kernel of sandpile dynamics on $\mathscr{T}_m$ satisfies

$$\mathrm{gap}_{\mathscr{T}_m} = (1+o(1))\frac{\min_j \gamma_j}{|\mathscr{T}_m|}. \tag{6.11}$$

The following theorem demonstrates a cut-off phenomenon in sandpile dynamics on general tiling graphs with either a torus or open boundary condition. Whereas the mixing of sandpiles with torus boundary condition is controlled by the spectral gap, when there is an open boundary condition, the mixing time is controlled by the spectral factor.

Theorem 6.3.2. *For a fixed tiling $\mathscr{T}$ in $\mathbb{R}^d$, sandpiles started from a recurrent state on $\mathbb{T}_m$ have asymptotic total variation mixing time*

$$t_{\mathrm{mix}}(\mathbb{T}_m) \sim \frac{d}{2\gamma}|\mathbb{T}_m|\log m \tag{6.12}$$

with a cut-off phenomenon as $m \to \infty$.

If the tiling $\mathscr{T}$ satisfies the reflection condition and condition A then sandpile dynamics started from a recurrent configuration on $\mathscr{T}_m$ have total variation mixing time

$$t_{\mathrm{mix}}(\mathscr{T}_m) \sim \frac{\Gamma}{2}|\mathscr{T}_m|\log m \tag{6.13}$$

with a cut-off phenomenon as $m \to \infty$.

Motivated by Theorem 6.3.2, if $\Gamma = \Gamma_0$ say that the bulk or top dimensional behavior controls the total variation mixing time, and otherwise that the boundary behavior controls the total variation mixing time. The proof of Theorem 6.3.2 will in fact generate a statistic which randomizes at the mixing time, and this statistic is either distributed throughout the graph, or concentrated near the boundary of the dimension controlling the spectral factor.

Corollary 6.3.3. *All plane tilings satisfying the reflection condition and condition A have total variation mixing time controlled by the bulk behavior.*

Proof. It suffices that $\Gamma_1 \leq \Gamma_0$. Indeed, the factor of 2^{-1} in γ_1 is canceled by the ratio $\frac{2}{2-1}$ of dimensions, and the anti-symmetry condition imposes an extra constraint on the harmonic modulo 1 function in the inf, so that $\frac{1}{2}\gamma_0 \leq \gamma_1$. □

In particular, Corollary 6.3.3 implies that asymptotic mixing time of sandpile dynamics on the square grid with open and periodic boundary condition are the same to top order.

Theorem 6.3.4. *The triangular tiling has periodic boundary spectral parameters*

$$\gamma_{\text{tri}} = 1.69416(6).$$

Remark. The digit in parenthesis indicates the last significant digit.

The determination of the Green's function in the tiling as opposed to lattice case is more involved. It is reduced to the lattice case by stopping a random walk on the tiling when it hits the period lattice, and using the resulting stopped measure to determine the Green's function restricted to the lattice.

6.4 Optimization problem and computer search

In this section the spectral parameters are determined by computer search for several tilings. Recall that

$$\gamma = \inf_{\substack{\xi \in \mathscr{H}(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}} 1 - c(\xi_x).$$

For two dimensional tilings set

$$\gamma_0 = \inf_{\substack{\xi \in \mathscr{H}(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}} 1 - c(\xi_x)$$

$$\gamma_1 = \frac{1}{2} \inf_{a \in \mathscr{L}} \inf_{\substack{\xi \in \mathscr{H}_a(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}_a} 1 - c(\xi_x)$$

$$\gamma_2 = \inf_{(a_1,a_2) \in \mathscr{C}} \inf_{\substack{\xi \in \mathscr{H}_{(a_1,a_2)}(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in Q_{(a_1,a_2)}} 1 - c(\xi_x).$$

In higher dimensions assume that the reflecting hyperplanes are built from an orthonormal system, and

$$\gamma_i = \inf_{\substack{S \subset \{1,2,\ldots,d\} \\ |S|=i}} \inf_{\substack{\xi \in \mathscr{H}_S(\mathscr{T}) \\ \xi \not\equiv 0 \bmod 1}} \sum_{x \in \mathscr{T}/\mathfrak{S}_S} 1 - c(\xi_x).$$

The following arguments index harmonic modulo 1 function ξ with its prevector $\nu = \Delta\xi$, which is simpler as the prevector is integer valued. This permits an approximate ordering on prevectors in terms of their norm, and the diameter of their support. The harmonic modulo 1 function is then recovered as $\xi = g * \nu$.

Lemma 6.4.1. *Let S be a finite or countable set and let $\xi \in \ell^2(S)$, $\|\xi\|_\infty \leq \frac{1}{2}$. Define*

$$f_S(\xi) = \sum_{x \in S} 1 - c(\xi_x). \tag{6.14}$$

Let $\alpha > 0$ and assume $\|\xi\|_2^2 \geq \alpha$. Then

$$2\pi^2\alpha\left(1 - \frac{\pi^2}{3}\alpha\right) \leq f_S(\xi) \leq 2\pi^2\|\xi\|_2^2. \tag{6.15}$$

Proof. The Taylor series approximation for $c(x)$ on $|x| \leq \frac{1}{2}$,

$$c(x) = 1 - 2\pi^2 x^2 + \frac{2\pi^4}{3}x^4 - \cdots$$

is an alternating series with decreasing increments after the term $2\pi^2 x^2$. Thus $f_S(\xi) \leq 2\pi^2 \|\xi\|_2^2$. Let $0 < \lambda \leq 1$ and let $\xi' = \lambda\xi$ satisfy $\|\xi'\|_2^2 = \alpha$. Then $f_S(\xi') \leq f_S(\xi)$. Furthermore,

$$\begin{aligned} f_S(\xi') &\geq 2\pi^2 \|\xi'\|_2^2 - \frac{2}{3}\pi^4 \|\xi'\|_4^4 \\ &\geq 2\pi^2\alpha - \frac{2}{3}\pi^4\alpha\|\xi'\|_\infty^2 \\ &\geq 2\pi^2\alpha - \frac{2}{3}\pi^4\alpha^2. \end{aligned}$$

□

The following lemma is used to estimate the functionals $f(\xi)$.

Lemma 6.4.2. *Let $R \subset \mathscr{T}$ and let $\xi : \mathscr{T} \to \left(-\frac{1}{2}, \frac{1}{2}\right]$. Let*

$$\|\xi\|_{2,R^c}^2 = \sum_{x \in \mathscr{T} \setminus R} \xi_x^2. \tag{6.16}$$

There is a number ϑ, $|\vartheta| \leq 1$ such that

$$f(\xi) = \sum_{x \in R} (1 - c(\xi_x)) + 2\pi^2 \|\xi\|_{2,R^c}^2 - \frac{\pi^4}{3} \|\xi\|_{2,R^c}^4 + \vartheta \frac{\pi^4}{3} \|\xi\|_{2,R^c}^4. \tag{6.17}$$

Proof. By Taylor approximation, for $x \in R^c$,

$$2\pi^2\xi_x^2 - \frac{2}{3}\pi^4\xi_x^4 \leq 1 - c(\xi_x) \leq 2\pi^2\xi_x^2.$$

Thus,

$$\begin{aligned} &\sum_{x \in R} (1 - c(\xi_x)) + 2\pi^2 \|\xi_x\|_{2,R^c}^2 - \frac{2}{3}\pi^4 \|\xi\|_{2,R^c}^4 \\ &\qquad \leq f(\xi) \leq \sum_{x \in R} (1 - c(\xi_x)) + 2\pi^2 \|\xi\|_{2,R^c}^2, \end{aligned}$$

from which the claim follows. □

In practice, Lemma 6.4.2 is applied by calculating ξ_x on R from the Fourier integral representations in Section **??** in a neighborhood of 0, and calculating $\|\xi\|_2^2$ by Parseval.

The following two optimization programs are used to obtain a lower bound for $f(\xi)$. Let $\xi = g * \nu$, $\|\xi\|_\infty \leq \frac{1}{2}$. Given a set $S \subset \mathscr{T}$, a lower bound for $f(\xi)$ is obtained as the solution of the optimization program $Q(S, \nu)$,

$$\begin{aligned}
&Q(S,\nu): \\
&\text{minimize:} \quad \sum_{d(w,S)\leq 1} 1 - c(x_w) \\
&\text{subject to:} \quad \forall u \in S,\ (\deg u) x_u - \sum_{d(w,u)=1} x_w = \nu_u \\
&\qquad -\frac{1}{2} \leq x_w \leq \frac{1}{2}.
\end{aligned}$$

A lower bound for $Q(S, \nu)$ is the relaxed optimization program with positive constraints $P(S, \nu)$

$$\begin{aligned}
&P(S,\nu): \\
&\text{minimize:} \quad \sum_{d(w,S)\leq 1} 1 - c(x_w) \\
&\text{subject to:} \quad \forall u \in S,\ (\deg u) x_u + \sum_{d(w,u)=1} x_w \geq \nu_u \\
&\qquad -\frac{1}{2} \leq x_w \leq \frac{1}{2}.
\end{aligned}$$

Note that the objective function is convex and with non-degenerate Hessian in the interior with the stronger condition $|x_w| \leq \frac{1}{4}$, and hence has a unique local minima there. In order to estimate $Q(S, \nu)$ and $P(S, \nu)$ numerically, the range $\frac{1}{4} \leq |x_w| \leq \frac{1}{2}$ was split into several equal size intervals and the objective function was approximated piecewise linearly on these, obtaining a lower bound for the minimum. The minima were compared with the variables constrained to lie in each interval. Denote $P_j(S, \nu)$ and $Q_j(S, \nu)$ the programs in which both $\left[-\frac{1}{2}, -\frac{1}{4}\right]$ and $\left[\frac{1}{4}, \frac{1}{2}\right]$ are split into j equal size intervals, and objective function interpolating linearly between the values of $c(x)$ on the endpoints. Note that the minimum of P_j and Q_j on each product of intervals is determined deterministically as a unique interior minimum or boundary value. In the examples considered in dimensions 3 and higher, $\|\xi\|_2^2$ was optimized rather than $f(\xi)$, and it was demonstrated that the extremal function is the same. Programs $Q'(S, \nu)$ and $P'(S, \nu)$ have the same constraints, but have objective function $\sum_{d(w,S)\leq 1} x_w^2$. Note that this objective function is convex.

Lemma 6.4.3. *Let $G = (V, E)$ be a graph and let $v \in V$ of degree at least 2. Let $|\nu_v| = 1$. The optimization problem $Q'(\{v\}, \nu)$ has value $\frac{1}{\deg(v)(\deg(v)+1)}$.*

Proof. Since the claimed value is smaller than the value on the boundary, it may be assumed that the optimum is achieved at an interior point. By Lagrange multipliers, there is a scalar λ such that $x_v = \lambda \deg v$ and $x_w = -\lambda$ for all $(v, w) \in E$. Thus $\lambda = \frac{1}{\deg(v)(\deg(v)+1)}$. The claim follows, since

$$\sum_{d(v,w)\leq 1} x_w^2 = \lambda^2 \deg(v)(\deg(v) + 1). \tag{6.18}$$

□

The optimization programs P, P_j, P', Q, Q_j, Q' satisfy the following monotonicity properties.

Lemma 6.4.4. *The programs P, P_j, P', Q, Q_j, Q' are monotone increasing in the set S. The programs P, P_j, P' are monotone increasing in the prevector $|\nu|$.*

Proof. This follows from constraint relaxation. □

The programs also satisfy the following additivity property.

Lemma 6.4.5. *Let $B(S) = \{u : d(u, S) \leq 1\}$ be the distance 1 enlargement of S. When $S_1, S_2, ..., S_k$ are some sets in $\mathscr{T}$ whose distance 1 enlargements $B(S_1), B(S_2), ..., B(S_k)$ are pairwise disjoint, then $\sum_{i=1}^k Q(S_i, \nu) \leq f(\xi)$ and $\sum_{i=1}^k Q'(S_i, \nu) \leq \|\xi\|_2^2$.*

Proof. Since the sets of variables are disjoint, the sum of the optimization programs can be considered to be a single optimization program, which is then satisfied by the optimizing solution ξ. □

Since the remaining programs P, P', P_j, P_j' are relaxations of Q and Q', the additivity property holds for these as well.

The strategy of the arguments is now described as follows. Say two points x_i, x_t in the support of ν are 2-path connected, or just connected for short, if there is a sequence of points $x_i = x_0, x_1, ..., x_n = x_t$ in the support of ν, such that the graph distance between x_i and x_{i+1} is at most 2. By the

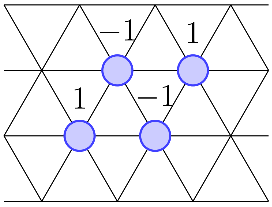


Fig. 6.3: The extremal configuration for the triangular lattice.

additivity lemma, the value of the optimization programs applied with S_i separated connected components of $\operatorname{supp} \nu$ is additive. Since the value of each optimization program is translation invariant and, for a fixed ν, monotone in S, all connected components with P or Q (resp. P', Q', P_j, Q_j) value at most a fixed constant can be enumerated by starting from a base configuration and adding connected points to the set S one at a time.

The configuration ν must be in C^ρ for $\xi \in \ell^2(\mathscr{T})$. Having enumerated all feasible connected components, the search is completed by considering all methods of gluing together several connected components which produce a $\nu \in C^\rho$.

Let the triangular lattice be generated by $v_1 = (1, 0)$ and $v_2 = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. Let $\xi^* = g * \nu^*$ with $\nu^* = \delta_0 - \delta_{v_1} - \delta_{v_2} + \delta_{v_1+v_2}$. The value

$$f(\xi^*) = 1.69416(5) \tag{6.19}$$

was estimated by Lemma 6.4.2 with

$$R = \{n_1 v_1 + n_2 v_2 : \max(|n_1|, |n_2|) \leq 10\}. \tag{6.20}$$

It is to be shown that $\gamma_{\text{tri}} = f(\xi^*)$.

First the case of a node of height 3 in the extremal prevector is ruled out.

Lemma 6.4.6. *Suppose* $|\nu_0| = 3$. *The optimization program* $P(\{0\}, |\nu|)$ *has value 2. In particular,* ν *does not achieve* γ_{tri}.

Proof. At the optimum, the largest value is x_0, since if x_w is larger for some w with $d(w, 0) = 1$ then the constraint may be improved by swapping x_0 and x_w. It follows that $|x_w| \leq \frac{1}{4}$ for $w \neq 0$ since otherwise the claimed bound would be exceeded. For a fixed x_0, the conditioned optimization problem is now convex with a unique local minimum, which by symmetry occurs with all variables equal. This reduces to minimizing $1 - c(x) + 6\left(1 - c\left(\frac{1}{2} - x\right)\right)$ for $0 \le x \le \frac{1}{2}$, which has minimum 2.

□

Next the possibility of a prevector with node of height at least 2 is ruled out.

Lemma 6.4.7. *If* $|\nu_0| = 2$*,* $P(\{0\}, |\nu|) \ge 1.4322$*. If* $|\nu_0| = 1$*,* $P(\{0\}, |\nu|) \ge 0.44256$*.*

Proof. These values were verified in SciPy. □

It follows that if the minimizing prevector has a node of height 2, it does not have any non-zero node at distance greater than 2 from the node of height 2, since otherwise the two optimization problems could be applied separately at the two nodes, and the total value would exceed γ_{tri}.

Up to rotation, there are two types of nodes at graph distance 2 from 0 in $\mathscr{T}$, $v_1 + v_2$ and $2v_1$. A non-zero node at distance two is ruled out by considering the following optimization problems.

Lemma 6.4.8. *Suppose* $|\nu_0| = 2$ *and* $|\nu_{v_1+v_2}| = 1$*. Then* $P(\{0, v_1 + v_2\}, |\nu|) \ge 1.83$*. If* $|\nu_0| = 2$ *and* $|\nu_{2v_1}| = 1$ *then* $P(\{0, 2v_1\}, |\nu|) \ge 1.85$*.*

Proof. These values were verified in SciPy. □

Note that $P(S, |\nu|)$ is increasing in $|\nu|$. The above lemmas prove that if the optimizing prevector ν has a node of height 2, then any non-zero node in ν is adjacent to the node of height 2. After translation and multiplying by ± 1, assume $\nu_0 = 2$. The case in which all six neighbors of 0 are non-zero is ruled out as follows.

Lemma 6.4.9. *Let* $|\nu_0| = 2$ *and* $|\nu_w| \ge 1$ *for each* w *with* $d(w, 0) = 1$*. Let* $S = \{w : d(w, 0) \le 1\}$*. Then* $P(S, |\nu|) \ge 1.9233$*.*

Proof. This was verified in SciPy. □

Similarly, there are not two adjacent nodes of height 2, as the following lemma verifies.

Lemma 6.4.10. *Suppose that* $|\nu_0|= 2$ *and* $|\nu_{v_1}|= 2$. *Then* $P(\{0, v_1\}, |\nu|) \geq 2.3$.

Proof. This was verified in SciPy. □

Since it is necessary that $\nu \in C^2(\mathscr{T})$ for $\xi \in \ell^2(\mathscr{T})$, the remaining possible configurations have an even number of non-zero nodes adjacent to 0. There must be at least 2, and when there are two, the configuration is, up to rotation, $\nu = -\delta_{-v_1} + 2\delta_0 - \delta_{v_1}$ which has $f(\xi) \geq 2.23$. No configuration with four non-zero nodes is in $C^2(\mathscr{T})$. This concludes the proof that there is not a node of height 2.

Next decompose the support of ν into 2-path connected components. The next stage in the argument reduces to the case of a single connected component. If there were four or more connected components, Lemma 6.4.7 could be applied at a node in each connected component, which obtains a value at least $4 \times 0.44256 > 1.76$. Hence there are at most 3 connected components, and since $\nu \in C^2(\mathscr{T})$, one must contain more than one node.

Lemma 6.4.11. *If* $|\nu_0|= 1$ *and* $|\nu_{v_1}|= 1$ *then*

$$P(\{0, v_1\}, |\nu|) \geq 0.6729. \tag{6.21}$$

If $|\nu_0|= 1$ *and* $|\nu_{v_1+v_2}|= 1$ *then*

$$P(\{0, v_1 + v_2\}, |\nu|) \geq 0.8509. \tag{6.22}$$

If $|\nu_0|= 1$ *and* $|\nu_{2v_1}|= 1$ *then*

$$P(\{0, 2v_1\}, |\nu|) \geq 0.8677. \tag{6.23}$$

Proof. These were verified in SciPy. □

It follows that if there are 3 connected components then the only possibility is that one has diameter 1 as in (6.21) and the other two are singletons, since otherwise the sum of the values of the programs exceeds γ. To remain in

$C^2(\mathscr{T})$, the configuration of diameter 1 has two nodes since the total number of nodes is even.

Lemma 6.4.12. *Let* $\nu_0 = 1$ *and* $\nu_w = 0$ *for* w *such that* $d(w, 0) = 1$. *Let* $S = \{w : d(w, 0) \leq 1\}$. *Then* $Q(\nu, S) \geq 0.9127$.

Proof. This was verified in SciPy. □

If there were an optimal configuration with 3 connected components, then the component with two adjacent nodes must have both nodes of equal sign for the configuration to be in $C^2(\mathscr{T})$. Thus the two singletons would be placed symmetrically opposite the center of the configuration of size 2 and have the same sign. Since they are disconnected, they have distance at least 3 from the component of size 2. It follows that Lemma 6.4.12 can be applied at each singleton so that the value exceeds γ_{tri}. This eliminates the case of 3 connected components.

Next suppose that there are two connected components. By applying (6.22) and (6.23) it follows that at least one of the connected components has diameter at most 1.

Lemma 6.4.13. *Suppose* $\nu_0 = \nu_{v_1} = 1$. *Then* $Q(\{0, v_1\}, \nu) \geq 1.1518$.

Proof. This was verified in SciPy. □

If one connected component has such a large Q value, then by Lemma 6.4.11, the other component can only be a singleton. The case of two connected components with one a singleton is deferred to the end of the proof. Thus consider the case of only connected components of size at least 2 in which adjacent nodes have opposite signs. It follows that one of the components of diameter 1 has size 2, with adjacent nodes of opposite sign.

Lemma 6.4.14. *Let* $\nu_0 = 1$, $\nu_{v_1} = -1$. *Let* $S = \{w : d(w, \{0, v_1\}) \leq 1\}$ *and assume* $\nu_w = 0$ *if* $d(w, \{0, v_1\}) = 1$. *Then* $Q(S, \nu) \geq 0.971$.

Proof. This was verified in SciPy. □

Combining Lemma 6.4.14 with Lemma 6.4.11 if one of the connected components has diameter greater than 1, then the component of diameter 1 has distance from it at most 3, hence exactly 3 since the components are not connected. As in the case of a singleton, this case is deferred to the end of the discussion. If both components have diameter 1, then to be in $C^2(\mathscr{T})$, both have size two and have adjacent nodes of opposite sign. Applying Lemma 6.4.14 to each, these are separated by at most distance 4. This reduces to a finite check, and none of the configurations achieves the optimum.

The argument above reduces to considering either prevectors with support that are 2 path connected, or prevectors of diameter greater than 1 which are connected at distance 2, together with a second connected component which is either a singleton or a pair of adjacent nodes of opposing signs. By combining Lemmas 6.4.12 and 6.4.14 with Lemma 6.4.11, it follows that if there is a second connected component it has distance exactly 3 from the component of diameter greater than 1.

The proof is now concluded by computer search. All connected components C up to translation and symmetry were enumerated, which satisfied one of the following three criteria, $P(C, 1) \leq \gamma_{\text{tri}} = 1.69416(5)$, $P(C, 1) \leq \gamma_{\text{tri}} - 0.44256$, $P(C, 1) \leq \gamma_{\text{tri}} - 0.6729$, with $\nu = 1$ indicating $\nu_x = 1$ for all x. The first list consists of all candidate supports which are connected and may give the optimum. By Lemmas 6.4.7 and 6.4.11, the latter two lists enumerate configurations which may be paired with a singleton or a pair of adjacent nodes. Since $P(C, 1)$ is increasing in C, the enumeration was performed by building configurations from the base $C = \{0\}$ adding neighbors at distance 1 or 2, until the appropriate limit was exceeded. The first list contains configurations with at most 7 vertices, the second list contains configurations with at most 5 vertices and the third list contains configurations with at most 4 vertices.

Note that a configuration which can appear with adjacent and opposite signed nodes and have a $C^2(\mathscr{T})$ assignment of signs must have an even number of nodes. Also, those of size 2 have already been considered. Only one configuration on 4 nodes, and no configurations on more nodes had a sufficiently small value of $P(C)$. The configuration on 4 nodes was, up to symmetries, $\{0, v_1, v_2, v_1 + v_2\}$. However, there is no assignment of signs which makes this configuration in $C^2(\mathscr{T})$ when paired with an adjacent pair of nodes with opposite signs. A connected component with 3 vertices cannot be assigned signs in such a way that a singleton can be added at distance 3 to make a configuration in $C^2(\mathscr{T})$, since the distance between the one pair of opposite signed nodes must match the other. There is a single configuration on 5 nodes with P value less than $\gamma_{\mathrm{tri}} - 0.44256$. There are 4 ways of assigning signs so that a singleton can be added that makes the configuration in $C^2(\mathscr{T})$. Each of these was tested and none give the extremal configuration. This reduces to the case of connected components. This finite check was performed in SciPy and obtains ν_0 and ξ_0 as claimed.

Bibliography

[1] Bak, P., C. Tang, and K. Wiesenfeld. “Self-organized criticality." *Physical Review A* 38.1 (1988): 364.

[2] Diaconis, Persi. “Group representations in probability and statistics.” *Lecture Notes-Monograph Series* 11 (1988): i-192.

[3] Serre, Jean-Pierre. *Linear Representations of Finite Groups.* Vol. 42. Graduate Texts in Mathematics, 1977.

[4] Cohn, Paul M. *Further Algebra and Applications.* Springer India, 2003.

[5] Cohn, Paul M. *Introduction to ring theory.* Springer 2004.

[6] Berndt, Rolf. *Representations of linear groups: an introduction based on examples from physics and number theory.* Vieweg 2007.

[7] Steinberg, Benjamin. *Representation theory of finite groups: an introductory approach.* Springer 2016.

[8] Grinstead, Charles M. and Snell, J. Laurie. *Grinstead and Snells Introduction to probability.* Editore 2000.

[9] Kijima, Masaaki. *Markov processes for stochastic modeling.* Chapman and Hall 1997.

[10] Ross, Sheldon M. *Introduction to probability models.* Academic Press 2014.

[11] Ross, Sheldon M. *A first course in probability.* Pearson 2019.

[12] Levin, David Asher and Peres, Yuval and Wilmer, Elizabeth Lee and Propp, James and Wilson, David B. *Markov chains and mixing times.* American Mathematical Society 2017.

[13] Durrett, Richard. *Probability: theory and examples.* Cambridge Univ. Press 2010.

[14] Holroyd, Alexander E. and Levine, Lionel and Mészáros, Karola and Peres, Yuyal and Propp, James and Wilson, David B. "Chip-Firing and Rotor-Routing on Directed Graphs." *In and Out of Equilibrium 2* (2008): 331.

[15] Corry, Scott and Perkinson, David. *Divisors and sandpiles: an introduction to chip-firing.* American Mathematical Society 2018.

[16] Hough, Robert and Son, Hyojeong. "Sandpile dynamics on periodic tiling graphs."