FOURIER ANALYSIS ON FINITE GROUPS AND APPLICATIONS TO RANDOM WALKS

SHIVESH MEHROTRA

Abstract. In this paper we build the required knowledge of Representation and Character Theory as well as Convolution needed to introduce Fourier Analysis on Finite Groups in generality. We culminate the paper by presenting an application of Fourier Analysis to Random Walks on Finite Groups. This paper assumes familiarity with linear algebra and integral calculus.

Contents

1. Introduction to Representation Theory

Broadly speaking, Representation Theory is the study of groups using matrices. One of the benefits of Representation Theory is that calculations and manipulations with matrices are much easier than with the original group elements. Furthermore, by mapping group elements to matrices in vector spaces V we can ask questions which shed more light on the group structure itself. Some examples of these questions include:

- Does our vector space V have subspaces which are invariant under the group action of G?
- What are these subspaces and what are they revealing about the structure of the group?
- What happens when there are no invariant subspaces of V under the action of G and can we classify these cases?

Representations of groups also help us understand and characterize functions which act on the group. In fact by using representations on groups we will be able to extend the familiar notion of Fourier Analysis to all finite groups. Thus, let us dive in.

Definition 1.1: Representation

A *representation* of a group G is a group homomorphism (preserves the group operation) $\phi : G \to GL(V)$ where V is a finite dimensional complex vector space (non-zero). The *degree* of ϕ is $\dim(V)$.

Remark 1.2. We often denote $\phi(g)$ for $g \in G$ as ϕ_g and use $\phi_g(v)$ or $\phi_g v$, for the application of ϕ_g on $v \in V$.

Example 1.3. For any group G we have the degree-one *trivial representation* $\phi : G \to \mathbb{C}^*$ via $\phi(g) = 1$ for all $g \in G$. Note that $\mathbb{C}^* = Gl(C)$ so equivalently we could say that $\phi : G \to Gl(C)$.

The following barrage of definitions will help us develop some basic notions in representation theory.

Definition 1.4: Homomorphism

Let $\phi : G \to GL(V)$ and $\psi : G \to GL(W)$ be representations. A *homomorphism* from ϕ to ψ is a linear map $T : V \to W$ such that $T\phi_g = \psi_g T$ for all $g \in G$. We say that T *intertwines* ϕ and ψ.

Remark 1.5. The set of homomorphisms from ϕ to ψ is denoted $\mathrm{Hom}_G(\phi, \psi)$.

Definition 1.6: Equivalence

We say that two representations $\phi : G \to GL(V)$ and $\psi : G \to GL(W)$ are *equivalent* if there is bijective homomorphism, equivalently an isomorphism, T which intertwines ϕ and ψ.

Remark 1.7. Visually we can see the relationship of homomorphism for a fixed $g \in G$:

$$\begin{array}{ccc} V & \xrightarrow{\phi_g} & V \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle T} \\ W & \xrightarrow[\psi_g]{} & W \end{array} .$$

Equivalence says that moving from the upper left to lower right of this diagram on either path results in the same mapping.

To motivate the following definitions: Often in mathematics it is very useful to have a notion of irreducible pieces. For example, the prime numbers are the building block of all other numbers. In linear algebra we have the upper triangular matrices from which we can extend results to all matrices. Similarly, in representation theory we have irreducible representations which can allude to fundamental properties of the groups they represent and which are the building blocks of all other representations.

Definition 1.8: Invariant subspace

Let $\phi : G \to GL(V)$ be a representation. We say that a subspace $W \leq V$ is *G-invariant* if for all $g \in G$ and $w \in W$ we have that $\phi_g w \in W$. That is $\phi_g(W) \subseteq W$.

Definition 1.9: Direct Sum of Representations

Let $\phi : G \to GL(V)$ and $\psi : G \to Gl(W)$ be representations of the group G. Their *direct sum* $\phi \oplus \psi : G \to GL(V \oplus W)$ is given by $(\phi \oplus \psi)_g(v, w) = (\phi_g v, \psi_g w)$.

Remark 1.10. We can also represent our direct sum in block matrix which yields:

$$(\phi \oplus \psi)_g = \begin{pmatrix} \phi_g & 0 \\ 0 & \psi_g \end{pmatrix}.$$

Definition 1.11: Reducibility

Let $\phi : G \to GL(V)$ be a non-zero representation of G. We say that ϕ is *irreducible* or *simple* if the only $G-invariant$ subspaces are V and $\{0\}$. We say ϕ is *completely reducible* if $V = V_1 \oplus \ldots \oplus V_n$ where each V_i is G-invariant and $\phi|_{V_i}$is irreducible. Equivalently we can say that ϕ is completely reducible if $\phi \sim \phi_1 \oplus \ldots \oplus \phi_n$ where each ϕ_i is irreducible. Last we say ϕ is *decomposable* if $V = V_1 \oplus V_2$ where V_1 and V_2 are non-zero G-invariant subspaces.

Remark 1.12. Note the distinction between *completely reducible* and *decomposable*. Decomposable is a weaker condition in the sense ϕ may not be irreducible on V_1 and V_2. That is there still may be non-zero *G-invariant* proper subspaces $W_1 \leq V_1$ and $W_2 \leq V_2$. Whereas if ϕ is *completely reducible* there are no smaller subspaces of the V_i's on which ϕ is $G-invariant$. To give an analogy we can think of the V_i's as atoms whereas in the case of decomposability V_1 and V_2 may be molecules.

Proposition 1.13. *If $\phi : G \to GL(V)$ is equivalent to an irreducible representation, then ϕ is irreducible.*

Proof. Let $\phi : G \to GL(V) \sim \psi : G \to GL(W)$ be representations of G where ψ is irreducible and T is the map which intertwines them. By way of contradiction suppose that ϕ is reducible. Then, $V = V_1 \oplus \ldots \oplus V_n$ where each V_i is G-invariant and $\phi|_{V_i}$ is irreducible. Since T is an isomorphism, this implies that $T(V) = T(V_1) \oplus \ldots \oplus T(V_n) = W$ where each $T(V_n)$ is G-invariant under ψ. But this is a contradiction since we know that ψ is irreducible. Thus ϕ must also be irreducible. □

Definition 1.14: Unitary

Let $\phi : G \to GL(V)$ be a representation where V is an inner product space with inner product denoted $\langle \cdot, \cdot \rangle$. We say ϕ is *unitary* if ϕ_g is a unitary matrix for all $g \in G$. That is for all $v, w \in V$ and $g \in G$ we have $\langle \phi_g(v), \phi_g(w) \rangle = \langle v, w \rangle$.

Proposition 1.15. *If $\phi : G \to GL(V)$ is a unitary representation then either ϕ is irreducible or decomposable.*

Proof. Suppose that ϕ isn't irreducible. Thus there exists $W \subseteq V$ such that W is $G-$*invariant*. And recall from linear algebra that we can write $V = W \oplus W^{\perp}$ where $W^{\perp}$ denotes the subspace of V such that if $w \in W$ and $w' \in W^{\perp}$ then $\langle w, w' \rangle = 0$. If we show that $W^{\perp}$ is $G-$*invariant* this will imply that ϕ is decomposable. Let $w \in W$ and $w' \in W^{\perp}$ be arbitrary and fix $g \in G$. We want to show that $\phi(w') \in W^{\perp}$. We see $\langle \phi_g w', w \rangle = \langle \phi_g^{-1} \phi_g w', \phi_g^{-1} w \rangle = \langle w', \phi_g^{-1} w \rangle = 0$ since we know that $\phi_g^{-1} w \in W$ because W is $G-$*invariant*. Thus, $\phi(w') \in W^{\perp}$ and so $W^{\perp}$ is $G-$*invariant* as desired. □

Proposition 1.16. *If $\phi : G \to GL(V)$ is a representation of a finite group, then ϕ is equivalent to a unitary representation.*

Proof. Denote $\dim(V)$ as n. Then we know that V is isomorphic to $\mathbb{C}^n$ via $T : V \to \mathbb{C}^n$. Thus we can define a representation $\psi : G \to GL_n(\mathbb{C})$ which is equivalent to ϕ via $\psi_g := T\phi_g T^{-1}$. If $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{C}^n$, then we can define a new inner product $(v_1, v_2) = \sum_{g \in G} \langle \psi_g(v_1), \psi_g(v_2) \rangle$. We can verify that $(\cdot, \cdot)$ is indeed an inner product and that ψ is a unitary representation with respect to this inner product. □

Remark 1.17. Notice that G must be a finite group in order for $\sum_{g \in G} \langle \psi_g(v_1), \psi_g(v_2) \rangle$ to be finite and define a true inner product.

Remark 1.18. Going forward when we refer to a representation on a finite group we will assume it to be unitary. Furthermore all results that will be presented are true up to equivalence of representations. That is, anything proven for a unitary representations will apply to all representations by Proposition 1.16.

Corollary 1.19. *If $\phi : G \to GL(V)$ is a non-zero representation of a finite group, then ϕ is irreducible or decomposable.*

The proof follows from Proposition 1.16 and Proposition 1.15.

Theorem 1.20: Maschke's Theorem

If $\phi : G \to Gl(V)$ is a representation of a finite group, then it is completely reducible.

Proof. We will proceed by strong induction on $\dim(V)$ and use Corollary 1.19. For the base case suppose that $\dim(V) = 1$. Then ϕ is irreducible as the only subspaces of V are itself and $\{0\}$. Now suppose that the claim holds for $\dim(V) \leq n$. Let $\dim(V) = n + 1$. If ϕ is irreducible then we are finished. Otherwise from Corollary 1.19 we know that ϕ is decomposable, that is $V = V_1 \oplus V_2$ where $\dim(V_1), \dim(V_2) \leq n$. Thus applying the inductive hypothesis to $\phi|_{V_1}$ and $\phi|_{V_2}$ we see that ϕ is completely reducible. □

Proposition 1.21. *Let $T : V \to W$ be in $Hom_G(\phi, \psi)$. Then $\ker(T)$ is a G-invariant subspace of V and $Im(T)$ is a G-invariant subspace of W.*

Proof. Let $v \in \ker(T)$. Then for all $g \in G$ we have $T\phi_g(v) = \psi_g T(v) = 0$. Thus $\phi_g(v) \in \ker(T)$.

Now let $w \in \mathrm{Im}(T)$ and suppose that $Tv = w$. Then for all $g \in G$ we have that $T\phi_g(v) = \psi_g(w)$ and so $\psi_g(w) \in \mathrm{Im}(T)$. □

Proposition 1.22. *Let $\phi : G \to GL(V)$ and $\psi : G \to GL(W)$ be representations. Then $Hom_G(\phi, \psi)$ is a subspace of $Hom(V, W)$.*

Proof. Let $T, R \in Hom_G(\phi, \psi)$ and $c, d \in \mathbb{C}$. Then we see that for all $g \in G$ we have:

$$(cT + dR)\phi_g = cT\phi_g + dR\phi_g = c\psi_g T + d\psi_g T = \psi_g(cT + dR).$$

□

Theorem 1.23: Schur's Lemma

Let ϕ, ψ be irreducible representations of G and let $T \in Hom_G(\phi, \psi)$. Then either $T = 0$ or T is invertible, that is T is an isomorphism (see Definitions 1.4 and 1.6). Also:

(1) If $\phi \not\sim \psi$ (i.e $T = 0$), then $Hom_G(\phi, \psi) = 0$.
(2) If $\phi = \psi$, then $T = \lambda I$ where $\lambda \in \mathbb{C}$.

Proof. We will start with the first claim. Let $\phi : G \to GL(V)$, and $\psi : G \to GL(W)$ be irreducible representations. Further suppose that $T \neq 0$. From Proposition 1.21 we know that $\ker(T)$ is a G-invariant subspace of V and $\mathrm{Im}(T)$ is a G-invariant subspace of W. Yet since ϕ is irreducible we know that V and $\{0\}$ are the only G-invariant subspaces of V. Thus since $T \neq 0$ we know that $\ker(T) \neq V$ and so $\ker(T) = \{0\}$ meaning that T is injective. Similarly since ψ is irreducible we know that the only G-invarient subspaces of W are W and $\{0\}$. Thus since $T \neq 0$ we see that $\mathrm{Im}(T) = W$ and T is a surjective map. Therefore if $T \neq 0$ then T is bijective and thus invertible. Claim (1) follows directly from Definition 1.6. For claim (2) if $\phi = \psi$, then $T \in Hom_G(\phi, \phi)$. The key observation is to see that $I \in Hom_G(\phi, \phi)$ where I is the identity matrix (see Defintion 1.4). So from Proposition 1.22 we see that $\lambda I - T \in Hom_G(\phi, \phi)$. But if we choose λ to be an eigenvalue of T then by definition $\lambda I - T$ is not invertible. Thus from above we must have $\lambda I - T = 0 \implies T = \lambda T$ as desired. □

Remark 1.24. Note in this proof we use the fact that V is a complex-vector space. Thus since $\mathbb{C}$ is algebraically closed we know that indeed such an eigenvalue λ exists.

We will now present corollaries to Theorem 1.23 which give neat results in linear algebra. Full proofs can be found Chapter 4 of [1].

Corollary 1.25. *Let G be a abelian group. Then any irreducible representation of G has degree-one.*

Corollary 1.26. *Let G be a finite abelian group and $\phi : G \to GL_N(\mathbb{C})$ be a representation. Then there exists an invertible matrix T such that $T\phi_g T^{-1}$ is diagonal for all $g \in G$.*

Corollary 1.27. *Let $A \in GL_m(\mathbb{C})$ be a matrix of finite order (i.e $A^n = 1$). Then A is diagonalizable and the eigenvalues of A are nth roots of unity.*

Definition 1.28: Group Algebra

Let G be a group. We denote the *group algebra of* G, $L(G)$, where $L(G) = \mathbb{C}^G = \{f | f : G \to \mathbb{C}\}$. Observe that $L(G)$ is an inner product space with addition and scalar multiplication. The inner product is defined by $\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{g \in G} \overline{f_1(g)} f_2(g)$.

For the following Theorem let $U_n(\mathbb{C})$ denote the unitary $n \times n$ matrices over $\mathbb{C}$. That is $UU^* = I$ where U^* is the conjugate transpose of U.

Theorem 1.29: Schur's Orthogonality Relations

Suppose that $\phi : G \to U_n(\mathbb{C})$ and $\psi : G \to U_m(\mathbb{C})$ are non-equivalent irreducible unitary representations. Let $\phi_{ij} : G \to C$ denote the map which sends $g \in G$ to the element in the ith row and jth column of ϕ_g and define ψ_{kl} similarly. Then,

(1) ϕ_{ij} and ψ_{kl} are orthogonal on $L(G)$. That is,

$$\langle \phi_{ij}, \psi_{kl} \rangle = \sum_{g \in G} \phi_{ij}(g)\overline{\psi_{kl}(g)} = 0.$$

(2) $\langle \phi_{ij}, \phi_{kl} \rangle \neq 0$ iff $i = k, j = l$.

Proof. (1) Let T be a linear map $\mathbb{C}^m \to \mathbb{C}^n$ and define $T' := \sum_{g \in G} \phi(g) T \psi(g^{-1})$. Let $h \in G$. Then,

$$\begin{aligned} \phi(h)T' &= \phi(h) \sum_{g \in G} \phi(g) T \psi(g^{-1}) \\ &= \sum_{g \in G} \phi(h)\phi(g) T \psi(g^{-1}) \\ &= \sum_{g \in G} \phi(hg) T \psi(g^{-1}) \end{aligned}$$

since we know that ϕ is a homomorphism. Now notice that we can re-index the summation by hg instead of g as we will still sum over all of G. Thus,

$$\begin{aligned} \sum_{g \in G} \phi(hg) T \psi(g^{-1}) &= \sum_{hg \in G} \phi(hg) T \psi((hg)^{-1}h) \\ &= \Big[\sum_{hg \in G} \phi(hg) T \psi((hg)^{-1}) \Big] \psi(h) \\ &= T'\psi(h). \end{aligned}$$

So, $\phi(h)T' = T'\psi(h)$ meaning that $T' \in Hom_G(\phi, \psi)$. Thus from Schur's Lemma 1.23 since $\phi \not\sim \psi$ we know that $T' = 0$. Recall that our only assumption is that T was linear. Thus choose T to have a 1 in the jth row and lth column and zeros elsewhere. Then $T'_{n,i} = 0$. Furthermore this entry is equal to $\sum_{g \in G} \phi_{ij}(g)\overline{\psi_{kl}(g)} = \langle \phi_{ij}, \psi_{kl} \rangle = 0$.

(2) Define T' identically as in (1) except using only ϕ. As above we see that $T' \in Hom_G(\phi, \phi)$. Thus from Schur's Lemma 1.23 we have that $T' = \lambda I$ where $\lambda \in \mathbb{C}$. Taking the trace (sum of the diagonal entries) of T' we have that $|G|\text{tr}(T) = \text{tr}(\lambda I)$ since $\text{tr}(AB) = \text{tr}(BA)$ and ϕ is a homomorphism. Thus, $|G|\text{tr}(T) = \lambda d$ where d is the degree of ϕ (see Definition 1.1). Identically to above let $T = E_{j,l}$ where every entry is 0 except the j, lth entry is 1. This choice yields that

$$T'_{n,i} = \sum_{g \in G} \phi_{ij}(g)\overline{\phi_{kl}(g)} = \langle \phi_{ij}, \phi_{kl} \rangle.$$

Now observe that $\lambda \neq 0$ iff $\text{tr}(T) = \text{tr}(E_{j,l}) \neq 0$ which occurs iff $j = l$. Furthermore if $\lambda \neq 0$, then $T'_{i,k} \neq 0$ iff $i = k$ since $T' = \lambda I$. Therefore $T'_{n,i} = \langle \phi_{ij}, \phi_{kl} \rangle \neq 0$ iff $i = k$ and $j = l$.

□

Remark 1.30. In fact we can improve on (2) and show that $\langle \phi_{ij}, \phi_{kl} \rangle = \begin{cases} \frac{1}{n} & i = k, j = l \\ 0 & else \end{cases}$.
We will not prove this as the proof isn't very illuminating, however the full proof can be found in [1].

Now we will define a character of a representation and introduce class functions.

2. Characters and Class Functions

Definition 2.1: Character

Let $\phi : G \to GL(V)$ be a representation. We define the *character* $\chi_\phi : G \to \mathbb{C}$ of ϕ to be $\chi_\phi(g) = \text{tr}(\phi_g)$. The character of an irreducible representation is called an *irreducible character* (see Definition 1.11).

Now we will present some key properties of characters.

Proposition 2.2. *Let $\phi : G \to GL(V)$ be a representation and χ its character. Then $\chi_\phi(e) = \deg(\phi)$ where $e \in G$ is the identity.*

Proof. We know that $\phi(e) = I$. So

$$\begin{aligned} \chi_\phi(e) &= \text{tr}(I) \\ &= \dim(V) \\ &= \deg(\phi). \end{aligned}$$

□

Proposition 2.3. *If ϕ and ψ are equivalent representations (see Definition 1.6), then $\chi_\phi = \chi_\psi$.*

Proof. Since we must choose a basis to calculate the trace of a matrix we can assume that $\phi, \psi : G \to GL_n(\mathbb{C})$. Furthermore since ϕ and ψ are equivalent we know there exists an invertible matrix $T \in GL_n(\mathbb{C})$ such that $T\phi_g T^{-1} = \psi_g$ for all $g \in G$. Then,

$$\begin{aligned} \chi_\psi(g) &= \text{tr}(\psi_g) \\ &= \text{tr}(T\phi_g T^{-1}) \\ &= \text{tr}(T^{-1}T\phi_g) \\ &= \text{tr}(\phi_g) \\ &= \chi_\phi(g) \end{aligned}$$

since $\text{tr}(AB) = \text{tr}(BA)$. □

Proposition 2.4. *Let $\phi : G \to GL(V)$ be a representation. Then for all $g, h \in G$, $\chi_\phi(g) = \chi_\phi(hgh^{-1})$.*

The proof uses the same idea as Proposition 2.3.

Remark 2.5. The above propositions show that characters are constant on conjugacy classes as well as equivalence classes of representations. In fact, being constant on conjugacy classes is very important in representation theory which leads to the following definition.

Definition 2.6: Class Function

A function $f : G \to \mathbb{C}$ is called a *class function* if $f(g) = f(hgh^{-1}), \forall g, h \in G$. The space of class functions is denoted $Z(L(G))$. Furthermore if C is a conjugacy class, we denote $f(C)$ to be the constant value of f over C.

Remark 2.7. One can show that $Z(L(G))$ is a subspace of $L(G)$, the group algebra of G (see Definition 1.28).

As we saw above characters are class functions. In fact irreducible characters form an orthonormal set of class functions as we see in the following Theorem.

Theorem 2.8: First orthogonality relations

Let ϕ, ψ be irreducible representations of G. Then

$$\langle \chi_\phi, \chi_\psi \rangle = \begin{cases} 1 & \phi \sim \psi \\ 0 & \phi \not\sim \psi \end{cases}.$$

That is irreducible characters comprise an orthonormal set of class functions.

Proof. Without loss of generality we can assume $\phi : G \to U_n(\mathbb{C})$ and $\psi : G \to U_m(\mathbb{C})$ are unitary (see Propositions 1.16 & 2.3). Computing we have

$$\begin{aligned} \langle \chi_\phi, \chi_\psi \rangle &= \frac{1}{|G|} \sum_{g \in G} \overline{\chi_\phi(g)} \chi_\psi(g) \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{i=1}^{n} \overline{\phi(g)_{i,i}} \sum_{j=1}^{m} \psi(g)_{j,j} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{|G|} \sum_{g \in G} \overline{\phi(g)_{i,i}} \psi(g)_{j,j} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \langle \phi(g)_{i,i}, \psi(g)_{j,j} \rangle. \end{aligned}$$

Thus from Schur's Orthogonality Relations 1.29 we know that $\langle \phi(g)_{i,i}, \psi(g)_{j,j} \rangle = 0$ if $\phi \not\sim \psi$ and so $\langle \chi_\phi, \chi_\psi \rangle = 0$. If $\phi \sim \psi$ then we may assume $\phi = \psi$ for the above calculation due to Proposition 2.3. Then from Remark 1.30 we know $\langle \phi_{i,i}, \phi_{j,j} \rangle = \begin{cases} \frac{1}{n} & i = j \\ 0 & i \neq j \end{cases}$. Therefore, $\langle \chi_\phi, \chi_\phi \rangle = \sum_{k=1}^{n} \langle \phi_{i,i}, \phi_{i,i} \rangle = \sum_{k=1}^{n} \frac{1}{n} = 1$ if $\phi \sim \psi$.

□

Remark 2.9. Thus we see that irreducible characters form an orthonormal set of class functions. Namely each irreducible character of G is in $Z(L(G))$ since each is a class function $\chi_\phi : G \to \mathbb{C}$, see Definition 1.28. In fact we can use this orthogonal set to form a basis of $Z(L(G))$, considering $Z(L(G))$ to be a complex inner product space. Utilizing this relationship will allow us to define Fourier series on finite groups. We will not show the proof of this fact in this paper as its set up is a bit long-winded, however for the curious see page 40 of [1].

Now that we have established the basics of representation theory and class functions we will define and motivate Convolution, an operation central to Fourier Analysis.

3. Convolution

We will now introduce convolution on $L(G)$. Convolution is the heart of Fourier analysis and is a widely used mathematical operation and has applications in computer science, differential equations, probability, statistics and many more fields. After we introduce some basic definitions we will briefly explore some of these applications. See Example 3.23 for one of the many motivations behind convolution.

Definition 3.1: Convolution on Finite Groups

Let G be a finite group and let $f, g \in L(G)$. Then the *convolution* $f * g(x) : G \to \mathbb{C}$ is defined to be

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

Remark 3.2. Later on we will mainly use Definition 3.1 for probability functions over finite groups.

We can also define the analogue for function spaces like $L^2(\mathbb{R})$.

Definition 3.3: Convolution on Function Spaces

Let $f, g \in L^2(\mathbb{R})$. Then the *convolution* is defined as

$$f * g(x) = \int_{\mathbb{R}} f(t)g(x-t)dt = \int_{\mathbb{R}} f(x-t)g(t)dt.$$

Definition 3.4: Discrete Convolution

For complex-valued functions f, g on $\mathbb{Z}$ we define the *discrete convolution* of f and g as

$$f * g(n) = \sum_{-\infty}^{M} f(n-m)g(m).$$

If g has finite support $\{-M, -M+1, \dots M\}$ we define the *finite discrete convolution* of f and g as

$$f * g(n) = \sum_{-M}^{M} f(n-m)g(m).$$

Remark 3.5. The convolution of sequences is defined by representing the sequence as functions over $\mathbb{Z}$. If the sequences represent the coefficients of polynomials as in Observation 3.7 then this is also known as the *Cauchy Product*.

Let's observe how convolution appears when asking the basic question of how to multiply together two polynomials.

Problem 3.6. *Given two polynomials* $A(x) = a_0 + a_1x + \ldots + a_nx^n$ *and* $B(x) = b_0 + b_1x + \ldots + b_nx^n$ *find* $C(x) = A(x) \cdot B(x)$.

Observation 3.7. Given two polynomials $A(x) = \sum_{i=0}^{m} a_i x^i$ and $B(x) = \sum_{j=0}^{n} b_j x^j$ we see that their product $C(x)$ is

$$C(x) = \sum_{i=0}^{m} a_i x^i \sum_{j=0}^{m} b_j x^j.$$

If we wish to rewrite this so grouping all the terms of the same degree together setting $k = i + j$ we have

$$C(x) = \sum_{i=0}^{m} \sum_{j=0}^{n} a_{k-j} b_j x^k.$$

Notice that we have rewritten the sum so that j depends on k and that k ranges from 0 to $m + n$. We can further re-index the summation in terms of strictly k to obtain:

$$C(x) = \sum_{k=0}^{m+n} \big[\sum_{j=0}^{n} a_{k-j} b_j\big] x^k. \tag{3.8}$$

Observe that $\big[\sum_{j=0}^{n} a_{k-j} b_j\big]$ is the discrete convolution of the coefficients of our polynomials as in Definition 3.4.

Thus if we represent the coefficients of our polynomials $A(x)$ and $B(x)$ as sequences we can use discrete convolution to find their product.

Now let us set up an example of convolution in probability.

3.1. **Convolution of Two Random Variables.** First let us introduce some basic definitions from probability theory.

Definition 3.9: Probability Space

A *probability space* or *probability triple* is a measure space with total measure 1. Recall that a measure space consists of three elements $(\Omega, \mathcal{F}, P)$:

(1) A sample space Ω which is the set of all possible outcomes.
(2) An event space $\mathcal{F}$ which is a set of events, where an event is a set of outcomes from Ω. Formally $\mathcal{F}$ is a sigma algebra such that $\mathcal{F} \subseteq 2^{\Omega}$.
(3) A probability measure, P which assigns an event $A \subseteq \mathcal{F}$ a probability between 0 and 1.

Remark 3.10. Recall that the measure P satisfies:

- $P(\emptyset) = 0$ and $P(\Omega) = 1$.

- $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ if the A_n's are disjoint. This property is known as *countable additivity* and we say that P is *countably additive*.

Remark 3.11. Differentiating the *event space* and *sample space* is subtle and is often a point of confusion. The *sample space* is the set of all possible outcomes, whereas the *event space* contains all sets of outcomes, i.e all possible subsets of the *sample space*.

Definition 3.12: Random Variable

A *random variable* is a measurable function $X : \Omega \to E$ (often $E = \mathbb{R}$). X maps possible outcomes in a sample space Ω to some measurable space E. If $\mathrm{Im}(X)$ is countable we say that X is a *discrete random variable*. If $\mathrm{Im}(X)$ is uncountably infinite (think of an interval) we say that X is a *continuous random variable*.

Here is an example which highlights the difference between a probability function P of a sample space and a random variable X on a the sample space.

Example 3.13. Consider the set of outcomes of a coin toss. So our sample space is simply $\{\text{heads}, \text{ tails}\}$. We know, $P(\text{heads}) = \frac{1}{2}$ and also $P(\text{tails}) = \frac{1}{2}$. Let $X : \Omega \to E$ be a random variable such that if the outcome is heads then X is 1 and otherwise X is zero. Now the difference between the probability function and X is clear. While the probability of heads is $\frac{1}{2}$, the value heads is assigned by X is 1.

Remark 3.14. A good way to think about a *random variable* is simply as a function which assigns numerical values to possible outcomes.

Now we introduce distribution functions and density functions.

Definition 3.15: Cumulative Distribution Function of a Random Variable

The *cumulative distribution function (CDF)* (often just distribution function) of a real-valued random variable $X : \Omega \to \mathbb{R}$ is denoted F_X and is defined by $F_X(x) = P(X \leq x)$ for $x \in \mathbb{R}$.

Remark 3.16. Notice that since X is real-valued, X is a continuous random variable, not discrete.

Example 3.17. Say we are interested in the weather and we want to know the chance that there is less than an inch of rainfall. Let X be a continuous random variable which outputs the amount of rainfall in inches on a given day. Notice that X is simply a variable which is assigning numerical values to outcomes, in this case inches to rainfall. Thus we want to know $P(X \leq 2)$, that is the probability that there is less than two inches of rainfall. We can use F_X, the distribution function of X to find $P(X \leq 2)$. Namely $F_X(2) = P(X \leq 2)$.

Definition 3.18: Probability Density Function

A *probability density function (PDF)* f_X of a random variable X is a non-negative Lebesgue integrable function such that $P(a \leq X \leq b) = \int_a^b f_X(x)dx$. We say X has density f_X.

Remark 3.19. Notice that if F_X is the cumulative distribution function of X then we have $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ where f_X is the density of X.

We now have the necessary background to set up the convolution of two continuous random variables.

Definition 3.20: Convolution of Continuous Random Variables

Let X, Y be continuous random variables with probability density functions $f_X(x)$ and $f_Y(y)$ defined over $\mathbb{R}$. Then the convolution $f_X * f_Y$ is given by

$$f_X * f_Y = \int_{\mathbb{R}} f_X(t-y)f_Y(y)dy = \int_{\mathbb{R}} f_Y(t-x)f_X(x)dx.$$

Remark 3.21. Notice that Definition 3.20 is a specific version of Definition 3.3. However we can now present a specific motivation behind convolution.

Definition 3.22: Independence of Continuous Random Variables

Two random variables X, Y with distribution functions F_X, F_Y are *independent* if and only if the combined random variable (X, Y) has a joint distribution function $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all x, y. Equivalently if X, Y have probability density functions f_x, f_y and the joint PDF $f_{x,y}$ exists then $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all x, y.

Example 3.23. Let $(\Omega, \mathcal{F}, P)$ be a probability triple and let X, Y be independent continuous real-valued random variables. Further suppose that $f_X(x)$ and $f_Y(x)$ are the corresponding density functions of X and Y. Now consider the random variable $Z = X + Y$. Notice that Z is perfectly well defined as X and Y are well defined. A natural question we might ask is what is the density function $f_Z(z)$ of Z. If X and Y are independent then,

$$f_Z(z) = f_X * f_Y(z) = \int_{\mathbb{R}} f_X(z-t)f_Y(t)dt = \int_{\mathbb{R}} f_X(t)f_Y(z-t)dt.$$

This example leads us to the next Theorem.

Theorem 3.24: Sum of Independent Random Variables

Let X and Y be independent continuous random variables. The probability density function of $Z = X + Y$ is the convolution of the probability density functions of X and Y. That is,

$$f_Z(z) = f_X * f_Y(z) = \int_{\mathbb{R}} f_X(z-t)f_Y(t)dt = \int_{\mathbb{R}} f_X(t)f_Y(z-t)dt.$$

Proof. We omit this proof as it is tangential to the rest of the material presented in this paper, but please see [2] for a full proof and further exploration of this Theorem. □

While convolution may have seemed unmotivated before we can now see one clear motivation from probability theory. In the following section will we use the machinery of Representation and Character Theory as well as Convolution we have developed to define Fourier Analysis on Finite Groups.

Remark 3.25. Going forward we will mainly use the notion of Convolution given in Definition 3.1. We will denote this $*$, and if we use $*$ without specifying assume we are referring to Definition 3.1.

4. Fourier Analysis on Finite Groups

4.1. **Fourier Analysis on Cyclic Groups.** We will begin by introducing a simpler case of Fourier Analysis, Fourier Analysis on Cyclic groups.

Definition 4.1: Periodic Function

A function $f : \mathbb{Z} \to C$ is *periodic* with period n if $f(x) = f(x+n), \forall x \in \mathbb{Z}$.

Remark 4.2. We can see that periodic functions with period n are in bijection with elements of $L(\mathbb{Z}/n)$ where $\mathbb{Z}/n$ denotes the integers modulo n (see Definition 1.28). The above definition is saying that f is constant on residue classes modulo n.

Now we can define the Fourier transform on cyclic groups. Recall that in Theorem 2.8 we showed that irreducible characters of any finite group G form an orthonormal set of class functions in $L(G)$, that is functions which are constant on conjugacy classes. Thus for any function $f : \mathbb{Z}/n \to \mathbb{C}$ we can decompose f into orthonormal components using the irreducible characters of $L(G)$. Specifically for $\mathbb{Z}/n$, the irreducible characters form a basis of $L(\mathbb{Z}/n)$. This decomposition is what the Fourier transform encapsulates.

Definition 4.3: Fourier Transform on Cyclic Groups

Let $f : \mathbb{Z}/n \to \mathbb{C}$. We define the *Fourier transform* $\hat{f} : \mathbb{Z}/n \to \mathbb{C}$ of f as:

$$\hat{f}(\overline{m}) = n\langle \chi_m, f\rangle = \sum_{k=0}^{n-1} e^{-2\pi imk} f(\overline{k}).$$

Remark 4.4. Notice that the Fourier Transform is a linear transformation $T : L(\mathbb{Z}/n) \to L(\mathbb{Z}/n)$ since inner products are linear in the second variable.

Theorem 4.5: Fourier Inversion on Cyclic groups

The Fourier transform is invertible. That is,

$$f = \frac{1}{n} \sum_{k=0}^{n-1} \hat{f}(\overline{k}) \chi_k.$$

Proof. See page 48 of [1] for further reading. We omit the proof here but provide the proof for abelian groups below. □

4.2. **Fourier Analysis on Abelian Groups.** We will now use character theory to introduce Fourier Analysis on abelian groups, a simpler case of the most general theory. Often analysis of abelian Groups is sufficient for most applications.

Observation 4.6. Consider an arbitrary finite abelian group $G = \{g_i\}_{i=1}^n$. Recall Definition 2.6 of class functions and notice that when G is abelian we have $Z(L(G)) = L(G)$ (see Definition 1.28). That is the space of class functions on G, $Z(L(G))$, is the same as the space of all functions on G, $L(G)$. Furthermore from Remark 2.9 we know that the irreducible characters of G, $\{\chi_i\}_{i=1}^n$ form an orthonormal set of class functions, and in fact are a basis of $Z(L(G)) = L(G)$. Thus we can define the Fourier Transform to encapsulate this.

Definition 4.7: Fourier Transform on Abelian Groups

Let $f : G \to \mathbb{C}$. The *Fourier transform* $\hat{f} : G \to \mathbb{C}$ is:

$$\hat{f}(g_i) = n\langle \chi_i, f\rangle = \sum_{g\in G} \overline{\chi_i(g)} f(g).$$

Remark 4.8. The complex numbers $n\langle \chi_i, f\rangle$ are called the *Fourier coefficients* of f.

Theorem 4.9: Fourier Inversion for Abelian Groups

If $f \in L(G)$, then

$$f = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(g_i)\chi_i.$$

Proof. From [1],

$$f = \sum_{i=1}^{n} \langle \chi_i, f\rangle \chi_i = \frac{1}{n}\sum_{i=1}^{n} n\langle \chi_i, f\rangle \chi_i = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(g_i)\chi_i$$

where the first equality is from decomposing f into orthonormal components since the irreducible characters form a basis for $L(G)$ when G is abelian. □

4.3. **Fourier Transform at a Representation.** Finally we define the Fourier Transform of a function on a finite group G at a representation ϕ.

Definition 4.10: Fourier Transform at a Representation

Let $f : G \to \mathbb{C}$ be a function on G. The *Fourier Transform* of f at the representation ϕ of G is the matrix:

$$\hat{f}(\phi) = \sum_{g\in G} f(g)\phi(g).$$

Remark 4.11. In the following section, f will be a probability distribution on G and we will denote the function P.

Now that we have presented Fourier Transforms on cyclic and abelian groups as well as at representations, we will delve into an application in probability theory: Random Walks on Finite Groups.

5. Random Walks on Finite Groups

Note that in this section when referring to the Fourier Transform and using the notation $\hat{f}$ it should be assumed we are working with Definition 4.10.

Let's start with the simplest example of a random walk on a finite group.

Example 5.1. Imagine $\mathbb{Z}/n$ as n points on a circle. The simplest random walk on this group is one where we start on one of the points and have probability $\frac{1}{2}$ to move left or right.

Already we can ask some very interesting questions such as:

- How many steps does it take to hit every point?
- How many steps does it take to reach point n?
- How many steps does it take for our distribution to be close to uniformly random?

Let's make close to uniformly random mathematically precise.

Remark 5.2. Note that in the following definition when we say "probability distribution" on a group G we are referring to a probability measure as in Definition 3.9. In fact in general when dealing with probability with reference to a group G one should think of G as the sample space, subsets of G as the event space (sigma algebra of G), and probability distributions as probability measures. Another important note is that "probability distribution" is often used over probability measure when the elements of the sample space are of great interest, as in the case where the sample space is discrete or as in our case when it is a group. When we are speaking about a general sample and event space the term probability measure is more common.

Definition 5.3: Variation Distance

Let G be a finite group. Let P and Q be probability distributions on G. Then the *variation distance* between P and Q is

$$\|P - Q\| = \max_{A \subseteq G} |P(A) - Q(A)|.$$

Remark 5.4. Intuitively the variation distance is the largest possible distance between probabilities assigned by the two probability distributions to the same event (subset of G).

The following definition is a frequently referenced probability distribution.

Definition 5.5: Uniform Distribution

Let U denote the *uniform probability distribution* over a group G. That is U is the probability distribution which assigns equal probability to each $g \in G$. Note that this only defines a measure because G is finite so we can assign non-zero measure to singletons.

Example 5.6. The easiest example of a discrete uniform distribution is the outcomes of rolling a fair six-sided dice.

We will now present some properties of variation distance without proof. Please see pages 21-22 of [3] for further reading.

Proposition 5.7. *Let P and Q be probability distributions on G. Then,*

$$\|P - Q\| = \frac{1}{2}\sum_{g\in G}|P(g) - Q(g)| = \frac{1}{2}\max_{\|f\|\leq 1}|P(f) - Q(f)|,$$

where $f : G \to \mathbb{R}$ is such that $|f(g)| \leq 1$ and $P(f) = \sum_{g\in G} P(g)f(g)$.

Remark 5.8. The quantity $P(f) = \sum_{g\in G} P(g)f(g)$ is commonly known as the *expected value* of f under P.

Proposition 5.9. *Let U denote the uniform distribution on G, and let $h : G \to G$ be injective (one to one), then*

$$\|P - U\| = \left\|Ph^{-1} - U\right\|$$

where $Ph^{-1}(A) = P(h^{-1}(A))$.

Remark 5.10. Sometimes *variation distance* is defined as the identity in Proposition 5.7.

We can now rigorously introduce close to uniformly random.

Problem 5.11. *Let P be a probability on a finite group G. Given any $\varepsilon > 0$, how large should k be so that $\left\|P^{*k} - U\right\| < \varepsilon$. P^{*k} denotes P convoluted with itself k times (see Definition 3.1).*

Remark 5.12. Recall from Definition 3.1 that $P * P(g) = \sum_{h\in G} P(gh^{-1})P(h)$. In the language of random walks $P * P(g)$ is the chance that a random walk on G generated using elements with weight $P(g)$ is at g after two steps. This is because first we would have to be at some element h followed by gh^{-1} to get to g. The same intuition holds for $P(g)^{*k}$. $P(g)^{*k}$ is the chance that our walk is at g after k steps.

We will begin approaching this question by developing bounds to approximate the variation distance. In order to present our first upper bound we will need to use Plancherel's Formula, a famous result in Fourier Analysis and Representation Theory.

Theorem 5.13: Plancherel's Formula

Let f and h be functions on a finite group G. Then,

$$\sum_{g\in G} f(g^{-1})h(g) = \frac{1}{|G|}\sum d_i \operatorname{Tr}(\hat{f}(\phi_i)\hat{h}(\phi_i))$$

where the second summation is over all non-trival irreducible representations ϕ_i of G and d_i is the degree of ϕ_i.

Proof. See page 13 of [3]. □

Now for our first upper bound.

Theorem 5.14: Upper Bound Lemma

Let P be a probability on the finite group G. Then

$$\|P - U\|^2 \leq \frac{1}{4} \sum^{*} d_\phi \operatorname{Tr}(\hat{P}(\phi)\hat{P}(\phi)^*)$$

where the sum is over all non-trival irreducible representations ϕ of G with degree d_ϕ and $\hat{P}(\phi)^*$ denotes the conjugate transpose of the Fourier transform $\hat{P}(\phi)$.

Proof. By Proposition 5.7 we know that

$$\|P - U\| = \frac{1}{2} \sum_{g \in G} |P(g) - U(g)|.$$

Thus,

$$\begin{aligned} 4\|P - U\|^2 &= \Big[\sum_{g \in G} |P(g) - U(g)|\Big]^2 \\ &\leq |G| \sum_{g \in G} |P(g) - U(g)|^2 \\ &= \sum^{*} d_\phi \operatorname{Tr}(\hat{P}(\phi)\hat{P}(\phi)^*). \end{aligned}$$

The inequality is from Cauchy-Schwarz and the last equality is follows from Plancherel's Formula 5.13 and Proposition 5.7. □

We have now developed the machinery to answer one of the questions in Example 5.1:

Problem 5.15. *Consider the elements of $\mathbb{Z}/m$ (integers mod m) as points around a circle. Define $P(-1) = P(1) = \frac{1}{2}$ and $P(i) = 0$ otherwise. That is, we have probability $\frac{1}{2}$ of moving either one step clockwise or counterclockwise around our circle at any given point. The following theorem shows that it takes more than m^2 steps for our distribution to be close to uniformly random.*

Theorem 5.16: Simple Random Walk on the Circle

When m is odd and greater than 7, we have, $\|P^{*n} - U\| \leq e^{-\alpha n/m^2}$ where $\alpha = \frac{\pi^2}{2}$.

The following proof is from [3].

Proof. The Fourier Transform of P is $\hat{P}(j) = \frac{1}{2}(e^{\frac{2\pi i j}{m}} + e^{\frac{-2\pi i j}{m}}) = \cos(\frac{2\pi j}{m})$. From the Upper Bound Lemma 5.14 we have

$$\|P^{*n} - U\|^2 \leq \frac{1}{4} \sum_{j=1}^{m-1} \cos(\frac{2\pi j}{m})^{2n} = \frac{1}{2} \sum_{j=1}^{\frac{m-1}{2}} \cos(\frac{\pi j}{m})^{2n}.$$

Now using the property of cosine that $\cos(x) \leq e^{\frac{-x^2}{2}}$ for $x \in [0, \frac{\pi}{2}]$ we have,

$$\begin{aligned}\|P^{*n} - U\|^2 \leq \frac{1}{2}\sum_{j=1}^{\frac{m-1}{2}} \cos(\frac{\pi j}{m})^{2n} &\leq \frac{1}{2} e^{\frac{-\pi^2 n}{m^2}} \sum_{j=1}^{\infty} e^{\frac{-\pi^2(j^2-1)n}{m^2}} \\ &\leq \frac{1}{2} e^{\frac{-\pi^2 n}{m^2}} \sum_{j=0}^{\infty} e^{\frac{-3\pi^2 jn}{m^2}} \\ &= \frac{1}{2} \cdot \frac{e^{\frac{-\pi^2 n}{m^2}}}{1 - e^{\frac{-3\pi^2 n}{m^2}}}.\end{aligned}$$

The above calculation holds for any n and any odd m. Furthermore when $n \geq m^2$ we have that $\left[2(1 - e^{\frac{-3\pi^2 n}{m^2}})\right]^{-1} < 1$ and so we have proved a stronger bound than claimed. □

So, we have shown that it takes more than m^2 steps for our random walk around the circle of $\mathbb{Z}/m$ to be close to uniformly distributed.

6. Further Reading

Fourier Analysis is a very broad subject with applications in various fields from signal processing to options pricing. As such there are many paths one can explore to learn and apply the material presented in this paper. We provide a few of these paths here.

6.1. **Random Walks.** If you enjoyed the material presented in the previous section, Diaconis [3] provides a multitude of other examples including:

- Random number generators base on the recurrence relation $X_{k+1} = aX_k + b$.
- Card shuffling.
- Random walks on the n cube.

6.2. **Ergodic theory.** Ergodic Theory is the study of statistical properties of dynamical systems. Fourier Analysis is frequent used throughout the field. See here for examples of how to use Fourier representation to prove ergodicity. Also please see Introduction to Ergodic Theory by Maryam Mirzakhani of MIT for more exciting topics including toral endormophisms and Bernoulli shifts.

6.3. **Fast Fourier Transform.** The Fast Fourier Transform (FFT) may be the single most important algorithm in computer science. The Fast Fourier Transform is a fast convolution algorithm that allows us to both multiply and add polynomials in time complexity $n\log(n)$. Please see The Fourier Transform on Finite Groups: Theory and Computation by Rohan Dandavati - written through this very REU program - for a thorough introduction including pseudo-code.

7. Acknowledgments

I would like to thank my mentor Maxwell Johnson for his invaluable guidance, suggestions, and patience while I was working on this paper. Without Max I wouldn't have been able to navigate this topic and understand the mathematics behind the Fourier Transform. I wish him the best of luck as he begins his Ph.D. I'd also like to thank Daniil Rudenko for his wonderful lectures and thought-provoking questions. Lastly, I'm very thankful to Peter May for organizing the REU and inviting such a large variety of speakers. Due to Dr. May,

I was exposed to a incredible assortment of mathematics which will guide my academic path in the following years. I'm very grateful that I was given this opportunity.

References

[1] B.Steinberg. Representation Theory of Finite Groups. Carleton University 2009.
[2] Alex Tsun. Multiple Random Variables. University of Washington.
[3] Persi Diaconis. Group Representations in Probability and Statistics. Harvard University 1998
[4] Lawrence Narici; Edward Beckenstein. Topological Vector Spaces. CRC Press 2011.
[5] Audrey Terras. Fourier Analysis on Finite Groups and Applications. Cambridge University Press 1999.
[6] Joseph K. Blitzstein and Jessica Hwang. Introduction to Probability. Harvard University and Stanford University 2019.