THE ALGEBRAIC STRUCTURE UNDERLYING FOURIER ANALYSIS

SIDDHARTH KOTHARI

Abstract. In this paper, we begin with the concept of a *projection* in the familiar vector space $\mathbb{R}^n$, and then extend it to an arbitrary inner product space. This foundational idea enables us to compute a line of best fit, approximate functions using polynomials, and decompose a function into an infinite series of sines and cosines, known as its Fourier series. We then extend the Fourier series to finite groups using representation and character theory. In the process, we will examine key results, including Maschke's theorem and Schur's orthogonality relations. Lastly, we provide a reasonably detailed guide to further extend this reasoning to compact groups using the Peter-Weyl theorem. For the most part, a background in lower-division linear algebra and introductory group theory will be sufficient, but towards the end, knowledge of measure theory and general topology will be necessary.

Contents

1. Introduction

In most texts on the subject, the *Fourier series* of a real-valued periodic function f is derived as follows. First, one *assumes* that we can write f as an infinite sum of the simple, the *fundamental*, sines and cosines. In symbols, we have that

$$f(x) = A_0 + \sum_{n=1}^{\infty} A_n \cos(nx) + \sum_{n=1}^{\infty} B_n \sin(nx),$$

where the A_i and B_i are some constants. Then, quite naturally, the goal is to compute those constants, which is achieved by some rapid fire integration, as carried out below. They start by multiplying the whole equation by the expression $\cos(mx)$,

$$f(x)\cos(mx) = A_0 \cos(mx) + \sum_{n=1}^{\infty} A_n \cos(nx)\cos(mx) + \sum_{n=1}^{\infty} B_n \sin(nx)\cos(mx),$$

and then integrating both sides from T to $T+2\pi$,[1] where T is any point in the function's domain,

$$\begin{aligned}\int_T^{T+2\pi} f(x)\cos(mx)\,dx = &\int_T^{T+2\pi} A_0\cos(mx)\,dx\\ &+\sum_{n=1}^{\infty} A_n\left(\int_T^{T+2\pi}\cos(nx)\cos(mx)\,dx\right)\\ &+\sum_{n=1}^{\infty} B_n\left(\int_T^{T+2\pi}\sin(nx)\cos(mx)\,dx\right).\end{aligned}$$

But why did we bother doing this? Oh, that's because the integral $\int_T^{T+2\pi}\cos(nx)\cos(mx)\,dx$ simplifies down to 0 when $m \neq n$; a fact commonly known as the *orthogonality* relations. Also, the other two integrals $\int_T^{T+2\pi}\sin(nx)\cos(mx)\,dx$ and $\int_T^{T+2\pi}\cos(mx)\,dx$ are always zero! This is good news for us, since we can extract the A_n for $n \geq 1$,

$$\int_T^{T+2\pi} f(x)\cos(nx)\,dx = A_n\int_T^{T+2\pi}\cos^2(nx)\,dx.$$

And since $\int_T^{T+2\pi}\cos^2(nx)\,dx = \pi$, we get that

$$A_n = \frac{1}{\pi}\int_T^{T+2\pi} f(x)\cos(nx)\,dx.$$

In a similar fashion, we can compute A_0 and the B_n to get

$$A_0 = \frac{1}{\pi}\int_T^{T+2\pi} f(x)\,dx \quad \text{and} \quad B_n = \frac{1}{\pi}\int_T^{T+2\pi} f(x)\sin(nx)\,dx.$$

Now, unarguably this is a quick, no-nonsense way of deriving the Fourier series, but I feel that quite a bit of the rich structure that underlies this is swept under the rug, making this procedure seem unmotivated, almost as a bolt from the blue. For example, while this equation

$$\int_T^{T+2\pi}\cos(nx)\cos(nx)\,dx = 0 \quad \textit{when } m \neq n,$$

and

$$\int_T^{T+2\pi}\sin(nx)\sin(mx)\,dx = 0 \quad \textit{when } m \neq n,$$

are crucial to this derivation, why are they called the *orthogonality relations*? Where are the perpendicular lines?

Moreover, this explanation misses the fact that the idea behind the Fourier series is the *exactly the same* as the idea behind *computing lines of best fit*, and *approximating arbitrary functions by polynomials*. We explore this unexpected relationship in section 2 of the paper, and in section 3, we take this to even greater heights, making an attempt to generalize this analysis of breaking up a function into simpler parts in the abstract setting of a group. Let's begin!

2. Fourier Series in the Real Setting

2.1. **The cornerstone.** Though we threw a flurry of mathematical buzzwords in the abstract, our journey starts with no more than an elementary theorem, one that could be plucked from any middle school math textbook.

Theorem 2.1. *Given a point P, and a line l, the shortest distance between the two is the length of the segment m, connecting the two such that it's perpendicular to l.*

[1] Note that here the period of f is taken to be 2π.

As it turns out, much of the paper is simply an in-depth version of this rudimentary, apparently *useless*, fact. While we do not present a formal proof of the theorem (silly thing to do given it's sheer *intuitiveness*), we do give a couple of proof sketches.

One uses the Pythagorean theorem. Start by drawing an arbitrary line m' from P to l, which is the hypotenuse of the right triangle PQQ' (see Figure 1). Since $(PQ)^2 + (QQ')^2 = (PQ')^2$, we clearly have that $PQ \leq PQ'$.

The other proof is one any calculus student would instantly churn out: first compute an expression for the length of m' in terms of the x coordinate of Q', then compute it's minimum value by setting its derivative to zero and solving the resulting equation.

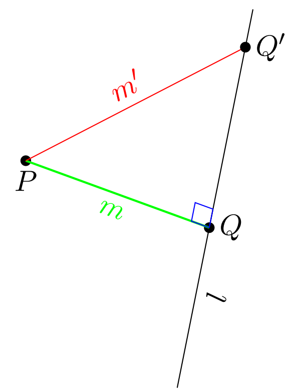

Figure 1. Diagram for the first proof of Theorem 2.1

Arguably, Theorem 2.1 isn't very useful, at least in the Euclidean geometry setting. It truly comes to life where we can talk about lengths, distances and angles generally, without necessarily referencing to tangible reality—exactly what a vector space [2] does.

2.2. **Projection in $\mathbb{R}^n$.** We start by presenting an apparently unrelated problem. Given n points in the plane, $(x_1, y_1), \cdots, (x_n, y_n)$, how would one compute a line, $y = mx + c$, that best approximates the data? Phrased differently, we want to *best* solve the system of n equations $y_i = mx_i + c$ for m and c, or the matrix-vector equation, $\mathbf{Ax} = \mathbf{b}$, for the vector $\mathbf{x} = (m, c)$ as given below.

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

But then, what do we *mean* by best solve? One way is that we *minimize* the length of the error vector, $\mathbf{e} := \mathbf{b} - \mathbf{Ax}$. Note that $\mathbf{Ax}$ cannot *leave* [3] a two dimensional linear subspace of $\mathbb{R}^n$, namely $C(\mathbf{A})$, or the column space of $\mathbf{A}$, which is an alternative way of representing the constraints on this problem. Hence, our question takes on the form:

Question 2.2. *Find the vector $\mathbf{p} \in \mathbb{R}^n$ such that $\mathbf{p} \in C(\mathbf{A})$ and $\|\mathbf{b} - \mathbf{p}\|$ is minimized.*

Now for the moment of insight: $C(\mathbf{A})$ is nothing but a plane, and $\mathbf{b}$, a vector (*possibly*) sticking out of it. What's the shortest distance between the two? Simple, the length of the vector that connects the two such that it's perpendicular to the plane! Did you notice what just happened? You were subconsciously relating back to Theorem 2.1! You were drawing intuition from it, even though we can't construct a formal proof of this particular version *just* using Theorem 2.1.

[2] Actually a *Hilbert space*, but we'll get to that later.

[3] More precisely, $\mathbf{Ax} \in C(\mathbf{A})$ for all $\mathbf{x} \in \mathbb{R}^n$. Note also that $C(\mathbf{A})$ doesn't always have to be two dimensional. In fact, it is one dimensional iff $x_i = 1$ for all i.

Now, we attempt to formalize this argument by first re-stating our conjecture in it's generalized form, deriving an expression for $\mathbf{p}$, and then proving that our $\mathbf{p}$ does indeed minimize the distance. Before we can proceed though, we must define *what we mean* when we say 'the vector is perpendicular to the plane'.

First, recall that two *non-zero* vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular, or *orthogonal*, if and only if $\cos\theta = 0$ where θ is the angle between the two vectors. Note that $\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ so $\cos\theta = 0 \iff \mathbf{x}\cdot\mathbf{y} = 0$. We frame this as Proposition 2.3.

Proposition 2.3. *Two vectors are orthogonal in $\mathbb{R}^n$ if and only if their dot product is zero. Note that the zero vector is orthogonal to any vector in $\mathbb{R}^n$.*

At this point, it is logical to say that a vector $\mathbf{x} \in \mathbb{R}^n$ is orthogonal to some linear subspace $\mathcal{S} \subseteq \mathbb{R}^n$ if it is orthogonal to every vector in $\mathcal{S}$.

Definition 2.4. Given a subspace $\mathcal{S} \subseteq \mathbb{R}^n$, a vector $\mathbf{x} \in \mathbb{R}^n$ is said to be *orthogonal* to $\mathcal{S}$ if $\mathbf{x}\cdot\mathbf{a} = 0$ for all $\mathbf{a} \in \mathcal{S}$. We denote this by $\mathbf{x} \perp \mathcal{S}$.

With those preliminaries taken care of, we can state our claim.

Claim 2.5. *Given a subspace $\mathcal{S} \subseteq \mathbb{R}^n$ and a vector $\mathbf{b}$, there exists a vector $\mathbf{p}$ so that $\|\mathbf{b}-\mathbf{p}\| = \min_{\mathbf{x}\in\mathcal{S}}\|\mathbf{b}-\mathbf{x}\|$ which is such that $\mathbf{b}-\mathbf{p} \perp \mathcal{S}$.*

As mentioned, we now compute an expression for $\mathbf{p}$, again, solely by drawing intuition from Theorem 2.1. Let's start by assigning a basis $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ for the subspace $\mathcal{S}$ (taken to be $m \le n$ dimensional). By the linearity of the dot product $\mathbf{b}-\mathbf{p} \perp \mathcal{S} \iff \mathbf{b}-\mathbf{p} \perp \mathbf{a}_i$ for all $1 \le i \le m$, i.e, $\mathbf{p}$ is orthogonal to each of the basis vectors for $\mathcal{S}$. Thus, we have an equation for each i

$$\mathbf{b}-\mathbf{p} \perp \mathbf{a}_i \iff \mathbf{a}_i\cdot(\mathbf{b}-\mathbf{p}) = 0,$$

which can be re-casted into a single matrix-vector equation

$$\mathbf{A}^T(\mathbf{b}-\mathbf{p}) = 0,$$

where $\mathbf{A}$ is the n by m matrix such that it's i^{th} column is $\mathbf{a}_i$. Observe that $\mathbf{p} \in \mathcal{S} \implies \mathbf{p} \in \operatorname{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$, meaning we can find unique scalars $x_1, \cdots, x_m$ such that $\mathbf{p} = \mathbf{a}_1 x_1 + \cdots + \mathbf{a}_m x_m = \mathbf{A}\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^m$ is such that it's i^{th} component is x_i. On substituting this into our previous equation we get

$$\mathbf{A}^T(\mathbf{b}-\mathbf{A}\mathbf{x}) = 0 \iff \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}.$$

To solve for $\mathbf{x}$, the coefficient vector, we can simply multiply both sides by $(\mathbf{A}^T\mathbf{A})^{-1}$ (which is indeed invertible for all matrices $\mathbf{A}$ whose columns are linearly independent; see the appendix for proof) which yields

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \tag{2.1}$$

Lastly, remembering that $\mathbf{p} = \mathbf{A}\mathbf{x}$, we get our sought after expression for $\mathbf{p}$

$$\mathbf{p} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \tag{2.2}$$

Remark 2.6. Equation 2.2 is known in the literature as the *projection formula*: the analogy here is that $\mathbf{p}$ is like the shadow of $\mathbf{b}$ on $\mathcal{S}$, commonly pictured as a plane. The vector $\mathbf{p}$ is called the *projection* of $\mathbf{b}$ onto the subspace $\mathcal{S}$, denoted by $\mathbf{p} = \operatorname{proj}_{\mathcal{S}}(\mathbf{b})$.

Remark 2.7. Notice that we have a linear operator $\mathbf{P} : \mathbb{R}^n \to \mathcal{S}$ given by $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. This is called the *projection operator onto the subspace* $\mathcal{S}$. Note that $\operatorname{proj}_{\mathcal{S}}(\mathbf{b}) = \mathbf{P}\mathbf{b}$.

Remark 2.8. It is of extreme importance that we show that $\mathbf{P}$ is independent of the choice of basis for $\mathcal{S}$—which is certainly not apparent from Equation 2.2. To see this, consider two bases $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$ with corresponding matrices $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$ and $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m]$. Note that there exists an invertible matrix $\mathbf{C}$ such that $\mathbf{B} = \mathbf{AC}$, called the *change of basis matrix*. The projection matrix using the first basis $\{\mathbf{a}_i\}$ is $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, and similarly, using the second basis $\{\mathbf{b}_i\}$ is $\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$. Next, substituting $\mathbf{B} = \mathbf{AC}$ into the expression for $\mathbf{P}_B$, we get

$$\begin{aligned}
\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T &= (\mathbf{AC})\left((\mathbf{AC})^T(\mathbf{AC})\right)^{-1}(\mathbf{AC})^T = \mathbf{AC}(\mathbf{C}^T\mathbf{A}^T\mathbf{AC})^{-1}\mathbf{C}^T\mathbf{A}^T \\
&= \mathbf{AC}\left(\mathbf{C}^{-1}(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{C}^T)^{-1}\right)\mathbf{C}^T\mathbf{A}^T = \mathbf{A}(\mathbf{CC}^{-1})(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{C}^T)^{-1}\mathbf{C}^T)\mathbf{A}^T \\
&= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{P}_A.
\end{aligned}$$

Thus, $\mathbf{P}_A = \mathbf{P}_B$, which shows that the Equation 2.2 is indeed independent of the choice of basis. Quite a relief!

Remark 2.9. In fact, if the basis $\{\mathbf{a}_1, \cdots, \mathbf{a}_m\}$ for $\mathcal{S}$ is an orthogonal one, then Equation 2.2 simplifies tremendously: $\mathbf{A}^T\mathbf{A}$ becomes a diagonal matrix whose i^{th} diagonal entry is $\|a_i\|^2$, meaning we have

$$\mathbf{p} = \mathbf{A}\begin{pmatrix} \|\mathbf{a}_1\|^{-2} & 0 & \cdots & 0 \\ 0 & \|\mathbf{a}_2\|^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{a}_n\|^{-2} \end{pmatrix}\mathbf{A}^T\mathbf{b} = \sum_{i=1}^{n} \frac{1}{\|\mathbf{a}_i\|^2}\mathbf{a}_i\mathbf{a}_i^T\mathbf{b} = \sum_{i=1}^{n} \frac{\mathbf{a}_i \cdot \mathbf{b}}{\mathbf{a}_i \cdot \mathbf{a}_i}\mathbf{a}_i = \sum_{i=1}^{n} \text{proj}_{\text{span}\{\mathbf{a}_i\}}(\mathbf{b}).$$

Essentially, in this case, projecting the vector $\mathbf{b}$ onto $\mathcal{S}$ boils down to the projecting $\mathbf{b}$ onto each of the individual one-dimensional linear subspaces spanned by each of the basis vectors, and then adding the results up. Keep this in mind!

Example. Before moving on, let's get our hands dirty and actually compute the line of best fit for a set of points, say $S = \{(1,2), (2,3), (3,5), (4,2), (5,6), (6,9), (7,11), (8,7), (9,10), (10,4)\}$, which is also shown in Figure 2. The corresponding matrix vector equation for S is given by

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \\ 7 & 1 \\ 8 & 1 \\ 9 & 1 \\ 10 & 1 \end{pmatrix}\begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 2 \\ 6 \\ 9 \\ 11 \\ 7 \\ 10 \\ 4 \end{pmatrix} \text{ where } \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \\ 7 & 1 \\ 8 & 1 \\ 9 & 1 \\ 10 & 1 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 2 \\ 6 \\ 9 \\ 11 \\ 7 \\ 10 \\ 4 \end{pmatrix}.$$

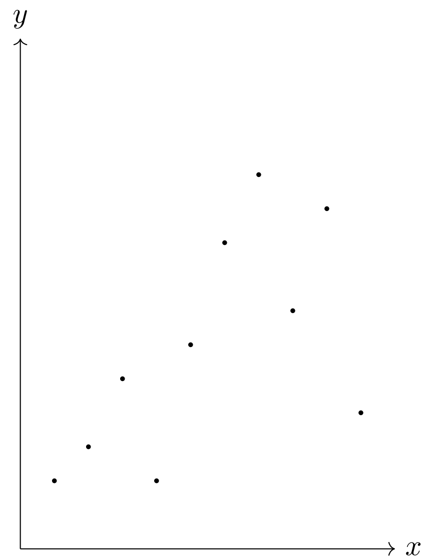

Figure 2. The points in S plotted on the xy plane.

Plugging our $\mathbf{A}$ and $\mathbf{b}$ into Equation 2.1 and simplifying

$$\mathbf{x} = \begin{pmatrix} 385 & 55 \\ 55 & 10 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \mathbf{b} = \begin{pmatrix} \frac{2}{165} & \frac{-1}{15} \\ \frac{-1}{15} & \frac{7}{15} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \mathbf{b}$$
$$= \begin{pmatrix} -\frac{3}{55} & -\frac{7}{165} & -\frac{1}{33} & -\frac{1}{55} & -\frac{1}{165} & \frac{1}{165} & \frac{1}{55} & \frac{1}{33} & \frac{7}{165} & \frac{3}{55} \\ \frac{2}{5} & \frac{1}{3} & \frac{4}{15} & \frac{1}{5} & \frac{2}{15} & \frac{1}{15} & 0 & -\frac{1}{15} & -\frac{2}{15} & -\frac{1}{5} \end{pmatrix} \mathbf{b} = \begin{pmatrix} \frac{118}{165} \\ \frac{7}{3} \end{pmatrix}.$$

Thus, our line should have a slope of $\frac{118}{165}$ and a $y-$intercept of $\frac{7}{3}$. Let's draw it!

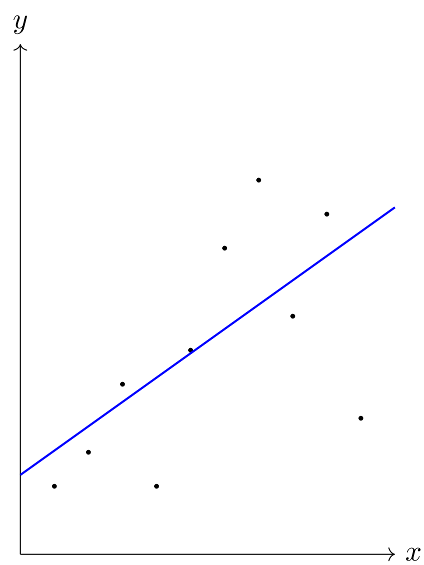

Figure 3. The points in S plotted on the xy plane along with the line of best fit.

There we have it—a pretty decent approximating line! Try experimenting with other data points!

2.3. **The Best Approximation Theorem.** To unlock the maximum potential of Theorem 2.1, we have to move away from $\mathbb{R}^n$ to an arbitrary vector space, V. However, as you might have guessed, V can't be any vector space, but one with additional structure—one that allows us to talk about closeness, lengths and angles. Recall that an inner product space is the perfect setting. As a quick recap, we state the only definitions that will be important to us.

Definition 2.10. Given a vector space V over a field $\mathbb{F}$, an *inner product* on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ that satisfies the following axioms:

- **Positive definiteness:** For all $v \in V$, we have $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0 \iff v = 0$.
- **Linearity in the first argument:** For all $v, u, w \in V$ and $\alpha, \beta \in \mathbb{F}$, we have $\langle \alpha v + \beta u, w \rangle = \alpha \langle v, w \rangle + \beta \langle u, w \rangle$.
- **Conjugate symmetry:** For all $v, w \in V$, we have $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

Definition 2.11. An *inner product space* is an ordered pair $(V, \langle \cdot, \cdot \rangle)$ where V is a vector space and $\langle \cdot, \cdot \rangle$ and inner product on V.

Remark 2.12. It's helpful to view the inner product as the generalization of the dot product on $\mathbb{R}^n$ —all the important *properties* that made the dot product the dot product find themselves in the *definition* of the inner product.

Taking inspiration from the fact that the length of $\mathbf{v} \in \mathbb{R}^n$ is equal to $\sqrt{\mathbf{v} \cdot \mathbf{v}}$, we define the norm of a vector in an inner product space V.

Definition 2.13. Given an inner product space $(V, \langle \cdot, \cdot \rangle)$, the map from $\| \cdot \| : V \to \mathbb{F}$ given by $\|v\| = \sqrt{\langle v, v \rangle}$ is called the *norm* on V.

Remark 2.14. The norm on V induces a metric $d : V \times V \to \mathbb{R}_{\geq 0}$ given by $d(x, y) = \|x - y\|$ for all $x, y \in V$.

Definition 2.15. Let V be an inner product space, $x, y \in V$ and $A, B \subseteq V$ Then, we say that

- x and y are *orthogonal* if $\langle x, y \rangle = 0$ and is denoted by $x \perp y$,
- x and A are *orthogonal* if $\langle x, a \rangle = 0$ for all $a \in A$ and is denoted by $x \perp A$,
- A and B are *orthogonal* if $\langle a, b \rangle = 0$ for all $a \in A$ and $b \in B$ and is denoted by $A \perp B$.

The concept of a Hilbert space ties the chain

$$\text{inner product} \to \text{norm} \to \text{metric}$$

together into a single structure.

Definition 2.16. An inner product space $(V, \langle \cdot, \cdot \rangle)$ is said to be a *Hilbert space* if the corresponding norm induces a metric d such that V is complete with respect to d. That is, every Cauchy sequence $(x_n)_{n \in \mathbb{N}}$ in V converges to a point $x \in V$.

We now give a few examples of inner product spaces—some might already be familiar.

Example. The usual real euclidean space $\mathbb{R}^n$ with the dot product as the inner product.

Example. The complex version of euclidean space, $\mathbb{C}^n = \{(z_1, \cdots, z_n) : z_i \in \mathbb{C}\}$ with vector addition and scalar multiplication given component wise, along with the complex version of the dot product,

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i \overline{w_i} \text{ for all } v, w \in \mathbb{C}^n.$$

Think about why we added the complex conjugate!

Example. The set $L^2([a, b])$ comprising of the real-valued integrable functions $f : [a, b] \to \mathbb{R}$. Here, vector addition and scalar multiplication are point-wise, and the inner product is given by

$$\langle f, g \rangle_{L^2} = \int_a^b f(x) g(x) \, dx \quad \text{for all } f, g \in L^2([a, b]).$$

The norm induced by the inner product is

$$\|f\|_{L^2} = \left(\int_a^b [f(x)]^2 \, dx\right)^{\frac{1}{2}} \quad \text{for all } f \in L^2([a,b]).$$

To ensure that the norm of any function is finite, we restrict $L^2([a,b])$ to functions such that $\|f\|_{L^2} = \left(\int_a^b [f(x)]^2 \, dx\right)^{\frac{1}{2}} < \infty$. In symbols,

$$L^2([a,b]) = \left\{ f : [a,b] \to \mathbb{R} \mid f \text{ is integrable and } \int_a^b |f(x)|^2 \, dx < \infty \right\}.$$

Note that elements of $L^2([a,b])$ are called *square integrable functions*.

All the three examples above are also Hilbert spaces. Try proving it! After this quick setup, we can state the generalized version of Claim 2.5.

Theorem 2.17. *Given a finite dimensional subspace $\mathcal{S}$ of an inner product space V, and a vector $v \in V$, there exists a vector $p \in \mathcal{S}$ such that $\|v - p\| = \min_{x \in \mathcal{S}} \|v - x\|$. Furthermore, p is unique, and is such that $v - p \perp \mathcal{S}$.*

Given our earlier reasoning about Claim 2.5 and Remark 2.9, we would expect that

$$p = \sum_{i=1}^{n} \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, \tag{2.3}$$

where $\{e_i\}_{i=1}^n$ is an orthogonal basis for $\mathcal{S}$—compare this with $\sum_{i=1}^n \frac{\mathbf{a}_i \cdot \mathbf{b}}{\mathbf{a}_i \cdot \mathbf{a}_i} \mathbf{a}_i$ we had earlier. With this in mind, we dive into the proof.

Proof. This proof is spilt into two parts: first, we will show that if $p \in \mathcal{S}$ is such that $v - p \perp \mathcal{S}$, then $\|v - p\| = \min_{x \in \mathcal{S}} \|v - x\|$, that is $\|v - x\| \geq \|v - p\|$; and then the expression $e = v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i$ is orthogonal to $\mathcal{S}$.

First, by the positive definiteness of the norm, $\|v - x\| \geq 0$ and $\|v - p\| \geq 0$, and so we have

$$\|v - x\| \geq \|v - p\| \iff \|v - x\|^2 \geq \|v - p\|^2.$$

Thus, it suffices to show $\|v - x\|^2 \geq \|v - p\|^2$ for all $x \in \mathcal{S}$. This expression, with the square, is easier to expand!

Start by considering an arbitrary $x \in \mathcal{S}$. We can write $x = p + (x - p)$, and since $\mathcal{S}$ is closed under vector addition, $x - p \in \mathcal{S}$ and so $x = p + y$ for $y \in \mathcal{S}$. Expanding $\|v - x\|^2$:

$$\|v - x\|^2 = \|v - (p + y)\|^2 = \|(v - p) - y\|^2.$$

Since $v - p \perp y$,

$$\|(v - p) - y\|^2 = \|v - p\|^2 + \|y\|^2,$$

by the Pythagorean theorem. As $\|y\|^2 \geq 0$:

$$\|v - x\|^2 = \|v - p\|^2 + \|y\|^2 \implies \|v - x\|^2 \geq \|v - p\|^2.$$

For the uniqueness part, we have to show that $\|v - x\| = \|v - p\| \iff x = p$. Again, since $\|v - x\| \geq 0$ and $\|v - p\| \geq 0$, we have

$$\|v - x\| = \|v - p\| \iff \|v - x\|^2 = \|v - p\|^2.$$

Thus, it suffices to show $\|v - x\|^2 = \|v - p\|^2 \iff x = p$, as done below

$$\|v - x\|^2 = \|v - p\|^2 \iff \|v - p\|^2 + \|y\|^2 = \|v - p\|^2 \iff \|y\|^2 = 0 \iff y = 0 \iff x = p.$$

Second, to show $v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i \perp \mathcal{S}$, we expand $\langle v-p, x\rangle$: note that $x \in \mathcal{S} \implies x = \sum_{i=1}^m a_i e_i$, for some scalars a_i. Thus,

$$\langle v-p, x\rangle = \left\langle v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, x \right\rangle = \left\langle v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, \sum_{j=1}^n a_j e_j \right\rangle = \sum_{j=1}^n \overline{a_j} \left\langle v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, e_j \right\rangle.$$

Now, see that

$$\begin{aligned}\left\langle v - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, e_j \right\rangle &= \langle v, e_j\rangle - \left\langle \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} e_i, e_j \right\rangle \\ &= \langle v, e_j\rangle - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} \langle e_i, e_j\rangle \\ &= \langle v, e_j\rangle - \sum_{i=1}^n \frac{\langle e_i, v\rangle}{\langle e_i, e_i\rangle} \delta_{ij} \langle e_i, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0.\end{aligned}$$

Hence $\langle v-p, x\rangle = 0$ for all $x \in \mathcal{S}$, which completes the proof. ■

As a follow-up, we look at two examples that highlight one of the central themes in math: *with generalization through abstraction comes power*. In fact, the second example sets the stage for the main content of this paper.

Example. Consider the inner product space $L^2([-1,1])$ as defined above. We'll attempt to approximate arbitrary functions using polynomials: first we'll project a $f \in L^2([-1,1])$ onto $\mathcal{P}_1 = \text{span}(1, x)$, the set of degree one polynomials, then onto $\mathcal{P}_2 = \text{span}(1, x, x^2)$, the degree two polynomials, and $\mathcal{P}_3 = \text{span}(1, x, x^2, x^3)$, continuing on, using polynomials with an ever increasing degree to build a sequence of polynomials—somehow, perhaps intuitively, one feels that the projection of f onto $\mathcal{P}_\infty$ will be f itself. Before we get started though, note that Equation 2.3 requires an orthogonal basis for the $\mathcal{P}_n$. Here, we use the *Gram-Schmidt algorithm* to orthogonalize the already existing basis $\{1, x, \cdots, x^n\}$, which is carried out in Table 1 for $n = 10$.

Original Basis Element	**Orthogonal Basis Element**
x^0	1
x^1	x
x^2	$x^2 - \frac{1}{3}$
x^3	$x^3 - \frac{3}{5}x$
x^4	$x^4 - \frac{3}{7}x^2 + \frac{3}{35}$
x^5	$x^5 - \frac{5}{9}x^3 + \frac{10}{63}x$
x^6	$x^6 - \frac{15}{11}x^4 + \frac{15}{77}x^2 - \frac{5}{231}$
x^7	$x^7 - \frac{7}{13}x^5 + \frac{35}{143}x^3 - \frac{21}{715}x$
x^8	$x^8 - \frac{4}{9}x^6 + \frac{12}{143}x^4 - \frac{12}{715}x^2 + \frac{1}{429}$
x^9	$x^9 - \frac{9}{17}x^7 + \frac{84}{323}x^5 - \frac{126}{2431}x^3 + \frac{9}{46189}x$
x^{10}	$x^{10} - \frac{5}{11}x^8 + \frac{15}{187}x^6 - \frac{45}{2431}x^4 + \frac{25}{46189}x^2 - \frac{5}{969969}$

Table 1. Orthogonal Polynomials using Gram-Schmidt Process on $\{1, x, x^2, \ldots, x^{10}\}$

These polynomials are known as the *Legendre polynomials*, and we use P_n to denote Legendre polynomial of degree n. With that out of the way, let's take $f(x) = e^x$, and get projecting! Invoking

Equation 2.3 we have,

$$p_n(x) = \text{proj}_{\mathcal{P}_n}(e^x) = \sum_{i=1}^{n} \frac{\int_{-1}^{1} P_i(x)e^x\, dx}{\int_{-1}^{1} P_i(x)P_i(x)\, dx} P_i(x).$$

The graphs of the resulting approximations from $n = 0$ to $n = 3$ are shown below.

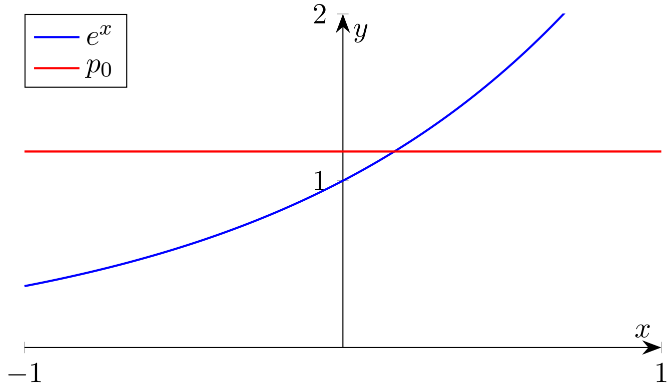

Figure 2. Approximation of e^x with $n = 0$.

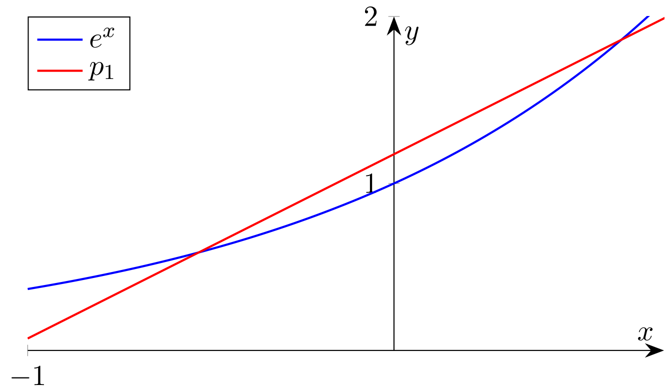

Figure 3. Approximation of e^x with $n = 1$.

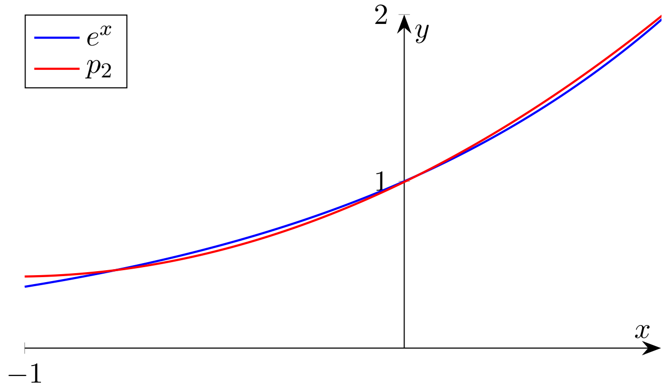

Figure 4. Approximation of e^x with $n = 2$.

Compare this with the corresponding Taylor series approximation for e^x, $T_2(x) = 1 + x + \frac{x^2}{2}$, as shown below.[4]

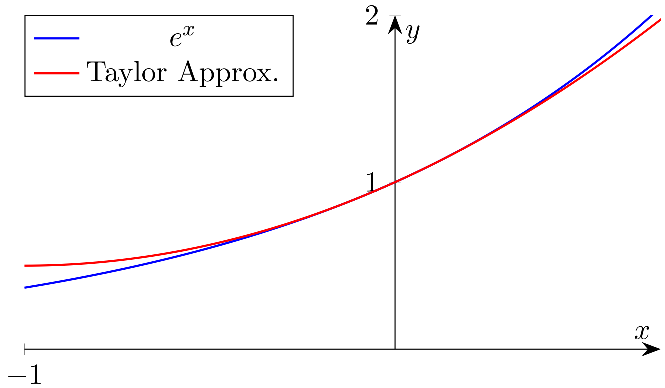

Figure 5. Degree 2 Taylor approximation of e^x.

[4]To make this more precise, one way would be to compare $\int_{-1}^{1} (e^x - p_2(x))\ dx$ and $\int_{-1}^{1} (e^x - T_2(x))\ dx$, the total area between the e^x and it's approximation.

In fact, by the time we get to $\text{proj}_{\mathcal{P}_3}(e^x)$, one can't even tell the difference between the two graphs!

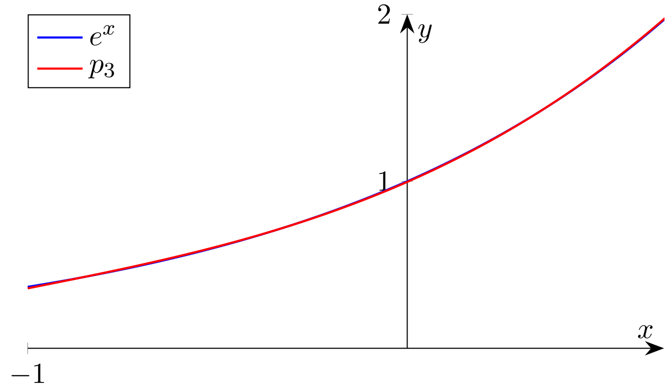

Figure 6. Approximation of e^x with $n = 3$.

Example. Let's turn the tables by approximating a polynomial, like $f(x) = x^2$, using *non-polynomial* functions—say using the fundamental cosines: $1, \cos x, \cos 2x, \cdots$. Before we dive into it, note that

$$\int_{-\pi}^{+\pi} \cos(nx)\cos(mx)\,dx = \pi\delta_{mn}$$

for all $m, n \in \mathbb{N}$, so $\{1, \cos nx : n \in \mathbb{N}\}$ is already an orthogonal set in $L^2([-\pi,\pi])$. Throwing in the appropriate normalizing factors, we get $\mathcal{C} = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos nx : n \in \mathbb{N}\right\}$ is an orthonormal set in $L([-\pi,\pi])$. Now, we'll project x^2 onto $\text{span}(\mathcal{C}_n)$ where

$$\mathcal{C}_n = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos mx : 1 \le m \le n\right\}$$

to get the n^{th} order approximation of x^2, which we denote by c_n. Again, we expect that as $n \to \infty$, we have $c_n \to x^2$. Using our projection formula,

$$\begin{aligned} c_n = \text{proj}_{\text{span}(\mathcal{C}_n)} &= \left\langle x^2, \frac{1}{\sqrt{2\pi}}\right\rangle \frac{1}{\sqrt{2\pi}} + \sum_{k=1}^{n} \left\langle x^2, \frac{1}{\sqrt{\pi}}\cos kx\right\rangle \frac{1}{\sqrt{\pi}}\cos kx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2\,dx + \frac{1}{\pi}\sum_{k=1}^{n}\int_{-\pi}^{\pi} x^2\cos(kx)\,dx\cos kx. \end{aligned}$$

Integrating by parts, $\int x^2\cos(kx)\,dx = \frac{x^2}{k}\sin(kx) + \frac{2x}{k^2}\cos(kx) - \frac{2}{k^3}\sin(kx) + C$. Computing the definite integral, $\int_{-\pi}^{\pi} x^2\cos(kx)\,dx = \frac{4\pi}{k^2}(-1)^k$. Picking off,

$$c_n = \frac{\pi^2}{3} + \frac{1}{\pi}\sum_{k=1}^{n}\frac{4\pi}{k^2}(-1)^k\cos kx = \frac{\pi^2}{3} + 4\sum_{k=1}^{n}\frac{1}{k^2}(-1)^k\cos kx.$$

Now for the burning question: do these c_n's approximate x^2 better as $n \to \infty$? Have a look at the graphs below.

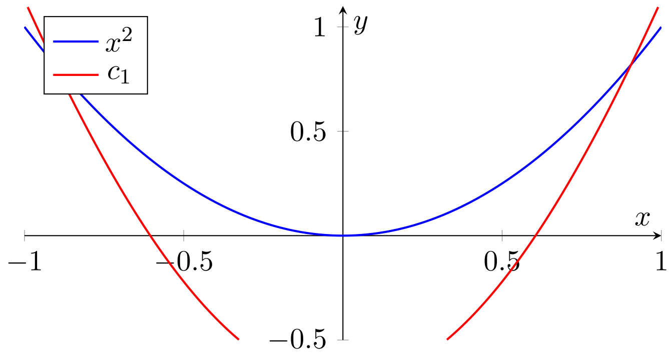

Figure 7. Approximation of x^2 with $n = 1$

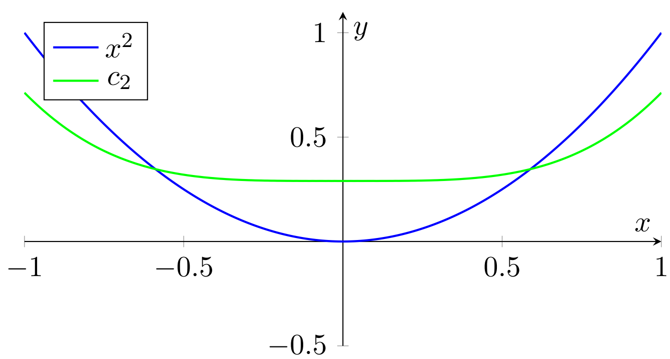

Figure 8. Approximation of x^2 with $n = 2$

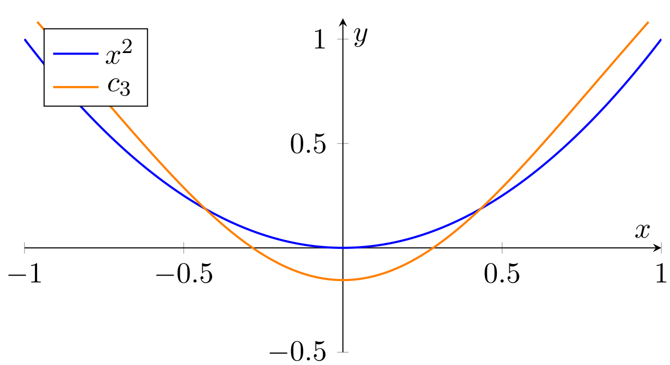

Figure 9. Approximation of x^2 with $n = 3$

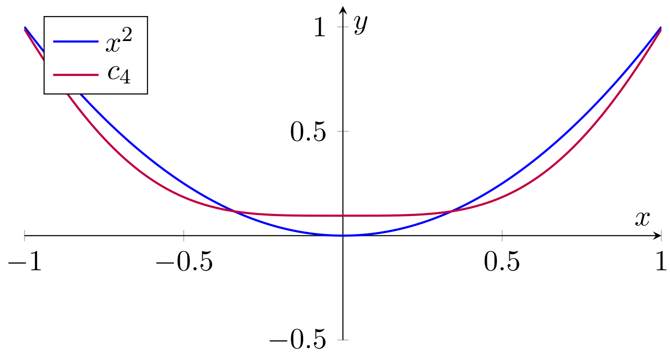

Figure 10. Approximation of x^2 with $n = 4$

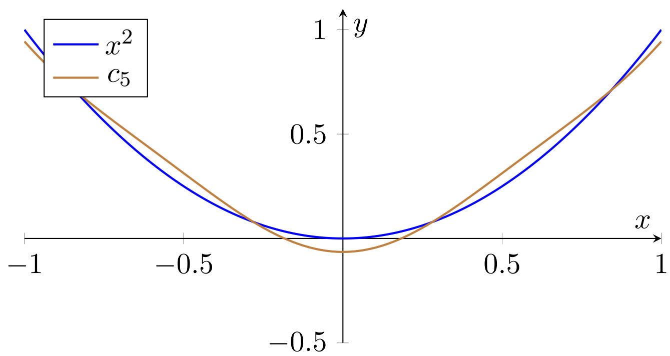

Figure 11. Approximation of x^2 with $n = 5$

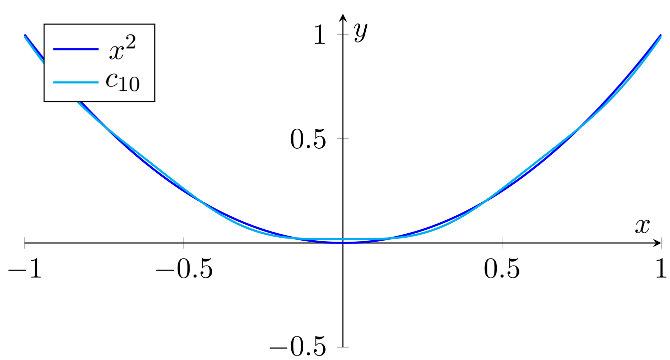

Figure 12. Approximation of x^2 with $n = 10$

Again, by the time we get to c_{10}, you can't tell the difference between x^2 and the approximating function! In fact, you may have recognized the expression for c_∞ as the Fourier series of x^2 from $-\pi$ to π!

Let's try approximating another function using $\mathcal{C}$, say $f(x) = x$. In this case, the n^{th} order approximation would be given by

$$c_n = \left\langle x, \frac{1}{\sqrt{2\pi}} \right\rangle + \frac{1}{\pi} \sum_{k=1}^{n} \langle x, \cos(kx) \rangle \cos(kx) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} x \, dx + \frac{1}{\pi} \sum_{k=1}^{n} \int_{-\pi}^{\pi} x \cos(kx) \, dx \cos(kx).$$

Before diving right in to bash out the integral $\int_{-\pi}^{\pi} x\cos(kx)\,dx$, note that $x\cos(kx)$ is an odd function *and* the domain of integration is symmetric about the origin. Hence the positive and negative areas on either side of the $y-$axis will completely cancel, meaning $\langle x, \cos(kx)\rangle = 0$ for all $n \in \mathbb{N}$. Reasoning identically, $\int_{-\pi}^{\pi} x\,dx = 0$ which means $\text{proj}_{\text{span}(\mathcal{C}_n)}(x) = 0$ for all $n \in \mathbb{N}$—that's definitely not what we wanted! $\mathcal{C}$ has failed!

This shows us that using $\mathcal{C}$ alone, it is not possible to approximate an arbitrary function. To approximate any function, we'll need the other trigonometric function, the sine.

Example. Just like the cosines, define

$$\mathcal{S} = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(nx) : n \in \mathbb{N}\right\}$$

which forms an orthonormal set in $L^2([-\pi,\pi])$. Again, we'll project x onto $\mathcal{S}_n = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(mx) : 1 \leq m \leq n\right\}$ to get s_n, the n^{th} order approximation of x.

$$\begin{aligned}\text{proj}_{\text{span}(\mathcal{S}_n)}(x) = s_n &= \left\langle x, \frac{1}{\sqrt{2\pi}}\right\rangle\frac{1}{\sqrt{2\pi}} + \sum_{k=1}^{n}\left\langle x, \frac{1}{\sqrt{\pi}}\sin kx\right\rangle\frac{1}{\sqrt{\pi}}\sin kx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} x\,dx + \frac{1}{\pi}\sum_{k=1}^{n}\int_{-\pi}^{\pi} x\sin(kx)\,dx\sin kx.\end{aligned}$$

Integrating by parts, $\int_{-\pi}^{\pi} x\sin(kx)\,dx = [-\frac{x}{k}\cos(kx) + \frac{1}{k^2}\sin(kx)]_{-\pi}^{+\pi} = -\frac{\pi}{k}\cos(k\pi) + \frac{2}{k^2}\sin(k\pi) = -\frac{2\pi(-1)^k}{k}$. Thus,

$$s_n = \sum_{k=1}^{n} -\frac{2(-1)^k}{k}\sin kx = 2\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\sin kx.$$

Use a graphing software to see what happens to s_n as $n \to \infty$!

Thus, the example of $f(x) = x$ tells us that we definitely require at least both the sines and cosines to approximate any function in $L^2([-\pi,\pi])$. To this end, define the orthornormal set $T \subseteq L^2([-\pi,\pi])$ as follows

$$\mathcal{T} := \mathcal{S} \cup \mathcal{C} = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos(nx), \frac{1}{\sqrt{\pi}}\sin(nx) : n \in \mathbb{N}\right\}$$

and

$$\mathcal{T}_N = \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos(nx), \frac{1}{\sqrt{\pi}}\sin(nx) : 1 \leq n \leq N\right\}.$$

The elements of $\mathcal{T}$ are called the *trigonometric polynomials.* Why?

That's all well, but the question still remains: how can we be sure that we can approximate *any* function using linear combinations of elements of $\mathcal{T}$?

This is a perfectly valid question, but it does turn out that any function *can* be approximated by $\mathcal{T}$ to an arbitrary degree of precision. That is, we have

Theorem 2.18. *The set of finite linear combinations of elements of $\mathcal{T}$ is dense in $L^2([-\pi,\pi])$ with respect to the metric induced by the norm $\|\cdot\|_{L^2}$.*

Remark 2.19. The proof of this theorem requires considerably more functional analysis backgrounds, so we omit it here. Moreover, it'll simply serve as a hurdle to the flow of this paper.

Recall that this is precisely the definition of a Hilbert basis, as defined below.

Definition 2.20. Let $\mathcal{H}$ be a Hilbert space. A *Hilbert basis* is a set $\{e_k\}_{k\in K} \subset \mathcal{H}$ such that

- $\{e_k\}_{k\in K}$ is an *orthonormal system.* That is, $\langle e_i, e_j\rangle = \delta_{ij}$ for all $i, j \in K$.

- $\{e_k\}_{k\in K}$ is *complete*. That is, the set of *finite* linear combinations of elements of $\{e_k\}_{k\in K}$ is dense in $\mathcal{H}$. In symbols, for every $v \in \mathcal{H}$ and $\varepsilon > 0$, there exists a finite subset $I \subseteq K$ and scalars α_i such that
$$\left\| v - \sum_{i\in I} \alpha_i e_i \right\| < \varepsilon.$$

At any rate, we would like to show that *given* that $\mathcal{T}$ is a Hilbert basis for $L^2([-\pi,\pi])$, the corresponding sequence of projections S_N, given by

$$S_N = \mathrm{proj}_{\mathrm{span}(\mathcal{T}_N)}(f) = A_0 + \sum_{n=1}^{N} A_n \cos nx + B_n \sin nx$$

where

- $A_0 = \frac{1}{\sqrt{2\pi}}\langle f, \frac{1}{\sqrt{2\pi}}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\, dx$,
- $A_n = \frac{1}{\sqrt{\pi}}\cos(nx)\langle f, \frac{1}{\sqrt{\pi}}\cos nx\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\, dx$ for all $n \in \mathbb{N}$,
- $B_n = \frac{1}{\sqrt{\pi}}\sin(nx)\langle f, \frac{1}{\sqrt{\pi}}\sin nx\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\, dx$ for all $n \in \mathbb{N}$,

converges to f in $L^2([-\pi,\pi])$. In general setting, this would take on the form of Proposition 2.21, which is what we prove.

Proposition 2.21. *Let $\{e_k\}_{k=1}^{\infty}$ be a countable Hilbert basis for the Hilbert space $\mathcal{H}$, and $v \in \mathcal{H}$. Then, the sequence $v_n = \mathrm{proj}_{\mathrm{span}(\{e_1,\cdots,e_n\})}(v) = \sum_{i=1}^{n}\langle v, e_i\rangle e_i$ converges with respect to the norm on $\mathcal{H}$ to v.*

Proof. We show that the sequence $S_n = \|v - \sum_{i=1}^{n}\langle v, e_i \rangle e_i\| \to 0$ as $n \to \infty$. Since $\{e_k\}_{k\in K}$ is complete in $\mathcal{H}$, given an $\varepsilon > 0$, there exists an $n \in \mathbb{N}$ and scalars $a_1, \cdots, a_n$ such that $\|v - \sum_{i=1}^{n} a_i e_i\|^2 < \varepsilon$. Now, if we can show that $\|v - \sum_{i=1}^{n} a_i e_i\|^2 \geq \|v - \sum_{i=1}^{n}\langle v, e_i\rangle e_i\|^2$ then we would have $\varepsilon > \|v - \sum_{i=1}^{n}\langle v, e_i\rangle e_i\|^2$, meaning $S_n^2 \to 0$ as $n \to \infty$ and since each S_n is non-negative, we must have that $S_n \to 0$ as $n \to \infty$. Thus, we show that $\|v - \sum_{i=1}^{n} a_i e_i\|^2 \geq \|v - \sum_{i=1}^{n}\langle v, e_i\rangle e_i\|^2$. First,

$$\begin{aligned}\left\| v - \sum_{i=1}^{n} a_i e_i \right\|^2 &= \langle v, v\rangle - \left\langle v, \sum_{i=1}^{n} a_i e_i \right\rangle - \left\langle \sum_{i=1}^{n} a_i e_i, v \right\rangle + \left\langle \sum_{i=1}^{n} a_i e_i, \sum_{i=1}^{n} a_i e_i \right\rangle \\ &= \|v\|^2 - \sum_{i=1}^{n} \overline{a_i}\langle v, e_i\rangle - \sum_{i=1}^{n} a_i \overline{\langle v, e_i\rangle} + \sum_{i=1}^{n} |a_i|^2.\end{aligned}$$

Next,

$$\begin{aligned}\left\| v - \sum_{i=1}^{n}\langle v, e_i\rangle e_i \right\|^2 &= \langle v, v\rangle - \left\langle v, \sum_{i=1}^{n}\langle v, e_i\rangle e_i \right\rangle - \left\langle \sum_{i=1}^{n}\langle v, e_i\rangle e_i, v \right\rangle + \left\langle \sum_{i=1}^{n}\langle v, e_i\rangle e_i, \sum_{i=1}^{n}\langle v, e_i\rangle e_i \right\rangle \\ &= \|v\|^2 - \sum_{i=1}^{n} \overline{\langle v, e_i\rangle}\langle v, e_i\rangle - \sum_{i=1}^{n}\langle v, e_i\rangle\langle e_i, v\rangle + \sum_{i=1}^{n} |\langle v, e_i\rangle|^2 \\ &= \|v\|^2 - \sum_{i=1}^{n} \overline{\langle v, e_i\rangle}\langle v, e_i\rangle - \sum_{i=1}^{n}\langle v, e_i\rangle\overline{\langle v, e_i\rangle} + \sum_{i=1}^{n} |\langle v, e_i\rangle|^2 \\ &= \|v\|^2 - \sum_{i=1}^{n} \overline{\langle v, e_i\rangle}\langle v, e_i\rangle.\end{aligned}$$

Finally, putting the two together,

$$\begin{aligned}\left\|v-\sum_{i=1}^{n}a_ie_i\right\|^2-\left\|v-\sum_{i=1}^{n}\langle v,e_i\rangle e_i\right\|^2 &= \sum_{i=1}^{n}|\langle v,e_i\rangle|^2-\sum_{i=1}^{n}\overline{a_i}\langle v,e_i\rangle-\sum_{i=1}^{n}a_i\overline{\langle v,e_i\rangle}+\sum_{i=1}^{n}|a_i|^2\\ &= \sum_{i=1}^{n}\overline{\langle v,e_i\rangle}\langle v,e_i\rangle-\overline{a_i}\langle v,e_i\rangle-a_i\overline{\langle v,e_i\rangle}+a_i\overline{a_i}\\ &= \sum_{i=1}^{n}\overline{\langle v,e_i\rangle}\big(\langle v,e_i\rangle-a_i\big)-\overline{a_i}\big(\langle v,e_i\rangle-a_i\big)\\ &= \sum_{i=1}^{n}|\langle v,e_i\rangle-a_i|^2\geq 0.\end{aligned}$$

■

Corollary 2.22. *Given a $f\in L^2([-\pi,\pi])$, the sequence of functions $(S_n)_{n\in\mathbb{N}}$ converges to f in the L^2 sense. That is,*

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}[S_n(x)-f(x)]^2\,dx=0.$$

Remark 2.23. Proposition 2.21 justifies the use of the word *basis* for what we call a *Hilbert basis*—we can write any $v\in\mathcal{H}$ as $\sum_{k=1}^{\infty}\langle e_k,v\rangle e_k$, which is exactly how what we would do with a *normal* orthonormal basis. In fact, another way to view the Fourier series is as a *change of basis transformation* in $L^2([-\pi,\pi])$. When we define a function $f\in L^2([-\pi,\pi])$ in the standard sense, we're actually writing down its coordinates with respect to the *useless* basis $B=\{\varphi_i\}_{i\in[-\pi,\pi]}$, where $\varphi_i(x)=\delta_{ix}$ for all $x\in[-\pi,\pi]$. Then, computing its Fourier series amounts to computing the coordinates of f with respect to the basis $\mathcal{T}\subset L^2([-\pi,\pi])$.

Explicitly,

$$\{f(i)\}_{i\in[-\pi,\pi]}\to\{\langle f,e_i\rangle\}_{e_i\in\mathcal{T}}.$$

While the value $f(i)$ tells us the contribution of φ_i to f, which is practically useless in this context, $\langle f,e_i\rangle$ quantifies the contribution of the periodic function e_i to f. This is nontrivial information, as it essentially decomposes the function f into its simplest frequency components. Note that the fact the Fourier series converges to f allows us to reconstruct the function f from it's simplest frequency components, $\langle f,e_i\rangle$, simply by summing them up after suitable multiplication.

Remark 2.24. This decomposing business makes more sense if we consider a $2\pi-$periodic function $f:\mathbb{R}\to\mathbb{R}$. In this special case, it's Fourier series on $[-\pi,\pi]$ is capable of describing f *completely throughout all of* $\mathbb{R}$: the Fourier series of f on any interval of the form $[n\pi,n\pi+2\pi]$ where $n\in\mathbb{Z}$ is exactly the same as it's Fourier series over $[-\pi,\pi]$! Then, it makes sense to say that $\sin(nx)$ and $\cos(nx)$ are indeed the simplest type of $2\pi-$periodic functions (try graphing them!), and hence we're decomposing an arbitrary $2\pi-$periodic function into a linear combination of the $\cos(nx)$'s and $\sin(nx)$'s! However, as already shown, we can extend our analysis to any $f\in L^2[-\pi,\pi]$.

Remark 2.25. Before we wrap up our analysis in this setting, I would like to transform the expression for the Fourier series of a function in a form that is more commonplace and compact.

First, using Euler's formula, we can rewrite $\cos(nx)=\frac{e^{inx}+e^{-inx}}{2}$ and $\sin(nx)=\frac{e^{inx}-e^{-inx}}{2i}$. Substituting this into the expression for S_N, we get

$$S_N(x)=A_0+\sum_{n=1}^{N}\left(A_n\frac{e^{inx}+e^{-inx}}{2}+B_n\frac{e^{inx}-e^{-inx}}{2i}\right)=A_0+\sum_{n=1}^{N}\left(\frac{A_n}{2}+\frac{B_n}{2i}\right)e^{inx}+\left(\frac{A_n}{2}-\frac{B_n}{2i}\right)e^{-inx}.$$

Now, we define C_n as

$$C_n = \frac{A_n}{2} + \frac{B_n}{2i} = \frac{1}{2}\Big(A_n - iB_n\Big),$$

and so

$$C_{-n} = \frac{A_n}{2} - \frac{B_n}{2i} = \frac{1}{2}\Big(A_n + iB_n\Big).$$

Hence, S_N can be written as

$$S_N(x) = A_0 + \sum_{n=1}^{N} C_n e^{inx} + C_{-n}e^{-inx} = \sum_{n=-N}^{N} C_n e^{inx}$$

where

$$C_n = \frac{1}{2}\left(A_n - iB_n\right) = \frac{1}{2}\left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,dx - i\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,dx\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx.$$

Proposition 2.26. *Let $f \in L^2([-\pi,\pi])$. Then the Fourier series to N terms of f can also be defined as*

$$S_N(x) = \sum_{n=-N}^{N} \widehat{f}(n)e^{inx} \text{ where } \widehat{f}(n) = C_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx \text{ for all } n \in \mathbb{Z}.$$

Remark 2.27. It is easy to generalize the formula mentioned above to compute the Fourier series of a function defined over *any* an arbitrary interval $[-a,a]$. First squeeze (or stretch!) the function to fit it in $[-\pi,\pi]$. Then, find the Fourier series of the modified function, and stretch (or squeeze) the Fourier series to fit it back into $[-a,+a]$. This is carried out explicitly below:

- The transformed function $g : [-\pi,\pi] \to \mathbb{R}$ is given by $g(x) = f(\frac{xa}{\pi})$. Note that $g(\pi) = f(a)$ and $g(-\pi) = f(-a)$, so the it makes sense.
- Then computing the Fourier coefficients C_n of g:

$$\begin{aligned} C_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} g(x)e^{-inx}\,dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f\Big(\frac{xa}{\pi}\Big)e^{-inx}\,dx \\ &= \frac{1}{2\pi}\int_{-a}^{a} f(u)e^{-\frac{i\pi nx}{a}}\frac{\pi}{a}\,du \\ &= \frac{1}{2a}\int_{-a}^{a} f(u)e^{-\frac{i\pi nx}{a}}\,du. \end{aligned}$$

  Where in the second line we made the substitution $u = x\frac{a}{\pi}$ to clear things up a bit. The Fourier series of g becomes: $S'_N(x) = \sum_{n=-N}^{N} C_n e^{-inx}$ where $C_n = \frac{1}{2a}\int_{-a}^{a} f(u)e^{\frac{-i\pi nx}{a}}\,du$.
- Finally, we de-transform (that is, squeeze or stretch) the Fourier series of g by replacing the x with an $\frac{\pi}{a}x$ to get S_N, to get the Fourier series of f over $[-a,a]$:

$$S_N(x) = \sum_{n=-N}^{N} \widehat{f}(n)e^{\frac{i\pi nx}{a}} \text{ where } \widehat{f}(n) = \frac{1}{2a}\int_{-a}^{a} f(x)e^{-\frac{i\pi nx}{a}}\,dx \text{ for all } n \in \mathbb{Z}.$$

Remark 2.28. Proposition 2.26 hints that $E = \left\{\frac{1}{\sqrt{2\pi}}e^{inx} : n \in \mathbb{Z}\right\}$ is a Hilbert basis for the complex version of $L^2([-\pi,\pi])$ —which can be precisely described as the Hilbert space of complex valued integrable functions $f : [-\pi,\pi] \to \mathbb{C}$ such that $\int_{-\pi}^{\pi} |f(x)|^2\,dx < \infty$ with the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx.$$

Again, this is true, but proving the completeness of E requires much more machinery.

3. Generalizing Fourier Series to Groups

After looking at the Fourier series of a function $f : [-\pi, \pi] \to \mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, we turn to the central question of this paper.

Question 3.1. *Is it possible to construct an analogue of the Fourier series of a function $f : G \to \mathbb{C}$ where G is any arbitrary group?*

Note that G could be anything: right from the humble $\mathbb{Z}/n\mathbb{Z}$, to $U(n)$, the group of n by n unitary matrices with complex entries, or even S_4, the symmetric group of order n. Essentially, we would like to break down an f, which might be dauntingly complicated at the moment, into simpler, more manageable parts, just as we decomposed a (periodic) function into a series of trigonometric polynomials, which represented the simplest type of periodic functions.

Before we begin, I'm sorry to disappoint you, but I must mention that we'll only be considering finite groups G in this paper. Even then, we require a good deal of representation theory, which we'll cover here. First, we'll start by defining where the functions $f : G \to \mathbb{C}$ live.

3.1. **The Group Algebra.** The set of all functions $f : G \to \mathbb{C}$ forms a vector space with addition and scalar multiplication given point-wise as usual. Taking inspiration from the inner products on $\mathbb{C}^n$ and $L([-\pi, \pi])$,

$$\langle v, v' \rangle = \sum_{i=1}^{n} v_i \overline{v'_i} \text{ for all } v, v' \in \mathbb{C}^n \qquad \langle f_1, f_2 \rangle = \int_{-\pi}^{\pi} f_1(x)\overline{f_2}\, dx \text{ for all } f_1, f_2 \in L^2([-\pi, \pi]),$$

we define for all $f_1, f_2 : G \to \mathbb{C}$,

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

We threw in a factor of $\frac{1}{|G|}$ because results in some neat cancellations, and so formula simplifications, as we'll see later.

Definition 3.2. The inner product space of functions $f : G \to \mathbb{C}$ as defined above is called the group algebra of G and is denoted by $\mathbb{C}[G]$.

Remark 3.3. The previous definition raises the question: why did we refer $\mathbb{C}[G]$ as the group algebra? Answering this requires a short detour, which we'll take as it will be important to us later.

In general, an *algebra* is a structure that is *both a ring and a vector space at the same time.* Stated more precisely,

Definition 3.4. A $k-$algebra is set equipped with operations $+ : A \times A \to A$, $\times : A \times A \to A$ and $\cdot : k \times A \to$ such that $(A, +, \cdot)$ forms a $k-$vector space and $(A, +, \times)$ a ring. Also, the two *multiplications*, $\cdot$ and $\times$, must be compatible with each other, that is, $\alpha(ab) = (\alpha a)b = a(\alpha b)$ for all $\alpha \in k$ and $a, b \in A$.

Algebras are all over the place: $\mathbb{C}$ is an $\mathbb{R}-$algebra, $\mathbb{F}[x]$ a $\mathbb{F}-$algebra, $\text{End}(V)$ where V is an $\mathbb{F}-$vector space an $\mathbb{F}-$algebra. To create an algebra from a group G, observe that one can 'encode', or map, a function $f : G \to \mathbb{C}$ as the summation $\sum_{g \in G} f(g)g$. Where does this summation live? Observe that it looks like a linear combination of elements of G, so it's an element of the vector space V over the field $\mathbb{C}$ with basis G. In symbols,

$$V = \left\{ \sum_{g \in G} a_g g : a_g \in \mathbb{C} \right\}.$$

The operations are defined as you would expect,

- **Vector addition:** $\sum_{g\in G} a_g g + \sum_{g\in G} b_g g = \sum_{g\in G}(a_g + b_g)g$,
- **Scalar multiplication:** $\alpha \cdot \left(\sum_{g\in G} a_g\right) = \sum_{g\in G}(\alpha a_g)g$ for all $\alpha \in \mathbb{C}$.

Constructed like this, it isn't too hard to define a multiplication operation on V—just use the distributive property;

$$\left(\sum_{g\in G} a_g g\right)\left(\sum_{h\in G} b_h h\right) = \sum_{g\in G}\sum_{h\in G} a_g b_h gh = \sum_{(g,h)\in G\times G} a_g b_h gh = \sum_{g'\in G}\left(\sum_{\substack{(x,y)\in G\times G \\ xy=g'}} a_x b_y g'\right),$$

meaning we have ourselves $\mathbb{C}-$algebra, called the *group algebra* over $\mathbb{C}$.

Now, we'll cover the required representation theory.

3.2. An Algebraic Interlude: Representation Theory.

3.2.1. *What is representation theory?* Roughly speaking, *representation theory* lies at the intersection of the two cornerstones of modern algebra: group theory, the study of symmetry and linear algebra, the study of maps that deform (vector) spaces in a certain, 'clean way'. More precisely, given a group, we hope to understand it better by studying it's interaction with a vector space.

Definition 3.5. A *representation* of a group G is an ordered pair (ρ, V), where V is a vector space and $\rho : G \to \mathrm{GL}(V)$ is a homomorphism of groups.

First, we scrutinize this definition, as is virtually the foundation for all the theory that follows. $\mathrm{GL}(V)$ is the general linear group of V, the set of invertible $\dim V$ by $\dim V$ matrices which forms a group under the usual matrix multiplication. Remember from linear algebra that these matrices describe invertible linear maps from V to itself (like a rotation of the plane), so here we are viewing a $g \in G$ as *acting on a vector space* via an invertible linear map, namely, $\rho(g)$.

With that out of the way, let us jump into some examples!

Example. Consider $G = \mathbb{Z}_n$, the cyclic group of order n. We could interpret the generator as a rotation by $2\pi/n$ about the origin, meaning G consists of the rotational symmetries of an $n-$gon, so it's reasonable to let $V = \mathbb{R}^2$, the plane, and

$$\rho(k) = \begin{bmatrix} \cos\frac{2\pi k}{n} & -\sin\frac{2\pi k}{n} \\ \sin\frac{2\pi k}{n} & \cos\frac{2\pi k}{n} \end{bmatrix}$$

for all $k \in \mathbb{Z}_n$, which is the matrix corresponding to the rotation.

Example. But who restricted us to rotations? There are *many* other ways to interpret the generator of $\mathbb{Z}_n$, like a representing *reflection* by a line tilted $2\pi/n$ from the positive $x-$axis. Then, $V = \mathbb{R}^2$ and

$$\rho(k) = \begin{bmatrix} \cos\frac{4\pi k}{n} & \sin\frac{4\pi k}{n} \\ \sin\frac{4\pi k}{n} & -\cos\frac{4\pi k}{n} \end{bmatrix}$$

for all $k \in \mathbb{Z}_n$.

Example. Stepping away form $\mathbb{Z}_n$, let $G = S_3$, the symmetric group on three elements. Representing this is quite straightforward: simply use the three by three permutation matrices! For concreteness, we have $V = \mathbb{R}^3$ and

$$\rho\Big((1)(2)(3)\Big) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \rho\Big((1)(23)\Big) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \rho\Big((13)(2)\Big) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\rho\Big((12)(3)\Big) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \rho\Big((123)\Big) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \rho\Big((132)\Big) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Notation. Note that $\rho(g)$ is used to denote the linear map $\rho(g) : V \to V$ and $\rho(g)(v)$ to the image of $v \in V$ under the map $\rho(g)$. Somewhat annoyingly, we will use only ρ or only V to denote the representation (ρ, V), whenever the other is clear from context.

Just like we can take the direct sum of two vector spaces and even groups, we can also do the same for representations.

Definition 3.6. Let (ρ_1, V_1) and (ρ_2, V_2) be representations of a group G. The direct sum of the two representations, denoted by $(\rho_1 \oplus \rho_2, V_1 \oplus V_2)$ is defined to be $(\rho_1 \oplus \rho_2)(g) = \big(\rho_1(g), \rho_2(g)\big) \in \mathrm{GL}(V_1 \oplus V_2)$ for all $g \in G$.

Remark 3.7. The matrix representation of the linear map $(\rho_1 \oplus \rho_2)(g)$ for a $g \in G$ is given by the block-diagonal matrix

$$\begin{bmatrix} \rho_1(g) & \mathbf{0} \\ \mathbf{0} & \rho_2(g) \end{bmatrix}.$$

A notion that will be extremely useful in what follows is the *dimension* of a representation.

Definition 3.8. The *dimension* of a representation $\rho : G \to \mathrm{GL}(V)$, is the dimension of the underlying vector space V.

3.2.2. *Decomposing Representations.* In this section, we study the structure of a representation in the hope to break it into simpler, more manageable parts. Just like in linear algebra we have *subspaces* and in group theory *subgroups*, we also have such a sub-structure for representations, called a *subrepresentation* (surprise, surprise!).

How would it look like? It should be a representation of a group G in its own right, while built out of a given representation (ρ, V). Thus, we can simply restrict the domain of $\rho(g) : V \to V$ for all $g \in G$ to some fixed vector subspace W of V. But are we guaranteed that $\rho(g)|_W \in \mathrm{GL}(W)$ for all $g \in G$? We are if and only if W is $G-$invariant—vectors in W should not be thrown out of it by $\rho(g)$.

Definition 3.9. Let (ρ, V) be a representation of a group G. A linear subspace W of V is said to be $G-$invariant if $\rho(g)(w) \in W$ for all $g \in G$ and $w \in W$.

Now we are ready to define the subrepresentation.

Definition 3.10. A *subrepresentation* of a representation (ρ, V) of a group G, is a $G-$invariant subspace $W \subseteq V$, along with the homomorphism obtained by restricting the domain of $\rho(g)$ to W for all $g \in G$.

With this, the concepts of reducible and irreducible representations naturally come up.

Definition 3.11. A representation (ρ, V) of a group G is said to be *irreducible* if its only subrepresentations are the trivial representation $W = \{0\}$, and itself. A subrepresentation is said to be *reducible* if it is not irreducible.

The first thing that comes to mind when one reads Definition 3.11 are prime numbers! In fact, we have Theorem 3.12 that tells us these irreducibles are indeed the building blocks, the atoms of representation theory [5].

[5] These type of theorems crop up a lot in different areas of algebra! For example in group theory, we have that any finite abelian group is isomorphic to a direct product of cyclic groups of prime power order (the 'atoms'), albeit this version is easier to prove.

Theorem 3.12. *Let V be a representation of a group G. Then either V is irreducible, or a direct sum of irreducible representations of G.*

How do we go about proving this? Recall again from linear algebra that given a subspace W of an inner product space V, we can construct it's orthogonal complement $W^\perp$ which is such that $V = W \oplus W^\perp$—a sort of decomposition of V. Now, if we can show that *if* $W \subseteq V$ is a subrepresentation of V *then* so is $W^\perp$, we should be able to complete the proof (with a dash of induction). To do so, we start with a definition.

Definition 3.13. Let V be a representation of a group G. An inner product $\langle \cdot, \cdot \rangle$ on V is said to be *unitary* if for all $g \in G$ and $v_1, v_2 \in V$ we have $\langle v_1, v_2 \rangle = \big\langle \rho(g)(v_1), \rho(g)(v_2) \big\rangle$.

This terminology is borrowed from linear algebra. A unitary linear map from a vector space to itself, say A, is one that preserves distances and angles (comprising only of rotations, both proper and improper), meaning the generalized dot product of two vectors doesn't change, which can be written as $\langle x, y \rangle = \langle Ax, Ay \rangle$ for all x and y in V.

Proposition 3.14. *Let W be a sub-representation of a representation V equipped with a unitary inner product $\langle \cdot, \cdot \rangle$ of a group G. Then $W^\perp$ is also a sub-representation of V.*

Proof. Assuming that W is $G-$invariant we must show that $W^\perp$ is also $G-$invariant. Start by considering $\langle \rho(g)(w'), w \rangle$ for any $g \in G$, $w' \in W^\perp$ and $w \in W$. Since the inner product is unitary we have that $\langle \rho(g)(w'), w \rangle = \rho(g^{-1}) \langle \rho(g)(w'), w \rangle = \langle w', \rho(g^{-1}) w \rangle = 0$, where the last equality follows since $\rho(g^{-1}) w \in W$. Thus, by the very definition of the orthogonal complement of a subspace, $\langle \rho(g)(w'), w \rangle = 0$ and $w \in W$ together imply that $\rho(g)(w') \in W^\perp$ which means that $W^\perp$ is $G-$invariant. Hence the restriction of ρ to $W^\perp$ is a sub-representation. ■

But hey—this only works out *if* the inner product is unitary! What if its not? Luckily for us, it turns out there is *nothing* exclusive about the gang of vector spaces that posses an unitary inner product. We can always build one up from a completely 'normal' inner product.

Proposition 3.15. *Let V be a representation of a group G. Given an arbitrary inner product $\langle \cdot, \cdot \rangle$ on V, it is possible to construct a unitary inner product on V, commonly denoted $\langle \cdot, \cdot \rangle_G$.*

Proof. Consider the inner product given by

$$\langle v_1, v_2 \rangle_G = \frac{1}{|G|} \sum_{g \in G} \big\langle \rho(g)(v_1), \rho(g)(v_2) \big\rangle.$$

It seems to be the 'averaged out' version of the original inner product. Checking that it is indeed an inner product is easy, and will just be a waste of space here, so we omit it. Next, we show that it's unitary, the important part

$$\begin{aligned} \big\langle \rho(h)(v_1), \rho(h)(v_2) \big\rangle_G &= \frac{1}{|G|} \sum_{g \in G} \Big\langle \rho(h)\big(\rho(g)(v_1)\big), \rho(h)\big(\rho(g)(v_2)\big) \Big\rangle \\ &= \frac{1}{|G|} \sum_{g \in G} \big\langle \rho(hg)(v_1), \rho(hg)(v_2) \big\rangle. \end{aligned}$$

Note that $g \mapsto gh$ is a bijection on G, so we can safely rewrite the expression as

$$\frac{1}{|G|} \sum_{g' \in G} \big\langle \rho(g')(v_1), \rho(g')(v_2) \big\rangle$$

which is, by definition, equal to $\langle v_1, v_2 \rangle_G$, completing the proof. ■

Now comes the time when we can compile all these small lemma's into the proof of Theorem 3.12, the main result of this section which is the analogue of the fundamental theorem of arithmetic for representations, as stated below.

Proof. As mentioned earlier, we induct on dim V.

- **Base case:** If dim $V = 1$, then there can't exist a proper subspace of V, implying V is irreducible.
- **Induction hypothesis:** Next, assume that the hypothesis holds for representations W such that dim $W < k$ where $k \in \mathbb{N}$. We have to prove the hypothesis for V such that dim $V = k$.
- **Inductive step:** Assume that V isn't irreducible, otherwise we'd be done. Then, there exists a sub-representation V', meaning we can apply Proposition 3.14 which says we have a sub-representation W' such that $V = W \oplus W'$. By the induction hypothesis both W and W' can be written as a direct product of irreducible representations (both their dimensions are less than k), which completes the proof.

■

Remark 3.16. Warning! About the proof of Theorem 3.12! We implicitly assumed that we can *always* equip a vector space with an inner product—though all of the vector spaces we'll encounter in this paper will be so as to say, 'productizable'.

Note that one might *feel that an irreducible representation for a group* is always one-dimensional, as they are the simplest of representations, but Example 3.2.2 tells us otherwise.

Example. Consider $G = S_3$. Recall the permutation representation of S_3 and note that $V_1 = \text{span}\{(1,1,1)\} \subset \mathbb{R}^3$ is a $G-$invariant subspace under the permutation representation. In fact, the permutation representation restricts to the identity on V_1, and hence is called the *trivial representation*. Maschke's theorem tells us that $V_2 = V_1^{\perp} = \text{span}\{(1,-1,0),(1,0,-1)\}$, which is two dimensional, must also be a $G-$invariant subspace and hence also a sub-representation, called the *standard representation*, of the permutation representation. It is left to the reader as an excersie to show that V_2 is indeed irreducible.

However, when G is abelian, we do indeed have that irreducible represtations are always one-dimensional. We defer it's proof to the next section, once we have the necessary tools.

3.2.3. *Maps Between Representations.* This part of the paper is best understood as a 'helper' section. We present theorems related to maps between representation spaces in quick succession, which while they might seem unmotivated or 'useless' at first, are of utmost importance. We begin with a definition of the type of maps we'll be interested in studying.

Definition 3.17. Let (ρ_V, V) and (ρ_W, W) be irreducible representations for a finite group G. A linear map, $\varphi : V \to W$, is said to be a $G-$equivariant map if for all $g \in G$ and $v \in V$, we have that $\varphi\big(\rho_V(g)(v)\big) = \rho_W(g)(\varphi(v))$. We denote the set of $G-$equivariant maps by $\text{Hom}_G(V, W)$.

The definition for an equivariant[6] map makes sense—the order of application of φ and ρ ought not to matter. In other words, it shouldn't matter what path you take to get to the bottom right W starting from the top left V, we say that the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ {\scriptstyle \rho_V(g)}\downarrow & & \downarrow{\scriptstyle \rho_W(g)} \\ V & \xrightarrow[\varphi]{} & W \end{array}$$

[6]We have a category! The objects are the representations of a fixed group G over a field $\mathbb{F}$, and the morphisms are the $G-$equivariant maps. Identity and composition are defined as usual. This category is denoted by $\text{Rep}_{\mathbb{F}}(G)$.

Why the notation Hom? That's because a *vector space homomorphism* is just a linear map between the two spaces. It might not be much of a surprise, but $\mathrm{Hom}_G(V_1, V_2)$ forms a vector space (with addition and multiplication given pointwise), so we can talk about it's bases and dimension. The definition of a $G-$equivariant map lets us talk about isomorphic representations.

Definition 3.18. Two representations (ρ_1, V_1) and (ρ_2, V_2) are said to be *isomorphic* if there exists a bijective $G-$equivariant map between the two.

Now, we straightaway state the result that will be important to us later, which is actually a corollary of *Schur's lemma.*

Proposition 3.19. *Let V_1 and V_2 be irreducible representations for a group G. Then,*

- *If $V_1 \ncong V_2$, then* $\dim_{\mathbb{F}} \mathrm{Hom}_G(V_1, V_2) = 0$.
- *If $V_1 \cong V_2$, then* $\dim_{\mathbb{F}} \mathrm{Hom}_G(V_1, V_2) = 1$.

Thus, one check the isomorphism of representations and the level of their underlying vector spaces. As mentioned, to prove Proposition 3.19 we will require *Schur's lemma*, which is split between the next couple of propositions.

Proposition 3.20 (Schur's Lemma Part 1)**.** *Given $\varphi \in \mathrm{Hom}_G(V, W)$, either $\varphi = 0$ or $\varphi =$ a vector space isomorphism.*

Proof. Assume that $\varphi \neq 0$. Then $\ker(\varphi)$ is a *proper* subset of V. Note that $\ker(\varphi)$ is a $G-$invariant subspace (exercise!), so the irreducibilty of V implies that $\ker(\varphi) = 0$, which shows that φ is injective. Next, using a similar argument, $\mathrm{im}(\varphi)$ cannot be the null set (otherwise $\varphi = 0$) and as $\mathrm{im}(\varphi)$ is a $G-$invariant subspace the irreducibility of W implies that $\mathrm{im}(\varphi) = W$, which shows φ is surjective. ■

Proposition 3.21 (Schur's Lemma Part 2)**.** *Given $\varphi \in \mathrm{Hom}_G(V, V)$, we have $\varphi = \lambda I$, a scalar multiple of the identity map.*

Proof. By Proposition 3.20, we have that $\varphi \in \mathrm{Aut}_G(V, V)$, that is, φ is an invertible, G-equivariant, linear map from V to itself. This, and the fact I used a 'λ' in the proposition statement should get you thinking about eigenvalues! In particular, since $\mathbb{F}$ is an algebraically closed field, φ will have at least one eigenvalue, say λ. The corresponding eigenspace is the subspace $W = \{v \in V : \varphi(v) = \lambda v\}$, which is G-invariant (a good exercise to verify!). Since this cannot be $\{0\}$, the only way for it to exist without contradicting the irreducibility of V is for $W = V$, which instantly completes the proof. ■

Notice how much of these proofs hinge on irreducibility of the representation space! Now we are ready to prove Proposition 3.19.

Proof. We prove the two parts as mentioned in the proposition.

- The first claim follows directly from Proposition 3.20: since V_1 is not isomorphic to V_2, the only homomorphism that can exist between the two is the 0 map.
- For the second part, consider two non-zero elements ϕ_1 and ϕ_2 of $\mathrm{Hom}_G(V_1, V_2)$. Note that $\phi_2^{-1}\phi_1$ is an element of $\mathrm{Hom}_G(V_1, V_1)$ by a proposition. Hence, by Proposition 3.21, $\phi_2^{-1}\phi_1 = \lambda I$ where I denotes the identity map from V_1 to itself and $\lambda \in \mathbb{F}$. Rearranging, we get $\phi_1 = \phi_2\lambda$, meaning all elements of $\mathrm{Hom}_G(V_1, V_2)$ are multiples of each other, which completes the proof.

■

Now, as promised, we return to the proof that an irreducible representation of an abelian group is always one-dimensional.

Proposition 3.22. *Let (ρ, V) be an irreducible representation of an abelian group G. Then $\dim V = 1$.*

Proof. Since $gh = hg$ for all $g, h \in G$, we have $\rho(gh) = \rho(hg) \implies \rho(g)\rho(h) = \rho(h)\rho(g)$. Written differently, $\rho(g)\big(\rho(h)(v)\big) = \rho(h)\big(\rho(g)(v)\big)$ for all $v \in V$ and $g, h \in G$. Does this ring a bell? Indeed, we have that $\rho(g)$ is a $G-$equivariant map! By Proposition 3.21 there exists a λ for all $g \in G$, such that $\rho(g) = \lambda I$. This implies that any subspace W of V is $G-$invariant, and so is a subrepresentation. At this stage, the irreducibility of V forces $\dim V = 1$. ∎

3.3. **Character Theory.** After embarking on a seemingly pointless, rambling journey in representation theory, we return to answer Question 3.1. First, a slight caveat: instead of considering an arbitrary $f \in \mathbb{C}[G]$, we will deal with class functions, for now, as defined below.

Definition 3.23. An function $f \in \mathbb{C}[G]$ is said to be a *class function* if $f(g) = f(hgh^{-1})$ for all $h, g \in G$. The set of class functions is denoted by $\mathcal{C}_G$.

Remark 3.24. Essentially, a class function is one that is constant over a conjugacy class —it does not discriminate between members of the same conjugacy class.

Remark 3.25. If G is an abelian group, say $\mathbb{Z}/n\mathbb{Z}$, the conjugacy classes are simply the singletons, so $\mathbb{C}[G] = \mathcal{C}_G$.

Definition 3.26. Let (ρ, V) be a representation of a group G. The *character* of the representation (ρ, V) is a function $\chi_V : G \to \mathbb{F}$ defined by $\chi_V(g) = \mathrm{Tr}\big(\rho(g)\big)$ for all $g \in G$. A character of an irreducible representation is said to be an *irreducible character*.

Proposition 3.27. *A character χ of a group G is a class function.*

Proof. Consider a representation (ρ, V) of a group G and let χ_V denote it's character. Then, $\chi_V(h^{-1}gh) = \mathrm{Tr}\big(\rho(h^{-1}gh)\big) = \mathrm{Tr}\big(\rho(h^{-1})\rho(g)\rho(h))\big)$. Since the trace is invariant under cyclical shifts, we get $\mathrm{Tr}\big(\rho(h^{-1})\rho(g)\rho(h))\big) = \mathrm{Tr}\big(\rho(g)\rho(h))\rho(h^{-1})\big) = \mathrm{Tr}\big(\rho(g)\rho(hh^{-1}))\big) = \mathrm{Tr}(\rho(g)) = \chi_V(g)$, which completes the proof. ∎

Now we present the central result of this whole paper! The one that allows us to generalize Fourier analysis beyond the realm of $\mathbb{R}$.

Theorem 3.28. *Let G be a finite group and $\mathcal{C}_G$ the vector space of complex-valued class functions defined on G. Then, the set of irreducible characters of G, denoted by $\mathcal{A} = \{\chi_\alpha\}_{\alpha \in A}$ forms an orthonormal basis for $\mathcal{C}_G$.*

But why is this result so important? As we shall see, these irreducible characters play the same role for $\mathbb{C}[G]$ as the exponential basis $E = \left\{\frac{1}{\sqrt{2\pi}}e^{inx} : n \in \mathbb{Z}\right\}$ does for $L^2([-\pi, \pi])$. Just like the density of E in $L^2([-\pi, \pi])$ allows us to decompose an $f : [-\pi, \pi] \to \mathbb{C}$ into a series of sines and cosines, the fact that $\mathcal{A}$, the set of irreducible characters of G, forms a basis for $\mathbb{C}[G]$ allows us to rewrite f in a simpler, more manageable way. To this end, we define the *Fourier series* of an $f \in \mathbb{C}[G]$.

Definition 3.29. Let $\{\chi_\alpha\}_{\alpha \in A}$ denote the set irreducible representations for a finite group G. The *Fourier coefficient* of $f \in \mathcal{C}_G$ corresponding to an irreducible character χ_α is defined by

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

The *Fourier series* of f is defined by

$$\sum_{\alpha \in A} \langle f, \chi \rangle \chi_\alpha(x).$$

Remark 3.30. Again, this is nothing but a *change of basis transformation*. Specifying an $f \in \mathcal{C}_G$ means writing down it's coordinates with respect to the 'trivial' basis $B = \{f_{\mathcal{O}_i} \in \mathcal{C}_G : i \in I\}$ where $\{\mathcal{O}_i\}_{i \in I}$ denotes the set of conjugacy classes of G, and $f_{\mathcal{O}_i}(g) = 1$ if $g \in \mathcal{O}_i$ and $f_{\mathcal{O}_i}(g) = 0$ if $g \notin \mathcal{O}_i$ for all $i \in I$, whereas $\{\langle f, \chi_\alpha \rangle\}_{\alpha \in A}$ are the coordinates of f with respect to $\mathcal{A}$, the 'character basis'. Each $\langle f, \chi_\alpha \rangle$ quantifies the contribution of χ_α to f, which effectively decomposes f into pieces. On the other hand, the *Fourier series* can be viewed as a *synthesis* formula, reconstructing f from it's Fourier coefficients.

Remark 3.31. Note that since B as defined above forms a basis for $\mathcal{C}_G$, Theorem 3.28 says that $|\mathcal{A}| = |B|$, or the number of distinct irreducible characters is equal to the number of conjugacy classes of G, which is finite for finite G.

That's great, but actually computing all the irreducible representations and then the corresponding characters looks daunting, at least at first glance. Luckily for us, if we restrict our attention to abelian groups, the irreducible characters can be completely described fairly easily.

Proposition 3.32. *An irreducible character of an abelian group G is a group homomorphism from G to $\mathbb{F}^\times$, the multiplicative group of the field $\mathbb{F}$.*

Proof. This proof relies heavily on Proposition 3.22: an irreducible representation (ρ, V), of an abelian group G is one-dimensional—$\rho(g)$ is nothing but a 1 by 1 matrix—essentially a scalar. Thus, quite obviously $\mathrm{Tr}\big(\rho(g)\big) = \rho(g)$ for all $g \in G$, which allows us to simplify $\chi_\rho(g_1 g_2)$:

$$\begin{aligned}
\chi_\rho(g_1 g_2) &= \mathrm{Tr}\big(\rho(g_1 g_2)\big) \\
&= \rho(g_1 g_2) \\
&= \rho(g_1)\rho(g_2) \\
&= \mathrm{Tr}\big(\rho(g_1)\big)\mathrm{Tr}\big(\rho(g_2)\big) \\
&= \chi_\rho(g_1)\chi_\rho(g_2).
\end{aligned}$$

Hence, each irreducible character is simply a group homomorphism to $\mathbb{F}^\times$—something that *seems* more manageable. ∎

Note that we can go the other way round as well: each group homomorphism $\phi : G \to \mathbb{F}^\times$ is an irreducible representation of G. To see this, simply note $\mathbb{F}^\times \cong \mathrm{Aut}(\mathbb{F})$, where $\mathbb{F}$ forms a one dimensional vector space over itself (essentially, linear maps from a field to itself can be represented by single numbers, which have the effect of squeezing/stretching the $\mathbb{F}$ number line). Hence, there is a one-to-one relationship between the two sets, so to talk of one is to talk of the other. With that out of the way, let's jump into an example!

Example. First, we'll have exactly N irreducible characters, and since $\mathbb{Z}/N\mathbb{Z}$ is a abelian, they'll be functions $\chi : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ such that $\chi(x+y) = \chi(x)\chi(y)$ for all x and y in $\mathbb{Z}/N\mathbb{Z}$—which naturally hints at the exponential function: we literally have $e^{a+b} = e^a e^b$ for real a and b, as we were taught so long ago. Hopefully then, it shouldn't be too surprising that $\chi_k(x) = e^{2\pi i k x/N}$ for $k = 0, \cdots, N-1$ are the irreducible characters for $\mathbb{Z}/N\mathbb{Z}$—this matches up with our experience in $\mathbb{R}$!

As before, the Fourier series of f will be given by

$$\sum_{\alpha \in A} \langle f, \chi \rangle \chi_\alpha(x),$$

which, in this case, becomes

$$\sum_{k=0}^{N-1} \langle f, e^{2\pi i k x/N} \rangle e^{2\pi i k x/N} = \sum_{k=0}^{N-1} X_i e^{2\pi i k x/N}$$

where

$$X_k = \frac{1}{|G|}\sum_{x\in\mathbb{Z}_N} f(x)\overline{\chi_k(x)} = \frac{1}{N}\sum_{n=0}^{N-1} f(n)e^{-2\pi ikn/N}.$$

Again, another way to think about it is simply as a change of basis linear transformation $\mathcal{F}:\mathbb{C}^N \to \mathbb{C}^N$ given by

$$\mathcal{F}(\mathbf{x}) = \mathcal{F}\big((x_0,\cdots,x_{N-1})\big) = (X_0,\cdots,X_{N-1}) \text{ where } X_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-2\pi ink/N}.$$

This, as you might already have recognized, is the Discrete Fourier Transform[7] of $\mathbf{x}$ —just one of the many, *somewhat identical*, offsprings of Definition 3.29!

Remark 3.33. Even though we can't apply this machinery when $G=\mathbb{R}/\mathbb{Z}$ (as $|G|=\infty$) to re-derrive the Fourier series of a $f\in L^2([-1,1])$, we can be cheeky and use the above analysis as $N\to\infty$. More precisely, the idea is to pick an $N\in\mathbb{N}$ and sample the function f at N points: $x_k=\frac{2k}{N-1}-1$ where k ranges from 0 to $N-1$. Thus, for each $N\in\mathbb{N}$, we have a function, $f_N:\mathbb{Z}/N\mathbb{Z}\to\mathbb{C}$ defined by $f_N(k)=f(x_k)$ for all $k\in\mathbb{Z}/N\mathbb{Z}$. It's quite clear that as $N\to\infty$, f_N approximates f better and better. Next, the Fourier series of f_N will be given by

$$\sum_{k=0}^{N-1}\langle f_N, e^{2\pi ikx/N}\rangle e^{2\pi ikx/N} \quad \text{where} \quad \langle f_N, e^{2\pi ikx/N}\rangle = \frac{1}{N}\sum_{n=0}^{N-1} f\left(\frac{2n}{N-1}-1\right)e^{-2\pi ikn/N}.$$

Try seeing what happens now when $N\to\infty$!

3.4. **Proof of Theorem 3.28.** As the title suggests, in this section, we prove Theorem 3.28. Ordinarily, it's proof would be broken into two parts: proving the orthogonality of the irreducible characters, and then showing that the number of irreducible characters is equal to the number of conjugacy classes of G, which would show that $\mathcal{A}$ forms a basis for $\mathcal{C}_G$, as B already does. While we do provide two distinct proofs for the first part, we do not show the second part, as it is beyond the scope of this paper.

3.4.1. *The First Proof.* For the first proof, observe that we have a 1 and a 0 in Theorem 3.28, which are hidden in the δ_{ij} in $\langle\chi_i,\chi_j\rangle=\delta_{ij}$, and we also have a 1 and a 0 in Theorem 3.19—and since it can be easily shown that isomorphic representations induce the same character, it suffices to show that $\langle\chi_V,\chi_W\rangle=\dim_{\mathbb{F}}\operatorname{Hom}_G(W,V)$ for irreducible representations V and W of G!

Theorem 3.34. *Let V and W be irreducible representations of a group G. Then $\langle\chi_V,\chi_W\rangle = \dim_{\mathbb{F}}\operatorname{Hom}_G(W,V)$.*

Proving this, however, requires a few preliminaries: namely, the *dual* and the *tensor* product of a representation, both of which might be familiar from linear algebra, where they're used to construct new vector spaces from old ones.

Definition 3.35 (Dual of a Representation)**.** Let (ρ,V) be a representation of a group G. The *dual representation* ρ is the representation (ρ^*,V^*) where V^* is the dual of the vector space V and $\rho^*:G\to\mathrm{GL}(V^*)$ defined by $\rho^*(g)(\phi)(v)=\phi(\rho(g^{-1})(v))$ for all $g\in G$ and $v\in V$.

Definition 3.36 (Tensor Product of Representations)**.** Let (ρ_1,V_1) and (ρ_2,V_2) be representations of a group G. Then, their *tensor product representation* is the representation $(\rho_1\otimes\rho_2,V_1\otimes V_2)$ where $V_1\otimes V_2$ is the usual tensor product of V_1 and V_2 and $(\rho_1\otimes\rho_2)(g)(v_1\otimes v_2)=\rho_1(g)(v_1)\otimes\rho_2(g)(v_2)$ for all $g\in G$, $v_1\in V_1$ and $v_2\in V_2$.

[7]Well, almost. In the actual discrete Fourier transform, that 'pesky' factor of $1/N$ is not present.

Remark 3.37. While this is only defined for *elementary tensors*, which are elements of $V \otimes W$ of the form $v \otimes w$ where $v \in V$ and $w \in W$, $\rho_{V \otimes W^*}$ automatically extends to *all* tensors since any element of $V \otimes W$ can be expressed as the sum of elementary tensors, $\sum_i v_i \otimes w_i$.

The existence of the following proposition is the very reason we talked about these two constructions in the first place; try expanding the expression for $\langle \chi_V, \chi_W \rangle$ to see why.

Proposition 3.38. *Let* (ρ, V), (ρ_1, V_1) *and* (ρ_2, V_2) *be representations of a group* G*. Then,* $\chi_{V^*} = \overline{\chi_V}$ *and* $\chi_{V_1 \otimes V_2} = \chi_{V_1}\chi_{V_2}$.

Proof. To prove the first claim, recall that $\rho^*(g) = \big(\rho(g^{-1})\big)^T$, where $\rho^* : V \to \mathrm{GL}(V^*)$ is the dual representation of $\rho : V \to \mathrm{GL}(V)$. Hence, $\chi_{V^*} = \mathrm{Tr}\big(\rho_{V^*}(g)\big) = \mathrm{Tr}\big(\big(\rho_V(g^{-1})\big)^T\big) = \mathrm{Tr}\big(\rho(g)\big)^{-1}$, since flipping a matrix about it's diagonal doesn't change the diagonal entries. Now, if $A = \rho(g)$ has eigenvalues λ_i, then

$$
\begin{aligned}
\mathrm{Tr}(A^{-1}) &= \sum_i \frac{1}{\lambda_i} \text{ as the trace of a matrix is the sum of it's eigenvalues} \\
&= \sum_i \overline{\lambda_i} = \overline{\mathrm{Tr}(A)} = \chi_V(g)
\end{aligned}
$$

where the second equality follows from the fact that all the eigenvalues of A are roots of unity: the finiteness of G means that for any $g \in G$ there exists an $m \in \mathbb{Z}$ such that $g^m = 1 \implies \rho(g^m) = I \implies A^m = I$, so the eigenvalues of A must be such that when raised to the mth power they equal the eigenvalues of I, the identity matrix, whose only eigenvalue is 1. The second claim, follows from the fact that the trace of the tensor product of two matrices is the product of the traces of the individual matrices. ■

Definition 3.39. Let (ρ, V) be a representation for a group G. Define the set $V^G := \{v \in G : \rho(g)(v) = v \text{ for all } g \in G\}$. In other words, V^G consists of the invariant elements of V.

As you might expect, V^G is a subspace.

Proposition 3.40. V^G *as defined above is a linear subspace of* V*.*

Proof. The proof is quote straightforward. Let $v_1 \in V^G \implies \rho(g)(v_1) = v_1$ and $v_2 \in V^G \implies \rho(g)(v_2) = v_2$ for all $g \in G$. Thus, $\rho(g)(v_1 + v_2) = \rho(g)(v_2) + \rho(g)(v_2) = v_1 + v_2$, so V^G is closed under vector addition. The proof for closure under scalar multiplication is virtually identical, so we omit it. ■

Thus, the dimension of V^G is well-defined. With those preliminaries out of the way, we're on track to show Theorem 3.34.

Proof. We begin by expanding and simplifying the expression for the inner product of χ_V and χ_W,

$$\begin{aligned}\langle \chi_V, \chi_W\rangle &= \sum_{g\in G} \chi_V(g)\overline{\chi_W(g)} \text{ by the definition of } \langle\cdot,\cdot\rangle\\ &= \sum_{g\in G} \chi_V(g)\chi_{W^*}(g) \text{ by Proposition 3.38}\\ &= \sum_{g\in G} \chi_{V\otimes W^*}(g) \text{ again by Proposition 3.38}\\ &= \sum_{g\in G} \mathrm{Tr}\big(\rho_{V\otimes W^*}(g)\big) \text{ definition of the character}\\ &= \mathrm{Tr}\left[\frac{1}{|G|}\sum_{g\in G}\rho_{V\otimes W^*}(g)\right]\\ &= \dim\big[(V\otimes W^*)^G\big],\end{aligned}$$

where the last equality follows from the fact that $\frac{1}{|G|}\sum_{g\in G}\rho_{V\otimes W^*}(g)(\cdot) : V\otimes W^* \to V\otimes W^*$ is a projection operator onto the subspace $(V\otimes W^*)^G$, and the trace of a projection operator is the dimension of the subspace it projects onto.

At this point we if can manage to show that $(V\otimes W^*)^G$ and $\mathrm{Hom}_G(W,V)$ are isomorphic as vector spaces over $\mathbb{F}$, we're done, since isomorphic vector spaces have the same dimension. ∎

Thus, it all boils down to establishing $(V\otimes W^*)^G \cong \mathrm{Hom}_G(W,V)$, which is not very difficult, but requires some careful constructions and A LOT of algebraic manipulation. We split this proof into two lemmas: in the first one we show that $V\otimes W^*\cong \mathrm{Hom}(W,V)$, and then extend the argument to show $(V\otimes W^*)^G\cong \mathrm{Hom}_G(W,V)$. Here goes.

Lemma 3.41. *Let V and W be finite dimensional vector spaces over the same field. Then*

$$\mathrm{Hom}(W,V)\cong V\otimes W^*$$

as vector spaces.

Recall another equivalent condition for two vector spaces V and W to be isomorphic: the existence of linear maps $f : V\to W$ and $g : W\to V$ such that $f\circ g : W\to W = \mathrm{id}_W$ and $g\circ f : V\to V\mathrm{id}_V$, in other words, f and g are inverses of one another.

Proof. First, define $\alpha : V\otimes W^*\to \mathrm{Hom}(W,V)$ by $\alpha(v\otimes\varphi) = T_{v,\varphi}$ for all $v\in V$ and $\varphi\in W^*$, where $T_{v,\varphi} : W\to V$ is such that $T_{v,\varphi}(w) = \varphi(w)v$ for all $w\in W$, and extend linearly [8]. That is, for all $v_i\in V$ and $w_i\in W$,

$$\alpha\Big(\sum_i v_i\otimes\varphi_i\Big) = \sum_i \alpha(v_i\otimes\varphi_i) = \sum_i T_{v_i\varphi_i}.$$

Recall that any element in $V\otimes W^*$ can be expressed as $\sum_i v_i\otimes\varphi_i$ with $v_i\in V$ and $\varphi_i\in W^*$, so α is defined for all elements in $V\otimes W^*$. Note that $T_{v,\varphi}(w_1+w_2) = \varphi(w_1+w_2)v = \big(\varphi(w_1)+\varphi(w_2)\big)v = \varphi(w_1)v+\varphi(w_2)v = T_{v,\varphi}(w_1)+T_{v,\varphi}(w_2)$, so $T_{v,\varphi}$ is linear and one can similarly prove homogeneity, meaning we indeed have $T_{v,\varphi}\in\mathrm{Hom}(V,W)$.

In the opposite direction, define $\beta : \mathrm{Hom}(W,V)\to V\otimes W^*$ by

$$\beta(L) = \sum_{i\in I} L(w_i)\otimes w_i^*$$

for all $L\in\mathrm{Hom}(W,V)$ where $\{w_i : i\in I\}$ is a basis for W and $\{w_i^* : i\in I\}$ is the corresponding dual basis for W^*. This time we clearly have $\beta(L)\in V\otimes W^*$. Next, we show that β is linear

[8] Hence α is linear by definition.

$$
\begin{aligned}
\beta(L_1+L_2) &= \sum_{i\in I}(L_1+L_2)(w_i)\otimes w_i^* \text{ definition of } \beta \\
&= \sum_{i\in I}\big(L_1(w_i)+L_2(w_i)\big)\otimes w_i^* \text{ by linearity of } L \\
&= \sum_{i\in I} L_1(w_i)\otimes w_i^* + L_2(w_i)\otimes w_i^* \text{ by linearity of } \otimes \\
&= \sum_{i\in I} L_1(w_i)\otimes w_i^* + \sum_{i\in I} L_2(w_i)\otimes w_i^* \\
&= \beta(L_1)+\beta(L_2) \text{ definition of } \beta.
\end{aligned}
$$

Again, the proof that β is homogeneous is essentially identical. Since α and β are both linear, all that is left is to show is that they are inverses of each other. First, we show that $\alpha\big(\beta(L)\big)(w) = L(w)$.

$$
\alpha\big(\beta(L)\big)(w) = \alpha\Big(\sum_{i\in I} L(w_i)\otimes w_i^*\Big)(w) = \sum_{i\in I}\alpha\big(L(w_i)\otimes w_i^*\big)(w) = \sum_{i\in I} w_i^*(w)L(w_i).
$$

Now, write w as $\sum_{i\in I} c_i w_i$. Then, $w_i^*(w) = c_i$ (by the definition of the dual basis). Picking off,

$$
\sum_{i\in I} w_i^*(w)L(w_i) = \sum_{i\in I} c_i L(w_i) = L\Big(\sum_{i\in I} c_i w_i\Big) = L(w).
$$

Second, we show that $\beta\big(\alpha(v\otimes\varphi)\big) = v\otimes\varphi$

$$
\beta(\alpha(v\otimes\varphi)) = \beta(T_{v,\varphi}) = \sum_{i\in I} T_{v,\varphi}(w_i)\otimes w_i^* = \sum_{i\in I}\varphi(w_i)v\otimes w_i^* = v\otimes\sum_{i\in I}\varphi(w_i)w_i^*.
$$

To see what is $\sum_{i\in I}\varphi(w_i)w_i^* \in W^*$, we evaluate it at a an arbitrary $w\in W$

$$
\Big(\sum_{i\in I}\varphi(w_i)w_i^*\Big)(w) = \sum_i \varphi(w_i)w_i^*(w) = \sum_i \varphi(w_i)c_i = \varphi\Big(\sum_{i\in I} c_i w_i\Big) = \varphi(w).
$$

Hence, $\sum_{i\in I}\varphi(w_i)w_i^* = \varphi$. Substituting this into our previous equations yields $\beta(\alpha(v\otimes\varphi)) = v\otimes\varphi$, completing the proof. ■

Remark 3.42. Even though it might not be apparent, the β used above is independent of the choice of basis for W; a fact that will be important to us while proving Lemma 3.43. To see why, consider a 'new' basis for W, namely $\{w_i'\}_{i\in I}$ such that $w_i' = L(w_i)$, where $L : W \to W$ is an invertible linear transformation. Explicitly, there exists a sqaure matrix A that describes L with respect to the 'old' basis $\{w_i\}_{i\in I}$, i.e, $w_i' = \sum_{j\in I} A_{ij}w_j$ where A_{ij} is the (i,j) entry in the matrix A. Since β also has the dual basis involved, one has to compute an expression for $\{w_i'^*\}_{i\in I}$, the *unique* basis dual to $\{w_i'\}_{i\in I}$ in terms of the $\{w_i^*\}_{i\in I}$, the older dual basis. Obviously we must have $w_i'^*(w_j') = \delta_{ij}$, and it can be shown without breaking sweat that $w_i'^* = \sum_{k\in I}(A^{-1})_{ki}w_k^*$. Now, we can substitute these transformed bases into β

$$\begin{aligned}\sum_{i\in I} T(w_i')\otimes w_i'^* &= \sum_{i\in I}\left(T\Big(\sum_{j\in I} A_{ij}w_j\Big)\otimes\sum_{k\in I}(A^{-1})_{ki}w_k^*\right) = \sum_{i\in I}\sum_{j\in I}\Big(A_{ij}T(w_j)\otimes\sum_{k\in I}(A^{-1})_{ki}w_k^*\Big)\\ &= \sum_{i\in I}\sum_{j\in I}\sum_{k\in I} A_{ij}T(w_j)\otimes(A^{-1})_{ki}w_k^* = \sum_{k\in I}\sum_{j\in I}\Big(\sum_{i\in I}\big(T(w_j)\otimes(A_{ij})(A^{-1})_{ki}w_k^*\big)\Big)\\ &= \sum_{k\in I}\sum_{j\in I}\left(T(w_j)\otimes w_k^*\sum_{i\in I}(A_{ij})(A^{-1})_{ki}\right) \text{ Note that} \sum_{i\in I}(A_{ij})(A^{-1})_{ki} = \delta_{jk}\\ &= \sum_{k\in I}\sum_{j\in I}\left(T(w_j)\otimes w_k^*\delta_{jk}\right) = \sum_{j\in I}\left(T(w_j)\otimes\sum_{k\in I} w_k^*\delta_{jk}\right) = \sum_{j\in I} T(w_j)\otimes w_j^*.\end{aligned}$$

Voilà! After shunting the summations around countless times, we got back our original expression for $\beta(T)$, in terms of the 'old' basis $\{w_i\}_{i\in I}$! Thus, β is indeed independent of the basis used. Keep this in mind!

Lemma 3.43. *Let V and W be representations for a finite group G. Then*

$$\mathrm{Hom}_G(W,V)\cong(V\otimes W^*)^G$$

as vector spaces.

Note that $\mathrm{Hom}_G(W,V)$ is a linear subspace of $\mathrm{Hom}(W,V)$ and similarly $(V\otimes W^*)^G$ is a subspace of $V\otimes W^*$. Thus, it suffices to show that the image of α, which is a linear map, when restricted to the domain $(V\otimes W^*)^G$ lies in $\mathrm{Hom}_G(W,V)$ and vice versa for β.

Proof. As mentioned, the proof is broken into two parts.

Stage one. First, it is required to show that if $x=\sum_i v_i\otimes\varphi_i\in(V\otimes W^*)^G$, we have $\alpha(x)=\sum_i\varphi_i(\cdot)v_i\in\mathrm{Hom}_G(W,V)\iff\alpha(x)\big(\rho_W(g^{-1})(w)\big)=\rho_V(g^{-1})\big(\alpha(x)(w)\big)$ for all $g\in G$ and $w\in W$, by the definition of a $G-$equivariant map, which is equivalent to proving

$$\alpha(x)(w)=\rho_V(g)\Big(\alpha(x)\big(\rho_W(g^{-1})(w)\big)\Big).$$

Expanding the left-hand side

$$\begin{aligned}\alpha(x)(w) &= \alpha\big(\rho_{V\otimes W^*}(g)(x)\big)(w) \text{ since } x \text{ is invariant under } \rho_{V\otimes W^*}\\ &= \alpha\Big(\sum_i\rho_V(g)(v_i)\otimes\rho_{W*}(g)(\varphi_i)\Big)(w) \text{ linearity of } \rho_{V\otimes W^*}\\ &= \sum_i\rho_{W*}(g)(\varphi_i)(w)\times\rho_V(g)(v_i) \text{ linearity and definition of } \alpha\\ &= \sum_i\varphi_i\big(\rho_W(g^{-1})(w)\big)\times\rho_V(g)(v_i) \text{ definition of dual representation.}\end{aligned}$$

One the other hand,

$$\begin{aligned}\rho_V(g)\Big(\alpha(x)\big(\rho_W(g^{-1})(w)\big)\Big) &= \rho_V(g)\Big(\sum_i \varphi_i\big(\rho_W(g^{-1})(w)\big)\times v_i\Big)\\ &= \sum_i \rho_V(g)\Big(\varphi_i\big(\rho_W(g^{-1})(w)\big)\times v_i\Big)\\ &= \sum_i \varphi_i\big(\rho_W(g^{-1})(w)\big)\times\rho_V(g)(v_i)\\ &= \alpha(x)(w).\end{aligned}$$

Thus, we have $x \in (V\otimes W^*)^G \implies \alpha(x) \in \mathrm{Hom}_G(W,V)$, which completes this part of the proof.

<u>**Stage two**</u>. Next, we show that $L \in \mathrm{Hom}_G(W,V) \implies \beta(L) \in (V\otimes W^*)^G$. Start by applying $\rho_{V\otimes W^*}(g)$ to our expression for β (in terms of the basis $\{w_i\}_{i\in I}$):

$$\begin{aligned}\rho_{V\otimes W^*}(g)\big(\beta(L)\big) = \rho_{V\otimes W^*}(g)\Big(\sum_{i\in I} L(w_i)\otimes w_i^*\Big) &= \sum_{i\in I}\rho_{V\otimes W^*}(g)(L(w_i)\otimes w_i^*)\\ &= \sum_{i\in I}\rho_V(g)\big(L(w_i)\big)\otimes\rho_{W^*}(g)(w_i^*) \text{ definition of the tensor representation}\\ &= \sum_{i\in I} L\big(\rho_W(g)(w_i)\big)\otimes\rho_{W^*}(g)(w_i^*) \text{ by equivariance of } L\\ &= \sum_{i\in I} L\big(\rho_W(g)(w_i)\big)\otimes w_i^*\rho_W(g^{-1}) \text{ definition of the dual of a representation.}\end{aligned}$$

Now for the moment of cancellation: we are transforming the basis $w_i \mapsto \rho_W(g)(w_i)$ and one can check that the corresponding dual basis transformation induced by it is $w_i^* \mapsto w_i^*\rho_W(g^{-1})$—exactly what is seen in the equation above. Since we already know that β is independent of the choice of basis, $\sum_{i\in I} L\big(\rho_W(g)(w_i)\big)\otimes w_i^*\rho_W(g^{-1}) = \beta(L)$, which completes the proof. ■

Ah! The monstrous proof is all but done! As discussed before even beginning the proof, the following corollary follows directly from Theorem 3.34.

Corollary 3.44. *Let $\{\chi_\alpha\}_{\alpha\in A}$ denote the set of distinct irreducible representations of a group G. Then, $\langle\chi_\alpha,\chi_\beta\rangle = \delta_{\alpha\beta}$.*

Even though this proof might have seemed lengthy, at every stage we knew precisely what to do: right from the very beginning, where it didn't take a genius to recognize characters are being complex conjugated (so we have duals) and multiplied (so tensor products). At any rate, as promised, I will be presenting another proof, which, in my opinion, isn't the best to simply *prove* the first part of Theorem 3.28, but rather opens the gate for a deeper dive.

3.4.2. *The Second Proof.* We start with the result that will get us on our way.

Theorem 3.45 (Schur's First Orthogonality Relation)**.** *Let G be a finite group, $\Gamma^{(i)}: G\to \mathrm{GL}(V_i)$ and $\Gamma^{(j)}: G\to\mathrm{GL}(V_j)$ irreducible representations of G. Then,*

$$\langle\Gamma^{(i)}_{\alpha\beta},\Gamma^{(j)}_{\mu\nu}\rangle = \frac{1}{|G|}\sum_{g\in G}\Gamma^{(i)}_{\alpha\beta}(g)\overline{\Gamma^{(j)}_{\mu\nu}(g)} = \frac{1}{\dim(\Gamma^{(i)})}\delta_{ij}\delta_{\alpha\mu}\delta_{\beta\nu}$$

This is quite the mouthful, so let's thoroughly unpack it for better understanding, and why it's related to Theorem 3.28. First, we have new notation: $\Gamma_{\alpha\beta}(g)$ denotes the (α,β) entry in the matrix $\Gamma(g)$, where $\Gamma: G\to\mathrm{GL}(V)$ is a representation of a group G—really then, $\Gamma_{\alpha\beta}(\cdot)$ is a function from G, to the underlying base field $\mathbb{F}$. What we're trying to establish is an orthogonality relation between certain elements (called the matrix coefficient functions) of $\mathbb{C}[G]$, the set of *all* functions form G to $\mathbb{F}$—not just the class functions. The relation is as one would expect: distinct matrix coefficient functions are orthogonal, and for the inner product of two such functions to be non-zero, all the three parameters: the representation ($i = j$), the row and column in the matrix ($\alpha = \mu$ and $\beta = \nu$) must be the same, i.e, the functions themselves must be identical. It makes logical to say that this theorem is a generalization of Theorem 3.28, as we're expanding to all functions now, as opposed to just class functions. We break the proof of Theorem 3.45 into two pieces varying one parameter $i/j, \alpha/\mu, \beta/\nu$ at a time.

Proposition 3.46. *Let $\Gamma^{(i)} : G \to \mathrm{GL}(V_i)$ and $\Gamma^{(j)} : G \to \mathrm{GL}(V_j)$ be distinct unitary irreducible representations for a finite group G, then*

$$\langle \Gamma^{(i)}_{\alpha\beta}, \Gamma^{(j)}_{\mu\nu} \rangle = \frac{1}{|G|} \sum_{g \in G} \Gamma^{(i)}_{\alpha\beta}(g) \overline{\Gamma^{(j)}_{\mu\nu}(g)} = 0$$

for all α, β, μ and ν.

Remark 3.47. Note that due to Proposition 3.15, we can assume without loss of generality that given a representation (ρ, V), $\rho(g)$ is unitary for all $g \in G$.

Proof. Let T be any linear map from V_j to V_i. First, we start by constructing a $G-$equivariant map $T' : V_j \to V_i$ from T, defined by $T' = \sum_{g \in G} \Gamma^{(i)}(g) \circ T \circ \Gamma^{(j)}(g^{-1})$. Note that to verify that T' is $G-$equivariant, we must show that $\Gamma^{(i)}(h) \circ T' = T' \circ \Gamma^{(j)}(h)$, for all $h \in G$, as we do below

$$\begin{aligned}
\Gamma^{(i)}(h) \circ T' &= \Gamma^{(i)}(h) \circ \frac{1}{|G|} \sum_{g \in G} \Gamma^{(i)}(g) \circ T \circ \Gamma^{(j)}(g^{-1}) \text{ definition of } T' \\
&= \frac{1}{|G|} \sum_{g \in G} \Gamma^{(i)}(hg) \circ T \circ \Gamma^{(j)}(g^{-1}) \text{ as } \Gamma^{(i)} \text{ is a homomorphism} \\
&= \frac{1}{|G|} \sum_{g \in G} \Gamma^{(i)}(hg) \circ T \circ \Gamma^{(j)}((hg)^{-1}h) \text{ re-writing } g^{-1} \text{ as } (gh)^{-1}h \\
&= \frac{1}{|G|} \sum_{g' \in G} \Gamma^{(i)}(g') \circ T \circ \Gamma^{(j)}((g')^{-1}h) \text{ note that } g \mapsto gh = g' \text{ is a bijection on } G \\
&= \frac{1}{|G|} \sum_{g' \in G} \Gamma^{(i)}(g') \circ T \circ \Gamma^{(j)}((g')^{-1}) \circ \Gamma^{(j)}(h) \text{ as } \Gamma^{(i)} \text{ is a homomorphism} \\
&= \frac{1}{|G|} \Big(\sum_{g' \in G} \Gamma^{(i)}(g') \circ T \circ \Gamma^{(j)}((g')^{-1}) \Big) \circ \Gamma^{(j)}(h) \text{ additivity of } \circ \\
&= T' \circ \Gamma^{(j)}(h) \text{ definition of } T'.
\end{aligned}$$

This shows that T' is a $G-$equivariant map, or $T' \in \mathrm{Hom}_G(V_j, V_i)$, and since we've assumed that $V_j \ncong V_i$, Schur's lemma tells us that $T' = 0$. Remember that T can be *any* linear map, and here define it to be such that it has zeros everywhere except for the (β, ν) entry. On the one hand we know that $T'_{mn} = 0$ for all m and n, but on the other hand

$$\begin{aligned}
T'_{\alpha\mu} &= \frac{1}{|G|}\sum_{g\in G}\big(\Gamma^{(i)}(g)\circ T\circ\Gamma^{(j)}(g^{-1})\big)_{\alpha\mu} \text{ as for matrices } A \text{ and } B,\ (A+B)_{\alpha\beta} = A_{\alpha\beta}+B_{\alpha\beta}\\
&= \frac{1}{|G|}\sum_{g\in G}\sum_{a}\sum_{b}\Gamma^{(i)}_{\alpha a}(g)T_{ab}\Gamma^{(j)}_{b\mu}(g^{-1}) \text{ as } (ABC)_{ij} = \sum_{k}\sum_{l}A_{il}B_{lk}C_{kj}\\
&= \frac{1}{|G|}\sum_{g\in G}\Gamma^{(i)}_{\alpha\beta}(g)\Gamma^{(j)}_{\nu\mu}(g^{-1}) \text{ as } T_{ab}=1 \text{ iff } a=\beta \text{ and } b=\nu\\
&= \frac{1}{|G|}\sum_{g\in G}\Gamma^{(i)}_{\alpha\beta}(g)\overline{\Gamma^{(j)}_{\mu\nu}(g)} \text{ as } \Gamma^{(j)}(g) \text{ is unitary}\\
&= \langle\Gamma^{(i)}_{\alpha\beta}(g),\Gamma^{(j)}_{\mu\nu}(g)\rangle.
\end{aligned}$$

Putting the two equalities together, we get $\langle\Gamma^{(i)}_{\alpha\beta},\Gamma^{(j)}_{\mu\nu}\rangle = 0$, for all α,β,μ and ν, which completes the proof. ■

Next, we progress onto varying the other two parameters (at the same time!): α/μ and β/ν.

Proposition 3.48. *Let $\Gamma : G \to \mathrm{GL}(V)$ be an irreducible representation for a group G. Then, $\langle\Gamma_{\alpha\beta},\Gamma_{\mu\nu}\rangle \neq 0 \iff \alpha=\mu$ and $\beta=\nu$.*

Proof. This proof is quite similar to the previous one—we consider any linear map $T : V \to V$, and build a $G-$equivariant map $T' : V \to T$ out of T, $T' = \frac{1}{|G|}\sum_{h\in G}\Gamma(g)\circ T\circ\Gamma(g^{-1})$. Just like last time, we set T to be the matrix such that it's zero everywhere except at (β,ν), when it's 1. This leads us to

$$T'_{\alpha\mu} = \frac{1}{|G|}\sum_{g\in G}\Gamma^{(i)}_{\alpha\beta}(g)\overline{\Gamma^{(j)}_{\mu\nu}(g)} = \langle\Gamma_{\alpha\beta},\Gamma_{\mu\nu}\rangle.$$

Since $T' \in \mathrm{Hom}_G(V,V)$, we can apply Schur's lemma, but this time we have that $T' = \lambda I$. Note that $\langle\Gamma_{\alpha\beta},\Gamma_{\mu\nu}\rangle \neq 0 \iff T_{\alpha\mu}\neq 0 \iff \lambda\neq 0$ and $\alpha=\mu$. Next, $\lambda\neq 0 \iff \mathrm{Tr}(T')\neq 0 \iff \mathrm{Tr}(T)\neq 0 \iff \beta=\nu$ where $\mathrm{Tr}(T') = \mathrm{Tr}(T)$ as the trace is linear and $\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$. This completes the proof. ■

Putting Proposition 3.46 and Proposition 3.48, we get that

$$\langle\Gamma^{(i)}_{\alpha\beta},\Gamma^{(j)}_{\mu\nu}\rangle = \begin{cases}0 & \text{if } i\neq j \text{ or } \alpha\neq\mu \text{ or } \beta\neq\nu,\\ \neq 0 & \text{if } i=j \text{ and } \alpha=\mu \text{ and } \beta=\nu.\end{cases}$$

All the remains to be shown that in the second case, the non-zero value the inner product takes on is $\frac{1}{\dim(\Gamma^{(i)})} = \frac{1}{\dim(\Gamma^{(j)})}$. However, the derivation is dry, devoid of any insight, so we omit it here. Nevertheless, assuming that it is true, we will show that the irreducible characters of a group are orthonormal.

Proof. We have to show that $\langle \chi_i, \chi_j \rangle = \delta_{ij}$, where $\{\chi_\alpha\}_{\alpha in A}$ denotes the set of distinct irreducible characters of G.

$$\begin{aligned}\langle \chi_i, \chi_j \rangle &= \frac{1}{|G|}\sum_{g\in G} \chi_i(g)\overline{\chi_j(g)} \\ &= \frac{1}{|G|}\sum_{g\in G}\Big(\sum_{\alpha}\Gamma^{(i)}_{\alpha\alpha}(g)\Big)\overline{\Big(\sum_{\beta}\Gamma^{(j)}_{\beta\beta}(g)\Big)} \\ &= \frac{1}{|G|}\sum_{g\in G}\sum_{\alpha,\beta}\Gamma^{(i)}_{\alpha\alpha}(g)\overline{\Gamma^{(j)}_{\beta\beta}(g)} \\ &= \sum_{\alpha,\beta}\frac{1}{|G|}\Gamma^{(i)}_{\alpha\alpha}(g)\overline{\Gamma^{(j)}_{\beta\beta}(g)} \\ &= \sum_{\alpha,\beta}\langle \Gamma^{(i)}_{\alpha\alpha}(g), \Gamma^{(j)}_{\beta\beta}(g)\rangle\end{aligned}$$

Now, assuming that $i \neq j$ we have that $\langle \Gamma^{(i)}_{\alpha\alpha}(g), \Gamma^{(j)}_{\beta\beta}(g)\rangle = 0$ for all α and β, so $\langle \chi_i, \chi_j \rangle = 0$. If $i = j$, then $\sum_{\alpha,\beta}\langle \Gamma^{(i)}_{\alpha\alpha}(g), \Gamma^{(j)}_{\beta\beta}(g)\rangle = \sum_{\alpha}\langle \Gamma^{(i)}_{\alpha\alpha}(g), \Gamma^{(i)}_{\alpha\alpha}(g)\rangle = \sum_{i=1}^{\dim(V_i)} \frac{1}{\dim(V_i)} = 1$, completing the proof. ■

Now, as soon as you learnt that the the matrix coefficients of irreducible representations are orthogonal, you must have guessed that they form a basis for $\mathbb{C}[G]$. This is true, and for the sake of completeness, we provide a proof. Since we already have a basis $B = \{\phi_g\}_{g\in G}$ where $\phi_g(h) = \delta_{gh}$, which is the version of the *useless* basis for $\mathbb{C}[G]$, all we have to show is that the number of irreducible matrix coefficients, which is $\sum_{\alpha\in A}\dim(V_i)$ is equal to $|B| = |G|$.

Proposition 3.49. *Let G be a group and $\{V_\alpha\}_{\alpha\in A}$ be the set of distinct irreducible representations of G. Then, $\sum_{\alpha\in A}(\dim V_\alpha)^2 = |G|$.*

Proof. Remember the group algebra $\mathbb{C}[G]$? We're going to define a representation of G, $\rho_R : G \to \mathbb{C}[G]$, called the *left regular representation.* It's defined quite simply: $\rho_R(g)(h) = gh$ for all $g \in G$ and $h \in G \subset \mathbb{C}[G]$, and extend linearly to all elements of $\mathbb{C}[G]$. That is,

$$\rho_R(g)\left(\sum_{h\in G} a_g h\right) = \sum_{h\in G} a_g \rho_R(g)(h) = \sum_{h\in G} a_g gh = g\sum_{h\in G} a_g h.$$

Note that

$$\chi_R(g) = \begin{cases} |G| & \text{if } g = e \\ 0 & \text{if } g \neq e\end{cases}.$$

This is because if g is not the identity, then $\rho_R(g)$ doesn't have any eigenvectors. Otherwise, if g is the identity, then $\rho_R(g)$ would be the $|G|$ by $|G|$ identity matrix, whose trace is clearly $|G|$.

Now, Maschke's theorem tells us that $\mathbb{C}[G] \cong \bigoplus_{\alpha\in A} V_\alpha^{\oplus n_\alpha} \implies \chi_R = \sum_{\alpha\in A} n_\alpha \chi_\alpha$ for some $n_\alpha \in \mathbb{N}$, and where $\{\chi_\alpha\}_{\alpha\in A}$ denotes the set of distinct irreducible characters of G as usual. To determine those n_α's, consider

$$\langle \chi_R, \chi_\beta \rangle = \left\langle \sum_{\alpha\in A} n_\alpha\chi_\alpha, \chi_\beta \right\rangle = \sum_{\alpha\in A} n_\alpha \langle \chi_\alpha, \chi_\beta\rangle = n_{\beta.}$$

On the other hand,

$$\langle \chi_R, \chi_\beta \rangle = \frac{1}{|G|}\sum_{g\in G}\chi_R(g)\overline{\chi_\beta(g)} = \frac{1}{|G|}(\chi_R(e)\overline{\chi_\beta(e)}) = \dim(V_\beta).$$

Putting the two together gives us

$$\mathbb{C}[G] \cong \bigoplus_{\alpha \in A} V_\alpha^{\oplus \dim(V_\alpha)}.$$

Taking the dimension of both the sides gives the desired result. ■

The following corollary summarizes the main result of this section.

Corollary 3.50. *Let G be a finite group and $\{(\rho_\alpha, V_\alpha)\}_{\alpha \in A}$ the set of distinct irreducible representations of G, and $\Gamma^{(i)}_{\mu\nu}(g)$ be the (μ, ν) matrix entry of $\rho_i(g)$. Then, the set*

$$\{\Gamma^{(i)}_{\mu\nu} : i \in A \text{ and } 1 \le \mu, \nu \le \dim(V_i)\}$$

is an orthogonal basis for $\mathbb{C}[G]$.

4. Conclusion

That's great, but as mentioned earlier, this machinery that we have developed to compute the Fourier series of a function $f \in \mathbb{C}[G]$ only works when G is finite; and, arguably, the most *interesting* groups out there such as $\mathbb{R}/\mathbb{Z}$, and $\mathrm{SO}(n)$ aren't finite. For starters, the inner product that we had defined on $\mathbb{C}[G]$:

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{g \in G} f_1(g)\overline{f_2(g)},$$

completely breaks down when $|G| = \infty$. This question, of summing infinitely many numbers, is a quintessential example of developing the integral over any set, as answered by measure theory. Almost instinctively, one would set

$$\langle f_1, f_2 \rangle = \int_G f_1(x)\overline{f_2(x)}\, d\mu(x),$$

but that raises another question: what is the measure μ here? What's the $\sigma-$algebra? Is it even guaranteed that a measure μ always exists on a group G?

To answer this question, we equip G with additional structure: specifically, we put a topology on G to turn it into a **topological group**—a structure that's simultaneously a group and a topological space. However, as a techincal side, note that the topology must be such that it's compatible with the pre-existing group structure, that is, the inversion map $g \mapsto g^{-1}$ and the multiplication map $(g, h) \mapsto gh$ must be continuous with respect to the topology on G, and the product topology on $G \times G$ respectively. Then, the sigma algebra is simply the one generated by the open sets of G, called the **Borel sigma algebra**.

It was proved by the revolutionary french mathematician André Weil, that if G is a *locally compact* topological group, that is, if there exists a compact neighbourhood for each point in G, then there exists a unique measure μ on G [9], that apart from satisfying a few technical conditions, is *invariant*, that is, translating a Borel set S of G around by a $g \in G$ doesn't affect the measure of S [10], exactly how we have $\lambda(B) = \lambda(a + B) = \lambda(\{a + b : b \in B\})$ where λ denotes the Lebesgue measure on $\mathbb{R}$, which can be viewed as a (locally compact) topological group under addition. This measure is called the **Haar measure**, named after the Hungarian mathematician Alfréd Haar, who first studied it in connection with Hilbert's fifth problem on Lie groups, which are special types of topological groups.

Equipped the Haar measure, we can now generalize Schur's orthogonality relations to a compact group, a series of results known as the **Peter-Weyl theorem**. Essentially, one version gives us an

[9]Note that the measure is unique up to multiplication by a scalar. Note also that we have two types of invariance of measures: a left invariant measure μ is such that $\mu(S) = \mu(gS) = \mu(\{gs : s \in S\})$ and a right invariant measure ν is such that $\nu(S) = \nu(Sg) = \nu(\{sg : g \in G\})$ for all $g \in G$ and open sets S.

[10]Throwing a ball up doesn't make it shrink spontaneously.

explicit formula for a basis for $L^2(G)$, the set of functions $f : G \to \mathbb{C}$ that are square-integrable with respect to the Haar measure [11]. We have:

Theorem 4.1. *The set of finite linear combinations of all matrix coefficients of all finite-dimensional irreducible unitary representations of a compact group G is dense in $L^2(G)$.*

Here as usual, the norm on $L^2(G)$ is the L^2 norm, the one induced by the inner product. That is,

$$\|f\|_{L^2} := \sqrt{\int_G |f(x)|^2 \, d\mu(x)}.$$

Note that a *matrix coefficient* is formally defined to be a function from G to $\mathbb{F}$ that is of the form $L \circ \rho$ where $\rho : G \to \mathrm{GL}(V)$ is a representation of G, and L is any linear map from $\mathrm{GL}(V)$ to $\mathbb{F}$, that is, $L \in (\mathrm{GL}(V))^*$. This definition single handily incorporates the trace and the individual matrix entry functions that we talked about earlier.

Since most of the familiar groups such as $\mathbb{R}/\mathbb{Z}$, $\mathrm{SO}(n)$ and $\mathrm{SU}(n)$ are all compact, we can directly apply the Peter-Weyl theorem in these contexts! Applying it to $\mathbb{R}/\mathbb{Z}$ gives us the usual Fourier series for periodic functions $f : \mathbb{R} \to \mathbb{C}$. On the other hand, computing the matrix coefficients of the irreducible unitary representations of $\mathrm{SO}(2)$, allows us to essentially decompose functions defined on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$, as it can be easily shown that $\mathrm{SO}(2)$ is homeomorphic to $\mathbb{S}^2$! These functions are known as **spherical harmonics**, and they're littered all over the place in physics, especially in quantum mechanics, where they can be used to describe the angular part of the wave functions of elementary particles. In fact, they even appear in the celebrated Schrödinger equation for the hydrogen atom! More concretely, just as we use the ordinary Fourier series to solve ordinary differential equations defined on $\mathbb{R}/\mathbb{Z}$, the unit circle, one could use spherical harmonics to solve differential equations defined on the sphere! Examples of such differential equations include the **Helmholtz equation**, which has numerous applications in science including in the wave equation, modelling diffusion and even optics. Yet more fancifully, one could extend this to $\mathrm{SO}(n)$, which can be visualized as a $n + 1$ dimensional sphere!

At the end of the day, all these methods of analysing a complex function by decomposing it into it's constituent parts boils down on one single theorem: The Peter-Weyl theorem!

Acknowledgements

I would like to thank Professor Simon Rubinstein-Salzedo for organizing this course, without which I wouldn't have been able to take my passion for Fourier analysis to this extent. I would also like to thank Paulina Paiz, my TA for this course, for her continual support and guidance throughout the course of this project.

References

[Dan18] Rohan Dandavati. The fourier transform on finite groups: theory and computation. *UChicago Math Summer Research Experience for Undergraduates (REU)*, 2018.

[Nac65] Leopoldo Nachbin. *The Haar Integral*. D. Van Nostrand Company, 1965.

[Tao11] Terrence Tao. The peter-weyl theorem, and non-abelian fourier analysis on compact groups. *Wordpress*, 2011.

[Wad21] Simon Wadsley. Representation theory. *University of Cambridge*, 2021.

[11] In symbols, $\int_G |f(x)|^2 \, d\mu(x) < \infty$.