ON ORTHOGONAL FUNCTIONS AND POLYNOMIALS

VARUN KATAKI

Abstract. The paper begins with introducing inner products and norms. Using inner products, a generalized form of perpendicularity known as orthogonality. Then, we define an inner product for a function space, letting us declare functions to be orthogonal. We then use Fourier series to decompose functions into orthogonal components. We show what properties these decompositions have and when. We then explore how functions can be recomposed through variants of the Shannon-Nyquist sampling theorem. We end by seeing what properties the Legendre polynomials have, and why we might want to use them.

1. Introduction

The Fourier series, first introduced by Joseph Fourier in his paper on heat [Poi08], has been used as a crucial technique in physical and abstract applications. This tool allows us to convert potentially complicated functions into an easier to work with form. The information we want to isolate may differ between applications, so it would be helpful to be able to create Fourier Series with respect to different functions. The notion of orthogonal functions allows us to do that.

Fourier series and transforms also play crucial roles in transferring signals, as the decomposition of waves into frequencies lets us understand their behavior more thoroughly. With some restrictions, we can completely recover waveforms from discrete samples, as shown by Shannon in his 1949 paper [Sha49]. This too can be modified to work with sets of orthogonal functions.

There are many different sets of orthogonal polynomials, each with their own use cases. One of the most common is $\{e^{inx} : n \in \mathbb{Z}\}$, as through Fourier series, it can accurately represent periodic functions, such as signals and waves. Legendre polynomials, a set of orthogonal polynomials specifically, have usage in spherical coordinates. Others, such as the Hermite or Chebyshev polynomials have their own use cases.

This paper will first go over inner product spaces, to provide fundamental tools for our exploration. Then we will define what it means for vectors in general to be orthogonal and orthonormal. We will then define our inner product for functions. With that, we can use Fourier series to decompose functions into orthogonal parts. Then, we can show what properties these series hold, such as what requirements need to be met for these series to converge to their original functions. Then, we discuss the Shannon-Nyquist sampling theorem, which will give us a way to recover functions using their Fourier series. Finally, we will explore why we might want to use these generalized tools in the first place by demonstrating properties of the Legendre polynomials.

Date: June 2026.

2. The Inner Product and Norm

We will first define the inner product, as it will give us the capabilities to determine orthogonality.

Definition 2.1 (inner product)**.** Let V be a vector space (over $\mathbb{C}$). An inner product on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ such that the following holds true:

(1) $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ for all $\mathbf{v} \in V \setminus \{\mathbf{0}\}$.
(2) $\langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$ for all $\mathbf{v}, \mathbf{u} \in V$.
(3) $\langle \mathbf{v} + \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$ for all $\mathbf{v}, \mathbf{u}, \mathbf{w} \in V$.
(4) $\langle \alpha\mathbf{v}, \mathbf{u} \rangle = \alpha\langle \mathbf{v}, \mathbf{u} \rangle$ for all $\alpha \in \mathbb{C}$ and $\mathbf{v}, \mathbf{u} \in V$.

A vector space V along with an inner product $\langle \cdot, \cdot \rangle$ is called an *inner product space*. A natural example of an inner product space is $\mathbb{R}^2$ with the basic dot product. A very helpful inequality related to inner product spaces is the Cauchy Schwarz inequality.

Proposition 2.2 (Cauchy-Schwarz)**.** *Let V be an inner product space and $\mathbf{v}, \mathbf{u} \in V$. Then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle}.$$

Proof. Let $\mathbf{v}, \mathbf{u} \in V$. Let α be such that $|\alpha| = 0$ and that $\alpha\langle \mathbf{u}, \mathbf{v} \rangle = |\langle \mathbf{u}, \mathbf{v} \rangle|$, which will always exist. Now, we define a function $p : \mathbb{R} \to \mathbb{R}$ by saying $p(t) = \langle t\alpha\mathbf{u} + \mathbf{v}, t\alpha\mathbf{u} + \mathbf{v} \rangle$. We know that for any real t, $p(t)$ must be non-negative by the first property of inner products and real by the second property. Now, using the linearity of the inner product, we have

$$\begin{aligned} p(t) &= \langle t\alpha\mathbf{u}, t\alpha\mathbf{u} \rangle + \langle t\alpha\mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, t\alpha\mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= t\alpha t\overline{\alpha}\langle \mathbf{u}, \mathbf{u} \rangle + t\alpha\langle \mathbf{u}, \mathbf{v} \rangle + t\overline{\alpha\langle \mathbf{u}, \mathbf{v} \rangle} + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle t^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| t + \langle \mathbf{v}, \mathbf{v} \rangle. \end{aligned}$$

If $\langle \mathbf{u}, \mathbf{u} \rangle \neq 0$, then $p(t)$ is a degree 2 polynomial, and since $p(t)$ never changes signs, The discriminant Δ of $p(t)$ must be non-positive. From here,

$$\Delta = 4\left(|\langle \mathbf{u}, \mathbf{v} \rangle|^2 - \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle}^2\right) \leq 0,$$

from which we can recover the Cauchy-Schwarz inequality. If it was the case that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$, then it must be that $\mathbf{u} = 0$, and so the equality holds true, completing the proof. ∎

It would also be helpful for our purposes to define the norm, a method of measuring "length".

Definition 2.3 (norm)**.** Let V be a vector space (over $\mathbb{C}$). A norm on V is a map $||\cdot|| : V \to \mathbb{R}$ such that the following holds true:

(1) $||\mathbf{v}|| \geq 0$ for all $\mathbf{v} \in V$ with $||\mathbf{v}|| = 0 \iff \mathbf{v} = \mathbf{0}$.
(2) $||\alpha\mathbf{v}|| = |\alpha| \cdot ||\mathbf{v}||$ for all $\alpha \in \mathbb{C}$ and all $\mathbf{v} \in V$.
(3) $||\mathbf{v} + \mathbf{u}|| \leq ||\mathbf{v}|| + ||\mathbf{u}||$ for all $\mathbf{v}, \mathbf{u} \in V$.

A vector space V along with a norm $|| \cdot ||$ is called a *normed vector space*. Now, to finish up this section, for some inner product space V, we can define a norm based upon our inner product.

Proposition 2.4. *Let V be an inner product space. The real valued function defined by $||\mathbf{v}|| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a norm on V.*

Proof. The first property of the norm is obvious by Definition 2.1. For the second property, note that $\langle \mathbf{v}, \alpha\mathbf{u}\rangle = \overline{\alpha}\langle \mathbf{v}, \mathbf{u}\rangle$. Now, $||\alpha\mathbf{v}|| = \sqrt{\langle \alpha\mathbf{v}, \alpha\mathbf{v}\rangle} = \sqrt{\alpha\overline{\alpha}\langle \mathbf{v}, \mathbf{v}\rangle} = |\alpha| \cdot ||\mathbf{v}||$. For the final property of the norm, notice

$$||\mathbf{v}+\mathbf{u}||^2 = \langle \mathbf{v}+\mathbf{u}, \mathbf{v}+\mathbf{u}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{u}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle + 2\cdot\Re\langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{u}\rangle. \tag{2.1}$$

By the Cauchy-Schwarz inequality,

$$\Re\langle \mathbf{v}, \mathbf{u}\rangle \leq |\Re\langle \mathbf{v}, \mathbf{u}\rangle| \leq |\langle \mathbf{v}, \mathbf{u}\rangle| \leq ||\mathbf{v}|| \cdot ||\mathbf{u}||. \tag{2.2}$$

Inserting (2.2) into (2.1) gives us

$$||\mathbf{v}+\mathbf{u}||^2 \leq ||\mathbf{v}||^2 + 2||\mathbf{v}|| \cdot ||\mathbf{u}|| + ||\mathbf{u}||^2 = (||\mathbf{v}|| + ||\mathbf{u}||)^2, \tag{2.3}$$

from which we can retrieve the third property by taking the square root. ∎

If we return to our example inner product space of $\mathbb{R}^2$, the norm defined by the inner product is the Euclidean norm.

3. Orthogonality and Orthornormality

Now we can define orthogonality. Orthogonality is a generalization of perpendicularity to more abstract inner product spaces, like the ones we will work with shortly.

Definition 3.1 (orthogonal, orthonormal)**.** Let V be an inner product space. We say that vectors $\mathbf{v}, \mathbf{u} \in V$ are *orthogonal* if $\langle \mathbf{v}, \mathbf{u}\rangle = 0$. We say that a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq V$ is orthogonal if $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ for all $1 \leq i, j \leq n, i \neq j$. We say that a set is *orthonormal* if it is orthogonal and $\langle \mathbf{v}_i, \mathbf{v}_i\rangle = 1$ for all $\mathbf{v}_i \in S$.

Orthogonal sets are already very convenient to work with, but orthonormal sets are even nicer. Luckily, we can very easily convert an orthogonal set into an orthonormal one. If we start with an orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, then the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_i = \frac{1}{||\mathbf{v}_i||}\mathbf{v}_i$ is orthonormal, as $||\mathbf{e}_i|| = \frac{1}{||\mathbf{v}_i||}||\mathbf{v}_i|| = 1$. An important property of orthogonal sets is their linear independence.

Proposition 3.2. *Let $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n\}$ be orthogonal and $\mathbf{0} \notin S$. S is linearly independent.*

Proof. Let

$$c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots + c_n\mathbf{e}_n = \mathbf{0}. \tag{3.1}$$

If we take the inner product of both sides with $\mathbf{e}_1$, we will end up with

$$c_1\langle \mathbf{e}_1, \mathbf{e}_1\rangle + c_2 \cdot 0 + \cdots + c_n \cdot 0 = 0. \tag{3.2}$$

By Definition 2.1, $\langle \mathbf{e}_1, \mathbf{e}_1\rangle > 0$, so $c_1 = 0$. We can repeat this with every other $\mathbf{e}_i$ to show that $c_i = 0$ for all $1 \leq i \leq n$, which completes the proof. ∎

Another important property is that if we have some vector $\mathbf{v}$ in the span of some orthogonal set S, then we can decompose it into a linear combination of S using the inner product. We even used this fact in our last proof! We took the inner product of our vector with a vector $\mathbf{e}_i$ in our orthogonal set and found the coefficient of $\mathbf{e}_i$.

4. L^2 Space

Now we can begin to address abstract vector spaces, like function spaces. We will first start by defining Hilbert spaces.

Definition 4.1. A Hilbert space H is a complete inner product space.

By "complete", we mean that every Cauchy sequence in H converges in H. What this will end up meaning for our function space is that every function will have a finite norm. However, we have yet to define what our function space will actually be, so it would be helpful if we could do that.

Definition 4.2 (L^2 space)**.** We say that L^2 space is the collection of all functions $f : \mathbb{R} \to \mathbb{C}$ such that $\int_{-\infty}^{\infty} |f(x)|^2 \, dx$ is finite. We say that $g : [a, b] \to \mathbb{C}$ is in $L^2([a, b])$ if $\int_a^b |f(x)|^2 \, dx$ is finite.

L^2 is also known as the *space of square-integrable functions*. For our purposes, we will assume the integral defining L^2 is over an arbitrary bounded interval instead of over $\mathbb{R}$ when mentioning L^2 if no interval is specified.

Remark 4.3. The L in L^2 refers to "Lebesgue", as L^2 space is actually defined using Lebesgue integration, rather than the more intuitive Riemann integration. For nearly all functions used in daily life, like polynomials and trigonometric functions, Lebesgue and Riemann integration will evaluate equally, but the Lebesgue integral is required for the completeness of L^2. See [Dhu24] for more.

Right now, L^2 is just a set of functions with little structure, which is very uninteresting. We can create an inner product on L^2, which will give us the tools to consider more interesting properties of this space.

Proposition 4.4. *The map* $\langle \cdot, \cdot \rangle : L^2 \times L^2 \to \mathbb{C}$ *defined by* $\langle f, g \rangle = \int_a^b f(x)\overline{g(x)} \, dx$ *is an inner product on* L^2.

The norm induced by this inner product is

$$||f|| = \sqrt{\langle f, f \rangle} = \sqrt{\int_a^b f(x)\overline{f(x)} \, dx} = \sqrt{\int_a^b |f(x)|^2 \, dx},$$

so we could alternatively define L^2 as the set of all functions f such that $||f|| < \infty$. This cements L^2 as a Hilbert space. The reason we want L^2 to be Hilbert is because it will guarantee the convergence of sequences of functions that we will encounter in later sections.

We can begin discussing orthogonal functions now that we have a nice function space with an inner product. A common and easy to work with orthogonal set is $\{e^{inx} : n \in \mathbb{Z}\}$. We will soon see why this set is used so often shortly. First, we should show that this is actually orthogonal.

Proposition 4.5. $\{e^{inx} : n \in \mathbb{Z}\}$ *is orthogonal in* $L^2([-\pi, \pi])$.

Proof. First, through Euler's identity, notice $\overline{e^{inx}} = \overline{\cos(nx) + i\sin(nx)} = \cos(nx) - i\sin(nx) = e^{-inx}$. Now, for $n, m \in \mathbb{Z}$,

$$\langle e^{inx}, e^{imx} \rangle = \int_{-\pi}^{\pi} e^{inx}\overline{e^{imx}} \, dx = \int_{-\pi}^{\pi} e^{i(n-m)x} \, dx. \tag{4.1}$$

If $n \neq m$, then (4.1) will evaluate to 0, so $\langle e^{inx}, e^{imx} \rangle = 0$, which completes the proof. ∎

It should be noted that $\langle e^{inx}, e^{inx}\rangle = 2\pi$, so $\{e^{inx} : n \in \mathbb{Z}\}$ is not orthonormal. We could consider $\{\frac{e^{inx}}{\sqrt{2\pi}} : n \in \mathbb{Z}\}$, but another option is to redefine our inner product by adding a coefficient of $\frac{1}{2\pi}$, so that $||e^{inx}|| = 1$ naturally. Simple coefficients preserve all desired properties of inner products. When we continue to discuss $\{e^{inx} : n \in \mathbb{Z}\}$ later, we will assume the modified inner product, but it will made clear when we are using this modification.

Another example of an orthogonal set is the Legendre polynomials $\{P_n(x)\}$ in $L^2([-1,1])$. We can construct this set by starting out with $P_0(x) = 1$, and say that $P_n(x)$ is the unique polynomial in real coefficients of degree n such that $P_n(1) = 1$ and that $\langle P_n, P_i\rangle = 0$ for $0 \leq i \leq n-1$. The first few Legendre polynomials are $1, x, \frac{3}{2}x^2 - \frac{1}{2}, \frac{5}{2}x^3 - \frac{3}{2}x, \ldots$, up to P_3. By construction, the Legendre polynomials are orthogonal, but they don't span L^2 as $f(x) = i$ is always imaginary, and $P_n(x)$ will always be real.

5. The Fourier Series

Now we will return to the idea of decomposing functions using inner products. For the entirety of this section, when we refer to L^2, we will assume $L^2([-\pi,\pi])$ with the modified inner product such that $\{e^{inx} : n \in \mathbb{Z}\}$ is orthonormal, as this section will focus on this orthogonal set specifically.

Definition 5.1. (Fourier coefficients, Fourier series) Let $f \in L^2$. We say that $c_n = \langle f, e^{inx}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx$ is the n-th Fourier coefficient of f. We say that $\sum_{-\infty}^{\infty} c_n e^{inx}$ is the Fourier series of f. To declare the Fourier series of g we write

$$g(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx}.$$

To make examples easier to visualize, these series can be equivalently be written as linear combinations of $\sin(nx)$ and $\cos(nx)$ due to Euler's identity. This representation still may have complex coefficients, so we should see when our coefficients are real.

Proposition 5.2. *Let $f(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx}$ for some $f \in L^2$ and $c_{-n} = \overline{c_n}$ for all n. Then the Fourier series of f in sines and cosines has all real coefficients.*

Proof. Given our Fourier series $f(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx}$, we can say

$$\sum_{-\infty}^{\infty} c_n e^{inx} = c_0 + \sum_{1}^{\infty} c_n e^{inx} + \overline{c_n} e^{-inx} \tag{5.1}$$

$$= c_0 + \sum_{1}^{\infty} c_n e^{inx} + \overline{c_n e^{inx}} \tag{5.2}$$

$$= c_0 + \sum_{1}^{\infty} 2\,\mathfrak{R}(c_n e^{inx}) \tag{5.3}$$

$$= c_0 + \sum_{1}^{\infty} 2\,\mathfrak{R}(c_n \cos(nx) + ic_n \sin(nx)) \tag{5.4}$$

$$= c_0 \cos(0x) + \sum_{1}^{\infty} 2\,\mathfrak{R}(c_n)\cos(nx) + 2\,\mathfrak{R}(ic_n)\sin(nx) \tag{5.5}$$

Since $c_0 = \overline{c_0}$, $c_0 \in \mathbb{R}$, and so all coefficients are real. ∎

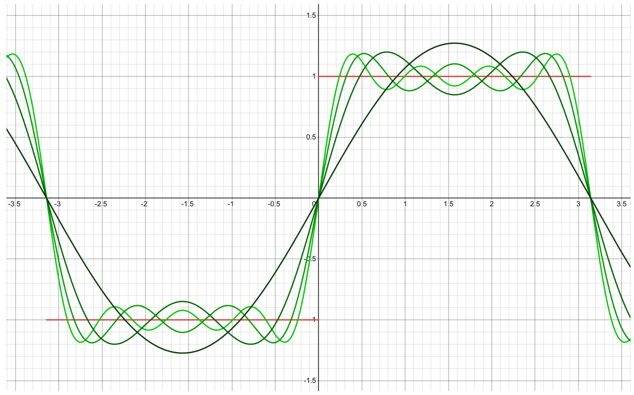

Figure 1. A graph of the square wave (red) and first four terms of its Fourier series expansion (green).

Now we will explore common examples of functions with Fourier series. The first one we will mention is the square wave, which is

$$f(x) = \begin{cases} 1 & 0 \leq x < \pi \\ -1 & -\pi \leq x < 0 \end{cases}$$

over our domain. Over $\mathbb{R}$, f would have a period of 2π. When we begin to calculate Fourier coefficients, we see

$$\begin{aligned} \langle f, e^{inx} \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx}\, dx \\ &= \frac{1}{2\pi} \int_{0}^{\pi} e^{-inx}\, dx - \frac{1}{2\pi} \int_{-\pi}^{0} e^{-inx}\, dx \\ &= \frac{1}{2\pi} \int_{0}^{\pi} i \sin(-nx)\, dx - \frac{1}{2\pi} \int_{-\pi}^{0} i \sin(-nx)\, dx. \end{aligned}$$

Notice how the cosines in Euler's identity are completely eliminated! What this means is that when we represent this Fourier series in sines and cosines, only sine terms will appear. When we finish evaluating every Fourier coefficient, we end up with

$$\frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\sin((2n-1)x)}{2n-1}$$

as the Fourier series of f. Figures 1 and 2 demonstrate varying degrees of this Fourier series.

We can look at our second example, the sawtooth wave, which is

$$g(x) = \frac{1}{\pi} x$$

over our interval. Over $\mathbb{R}$, the sawtooth wave would have a period of 2π, just like the square wave. Calculating the Fourier coefficients of g will, similarly to f, have the cosine terms of e^{inx} cancel out, meaning the Fourier series will once again contain only sine terms. The Fourier series of g evaluates to

$$-\frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \sin(nx).$$

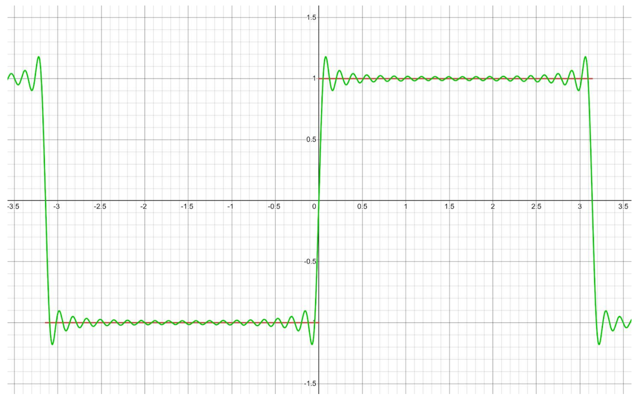

Figure 2. A graph of the square wave (red) and its Fourier series up to the first 20 terms (green).

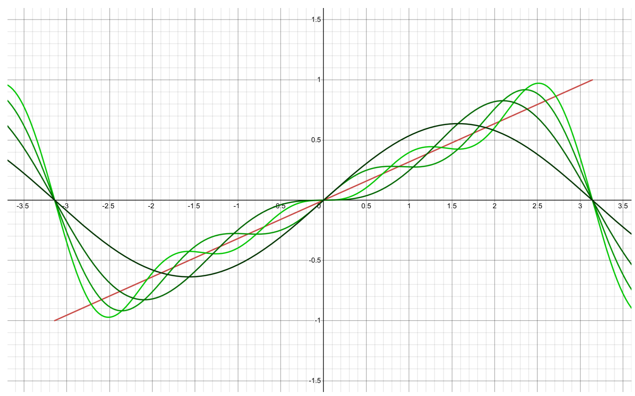

Figure 3. A graph of the sawtooth wave (red) and first four terms of its Fourier series expansion (green).

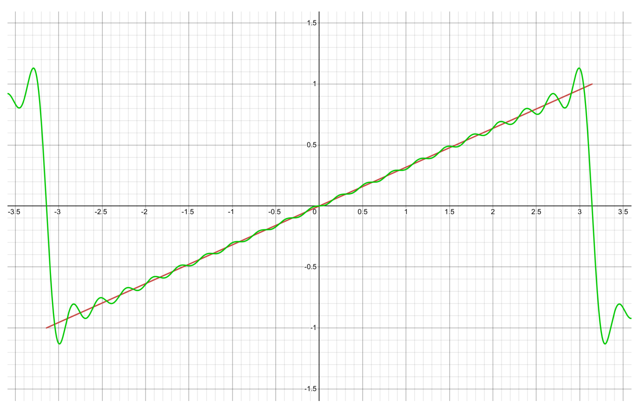

Figure 4. A graph of the sawtooth wave (red) and its Fourier series up to the first 20 terms (green).

Figures 3 and 4 demonstrate varying degrees of this Fourier series.

Remark 5.3. You may notice that even in our higher degree Fourier series, there are small spikes near the discontinuities of each wave. Initially, you may expect this to disappear after enough terms, but this isn't the case! These spikes of Fourier series at discontinuities is known as the Gibbs phenomenon. The spikes are approximately 9% of the length of the jump at the discontinuity. See [GS97] for more.

6. Fourier Series Relative to Other Systems

In the last section we defined Fourier series as retrieving coefficients of functions with respect to $\{e^{inx} : n \in \mathbb{Z}\}$. This process of extracting coefficients through inner products isn't unique to $\{e^{inx} : n \in \mathbb{Z}\}$, as we could theoretically do this exact same process with any orthogonal set of functions. And so, we will generalize the Fourier series process to any orthogonal.

Definition 6.1. Let $\{\phi_n\} \subseteq L^2$ be a countable set of orthonormal functions. The Fourier coefficients of f with respect to $\{\phi_n\}$ are $c_n = \langle f, \phi_n \rangle = \int_a^b f(x)\overline{\phi_n(x)}\, dx$ for some $f \in L^2$. We write

$$g(x) \sim \sum_{n=1}^{\infty} c_n \phi_n(x)$$

to be the Fourier series of g with respect to $\{\phi_n\}$ for some $g \in L^2$.

For this section, we will be using the Legendre polynomials $P_n(x)$. First, we will have to normalize them, and we will say $Q_n(x) = \frac{P_n(x)}{||P_n||}$ is our orthonormal set. Now, we can try to calculate our Fourier coefficients for the square wave (now with a period of 2 instead of 2π to fit within the bounds of our integral). Say the function f is our square wave. Then,

- $\langle f, Q_0 \rangle = \int_{-1}^{1} f(x)Q_0(x)\, dx = 0$
- $\langle f, Q_1 \rangle = \int_{-1}^{1} f(x)Q_1(x)\, dx \approx 1.2247$
- $\langle f, Q_2 \rangle = \int_{-1}^{1} f(x)Q_2(x)\, dx = 0$
- $\langle f, Q_3 \rangle = \int_{-1}^{1} f(x)Q_3(x)\, dx \approx -0.4677$

are our first few Fourier coefficients. The Fourier series of f with respect to $\{Q_n(x)\}$ up to Q_3 is depicted in Figure 5. It is not extremely impressive, but adding more terms will increase its accuracy. Figure 6 shows the same Fourier series up to Q_7, and the improvement is visually noticeable.

If we try this again now with our sawtooth wave, This Fourier series becomes very uninteresting, as the way we described the sawtooth would mean every coefficient would become 0 except for c_1, as our sawtooth is just Q_1 times some coefficient, so it would be orthogonal to every other Q_n.

7. Bessel's Inequality

After exploring Fourier series in the last two sections, we can move on to discussing properties of general Fourier series. The property we will focus on in this section is Bessel's Inequality.

Theorem 7.1 (Bessel's inequality)**.** *Let $\{\phi_n\}$ be orthonormal and $f(x) \sim \sum_{n=1}^{\infty} c_n\phi_n(x)$ for some $f \in L^2$. Then*

$$\sum_{n=1}^{\infty} |c_n|^2 \leq ||f||^2$$

.

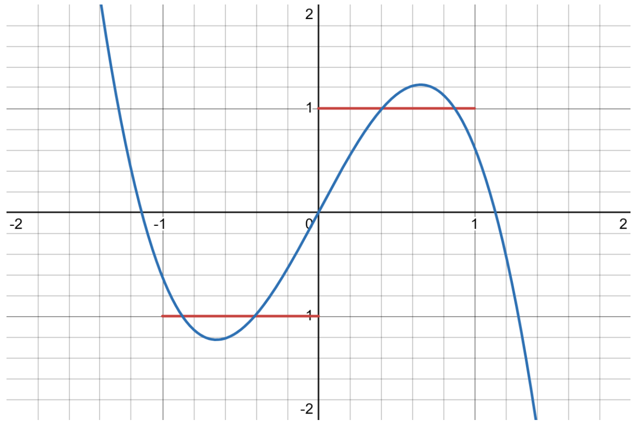

Figure 5. A graph of the square wave (red) and its Fourier series with respect to the normalized Legendre polynomials up to Q_3 (blue)

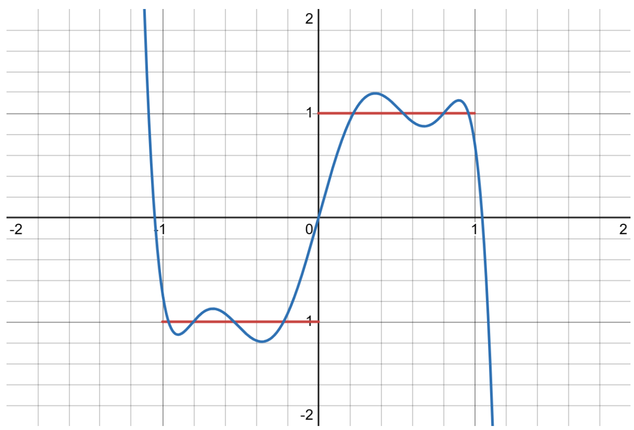

Figure 6. A graph of the square wave (red) and its Fourier series with respect to the normalized Legendre polynomials up to Q_7 (blue)

We will prove a case for partial sums first to help make this proof easier.

Lemma 7.2. *Let $\{\phi_n\}$ be orthonormal, $s_n(x) = \sum_{m=1}^{n} c_m\phi_n(x)$ be the nth partial sum of the Fourier series of f, and $t_n(x) = \sum_{m=1}^{n} d_m\phi_n(x)$ be some other linear combination of $\{\phi_n\}$. Then*

$$||f - s_n||^2 \leq ||f - t_n||^2$$

.

Proof. First, we see that

$$\langle f, t_n\rangle = \int_a^b f(x)\sum_{m=1}^{n}\overline{d_m\phi_m(x)}\,dx = \sum_{m=1}^{n}\overline{d_m}\int_a^b f(x)\overline{\phi_m(x)}\,dx = \sum_{m=1}^{n} c_m\overline{d_m} \tag{7.1}$$

through the definition of Fourier coefficients. Also,

$$\langle t_n, t_n\rangle = \int_a^b \sum_{m=1}^{n} d_m\phi_m(x)\sum_{m=1}^{n}\overline{d_m\phi_m(x)}\,dx = \sum_{m=1}^{n} d_m\overline{d_m} \tag{7.2}$$

through the orthonormality of $\{\phi_n\}$. Now,

$$\begin{aligned}||f-t_n||^2 &= \langle f-t_n, f-t_n\rangle \\ &= \langle f,f\rangle - \langle f,t_n\rangle - \langle t_n,f\rangle + \langle t_n,t_n\rangle \\ &= \langle f,f\rangle - \sum_{m=1}^{n} c_m\overline{d_m} - \sum_{m=1}^{n} d_m\overline{c_m} - \sum_{m=1}^{n} d_m\overline{d_m} \\ &= \langle f,f\rangle - \sum_{m=1}^{n}|c_m|^2 + \sum_{m=1}^{n}|d_m-c_m|^2.\end{aligned}$$

This means that $||f-t_n||^2$ is minimized when $d_m = c_m$. Letting $d_m = c_m$, we get

$$||f-t_n||^2 = ||f||^2 - \sum_{m=1}^{n}|c_m|^2 \tag{7.3}$$

Since $||f-t_n||^2 \geq 0$, we have

$$\sum_{m=1}^{n}|c_m|^2 \leq ||f||^2. \tag{7.4}$$

■

Now that we have this result, Bessel's inequality becomes very easy to prove.

Proof of Bessel's inequality. Let $n \to \infty$ in (7.4), we will get the desired inequality. ■

Since L^2 is a Hilbert space, Bessel's inequality has a nice result.

Corollary 7.3. *Let $\{\phi_n\}$ be orthonormal and $f(x) \sim \sum_{n=1}^{\infty} c_n\phi_n(x)$ for some $f \in L^2$. Then,*

$$\lim_{n\to\infty} c_n = 0.$$

8. Orthonormal Bases and Parseval's Identity

Throughout this entire paper, we have been assuming that our Fourier series will converge to our functions, when that isn't necessarily the case. For example, consider $\{e^{inx} : n \in \mathbb{N}\}$. This is a countable orthonormal subset of L^2, but if you attempt to take the Fourier series of e^{-ix} with respect to that set, you will get 0, so it won't always converge back. However, how do we know that the Fourier series with respect to $\{e^{inx} : n \in \mathbb{Z}\}$ will? We will now show that the convergence of the Fourier series of f with respect to $\{\phi_n\}$ to f is guaranteed if linear combinations of $\{\phi_n\}$ is dense in L^2.

Theorem 8.1. *Let $\{\phi_n\}$ be orthonormal and linear combinations of $\{\phi_n\}$ be dense in L^2. Let $f(x) \sim \sum_{n=1}^{\infty} c_n\phi_n(x)$ for some $f \in L^2$. Then,*

$$\lim_{n\to\infty} ||f - s_n(f)||^2 = 0,$$

where $s_n(f)$ denotes the nth partial sum of the Fourier series of f.

Proof. Let $\varepsilon > 0$. By the completeness of L^2, we can find some $h \in L^2$ such that $h \neq f$ and

$$||f-h|| < \frac{\varepsilon}{3}. \tag{8.1}$$

By the density of the span of $\{\phi_n\}$ in L^2, there is some linear combination of $\{\phi_n\}$ which we will call P such that $||h - P|| < \frac{\varepsilon}{3}$. Then, say P has degree N (the latest ϕ_n used is ϕ_N). Then,

$$||h - s_n(h)|| \leq ||h - P|| < \frac{\varepsilon}{3} \tag{8.2}$$

for $n \geq N$. Now,

$$||s_n(h) - s_n(f)|| = ||s_n(h - f)|| \leq ||h - f|| < \frac{\varepsilon}{3}. \tag{8.3}$$

Now combining the triangle inequality and (8.1), (8.2), and (8.3), we get

$$||f - s_n(f)|| < \varepsilon \tag{8.4}$$

for $n \geq N$, which completes the proof. ∎

This property is known as the *completeness of* $\{\phi_n\}$. Completeness of $\{\phi_n\}$ allows it to function similarly to a basis. Another fact that relies on the density of the span of $\{\phi_n\}$ is Parseval's identity, which is a generalization of the Pythagorean identity.

Theorem 8.2 (Parseval's identity)**.** *Let* $\{\phi_n\}$ *be orthonormal and linear combinations of* $\{\phi_n\}$ *be dense in* L^2. *Let* $f(x) \sim \sum_{n=1}^{\infty} c_n \phi_n(x)$. *Then,*

$$||f||^2 = \sum_{n=1}^{\infty} |c_n|^2.$$

Proof. First, notice that

$$\langle s_n(f), f \rangle = \sum_{m=1}^{n} c_m \langle \phi_m, f \rangle = \sum_{m=1}^{n} |c_m|^2 \tag{8.5}$$

and by the Cauchy-Schwarz inequality,

$$|||f||^2 - \langle s_n(f), f \rangle| = |\langle f - s_n(f), f \rangle| \leq ||f - s_n(f)|| \cdot ||f||. \tag{8.6}$$

We already know that the right hand side of (8.6) tends to 0 as $n \to \infty$, so by inserting (8.5) into the left hand side of (8.6), we get

$$|||f||^2 - \sum_{m=1}^{n} |c_m|^2| \to 0 \tag{8.7}$$

as $n \to \infty$, completing the proof. ∎

9. The Shannon-Nyquist Sampling Theorem

In the previous sections, we have discussed how and when Fourier series can be used to break down functions into its components, but could the same techniques be used to reconstruct an original function? Capabilites such as this have extremely useful applications, as minimizing lost data is crucial. Here, we will explore how much information of a function is needed to recover it from its Fourier series.

The original Shannon-Nyquist sampling theorem states the following:

Theorem 9.1 (Shannon-Nyquist)**.** *If a function* $f \in L^2$ *contains no frequencies higher than* B*, Then* f *can be completely determined by sampled points spaced less than* $\frac{1}{2B}$ *apart.*

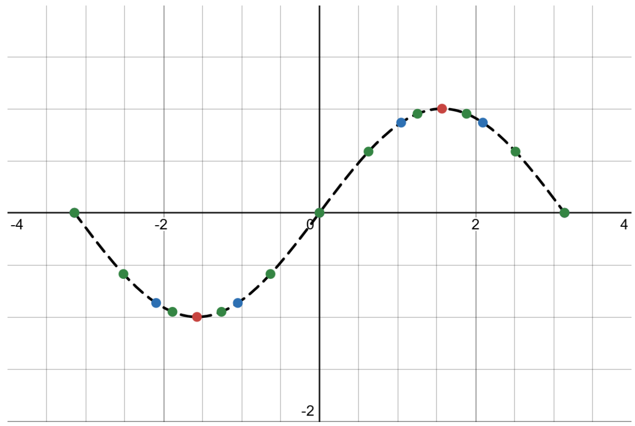

Figure 7. A sine wave sampled at rates of $\frac{\pi}{2}$ (red), $\frac{\pi}{3}$ (blue), and $\frac{\pi}{5}$ (green).

Informally, this states that for a function with frequencies less than B, then it can be reconstructed by samples taken at a rate of $\frac{1}{2B}$. This version of the statement requires use of the Fourier transform instead of the Fourier series. Unlike the series, the Fourier transform cannot be easily generalized to any orthogonal set of functions, so what we will work with is a slightly modified and weaker version of the theorem. Similar to earlier in this paper, we will introduce a case involving $\{e^{inx} : n \in \mathbb{Z}\}$ as our orthonormal set, and then expand from there.

Theorem 9.2 (Shannon-Nyquist, modified)**.** *If a function $f \in L^2$ with $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$ has $c_N = 0$ for all $|N| \geq B$, then the Fourier series of f can be completely determined by sampled points spaced less than $\frac{\pi}{B}$ apart.*

Remark 9.3. The Fourier transform can be informally described as turning a function of time into a function of frequency. If f is our original function, then $\hat{f}$ is the function after the Fourier transform. $\hat{f}$ is defined for all real numbers, while ϕ_n is only defined for natural n. This is where the trouble in generalization comes in. The extra factor of 2π in the modified theorem originates from the fact the Fourier transform has coefficients of 2π in its exponentiation. See Shannon's original paper [Sha49] for the usage of the transform.

Before we can begin work, we never covered what we meant by sampling. When we take a sample of some function f, we simply take the value $f(x)$ at some point x in the domain of f. When we take sample at different rates, we will end up taking more or less values of f depending on how we change the sampling interval. Figure 7 demonstrates the effects of differing sampling rates. With this in mind, Shannon-Nyquist essentially states that more complex functions require more samples in order to prevent loss.

Proof of Theorem 9.2. Let $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$ be such that f satisfies the statement of Theorem 9.2. Then,

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx} = \sum_{n=-B+1}^{B-1} c_n e^{inx}$$

as $c_N = 0$ for all $|N| \geq B$. If we take our samples at a rate of $\frac{\pi}{B}$, then we will have

$$f\left(\frac{\pi m}{B}\right) = \sum_{n=-B+1}^{B-1} c_n e^{in\frac{\pi m}{B}},$$

which gives us a system of equations in $2B-1$ variables $c_{-B+1}, \ldots, c_{B-1}$ and $2B-1$ unique equations as there are $2B-1$ possible values of $m \in \mathbb{Z}$ such that $\frac{\pi m}{B} \in [-\pi, \pi]$. This lets us determine the coefficients $c_{-B+1}, \ldots, c_{B-1}$ fully. Since every other coefficient of the Fourier series of f is 0, we have determined $\sum_{n=-\infty}^{\infty} c_n e^{inx}$, as desired. ■

You might have noticed that we require that if we want to determine some function f with a sampling rate of $\frac{\pi}{B}$, then $c_B, c_{-B} = 0$ must be the case in order to be able to reconstruct f. This requirement is necessary for reconstruction. You won't be capable of telling apart two functions with a sample rate of $\frac{\pi}{B}$ if they both have a highest frequency of B. As an example, consider $\cos(x)$ and $\cos(x) - \sin(x)$. These both have a highest frequency of 1, but if we attempt to take samples at a rate of π, we see that both sets of samples are identical, necessitating a higher sampling rate. An example of this can be found in Figure 8.

Now we can generalize this theorem to any orthogonal set. The statement and proof are quite similar, so not much should change.

Theorem 9.4. *Let $\{\phi_n\}$ be orthonormal in $L^2([a,b])$ If a function $f \in L^2$ with $f(x) \sim \sum_{n=-\infty}^{\infty} c_n \phi_n(x)$ has $c_N = 0$ for all $|N| \geq B$, then the Fourier series of f with respect to $\{\phi_n\}$ can be completely determined by sampled points spaced less than $\frac{b-a}{2B}$ apart.*

Proof. Let $f(x) \sim \sum_{n=-\infty}^{\infty} c_n \phi_n(x)$ be such that f satisfies the statement of Theorem 9.4. Then,

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n \phi_n(x) = \sum_{n=-B+1}^{B-1} c_n \phi_n(x)$$

as $c_N = 0$ for all $|N| \geq B$. If we take our samples at a rate of $\frac{b-a}{2B}$, then we will have

$$f\left(a + \frac{(b-a)m}{2B}\right) = \sum_{n=-B+1}^{B-1} c_n \phi_n \left(a + \frac{(b-a)m}{2B}\right),$$

which gives us a system of equations in $2B-1$ variables $c_{-B+1}, \ldots, c_{B-1}$ and $2B-1$ unique equations as there are $2B-1$ possible values of $m \in \mathbb{Z}$ such that $a + \frac{(b-a)m}{2B} \in [a,b]$. This lets us determine the coefficients $c_{-B+1}, \ldots, c_{B-1}$ fully. Since every other coefficient of the Fourier series of f is 0, we have determined $\sum_{n=-\infty}^{\infty} c_n \phi_n(x)$, as desired. ■

10. Legendre Polynomials

Throughout this paper, we have introduced different use cases of different orthogonal sets. However, we usually had to start by defining a tool for $\{e^{inx} : n \in \mathbb{Z}\}$, and then generalize to any orthonormal set $\{\phi_n\}$. Naturally, one may wonder why we need the generalized tools if the original ones all function well. What properties could $\{\phi_n\}$ have to be used over $\{e^{inx} : n \in \mathbb{Z}\}$? In this section, we will further discuss special properties of the Legendre polynomials $\{P_n\}$.

Definition 10.1. Consider the differential equation

$$(1-x^2)y'' - 2xy' + \lambda y = 0 \tag{10.1}$$

for $x \in [-1, 1]$ and some $\lambda \in \mathbb{R}$. We will call the polynomial solutions to this $\{R_n\}$.

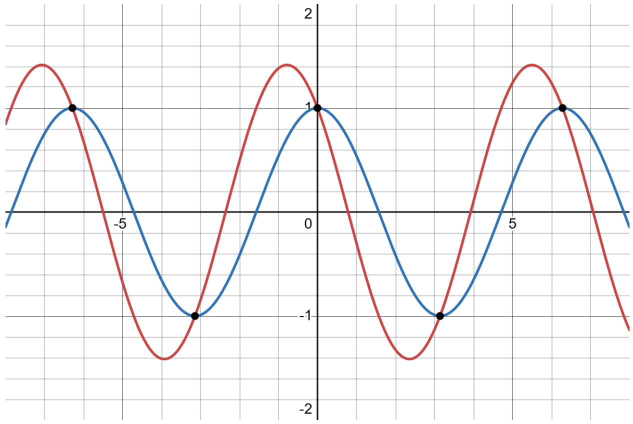

Figure 8. $\cos(x)$ (blue) and $\cos(x) - \sin(x)$ (red) being sampled at a rate of π. Both sample sets are identical.

We want to say that $\{R_n\}$ is the same as $\{P_n\}$, but doing so right now would be absurd. A crucial aspect of the definition of $\{P_n\}$ is orthogonality. We'll be moving in the right place to start with proving the orthogonality of $\{R_n\}$.

Proof. First, notice that $(1-x^2)y'' - 2xy' + \lambda y = 0$ can be rewritten as

$$((1-x^2)y')' + \lambda y = 0,$$

and $\{R_n\}$ will still satisfy the equation (with the correct lambda value). Now, consider distinct $R_n, R_m \in \{R_n\}$ with corresponding distinct λ_n, λ_m. From this we see that

$$\lambda_n R_n(x) = ((x^2-1)R_n'(x))'$$
$$\lambda_m R_m(x) = ((x^2-1)R_m'(x))'$$

is satisfied. Now, if we multiply the first equation by R_m, the second equation by R_n, and subtract the two, we get

$$\begin{aligned}(\lambda_n - \lambda_m)R_n(x)R_m(x) &= ((x^2-1)R_n'(x))'R_m(x) - ((x^2-1)R_m'(x))'R_n(x)\\ &= ((x^2-1)R_n'(x)R_m(x) - (x^2-1)R_m'(x)R_n(x))'\\ &= ((x^2-1)(R_n'(x)R_m(x) - R_m'(x)R_n(x)))'.\end{aligned}$$

Now, when we take the inner product, we have

$$\begin{aligned}\langle R_n, R_m\rangle &= \int_{-1}^{1} R_n(x)\overline{R_m(x)}\,dx\\ &= \frac{1}{\lambda_n - \lambda_m}\int_{-1}^{1}(\lambda_n - \lambda_m)R_n(x)R_m(x)\,dx\\ &= \frac{1}{\lambda_n - \lambda_m}[(x^2-1)(R_n'(x)R_m(x) - R_m'(x)R_n(x))]_{-1}^{1}\\ &= 0,\end{aligned}$$

which completes the proof. ■

We still aren't done here however. This proof assumes that for two different R_n and R_m, they have two distinct λ_n and λ_m such that they satisfy the original equation. If $\lambda_n = \lambda_m$, then we would end up dividing by 0. Even worse, if R_n solves (10.1), so does KR_n for any constant K. Another issue one might take is that we give an order to the polynomials $R_0, R_1, R_2, \ldots$, but never define it. Rodrigues' Formula will help us address all of those issues and link $\{R_n\}$ with $\{P_n\}$.

Theorem 10.2 (Rodrigues' Formula)**.** *For $n \in \mathbb{N}$, we have*

$$R_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}[(x^2-1)^n],$$

which solves (10.1) *with $\lambda = n(n+1)$.*

Proof. Let $y = (x^2-1)^n$. Now we claim that for the k-th derivative of y $y^{(k)}(x)$,

$$(1-x^2)\frac{d^2y^{(k)}}{dx^2} + 2(n-k-1)x\frac{dy^{(k)}}{dx} + (2n-k)(k+1)y^{(k)} = 0 \tag{10.2}$$

holds true. We will show this through induction. We start with $k = 0$, where $y^{(0)} = y$. We have

$$y' = 2nx(x^2-1)^{n-1},$$

which implies

$$(1-x^2)^{n-1}y' + 2nxy = 0.$$

After differentiation, we get

$$(1-x^2)y'' + 2(n-1)xy' + 2nxy = 0,$$

so (10.2) holds for the $k = 0$ case. Now, assume (10.2) holds for up to $k-1$. We can then rewrite (10.2) for $k-1$ as

$$(1-x^2)\frac{dy^{(k)}}{dx} + 2(n-k)xy^{(k)} + (2n-k+1)ky^{(k-1)} = 0,$$

which is true due to the inductive hypothesis. If we differentiate this, we will get (10.2) back exactly. Now that we have proven (10.2), we let $k = n$, and obtain

$$(1-x^2)\frac{d^2y^{(n)}}{dx^2} - 2x\frac{dy^{(n)}}{dx} + n(n+1)y^{(n)} = 0,$$

which solves (10.1) with $\lambda = n(n+1)$. Since $y^{(n)}$ is a polynomial of degree n, is orthogonal to all $\frac{d^k}{dx^k}[(x^2-1)^k]$ for all $0 \leq k \leq n-1$ (as they will have distinct λ values), and that letting $k = 0$ will obviously give some multiple of the constant function, we can deduce that $y^{(n)}$ is multiple of P_n. Using the fact that the leading coefficient c_n of P_n is $\frac{(2n!)}{2^n(n!)^2}$, we can find some value K such that $Ky^{(n)}(x) = P_n(x)$. The coefficient of x^n in $y^{(n)}$ can be found by

$$\frac{d^n(x^{2n})}{dx^n} = (2n)(2n-1)\cdots(2n-n+1)x^n = \frac{(2n)!}{n!}x^n,$$

so by simply solving for K, we get

$$K = \frac{(2n)!n!}{2^n(n!)^2(2n)!} = \frac{1}{2^n n!},$$

which completes the proof. ∎

Remark 10.3. This proof uses the fact that the leading coefficient of P_n is $\frac{(2n!)}{2^n(n!)^2}$. This fact relies on usage of generating functions, which is beyond this paper. See [Lim22] for more on Legendre polynomials.

Now we can be satisfied with $\{P_n\}$ solving (10.1), which is can make the Legendre polynomials useful.

Acknowledgments

I would like to thank Jay Paek for his helpful contributions and Simon Rubinstein-Salzedo for the opportunity to write this paper through Euler Circle.

References

[Dhu24] Sanjeev Dhurandhar. Hilbert spaces. In *Understanding Mathematical Concepts in Physics: Insights from Geometrical and Numerical Approaches*, pages 23–51. Springer, 2024.

[GS97] David Gottlieb and Chi-Wang Shu. On the gibbs phenomenon and its resolution. *SIAM review*, 39(4):644–668, 1997.

[Lim22] Fábio MS Lima. Lecture notes on legendre polynomials: their origin and main properties. *arXiv preprint arXiv:2210.10942*, 2022.

[Poi08] Simeon Denis Poisson. Mémoire sur la propagation de la chaleur dans les corps solides. *Nouveau Bulletin des Sciences par la Société philomatique de Paris, tI*, pages 112–116, 1808.

[Sha49] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.