

# 1 The Weierstrass Approximation Theorem

The fact that the Fourier Series approximates any periodic function, and that the corresponding basis is complete can be proved in multiple ways. One of these ways makes use of the Weierstrass approximation theorem. Here, I will prove a very important lemma required to prove the Weierstrass approximation theorem.

**Theorem 1** (Weierstrass approximation theorem). *If  $[a, b]$  is an interval,  $f : [a, b] \rightarrow \mathbf{R}$  is a continuous function, and  $\varepsilon > 0$ , then there exists a polynomial  $P$  on  $[a, b]$  such that  $|P(x) - f(x)| \leq \varepsilon$  for all  $x \in [a, b]$ .*

**Definition 2** (Approximation to the identity). Let  $\varepsilon > 0$  and  $0 < \delta < 1$ . A function  $f : \mathbf{R} \rightarrow \mathbf{R}$  is said to be an  $(\varepsilon, \delta)$ -approximation to the identity if it obeys the following three properties:

- (a) If  $x \notin [-1, 1]$ , then  $f(x) = 0$ ; and  $f(x) \geq 0$  for all  $x \in [-1, 1]$ .
- (b)  $f$  is continuous, and  $\int_{-\infty}^{\infty} f = 1$ .
- (c)  $|f(x)| \leq \varepsilon$  for all  $\delta \leq |x| \leq 1$ .

*Remark.*  $\int_{-\infty}^{\infty} f = \int_{-1}^1 f$ , since  $f(x) = 0$  whenever  $x \in (-\infty, -1) \cup (1, \infty)$ .

The lemma that we want to prove is given below which provides the error bound between a continuous function and its convolution with a function that is an approximation to the identity. This can then be combined with other results to prove the Weierstrass approximation theorem.

**Lemma 3.** *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a continuous function such that  $x \notin [0, 1] \implies f(x) = 0$ , which is bounded by some  $M > 0$  (i.e.,  $|f(x)| \leq M$  for all  $x \in \mathbf{R}$ ), and let  $\varepsilon > 0$  and  $0 < \delta < 1$  be such that one has  $|f(x) - f(z)| < \varepsilon$  whenever  $x, z \in \mathbf{R}$  and  $|x - z| < \delta$ . Let  $g$  be any  $(\varepsilon, \delta)$ -approximation to the identity. Then we have*

$$|f * g(x) - f(x)| \leq (1 + 4M)\varepsilon$$

for all  $x \in [0, 1]$ , where  $f * g(x)$  is called the convolution of  $f$  and  $g$  and is defined as

$$\int_{-\infty}^{\infty} f(x - y)g(y)dy.$$

*Proof.* We start with observing that  $\int_{-\infty}^{\infty} f(x - y)g(y)dy = \int_{-1}^1 f(x - y)g(y)dy$ .

$$\begin{aligned} |f * g(x) - f(x)| &= \left| \int_{-1}^1 f(x - y)g(y)dy - f(x) \right| \\ &= \left| \int_{-1}^1 f(x - y)g(y)dy - \int_{-1}^1 f(x)g(y)dy \right| \\ &\quad \text{(because of property (b) of Def. 2)} \\ &\leq \int_{-1}^1 |f(x - y) - f(x)|g(y)dy \end{aligned}$$

Now, we have to split the integral as follows.

$$\begin{aligned} I &= \int_{-1}^1 |f(x - y) - f(x)|g(y)dy \\ &= \int_{-1}^{-\delta} |f(x - y) - f(x)|g(y)dy + \int_{-\delta}^{\delta} |f(x - y) - f(x)|g(y)dy + \int_{\delta}^1 |f(x - y) - f(x)|g(y)dy \end{aligned}$$

Let's call the three integrals on the RHS  $I_1$ ,  $I_2$ , and  $I_3$  respectively.

Let's first look at  $I_2$ . Observe that

P1: within this integral, the variable of integral  $y$  ranges from  $-\delta$  to  $\delta$ , i.e.,  $|y| \leq \delta$ .

P2:  $|x - (x - y)| = |y|$ .

From P1 and P2, and from our assumption, we know that  $|f(x - y) - f(x)| < \varepsilon$  (take  $z = x - y$ ).

$$\begin{aligned} I_2 &\leq \int_{-\delta}^{\delta} \varepsilon g(y) dy \\ &= \varepsilon \int_{-\delta}^{\delta} g(y) dy \\ &\leq \varepsilon \quad (\text{because the maximum value of } \int_{-\delta}^{\delta} g(y) dy \text{ is } 1) \end{aligned}$$

Now, let's look at  $I_3$ . Observe that

P3: We cannot bound  $|f(x - y) - f(x)|$  by  $\varepsilon$  this time. But observe the farthest apart these two values can be is  $2M$  if one of them is  $M$  and the other is  $-M$  since  $|f(x)| \leq M$ , by assumption. This means that  $|f(x - y) - f(y)| \leq 2M$ .

P4:  $g(y) \leq \varepsilon$  by property (c) of Def. 2.

From P3 and P4 we can deduce the following.

$$\begin{aligned} I_3 &\leq \int_{\delta}^1 2M\varepsilon dy \\ &= 2M\varepsilon \int_{\delta}^1 dy = 2M\varepsilon(1 - \delta) \\ &\leq 2M\varepsilon \quad (\text{because } 0 < \delta \implies 1 - \delta < 1 \implies 2M\varepsilon \cdot (1 - \delta) < 2M\varepsilon \cdot 1) \end{aligned}$$

With similar arguments, we can say that  $I_1 \leq 2M\varepsilon$ . Now, let's combine the three inequalities back.

$$\begin{aligned} I &= I_1 + I_2 + I_3 \\ &\leq 2M\varepsilon + \varepsilon + 2M\varepsilon \\ &= (1 + 4M)\varepsilon \end{aligned}$$

□

## 2 Polynomials Can be Approximations to The Identity

The Weierstrass approximation theorem follows from Lemma 3 by setting  $\varepsilon$  as small as desired and finding an appropriate  $(\varepsilon, \delta)$ -approximation to the identity which makes  $f * g$  a polynomial. This is easier than it looks because of the fact that if  $g$  is a polynomial, then so is  $f * g$  (the domain may be different, which is why we have to carefully choose the domains of  $f$  and  $g$ ).

We will prove this fact later, but we also need to prove that polynomials can be approximations to the identity. This is in fact true, and is stated as the following lemma.

**Lemma 4** (Polynomials can approximate the identity). *For every  $\varepsilon > 0$  and  $0 < \delta < 1$  there exists an  $(\varepsilon, \delta)$ -approximation to the identity which is a polynomial  $P$  on  $[-1, 1]$ .*

The proof requires several intermediate facts, which we will detail out here.

**Lemma 5** (Bernoulli's inequality). *Let  $y \in \mathbf{R}$  and  $0 < y < 1$ , and  $n \in \mathbf{N}$  with  $n \geq 0$ ,*

$$(1 - y)^n \geq 1 - ny.$$

*Proof.* Use induction. □

**Lemma 6.** *Let  $x \in \mathbf{R}$  and  $n \in \mathbf{N}$  with  $n \geq 1$ , then*

$$\int_{-1}^1 (1 - x^2)^n dx \geq \frac{1}{\sqrt{n}}$$

*Proof.* Notice that the integral is symmetric, and thus equal to  $2 \int_0^1 (1 - x^2)^n dx$ . Now, we can split the interval of integration  $[0, 1]$  into  $[0, 1/\sqrt{n}]$  and  $[1/\sqrt{n}, 1]$ . We will label the integrals over these intervals  $I_1$  and  $I_2$  respectively.

$$I = \int_0^1 (1 - x^2)^n dx = \int_0^{1/\sqrt{n}} (1 - x^2)^n dx + \int_{1/\sqrt{n}}^1 (1 - x^2)^n dx = I_1 + I_2$$

Now, we will look at each integrals separately. We will need the fact that  $x \geq 1/\sqrt{n} \implies x^2 \geq 1/n$  for  $I_2$ .

$$\begin{aligned} I_2 &= \int_{1/\sqrt{n}}^1 (1 - x^2)^n dx \\ &\geq \int_{1/\sqrt{n}}^1 (1 - nx^2) dx \\ &\geq \int_{1/\sqrt{n}}^1 \left(1 - n \left(\frac{1}{n}\right)\right) dx \\ &= 0 \end{aligned}$$

Let's look at  $I_1$  now.

$$\begin{aligned} I_1 &= \int_0^{1/\sqrt{n}} (1 - x^2)^n dx \\ &\geq \int_0^{1/\sqrt{n}} (1 - nx) dx \\ &= \left[ x - \frac{nx^2}{2} \right]_0^{1/\sqrt{n}} \\ &= \frac{1}{\sqrt{n}} - \frac{1}{2\sqrt{n}} \\ &= \frac{1}{2\sqrt{n}} \end{aligned}$$

Combining the two gives the following.

$$\begin{aligned} I &= I_1 + I_2 \geq \frac{1}{2\sqrt{n}} \\ \implies 2 \int_0^1 (1 - x^2)^n dx &\geq \frac{1}{\sqrt{n}} \geq \frac{1}{\sqrt{n}} \end{aligned}$$

□

*Proof of Lemma 4.* This is going to be a constructive proof. We will show how to construct one polynomial which approximates the identity. First, choose  $P$  of the form

$$P(x) = \begin{cases} c(1 - x^2)^N, & x \in [-1, 1] \\ 0, & \text{otherwise} \end{cases}$$

where  $c \in \mathbf{R}$  has to be chosen so that  $\int_{-1}^1 P = 1$ . This can be done since  $\int_{-1}^1 (1 - x^2)^n dx$  is positive. It is now trivial to verify that this  $P$  will satisfy properties (a) and (b) of Def. 2. We just need to verify property (c), and this is where  $N \in \mathbf{N}$  will come into play. We have used  $N$  to suggest that  $N$  will be a large natural number, and we will need to take a large natural number for  $N$ .

We will split our analysis into two halves, that is instead of talking about when  $x \in [-1, 1]$  we will analyze when  $x \in [0, 1]$  and when  $x \in [-1, 0]$ , because the arguments are symmetric. First, note that  $x^2$  is an increasing function of  $x$  in  $[0, 1]$ , that is if  $x_0 \leq x_1$  then  $x_0^2 \leq x_1^2$ . So, once we fix the value of  $\delta$ , we know that  $\delta^2 \leq x^2$  for any  $x \in [\delta, 1]$ . This means that  $c(1 - \delta^2)^N \geq c(1 - x^2)^N$  ( $c$  is positive). So, instead of bounding  $c(1 - x^2)^N \leq \varepsilon$ , we will bound  $c(1 - \delta^2)^N \leq \varepsilon$  by setting an appropriate value of  $N$ . To bound this, just choose a large  $N$  such that  $c(1 - \delta^2)^N = \varepsilon$ . This happens precisely when  $N = (\log \varepsilon - \log c) / \log(1 - \delta^2)$ .

To summarize, with the above form of  $P$  and given any  $\varepsilon > 0$  and  $0 < \delta < 1$ , we can always choose  $c \in \mathbf{R}$  and  $N \in \mathbf{N}$  such that  $P$  is an approximation to the identity.  $\square$

*Remark.*  $P$  above is an approximation to the Dirac delta function. It is beneficial to plot the function in a graphing calculator like Desmos or GeoGebra.

### 3 Convolution with Polynomials Give Polynomials

**Definition 7** (Supported functions). Let  $[a, b]$  be an interval. A function  $f : \mathbf{R} \rightarrow \mathbf{R}$  is said to be *supported* on  $[a, b]$  iff  $f(x) = 0$  for all  $x \notin [a, b]$ .

**Lemma 8.** Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a continuous function supported on  $[0, 1]$ , and let  $g : \mathbf{R} \rightarrow \mathbf{R}$  be a continuous function supported on  $[-1, 1]$  which is a polynomial on  $[-1, 1]$ . Then  $f * g$  is a polynomial on  $[0, 1]$ . (Note however that it may be non-polynomial outside  $[0, 1]$ .)

*Proof.* Since  $g$  is a polynomial on  $[-1, 1]$ , we may find an integer  $n \geq 0$  and real numbers  $c_0, c_1, \dots, c_n$  such that

$$g(x) = \sum_{j=0}^n c_j x^j \text{ for all } x \in [-1, 1].$$

On the other hand, for all  $x \in [0, 1]$ , we have

$$f * g(x) = \int_{-\infty}^{\infty} f(y)g(x - y)dy = \int_{[0,1]} f(y)g(x - y)dy$$

since  $f$  is supported on  $[0, 1]$ . Since  $x \in [0, 1]$  and the variable of integration  $y$  is also on  $[0, 1]$ , we have  $x - y \in [-1, 1]$ . Thus we may substitute in our formula for  $g$  to obtain

$$f * g(x) = \int_{[0,1]} f(y) \sum_{j=0}^n c_j (x - y)^j dy.$$

We expand this using the binomial formula to obtain

$$f * g(x) = \int_{[0,1]} f(y) \sum_{j=0}^n c_j \sum_{k=0}^j \frac{j!}{k!(j-k)!} x^k (-y)^{j-k} dy.$$

We can interchange the two summations to obtain

$$f * g(x) = \int_{[0,1]} f(y) \sum_{k=0}^n \sum_{j=k}^n c_j \frac{j!}{k!(j-k)!} x^k (-y)^{j-k} dy.$$

The limits of the summation change, and a justification for this is given at the end of the proof. Now we interchange the  $k$  summation with the integral, and observe that  $x^k$  is independent of  $y$ , to obtain

$$f * g(x) = \sum_{k=0}^n x^k \int_{[0,1]} f(y) \sum_{j=k}^n c_j \frac{j!}{k!(j-k)!} (-y)^{j-k} dy.$$

If we define

$$C_k \equiv \int_{[0,1]} f(y) \sum_{j=k}^n c_j \frac{j!}{k!(j-k)!} (-y)^{j-k} dy$$

for each  $k = 0, \dots, n$ , then  $C_k$  is a number which is independent of  $x$ , and we have

$$f * g(x) = \sum_{k=0}^n C_k x^k$$

for all  $x \in [0, 1]$ . Thus  $f * g$  is a polynomial on  $[0, 1]$ .

The (informal) justification for why the limits of the summation change as they do can be seen by looking at the following table for an example value of  $n = 5$ .

| $j/k$ | 0 | 1 | 2 | 3 | 4 | 5 |
|-------|---|---|---|---|---|---|
| 0     | • | • | • | • | • | • |
| 1     |   | • | • | • | • | • |
| 2     |   |   | • | • | • | • |
| 3     |   |   |   | • | • | • |
| 4     |   |   |   |   | • | • |
| 5     |   |   |   |   |   | • |

We sum over the  $(j, k)$  combinations indicated by •

In other words, the two sums  $\sum_{j=0}^n \sum_{k=0}^j$  can be written as a single sum over all  $(j, k)$  pairs satisfying  $j \leq k$ , i.e.  $\sum_{0 \leq j \leq k \leq n}$ , which can then be re-written as  $\sum_{k=0}^n \sum_{j=k}^n$ .  $\square$

## 4 How to Use Weierstrass Approximation Theorem for Fourier Analysis

We wanted to prove that the Fourier Series approximates any periodic function. “The Fourier Series of a periodic function  $f$ ” is to  $f$  what is “a sequence of rational numbers that approximate an irrational number  $x$ ” to  $x$ . An alternative view of the Fourier Series that we will be interested in arises from looking at a Hilbert space of functions. A Hilbert space is a vector space with an inner product such that the vector space is also complete with respect to the metric induced by the inner product, i.e. all Cauchy Sequence of vectors converge to limits that are within the space itself. The rational numbers are not complete because some converging sequences of rational numbers may converge to an irrational number. The real numbers are complete because the limits of a sequence of all real numbers (rational or irrational) is again a real number.

The space we will be looking at is called  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  and is the space of all functions  $f$  such that:

1.  $f$  is continuous and  $f : \mathbf{R} \rightarrow \mathbf{C}$ , and
2.  $f$  is periodic with a period of 1, i.e.  $f(x + 1) = f(x) \forall x \in \mathbf{R}$ .

The inner product defined in this space is  $\langle f, g \rangle = \int_{[0,1]} f(x) \overline{g(x)} dx$  for  $f, g \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  where  $\overline{g(x)}$  is the complex conjugate of  $g(x)$ . In fact, the number  $\sqrt{\langle f, f \rangle}$  is the  $L^2$ -norm of  $f$  (also written as  $\|f\|_2$ ), which suggests that this space is a subspace of the  $L^2$ -space. We also have  $L^\infty$ -norm of  $f$  (written as  $\|f\|_\infty$ ) defined as  $\sup_{x \in \mathbf{R}} |f(x)|$ , where  $|z| = \sqrt{z\bar{z}}$  for  $z \in \mathbf{C}$ . And we have the general inequality  $0 \leq \|f\|_2 \leq \|f\|_\infty$ . We also have  $\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$  called the triangle inequality, and a related identity called the Pythagoras’ theorem: If  $\langle f, g \rangle = 0$  then  $\|f + g\|_2^2 = \|f\|_2^2 + \|g\|_2^2$

We will also need to use some properties of inner products in future, therefore I list them here.

**Property 1:** (Hermitian Property)  $\langle f, g \rangle = \overline{\langle g, f \rangle}$ .

**Property 2:** (Positivity)  $\langle f, f \rangle > 0$  if  $f \neq 0$  and  $\langle f, f \rangle = 0$  otherwise.

**Property 3:** (Linearity in the first variable)  $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$  and  $\langle cf, g \rangle = c\langle f, g \rangle$ , where  $c \in \mathbf{C}$ .

Once we have a vector space, we also have its basis. A basis for  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  is  $e_n$  for  $n \in \mathbf{Z}$ , where  $e_n = e^{2\pi i n x}$ . Once we have a basis, we may express any element of  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  as a linear combination of the basis “vectors” if the basis is orthonormal (which this one is) using the fact that

$$f = \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n$$

*Remark.* We may look at functions with a different period, then the space also changes, and the basis will change as well.

*Remark.* Orthonormal basis means that for any two basis vectors  $e_n$  and  $e_m$ ,  $n \neq m$ ,  $\langle e_n, e_m \rangle = 0$  and  $\langle e_n, e_n \rangle = 1$ .

The series on the right hand side of the above equation is the Fourier Series of  $f$ . We shall formalize this with a definition.

**Definition 9** (Fourier series). Let  $f : \mathbf{R} \rightarrow \mathbf{C}$  be a continuous periodic function with a period of 1. Then we define its *Fourier series* as following.

$$F(f) \equiv \sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n$$

For convenience, we will define  $\hat{f}(n) \equiv \langle f, e_n \rangle$ .

Now that we have defined what a Fourier series is, we can come back to our discussion of the Weierstrass approximation theorem that applies to this space, and that eventually leads to the Fourier series. But first we need to discuss about trigonometric polynomials, which will be followed by a discussion of periodic convolutions, periodic approximations to the identity and proving a variant of the Weierstrass approximation theorem.

## 4.1 Trigonometric Polynomials

**Definition 10** (Characters). For every integer  $n$ , we let  $e_n \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  denote the function

$$e_n(x) = e^{2\pi i n x}.$$

This is sometimes referred to as the *character with frequency  $n$* . These are the same functions which form the basis of  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  as discussed above.

**Definition 11** (Trigonometric polynomials). A function  $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  is said to be a *trigonometric polynomial* if we can write  $f = \sum_{n=-N}^N c_n e_n$  for some integer  $N \geq 0$  and some complex numbers  $(c_n)_{n=-N}^N$ .

## 4.2 Periodic Convolutions, Periodic Approximation to the Identity, and Weierstrass

**Theorem 12.** Let  $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ , and let  $\varepsilon > 0$ . Then there exists a trigonometric polynomial  $P$  such that  $\|f - P\|_\infty \leq \varepsilon$ .

This theorem says that any continuous periodic function can be uniformly approximated by trigonometric polynomials. If  $P(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  denotes the space of all trigonometric polynomials, then the closure of  $P(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  in the  $L^\infty$  metric is  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ . To prove the above theorem, we will follow almost the same template as in the former part of this document. Therefore, we need to talk about what convolutions mean for periodic functions.

**Definition 13** (Periodic convolution). Let  $f, g \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ . Then we define *periodic convolution*  $f * g : \mathbf{R} \rightarrow \mathbf{C}$  of  $f$  and  $g$  by the formula

$$f * g(x) \equiv \int_{[0,1]} f(y)g(x-y)dy.$$

The convolutions can be related back to the Fourier series by observing that  $f * e_n = \langle f, e_n \rangle e_n$ . The following computation shows this.

$$\begin{aligned} f * e_n(x) &= \int_{[0,1]} f(y)e^{2\pi i n(x-y)}dy \\ &= e^{2\pi i n x} \int_{[0,1]} f(y)e^{-2\pi i n y}dy \\ &= \hat{f}(n)e^{2\pi i n x} = \hat{f}(n)e_n. \end{aligned}$$

The key observation that the periodic convolution of any function in  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  with a trigonometric polynomial, is again a trigonometric polynomial.

**Definition 14** (Periodic approximation to the identity). Let  $\varepsilon > 0$  and  $0 < \delta < 1/2$ . A function  $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$  is said to be a *periodic  $(\varepsilon, \delta)$ -approximation to the identity* if the following properties are true:

1.  $f(x) \geq 0$  for all  $x \in \mathbf{R}$ , and  $\int_{[0,1]} f = 1$ .
2. We have  $f(x) < \varepsilon$  for all  $\delta \leq |x| \leq 1 - \delta$ .

Note that now the “tails” of the function lie between  $\delta$  and  $1 - \delta$  instead of  $\delta$  and 1 like in the case of the polynomials.

**Lemma 15.** *For every  $\varepsilon > 0$  and  $0 < \delta < 1/2$ , there exists a trigonometric polynomial  $P$  which is an  $(\varepsilon, \delta)$ -approximation to the identity.*

*Proof.* This is a constructional proof. We show the existence of such a trigonometric polynomial. Such a trigonometric polynomial is called the *Fejér kernel*. Define the Fejér kernel  $F_N$  to be the function

$$F_N = \sum_{n=-N}^N \left(1 - \frac{|n|}{N}\right) e_n.$$

We can see that  $F_N$  is a trigonometric polynomial.  $F_N$  can be equivalently defined as

$$F_N = \frac{1}{N} \left| \sum_{n=0}^{N-1} e_n \right|^2.$$

The justification of this is given towards the end of the proof.

Now, we note that when  $x$  is not an integer, the summation is a geometric series with the common ratio  $r = e_1 = e^{2\pi i x}$ , and  $a = e_0 = 1$ . For a geometric series  $\sum_{j=0}^n ar^j$ , we know that the sum is  $a \frac{1-r^{n+1}}{1-r}$ . Therefore,

$$\sum_{n=0}^{N-1} e_n(x) = \frac{e_N - e_0}{e_1 - e_0} = \frac{e_{N/2}(e_{N/2} - e_{-N/2})}{e_{-1/2}(e_{1/2} - e_{-1/2})}.$$

Using the fact that  $2i \sin \theta = e^{i\theta} - e^{-i\theta}$ , we get  $e_{N/2} - e_{-N/2} = e^{\pi i N x} - e^{-\pi i N x} = 2i \sin(\pi N x)$ . This gives

$$\sum_{n=0}^{N-1} e_n(x) = \frac{e^{\pi i (N-1)x} \sin(\pi N x)}{\sin(\pi x)}.$$

We can now compute  $F_N$  as follows.

$$\begin{aligned} F_N(x) &= \frac{1}{N} \left| \frac{e^{\pi i (N-1)x} \sin(\pi N x)}{\sin(\pi x)} \right|^2 \\ &= \frac{1}{N} \frac{e^{\pi i (N-1)x} \sin(\pi N x)}{\sin(\pi x)} \cdot \frac{e^{-\pi i (N-1)x} \sin(\pi N x)}{\sin(\pi x)} \\ &= \frac{\sin(\pi N x)^2}{N \sin(\pi x)^2} \end{aligned}$$

When  $x$  is an integer,  $F_N = N$  by direct computation. In either case we can verify that  $F_N(x) \geq 0$  for any  $x$ . Thus we have verified a part of the first property of approximation to the identity.

Next, observe that  $\int_0^1 e_n = \int_0^1 e^{2\pi i n x} = 1/(2\pi i n) [e^{2\pi i n x}]_0^1 = 1/(2\pi i n) [e^{2\pi i n} - 1] = 1/(2n\pi i) (\cos(2n\pi) + i \sin(2n\pi) - 1) = 0$  when  $n \neq 0$  and  $\int_0^1 e_0 = \int_0^1 1 = 1$ . This implies the following.

$$\int_{[0,1]} F_N(x) dx = \sum_{n=-N}^N \left(1 - \frac{|n|}{N}\right) \int_{[0,1]} e_n = \left(1 - \frac{|0|}{N}\right) \int_{[0,1]} e_0 = 1.$$

Now, we only need to verify the second property, but note that  $\delta$  is not an integer. Therefore,

$$F_N(x) = \frac{\sin(\pi Nx)^2}{N \sin(\pi x)^2} \leq \frac{1}{N \sin(\pi x)^2} \leq \frac{1}{N \sin(\pi \delta)^2}$$

whenever  $\delta < |x| < 1 - \delta$  because  $\sin$  is increasing on  $[0, \pi/2]$  and decreasing on  $[\pi/2, \pi]$ . Thus by choosing  $N$  large enough, we can make  $F_N(x) \leq \varepsilon$  for all  $\delta < |x| < 1 - \delta$ .

To complete the proof, we now prove that the alternative form of  $F_N$  is correct in the following.

$$\begin{aligned} \left| \sum_{n=0}^N e_n \right|^2 &= \sum_{n=0}^{N-1} e_n \sum_{m=0}^{N-1} e_{-m} \\ &= \sum_{n=0}^{N-1} e_n \sum_{m=0}^{N-1} e_{-(N-1)+m} \end{aligned}$$

(We have rewritten the sum loop in reverse, but its still the same sum.)

$$= e_{-(N-1)} \left( \sum_{n=0}^{N-1} e_n \right)^2$$

(because  $e_n e_m = e_{n+m}$  as seen by definition of  $e_n$ )

We can now multiply the two sums to get

$$\left( \sum_{n=0}^{N-1} e_n \right)^2 = \sum_{n=0}^{2N-2} c_n e_n$$

where  $c_n$  is the number of pairs  $(k, l)$  such that  $k, l \in \{0, 1, \dots, N-1\}$  and  $k + l = n$ . This is the of the  $n^{\text{th}}$  diagonal of an  $N \times N$  matrix (0-indexing). There are a total of  $2N - 1$  diagonals. The longest diagonal is of length  $N$ , smallest of length 1, and the lengths first increase until it reaches  $N$  and then decrease again until they reach 1. Therefore, the length of diagonal  $n$ ,  $0 \leq n \leq N - 1$  is  $n + 1$ , and the length of diagonals  $n$ ,  $N - 1 < n \leq 2N - 1$  is  $(2N - 1) - n$ . This can be succinctly written as  $c_n = N - |n - (N - 1)|$ .

$$\begin{aligned} \frac{1}{N} \left| \sum_{n=0}^{N-1} e_n \right|^2 &= \frac{e_{-(N-1)}}{N} \left( \sum_{n=0}^{N-1} e_n \right)^2 \\ &= \frac{e_{-(N-1)}}{N} \sum_{n=0}^{2N-2} (N - |n - (N - 1)|) e_n \\ &= \frac{1}{N} \sum_{n=0}^{2N-2} (N - |n - (N - 1)|) e_{n-(N-1)} \\ &= \frac{1}{N} \sum_{m=-(N-1)}^{N-1} (N - |m|) e_m \quad (\text{where } m = n - (N - 1)) \\ &= \frac{1}{N} \sum_{n=-N}^N (N - |n|) e_n \quad (\text{because } N - |\pm N| = 0) \\ &= \sum_{n=-N}^N \left( 1 - \frac{|n|}{N} \right) e_n \end{aligned}$$

□

*Proof of theorem 12.* Let  $f$  be any element of  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ ; we know that  $f$  is bounded, so that we have some  $M > 0$  such that  $|f(x)| \leq M$  for all  $x \in \mathbf{R}$ .

Let  $\varepsilon > 0$  be arbitrary. Since  $f$  is uniformly continuous, there exists a  $\delta > 0$  such that  $|f(x) - f(y)| \leq \varepsilon$  whenever  $|x - y| \leq \delta$ . Now use lemma 15 to find a trigonometric polynomial  $P$  which is a  $(\varepsilon, \delta)$ -approximation to the identity. Then  $f * P$  is also a trigonometric polynomial. We will now estimate  $\|f - f * P\|_\infty$ .

Let  $x$  be any real number. We have

$$\begin{aligned} |f(x) - f * P(x)| &= |f(x) - P * f(x)| \\ &= \left| f(x) - \int_{[0,1]} f(x-y)P(y)dy \right| \\ &= \left| \int_{[0,1]} f(x)P(y)dy - \int_{[0,1]} f(x-y)P(y)dy \right| \\ &= \left| \int_{[0,1]} f(x) - f(x-y)P(y)dy \right| \\ &\leq \int_{[0,1]} |f(x) - f(x-y)|P(y)dy \end{aligned}$$

The right-hand side can be split as

$$\int_{[0,\delta]} |f(x) - f(x-y)|P(y)dy + \int_{[\delta,1-\delta]} |f(x) - f(x-y)|P(y)dy + \int_{[1-\delta,1]} |f(x) - f(x-y)|P(y)dy$$

which can be bound from above by

$$\begin{aligned} &\leq \int_{[0,\delta]} \varepsilon P(y)dy + \int_{[\delta,1-\delta]} 2M\varepsilon dy + \int_{[1-\delta,1]} \varepsilon P(y)dy \\ &\leq \varepsilon + 2M\varepsilon + \varepsilon \\ &= (2M + 2)\varepsilon. \end{aligned}$$

Thus we have  $\|f - f * P\|_\infty \leq (2M + 2)\varepsilon$ . Since  $M$  is fixed and  $\varepsilon$  is arbitrary, we can thus make  $f * P$  arbitrarily close to  $f$  in sup norm, which proves the periodic Weierstrass approximation theorem.  $\square$

### 4.3 The Fourier and Plancherel Theorems

**Theorem 16** (Fourier theorem). *For any  $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ , the series  $\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n$  converges in  $L^2$  metric to  $f$ . In other words, we have*

$$\lim_{N \rightarrow \infty} \left\| f - \sum_{n=-N}^N \hat{f}(n)e_n \right\|_2 = 0.$$

*Proof.* By the Weierstrass approximation theorem, we can find a trigonometric polynomial  $P = \sum_{n=-N_0}^{N_0} c_n e_n$  such that  $\|f - P\|_\infty \leq \varepsilon$ , for some  $N_0 > 0$ . In particular we have  $\|f - P\|_2 \leq \varepsilon$ .

Now, let  $N > N_0$ , and let  $F_N \equiv \sum_{n=-N}^N \hat{f}(n)e_n$ . We claim that  $\|f - F_N\|_2 \leq \varepsilon$ . First observe that for any  $|m| \leq N$ , we have

$$\langle f - F_N, e_m \rangle = \langle f, e_m \rangle - \sum_{n=-N}^N \hat{f}(n) \langle e_n, e_m \rangle = \hat{f}(m) - \hat{f}(m) = 0,$$

because  $\langle e_n, e_m \rangle = 0$  when  $n \neq m$  and 1 otherwise. Since both  $F_N$  and  $P$  are trigonometric polynomials, we can write  $F_N - P$  as a linear combination of  $e_n$ 's. Which implies

$$\langle f - F_N, F_N - P \rangle = 0.$$

By the Pythagoras' theorem we have

$$\|f - P\|_2^2 = \|f - F_N\|_2^2 + \|F_N - P\|_2^2$$

and in particular

$$\|f - F_N\|_2 \leq \|f - P\|_2 \leq \varepsilon$$

as desired.  $\square$

**Theorem 17** (Plancherel theorem). *For any  $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ , the series  $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$  is absolutely convergent, and*

$$\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

*This theorem is also known as Parseval's theorem.*

*Proof.* Let  $\varepsilon > 0$ . By the Fourier theorem we know that

$$\left\| f - \sum_{n=-N}^N \hat{f}(n)e_n \right\|_2 \leq \varepsilon$$

if  $N$  is large enough (depending on  $\varepsilon$ ). In particular, by the triangle inequality this implies that

$$\|f\|_2 - \varepsilon \leq \left\| \sum_{n=-N}^N \hat{f}(n)e_n \right\|_2 \leq \|f\|_2 + \varepsilon.$$

However, since the series is a usual polynomial element in  $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ , we know that

$$\left\| \sum_{n=-N}^N \hat{f}(n)e_n \right\|_2^2 = \sum_{n=-N}^N |\hat{f}(n)|^2$$

and hence

$$(\|f\|_2 - \varepsilon)^2 \leq \sum_{n=-N}^N |\hat{f}(n)|^2 \leq (\|f\|_2 + \varepsilon)^2.$$

Take the limit supremum to obtain

$$(\|f\|_2 - \varepsilon)^2 \leq \limsup_{N \rightarrow \infty} \sum_{n=-N}^N |\hat{f}(n)|^2 \leq (\|f\|_2 + \varepsilon)^2.$$

Since,  $\varepsilon$  is arbitrary, this inequality also holds for  $\varepsilon = 0$ , and therefore by the squeeze test

$$\limsup_{N \rightarrow \infty} \sum_{n=-N}^N |\hat{f}(n)|^2 = \|f\|_2^2.$$

Materials on squeeze test and lim sup.  $\square$

Finally, without proving, the generalized version of the Fourier series that applies to a periodic function with a period  $L$  requires the characters

$$e_n = \frac{1}{\sqrt{L}} e^{2\pi i n x / L}.$$

One of the most common periods is  $L = 2\pi$ , where  $e_n = \frac{1}{\sqrt{2\pi}} e^{i n x}$  is used to compute the Fourier series. The interval of integration in the inner product also changes from  $[0, 1]$  to  $[0, L]$ .

The Fourier series and the Fejer kernel are related via the Dirichlet kernel. The Fejer kernel can be written in terms of the Dirichlet kernel, and so can the partial sums of the Fourier series. Moreover, the Cesàro sum of the Fourier series can be written in terms of the Fejer kernel. The details can be found here (William Wu, “Fourier Series and Fejér Theorem”, 2004, ocf.berkeley.edu).

I will discuss one other thing here which I borrow directly from the above reference of William Wu.

## 4.4 Cesaro Convergence

The Fourier series is an *approximation* of the original function, therefore mathematicians were interested in finding an answer to how does the series approximate the function? This and similar questions of Fourier series convergence led to real analysis.

One notion of convergence between functions is  $L^2$ -convergence, or convergence in the mean. The statement is that

$$\lim_{N \rightarrow \infty} \int_0^1 \left| f(x) - \sum_{n=-N}^N \hat{f}(n) e^{2\pi i n x} \right|^2 dx = 0.$$

But this only says that the average deviation from the function value is 0 over the interval  $[0, 1]$ , however for a fixed  $x$  in  $[0, 1]$ , there are no guarantees on the difference between  $f(x)$  and the series approximation at  $x$ . Therefore, mathematicians turned to *pointwise* convergence in which we demand that at each point  $x \in [0, 1]$ , the series approximation converges to  $f(x)$ . Jordan’s Pointwise Convergence Theorem states that if  $f$  is sectionally continuous and  $x_0$  is such that one-sided derivatives  $f'(x_0^+)$  and  $f'(x_0^-)$  both exists, then the Fourier series  $\sum_n \hat{f}(n) e^{2\pi i n x_0}$  converges to  $f(x_0)$ .

This is interesting, but inappropriate in some cases. An example is the sequence of functions defined by  $g_n(x) : x \rightarrow x^n$  for  $x \in [0, 1]$ . Then  $(g_n)$  converges to pointwise to a function  $h$  that equals 0 for  $x \in [0, 1)$ , but equals 1 for  $x = 1$ . Thus although  $(g_n)$  consists only of continuous functions, the limit function is discontinuous.

To avoid such problems, we desire the even stronger notion of *uniform* convergence, such that the rate at which the series converges is identical for all points in  $[0, 1]$ . By adopting the metric

$$d(f, g) = \sup\{|f(t) - g(t)| : t \in [0, 1]\}$$

over the space of continuous functions from  $[0, 1]$  to  $\mathbf{R}$ , we can force convergence to imply uniform convergence, simply by definition. This metric space is denoted by  $C([0, 1], \mathbf{R})$ . This is the space we have been working with so far. It can also be proven that this space is a vector space, and thus the concept of a series is well-defined. Fejer’s theorem then states the following.

**Theorem 18** (Fejér’s Theorem). *Let  $f : [-\pi, \pi] \rightarrow \mathbf{R}$  be a continuous function with  $f(-\pi) = f(\pi)$ . Then the Fourier series of  $f$   $(C, 1)$ -converges to  $f$  in  $C([-\pi, \pi], \mathbf{R})$ , where  $C([-\pi, \pi], \mathbf{R})$  is the metric space of continuous functions from  $[-\pi, \pi]$  to  $\mathbf{R}$ .*

Without imposing any additional conditions on  $f$  aside from being continuous and periodic, Fejér's theorem shows that Fourier series can still achieve uniform convergence, granted that we instead consider the arithmetic means of partial Fourier sums (Cesaro sums). We will be talking with respect to  $C([-\pi, \pi], \mathbf{R})$  so that we can talk about the theorem in its original form. In  $C([-\pi, \pi], \mathbf{R})$ , the Fejér kernel has the property that for a fixed  $\delta > 0$ ,  $\lim_{n \rightarrow \infty} \int_{\delta \leq |x| \leq \pi} F_n(x) dx = 0$ .

Lets denote the  $n^{\text{th}}$  partial sum of the Fourier series over a period of  $2\pi$  by  $s_n$ , and denote the corresponding  $n^{\text{th}}$  Cesaro sum by  $\sigma_n$ .

$$s_n(x) = \sum_{k=-n}^n \hat{f}(k) e^{ikx}$$

$$\sigma_n(x) = \frac{1}{n+1} \sum_{k=0}^n s_k(x)$$

We can rewrite these expressions in terms of the Fejér kernel by substituting the integral form of  $\hat{f}(k)$  to get

$$\begin{aligned} s_n(x) &= \sum_{k=-n}^n \left[ \frac{1}{2\pi} \int_{[-\pi, \pi]} f(t) e^{-ikt} dt \right] e^{ikx} \\ &= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(t) \sum_{k=-n}^n e^{ik(x-t)} dt \\ &= \frac{1}{2\pi} \int_{[-\pi, \pi]} f(t) D_n(x-t) dt \end{aligned}$$

where  $D_n$  is the Dirichlet kernel. With a change of variables, we get

$$s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_n(t) dt.$$

Using, this we can rewrite the Cesaro sum  $\sigma_n$  as

$$\begin{aligned} \sigma_n(x) &= \frac{1}{n+1} \sum_{k=0}^n \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_k(t) dt \right] \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) \left[ \frac{1}{n+1} \sum_{k=0}^n D_k(t) \right] dt \end{aligned}$$

Notice that the mean of the Dirichlet kernel is the Fejér kernel. This can also be seen in the definition of the Fejér kernel introduced earlier. Thus

$$\sigma_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) F_n(t) dt$$

which is the convolution of  $f$  with the Fejér kernel. We already proved that this converges to  $f$ . The steps are precisely to compute  $\sigma_n(x) - f(x)$ , then applying the triangle inequality, then combine it with the non-negativity of the Fejér kernel. We have already done this above (for a different period,  $L = 1$ ).