Chapter Eleven

Introduction to Fourier Analysis

Everyone has experience representing quantities in terms of a basis set. Many cities are laid out in a grid, and locations are represented by x blocks east and y blocks north. We represent this mathematically by writing any point in the plane $\mathbb{R}^2$ as $x\binom{1}{0} + y\binom{0}{1}$, with $x, y \in \mathbb{R}$. Here $\binom{1}{0}$ is the unit vector in the east direction, $\binom{0}{1}$ is the unit vector in the north direction, and our "basis" is the set of the two unit vectors.

Another example is Taylor series expansions (see §A.2.3). For "nice" functions (although see Exercise A.2.7)

$$f(x) = a_0 \cdot 1 + a_1 \cdot x + a_2 \cdot x^2 + \cdots, \tag{11.1}$$

at least for x sufficiently close to 0. We are expanding $f(x)$ in terms of $\{1, x, x^2, x^3, x^4, \dots\}$. Unlike our example in the plane, here we have an infinite dimensional space. Because there are infinitely many directions, we will have to work to show how to deal with issues such as convergence (does any such series converge to a nice function?) and representation (can any nice function be written as such a series?).

Fourier Analysis is concerned with expanding periodic functions (often with period 1) in terms of the Fourier basis $e_n(x) = e^{2\pi i n x}$, $n \in \mathbb{Z}$. Recall a function f is **periodic** of period a if $f(x + a) = f(x)$ for all x. This basis turns out to be extremely useful for a variety of problems; we will see several instances below, ranging from the equidistribution of sequences in Chapter 12 (if $\alpha \notin \mathbb{Q}$ then $n\alpha \bmod 1$ is equidistributed in $[0, 1]$) to Goldbach-type problems in Chapter 13 (representing numbers as the sum of primes). While the investigations in Chapter 12 require many of the technical results on convergence proved below, in Chapter 13 all we really need is the notation and results of §11.1. We will also prove Poisson Summation, one of the most useful techniques in number theory with applications ranging from proving the functional equation of $\zeta(s)$ (Theorem 3.1.20) to investigating digit bias (see §9.4.2). Finally, when we investigate the distribution of zeros of L-functions in Chapter 18, we shall use Fourier analysis to derive formulas connecting sums over zeros to sums over primes.

It is a deep problem to determine what functions are given by an expansion in the Fourier basis. We prove for many "nice" periodic functions that f is equal to its expansion in the Fourier basis. This is but one of many applications of Fourier analysis; others include solving the heat and wave equations (how systems evolve with time), the isoperimetric inequality (of all smooth closed curves in the plane, for a given perimeter the circle encloses the greatest area), the uncertainty principle (one cannot localize arbitrarily well a function and its Fourier transform), computation of special values of $\zeta(s)$ and L-functions (Chapter 3, and is related to a proof that there are infinitely many primes), the Central Limit Theorem (Chapter 8) and Poissonian behavior of $n^d\alpha \bmod 1$ (Chapter 12), to name a few. We sketch some of these applications in §11.6. For a comprehensive treatment of Fourier analysis, see [Be, SS1, Zy].

11.1 INNER PRODUCT OF FUNCTIONS

For $x \in \mathbb{C}$ we define the exponential function by means of the series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{11.2}$$

which converges everywhere; see §5.4 for more properties of e. Given the Taylor series expansion of $\sin x$ and $\cos x$, we can verify the identity

$$e^{ix} = \cos x + i \sin x. \tag{11.3}$$

Exercise 11.1.1. *Prove e^x converges for all $x \in \mathbb{R}$ (even better, for all $x \in \mathbb{C}$). Show the series for e^x also equals*

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n, \tag{11.4}$$

which you may remember from compound interest problems.

Exercise 11.1.2. *Prove, using the series definition, that $e^{x+y} = e^x e^y$ and calculate the derivative of e^x.*

Recall the definition of the **inner** or **dot product**: for two complex-valued vectors $\vec{v} = (v_1, \cdots, v_n)$, $\vec{w} = (w_1, \cdots, w_n)$, we define the inner product $\vec{v} \cdot \vec{w}$ (also denoted $\langle \vec{v}, \vec{w} \rangle$) by

$$\vec{v} \cdot \vec{w} = \langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i \overline{w}_i, \tag{11.5}$$

where $\overline{z}$ is the complex conjugate of z (if $z = x + iy$, $\overline{z} = x - iy$). The length of a vector $\vec{v}$ is

$$|\vec{v}| = \sqrt{\langle \vec{v}, \vec{v} \rangle}. \tag{11.6}$$

We generalize the dot product to functions. Let $f : \mathbb{R} \to \mathbb{C}$, say $f(x) = u(x) + iv(x)$ with u, v real valued functions. We define

$$\int_a^b f(x)dx = \int_a^b u(x)dx + i \int_a^b v(x)dx. \tag{11.7}$$

For a complex number z, $|z|^2 = z\overline{z}$. We will see the generalization of length to a function will be $\int f(x)\overline{f(x)}dx$ (while $\int f(x)^2 dx$ can be zero (or even negative) for complex valued f, the first integral is always non-negative).

For definiteness, assume f and g are functions from $[0, 1]$ to $\mathbb{C}$. Divide the interval $[0, 1]$ into n equal pieces. Then we can associate to f a vector in $\mathbb{R}^n$ by

$$f(x) \longleftrightarrow \left(f(0), f\left(\frac{1}{n}\right), \ldots, f\left(\frac{n-1}{n}\right)\right), \tag{11.8}$$

and similarly for g. Call these vectors f_n and g_n. As before, we consider

$$\langle f_n, g_n \rangle = \sum_{j=0}^{n-1} f\left(\frac{j}{n}\right) \cdot \overline{g}\left(\frac{j}{n}\right); \tag{11.9}$$

note if $f = g$ this sum is real and non-negative. In general, as we continue to divide the interval ($n \to \infty$), the above sum diverges. For example, if f and g are identically 1, the above sum is n. We expect that the inner product of the constant function on the unit interval with itself (its length) should be 1.

There is a natural re-scaling: we multiply each term in the sum by $\frac{1}{n}$, the size of the subinterval. Note for the constant function the sum is now independent of n. Thus, for good f and g, we are led to define $\langle f, g \rangle$ by a Riemann integral

$$\langle f, g \rangle = \lim_{n\to\infty} \sum_{j=0}^{n-1} f\left(\frac{j}{n}\right) \cdot \overline{g}\left(\frac{j}{n}\right) \frac{1}{n} = \int_0^1 f(x)\overline{g(x)}dx. \tag{11.10}$$

Taking the complex conjugate of g ensures that $\langle f, f \rangle$ is non-negative. Here, "good" means any class of functions such that the Riemann integral converges (for example, continuous or piecewise continuous functions).

Exercise 11.1.3. *Let f, g and h be continuous functions on $[0,1]$, and $a, b \in \mathbb{C}$. Prove*

1. *$\langle f, f \rangle \geq 0$, and equals 0 if and only if f is identically zero;*
2. *$\langle f, g \rangle = \overline{\langle g, f \rangle}$;*
3. *$\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$.*

Exercise 11.1.4. *Find a vector $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{C}^2$ such that $v_1^2 + v_2^2 = 0$, but $\langle \vec{v}, \vec{v} \rangle \neq 0$.*

Definition 11.1.5 (Orthogonal)**.** *Two continuous functions on $[0,1]$ are orthogonal (or perpendicular) if their inner product equals zero.*

Exercise 11.1.6. *Prove x^n and x^m are not perpendicular on $[0,1]$. Find a $c \in \mathbb{R}$ such that $x^n - cx^m$ is perpendicular to x^m; c is related to the projection of x^n in the direction of x^m.*

We will see that the exponential function behaves very nicely under the inner product. For $n \in \mathbb{Z}$, define

$$e_n(x) = e^{2\pi i n x}. \tag{11.11}$$

Exercise 11.1.7 (Important)**.** *Show for $m, n \in \mathbb{Z}$ that*

$$\langle e_m(x), e_n(x) \rangle = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{otherwise.} \end{cases} \tag{11.12}$$

Thus $\{\dots, e_{-1}(x), e_0(x), e_1(x), \dots\}$ is a set of mutually perpendicular, unit-length functions. By an **orthogonal set** we mean a set of vectors or functions which are mutually perpendicular; if additionally each vector or function has unit length, we say the set is **orthonormal**. Thus the Fourier basis is an orthonormal set.

Much care is needed, however, in expanding a general periodic function in terms of the $e_n(x)$'s. First, we have issues arising because we have infinitely many basis functions – we must show the infinite series converge, and further that it converges to the initial function (and this is, sadly, not always the case). This is very different from the case of the plane, when we had just two basis vectors, $\binom{1}{0}$ and $\binom{0}{1}$. Second, we have questions of completeness. Is the above list of the $e_n(x)$'s complete? Do these infinitely many functions capture everything, or do we need to add more functions which are orthogonal to each $e_n(x)$?

Definition 11.1.8 (Periodic). *A function $f(x)$ is periodic with period a if for all $x \in \mathbb{R}$, $f(x + a) = f(x)$.*

Note the exponential functions $e_n(x)$ are periodic with period 1. Thus, if f is periodic with period 1, it makes no difference if we study f on $[0, 1]$ or $[-\frac{1}{2}, \frac{1}{2}]$ or, more generally, on *any* interval of length one.

Exercise 11.1.9. *Let f and g be periodic functions with period a. Prove $\alpha f(x) + \beta g(x)$ is periodic with period a.*

Definition 11.1.10 (Even, Odd). *A function $f(x)$ is even (resp., odd) if $f(x) = f(-x)$ (if $f(x) = -f(-x)$).*

Exercise 11.1.11. *Prove any function can be written as the sum of an even and an odd function.*

11.2 FOURIER SERIES

11.2.1 Introduction

Let f be continuous and periodic on $\mathbb{R}$ with period one. Define the n^{th} **Fourier coefficient** $\widehat{f}(n)$ of f to be

$$\widehat{f}(n) = \langle f(x), e_n(x) \rangle = \int_0^1 f(x) e^{-2\pi i n x} dx. \tag{11.13}$$

Returning to the intuition of $\mathbb{R}^m$, we can think of the $e_n(x)$'s as an infinite set of perpendicular unit directions. The above is simply the projection of f in the direction of $e_n(x)$. Often one writes a_n for $\widehat{f}(n)$.

Exercise 11.2.1. *Show*

$$\langle f(x) - \widehat{f}(n) e_n(x), e_n(x) \rangle = 0. \tag{11.14}$$

This agrees with our intuition: after removing the projection in a certain direction, what is left is perpendicular to that direction.

The N^{th} **partial Fourier series** of f is

$$S_N(x) = \sum_{n=-N}^{N} \widehat{f}(n) e_n(x). \tag{11.15}$$

Exercise 11.2.2. *Prove*

1. $\langle f(x) - S_N(x), e_n(x)\rangle = 0$ *if* $|n| \leq N$.
2. $|\widehat{f}(n)| \leq \int_0^1 |f(x)| dx$.
3. *Bessel's Inequality: if* $\langle f, f\rangle < \infty$ *then* $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 \leq \langle f, f\rangle$.
4. *Riemann-Lebesgue Lemma: if* $\langle f, f\rangle < \infty$ *then* $\lim_{|n|\to\infty} \widehat{f}(n) = 0$ *(this holds for more general* f*; it suffices that* $\int_0^1 |f(x)| dx < \infty$*).*
5. *Assume* f *is differentiable* k *times; integrating by parts, show* $|\widehat{f}(n)| \ll \frac{1}{n^k}$ *and the constant depends only on* f *and its first* k *derivatives.*

As $\langle f(x) - S_N(x), e_n(x)\rangle = 0$ if $|n| \leq N$, we might think that we just have to let N tend to infinity to obtain a series $S_\infty(x)$ such that

$$\langle f(x) - S_\infty(x), e_n(x)\rangle = 0. \tag{11.16}$$

Assume that a periodic function $g(x)$ is orthogonal to $e_n(x)$ for every n if and only if $g(x)$ is identically zero. Then $f(x) - S_\infty(x) = 0$, and hence $f(x) = S_\infty(x)$: we have expressed $f(x)$ as a sum of exponentials! We must be very careful. We have just glossed over the two central issues – completeness (are the $e_n(x)$'s all the "directions"?) and, even worse, convergence (do the sums agree with f for all x?). For many f, the Fourier series does converge pointwise, but much care is required to prove such results. By looking at modified Fourier series, we will easily give examples of finite approximations to f with good pointwise convergence.

Exercise 11.2.3. *Let* $h(x) = f(x) + g(x)$. *Does* $\widehat{h}(n) = \widehat{f}(n) + \widehat{g}(n)$*? Let* $k(x) = f(x)g(x)$. *Does* $\widehat{k}(n) = \widehat{f}(n)\widehat{g}(n)$*?*

Remark 11.2.4. In many of our theorems below we assume that $\langle f, f\rangle = \int_0^1 |f(x)|^2 dx < \infty$. This is a natural condition, as the Cauchy-Schwarz inequality (Appendix A.6) implies that if f and g are two such functions, $\langle f, g\rangle < \infty$ and $\int_0^1 |f(x)| dx < \infty$; the second statement is false if $f, g : \mathbb{R} \to \mathbb{C}$ and not $[0, 1] \to \mathbb{C}$. If $\int_0^1 |f(x)|^r dx < \infty$, one often writes $f \in L^r([0, 1])$.

Exercise 11.2.5. *Remark 11.2.4 shows that if* $\langle f, f\rangle, \langle g, g\rangle < \infty$ *then the dot product of* f *and* g *exists:* $\langle f, g\rangle < \infty$. *Do there exist* $f, g : [0, 1] \to \mathbb{C}$ *such that* $\int_0^1 |f(x)| dx, \int_0^1 |g(x)| dx < \infty$ *but* $\int_0^1 f(x)\overline{g}(x) dx = \infty$*? Is* $f \in L^2([0, 1])$ *a stronger or an equivalent assumption as* $f \in L^1([0, 1])$*?*

11.2.2 Approximations to the Identity

We assume the reader is familiar with the basics of probability functions (see Chapter 8, especially §8.2.3). A sequence $A_1(x), A_2(x), A_3(x), \dots$ of functions is an **approximation to the identity on** $[0, 1]$ if

1. for all x and N, $A_N(x) \geq 0$;
2. for all N, $\int_0^1 A_N(x)dx = 1$;
3. for all δ, $0 < \delta < \frac{1}{2}$, $\lim_{N\to\infty} \int_\delta^{1-\delta} A_N(x)dx = 0$.

Similar definitions hold with $[0, 1]$ replaced by other intervals; it is often more convenient to work on $[-\frac{1}{2}, \frac{1}{2}]$, replacing the third condition with

$$\lim_{N\to\infty} \int_{|x|>\delta} A_N(x)dx = 0 \quad \text{if} \quad 0 < \delta < \frac{1}{2}. \tag{11.17}$$

We could also replace $[0, 1]$ with $\mathbb{R}$, which would make the third condition

$$\lim_{N\to\infty} \int_{-\infty}^{-\delta} + \int_\delta^\infty A_N(x)dx = 0 \quad \text{if} \quad \delta > 0. \tag{11.18}$$

There is a natural interpretation of these conditions. The first two (non-negativity and integral equals 1) allow us to think of each $A_N(x)$ as a probability distribution. The third condition states that, as $N \to \infty$, all of the probability is concentrated arbitrarily close to one point. Physically, we may regard the $A_n(x)$'s as densities for a unit mass of smaller and smaller width. In the limit, we obtain a **unit point mass**; it will have finite mass, but infinite density at one point, and zero density elsewhere.

Exercise 11.2.6. *Define*

$$A_N(x) = \begin{cases} N & \text{for } |x| \leq \frac{1}{N} \\ 0 & \text{otherwise.} \end{cases} \tag{11.19}$$

Prove A_N is an approximation to the identity on $[-\frac{1}{2}, \frac{1}{2}]$. If f is continuously differentiable and periodic with period 1, calculate

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)A_N(x)dx. \tag{11.20}$$

Exercise 11.2.7. *Let $A(x)$ be a non-negative function with $\int_{\mathbb{R}} A(x)dx = 1$. Prove $A_N(x) = N \cdot A(Nx)$ is an approximation to the identity on $\mathbb{R}$.*

Exercise 11.2.8 (Important)**.** *Let $A_N(x)$ be an approximation to the identity on $[-\frac{1}{2}, \frac{1}{2}]$. Let $f(x)$ be a continuous function on $[-\frac{1}{2}, \frac{1}{2}]$. Prove*

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)A_N(x)dx = f(0). \tag{11.21}$$

By Exercise 11.2.8, in the limit the functions $A_N(x)$ are acting like unit point masses at the origin.

Definition 11.2.9 (Dirac Delta Functional). *We define a map from continuous complex valued functions to the complex numbers by $\delta(f) = f(0)$. We often write this in the more suggestive notation*

$$\int f(x)\delta(x)dx = f(0), \tag{11.22}$$

where the integration will usually be over $[0,1]$, $[-\frac{1}{2},\frac{1}{2}]$ or $\mathbb{R}$.

By a standard abuse of notation, we often call $\delta(x)$ the delta function. We can consider the probability densities $A_n(x)dx$ and $\delta(x)dx$. For $A_N(x)dx$, as $N \to \infty$ almost all the probability (mass) is concentrated in a narrower and narrower band about the origin; $\delta(x)dx$ is the limit with all the mass at one point. It is a discrete (as opposed to continuous) probability measure, with infinite density but finite mass. Note that $\delta(x-a)$ acts like a unit point mass; however, instead of having its mass concentrated at the origin, it is now concentrated at a.

11.2.3 Dirichlet and Fejér Kernels

We define two functions which will be useful in investigating convergence of Fourier series. Set

$$\begin{aligned} D_N(x) &:= \sum_{n=-N}^{N} e_n(x) = \frac{\sin((2N+1)\pi x)}{\sin \pi x} \\ F_N(x) &:= \frac{1}{N}\sum_{n=0}^{N-1} D_n(x) = \frac{\sin^2(N\pi x)}{N\sin^2 \pi x}. \end{aligned} \tag{11.23}$$

Exercise 11.2.10. *Prove the two formulas above. The geometric series formula will be helpful:*

$$\sum_{n=N}^{M} r^n = \frac{r^N - r^{M+1}}{1-r}. \tag{11.24}$$

Here F stands for Fejér, D for Dirichlet. $F_N(x)$ and $D_N(x)$ are two important examples of (integral) **kernels**. By integrating a function against a kernel, we obtain a new function related to the original. We will study integrals of the form

$$g(x) = \int_0^1 f(y)K(x-y)dy. \tag{11.25}$$

Such an integral is called the **convolution** of f and K. The Fejér and Dirichlet kernels yield new functions related to the Fourier expansion of $f(x)$.

Theorem 11.2.11. *The Fejér kernels $F_1(x), F_2(x), F_3(x), \ldots$ are an approximation to the identity on $[0,1]$.*

Proof. The first property is immediate. The second follows from the observation that $F_N(x)$ can be written as

$$F_N(x) = e_0(x) + \frac{N-1}{N}\left(e_{-1}(x) + e_1(x)\right) + \cdots, \tag{11.26}$$

and all integrals are zero but the first, which is 1. To prove the third property, note that $F_N(x) \le \frac{1}{N \sin^2 \pi\delta}$ for $\delta \le x \le 1-\delta$. $\square$

Exercise 11.2.12. *Show that the Dirichlet kernels are* not *an approximation to the identity. How large are* $\int_0^1 |D_N(x)|dx$ *and* $\int_0^1 D_N(x)^2 dx$*?*

Let f be a continuous, periodic function on $\mathbb{R}$ with period one. We may consider f as a function on just $[0,1]$, with $f(0) = f(1)$. Define

$$T_N(x) = \int_0^1 f(y)F_N(x-y)dy. \tag{11.27}$$

In other words, $T_N(x)$ is the integral transform of $f(x)$ with respect to the Fejér kernel. We show below that, for many f, $T_N(x)$ has good convergence properties to $f(x)$. To do so requires some basic facts from analysis, which are recalled in Appendix A.3.

11.3 CONVERGENCE OF FOURIER SERIES

We investigate when the Fourier series converges to the original function. For continuous functions, a related series always converges. An important application is that instead of proving results for "general" f, it often suffices to prove results for Fourier series (see Chapter 12).

11.3.1 Convergence of Fejér Series to f

Theorem 11.3.1 (Fejér). *Let $f(x)$ be a continuous, periodic function on $[0,1]$. Given $\epsilon > 0$ there exists an N_0 such that for all $N > N_0$,*

$$|f(x) - T_N(x)| \le \epsilon \tag{11.28}$$

for every $x \in [0,1]$. Hence as $N \to \infty$, $T_N f(x) \to f(x)$.

Proof. The starting point of the proof is multiplying by 1 in a clever way, a very powerful technique. We have

$$f(x) = f(x)\int_0^1 F_N(y)dy = \int_0^1 f(x)F_N(y)dy; \tag{11.29}$$

this is true as $F_N(y)$ is an approximation to the identity and thus integrates to 1.

For any positive N and $\delta \in (0, 1/2)$,

$$\begin{aligned} T_N(x) - f(x) &= \int_0^1 f(x-y)F_N(y)dy - f(x)\cdot 1 \\ &= \int_0^1 f(x-y)F_N(y)dy - \int_0^1 f(x)F_N(y)dy \\ &\qquad \text{(by property 2 of } F_N\text{)} \\ &= \int_0^\delta \left(f(x-y) - f(x)\right) F_N(y)dy \\ &\quad + \int_\delta^{1-\delta} \left(f(x-y) - f(x)\right) F_N(y)dy \\ &\quad + \int_{1-\delta}^1 \left(f(x-y) - f(x)\right) F_N(y)dy. \end{aligned} \tag{11.30}$$

As the $F_N(x)$'s are an approximation to the identity, we find

$$\left|\int_\delta^{1-\delta} \left(f(x-y) - f(x)\right) F_N(y)dy\right| \ \le\ 2\max|f(x)| \cdot \int_\delta^{1-\delta} F_N(y)dy. \tag{11.31}$$

By Theorem A.3.13, $f(x)$ is bounded, so there exists a B such that $\max|f(x)| \le B$. Since

$$\lim_{N\to\infty} \int_\delta^{1-\delta} F_N(y)dy = 0, \tag{11.32}$$

we obtain

$$\lim_{N\to\infty} \int_\delta^{1-\delta} (f(x-y) - f(x))F_N(y)dy = 0. \tag{11.33}$$

Thus, by choosing N large enough (where large depends on δ), we can ensure that this piece is at most $\frac{\epsilon}{3}$.

It remains to estimate what happens near zero, *and this is where we use f is periodic*. Since f is continuous and $[0, 1]$ is a finite closed interval, f is uniformly continuous (Theorem A.3.7). Thus we can choose δ small enough that $|f(x-y) - f(x)| < \frac{\epsilon}{3}$ for any x and any positive $y < \delta$. Then

$$\left|\int_0^\delta \left(f(x-y) - f(x)\right) F_N(y)dy\right| \ \le\ \int_0^\delta \frac{\epsilon}{3} F_N(y)dy \ \le\ \frac{\epsilon}{3}\int_0^1 F_N(y)dy \ \le\ \frac{\epsilon}{3}. \tag{11.34}$$

Similarly

$$\left|\int_{1-\delta}^1 \left(f(x-y) - f(x)\right) F_N(y)dy\right| \ \le\ \frac{\epsilon}{3}. \tag{11.35}$$

Therefore

$$|T_N(x) - f(x)| \ \le\ \epsilon \tag{11.36}$$

for all N sufficiently large. □

Remark 11.3.2. Where did we use f periodic? Recall we had expressions such as $f(x-y)-f(x)$. For example, if $x = .001$ and $y = .002$, we have $f(-.001) - f(.001)$. The periodicity of f allows us to extend f to a continuous function on $\mathbb{R}$.

One often uses the interval $[-\frac{1}{2}, \frac{1}{2}]$ instead of $[0,1]$; the proof follows analogously. Proofs of this nature are often called *three epsilon proofs*; splitting the interval as above is a common technique for analyzing such functions.

Definition 11.3.3 (Trigonometric Polynomials). *Any finite linear combination of the functions $e_n(x)$ is called a trigonometric polynomial.*

From Fejér's Theorem (Theorem 11.3.1) we immediately obtain the

Theorem 11.3.4 (Weierstrass Approximation Theorem). *Any continuous periodic function can be uniformly approximated by trigonometric polynomials.*

Remark 11.3.5. Weierstrass proved (many years before Fejér) that if f is continuous on $[a,b]$, then for any $\epsilon > 0$ there is a polynomial $p(x)$ such that $|f(x)-p(x)| < \epsilon$ for all $x \in [a,b]$. This important theorem has been extended numerous times (see, for example, the Stone-Weierstrass Theorem in [Rud]).

Exercise 11.3.6. *Prove the Weierstrass Approximation Theorem implies the original version of Weierstrass' Theorem (see Remark 11.3.5).*

We have shown the following: if f is a continuous, periodic function, given any $\epsilon > 0$ we can find an N_0 such that for $N > N_0$, $T_N(x)$ is within ϵ of $f(x)$. As ϵ was arbitrary, as $N \to \infty$, $T_N(x) \to f(x)$.

11.3.2 Pointwise Convergence of Fourier Series

Theorem 11.3.1 shows that given a continuous, periodic f, the Fejér series $T_N(x)$ converges pointwise to $f(x)$. The Fejér series is a weighted Fourier series, though; what can be said about pointwise convergence of the initial Fourier series to $f(x)$?

Recall $\widehat{f}(n)$ is the n^{th} Fourier coefficient of $f(x)$. Consider the Fourier series

$$S_N(x) = \sum_{n=-N}^{N} \widehat{f}(n) e^{2\pi i n x}. \tag{11.37}$$

Exercise 11.3.7. *Let $f(x)$ be periodic function with period 1. Show*

$$S_N(x_0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x) D_N(x-x_0)dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_0-x) D_N(x)dx. \tag{11.38}$$

Theorem 11.3.8 (Dirichlet). *Suppose*

1. *$f(x)$ is real valued and periodic with period 1;*
2. *$|f(x)|$ is bounded;*
3. *$f(x)$ is differentiable at x_0.*

Then $\lim_{N\to\infty} S_N(x_0) = f(x_0)$.

Proof. Let $D_N(x)$ be the Dirichlet kernel. Previously we have shown that $D_N(x) = \frac{\sin((2N+1)\pi x)}{\sin(\pi x)}$ and $\int_{-\frac{1}{2}}^{\frac{1}{2}} D_N(x)dx = 1$. Thus

$$\begin{aligned}
f(x_0) - S_N(x_0) &= f(x_0)\int_{-\frac{1}{2}}^{\frac{1}{2}} D_N(x)dx - \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_0 - x)D_N(x)dx \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[f(x_0) - f(x_0 - x)\right] D_N(x)dx \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{f(x_0) - f(x_0 - x)}{\sin(\pi x)} \cdot \sin((2N+1)\pi x)dx \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} g_{x_0}(x)\sin((2N+1)\pi x)dx. \qquad (11.39)
\end{aligned}$$

We claim $g_{x_0}(x) = \frac{f(x_0)-f(x_0-x)}{\sin(\pi x)}$ is bounded. As f is bounded, the numerator is bounded. The denominator is only troublesome near $x = 0$; however, as f is differentiable at x_0,

$$\lim_{x\to 0} \frac{f(x_0 + x) - f(x_0)}{x} = f'(x_0). \qquad (11.40)$$

Multiplying by 1 in a clever way (one of the most useful proof techniques) gives

$$\lim_{x\to 0} \frac{f(x_0 + x) - f(x_0)}{\sin(\pi x)} = \lim_{x\to 0} \frac{f(x_0 + x) - f(x_0)}{\pi x} \cdot \frac{\pi x}{\sin(\pi x)} = \frac{f'(x_0)}{\pi}, \qquad (11.41)$$

where we used L'Hospital's rule to conclude that $\lim_{x\to 0} \frac{\pi x}{\sin(\pi x)} = 1$. Therefore $g_{x_0}(x)$ is bounded everywhere, say by B. As g_{x_0} is a bounded function, it is square-integrable, and thus the Riemann-Lebesgue Lemma (see Exercise 11.2.2) implies that its Fourier coefficients tend to zero. This completes the proof, as

$$i\int_{-\frac{1}{2}}^{\frac{1}{2}} g_{x_0}(x)\sin((2N+1)\pi x)dx = \Im\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} g_{x_0}(x)e^{2\pi i(2N+1)x}dx\right); \qquad (11.42)$$

thus our integral is just the imaginary part of the $2N+1^{\text{st}}$ Fourier coefficient, which tends to zero as $N \to \infty$. Hence as $N \to \infty$, $S_N(x_0)$ converges (pointwise) to $f(x_0)$.

□

Remark 11.3.9. If f is twice differentiable, by Exercise 11.2.2 $\widehat{f}(n) \ll \frac{1}{n^2}$ and the series $S_N(x)$ has good convergence properties.

What can be said about pointwise convergence for general functions? It is possible for the Fourier series of a continuous function to diverge at a point (see §2.2 of [SS1]). Kolmogorov [Kol] (1926) constructed a function such that $\int_0^1 |f(x)|dx$ is finite and the Fourier series diverges everywhere; however, if $\int_0^1 |f(x)|^2 dx < \infty$, the story is completely different. For such f, Carleson proved that for almost all $x \in [0, 1]$ the Fourier series converges to the original function (see [Ca, Fef]).

Exercise 11.3.10. *Let $\widehat{f}(n) = \frac{1}{2^{|n|}}$. Does $\sum_{-\infty}^{\infty} \widehat{f}(n)e_n(x)$ converge to a continuous, differentiable function? If so, is there a simple expression for that function?*

11.3.3 Parseval's Identity

Theorem 11.3.11 (Parseval's Identity). *Assume $\int_0^1 |f(x)|^2 dx < \infty$. Then*

$$\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 = \int_0^1 |f(x)|^2 dx. \tag{11.43}$$

In other words, Bessel's Inequality (Exercise 11.2.2) is an equality for such f.

We sketch the proof for continuous functions. From the definition of $S_N(x)$ we obtain Bessel's Inequality:

$$0 \leq \int_0^1 (f(x) - S_N(x))\overline{(f(x) - S_N(x))}dx = \int_0^1 |f(x)|^2 - \sum_{|n|\leq N} |\widehat{f}(n)|^2. \tag{11.44}$$

Rearranging yields

$$\int_0^1 |f(x)|^2 dx = \int_0^1 |f(x) - S_N(x)|^2 dx + \sum_{|n|\leq N} |\widehat{f}(n)|^2. \tag{11.45}$$

To complete the proof, we need only show that as $N \to \infty$, $\int_0^1 |f(x) - S_N(x)|^2 dx \to 0$. Note Bessel's Inequality, (11.45), immediately implies

$$\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 < \int_0^1 |f(x)|^2 dx < \infty; \tag{11.46}$$

therefore the sum converges and given any ϵ, for N sufficiently large

$$\sum_{|n|>\sqrt{N}} |\widehat{f}(n)|^2 < \epsilon. \tag{11.47}$$

In the proof of Theorem 11.3.8 we multiplied by 1; we now do another common trick: adding zero in a clever way. By Theorem 11.3.1, given any $\epsilon > 0$ there exists an N_0 such that for all $N > N_0$, $|f(x) - T_N(x)| < \epsilon$. We apply the inequality $|a + b|^2 \leq 4|a|^2 + 4|b|^2$ to $|f(x) - S_N(x)|$ with $a = f(x) - T_N(x)$ and $b = T_N(x) - S_N(x)$. We have

$$\int_0^1 |f(x) - S_N(x)|^2 dx \leq 4\int_0^1 |f(x) - T_N(x)|^2 dx + 4\int_0^1 |T_N(x) - S_N(x)|^2 dx. \tag{11.48}$$

The first term on the right is at most $4\epsilon^2$. To handle the second integral, note

$$T_N(x) = \sum_{n=-N}^{N} \frac{N - |n|}{N} \widehat{f}(n) e^{2\pi i n x}, \tag{11.49}$$

which implies

$$\int_0^1 |T_N(x) - S_N(x)|^2 dx = \sum_{n=-N}^{N} \frac{|n|^2}{N^2} |\widehat{f}(n)|^2. \tag{11.50}$$

Since f is continuous, f is bounded (Theorem A.3.13), hence by Exercise 11.2.2 $\widehat{f}(n)$ is bounded, say by B. The sum in (11.50) can be made arbitrarily small (the terms with $|n| \leq \sqrt{N}$ contribute at most $\frac{2B^2}{N}$, and the remaining contributes at most ϵ by (11.47)).

Exercise 11.3.12. *Fill in the details for the above proof. Prove the result for all f satisfying $\int_0^1 |f(x)|^2 dx < \infty$.*

Exercise 11.3.13. *If $\int_0^1 |f(x)|^2 dx < \infty$, show Bessel's Inequality implies there exists a B such that $|\widehat{f}(n)| \leq B$ for all n.*

Exercise 11.3.14. *Though we used $|a + b|^2 \leq 4|a|^2 + 4|b|^2$, any bound of the form $c|a|^2 + c|b|^2$ would suffice. What is the smallest c that works for all $a, b \in \mathbb{C}$?*

11.3.4 Sums of Series

One common application of pointwise convergence and Parseval's identity is to evaluate infinite sums. For example, if we know at some point x_0 that $S_N(x_0) \to f(x_0)$, we obtain

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{2\pi i n x_0} = f(x_0). \tag{11.51}$$

Additionally, if $\int_0^1 |f(x)|^2 dx < \infty$ we obtain

$$\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 = \int_0^1 |f(x)|^2 dx. \tag{11.52}$$

Thus, if the terms in a series correspond to Fourier coefficients of a "nice" function, we can evaluate the series.

Exercise 11.3.15. *Let $f(x) = \frac{1}{2} - |x|$ on $[-\frac{1}{2}, \frac{1}{2}]$. Calculate $\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2}$. Use this to deduce the value of $\sum_{n=1}^{\infty} \frac{1}{n^2}$. This is often denoted $\zeta(2)$ (see Exercise 3.1.7). See [BP] for connections with continued fractions, and [Kar] for connections with quadratic reciprocity.*

Exercise 11.3.16. *Let $f(x) = x$ on $[0, 1]$. Evaluate $\sum_{n=1}^{\infty} \frac{1}{n^2}$.*

Exercise 11.3.17. *Let $f(x) = x$ on $[-\frac{1}{2}, \frac{1}{2}]$. Prove $\frac{\pi}{4} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)^2}$. See also Exercise 3.3.29; see Chapter 11 of [BB] or [Sc] for a history of calculations of π.*

Exercise 11.3.18. *Find a function to determine $\sum_{n=1}^{\infty} \frac{1}{n^4}$; compare your answer with Exercise 3.1.26.*

11.4 APPLICATIONS OF THE FOURIER TRANSFORM

To each periodic function (say with period 1) we associate its Fourier coefficients

$$\widehat{f}(n) = \int_0^1 f(x) e^{-2\pi i n x}. \tag{11.53}$$

Note the integral is well defined for all n if $\int_0^1 |f(x)|dx < \infty$. If f is continuous and differentiable, we have seen that we can recover f from its Fourier coefficients $\widehat{f}(n)$. We briefly discuss the generalization to non-periodic functions on all of $\mathbb{R}$, the Fourier transform; see [SS1] for complete details.

We give two applications. The first is Poisson Summation, which relates sums of f at integers to sums of its Fourier transform at integers. Often this converts a long, slowly decaying sum to a short, rapidly decaying one (see for example Theorem 3.1.20 on the functional equation of $\zeta(s)$, as well as Theorem 9.4.2 which shows there is digit bias in Geometric Brownian Motions). Poisson Summation is one of the most important tools in a number theorist's arsenal. As a second application we sketch the proof of the Central Limit Theorem (Theorem 8.4.1).

The Fourier transform also appears in Chapter 18 when we investigate the zeros of L-functions. When we derive formulas relating sums of a function f at zeros of an L-function to sums of the product of the coefficients of the L-function times the Fourier transform of f at primes, we shall use some properties of the Fourier transform to show that the sums converge. Relations like this are the starting point of many investigations of properties of zeros of L-functions, primarily because often we can evaluate the sums of the Fourier transform at primes and then use that knowledge to glean information about the zeros.

11.4.1 Fourier Transform

We define the **Fourier transform** by

$$\widehat{f}(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x y} dx; \tag{11.54}$$

(sometimes the Fourier transform is defined with e^{-ixy} instead of $e^{-2\pi i x y}$). Instead of countably many Fourier coefficients, we now have one for each $y \in \mathbb{R}$. While $\widehat{f}(y)$ is well defined whenever $\int_{-\infty}^{\infty} |f(x)|dx < \infty$, much more is true for functions with $\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$.

The **Schwartz Space** $\mathcal{S}(\mathbb{R})$ is the space of all infinitely differentiable functions whose derivatives are rapidly decreasing. Explicitly,

$$\forall j, k \geq 0, \quad \sup_{x \in \mathbb{R}} (|x| + 1)^j |f^{(k)}(x)| < \infty. \tag{11.55}$$

Thus as $|x| \to \infty$, f and all its derivatives decay faster than any polynomial. One can show the Fourier transform of a Schwartz function is a Schwartz function, and

Theorem 11.4.1 (Fourier Inversion Formula). *For $f \in \mathcal{S}(\mathbb{R})$,*

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(y) e^{2\pi i x y} dy. \tag{11.56}$$

In fact, for any $f \in \mathcal{S}(\mathbb{R})$,

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\widehat{f}(y)|^2 dy. \tag{11.57}$$

Definition 11.4.2 (Compact Support). *A function $f : \mathbb{R} \to \mathbb{C}$ has compact support if there is a finite closed interval $[a, b]$ such that for all $x \notin [a, b]$, $f(x) = 0$.*

Remark 11.4.3 (Advanced)**.** Schwartz functions with compact support are extremely useful in many arguments. It can be shown that given any continuous function g on a finite closed interval $[a, b]$, there is a Schwartz function f with compact support arbitrarily close to g; i.e., for all $x \in [a, b]$, $|f(x) - g(x)| < \epsilon$. Similarly, given any such continuous function g, one can find a sum of step functions of intervals arbitrarily close to g (in the same sense as above). Often, to prove a result for step functions it suffices to prove the result for continuous functions, which is the same as proving the result for Schwartz functions. Schwartz functions are infinitely differentiable and as the Fourier Inversion formula holds, we can pass to the Fourier transform space, which is sometimes easier to study.

Exercise 11.4.4. *Show the Gaussian $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ is in $\mathcal{S}(\mathbb{R})$ for any $\mu, \sigma \in \mathbb{R}$.*

Exercise 11.4.5. *Let $f(x)$ be a Schwartz function with compact support contained in $[-\sigma, \sigma]$ and denote its Fourier transform by $\widehat{f}(y)$. Prove for any integer $A > 0$ that $|\widehat{f}(y)| \le c_f y^{-A}$, where the constant c_f depends only on f, its derivatives and σ. As such a bound is useless at $y = 0$, one often derives bounds of the form $|\widehat{f}(y)| \le \frac{\tilde{c}_f}{(1+|y|)^A}$.*

11.4.2 Poisson Summation

We say a function $f(x)$ decays like x^{-a} if there are constants x_0 and C such that for all $|x| > x_0$, $|f(x)| \le C/|x|^a$.

Theorem 11.4.6 (Poisson Summation)**.** *Assume f is twice continuously differentiable and that f, f' and f'' decay like $x^{-(1+\eta)}$ for some $\eta > 0$. Then*

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n), \tag{11.58}$$

where $\hat{f}$ is the Fourier transform of f.

The theorem is true for more general f. We confine ourselves to this useful case, whose conditions are often met in applications. See, for example, Theorem 9.4.2, where Poisson Summation allowed us to replace a long slowly decaying sum with just one term (plus a negligible error), as well as Theorem 3.1.20, where we proved the functional equation of $\zeta(s)$.

It is natural to study $F(x) = \sum_{n\in\mathbb{Z}} f(x+n)$; the theorem follows from understanding $F(0)$. As $F(x)$ is periodic with period 1, it is natural to try to apply our results on Fourier series to approximate F. To use the results from §11.3.2, we need F to be continuously differentiable on $[0, 1]$; however, before we show F is differentiable, we must first show F is continuous and well defined! For example, consider $f(x)$ to be narrow spikes at the integers, say of height n and width $\frac{1}{n^4}$ centered around $x = n$. Note the sum $F(0)$ does not exist.

Exercise 11.4.7. *Consider*

$$f(x) = \begin{cases} n^6\left(\frac{1}{n^4} - |n - x|\right) & \text{if } |x - n| \le \frac{1}{n^4} \text{ for some } n \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases} \tag{11.59}$$

Show $f(x)$ is continuous but $F(0)$ is undefined. Show $F(x)$ converges and is well defined for any $x \notin \mathbb{Z}$.

Lemma 11.4.8. *If $g(x)$ decays like $x^{-(1+\eta)}$ for some $\eta > 0$, then $G(x) = \sum_{n\in\mathbb{Z}} g(x+n)$ converges for all x, and is continuous.*

Exercise 11.4.9. *Prove Lemma 11.4.8.*

Lemma 11.4.10. *If f, f', f'' decay like $x^{-(1+\eta)}$, then $F'(x)$ is continuously differentiable.*

Proof. The natural candidate for $F'(x)$ is $\sum_{n\in\mathbb{Z}} f'(x+n)$. This infinite sum exists and is continuous by Lemma 11.4.8 (applied to f'); it suffices to show that this sum equals $F'(x)$. We shall denote this infinite sum by $F'(x)$, and we now justify this notation by showing it does equal the derivative of $F(x)$. To see this, it suffices to show that for any $\epsilon > 0$ there is a $\delta_0 > 0$ such that for all $|\delta| < \delta_0$ we have

$$\left|\frac{F(x+\delta)-F(x)}{\delta} - F'(x)\right| < \epsilon. \tag{11.60}$$

Without loss of generality we may assume $\delta < 1$. We may write

$$\begin{aligned}\frac{F(x+\delta)-F(x)}{\delta} - F'(x) &= \sum_{|n|\le N}\left[\frac{f(x+n+\delta)-f(x+n)}{\delta} - f'(x+n)\right]\\ &\quad + \sum_{|n|>N}\left[\frac{f(x+n+\delta)-f(x+n)}{\delta} - f'(x+n)\right].\end{aligned} \tag{11.61}$$

By the Mean Value Theorem (Theorem A.2.2),

$$f(x+n+\delta) - f(x+n) = \delta\cdot f'(x+n+c_n),\quad c_n\in[0,\delta], \tag{11.62}$$

and similarly with f replaced by f'. Therefore

$$\begin{aligned}\frac{F(x+\delta)-F(x)}{\delta} - F'(x) &= \sum_{|n|\le N}\left[f'(x+n+c_n) - f'(x+n)\right]\\ &\quad + \sum_{|n|>N}\left[f'(x+n+c_n) - f'(x+n)\right]\\ &= \sum_{|n|\le N}\left[f'(x+n+c_n) - f'(x+n)\right]\\ &\quad + \sum_{|n|>N} f''(x+n+d_n)\end{aligned} \tag{11.63}$$

with $d_n \in [0, \delta]$. Since f'' is of rapid decay, x is fixed and $\delta_0 < 1$, by taking N sufficiently large we can make

$$\sum_{|n|>N} f''(x+n+d_n) \le \sum_{|n|>N}\frac{C}{n^{1+\eta}} \le \frac{2C}{\eta(N-1)^{\eta}} < \frac{\epsilon}{2}. \tag{11.64}$$

Since f' is continuous at $x+n$, we can find $\delta_n < 1$ such that if $|c_n| < \delta_n$ then $|f'(x+n+c_n) - f'(x+n)| < \frac{\epsilon}{2(2N+1)}$. Letting $\delta = \min(1, \delta_0, \min_{|n|\geq N}\delta_n)$, we find

$$\sum_{|n|\leq N}[f'(x+n+c_n) - f'(x+n)] \;<\; \sum_{|n|\leq N}\frac{\epsilon}{2(2N+1)} \;<\; \frac{\epsilon}{2}, \tag{11.65}$$

completing the proof. $\square$

Exercise 11.4.11. *For what weaker assumptions on f, f', f'' is the conclusion of Lemma 11.4.10 still true?*

We have shown that the assumptions in Theorem 11.4.6 imply that F, F' exist and are continuous, and clearly F is periodic of period 1. Let $\widehat{F}(m)$ be the m^{th} Fourier coefficient of F:

$$\widehat{F}(m) = \int_0^1 F(x)e^{-2\pi imx}dx. \tag{11.66}$$

Because F is continuously differentiable for all x, by Theorem 11.3.8

$$F(x) = \sum_{m=-\infty}^{\infty}\widehat{F}(m)e^{2\pi imx}. \tag{11.67}$$

In particular,

$$F(0) = \sum_{m=-\infty}^{\infty}\widehat{F}(m). \tag{11.68}$$

As $F(0) = \sum_n f(n)$, from (11.68) it suffices to show $\widehat{F}(m) = \widehat{f}(m)$. While we use the same notation for the Fourier Transform and the Fourier coefficients, the Fourier Transform is for a function defined on $\mathbb{R}$ (such as f) and the Fourier coefficients are for periodic functions (such as F). We have

$$\widehat{F}(m) = \int_0^1\sum_{n=-\infty}^{\infty}f(x+n)e^{-2\pi imx}dx. \tag{11.69}$$

By Fubini's Theorem (Theorem A.2.8), if we take absolute values of the integrand and either $\int\sum|*|$ or $\sum\int|*|$ exists, then we can interchange order of summation and integration. One must be careful, as it is not always possible to interchange orders. See Exercise 11.4.12.

Hence in the integral-sum for $\widehat{F}(m)$, we find

$$\begin{aligned}\sum_{n=-\infty}^{\infty}\int_0^1|f(x+n)e^{-2\pi imx}| &= \sum_{n=-\infty}^{\infty}\int_0^1|f(x+n)e^{-2\pi imx}|dx\\ &= \sum_{n=-\infty}^{\infty}\int_0^1|f(x+n)e^{-2\pi im(x+n)}e^{2\pi imn}|dx\\ &= \int_{-\infty}^{\infty}|f(x)e^{-2\pi imx}|dx\\ &= \int_{-\infty}^{\infty}|f(x)|dx \;<\; \infty,\end{aligned} \tag{11.70}$$

as f decays like $x^{-(1+\eta)}$. We can therefore interchange the order of integration and summation. Removing the absolute values above gives $\hat{f}(m)$, the Fourier Transform of f evaluated at m. We have shown $\widehat{F}(m) = \hat{f}(m)$, and substituting into (11.68) completes the proof of Theorem 11.4.6. $\square$

While the following exercise is not needed for the investigations above, it indicates how dangerous interchanging orders of summation and integration can be. The reader is advised to study and remember this example!

Exercise 11.4.12. *One cannot always interchange orders of integration. For simplicity, we give a sequence a_{mn} such that $\sum_m(\sum_n a_{m,n}) \neq \sum_n(\sum_m a_{m,n})$. For $m, n \geq 0$ let*

$$a_{m,n} = \begin{cases} 1 & \text{if } n = m \\ -1 & \text{if } n = m+1 \\ 0 & \text{otherwise.} \end{cases} \tag{11.71}$$

Show that the two different orders of summation yield different answers (the reason for this is that the sum of the absolute value of the terms diverges).

Remark 11.4.13. In many problems one wants to interchange a limit and an integral; unfortunately, it is not always the case that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty}\lim_{n\to\infty} f_n(x)dx, \tag{11.72}$$

even if $f_n(x)$ and f are continuous. For example, let

$$f_n(x) = \begin{cases} 0 & \text{if } |x| > 1/n \\ n - 2n^2\left||x| - \frac{1}{2n}\right| & \text{if } |x| \leq 1/n. \end{cases} \tag{11.73}$$

Show $\int_{-\infty}^{\infty} f_n(x)dx = 1$ but $\lim_{n\to\infty} f_n(x) = 0$.

Exercise 11.4.14. *The example in the previous remark has $\lim_{n\to\infty}\max_x |f_n(x)| = \infty$; in other words, there is no M such that $|f_n(x)| \leq M$ for all M and x. Find a family of functions $f_n(x)$ such that*

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dx \neq \int_{-\infty}^{\infty}\lim_{n\to\infty} f_n(x)dx \tag{11.74}$$

and each $f_n(x)$ and $f(x)$ is continuous and $|f_n(x)|, |f(x)| \leq M$ for some M and all x.

11.4.3 Convolutions and Probability Theory

An important property of both Fourier series and Fourier transforms is that they behave nicely under convolution. We denote the convolution of two functions f and g by $h = f * g$, where

$$h(y) = \int_I f(x)g(y-x)dx = \int_I f(x-y)g(x)dx \tag{11.75}$$

and $I = [0,1]$ if f, g are periodic of period 1 and $I = \mathbb{R}$ if $f, g : \mathbb{R} \to \mathbb{C}$. We assume the reader is familiar with the Cauchy-Schwarz inequality (see Appendix A.6). Recall $\langle f, g\rangle = \int_I f(x)\overline{g}(x)dx$, with I as above.

Exercise 11.4.15. *Let f, g be continuous functions on $I = [0, 1]$ or $I = \mathbb{R}$. Show if $\langle f, f\rangle, \langle g, g\rangle < \infty$ then $h = f * g$ exists.* Hint: *Use the Cauchy-Schwarz inequality. Show further that $\widehat{h}(n) = \widehat{f}(n)\widehat{g}(n)$ if $I = [0, 1]$ or if $I = \mathbb{R}$. Thus the Fourier transform converts convolution to multiplication.*

We can now return to the proof of the Central Limit Theorem, Theorem 8.4.1. We assume the reader is familiar with the notations from Chapter 8. The following example is the starting point to the proof of the Central Limit Theorem. Let p be a probability density on $\mathbb{R}$ such that $\langle p, p\rangle < \infty$. Let X_1 and X_2 be two random variables chosen independently with probability density p. Thus the probability of $X_i \in [x, x + \Delta x]$ is $\int_x^{x+\Delta x} p(t)dt$, which is approximately $p(x)\Delta x$. The probability that $X_1 + X_2 \in [x, x + \Delta x]$ is just

$$\int_{x_1 = -\infty}^{\infty} \int_{x_2 = x - x_1}^{x+\Delta x - x_1} p(x_1)p(x_2)dx_2dx_1. \tag{11.76}$$

As $\Delta x \to 0$ we obtain the convolution of p with itself, and find

$$\text{Prob}(X_1 + X_2 \in [a, b]) = \int_a^b (p * p)(z)dz. \tag{11.77}$$

We must justify our use of the word "probability" in (11.77); namely, we must show $p * p$ is a probability density. Clearly $(p * p)(z) \geq 0$, and for any two f, g with $\langle f, f\rangle, \langle g, g\rangle < \infty$,

$$\begin{aligned}
\int_{-\infty}^{\infty} (f * g)(x)dx &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-y)g(y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-y)g(y)dxdy \\
&= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{\infty} f(x-y)dx\right)dy \\
&= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{\infty} f(t)dt\right)dy.
\end{aligned} \tag{11.78}$$

If we take $f = g = p$, the last integrals are 1. We used general f and g as the above arguments yield

Lemma 11.4.16. *The convolution of two "nice" probability densities is a probability density.*

Exercise 11.4.17. *Prove* (11.77).

Exercise 11.4.18 (Important). *If for all $i = 1, 2, \dots$ we have $\langle f_i, f_i\rangle < \infty$, prove for all i and j that $\langle f_i * f_j, f_i * f_j\rangle < \infty$. What about $f_1 * (f_2 * f_3)$ (and so on)? Prove $f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$. Therefore convolution is associative, and we may write $f_1 * \cdots * f_N$ for the convolution of N functions.*

Exercise 11.4.19. *Suppose $X_1, \dots, X_N$ are i.i.d.r.v. from a probability distribution p on $\mathbb{R}$. Determine the probability that $X_1 + \cdots + X_N \in [a, b]$. What must be assumed about p for the integrals to converge?*

11.5 CENTRAL LIMIT THEOREM

As another application of Fourier analysis, we sketch the proof of the Central Limit Theorem (Theorem 8.4.1). We highlight the key steps, but we do not provide detailed justifications (which would require several standard lemmas about the Fourier transform; see for example [SS1]).

For simplicity, we consider the case where we have a probability density p on $\mathbb{R}$ that has mean zero, variance one, finite third moment and is of sufficiently rapid decay so that all convolution integrals that arise converge; see Exercise 15.1.6. Specifically, let p be an infinitely differentiable function satisfying

$$\int_{-\infty}^{\infty} x p(x) dx = 0, \quad \int_{-\infty}^{\infty} x^2 p(x) dx = 1, \quad \int_{-\infty}^{\infty} |x|^3 p(x) dx < \infty. \tag{11.79}$$

Assume $X_1, X_2, \dots$ are independent identically distributed random variables (i.i.d.r.v.) drawn from p; thus, $\text{Prob}(X_i \in [a, b]) = \int_a^b p(x) dx$. Define $S_N = \sum_{i=1}^N X_i$. Recall the standard Gaussian (mean zero, variance one) is $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Theorem 11.5.1 (Central Limit Theorem)**.** *Let X_i, S_N be as above and assume the third moment of each X_i is finite. Then $\frac{S_N}{\sqrt{N}}$ converges in probability to the standard Gaussian:*

$$\lim_{N\to\infty} \text{Prob}\left(\frac{S_N}{\sqrt{N}} \in [a, b]\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \tag{11.80}$$

We sketch the proof. The Fourier transform of p is

$$\widehat{p}(y) = \int_{-\infty}^{\infty} p(x) e^{-2\pi i x y} dx. \tag{11.81}$$

Clearly, $|\widehat{p}(y)| \le \int_{-\infty}^{\infty} p(x) dx = 1$, and $\widehat{p}(0) = \int_{-\infty}^{\infty} p(x) dx = 1$.

Exercise 11.5.2. *One useful property of the Fourier transform is that the derivative of $\widehat{g}$ is the Fourier transform of $2\pi i x g(x)$; thus, differentiation (hard) is converted to multiplication (easy). Explicitly, show*

$$\widehat{g}'(y) = \int_{-\infty}^{\infty} 2\pi i x \cdot g(x) e^{-2\pi i x y} dx. \tag{11.82}$$

If g is a probability density, note $\widehat{g}'(0) = -2\pi i \mathbb{E}[x]$ and $\widehat{g}''(0) = -4\pi^2 \mathbb{E}[x^2]$.

The above exercise shows why it is, at least potentially, natural to use the Fourier transform to analyze probability distributions. The mean and variance (and the higher moments) are simple multiples of the derivatives of $\widehat{p}$ at zero. By Exercise 11.5.2, as p has mean zero and variance one, $\widehat{p}'(0) = 0$, $\widehat{p}''(0) = -4\pi^2$. We Taylor expand $\widehat{p}$ (we do not justify that such an expansion exists and converges; however, in most problems of interest this can be checked directly, and this is the reason we need technical conditions about the higher moments of p), and find near the origin that

$$\widehat{p}(y) = 1 + \frac{p''(0)}{2} y^2 + \cdots = 1 - 2\pi^2 y^2 + O(y^3). \tag{11.83}$$

Near the origin, the above shows $\widehat{p}$ looks like a concave down parabola.

By Exercises 11.4.15, 11.4.18 and 11.4.19, we have the following:

1. The probability that $X_1 + \cdots + X_N \in [a, b]$ is $\int_a^b (p * \cdots * p)(z)dz$.
2. The Fourier transform converts convolution to multiplication. If $\mathrm{FT}[f](y)$ denotes the Fourier transform of f evaluated at y, then we have

$$\mathrm{FT}[p * \cdots * p](y) = \widehat{p}(y) \cdots \widehat{p}(y). \tag{11.84}$$

However, we do not want to study the distribution of $X_1 + \cdots + X_N = x$, but rather the distribution of $S_N = \frac{X_1 + \cdots + X_N}{\sqrt{N}} = x$.

Exercise 11.5.3. *If $B(x) = A(cx)$ for some fixed $c \neq 0$, show $\widehat{B}(y) = \frac{1}{c}\widehat{A}\left(\frac{y}{c}\right)$.*

Exercise 11.5.4. *Show that if the probability density of $X_1 + \cdots + X_N = x$ is $(p * \cdots * p)(x)$ (i.e., the distribution of the sum is given by $p * \cdots * p$), then the probability density of $\frac{X_1 + \cdots + X_N}{\sqrt{N}} = x$ is $(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})$. By Exercise 11.5.3, show*

$$FT\left[(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})\right](y) = \left[\widehat{p}\left(\frac{y}{\sqrt{N}}\right)\right]^N. \tag{11.85}$$

The previous exercises allow us to determine the Fourier transform of the distribution of S_N. It is just $\left[\widehat{p}\left(\frac{y}{\sqrt{N}}\right)\right]^N$. We take the limit as $N \to \infty$ for **fixed** y. From (11.83), $\widehat{p}(y) = 1 - 2\pi^2 y^2 + O(y^3)$. Thus we have to study

$$\left[1 - \frac{2\pi^2 y^2}{N} + O\left(\frac{y^3}{N^{3/2}}\right)\right]^N. \tag{11.86}$$

Exercise 11.5.5. *Show for any fixed y that*

$$\lim_{N\to\infty}\left[1 - \frac{2\pi^2 y^2}{N} + O\left(\frac{y^3}{N^{3/2}}\right)\right]^N = e^{-2\pi^2 y^2}. \tag{11.87}$$

Exercise 11.5.6. *Show that the Fourier transform of $e^{-2\pi^2 y^2}$ at x is $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.* Hint: *This problem requires contour integration from complex analysis.*

We would like to conclude that as the Fourier transform of the distribution of S_N converges to $e^{-2\pi^2 y^2}$ and the Fourier transform of $e^{-2\pi^2 y^2}$ is $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, then the distribution of S_N equalling x converges to $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Justifying these statements requires some results from complex analysis. We refer the reader to [Fe] for details.

The key point in the proof is that we used Fourier Analysis to study the sum of independent identically distributed random variables, as Fourier transforms convert convolution to multiplication. The universality is due to the fact that *only* terms up to the second order contribute in the Taylor expansions. Explicitly, for "nice" p the distribution of S_N converges to the standard Gaussian, independent of the fine structure of p. The fact that p has mean zero and variance one is really just a normalization to study all probability distributions on a similar scale; see Exercise 15.1.6.

The higher order terms are important in determining the *rate* of convergence in the Central Limit Theorem (see [Fe] for details and [KonMi] for an application to Benford's Law).

Exercise 11.5.7. *Modify the proof to deal with the case of p having mean μ and variance σ^2.*

Exercise 11.5.8. *For reasonable assumptions on p, estimate the rate of convergence to the Gaussian.*

Exercise 11.5.9. *Let p_1, p_2 be two probability densities satisfying* (11.79). *Consider $S_N = X_1 + \cdots + X_N$, where for each i, X_1 is equally likely to be drawn randomly from p_1* or *p_2. Show the Central Limit Theorem is still true in this case. What if we instead had a fixed, finite number of such distributions $p_1, \dots, p_k$, and for each i we draw X_i from p_j with probability q_j (of course, $q_1 + \cdots + q_k = 1$)?*

11.6 ADVANCED TOPICS

Below we briefly highlight additional applications of Fourier Series and the Fourier Transform. The first problem complements Dirichlet's Theorem (Theorem 11.3.8) by describing what can go wrong at points where the function is discontinuous. We then give an example of a continuous function that is nowhere differentiable, followed by a proof that of all smooth curves with a given perimeter a circle encloses the most area. We then end with some applications to differential equations. Several of these problems require more mathematical pre-requisites. For more details, see for example [Be, SS1].

Exercise 11.6.1 (Gibbs Phenomenon)**.** *Define a periodic with period 1 function by*

$$f(x) = \begin{cases} -1 & \text{if } -\frac{1}{2} \leq x < 0 \\ 1 & \text{if } 0 \leq x < \frac{1}{2}. \end{cases} \tag{11.88}$$

Prove that the Fourier coefficients are

$$\widehat{f}(n) = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{4}{n\pi i} & \text{if } n \text{ is odd.} \end{cases} \tag{11.89}$$

Show that the N^{th} partial Fourier series $S_N(x)$ converges pointwise to $f(x)$ wherever f is continuous, but overshoots and undershoots for x near 0. Hint: *Express the series expansion for $S_N(x)$ as a sum of sines. Note $\frac{\sin(2m\pi x)}{2m\pi} = \int_0^x \cos(2m\pi t)dt$. Express this as the real part of a geometric series of complex exponentials, and use the geometric series formula. This will lead to*

$$S_{2N-1}(x) = 8\int_0^x \Re\left(\frac{1}{2i}\frac{e^{4n\pi i t}-1}{\sin(2\pi t)}\right)dt = 4\int_0^x \frac{\sin(4n\pi t)}{\sin(2\pi t)}dt, \tag{11.90}$$

which is about 1.179 (or an overshoot of about 18%) when $x = \frac{1}{4n\pi}$. What can you say about the Fejér series $T_N(x)$ for x near 0?

Exercise 11.6.2 (Nowhere Differentiable Function)**.** *Weierstrass constructed a continuous but nowhere differentiable function! We give a modified example and sketch the proof. Consider*

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(2^n \cdot 2\pi x), \quad \frac{1}{2} < a < 1. \tag{11.91}$$

Show f is continuous but nowhere differentiable. Hint: *First show $|a| < 1$ implies f is continuous. Our claim on f follows from: if a periodic continuous function g is differentiable at x_0 and $\widehat{g}(n) = 0$ unless $n = \pm 2^m$, then there exists C such that for all n, $|\widehat{g}(n)| \le Cn2^{-n}$. To see this, show it suffices to consider $x_0 = 0$ and $g(0) = 0$. Our assumptions imply that $(g, e_m) = 0$ if $2^{n-1} < m < 2^{n+1}$ and $m \neq 2^n$. We have $\widehat{g}(2^n) = (g, e_{2^n} F_{2^{n-1}}(x))$ where F_N is the Fejér kernel. The claim follows from bounding the integral $(g, e_{2^n} F_{2^{n-1}}(x))$. In fact, more is true: Baire showed that, in a certain sense, "most" continuous functions are nowhere differentiable! See, for example, [Fol].*

Exercise 11.6.3 (Isoperimetric Inequality). *Let $\gamma(t) = (x(t), y(t))$ be a smooth closed curve in the plane; we may assume it is parametrized by arc length and has length 1. Prove the enclosed area A is largest when $\gamma(t)$ is a circle.* Hint: *By Green's Theorem (Theorem A.2.9),*

$$\oint_\gamma x dy - y dx = 2\text{Area}(A). \tag{11.92}$$

The assumptions on $\gamma(t)$ imply $x(t), y(t)$ are periodic functions with Fourier series expansions and $\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 1$. Integrate this equality from $t = 0$ to $t = 1$ to obtain a relation among the Fourier coefficients of $\frac{dx}{dt}$ and $\frac{dx}{dt}$ (which are related to those of $x(t)$ and $y(t)$); (11.92) *gives another relation among the Fourier coefficients. These relations imply $4\pi\text{Area}(A) \le 1$ with strict inequality unless the Fourier coefficients vanish for $|n| > 1$. After some algebra, one finds this implies we have a strict inequality unless γ is a circle.*

Exercise 11.6.4 (Applications to Differential Equations). *One reason for the introduction of Fourier series was to solve differential equations. Consider the vibrating string problem: a unit string with endpoints fixed is stretched into some initial position and then released; describe its motion as time passes. Let $u(x,t)$ denote the vertical displacement from the rest position x units from the left endpoint at time t. For all t we have $u(0,t) = u(1,t) = 0$ as the endpoints are fixed. Ignoring gravity and friction, for small displacements Newton's laws imply*

$$\frac{\partial^2 u(x,t)}{\partial x^2} = c^2 \frac{\partial^2 u(x,t)}{\partial t^2}, \tag{11.93}$$

where c depends on the tension and density of the string. Guessing a solution of the form

$$u(x,t) = \sum_{n=1}^{\infty} a_n(t) \sin(n\pi x), \tag{11.94}$$

solve for $a_n(t)$.

One can also study problems on $\mathbb{R}$ by using the Fourier Transform. Its use stems from the fact that it converts multiplication to differentiation, and vice versa: if $g(x) = f'(x)$ and $h(x) = xf(x)$, prove that $\widehat{g}(y) = 2\pi i y \widehat{f}(y)$ and $\frac{d\widehat{f}(y)}{dy} = -2\pi i \widehat{h}(y)$. This and Fourier Inversion allow us to solve problems such as the heat equation

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2}, \quad x \in \mathbb{R},\ t > 0 \tag{11.95}$$

with initial conditions $u(x, 0) \quad f(x)$.