Harmonic Analysis: from Fourier to Haar

María Cristina Pereyra

Lesley A. Ward

Department of Mathematics and Statistics, MSC03 2150, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
E-mail address: crisp@math.unm.edu

School of Mathematics and Statistics, University of South Australia, Mawson Lakes SA 5095, Australia
E-mail address: Lesley.Ward@unisa.edu.au

2000 *Mathematics Subject Classification.* Primary 42-xx

CHAPTER 4

Summability methods: Dirichlet kernels and Fejér kernels

Pointwise convergence of partial Fourier sums for continuous functions was ruled out by the Du Bois-Reymond example. However, in an attempt to obtain a convergence result for all continuous functions, mathematicians devised various averaging or *summability methods*. These methods require only knowledge of the Fourier coefficients in order to recover a continuous function as the sum of an appropriate trigonometric series. Along the way, a number of very important approximation techniques in analysis were developed, in particular *convolutions* (Section 4.1), and *approximations of the identity*, also known as *good kernels* (Sections 4.4 and 7.5). Roughly speaking, an approximation of the identity is a sequence of functions whose mass is concentrating near the origin (very much like a delta function[1]); a precise definition is given below. We discuss these techniques in detail in the context of 2π-periodic functions. Later we will revisit these ideas in the context of functions defined on the line (Section 7.5).

Here we describe several kernels that arise naturally in the theory of Fourier series. Some of them (the Fejér and Poisson[2] kernels) are good kernels that generate approximations of the identity; another (the Dirichlet kernel) is equally important but not good in this sense. As we will see, the fact that the Dirichlet kernel is not a good kernel accounts for the difficulties in achieving pointwise convergence for continuous functions.

For a captivating account of the history surrounding the use of summability methods for Fourier series, see [**KL**, Chapter 7].

4.1. Partial Fourier sums and the Dirichlet kernel D_N

We compute an integral representation for $S_N f$, the N^{th} partial Fourier sum of a suitable function f. Note that in this calculation we may interchange the summation and the integral, since there are only finitely many terms in the sum. The notations D_N and $*$ will be explained below. Recall that the N^{th} partial Fourier sum of an integrable function $f : \mathbb{T} \to \mathbb{C}$ is given by

$$S_N f(\theta) = \sum_{|n|\leq N} a_n e^{in\theta},$$

where the Fourier coefficients are calculated using the formula

$$a_n := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iny} dy.$$

[1]The so-called delta function is not a function at all, but rather an example of the kind of generalized function known as a distribution. We define it in Section 8.4.

[2]Named after the French mathematician Simeón Denis Poisson (1781–1840).

Inserting the formula for a_n into the right-hand side and using the linearity of the integral and the properties of the exponential, we obtain a formula for $S_N f$ as an integral involving $f(y)$:

$$\begin{aligned} S_N f(\theta) &= \sum_{|n|\leq N} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iny} dy \right) e^{in\theta} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \sum_{|n|\leq N} e^{-iny} e^{in\theta}\, dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \sum_{|n|\leq N} e^{in(\theta-y)}\, dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_N(\theta - y)\, dy \\ &=: (f * D_N)(\theta). \end{aligned}$$

Here we were led to define the function

$$D_N(\theta) := \sum_{|n|\leq N} e^{in\theta}. \tag{4.1}$$

This function D_N is called the *Dirichlet kernel*. Notice that D_N does not depend on f. Also, D_N is a trigonometric polynomial of degree N, with coefficients equal to 1 for $-N \leq n \leq N$, and 0 for all other values of n. The Dirichlet kernel D_N is 2π-periodic.

In the last line of our calculation we used the idea of *convolving* two functions. If f, $g : \mathbb{T} \to \mathbb{C}$ are integrable, their *convolution* is the new integrable[3] function $f * g : \mathbb{T} \to \mathbb{C}$ given by

$$(f * g)(\theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(\theta - y)\, dy.$$

We study convolution in more detail in Section 4.3. With these definitions, we have shown that the N^{th} partial Fourier sum of f is given by

$$S_N f(\theta) = (f * D_N)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_N(\theta - y)\, dy. \tag{4.2}$$

The Dirichlet kernel and convolution are both essential in what follows. We discuss their properties, beginning with the Dirichlet kernel.

Aside 4.1. *In this chapter we use the term* integrable *without specifying whether we mean Riemann-integrable or Lebesgue-integrable. Most results do hold for both types of integrability. Some results we have stated and proved specifically for integrable and bounded functions. We do use in our proofs the assumption that the functions are bounded. If the functions are Riemann integrable then they are bounded by definition, but this is not true for Lebesgue-integrable functions. In those instances where we relied on boundedness a different argument is required to verify the result for all Lebesgue-integrable functions, and we point the reader to appropriate references.*

[3]Riemann-integrable functions on bounded intervals are closed under translations, reflections and products. Hence $\mathcal{R}(\mathbb{T})$ is closed under convolutions. On the other hand, the set of Lebesgue-integrable functions on $\mathbb{T}$ is not closed under products; however, it is closed under convolutions.

4.2. Properties of the Dirichlet kernel

We begin with a formula for the Dirichlet kernel in terms of cosines. Using the formula $2\cos x = e^{ix} + e^{-ix}$, we see that

$$\begin{aligned} D_N(\theta) &= e^{-iN\theta} + e^{-i(N-1)\theta} + \cdots + e^{-i\theta} + e^{0} + e^{i\theta} + \cdots + e^{i(N-1)\theta} + e^{iN\theta} \\ &= 1 + 2\sum_{n=1}^{N} \cos(n\theta). \end{aligned}$$

In particular, the Dirichlet kernel D_N is real-valued and even.

Next, we obtain a convenient and concise closed formula for D_N as a quotient of two sines:

$$\begin{aligned} D_N(\theta) &= \sum_{|n|\leq N} e^{in\theta} \\ &= e^{-iN\theta} \sum_{n=0}^{2N} (e^{i\theta})^n \\ &= e^{-iN\theta}\, \frac{(e^{i\theta})^{2N+1} - 1}{e^{i\theta} - 1} \qquad \text{(sum of a finite geometric series)} \\ &= \frac{e^{i(N+1)\theta} - e^{-iN\theta}}{e^{i\theta} - 1} \cdot \frac{e^{-i\theta/2}}{e^{-i\theta/2}} \\ &= \frac{e^{i(2N+1)\theta/2} - e^{-i(2N+1)\theta/2}}{e^{i\theta/2} - e^{-i\theta/2}}\,, \end{aligned}$$

Here we have used the typical analysis trick of multiplying and dividing by the same quantity, in this case $e^{-i\theta/2}$. Now we use the formula $2i\sin\theta = e^{i\theta} - e^{-i\theta}$ to conclude that

$$D_N(\theta) = \frac{\sin\left((2N+1)\theta/2\right)}{\sin\left(\theta/2\right)}\,. \tag{4.3}$$

The quotient of sines has a removable singularity at the origin $\theta = 0$. We fill in the appropriate value $D_N(0) = 2N + 1$, and use it without further comment. The kernel D_N has no other singularities on $[-\pi, \pi]$, and in fact D_N is C^∞ on the unit circle $\mathbb{T}$.

FIGURE 4.1. Graphs of Dirichlet kernels D_N for $N = 1$, 3, and 8.

In Figure 4.1 we see that the oscillations of D_N, as $\sin\left((2N+1)\theta/2\right)$ varies between 1 and -1, lie within the envelope $\pm 1/\sin\left(\theta/2\right)$. Also, D_N takes some positive and some negative values, D_N is even, and D_N achieves its maximum value of $2N + 1$ at $\theta = 0$.

THEOREM 4.2. *The Dirichlet kernel has the following properties.*

(i) $(f * D_N)(\theta) = S_N f(\theta)$, *for* $\theta \in \mathbb{T}$ *and integrable* $f : \mathbb{T} \to \mathbb{C}$.

(ii) D_N *is an even function of* θ.

(iii) D_N *has mean value 1 for all* N*:*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\theta)\, d\theta = 1. \tag{4.4}$$

(iv) *However, the integral of the absolute value of D_N depends on N, and in fact*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(\theta)|\,d\theta \approx \log N. \tag{4.5}$$

We are using the notation $A_N \approx B_N$ (to be read A_N is comparable to B_N) if there exist constants $c, C > 0$ such that for all N $cB_N \leq A_N \leq CB_N$.

It may seem surprising that the averages of D_N can be uniformly bounded while the averages of $|D_N|$ grow like $\log N$; see Figure 4.2. This growth is possible because of the increasing oscillation of D_N with N, which allows the total area between the graph of D_N and the θ-axis to grow without bound as $N \to \infty$, while cancellation of the positive and negative parts of this area keeps the integral of D_N constant.

FIGURE 4.2. The value of the integral of the absolute value of the Dirichlet kernel D_N, for $0 \leq N \leq 15$. Note the logarithmic growth.

PROOF. We have already established properties (i) and (ii). For property (iii), note that

(4.6)
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\theta)\,d\theta = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \sum_{|n|\leq N} e^{in\theta}\right] d\theta = \sum_{|n|\leq N}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{in\theta}\,d\theta\right] = 1,$$

since the integrals in the second last term are zero unless $n = 0$, and 1 in that case.

Verifying property (iv) is left to the reader; see Exercise 4.3. See Figure 4.2 for numerical evidence. □

EXERCISE 4.3. Justify equation (4.5) analytically. It may be useful to remember that by the integral test,

$$\sum_{N=1}^{N} \frac{1}{n} \approx \int_1^N \frac{1}{x}\,dx = \log N.$$

◊

4.3. Convolution

Convolution can be thought of as a way to create new functions from old ones, alongside the familiar methods of addition, multiplication, and composition of functions. We recall the definition.

DEFINITION 4.4. Given two integrable, 2π-periodic functions f, $g : \mathbb{T} \to \mathbb{C}$, their *convolution on* $\mathbb{T}$ is the new function $f * g : \mathbb{T} \to \mathbb{C}$ given by

$$(f * g)(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)g(\theta - y)\,dy. \tag{4.7}$$

◊

Notice that we integrate over the dummy variable y, leaving a function of θ. The point $\theta - y$ will not always lie in $[-\pi, \pi)$, but since the integrand is 2π-periodic ($g(\theta) = g(\theta + 2\pi k)$ for all integers k) we still know the values of the integrand everywhere on $\mathbb{R}$. Pictorially, we compute $f * g$ at a point θ by reflecting the graph of $g(y)$ in the vertical axis and translating it to the left by θ units to obtain $g(\theta - y)$, multiplying by $f(y)$, then finding the area under the graph of the resulting product $f(y)g(\theta - y)$ over $y \in [-\pi, \pi)$ and dividing by the normalizing constant 2π. See Figures 4.3 and 4.4.

FIGURE 4.3. Reflection and translation. Graphs over $[-3\pi, 3\pi]$ of a function $g(x)$, its reflection $g(-x)$, and its translated reflection $g(1.5 - x)$. To clarify their 2π-periodic nature, the functions are graphed over $[-3\pi, 3\pi]$ rather than over $[-\pi, \pi]$.

FIGURE 4.4. The ingredients of a convolution. Graphs over $[-\pi, \pi]$ of the same function $g(x)$ as in Figure 4.3, its translated reflection $g(1.5 - x)$, another function $f(x)$, and the product $f(x)g(1.5 - x)$. The value $f * g(1.5)$ of the convolution $f * g(t)$ at $t = 1.5$ is the area under the graph of $f(x)g(1.5 - x)$, divided by $2 * \pi$.

Notice that if the periodic function g is integrable on $\mathbb{T}$, then so is the new periodic function $h(y) := g(\theta - y)$. Also, the product of two Riemann-integrable periodic functions on $\mathbb{T}$ is also periodic and integrable. See Section 2.1.1. This last statement is not true for Lebesgue-integrable functions. Nevertheless the convolution of two functions $f, g \in L^1(\mathbb{T})$ is always a function in $L^1(\mathbb{T})$, and the following inequality[4] holds, see Exercise 4.11,

$$\|f * g\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})} \|g\|_{L^1(\mathbb{T})}. \tag{4.8}$$

The convolution of two integrable functions is always integrable, and so we can calculate its Fourier coefficients.

EXAMPLE 4.5. (*Convolving with a Characteristic Function*) We compute a convolution in the special case where one of the convolution factors is the characteristic function $\chi_{[a,b]}(x)$ of a closed interval $[a, b]$ in $\mathbb{T}$. See Figure 4.5. Recall that the characteristic function is defined by

$$\chi_{[a,b]}(\theta) = \begin{cases} 1, & \text{if } \theta \in [a, b]; \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

FIGURE 4.5. Graph of the characteristic function $\chi_{[a,b]}$ of the interval $[a, b] = [-0.5, 2.3]$.

[4]This is an instance of Young's inequality; see Chapters 7 and 11.

Using the observation that

$$a \leq \theta - y \leq b \qquad \Longleftrightarrow \qquad \theta - b \leq y \leq \theta - a,$$

we see that

$$\begin{aligned}(f * \chi_{[a,b]})(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\chi_{[a,b]}(\theta - y)\,dy \\ &= \frac{1}{2\pi}\int_{\theta-b}^{\theta-a} f(y)\,dy \\ &= \frac{b-a}{2\pi}\,\frac{1}{b-a}\int_{\theta-b}^{\theta-a} f(y)\,dy.\end{aligned}$$

So the convolution of f with $\chi_{[a,b]}$, evaluated at θ, turns out to be a multiple of the average value of f on the reflected and translated interval $[\theta - b, \theta - a]$. ◊

EXERCISE 4.6. (*Convolving Again and Again*) Let $f = \chi_{[0,1]}$. Compute $f * f$ and $f * f * f$. Use the following MATLAB script [***script has to be made] to plot f, $f * f$, and $f * f * f$. Notice how the smoothness improves as we take more convolutions. ◊

This is an example of the so-called *smoothing properties* of convolution. See Section 4.3.2.

4.3.1. Properties of convolution. We summarize the main properties of convolution.

THEOREM 4.7. *Let f, g, $h : \mathbb{T} \to \mathbb{C}$ be 2π-periodic integrable functions, and let $c \in \mathbb{C}$ be a constant. Then*

(i) $f * g = g * f$ (commutative)
(ii) $f * (g + h) = (f * g) + (f * h)$ (distributive)
(iii) $(cf) * g = c(f * g) = f * (cg)$ (homogeneous)
(iv) $(f * g) * h = f * (g * h)$ (associative)
(v) $\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n)$ (Fourier process converts convolution to multiplication)

The first four properties are obtained by manipulating the integrals (change of variables, interchanging integrals, etc). The last property requires some work. The idea is first to prove these properties for *continuous* functions, and then to extend to *integrable* functions, by approximating them with continuous functions. This is an example of the usefulness of approximating functions by nicer functions[5]. We will use the following approximation lemmas.

LEMMA 4.8. *Suppose $f : \mathbb{T} \to \mathbb{C}$ is an integrable and bounded function. Let B be a bound for f. In other words, $|f(\theta)| \leq B$ for all $\theta \in \mathbb{T}$. Then there is a sequence $\{f_k\}_{k=1}^{\infty}$ of continuous functions $f_k : \mathbb{T} \to \mathbb{C}$ such that*

(i) $\sup_{\theta\in\mathbb{T}} |f_k(\theta)| \leq B$ *for all* $k \in \mathbb{N}$, *and*
(ii) $\int_{-\pi}^{\pi} |f(\theta) - f_k(\theta)|\,d\theta \to 0$ *as* $k \to \infty$.

[5]Approximation by nice or simpler functions, such as continuous functions, trigonometric polynomials, or step functions, is a recurrent theme in analysis and in this book. We already encountered this theme in Chapter 2 and we will encounter it again in the context of mean-square convergence of Fourier series, Fourier integrals, Weierstrass's Theorem, and orthogonal bases, in particular the Haar basis and wavelet bases.

Part (i) of the lemma says that the f_k are also bounded by B, and part (ii) says, in language we already met in Section 2, that the f_k converge to f in $L^1(\mathbb{T})$. If $f \in \mathcal{R}(\mathbb{T})$, then it is bounded, and the Lemma says that *the continuous functions on* $\mathbb{T}$ *are dense in the set of Riemann-integrable functions on* $\mathbb{T}$. See Theorem 2.65, in particular Remark 2.66.

LEMMA 4.9. *Suppose $f, g : \mathbb{T} \to \mathbb{C}$ are bounded and integrable functions (bounded by B). Let $\{f_k, g_k\}_{k=1}^{\infty}$ be sequences of continuous functions (bounded by B) approximating f and g respectively in the L^1-norm, as in Lemma 4.8. Then $f_k * g_k$ converges uniformly to $f * g$.*

These two lemmas work for functions that are assumed to be Riemann integrable (hence bounded) or for bounded Lebesgue integrable functions.

We will prove Lemma 4.9 after we prove Theorem 4.7.

PROOF OF THEOREM 4.7. Property (i), rewritten slightly, says that

$$\begin{aligned} (f * g)(\theta) &:= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(\theta - y)\, dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta - y) g(y)\, dy, \end{aligned} \tag{4.10}$$

so that 'the $\theta - y$ can go with g *or* with f'. Equation (4.10) follows from the change of variables $y' = \theta - y$, and the 2π-periodicity of the integrand.

Properties (ii) and (iii) show that convolution is linear in the second variable, coupled with the commutativity Property (i), one deduces the linearity in the first variable as well. Hence convolution is a bilinear operation. Properties (ii) and (iii) are consequences of the linearity of the integral. Property (iv), associativity, is left as an exercise for the reader.

We prove property (v), that the n^{th} Fourier coefficient of the convolution $f * g$ is the product of the n^{th} coefficients of f and of g. For *continuous* f and g, we may interchange the order of integration by Fubini's Theorem 2.59 for continuous functions, obtaining

$$\begin{aligned} \widehat{f * g}(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f * g)(\theta)\, e^{in\theta}\, d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(\theta - y)\, dy \right] e^{-in\theta}\, d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta - y) e^{-in\theta}\, d\theta \right] dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\, e^{-iny} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta - y) e^{-in\theta}\, e^{iny}\, d\theta \right] dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\, e^{-iny} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta - y) e^{-in(\theta - y)}\, d\theta \right] dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\, e^{-iny} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta') e^{-in(\theta')}\, d\theta' \right] dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)\, e^{-iny}\, \widehat{g}\,(n)\, dy \\ &= \widehat{f}\,(n)\, \widehat{g}\,(n). \end{aligned}$$

In the second last line we changed variables to $\theta' = \theta - y$. Since the integrand is 2π-periodic we may keep the limits of integration unchanged.

We have established property (v) for continuous functions. Now suppose f and g are integrable and bounded. Take sequences $\{f_k\}$ and $\{g_k\}$ of continuous functions such that $f_k \to f$, $g_k \to g$, as in the Approximation Lemma (Lemma 4.8).

First, $\widehat{f_k}(n) \to \widehat{f}(n)$ for each $n \in \mathbb{Z}$, since by Lemma 4.8(ii)

$$|\widehat{f_k}(n) - \widehat{f}(n)| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} [f_k(\theta) - f(\theta)]\, e^{-in\theta}\, d\theta \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(\theta) - f(\theta)|\, d\theta \to 0$$

as $k \to \infty$. Similarly $\widehat{g_k}(n) \to \widehat{g}(n)$ for each $n \in \mathbb{Z}$.

Second, by property (v) for continuous functions,

$$\widehat{f_k * g_k}(n) = \widehat{f_k}(n)\, \widehat{g_k}(n)$$

for each $n \in \mathbb{Z}$ and each $k \in \mathbb{N}$.

Third, we show that as $k \to \infty$

$$\widehat{f_k * g_k}(n) \to \widehat{f * g}(n)$$

for each $n \in \mathbb{Z}$. For

$$\begin{aligned}
\lim_{k\to\infty} \widehat{f_k * g_k}(n) &= \lim_{k\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} (f_k * g_k)(\theta)\, e^{in\theta}\, d\theta \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \lim_{k\to\infty} (f_k * g_k)(\theta)\, e^{in\theta}\, d\theta \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[(f * g)(\theta)\, e^{in\theta}\right] d\theta \\
&= \widehat{f * g}(n).
\end{aligned}$$

Here we can move the limit inside the integral since, by Lemma 4.9, the integrands converge uniformly on $\mathbb{T}$ (see Theorem 2.49).

Hence

$$\widehat{f * g}(n) = \lim_{k\to\infty} \widehat{f_k * g_k}(n) = \left(\lim_{k\to\infty} \widehat{f_k}(n)\right) \left(\lim_{k\to\infty} \widehat{g_k}(n)\right) = \widehat{f}(n)\, \widehat{g}(n),$$

as required.

For Lebesgue-integrable functions, the L^1-version of Fubini's Theorem is needed to carry out the calculation we did for continuous functions. We then do not need to use the approximation argument explicitly, because it is used implicitly in the proof of Fubini's theorem for L^1 functions. □

EXERCISE 4.10. (*Convolution is Associative*) Prove property (iv). ◊

PROOF OF LEMMA 4.9. We show that $f_k * g_k \to f * g$ uniformly on $\mathbb{T}$. Recall that f_k, g_k and f, g are all periodic functions bounded by $B > 0$ on $\mathbb{T}$, and that

$\int_{-\pi}^{\pi} |f_k - f| \to 0$, and $\int_{-\pi}^{\pi} |g_k - g| \to 0$ as $k \to \infty$ by Lemma 4.8, hence

$$\begin{aligned}
&|(f_k * g_k)(\theta) - (f * g)(\theta)| \\
&\quad \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y)g_k(\theta - y) - f(y)g(\theta - y)|\, dy \\
&\quad \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y)|\, |g_k(\theta - y) - g(\theta - y)|\, dy \\
&\qquad\qquad + \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, |g(\theta - y)|\, dy \\
&\quad \leq \frac{B}{2\pi} \left(\int_{-\pi}^{\pi} |g_k(\theta - y) - g(\theta - y)|\, dy + \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, dy \right) \\
&\quad = \frac{B}{2\pi} \left(\int_{-\pi}^{\pi} |g_k(y) - g(y)|\, dy + \int_{-\pi}^{\pi} |f_k(y) - f(y)|\, dy \right) \\
&\quad \to 0
\end{aligned}$$

as $k \to \infty$. The last integrals are independent of θ because of the periodicity of g and g_k, and so the convergence is uniform. Also, the second inequality is a consequence of the *Triangle Inequality*[6] for complex numbers and the *additivity*[7] of the integral. □

EXERCISE 4.11. Use Property (v) at $n = 0$ to show inequality 4.8.

4.3.2. Convolution is a smoothing operation. Convolution is a so-called *smoothing operation*. We will see that the smoothness of the convolution is the combined smoothness of the convolved functions (see Exercise 4.14). (This is how in theory you would like marriages to be: the marriage combines the smoothness of each partner to produce an even smoother couple).

The following result is a precursor of that principle. We start with at least one continuous function and the convolution preserves continuity. It is not apparent that there has been any improvement in the smoothness of the convolution compared to the smoothness of the convolved functions. However an approximation argument allows us to start with a pair of integrable and bounded functions not necessarilly continuous, and their convolution will be continuous.

LEMMA 4.12. *If f, $g \in C(\mathbb{T})$, then $f * g \in C(\mathbb{T})$. In fact, if $f \in C(\mathbb{T})$ and g is integrable, then $f * g \in C(\mathbb{T})$.*

PROOF. Given $f \in C(\mathbb{T})$, and g integrable on $\mathbb{T}$. First, f is uniformly continuous. That is, given $\varepsilon > 0$ there is $\delta > 0$ such that if $|h| < \delta$, then $|f(\theta+h)-f(\theta)| < \varepsilon$ for all $\theta \in \mathbb{T}$. Second, by the linearity of the integral,

$$(f * g)(\theta + h) - (f * g)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \big(f(\theta + h - y) - f(\theta - y)\big) g(y)\, dy.$$

[6] Namely, for a, $b \in \mathbb{C}$, $|a + b| \leq |a| + |b|$.

[7] Additivity means that if f, $g \in L^1(\mathbb{T})$, then $\int_{-\pi}^{\pi} (f + g) = \int_{-\pi}^{\pi} f + \int_{-\pi}^{\pi} g$.

Third, using the triangle inequality for integrals[8], and the uniform continuity of f, we conclude that for all $|h| < \delta$

$$\begin{aligned} |(f*g)(\theta+h)-(f*g)(\theta)| &\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\big|f(\theta+h-y)-f(\theta-y)\big||g(y)|\,dy \\ &\leq \frac{\varepsilon}{2\pi}\int_{-\pi}^{\pi}|g(y)| = \frac{\varepsilon}{2\pi}\|g\|_{L^1(\mathbb{T})}. \end{aligned}$$

Therefore the convolution of a continuous function and an integrable function is continuous. In particular the convolution of two continuous functions is continuous. □

ASIDE 4.13. *Let f and g be integrable and bounded functions, and let f_k and g_k be continuous and bounded functions approximating f and g as in Lemma 4.8. Then $f_k * g_k$ is continuous (by Lemma 4.12), and $f * g$ is the uniform limit of continuous functions (by Lemma 4.9). We conclude, by Theorem 2.55 that $f * g$ is continuous.*

We have shown (see Aside 4.13)

> *The convolution of two integrable and bounded functions on $\mathbb{T}$ is continuous.*

This is an example of how *convolution improves smoothness*. Even more is true: if the convolved functions are smooth then the convolution absorbs the smoothness from each of them, as the following exercise illustrates.

EXERCISE 4.14. (*Convolution Improves Smoothness.*) Suppose $f, g : \mathbb{T} \to \mathbb{C}$, $f \in C^k(\mathbb{T})$, $g \in C^m(\mathbb{T})$. Show that $f * g \in C^{k+m}(\mathbb{T})$. Furthermore the following formula holds:

$$(f*g)^{(k+m)} = f^{(k)} * g^{(m)}.$$

It suffices to assume that the functions $f^{(k)}$ and $g^{(m)}$ are bounded and integrable to conclude that $f * g \in C^{k+m}(\mathbb{T})$.

Hint: Check first for $k = 1$, $m = 0$, then by induction on k check for all $k \geq 0$ and $m = 0$. Now use commutativity of the convolution to verify the statement for $k = 0$ and any $m \geq 0$. Finally put all these facts together to get the desired conclusion. ♢

Another instance of how convolution keeps "the best features" from each function is the content of the following exercise.

EXERCISE 4.15. (*Convolution with a Trigonometric Polynomial Yields a Trigonometric Polynomial.*) Show that if f is integrable and $P \in \mathcal{P}_N$ then $f * P \in \mathcal{P}_N$. ♢

Property (v) in Theorem 4.7 is another instance of the time–frequency dictionary. Earlier, we showed that *differentiation is transformed into polynomial multiplication*. Here we show that *convolution is transformed into the ordinary product*. In the next exercise we have yet another instance of this interplay: *translations are transformed into modulations.*

---

[8]That is, the fact that $|\int_T f| \leq \int_T |f|$.

EXERCISE 4.16. (*Translation Corresponds to Modulation.*) Prove that the process of computing Fourier coefficients converts *translation* to *modulation*. In other words, if f is 2π-periodic and integrable on $\mathbb{T}$, and the translation operator τ_h is defined by

$$\tau_h f(\theta) := f(\theta - h) \qquad \text{for } h \in \mathbb{R},$$

then $\tau_h f$ is also 2π-periodic and integrable on $\mathbb{T}$. Moreover,

$$\widehat{\tau_h f}(n) = e^{-inh}\widehat{f}(n).$$

So translation of f by the amount h has the effect of multiplying the n^{th} Fourier coefficient of f by e^{inh}, for $n \in \mathbb{Z}$. $\Diamond$

We summarize the *time–frequency dictionary* for Fourier coefficients of periodic functions in Table 4.1. Note that f is treated as periodic.

TABLE 4.1. A time–frequency dictionary for Fourier series.

Time/Space $\theta \in \mathbb{T}$	**Frequency** $n \in \mathbb{Z}$
derivative	polynomial
$f'(\theta)$	$\widehat{f'}(n) = in\widehat{f}(n)$
circular convolution	product
$(f * g)(\theta)$	$\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n)$
translation/shift	modulation
$\tau_h f(\theta) = f(\theta - h)$	$\widehat{\tau_h f}(n) = e^{-isn}\widehat{f}(n)$

We will study other versions of convolution. For convolution of vectors in $\mathbb{C}^n$, see Chapter 6, and for convolution of integrable functions defined on the whole real line and not necessarily periodic, see Section 7.5. In these contexts there is a Fourier theory and the same phenomenon occurs: the Fourier transform converts convolution to multiplication.

4.4. Good kernels, or approximations of the identity

DEFINITION 4.17. A family $\{K_n\}_{n=1}^{\infty}$ of real-valued integrable functions on the circle, $K_n : \mathbb{T} \to \mathbb{R}$, is a *family of good kernels* if it satisfies these three properties:

(a) The K_n all have mean value 1:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(\theta)\,d\theta = 1 \qquad \text{for all } n \in \mathbb{N}.$$

(b) There is some $M > 0$ such that

$$\int_{-\pi}^{\pi} |K_n(\theta)|\,d\theta \le M \qquad \text{for all } n \in \mathbb{N}.$$

(c) For each $\delta > 0$,

$$\int_{\delta \le |\theta| \le \pi} |K_n(\theta)|\,d\theta \to 0 \qquad \text{as } n \to \infty.$$

$\Diamond$

For brevity, we may simply say that K_n is a good kernel, without explicitly mentioning the family $\{K_n\}_{n=1}^{\infty}$ of kernels.

A family of good kernels is often called an *approximation of the identity.* After Theorem 4.21 below, we explain why.

Condition (b) says that the mean values of the absolute value $|K_n|$ are uniformly bounded. Condition (c) says that for each fixed positive δ, the total (unsigned) area between the graph of K_n and the θ-axis, more than distance δ from the origin, tends to zero. A less precise way to say this is that as $n \to \infty$, most of the 'mass' of K_n concentrates as near to $\theta = 0$ as we please.

ASIDE 4.18. *Some authors use a more restrictive definition of approximations of the identity, replacing conditions* (b) *and* (c) *by*

(b') K_n *is positive:*

$$K_n(\theta) \geq 0 \qquad \text{for all } \theta \in \mathbb{T}, \text{ for all } n \in \mathbb{N}.$$

(c') *For each* $\delta > 0$, $K_n(\theta) \to 0$ *uniformly on* $\delta \leq |\theta| \leq \pi$.

EXERCISE 4.19. Show that a family of kernels satisfying (a), (b'), and (c') is a family of good kernels according to our definition. Show that conditions (a) and (b') imply condition (b), and condition (c') implies condition (c). $\lozenge$

A canonical way to generate a good family of kernels from one function and its dilations is described in the following exercise.

EXERCISE 4.20. Suppose K is a continuous function on $\mathbb{R}$ that is zero for all $|\theta| \geq \pi$. Assume $\int_{-\pi}^{\pi} |K(\theta)|\, d\theta = 1$. Let $K_n(\theta) := nK(n\theta)$ for $-\pi \leq \theta \leq \pi$. Verify that $\{K_n\}_{n\geq 1}$ is a family of good kernels in $\mathbb{T}$. $\lozenge$

Why is a family of kernels satisfying conditions (a), (b), and (c) called *good*? In the context of Fourier series, because such a family allows us to recover the values of a continuous function $f : \mathbb{T} \to \mathbb{C}$ from its Fourier coefficients. This mysterious statement will become clear in Section 4.5. In the present Fourier-free context, the kernels are called good because they provide a constructive way of producing approximating functions that are smoother than the limit function.

THEOREM 4.21. *Let* $\{K_n(\theta)\}_{n=1}^{\infty}$ *be a family of good kernels,* $K_n : \mathbb{T} \to \mathbb{R}$, *and let* $f : \mathbb{T} \to \mathbb{C}$ *be integrable and bounded. Then*

$$\lim_{n\to\infty} (f * K_n)(\theta) = f(\theta)$$

at all points of continuity of f. *Furthermore, if* f *is continuous on the whole circle* $\mathbb{T}$, *then* $f * K_n \to f$ *uniformly on* $\mathbb{T}$.

This result explains the use of the term *approximation of the identity* for a family of good kernels K_n, for it says that the convolutions of these kernels K_n with f converge, as $n \to \infty$, to the function f again. Moreover, a priori, $f * K_n$ is at the very least continuous. If we can locate a kernel that is k-times differentiable then $f * K_n$ will be at least C^k, see Exercise 4.14. The second part of the Theorem 4.21 will say that we can approximate uniformly continuous functions by k-times continuously differentiable functions. If we could find a kernel that is a trigonometric polynomial, then $f * K_n$ will be itself a trigonometric polynomial, see Exercise 4.15 The second part of the Theorem 4.21 will say that we can approximate uniformly continuous functions by trigonometric polynomials, and Weierstrass' Theorem 3.4 will be proved.

FIGURE 4.6. Graph of $K_n(y)$ for large n—mass concentrates at origin. Graph of $K_n(\theta - y)$—mass concentrates at θ. Graph of $f(y)K_n(\theta - y)$ illustrating how the integral only 'sees' the part of f near $y = \theta$.

Before proving Theorem 4.21, let us indulge in a flight of fancy. Suppose for a moment that the Dirichlet kernel D_N were a good kernel. (Aside: The subscript change from n in K_n to N in D_N is just to echo the notation in our earlier discussion of the Dirichlet kernel.) Then Theorem 4.21 would imply that the Fourier partial sums $S_N f$ of a continuous function f on $\mathbb{T}$ would converge to f, since we would have

$$S_N f(\theta) = \sum_{|n| \leq N} \widehat{f}(n) = (f * D_N)(\theta) \to f(\theta)$$

at all points $\theta \in \mathbb{T}$.

We already know that this dream cannot be real, since $S_N f(0) \not\to f(0)$ for du Bois-Reymond's function, which is continuous on $\mathbb{T}$. However, we will see below that we can salvage something from what we learn about good kernels. In particular, we *can* recover $f(\theta)$ at points of continuity θ of f, using a modification of the Fourier partial sums. (The only information we need is the Fourier coefficients.) This is Fejér's Theorem about Cesàro sums, discussed in Section 4.5 below, and already mentioned in Chapter 3.

EXERCISE 4.22. (*The Dirichlet Kernel is Not a Good Kernel*) Determine which of the properties of a good kernel fail for the Dirichlet kernel D_N. $\Diamond$

In preparation, let us continue our study of good kernels. We begin with the question 'Why should $f * K_n(\theta)$ converge to $f(\theta)$ anyway?' As n increases, the mass of $K_n(y)$ becomes concentrated near $y = 0$, and so the mass of $K_n(\theta - y)$ becomes concentrated near $y = \theta$. So the integral in $f * K_n(\theta)$ only 'sees' the part of f near $y = \theta$. See Figure 4.6.

Here is a model example.

EXAMPLE 4.23. (*A Good Kernel Made up of Characteristic Functions*) Consider the kernels

$$K_n(\theta) = n\chi_{[-\pi/n, \pi/n]}(\theta), \qquad n \in \mathbb{N}.$$

They are pictured in Figure 4.7 for several values of n.

FIGURE 4.7. Graphs of rectangular kernels K_2, K_4, and K_8.

It is straightforward to check that these functions K_n are a family of good kernels. Further, for a fixed $\theta \in \mathbb{T}$,

$$\begin{aligned}(f * K_n)(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\, n\chi_{[-\pi/n,\pi/n]}(\theta - y)\, dy \\ &= \frac{n}{2\pi}\int_{\theta-(\pi/n)}^{\theta+(\pi/n)} f(y)\, dy \\ &= \text{average value of } f \text{ on } \left[\theta - \frac{\pi}{n}, \theta + \frac{\pi}{n}\right],\end{aligned}$$

as for convolution with the characteristic function in Example 4.5. We expect this average value over a tiny interval centered at θ to converge to the value $f(\theta)$ as $n \to \infty$. This is certainly true if f is continuous at θ, by an application of the Fundamental Theorem of Calculus. In fact, if f is continuous, let $F(\theta) := \int_{-\pi}^{\theta} f(y)\, dy$. Then $F'(\theta) = f(\theta)$, and by the Fundamental Theorem of Calculus,

$$\frac{n}{2\pi}\int_{\theta-(\pi/n)}^{\theta+(\pi/n)} f(y)\, dy = \frac{n}{2\pi}\left[F\big(\theta + (\pi/n)\big) - F\big(\theta - (\pi/n)\big)\right].$$

It is not hard to see that the limit as $n \to \infty$ of the right-hand side is exactly $F'(\theta) = f(\theta)$. ♢

We are now ready to return to the approximation property enjoyed by good kernels.

Proof of Theorem 4.21. First $f : \mathbb{T} \to \mathbb{C}$ is assumed to be integrable and bounded. Let B be a bound for f on $\mathbb{T}$. Suppose f is continuous at some point $\theta \in \mathbb{T}$. Fix $\varepsilon > 0$. Choose δ such that

$$|f(\theta - y) - f(\theta)| < \varepsilon \qquad \text{whenever } |y| < \delta. \tag{4.11}$$

Now

$$\begin{aligned}(K_n * f)(\theta) - f(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta - y)\, dy - f(\theta) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta - y)\, dy - f(\theta)\,\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\, dy \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta - y)\, dy - \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y) f(\theta)\, dy \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\left[f(\theta - y) - f(\theta)\right]\, dy.\end{aligned}$$

In the second line we were able to multiply $f(\theta)$ by $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\, dy$ (another typical analysis trick), since this quantity is 1 by condition (a) on good kernels.

Yet another common technique in analysis is to estimate an integral (that is, to find an upper bound for the absolute value of the integral) by splitting the domain of integration into two regions, typically one region where some quantity is small and one where it is large, and controlling the two resulting integrals by different methods. We use this technique now, splitting the domain $\mathbb{T}$ into the regions where

$|y| \leq \delta$ and $\delta \leq |y| \leq \pi$:

$$\begin{aligned}
|(K_n * f)(\theta) - f(\theta)| &\leq \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(y) \, [f(\theta - y) - f(\theta)] \, dy \right| \\
&\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |K_n(y)| \, |f(\theta - y) - f(\theta)| \, dy \\
&= \frac{1}{2\pi} \int_{|y| \leq \delta} |K_n(y)| \, |f(\theta - y) - f(\theta)| \, dy \\
&\quad + \frac{1}{2\pi} \int_{\delta \leq |y| \leq \pi} |K_n(y)| \, |f(\theta - y) - f(\theta)| \, dy. \qquad (4.12)
\end{aligned}$$

For $y \leq \delta$ we have $|f(\theta - y) - f(\theta)| < \varepsilon$, and so

$$\begin{aligned}
\frac{1}{2\pi} \int_{|y| \leq \delta} |K_n(y)||f(\theta - y) - f(\theta)| \, dy &\leq \frac{\varepsilon}{2\pi} \int_{|y| \leq \delta} |K_n(y)| \, dy \\
&\leq \frac{\varepsilon}{2\pi} \int_{-\pi}^{\pi} |K_n(y)| \, dy \\
&\leq \frac{\varepsilon M}{2\pi}.
\end{aligned}$$

We have used condition (b) on good kernels in the last inequality.

Condition (c) on good kernels implies that for our δ,

$$\int_{\delta \leq |y| \leq \pi} |K_n(y)| \, dy \to 0$$

as $n \to \infty$. So there is some N such that for all $n \geq N$,

$$\int_{\delta \leq |y| \leq \pi} |K_n(y)| \, dy \leq \varepsilon.$$

Also, $|f(\theta - y) - f(\theta)| \leq 2B$. So we can estimate the second integral in the last line of equation (4.12), for $n \geq N$, by

$$\frac{1}{2\pi} \int_{\delta \leq |y| \leq \pi} |K_n(y)| \, |f(\theta - y) - f(\theta)| \, dy \leq \frac{2B\varepsilon}{2\pi}.$$

Hence for $n \geq N$, we have

$$|(K_n * f)(\theta) - f(\theta)| \leq (M + 2B) \, \varepsilon/(2\pi). \qquad (4.13)$$

This proves the first part of the theorem.

Assume now that f is continuous on $\mathbb{T}$, then it is uniformly continuous. This means that fixed $\varepsilon > 0$ there exists $\delta > 0$ such that (4.11) holds for all $\theta \in \mathbb{T}$. We can repeat verbatim the previous argument and conclude that the estimate (4.13) holds for all $\theta \in \mathbb{T}$, that is $K_n * f \to f$ uniformly. □

EXERCISE 4.24. Check that if F is continuously differentiable at θ and $h > 0$, then

$$\lim_{h \to 0} \frac{F(\theta + h) - F(\theta - h)}{2h} = F'(\theta).$$

◊

For an arbitrary integrable function f, the limit as n goes to infinity of the convolution with the good kernel $\tilde{K}_n$ does pick out the value of f at points of continuity θ (in other words, at all points θ where f is continuous). The L^1-version of this averaging limiting result goes by the name *Lebesgue Differentiation Theorem.*

THEOREM 4.25 (Lebesgue Differentiation Theorem). *If* $f \in L^1(\mathbb{T})$ *then*

$$\lim_{h\to 0} \frac{1}{2h}\int_{\theta-h}^{\theta+h} f(y)\,dy = f(\theta) \qquad \text{for a.e. } \theta \in \mathbb{T}.$$

Proofs of this result require a thorough understanding of measure theory. For example, see [**SS2**, Chapter 3, Section 1.2].

4.5. Fejér kernels and Cesàro means

Returning to an earlier question, can we recover the values of an integrable function f from knowledge of its Fourier coefficients? Perhaps just at points where f is continuous?

Du Bois-Reymond's example shows that even if f is continuous on the whole circle $\mathbb{T}$, we cannot hope to recover f by taking the limit as $n \to \infty$ of the traditional partial sums $S_N f$. However, Fejér discovered that we *can* always recover a *continuous* function f from its Fourier coefficients, if we use a different method of summation, which had been developed by Cesàro about ten years earlier. If f is just integrable, the method recovers f at all points where f is continuous.

Fejér's method boils down to using *Cesàro partial sums* $\sigma_N f$ corresponding to convolution with a particular family of good kernels, now called the Fejér kernels. The Cesàro partial sums, also known as *Cesàro means*, are defined by

$$\begin{aligned}
\sigma_N f(\theta) &:= \frac{S_0 f(\theta) + S_1 f(\theta) + \cdots + S_{N-1} f(\theta)}{N} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} S_n f(\theta) \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{|k|\le n} a_k e^{ik\theta},
\end{aligned}$$

where a_n is the n^{th}-Fourier coefficient of f. Thus $\sigma_N f$ is the average of the first N partial Fourier sums $S_n f$, and it is a 2π-periodic trigonometric polynomial (see also Exercise 4.15). Interchanging the order of summation, and counting the number of appearances of each Fourier summand $a_k e^{ik\theta}$ gives a representation of $\sigma_N f$ as a weighted average of the Fourier coefficients of f corresponding to frequencies $|k| \le N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. More precisely,

$$\sigma_N f(\theta) = \sum_{|k|\le N-1} \frac{N-|k|}{N} a_k e^{ik\theta}.$$

The *Fejér kernel* $F_N(\theta)$ is defined to be the average of the Dirichlet kernels,

$$F_N f(\theta) := \frac{D_0(\theta) + D_1(\theta) + \cdots + D_{N-1}(\theta)}{N}. \tag{4.14}$$

By the same calculations performed above we conclude that the Fejér kernel can be written as a weighted average of the trigonometric functions corresponding to frequencies $|k| \leq N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. More precisely,

$$F_N f(\theta) = \sum_{|n| \leq N} \left(\frac{N - |n|}{N} \right) e^{in\theta}.$$

What is not so obvious is that there is a closed formula for the Fejér kernel, namely

$$F_N f(\theta) = \frac{1}{N} \left[\frac{\sin(N\theta/2)}{\sin(\theta/2)} \right]^2 . \tag{4.15}$$

EXERCISE 4.26. Verify (4.15). You might want to use the closed formula (4.3) for the Dirichlet kernel. ♢

Compare with formula (4.3) for the Dirichlet kernel in terms of sines. Notice that, unlike the Dirichlet kernel, the Fejér kernel is non-negative. See Figure 4.8.

FIGURE 4.8. Graphs of Fejér kernels $F_N(x)$ for $N = 1$, 3, and 5. [*** Maybe lower horizontal line to x-axis.

EXERCISE 4.27. (*The Fejér Kernel is a Good Kernel*) Verify that the Fejér kernel F_N is a good kernel, and that

$$\sigma_N f = F_N * f.$$

♢

By Theorem 4.21 on good kernels, it follows that the Cesàro sums of f converge to f wherever f is continuous.

THEOREM 4.28 (Fejér, 1900). *If $f : \mathbb{T} \to \mathbb{C}$ is integrable and bounded, then*

i) *at all points of continuity of f there is pointwise convergence of the Cesàro sums, in other words if f is continuous at θ then*

$$f(\theta) = \lim_{N\to\infty} (f * F_N)(\theta) = \lim_{N\to\infty} \sigma_N f(\theta);$$

ii) *if f is continuous on $\mathbb{T}$, then $\sigma_N f \to f$ uniformly on $\mathbb{T}$.*

Thus we can indeed recover a continuous function f from knowledge of its Fourier coefficients.

Let us reiterate that the Cesàro sums give a way to approximate continuous functions on $\mathbb{T}$ uniformly by trigonometric polynomials. That is Weierstrass's celebrated Theorem 3.4, which we stated in Section 3.1 as an appetizer.

Fejér's Theorem gives a proof of the uniqueness of Fourier coefficients for continuous and periodic functions. More precisely, if $f, g \in C(\mathbb{T})$ and $\widehat{f}(n) = \widehat{g}(n)$ for all $n \in \mathbb{Z}$, then $f = g$. It suffices to verify the above statement for $g = 0$, which we did at the end of Section 3.1.

EXERCISE 4.29. The uniqueness principle for continuous functions on the circle states that

If $f \in C(\mathbb{T})$ and all its Fourier coefficients vanish, then $f = 0$.

Show that the uniqueness principle implies that if f, $g \in C(\mathbb{T})$ and $\widehat{f}(n) = \widehat{g}(n)$ for all $n \in \mathbb{Z}$, then $f = g$. ◊

A similar argument shows that if f is Riemann-integrable and all its Fourier coefficients are zero, then at all points of continuity of f, $f(\theta) = \lim_{N\to\infty} \sigma_N f(\theta) = 0$. We reproduce the argument in this setting: suppose $f : \mathbb{T} \to \mathbb{C}$ is Riemann-integrable, and all its Fourier coefficients are zero, then all the partial Fourier sums are zero, and so are the Cesàro means, $\sigma_N f = 0$. But the Cesàro means converge to f, at all points of continuity, and we are done. Lebesgue's Theorem 2.29 says that f is Riemann integrable if and only if it is continuous almost everywhere, so the conclusion can be strengthened to $f = 0$ a.e.

With the machinery of the Lebesgue integral one can prove the same result for Lebesgue-integrable functions.

THEOREM 4.30 (Uniqueness Principle). *If $f \in L^1(\mathbb{T})$ and $\widehat{f}\,(n) = 0$ for all n, then $f = 0$ a.e.*

4.6. Poisson kernels and Abel means

The Poisson kernel is another good kernel. Convolution of a function f with the Poisson kernel yields a new quantity known as the *Abel mean*, analogous to the way in which convolution with the Fejér kernel yields the Cesàro mean. Some differences are that the Poisson kernel is indexed by a continuous rather than a discrete parameter, and that each Abel mean involves all the Fourier coefficients rather than just those for $|n| \leq N$. See [**SS1**, Sections 5.3, 5.4] for a fuller discussion, including the notion of Abel summability of the Fourier series to f at points of continuity of the integrable function f, and an application to solving the heat equation on the unit disk.

DEFINITION 4.31. The *Poisson kernel* $P_r(\theta)$ is defined for $r \in [0, 1)$ by

$$P_r(\theta) := \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} = \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

◊

Notice that the Poisson kernel is indexed by all real numbers r between 0 and 1, not by the discrete positive integers N as in the Dirichlet and Fejér kernels. We are interested now in the behavior as r increases towards 1, instead of $N \to \infty$. The Poisson kernel is non-negative (for $r \in [0, 1)$). The series for $P_r(\theta)$ is absolutely convergent and uniformly convergent.

FIGURE 4.9. Graphs of Poisson kernels $P_r(x)$ for $r = 1/2$, $2/3$, and $9/10$.

EXERCISE 4.32. Verify that the two formulae in our definition of the Poisson kernel are actually equal. ◊

EXERCISE 4.33. (*The Poisson Kernel is a Good Kernel*) Verify that the Poisson kernel $P_r(\theta)$ is a good kernel. (Modify the definition of *good kernel* to take account of the change in index from $n \in \mathbb{N}$ to $r \in [0, 1)$.) ◊

DEFINITION 4.34. Let f be a function with Fourier series $\sum_{n\in\mathbb{Z}} a_n e^{in\theta}$. The r^{th}-*Abel mean* of f is defined by

$$A_r f(\theta) := \sum_{n=-\infty}^{\infty} r^{|n|} a_n e^{in\theta},$$

for $r \in [0,1)$. $\Diamond$

Thus the Abel mean is formed from f by multiplying the Fourier coefficients a_n by the corresponding factor $r^{|n|}$.

The Abel mean arises from the Poisson kernel by convolution, just as the Cesàro mean arises from the Fejér kernel by convolution.

EXERCISE 4.35. Verify that for integrable functions $f : \mathbb{T} \to \mathbb{C}$,

$$A_r f(\theta) = (P_r * f)(\theta). \tag{4.16}$$

$\Diamond$

It follows from Theorem 4.21 (modified for the continuous index $r \in [0,1)$) that if $f : \mathbb{T} \to \mathbb{C}$ is Riemann integrable, and f is continuous at $\theta \in \mathbb{T}$, then

$$\begin{aligned} f(\theta) &= \lim_{r\to 1^-} (f * P_r)(\theta) \\ &= \lim_{r\to 1^-} A_r f(\theta). \end{aligned}$$

Thus we can recover the values of f (at points of continuity of f) from the Abel means of f, just as we can from the Cesàro means. If f is continuous the convergence is uniform.

EXERCISE 4.36. To plot the kernels $\widetilde{K}_n$ and the Féjer kernels $\{F_n\}$, we used the following MATLAB script [****script has to be written]. Modify the script to plot the Poisson kernels. Hint: replace the continuous parameter r by the discrete parameter $r_n = 1 - 1/n$. $\Diamond$

4.7. Excursion into $L^p(\mathbb{T})$

We will be concerned with other modes of convergence for the convolution of a family of good kernels and a function. For example, convergence in $L^2(\mathbb{T})$, and also in $L^p(\mathbb{T})$.

THEOREM 4.37. *If $f \in C(\mathbb{T})$ and $\{K_n\}_{n\geq 1}$ is a family of good kernels, then*

$$\lim_{n\to\infty} \|f * K_n - f\|_{L^p(\mathbb{T})} = 0.$$

EXERCISE 4.38. Prove Theorem 4.37. Just notice that you can control the L^p-norm with the L^∞-norm on $\mathbb{T}$; see Exercise 2.21. Recall the relations between different modes of convergence illustrated in Figure 2.5. $\Diamond$

In particular this theorem implies the convergence in $L^p(\mathbb{T})$ of the Cesàro means $\sigma_N f$ and the Abel means $A_r f$ for continuous functions $f \in C(\mathbb{T})$. We state explicitly the mean-square convergence for continuous functions and the Cesàro means.

THEOREM 4.39. *If $f \in C(\mathbb{T})$ then*

$$\lim_{N\to\infty} \|\sigma_N f - f\|_{L^2(\mathbb{T})} = 0.$$

In other words the Cesàro means $\sigma_N f$ converge to f in the L^2-sense.

In the next chapter we are interested in proving the mean-square convergence of the partial Fourier sums of f. Since we have information about the mean-square convergence of the Cesàro means, if we had an analogue to Lemma 3.9 in $L^2(\mathbb{T})$, and we knew that $S_N f$, the partial Fourier sums, converge in L^2-sense, we could deduce indirectly, as we did in Section 3.2.1 that they have no other choice than to converge to f, the limit of the Cesàro sums. We will prove that $\|S_N f - f\|_{L^2(\mathbb{T})} \to 0$, and Theorem 4.39 will play a rôle, however we will not follow the route sketched in this paragraph. We will first understand better the geometry of $L^2(\mathbb{T})$, and of $S_N f$, in particular we will see that $S_N f$ is the "best approximation to f in the L^2-norm in the space $\mathcal{P}_N(\mathbb{T})$ of trigonometric polynomials of degree N".

Just for fun, here is the analogue of Lemma 3.9 in $L^p(\mathbb{T})$.

Lemma 4.40. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions in $L^p(\mathbb{T})$. Assume the sequence is convergent in $L^p(\mathbb{T})$ to f (since $L^p(\mathbb{T})$ is complete, the limit function f must be in $L^p(\mathbb{T})$). Then the sequence of averages*

$$F_n(\theta) = \frac{f_1(\theta) + \cdots + f_n(\theta)}{n}$$

converges in $L^p(\mathbb{T})$ and to the same limit function f.

Exercise 4.41. Prove Lemma 4.40. $\Diamond$

In Section 3.2 we proved the Riemann–Lebesgue Lemma for continuous functions; see Lemma 3.19. We can use an approximation argument to prove the lemma for integrable functions.

Lemma 4.42 (Riemann-Lebesgue Lemma for Integrable Functions). *If $f \in L^1(\mathbb{T})$, then*

$$\widehat{f}(n) \to 0 \qquad \text{as } |n| \to \infty.$$

Proof. The result holds for $f \in C(\mathbb{T})$ by Lemma 3.19. Given $f \in L^1(\mathbb{T})$, there is a sequence of continuous functions $\{f_k\}_{k\in\mathbb{N}}$ such that $\|f_k - f\|_{L^1(\mathbb{T})} \to 0$ as $k \to \infty$; see Theorem 2.70. In particular their Fourier coefficients are close: for each fixed n,

$$|\widehat{f}(n) - \widehat{f_k}(n)| \leq \frac{1}{2\pi}\|f_k - f\|_{L^1(\mathbb{T})} \to 0 \qquad \text{as } k \to \infty. \tag{4.17}$$

Now we estimate the size of the Fourier coefficients of f, knowing that for the continuous functions f_k, $\widehat{f_k}(n) \to 0$ as $|n| \to \infty$. Fix $\varepsilon > 0$. There is some $K > 0$ such that for all $k > K$

$$\|f_k - f\|_{L^1(\mathbb{T})} < \pi\varepsilon.$$

Hence, by inequality (4.17), for all $n > 0$ and all $k > K$

$$|\widehat{f}(n) - \widehat{f_k}(n)| < \varepsilon/2.$$

Fix $k > K$. There is an $N > 0$ such that for all $|n| > N$ and for the particular $k > K$,

$$|\widehat{f_k}(n)| < \varepsilon/2.$$

Finally, by the triangle inequality for real numbers, we see that for all $|n| > N$,

$$|\widehat{f}(n)| \leq |\widehat{f}(n) - \widehat{f_k}(n)| + |\widehat{f_k}(n)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since our argument is valid for all positive ε, we have shown that $\lim\limits_{|n|\to\infty} |\widehat{f}(n)| = 0$ for all $f \in L^1(\mathbb{T})$, proving the lemma. $\square$

Since $L^p(\mathbb{T}) \subset L^1(\mathbb{T})$, the same must hold for functions in $L^p(\mathbb{T})$. One could run a direct proof in complete analogy to the proof for $L^1(\mathbb{T})$, provided (4.17) holds with L^1-norm replaced by L^p-norm, which is the content of the next exercise.

EXERCISE 4.43. Prove that if $f \in L^p(\mathbb{T})$ and $f_k \in C(\mathbb{T})$ with $\|f_k - f\|_{L^p(\mathbb{T})} \to 0$ as $|n| \to \infty$, then for $p \neq 2$

$$|\widehat{f}(n) - \widehat{f_k}(n)| \leq (2\pi)^{1-1/p} \|f_k - f\|_{L^p(\mathbb{T})} \to 0 \qquad \text{as } |n| \to \infty.$$

For $p = 1$ this proves (4.17). For $p = 2$ the constant is 1 instead of $(2\pi)^{1/2}$, because of the normalization constant in the definition of the L^2-norm. Hint: Hölder's Inequality 2.22 may be helpful. $\Diamond$