# A Little Harmonic Analysis

Carl D. Offner

This is an introduction to some fundamental ideas of harmonic analysis. It assumes that the reader knows the basics of measure theory and the Lebesgue integral, and knows a little (not much more than the definitions) about Banach and Hilbert spaces. Some of this is reviewed in Section 2.1.

## Contents

# 1 The three classical domains

Classical 1-dimensional harmonic analysis takes place on three domains:

- The unit circle $C$ in the complex plane. Usually, this is represented as the real numbers $\mathbf{R} \bmod 2\pi$, and functions on $C$ are represented as periodic functions on $\mathbf{R}$ with period $2\pi$. Another notation for this domain, and the main one we use in these notes, is $\mathbf{T}$, the 1-dimensional torus.
- The real numbers $\mathbf{R}$.
- The integers $\mathbf{Z}$.

The important facts about all these domains from the standpoint of harmonic analysis are that

- They are all abelian groups under addition.
- They each have a topology that is compatible with the group structure in the sense that for each of these domains $G$, addition as a map from $G \times G \to G$ is continuous. That is, each of these domains is a topological group.
- They each have an invariant measure—that is, a measure that is invariant under translation. The measure on $\mathbf{Z}$ is just the discrete measure assigning a mass of 1 to each point, and the measure on $\mathbf{T}$ and on $\mathbf{R}$ is just any positive multiple of Lebesgue measure.

  To say these measures are invariant is just to say that for any set $E$ and any element $x$, the measure of $E + x$ is the same as the measure of $E$. Such a measure can be proved to exist on any locally compact abelian group, where it is called Haar measure, and harmonic analysis has been successfully abstracted to work in that context.

We shall use $G$ (for "group") to refer to any of these domains, when we are proving something that is true of all of them. The measure on $G$ will be denoted by $d\mu$.

When we are dealing specifically with $\mathbf{T}$, we will parametrize it by $\theta$, and Lebesgue measure on $\mathbf{T}$ will be $d\theta$. Similarly, $\mathbf{R}$ will be parametrized by $x$, $t$, or similar variables, and Lebesgue measure on $\mathbf{R}$ will be simply $dx$, or $dt$, and so on. The actual measure we use for $\mu$ will be a multiple of Lebesgue measure. To be precise, if $\lambda$ denotes Lebesgue measure on $\mathbf{T}$ or $\mathbf{R}$, let us reserve the letter $m$ to denote the positive number such that $d\mu = m\, d\lambda$. We will discuss this further in Section 6.

We noted above that the topology and the group structure of $G$ are related by the fact that the group operation is continuous as a map from $G \times G$ to $G$. We also noted that the measure and the group structure of $G$ are related by the fact that the measure is translation-invariant.

There is a third relation—between the measure and the topology of $G$—that is somewhat more subtle, but is equally important. A standard way of constructing Lebesgue measure on **R**(or **T**) is to start with the length function on half-open intervals (i.e., intervals of the form $[a, b)$) and extend it to the $\sigma$-algebra of sets generated by this family of intervals. This $\sigma$-algebra is just the family of Borel sets in **R** or **T**. It is a standard result of measure theory that because of this construction, any Borel set $E$ of finite measure can be approximated by a set $A$ that is a finite union of half-open intervals, in the sense that $\mu(E\Delta A)$ is small (i.e., "given $\epsilon > 0$, there is an $A$..."), where $E\Delta A$ denotes the symmetric difference $(E - A) \cup (A - E)$. This fact, which says that in some sense measurable sets are not too different from simple unions of intervals, is often expressed in this context by saying that Lebesgue measure is *regular*.

# 2 Function spaces

The following spaces of functions[1] on $G$ are important for us:

## 2.1 The $L^p$ spaces

The $L^p$ spaces are defined for $1 \leq p < \infty$ by

$$L^p(G) = \left\{ f : \int_G |f|^p \, d\mu < \infty \right\}$$

$L^p(G)$ is a Banach space (see below) under the norm

$$\|f\|_p = \left( \int_G |f|^p \, d\mu \right)^{1/p}$$

As a limiting case, $L^\infty(G)$ is the space of bounded measurable functions on $G$ with the norm

$$\|f\|_\infty = \operatorname{ess\,sup} |f| = \inf \{r : \mu\{x : |f(x)| > r\} = 0\}$$

If $f$ is actually continuous, this is just the usual supremum of $|f|$. The definition above takes account of the possibility that there is a set of measure 0 on which $|f|$ has large values—such a set can be ignored in computing $\|f\|_\infty$.

In all these $L^p$ spaces, a function is really only determined up to a set of measure 0, since the integral fails to distinguish functions differing on such sets. So "really", elements of $L^p$ are equivalence classes of functions, two functions being in the same equivalence class iff they are equal almost everywhere with respect to the measure $\mu$. Analysts, however, almost never think in terms of these equivalence classes—we talk of the functions directly, bearing in mind that we can't really talk of the value of the function at a point unless we know something more about the function (for instance, that it is continuous).

The Cauchy-Schwarz inequality states that if $f$ and $g$ are in $L^2$, then their product $fg$ is in $L^1$, and $\|fg\|_1 \leq \|f\|_2 \|g\|_2$. Here is the standard short proof of this:

[1] The word "function" without any further qualification refers to a measurable function with values in **C**.

First, this is clearly true when $g = 0$. (By this we mean, of course, that $g$ is 0 almost everywhere.) So we can assume that $g$ is not almost everywhere 0, and therefore that $\int_G |g|\, d\mu > 0$ and also $\int_G |g|^2\, d\mu > 0$.

Now for any $t \in \mathbf{R}$, $(|f| - t\,|g|)^2 \geq 0$, so we have

$$0 \leq \int_G (|f| - t\,|g|)^2\, d\mu = \int_G |f|^2\, d\mu - 2t \int_G |fg|\, d\mu + t^2 \int_G |g|^2\, d\mu$$

This is a quadratic function of $t$, and it attains its minimum when

$$t = \frac{\int_G |fg|\, d\mu}{\int_G |g|^2\, d\mu}$$

Therefore, substituting in this value for $t$, we have

$$0 \leq \int_G |f|^2\, d\mu - 2\frac{\left(\int_G |fg|\, d\mu\right)^2}{\int_G |g|^2\, d\mu} + \frac{\left(\int_G |fg|\, d\mu\right)^2}{\int_G |g|^2\, d\mu}$$

and therefore

$$\left(\int_G |fg|\, d\mu\right)^2 \leq \int_G |f|^2\, d\mu \int_G |g|^2\, d\mu$$

and we are done.

There is a generalization of the Cauchy-Schwarz inequality known as Hölder's inequality: for any $p$ such that $1 \leq p \leq \infty$ we define its *conjugate exponent* $q$ by

$$\frac{1}{p} + \frac{1}{q} = 1$$

with the convention that $1/\infty = 0$. Then also $1 \leq q \leq \infty$, and $p$ is reciprocally the conjugate exponent of $q$. Hölder's inequality states that if $p$ and $q$ are conjugate exponents and if $f \in L^p(G)$ and $g \in L^q(G)$ then their product $fg$ is in $L^1(G)$ and

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

The proof of this is not much more complicated than the proof of the Cauchy-Schwarz inequality, but we will not need Hölder's inequality here[2], so we omit it.

We have referred to the function $f \mapsto \|f\|_p$ as a norm. It certainly satisfies $\|af\|_p = |a|\,\|f\|_p$ for any $a \in \mathbf{C}$, and $\|f\|_p = 0 \iff f = 0$ almost everywhere. To complete the proof that $\|\cdot\|_p$ really is a norm, we have to show that the triangle inequality holds. This is easy to see for $p = 1$ and $p = \infty$.

[2]For our purposes here, the only $L^p$ spaces we need are those for which $p$ is 1, 2, or $\infty$.

For $p = 2$ it follows from the Cauchy-Schwarz inequality:

$$\begin{aligned}\|f+g\|_2^2 &= \left(\int_G |f+g|^2 \, d\mu\right) \\ &= \int_G (|f|^2 + 2\Re f\bar{g} + |g|)^2 \, d\mu \\ &\le \|f\|_2^2 + 2\|f\|_2 \|g\|_2 + \|g\|_2^2 \\ &= (\|f\|_2 + \|g\|_2)^2\end{aligned}$$

For any other $p$ between 1 and $\infty$, the triangle inequality (which becomes known in that context as Minkowski's inequality) follows in a somewhat more complicated fashion from Hölder's inequality. We omit that derivation.

By far the nicest behaved $L^p$ space is $L^2$, because it is a Hilbert space: we define the inner product of two functions $f$ and $g$ to be

$$(f, g) = \int_G f\bar{g} \, d\mu$$

so $\|f\|_2^2 = (f, f)$. No other $L^p$ space is a Hilbert space.

The Cauchy-Schwarz and Hölder inequalities are just the tip of the iceberg, which is the following fundamental fact:

**2.1 Theorem** *If $1 \le p < \infty$, and if $q$ is the conjugate exponent of $p$, then there is an isometric isomorphism*[3] *between $L^q(G)$ and the dual space of $L^p(G)$.*

*Under this correspondence, each linear functional $T \in L^p(G)^*$ corresponds to a function $g \in L^q(G)$ in such a way that for each $f \in L^p(G)$,*

$$Tf = (f, g) = \int_G f\bar{g} \, d\mu$$

*and we have $\|T\| = \|g\|_q$, where the operator norm $\|T\|$ is defined by $\|T\| = \sup\left\{|Tf| : \|f\|_p = 1\right\}$.*

For $p = 2$ this is immediate: $L^2$ is a Hilbert space, and any Hilbert space is naturally isomorphic to its dual space. But for any other exponent $p$, the result is deep—it is equivalent to the Radon-Nikodym theorem in measure theory. We won't prove this theorem here.

This theorem in turn can be used to prove the following result, which is sometimes called the "Landau resonance theorem":

**2.2 Theorem** *If*

- $1 \le p < \infty$, *and $q$ is the conjugate exponent of $p$,*
- *$f$ is a function on $G$*

[3] The correspondence is usually set up, as it is here, so as to be conjugate-linear, so is often referred to as a conjugate-isomorphism.

*and if* $fg \in L^1(G)$ *for every* $g \in L^q(G)$, *then* $f \in L^p(G)$. *Further, in such a case,*

$$\|f\|_p = \sup\left\{\frac{\|fg\|_1}{\|g\|_q} : g \in L^q(G)\right\}$$

We won't prove this theorem either, although we will use it below. The proof is fairly straightforward.

Since $\mathbf{T}$ has finite measure,

$$L^\infty(\mathbf{T}) \subset L^2(\mathbf{T}) \subset L^1(\mathbf{T})$$

Each of these inclusions is easy to see: if $f \in L^\infty(\mathbf{T})$, then certainly any power of $|f|$ has a finite integral on $\mathbf{T}$, and so in particular $f \in L^2(\mathbf{T})$. And if $f \in L^2(\mathbf{T})$, then by the Cauchy-Schwarz inequality[4],

$$\begin{aligned}
\|f\|_1 &= \int_{\mathbf{T}} |f(\theta)|\, d\mu(\theta) \\
&\le \left(\int_{\mathbf{T}} 1^2\, d\mu(\theta)\right)^{1/2} \left(\int_{\mathbf{T}} |f(\theta)|^2\, d\mu(\theta)\right)^{1/2} \\
&= \sqrt{2\pi m}\, \|f\|_2 \\
&\le \infty
\end{aligned}$$

so $f \in L^1(\mathbf{T})$. An equivalent way to write the above computation is this:

$$\|f\|_1 = \|f \cdot 1\|_1 \le \|f\|_2\, \|1\|_2$$

and $\|1\|_2 < \infty$ because $\mathbf{T}$ has finite measure.

There is a more general statement of this: if $1 \le r < s \le \infty$, then $L^r(\mathbf{T}) \supseteq L^s(\mathbf{T})$. This result follows similarly from Hölder's inequality, but the result for $L^\infty$, $L^2$, and $L^1$ is all we need here.

The opposite inclusions are true for $\mathbf{Z}$, since it is discrete with the counting measure: if $1 \le r < s \le \infty$, then $L^r(\mathbf{Z}) \subset L^s(\mathbf{Z})$. In particular,

$$L^1(\mathbf{Z}) \subset L^2(\mathbf{Z}) \subset L^\infty(\mathbf{Z})$$

We won't need these results here.

On $\mathbf{R}$, there is no relation between any of the $L^p$ spaces—they all contain (for instance) all continuous functions of compact support, so they have a non-trivial intersection. But no one of them includes any other. This makes harmonic analysis on $\mathbf{R}$ a more delicate matter than on $\mathbf{T}$ or $\mathbf{Z}$.

There is one final result we need. We mentioned above that Lebesgue measure on $\mathbf{R}$ and $\mathbf{T}$ is regular—in particular, that any measurable set can be approximated arbitrarily closely in measure by a finite union of half-open intervals.

It then follows (pretty immediately from the definition of the integral) that the set of finite linear combinations of characteristic functions of half-open intervals is dense in $L^p$ provided $1 \le p < \infty$.

[4] Remember that we are using the letter $m$ to denote the positive number such that $d\mu = m\, d\lambda$, where $\lambda$ denotes Lebesgue measure.

## 2.2 Spaces of continuous functions

We define

$$C_b(G) = \text{the set of bounded continuous functions on } G$$
$$C_u(G) = \text{the set of bounded uniformly continuous functions on } G$$

These spaces are Banach spaces under the uniform norm $\|f\|_\infty$.

Since $\mathbf{T}$ is compact, any continuous function on $\mathbf{T}$ is uniformly continuous, and so

$$C_b(\mathbf{T}) = C_u(\mathbf{T})$$

On $\mathbf{Z}$, all functions are continuous, and so

$$C_b(\mathbf{Z}) = C_u(\mathbf{Z}) = L^\infty(\mathbf{Z})$$

## 2.3 The continuity of translation

Let us consider the operation of translation on each of our function spaces. That is, for each $t \in G$, define an operator $T_t$ taking the function $f$ to the function $T_t f$, defined by $(T_t f)(x) = f(x - t)$. That is, $T_t$ translates $f$ to the right by $t$.

Clearly, $T_t$ maps each of the function spaces we have defined above into itself. Let us change our notation slightly: for any function $f$, define $f_t = T_t f$. For each $f$, the function $t \mapsto f_t$ is a map from $G$ into some function space.

Now we can ask if this map is continuous as a function of $t$. That is, for each $f$, is the function $t \mapsto f_t$ a continuous map from $G$ to the function space in question?

Well, first of all, this map is trivially continuous when $G = \mathbf{Z}$, so we can forget about that case. Now when $G = \mathbf{R}$ and the function space is $C_b(\mathbf{R})$, the map is not in general continuous. For instance, consider the function $f(x) = \sin x^2$. It is not true that $\|f_t - f\|_\infty \to 0$ as $t \to 0$.

However, if $f \in C_u(G)$, then it *is* true that $\|f_t - f\|_\infty \to 0$ as $t \to 0$—this is just the definition of uniform continuity. So translation is continuous on $C_u(G)$.

As far as the $L^p$ spaces go, translation is not continuous on $L^\infty(G)$, for the same reason as for $C_b$. In fact, here we even have a simpler counterexample—just take $f$ to be the characteristic function of $(0, \infty)$.

It is a remarkable fact, however, that translation *is* continuous on $L^p$, provided that $1 \le p < \infty$.

**2.3 Theorem** *Translation is continuous in $L^p$. Precisely, if $1 \le p < \infty$, then for each $s \in G$, $\|f_t - f_s\|_p \to 0$ as $t \to s$. Furthermore the convergence is uniform in $s$.*

Proof. Since $\|f_t - f_s\|_p = \|f_{t-s} - f\|_p$, the entire theorem will be proved if we can show that $\|f_t - f\|_p \to 0$ as $t \to 0$.

We have already noted that the set of finite linear combinations of characteristic functions of half-open intervals is dense in $L^p$.

So let $\phi$ be a finite linear combination of characteristic functions of half-open intervals approximating $f$ in $L^p$. For each half-open interval $[a, b)$, we have[5]

$$\left\| T_t(\chi_{[a,b)}) - \chi_{[a,b)} \right\|_p \leq (2m\,|t|)^{1/p}$$

and this tends to 0 as $t \to 0$. Hence the same is true for $\phi$. But then

$$\begin{aligned} \|f_t - f\|_p &\leq \|f_t - \phi_t\|_p + \|\phi_t - \phi\|_p + \|\phi - f\|_p \\ &= 2\,\|f - \phi\|_p + \|\phi_t - \phi\|_p \end{aligned}$$

which proves the theorem. □

# 3 Convolutions

Let $G$ denote any of the three classical domains, and let $\mu$ denote its invariant measure.

**3.1 Lemma** *If $f$ is a measurable function from $G$ to $\mathbf{C}$, then the function on $G \times G$ to $\mathbf{C}$ defined by $(x, y) \to f(x + y)$ is measurable.*

Proof. This function is the composition $f \circ a$ of $f$ with the function $a(x, y) = x + y$. $f$ is measurable by assumption. $a$ is measurable because it is continuous (since $G$ is a topological group). Hence $f \circ a$, as the composition of two measurable functions, is also measurable. □

**3.2 Theorem** *If $f$ and $g$ are in $L^1(G)$, then the integral $\int f(x - t)g(t)\, d\mu(t)$ exists for almost all $x$, and defines a function $f * g$, called the* convolution *of $f$ and $g$. The convolution has the following properties:*

1. *$f * g$ is in $L^1(G)$, and in fact $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$.*
2. *Convolution is commutative.*
3. *Convolution is associative.*
4. *Convolution is linear in each variable.*

Proof. We know that $f(x - t)g(t)$ is a measurable function on $G \times G$. Replacing $f$ and $g$ by their absolute values $|f|$ and $|g|$, Fubini's theorem together with translation invariance of the integral on $G$ then shows that $f(x - t)g(t)$ is in $L^1(G \times G)$: for we have

$$\begin{aligned} \int\int |f(x - t)g(t)|\, d\mu(t)\, d\mu(x) &= \int\int |f(x - t)||g(t)|\, d\mu(x)\, d\mu(t) \\ &= \int \|f\|_1\, g(t)\, d\mu(t) \\ &= \|f\|_1 \|g\|_1 \end{aligned}$$

This in turn proves item 1. The other items are straightforward. □

[5] Again recall that $m$ denotes the positive number such that $d\mu = m\, d\lambda$, where $\lambda$ denotes Lebesgue measure.

Convolution also acts nicely on other function spaces. Here are a few examples of this, which we will need later:

**3.3 Theorem** *If $p$ and $q$ are conjugate exponents with $1 \le p \le \infty$, and if $f \in L^p$ and $g \in L^q$, then $f * g \in C_u$ (that is, it is a bounded uniformly continuous function), and*

$$\|f * g\|_\infty \le \|f\|_p \|g\|_q$$

PROOF. By Hölder's inequality, $|f * g(x)| \le \int |f(x-t)g(t)|\, d\mu(t) \le \|f\|_p \|g\|_q$,[6] and so $f * g \in L^\infty$ and $\|f * g\|_\infty \le \|f\|_p \|g\|_q$.

Now at least one of $p$ and $q$ is not $\infty$. Without loss of generality, let us assume that $p < \infty$. Then

$$|f * g(x-s) - f * g(x)| \le \int |f_s(x-t) - f(x-t)||g(t)|\, d\mu(t) \le \|f_s - f\|_p \|g\|_q \to 0$$

as $s \to 0$, uniformly in $x$, the limit by continuity of translation (Theorem 2.3). Thus, $f * g$ is uniformly continuous. □

If $f : G \to \mathbf{C}$ we define $\tilde{f}(x) = \overline{f}(-x)$. The map $f \to \tilde{f}$ is an isometry on $L^p$ for all $p \ge 1$. Further, considering convolution by $f$ as an operator on $L^p$(this is just formal until after the proof of the next theorem), convolution by $\tilde{f}$ corresponds to the formal adjoint of that operator. That is, if the pairing between $L^p$ and $L^q$ is given as usual by $(f, g) = \int_G f\bar{g}\, d\mu$, we have (formally at least)

$$(f * g, h) = (g, \tilde{f} * h)$$

This is just to say that formally, we have

$$\begin{aligned} \int \left( \int f(x-t)g(t)\, d\mu(t) \right) \overline{h(x)}\, d\mu(x) &= \int g(t) \left( \int f(x-t)\overline{h(x)}\, d\mu(x) \right) d\mu(t) \\ &= \int g(t) \overline{\left( \int \tilde{f}(t-x)h(x)\, d\mu(x) \right)}\, d\mu(t) \end{aligned}$$

We will show that this is actually true when used reasonably. The notation itself is useful in the proof of the next theorem.

**3.4 Theorem** *If $f \in L^1$ and $g \in L^p$ $(1 \le p \le \infty)$, then $f * g \in L^p$ and $\|f * g\|_p \le \|f\|_1 \|g\|_p$.*

PROOF. The case $p = 1$ has already been taken care of by Theorem 3.2, and the case $p = \infty$ is handled by Theorem 3.3. Thus, we may assume that $1 < p < \infty$. Let $q$ be the conjugate exponent for $p$, so $1/p + 1/q = 1$ and $1 < q < \infty$. If $h$ is any function in $L^q$, then

$$\int |f * g(x)h(x)|\, d\mu(x) = \int |f(x)\tilde{g} * h(x)|\, d\mu(x) \le \|f\|_1 \|\tilde{g} * h\|_\infty \le \|f\|_1 \|g\|_p \|h\|_q$$

by the previous theorem. Thus by the Landau resonance theorem (Theorem 2.2), $f * g$ must be in $L^p$ with norm at most $\|f\|_1 \|g\|_p$. □

[6] Note that this inequality holds everywhere, not just almost everywhere. Of course, since $f * g$ will be shown to be continuous, this has to be the case.

Theorem 3.3 can be used to give an elegant though non-elementary proof of the fact that if $E$ is a measurable subset of $\mathbf{R}$ whose Lebesgue measure $\mu(E)$ is greater than 0, then the difference set $E - E$ contains an open interval.

**3.5 Theorem** *If $E$ is a Lebesgue measurable subset of $\mathbf{R}$ such that $\mu(E) > 0$, then the difference set $E - E$ contains an open set.*

Proof. We may assume that $0 < \mu(E) < \infty$; otherwise replace $E$ by a subset with finite measure; proving the theorem for that subset proves it for $E$.

Now $\chi_E$ and $\tilde{\chi}_E$ are in $L^2$; in fact, $\|\chi_E\|_2 = \|\tilde{\chi}_E\|_2 = \mu(E)^{1/2}$. We know by the theorem that $\chi_E * \tilde{\chi}_E$ is continuous, and $\chi_E * \tilde{\chi}_E(0) = \int \chi_E^2(x)\, d\mu(x) = \mu(E) > 0$. Hence there is an open set containing 0 on which $\chi_E * \tilde{\chi}_E > 0$. But any $x$ for which $\chi_E * \tilde{\chi}_E(x) > 0$ is an element of the difference set of $E$. $\square$

# 4 Approximate identities

Theorem 3.2 shows that convolution by an $L^1$ function defines a bounded linear map from $L^1$ to $L^1$. In fact, convolution makes $L^1$ into what is called a *Banach algebra*—a Banach space having a bilinear[7] and associative multiplication (denoted by $*$, say) such that the norm satisfies

$$\|f * g\| \leq \|f\| \, \|g\|$$

Here the multiplication is just convolution, and it is commutative, although in Banach algebras in general it does not need to be.

There also does not need to be a multiplicative identity, although sometimes there is. For instance, on $\mathbf{Z}$, the delta function

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

is a convolution identity: for any $f \in L^1(\mathbf{Z})$, we have

$$f * \delta(n) = \sum_{i=-\infty}^{\infty} f(n-i)\delta(i) = f(n)$$

On the other hand, there is no convolution identity in $L^1(\mathbf{R})$ or $L^1(\mathbf{T})$—such a function would be the "physicists' delta function", which *can* be represented as a measure of mass 1 concentrated at 0, but cannot be represented as a function in $L^1$.

Why do we care about convolution identities? Well, it turns out that convolutions are powerful ways of filtering out information about a function. For example, suppose we have a function $f$ in $L^1$. This function may be quite irregular, containing all sorts of kinks and spikes. Suppose we want to create a smoothed version of this function—a function that is more well-behaved and that is a good approximation to this function. One standard thing to do is to take a moving average: let us

[7]A consequence of bilinearity is that multiplication distributes over addition in both operands.

for the moment simplify things by using Lebesgue measure (i.e., we let $m = 1$, and $d\mu(t) = dt$). Let $s > 0$ and define the function $S_s(f)$ ($S$ for "smooth") by

$$S_s(f)(x) = \frac{1}{2s} \int_{x-s}^{x+s} f(t)\, dt$$

Note that this definition makes sense either on $\mathbf{R}$ or on $\mathbf{T}$. Then intuitively,

- $S_s(f)$ is close to $f$.
- $S_s(f)$ gets closer to $f$ as $s \downarrow 0$.
- $S_s(f)$ is more well-behaved than $f$.

Further, if we define the function

$$\phi_s(t) = \frac{1}{2s} \chi_{[-s,s]}(t)$$

then $S_s(f)$ is just $f * \phi_s$.

Now note that the function $\phi_s$ approaches the "delta function" as $s \downarrow 0$. That is, it looks more and more like a spike at 0, and the area under the function is always 1. (See Figure 1.)

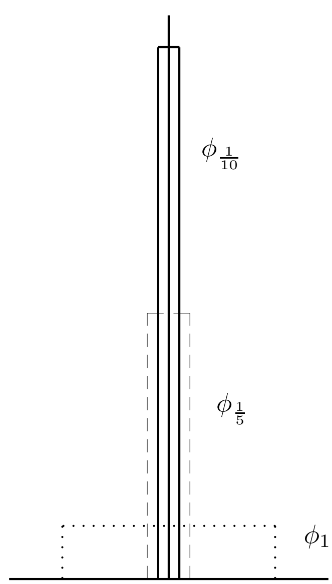


Figure 1: The moving average kernel $\phi_s$.

Because of this behavior, the family $\{\phi_s\}$ is called an *approximate identity.* In general, we make the following definition:

**Definition** *An* approximate identity *on $G$ ($G$ here being either* $\mathbf{R}$ *or* $\mathbf{T}$*, and we're back to using the measure $\mu$) is a family $\{\phi_s\}$ of real-valued functions such that*

1. $\phi_s \geq 0$.
2. $\int \phi_s(x)\, d\mu(x) = 1$.
3. *For each* $\delta > 0$, $\lim_{s\to 0} \int_{|x|\leq\delta} \phi_s(x)\, d\mu(x) = 1$.

   *Equivalently, for each* $\delta > 0$, $\lim_{s\to 0} \int_{|x|>\delta} \phi_s(x)\, d\mu(x) = 0$.

Sometimes the notation is such that the limit occurs as $s \to \infty$ or as $s \uparrow 1$. These are just trivial notational changes. And sometimes the family is actually a sequence $\{\phi_n\}$, and the limit occurs as $n \to \infty$.

**4.1 Lemma** *If $\{\phi_s\}$ is an approximate identity on $G$ and if $f \in L^\infty$ is such that* $\lim_{t\to 0} f(t) = 0$*, then* $\lim_{s\to 0} \int f(t)\phi_s(t)\, d\mu(t) \to 0$.

Proof. For each $\epsilon > 0$ there is a $\delta > 0$ such that $|f(t)| < \epsilon$ for $|t| < \delta$. Then

$$\begin{aligned}\left|\int f(t)\phi_s(t)\, d\mu(t)\right| &\leq \int |f(t)\phi_s(t)|\, d\mu(t) \\ &= \int_{|t|<\delta} + \int_{|t|>\delta} \\ &\leq \epsilon + \|f\|_\infty \int_{|t|>\delta} \phi_s(t)\, d\mu(t)\end{aligned}$$

So first make $\epsilon$ small and then make $s$ small. □

**4.2 Theorem** *If $\{\phi_s\}$ is an approximate identity on $G$, then*

1. *If $f \in C_b$ (i.e. $f$ is a bounded continuous function) then $f * \phi_s \in C_u$ (it's bounded and uniformly continuous) and $f * \phi_s \to f$ pointwise.*
2. *If $f \in C_u$ (i.e. $f$ is bounded and uniformly continuous) then $f * \phi_s \in C_u$ and $f * \phi_s \to f$ uniformly.*
3. *If $1 \leq p < \infty$ and $f \in L^p$ then $f * \phi_s \in L^p$ and $f * \phi_s \to f$ in $L^p$.*

Proof. 1. $f * \phi_s \in C_u$ by Theorem 3.3. For a fixed $x \in G$,

$$\left|\int f(x-t)\phi_s(t)\, d\mu(t) - f(x)\right| \leq \int |f(x-t) - f(x)|\phi_s(t)\, d\mu(t)$$

The lemma then yields convergence at $x$.

2. If $f$ is uniformly continuous, the family of functions $g_x(t) = f(x-t) - f(x)$ converges to 0 as $t \to 0$ uniformly in $x$, so the convergence in part 1 is uniform.

3. If $f \in L^1$, we have (using Fubini's Theorem)

$$\begin{aligned}\|f * \phi_s - f\|_1 &= \int \left| \int \big(f(x-t) - f(x)\big)\phi_s(t)\,d\mu(t) \right| \,d\mu(x) \\ &\le \int \left( \int \big|f(x-t) - f(x)\big|\phi_s(t)\,d\mu(t) \right) d\mu(x) \\ &= \int \|f_t - f\|_1 \,\phi_s(t)\,d\mu(t)\end{aligned}$$

which converges to 0 as $s \to 0$, by the lemma and the fact that translation is continuous in $L^1$.

When $1 < p < \infty$, we use the Landau resonance theorem (Theorem 2.2). (Actually, we can use the same argument when $p = 1$; it's just that the proof just given for $p = 1$ is simpler.) Let $q$ be the conjugate exponent for $p$. We know by the theorem that

$$\|f * \phi_s - f\|_p = \sup_{\|g\|_q \le 1} \int \big(f * \phi_s(x) - f(x)\big)\bar{g}(x)\,d\mu(x)$$

Now

$$\begin{aligned}\int \big(f * \phi_s(x) - f(x)\big)\bar{g}(x)\,d\mu(x) &= \int \left( \int f(x-t)\phi_s(t)\,d\mu(t) - f(x) \right) \bar{g}(x)\,d\mu(x) \\ &= \int \left( \int \big(f(x-t) - f(x)\big)\phi_s(t)\,d\mu(t) \right) \bar{g}(x)\,d\mu(x) \\ &= \int \left( \int \big(f(x-t) - f(x)\big)\bar{g}(x)\,d\mu(x) \right) \phi_s(t)\,d\mu(t) \\ &\le \|g\|_q \int \|f_t - f\|_p \,\phi_s(t)\,d\mu(t)\end{aligned}$$

Hence

$$\|f * \phi_s - f\|_p \le \int \|f_t - f\|_p \,\phi_s(t)\,d\mu(t) \to 0$$

as $s \to 0$, by the lemma and the fact that translation is continuous in $L^p$. □

As a corollary, we can prove that continuous functions are dense in $L^p$ for $1 \le p < \infty$:

**4.3 Theorem** *If $f \in L^p$ with $1 \le p < \infty$, there is a sequence $\{f_n\}$ of continuous functions with compact support converging to $f$ in $L^p$. That is, the set of continuous functions with compact support (which are all in $L^p$) is dense in $L^p$.*

Proof. First approximate $f$ by a function in $g$ in $L^p$ with compact support. Then take an approximate identity $\{\phi_n\}$, all of whose members have compact support. Each function $g * \phi_n$ is then a continuous function of compact support, and $g * \phi_n \to g$ in $L^p$. □

# 5 Why we like convolutions

Here is an extended quotation from the little out-of-print book *Smoothing and Approximation of Functions* by Harold S. Shapiro (1969). The book itself is a tiny gem, and this quotation explains the way in which analysts have learned to use convolutions. By $K_\lambda$, Shapiro is referring to a family of functions of the form $\lambda K(\lambda x)$ that constitutes an approximate identity as $\lambda \to \infty$. $\hat{K}$ refers to the Fourier transform of $K$, which we will get to later. The quotation is from pages 8-10 of the book:

> Our main purpose, as we have said, is to illustrate the use of convolutions in approximation problems. In particular, what is perhaps the most powerful and general known method for generating approximations to $f$ may be summarized thus: "convolve $f$ with a peaking kernel". The reasons for the dazzling versatility of this method may be summed up as follows:
>
> a) Convolution is a *smoothness-increasing operation.* That is, if $g$ is integrable and of norm one, $f * g$ is at least as smooth as $f$ by just about any conceivable test (modulus of continuity, moduli of smoothness of higher order, number of derivatives, total variation, etc.). This isn't too surprising perhaps if we think of convolution as a (generalized) moving average.
>
> b) Various special structural properties of a function $f$ (e.g., having a given period, or being a trigonometric polynomial of degree not exceeding $n$) are likewise inherited by $f * g$.
>
> At bottom a) and b) are the same: very roughly, they say that properties based on the *translation group* are *hereditary* under convolution. And because of the commutativity of convolution, their presence in *either factor* ensures their presence in the convolution product. Thus, convolution is like a marriage in which (unlike real life) the "best" properties of each parent are inherited by the offspring (i.e. differentiability, periodicity, etc. are "dominant genes"). Thus, suppose a lowly bounded measurable function $f$ is convolved with an integrable function $g$ which happens to have 100 derivatives. The resulting function has again 100 derivatives, but moreover resembles $f$ if $g$ is chosen to be a peaking kernel (for instance, if $g$ is $K_\lambda$ for large $\lambda$ and suitable $K$, then $f * K_\lambda$ tends almost everywhere to $f$). If $f$ moreover has period $2\pi$, so have all the approximating functions. And if, in addition, $\hat{K} = 0$ for $|x| \geq 1$ (so that $K_\lambda$ has a Fourier transform vanishing for $|x| \geq \lambda$), this property too is inherited by $f * K_\lambda$ which must, therefore, be a trigonometric polynomial of degree less than $\lambda$.
>
> Moreover, convolutions have other properties which make them technically very nice to work with. For instance, if $f$ and $g$ are differentiable we can for the derivative of $f * g$ take our choice of the expressions $f' * g$ and $f * g'$. For higher derivatives there is still greater freedom. Also the close tie-in with Fourier transforms puts powerful techniques from harmonic analysis at our disposal. Furthermore, the asymptotic behavior of convolutions is often easy to estimate.
>
> In addition, it turns out (although this is far from obvious *a priori*) that under suitable restrictions the operations of passing from a function to its derivative or its (suitably normalized) primitive may be realized as convolutions with suitable kernels. These facts enhance the importance of convolutions, and play an essential role in the theory which follows.
>
> Finally, although this plays only a minor role in the present book, the notion of convolution admits of far reaching and fruitful generalization: not only can functions be convolved with one another, but more general entities (functionals) can meaningfully

be convolved with functions, and under suitable circumstances, with one another. Of course, such "convolutions" cannot any longer be interpreted as "smoothing" operations.

# 6 Fourier analysis on T

## 6.1 The invariant measure on T

Now we need to be specific about the measure we are going to use on **T**. This is really just a matter of convenience. Unfortunately, no matter which measure we pick, there are always factors of $2\pi$ or $\sqrt{2\pi}$ that clutter up our equations. Physicists have the ability to use units in which the speed of light is 1, and this simplifies things for them. Unfortunately, we can't let $2\pi$ be 1, and we have to make an arbitrary choice. The choice we use is the conventional one, and it is probably used because it fits in most naturally with applications to complex function theory, where we identify **T** with the unit circle in the complex plane.

So here is the choice we make: On **T**, we will use

$$d\mu = \frac{1}{2\pi}\, d\lambda$$

or, more simply, $d\mu(\theta) = d\theta/2\pi$. This makes the total measure of **T** equal to 1. Note that this makes

$$\|f\|_p = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|^p \; d\theta\right)^{1/p}$$

The inner product in $L^2$ is given by

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\, d\theta$$

Convolution becomes

$$(f * g)(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta - \phi)g(\phi)\, d\phi$$

And finally, the conditions for $\phi_s$ to be an approximate identity become:

1. $\phi_s \geq 0$.

2. $\int \phi_s(\theta)\, d\theta = 2\pi$.

3. For each $\delta > 0$, $\lim_{s\to 0} \int_{|x|\leq\delta} \phi_s(\theta)\, d\theta = 2\pi$.

   Or equivalently, $\lim_{s\to 0} \int_{\delta\leq|x|\leq\pi} \phi_s(\theta)\, d\theta = 0$.

## 6.2 Fourier coefficients

We already noted that $L^2(\mathbf{T})$ is a Hilbert space. In this space, the functions

$$e_n(\theta) = e^{in\theta}$$

are orthonormal. This follows immediately from the observation that

$$\int_{-\pi}^{\pi} e^{in\theta}\, d\theta = \int_0^{2\pi} e^{in\theta}\, d\theta = \begin{cases} 2\pi & n = 0 \\ 0 & n \neq 0 \end{cases}$$

(Note that any other choice of $d\mu$ would mean that we would have to define $e_n$ to be a non-trivial multiple of $e^{in\theta}$ in order for $e_n$ to have norm 1.)

Thus, any function $f \in L^2(\mathbf{T})$ has a set of *Fourier coefficients*

$$\hat{f}(n) = (f, e_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta}\, d\theta$$

and we know the following, just by elementary Hilbert space arguments:

- The sum $\sum_{-\infty}^{\infty} \hat{f}(n) e_n$ converges in $L^2(\mathbf{T})$ to the orthogonal projection of $f$ on the subspace of $L^2(\mathbf{T})$ generated by the $\{e_n\}$. (This subspace is the closure in $L^2(\mathbf{T})$ of the set of finite linear combinations of elements of $\{e_n\}$.)
- Consequently, we have for any $f \in L^2(\mathbf{T})$, the following inequality, known as *Bessel's inequality*:

$$\|f\|_2 \geq \sum_{-\infty}^{\infty} \left|\hat{f}(n)\right|^2$$

If furthermore the set of functions $\{e_n\}$ is an orthonormal basis of $L^2(\mathbf{T})$—that is, if finite linear combinations of these functions are dense in $L^2(\mathbf{T})$—then we can recover $f$ in terms of its Fourier expansion:

$$f(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{in\theta}$$

where, again, the convergence of the series is understood in the sense of convergence of functions in $L^2(\mathbf{T})$. That is,

$$\lim_{n\to\infty} \left\| f - \sum_{k=-n}^{n} \hat{f}(k) e_k \right\|_2 = 0$$

In this case, Bessel's inequality will become an actual equality. So we have a question:

- Is this true? That is, is the orthonormal set $\{e_n\}$ a basis for $L^2(\mathbf{T})$?

Now actually, the Fourier transform (that is, the map from $f$ to $\hat{f}$) is naturally defined on $L^1(\mathbf{T})$, which properly includes $L^2(\mathbf{T})$. (This is just because the integral that defines $\hat{f}(n)$ makes sense for any $f \in L^1(\mathbf{T})$.) For such functions, we can still form a Fourier series, formally at least, and we can ask

- Is the mapping from $f$ to $\hat{f}$ 1-1? That is, is $f$ determined by its Fourier coefficients?

- If so, how can we retrieve $f$ from its Fourier coefficients? A more classical way of asking this question is to ask under what conditions and in what sense the Fourier series of $f$ converges to $f$.

## 6.3 Fejér's theorem

After several false starts, some answers to these questions were found, first by Dirichlet in 1829. These answers were stated in terms of pointwise convergence and assumed some regularity of the function $f$. For instance, if $f$ is at least twice differentiable, then its Fourier series can be shown to converge to $f$ pointwise. A small industry developed of finding successively weaker conditions on $f$ that would ensure that its Fourier series converged to it. Dirichlet himself believed that it would be possible to prove that the Fourier series of any continuous function converges at each point to that function.

However, understanding pointwise convergence turns out to be inherently difficult, and in addition, pointwise convergence itself does not really address some important questions. The first problem that was found—and it was a big one—was that there are in fact functions that are continuous at all points, but whose Fourier series does not converge at one or more points. This was shown in an explicit construction by du Bois Reymond in 1876.

The situation then was this: on the one hand, Fourier series had proved to be immensely powerful and useful in mathematics and mathematical physics—mathematicians and physicists both felt that in some sense the Fourier series of a function gives an accurate representation of that function; on the other hand, convergence of a Fourier series could not be assured for continuous functions.

The breakthrough came in 1900, and was due to Fejér. But before seeing what he did, let us first see where the problem lay:

If we denote the Fourier coefficients of $f$ by $a_n = \hat{f}(n)$, and if we denote the $n^{\text{th}}$ partial sum of the Fourier series for $f$ by $s_n$:

$$s_n(\theta) = \sum_{k=-n}^{n} a_k e^{ik\theta}$$

then we have

$$\begin{aligned} s_n(\theta) &= \frac{1}{2\pi}\sum_{k=-n}^{n} e^{ik\theta}\int_{-\pi}^{\pi} f(\phi)e^{-ik\phi}\,d\phi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)\left(\sum_{k=-n}^{n} e^{ik(\theta-\phi)}\right)d\phi \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)D_n(\theta-\phi)\,d\phi \\ &= (f*D_n)(\theta) \end{aligned}$$

where by summing the geometric series, we find that

$$D_n(\theta) = \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin\left(n+\frac{1}{2}\right)\theta}{\sin\frac{1}{2}\theta}$$

$D_n(\theta)$ is called the *Dirichlet kernel*. Figure 2 shows what this family of functions looks like.

We can see that, although $D_n(\theta)$ appears to become more concentrated at 0 as $n$ becomes larger, it has some fatal flaws, any one of which prevents it from being an approximate identity:

- It takes on both positive and negative values.
- It does not really tend to 0 outside a neighborhood of 0—it just oscillates faster and faster there.
- Although $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(\theta)\,d\theta$ does equal 1, the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(\theta)|\,d\theta$ is not even bounded—it tends to $\infty$ roughly as $\log n$.

Because of the second of these three points, any proof of convergence has to rely on cancellation of positive and negative values of the integrand in the convolution $f*D_n$. This makes reasoning about pointwise convergence quite difficult. Dirichlet's argument was a real masterpiece of hard analysis, and led ultimately to the modern definition of bounded variation.

What Fejér did was this: Instead of considering the partial sums directly, he considered averages of them. He considered the *Cesàro means* defined by

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

By a similar computation as before, we have

$$\begin{aligned} \sigma_n(\theta) &= \frac{1}{n}\sum_{k=0}^{n-1}(f*D_k)(\theta) \\ &= \frac{1}{n}\left(f*\sum_{k=0}^{n-1} D_k\right)(\theta) \\ &= (f*K_n)(\theta) \end{aligned}$$

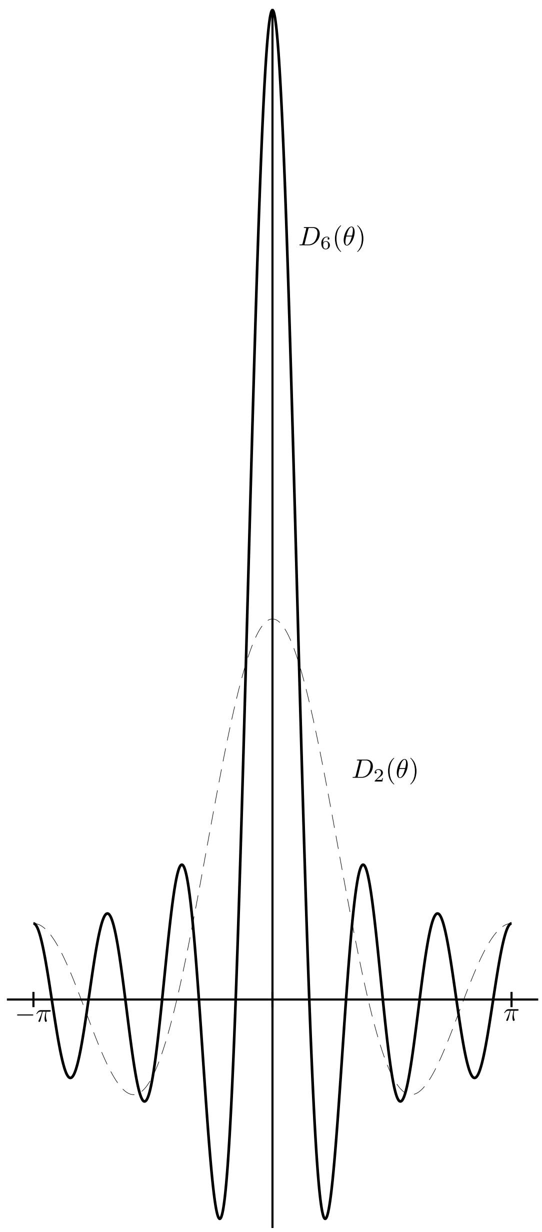


Figure 2: The Dirichlet kernel

where

$$\begin{aligned} K_n(\theta) &= \frac{1}{n}\sum_{k=0}^{n-1} D_k \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\frac{\sin\left(k+\frac{1}{2}\right)\theta}{\sin\frac{1}{2}\theta} \\ &= \frac{1}{n}\frac{1}{\sin\frac{1}{2}\theta}\Im\sum_{k=0}^{n-1} e^{i\left(k+\frac{1}{2}\right)\theta} \\ &= \frac{1}{n}\left(\frac{\sin\frac{1}{2}n\theta}{\sin\frac{1}{2}\theta}\right)^2 \end{aligned}$$

where the last line is found by summing the geometric series in the line above. $K_n$ is called the *Fejér kernel*. Figure 3 shows three members of this family.

The Fejér kernel looks like a real approximate identity, and in fact it is:

- $K_n \geq 0$. This is obvious from the formula for $K_n$.
- $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(\theta)\,d\theta = 1$. This is a straightforward computation. The easiest way to see it is to note that it simply states that the $n^{\text{th}}$ Cesàro mean of the Fourier series for the constant function 1 is 1.
- For each $\delta > 0$, $\lim_{n\to\infty}\int_{\delta<|\theta|<\pi} K_n(\theta)\,d\theta = 0$. This is because the numerator in $K_n$ is bounded above by 1, and for $\theta$ restricted to be outside of a neighborhood of 0 (mod $2\pi$), the sin in the denominator is bounded away from 0, so the $n$ in the denominator controls the growth.

  To be precise: Given any $0 < \delta < \pi$, if $\delta < \theta < \pi - \delta$, we have $\left(\sin\frac{1}{2}\theta\right)^2 \geq \left(\sin\frac{1}{2}\delta\right)^2$ so

  $$|K_n(\theta)| \leq \frac{1}{n\left(\sin\frac{1}{2}\delta\right)^2}$$

  on that interval, and so for each $\delta > 0$, $\int_{\delta<|\theta|<\pi} K_n(\theta)\,d\theta \to 0$ as $n \to \infty$

This answers our questions at the beginning of this section: If $f \in L^2(\mathbf{T})$, then $f * K_n \to f$ in $L^2$, and therefore the set $\{e_n\}$ forms a basis for $L^2$. Therefore in fact the partial sums of the Fourier series for $f$ also converge to $f$ in $L^2$. Further, again since the set $\{e_n\}$ is an orthonormal basis of $L^2(\mathbf{T})$, Bessel's inequality is replaced by *Parseval's formula*:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\theta)|^2\,d\theta = \|f\|_2^2 = \sum_{n=-\infty}^{\infty}\left|\hat{f}(n)\right|^2$$

We also know that the converse is true: if $\{a_n : -\infty < n < \infty\}$ is any square summable sequence then $\sum a_n e_n$ converges[8] in $L^2(\mathbf{T})$ to a function $f$ whose Fourier coefficients are $\hat{f}(n) = a_n$. Thus,

[8] This uses (as do all the preceding mentions of convergence in $L^2$) the fact that $L^2$ is complete—as are all the $L^p$ spaces. This remarkable fact was known historically as the Riesz-Fischer theorem. It was proved first in this context.

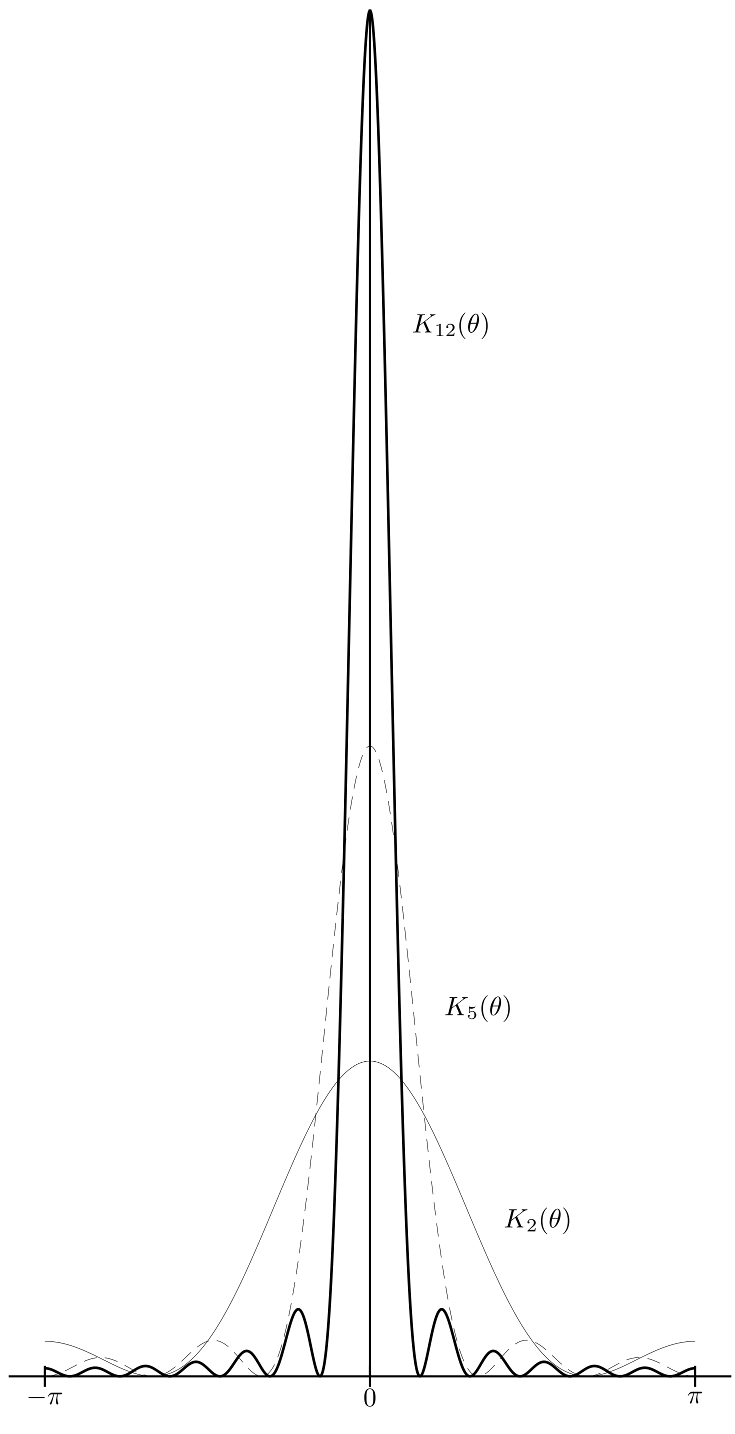


Figure 3: The Fejér kernel

thinking of the set of Fourier coefficients as a function on $\mathbf{Z}$, we see that the Fourier transform $f \to \hat{f}$ is a unitary isomorphism (i.e., linear, bijective and norm-preserving) from $L^2(\mathbf{T})$ to $L^2(\mathbf{Z})$. It is for this reason that the $L^2$ theory of the Fourier transform is so well-behaved and understandable.

In addition, if $f \in L^1(\mathbf{T})$, the Cesàro means of the Fourier series of $f$ converge to $f$ in $L^1$. Therefore, $f$ is determined by its Fourier series and the map $f \to \hat{f}$ is 1-1. Now by the formula for Fourier coefficients, we have

$$\left|\hat{f}(k)\right| = \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-ik\theta}\,d\theta\right| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|\,d\theta = \|f\|_1$$

So the Fourier transform $f \to \hat{f}$ maps $L^1(\mathbf{T})$ into $L^\infty(\mathbf{Z})$, and we just saw that the map is 1-1. But the map is not onto, nor is it norm-preserving.

We also see that if $f \in C(\mathbf{T})$, then the Cesàro means of the Fourier series of $f$ converge uniformly to $f$. (This was actually the original setting for Fejér's theorem.)

# 7 Convolutions and the Fourier transform

As we have seen, the Fourier transform $f \to \hat{f}$ takes functions on $\mathbf{T}$ to functions on $\mathbf{Z}$. Scientists and engineers often think of $\mathbf{T}$ as representing space or time, and $\mathbf{Z}$ as the "frequency domain", in the sense that each Fourier coefficient $\hat{f}(n)$ is the amplitude of a particular frequency in the Fourier series for $f$. This correspondence between functions on these two domains is worth looking at closely. We have already seen that the correspondence, when taken as a map on $L^2(\mathbf{T})$, is a unitary isomorphism with $L^2(\mathbf{Z})$.

To go further, there is an important formula that relates convolutions and Fourier transforms:

**7.1 Theorem** *If $f$ and $g$ are both in $L^1(\mathbf{T})$, then*

$$\widehat{f * g}(n) = \hat{f}(n)\hat{g}(n)$$

Proof.

$$\begin{aligned}
\widehat{f * g}(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f * g(\theta)e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta-\phi)g(\phi)\,d\phi\right)e^{-in\theta}\,d\theta \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta-\phi)e^{-in(\theta-\phi)}\,d\theta\right)g(\phi)e^{-in\phi}\,d\phi \\
&= \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)e^{-in\psi}\,d\psi\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\phi)e^{-in\phi}\,d\phi\right) \\
&= \hat{f}(n)\hat{g}(n)
\end{aligned}$$

□

We already noted that $L^1(\mathbf{T})$ is an algebra with convolution as the multiplication, and we know that the Fourier transform maps $L^1(\mathbf{T})$ into $L^\infty(\mathbf{Z})$. Now $L^\infty(\mathbf{Z})$ is also an algebra, under pointwise multiplication. This theorem shows that the Fourier transform is actually an algebra homomorphism from $L^1(\mathbf{T})$ into $L^\infty(\mathbf{Z})$.

Let us see how this formula can be used:

## 7.1 Pointwise convergence

Pointwise convergence is determined, as we have seen, by the properties of the Dirichlet kernel

$$D_n(\theta) = \frac{\sin\left(n+\frac{1}{2}\right)\theta}{\sin\frac{1}{2}\theta}$$

Since

$$D_n(\theta) = \sum_{k=-n}^{n} e^{ik\theta}$$

we know that

$$\widehat{D_n}(i) = \begin{cases} 1 & -n \le i \le n \\ 0 & \text{otherwise} \end{cases}$$

That is, $\widehat{D_n}$ is the characteristic function of the interval $[-n, n]$ on $\mathbf{Z}$. Of course, this fits in exactly with Theorem 7.1:

$$\widehat{s_n}(i) = \widehat{f * D_n}(i) = \hat{f}(i)\widehat{D_n}(i) = \begin{cases} \hat{f}(i) & n \le i \le n \\ 0 & \text{otherwise} \end{cases}$$

which is just how $s_n$ is defined: it is the function whose non-zero Fourier coefficients are just a subset (the subset with indices $-n$ to $n$) of those of $f$.

## 7.2 Cesàro summation

Cesàro summation (usually called "(C,1) summation" or "(C,1) summability") is determined by the properties of the Fejér kernel

$$K_n(\theta) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(\theta)$$

We therefore have

$$\widehat{K_n}(i) = \frac{1}{n}\sum_{k=0}^{n-1} \widehat{D_k}(i) = \begin{cases} 1 - \dfrac{i}{n} & -n \le i \le n \\ 0 & \text{otherwise} \end{cases}$$

That is, $\widehat{K_n}$ is a "tent function" on $\mathbf{Z}$, having its vertex at 0 (where its value is 1), and decreasing linearly in both directions until it reaches 0 at $i = \pm n$.

Thus, we have for the $n^{\text{th}}$ Cesàro mean of the Fourier series for $f$,

$$\widehat{\sigma_n}(i) = \widehat{f * K_n}(i) = \hat{f}(i)\widehat{K_n}(i) = \begin{cases} \left(1 - \frac{i}{n}\right)\hat{f}(i) & -n \leq i \leq n \\ 0 & \text{otherwise} \end{cases}$$

So the Cesàro means take account of the Fourier coefficients between $-(n-1)$ and $n-1$, but give less weight to the ones with higher frequency. This can be thought of as one reason why (C,1) summation is more well-behaved than plain pointwise convergence.

Note that as $K_n$ becomes more concentrated at 0 (i.e., as $n$ becomes large), its Fourier transform $\widehat{K_n}$ becomes more spread out. This phenomenon is quite general, and there are many ways of making it precise. The intuition behind it is that for a function to be concentrated near a point, it must have a large derivative near that point, which in turn means that higher-order harmonics need to be present in its Fourier expansion. This general property of the Fourier transform is known as the "uncertainty principle". (One of the ways of expressing it precisely is equivalent to the uncertainty principle of quantum mechanics.)

Another way to think of the uncertainty principle as applied to this case is that as a function $f$ becomes more and more like the physicists' delta function $\delta$, the sequence of its Fourier coefficients becomes more and more like the sequence of Fourier coefficients of $\delta$. (This can actually be interpreted to make sense, by regarding $\delta$ as a measure.) That sequence is identically 1, because formally, we have

$$\hat{\delta}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \delta(\theta)e^{-in\theta}\,d\theta = 1$$

This fits in with Theorem 7.1: as $f$ approaches the identity in the convolution algebra $L^1$, $\hat{f}$ approaches the identity in the algebra (under pointwise multiplication) $L^\infty$.

## 7.3 Abel summation

One might try to come up with other methods of summing Fourier series, by looking at other ways of weighting the Fourier coefficients. For instance, we could try having the weights decrease exponentially: Suppose we had a function $P_r(\theta)$ such that for $0 < r < 1$,

$$\widehat{P_r}(k) = r^{|k|}$$

Then

$$\widehat{f * P_r}(k) = \hat{f}(k)\widehat{P_r}(k) = r^{|k|}\hat{f}(k)$$

and we might hope that $P_r$ is an approximate identity as $r \uparrow 1$ and that this form of summation is actually useful.

This is all true. The form of summation is known as *Abel summation*, and the kernel $P_r(\theta)$ is the

*Poisson kernel.* We can see what $P_r(\theta)$ is by summing its Fourier series explicitly:

$$\begin{aligned} P_r(\theta) &= \sum_{k=-\infty}^{\infty} r^{|k|} e^{ik\theta} \\ &= \sum_{k=-\infty}^{-1} r^{|k|} e^{ik\theta} + \sum_{k=0}^{\infty} r^{|k|} e^{ik\theta} \\ &= \sum_{k=1}^{\infty} r^k e^{-ik\theta} + \sum_{k=0}^{\infty} r^k e^{ik\theta} \\ &= \frac{re^{-i\theta}}{1-re^{-i\theta}} + \frac{1}{1-re^{i\theta}} \\ &= \frac{1-r^2}{1-2r\cos\theta + r^2} \end{aligned}$$

For any value of $r$, the convergence is absolute and uniform in $\theta$. Thus $\frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta) e^{-in\theta}\, d\theta$ can be integrated term-by-term, and this shows that $P_r$ in fact has the Fourier coefficients we started with.

We still have to show that $P_r(\theta)$ is an approximate identity:

- $P_r(\theta) \geq 0$. This is immediate.
- $\int_{-\pi}^{\pi} P_r(\theta) = 2\pi$. This just says that $\widehat{P_r}(0) = 1$.
- For each $\delta > 0$, $\lim_{r\uparrow 1} \int_{\delta<|\theta|<\pi} P_r(\theta)\, d\theta = 0$. The proof of this is more or less the same as that for $K_n$: if $\delta < |\theta| < \pi$, we have $\cos\theta < \cos\delta < 1$. Therefore,

$$P_r(\theta) \leq \frac{1-r^2}{1-2r\cos\delta + r^2}$$

  As $r \uparrow 1$, the numerator of the fraction on the right approaches 0, and the denominator approaches $2(1-\cos\delta) > 0$, so $P_r(\theta) \to 0$ as $r \uparrow 1$.

so $P_r(\theta)$ is in fact an approximate identity. The Poisson kernel is important because it solves the Dirichlet problem for the unit disc in $\mathbf{C}$: say $f$ is a continuous function on $\mathbf{T}$, and define $F$ on the unit disc $C$ by wrapping $f$ around it:

$$F(e^{i\theta}) = f(\theta)$$

then (with $z = re^{i\theta}$) $P_r * f(\theta)$ is a harmonic function in the unit disc $|z| < 1$ such that it approaches $F$ radially (in fact, non-tangentially) at every point of $C$.

Figure 4 shows what this kernel looks like. Figure 5 shows how the Fourier transforms of the Fejér and Poisson kernels spread out and approach the constant sequence 1 as the kernels become more concentrated at 0.

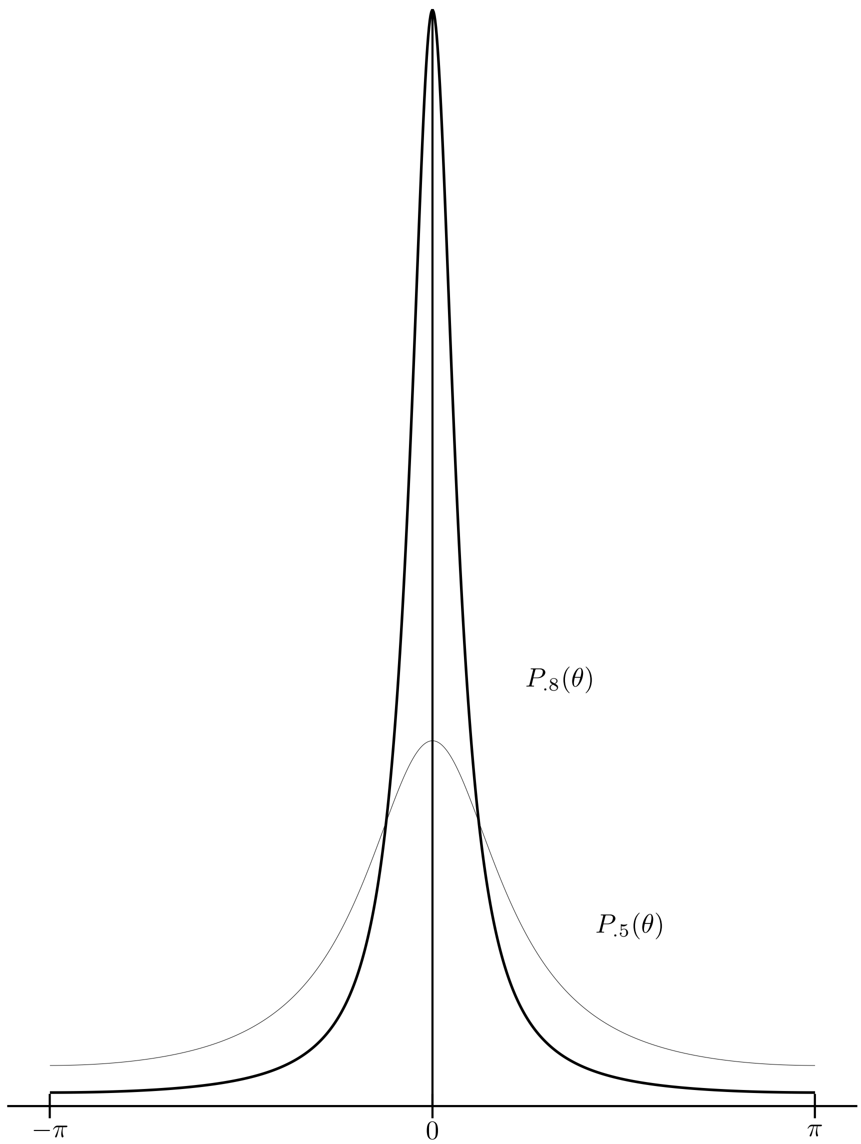


Figure 4: The Poisson kernel

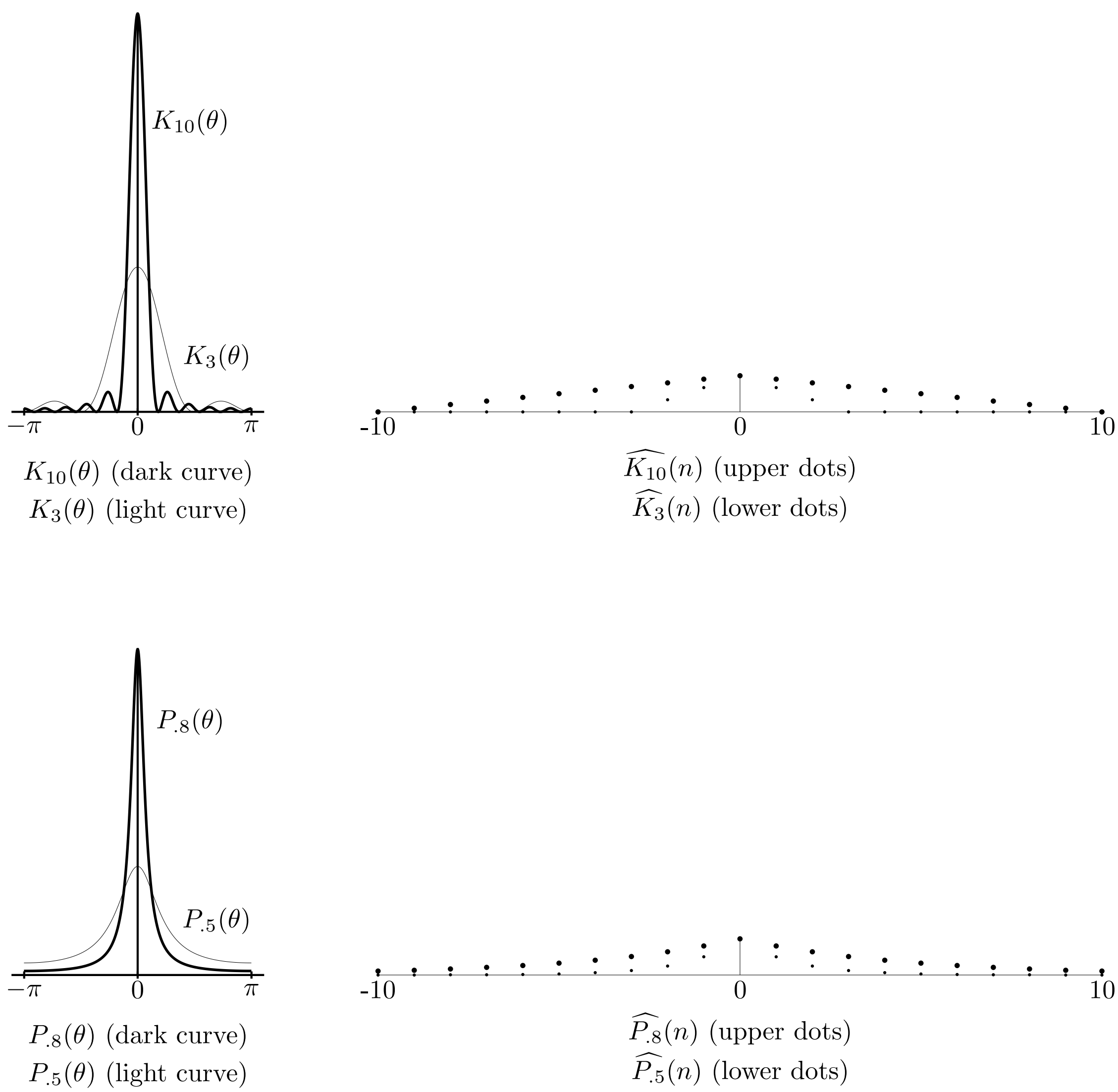


Figure 5: The Fejér and Poisson kernels and their Fourier transforms. As the kernels become more concentrated at 0, the sequence of their Fourier coefficients spreads out and approaches the constant sequence 1. This is an example of the uncertainty principle in harmonic analysis.

# 8 A historical sketch

Here is an outline of the early history of harmonic analysis, much of it from secondary sources. It is limited in scope. I have restricted myself to providing a historical context for the mathematical content of this exposition. In particular, I have completely ignored the work of Riemann and Cantor.

## 8.1 The eighteenth century

**1747** D'Alembert derived the partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

that describes the displacement $u(t, x)$ of a violin string as a function of time $t$ and distance $x$ along the string. He then showed that a general solution to this equation has the form

$$u(x, t) = f(t + x) + g(t - x)$$

If the string is normalized so that it has length 1, then from the requirements that $u(0, t) = u(1, t) = 0$ (i.e., that the displacement always be 0 at the endpoints), we get first (looking at $u(0, t)$) that $g \equiv -f$, so

$$u(x, t) = f(t + x) - f(t - x)$$

Then (looking at $u(1, t)$) we see that $f$ must be periodic of period 2:

$$f(t + 2) = f(t)$$

D'Alembert then noted that there were many functions of this form, including trigonometric functions. Implicitly, however, the functions he considered all had the form of a single analytic expression.

**1748** Euler, after seeing d'Alembert's paper, published equivalent results, with one significant difference: he was willing to allow as a function $f$ any continuous curve that could be drawn by hand. In particular, it might be given by different expressions over different parts of the interval, and its graph might include corners. Euler may have come to this definition by considering that the initial displacement of a vibrating string could be pretty arbitrary.

Euler felt, quite rightly, that his notion of function was considerably more general than that of d'Alembert.

These results focused attention on the problem of characterizing periodic functions. Everyone agreed that trigonometric functions were bona fide functions, and were periodic. And everyone agreed that finite linear combinations of them (what we now call trigonometric polynomials) were also.

In 1749, Euler considered starting with the function $f(t) = \sin n\pi t$ (where $n$ is a positive integer). Then

$$f(t + x) - f(t - x) = 2 \sin n\pi x \cos n\pi t$$

and Euler pointed out that any linear combination of functions of this form is a solution of d'Alembert's equation.

**1753** Daniel Bernoulli proposed that (assuming the string has length 1) the initial shape of the string can be expanded in a sine series containing *infinitely* many terms:

$$f(x) = \sum_{n=1}^{\infty} a_n \sin n\pi x$$

which gives

$$u(t, x) = 2 \sum_{n=1}^{\infty} a_n \sin n\pi x \cos n\pi t$$

That is, Bernoulli suggested that such a series provided a *general* solution of d'Alembert's equation. The numbers $n$ were explicitly understood to represent the frequencies of the simple modes of vibration of the string.

Euler immediately criticized this result, since it seemed to imply that any function could be represented as a trigonometric series, and Euler felt this was clearly absurd.

**1759** Lagrange performed a somewhat similar analysis, arriving at his result by replacing the string by a discrete set of masses tied together by a massless string, solving the linear equations thus generated, and proceeding to the limit as the number of masses tends to infinity. However, he never actually came up with a trigonometric series representation, and in fact, he felt strongly that such a representation was impossible. He believed throughout his life that any expression such as a trigonometric series could not possibly give rise to a graph that vanished on a subinterval without vanishing everywhere.

**The 18$^{\text{th}}$ century concept of function.** An extensive argument developed between d'Alembert, Euler, Bernoulli, and Lagrange concerning the meaning and validity of these results. The controversy, in addition to being quite bitter, was confused and confusing, because the notion of function in the eighteenth century was quite different from what it is today.

The word "function" seems to have been used first around the time of Leibnitz. In the eighteenth century, it was used to refer to algebraic expressions, and also to functions defined by means of an infinite sequence of algebraic operations, such as power series and some infinite products. There was no clear definition of function, however—and implicit functions (e.g., functions defined as solutions of algebraic equations) were also recognized. All such functions were in some sense "rigid"—generally speaking, they were analytic functions[9] and were determined by their values on a small interval. Mathematicians were used to operating with these functions in a formal way. It is fair to say that the primary meaning of "function" was similar to what we now mean by "expression".

Nevertheless, it was apparent that other kinds of functional dependence might appear, in nature at least. So by the eighteenth century, the term "continuous function" was used to denote the kind of function I have just described. The term "discontinuous function" or "mechanical function" was used to denote a function that was more general in some sense. For instance, it might be patched together from functions that were "continuous" on neighboring intervals. In Euler's usage, it was even more general than this—perhaps what we might now call "piecewise smooth". But in all cases, these functions were continuous in the modern sense—one could "draw" them.

[9] This of course is a notion that did not exist at that time.

The eighteenth century was not a time in which mathematicians were accustomed to defining their terms, and so in fact, these usages were not at all consistent. In addition, these mathematicians produced results that blatantly contradicted their own fiercely held beliefs, with no apparent recognition of this problem. And the disputes between these scholars were heated and at times truly unpleasant.

Still, the disagreement between Euler and d'Alembert had some real content[10]. In essence, d'Alembert was saying that the only functions that could be effectively dealt with mathematically were functions that had derivatives. There is something to this point. The wave equation, after all, is a differential equation. From this equation, d'Alembert had derived a functional equation, and Euler pointed out that a function did not have to be smooth at all points to satisfy this equation. On the other hand, it does not seem right that such a function should be allowed as a solution to a differential equation. Euler had really come upon a striking phenomenon: there are some differential equations that in a very natural sense admit non-differentiable solutions. Furthermore, these solutions are physically meaningful. But this phenomenon did not really begin to be understood until the time of Riemann, and Euler was far ahead of his time in this area.

In this argument, one can see the germ of the distinction between analytic function theory and the theory of "functions of a real variable". Mathematicians in the eighteenth century were quite accustomed to representing functions as power series, and power series *are* determined locally. It was the need to consider functions arising from physical problems, and the behavior of trigonometric series—which are not determined locally—that led to the controversy and ultimately to our modern notion of function.

It seems to me, following the history into the next century, that the function concept developed in parallel with ideas of continuity and discontinuity—roughly speaking, as more kinds of discontinuous functions were needed or discovered, the notion of function was further refined.

All this, however, did not stop Euler, d'Alembert, and Bernoulli from finding cases in which trigonometric series did actually represent specific functions such as $x$, $x^2$, and so on, on a finite interval. (Amazingly from our vantage point, no one noticed the problem this caused with their notion of function.) This amounted to finding the Fourier coefficients of particular functions. These investigations were at first *ad hoc*, since the general formula had not been discovered. But in 1757, Clairaut succeeded in finding the general formula for the coefficients in a cosine series. He derived it by a process of interpolation, and claimed that it would work for any function.

This culminated in Euler's remarkable discovery of 1777:

**1777** Euler showed that the formula

$$a_n = 2\int_0^1 f(x)\sin n\pi x\,dx$$

for computing the coefficients in a sine series could be derived by integrating term-by-term. In today's language, he discovered the orthogonality of the trigonometric functions. His paper was not published until 1793, ten years after his death.

[10] I am following here the excellent discussions in Truesdell (1960) and Luzin (1998).

## 8.2 The nineteenth century

**1807** Fourier, unaware of Euler's results, rediscovered the formula for what are now known as "Fourier coefficients". In his presentation to the French Academy on the theory of heat[11], he

> ... laid down the proposition that an arbitrary function given graphically by means of a curve, which may be broken by (ordinary) discontinuities, is capable of representation by means of a single trigonometrical series. This theorem is said to have been received by Lagrange with astonishment and incredulity (Hobson, 1927 (Volume I, 3d edition); 1926 (Volume II, 2nd edition), Volume 2, p. 480).

... as well as hostility. Lagrange, who died in 1813, prevented Fourier's papers from being published. In consequence, Fourier's results were not widely disseminated until the publication of his *Théorie analytique de la chaleur*, in 1822.

The functions that Fourier was willing to consider not only had corners or cusps in their graphs—they also had jump discontinuities. Such functions did not make physical sense for a vibrating string, but were reasonable initial conditions for problems involving heat conduction. It was Fourier's contention that such functions could be represented as trigonometric series that was so surprising and controversial.

Fourier gave some indications of how he thought that his trigonometric series could be proved to converge. One way involved replacing the trigonometric functions by their power series. Hardy has an interesting discussion of this in his book *Divergent Series*. Fourier's other suggestion was to write an expression for the sum of the first $n$ terms of the series. In this way, he came up explicitly with the "Dirichlet kernel". He did not pursue this method, however; it was left to Dirichlet to take up where Fourier had left off.

**1823** Poisson tried to prove Fourier's theorem. He claimed, in our notation, that $P_r * f(\theta) \to f(\theta)$ if $f$ is continuous, and hence that for continuous functions

$$\lim_{r\uparrow 1} \sum_{n=-\infty}^{\infty} r^{|n|} \hat{f}(n) e^{in\theta} = f(\theta)$$

Poisson evidently believed that this proved the theorem—that one could pass to the limit by simply setting $r = 1$ without any further justification.

Actually, Poisson didn't even quite prove this. His argument that $P_r * f(\theta) \to f(\theta)$ was badly botched by today's standards; in fact it's hard to make much sense of it. In 1872 Schwarz fixed up this part of the proof.

Cauchy didn't think much of Poisson's proof. He himself gave two proofs of the theorem, but his proofs were also flawed.

**1826** Abel (1826) proved what is now called "Abel's theorem": if the series $\sum_{k=1}^{\infty} a_n$ converges (i.e., if the sequence of partial sums converges), say to the limit $S$, then the power series $\sum_{k=1}^{\infty} a_n x^n$ converges for all $|x| < 1$ and $\lim_{x\uparrow 1} \sum_{k=1}^{\infty} a_n x^n = S$. (This would now be called the consistency theorem for Abel summability—that is, if a series is convergent in the ordinary sense, then it is Abel summable to the same value.) Abel may have been at least partially

[11] An annotated copy of this paper is in Grattan-Guinness (1972).

motivated by Poisson's attempted proof of Fourier's theorem—at any rate, his paper does refer to expressions of the form

$$\sum_{n=1}^{\infty}(a_n \sin\theta + b_n \cos\theta)r^n$$

One of the significant aspects of Abel's paper was that he had to take careful account of the possible non-absolute convergence of the series $\sum a_n$—in fact, if the series is absolutely convergent, a much simpler proof would suffice.

In this paper, Abel also pointed out that Cauchy's "proof" that the sum of a series of continuous functions is necessarily continuous cannot be correct; he gives a Fourier series that converges to a discontinuous function as a counterexample.

**Did Abel prove "Abel's theorem"?** In 1863, Liouville published a short note (Dirichlet, 1863) in which he stated that he found Abel's proof difficult to understand, and that Dirichlet had helped him by showing him another proof, which he reproduced. (This was after Dirichlet's death.) I think it is possible that Liouville was being overly modest and wanted an excuse for publishing Dirichlet's quite elegant argument.

But the matter didn't die, it seems. Grattan-Guinness (1970) refers to Liouville's note, and goes to some pains to show that Abel's proof is incorrect. There is in fact a little sloppiness in Abel's proof. The issue is this (I am abstracting things just a little, and changing the notation): Abel points out, following Cauchy, that a series $\sum a_i$ converges if and only if the family of finite sums

$$\sum_{i=m}^{m+n} a_i$$

approaches 0 with $m$. (This is now called the "Cauchy criterion".) In using this in the proof of his Theorem IV (which is what is now called "Abel's theorem"), he should thus write

$$\left|\sum_{i=m}^{m+n} a_i\right| < \epsilon$$

However, Abel writes this without the absolute value signs[12]. It's not at all hard to see that this is easily fixed, and that is what all subsequent authors have done in giving Abel's proof. Grattan-Guiness, however, thinks that one would have to introduce absolute values like this:

$$\sum_{i=m}^{m+n} |a_i| < \epsilon$$

which of course is wrong—this would be correct only if the series were absolutely convergent to begin with. Grattan-Guinness then concludes that the proof is unsalvageable.

Bottazzini (1986), similarly, refers to Liouville's note. Without reproducing Abel's proof, he then quotes the start of Dirichlet's proof, but leaves out the two or three lines that drive the

[12]He would really have had to bound it explicitly above and below, because the term "absolute value", and the associated notation, did not exist at the time. It was introduced much later, by Weierstrass.

point home. While he does not actually say that Abel's proof is wrong, he does say (correctly) that the subsequent Theorem V in Abel's paper is incomplete (because it implicitly assumed uniform convergence)—except that he doesn't make it clear that he is talking about Theorem V rather than Theorem IV. He then quotes Hardy as saying that Abel's proof contained within it the germ of the idea of uniform convergence, and says that Hardy was therefore mistaken.

The problem with this is that Hardy was referring to the proof of Theorem IV, not Theorem V. Now it is true that Abel's theorem—Theorem IV—is today usually proved with some mention of uniform convergence. Abel did not have that notion, but his proof is correct nonetheless, and that is what Hardy meant. There is no doubt about Hardy's meaning—Hardy was an expert in series and summability, and had taken a particular interest in this proof. So while Bottazzini doesn't actually say that Abel's proof is wrong, it would be hard to read him and not come away with that impression.

So both these writers are confused, and Abel really did give a correct proof.

**1829** Dirichlet gave the first rigorous proof of Fourier's theorem. Riemann (1868) says that this paper was the first in which the fact that Fourier series are non-absolutely convergent was noticed and dealt with correctly.

Riemann states this in the historical introduction to his paper on Fourier analysis, which was his *Habilitationsschrift* (probationary essay) at Göttingen in 1854. It remained unpublished until 1868, when Dedekind discovered it after Riemann's death. Riemann actually says more: He says that Dirichlet was the first person to discover the phenomenon of conditional convergence. This, however, cannot be true. Cauchy was evidently aware of the difference between absolutely and conditionally convergent series, and we already noted that Abel took careful account of this distinction in 1826. It is true, however, that Dirichlet begins his paper by pointing out that one of Cauchy's attempted proofs of convergence for Fourier series fails on just this point[13]. Riemann based his history in this paper on a long conversation with Dirichlet, so this probably represents Dirichlet's memory of the essential difficulty in the proof.

Dirichlet's proof worked for functions that were "piecewise monotonic", and guaranteed convergence at points of continuity of such functions. Dirichlet believed the proof could be extended to prove convergence for *any* continuous function.

Dirichlet's proof was extensively analyzed, and his sufficient condition was relaxed or modified in several different directions by a number of mathematicians, notably Lipschitz (1864; this is where the "Lipschitz condition" originated), Dini (1872 and 1880), and Jordan (1881), who gave the modern definition of bounded variation, and showed that it was the natural condition for Dirichlet's original argument. Lebesgue (1905) gives an overview of these different conditions and gives a still more general one that subsumes them all.

The assumption throughout much of this period was that Fourier series really do converge pointwise for continuous functions, and it was just a matter of finding the proof.

**Dirichlet's concept of function.** Dirichlet's paper also marked a significant advance in the direction of the modern notion of function: In this paper, he gave his famous example of the characteristic function of the rationals, or, as he actually put it, a function which takes one value on the rationals and a different value on the irrationals. In a reworking of this paper in 1837, he gave a definition of a continuous function that is quite close to the modern one:

[13] Cauchy's other proof, which Dirichlet was not aware of, was based on the Cauchy integral theorem and so really only applies to analytic functions.

> Let $a$ and $b$ be two fixed numbers, and let $x$ be a variable which is to take on all values between $a$ and $b$. Now if to each $x$ there corresponds a single finite $y$ in such a way that, while $x$ runs continuously through the interval from $a$ to $b$, $y = f(x)$ also varies gradually, then $y$ is called a continuous function of $x$ for this interval. It is not at all necessary here that $y$ depend on $x$ according to the same [formal] law in the whole interval; one does not have to think at all of a dependence expressible by mathematical operations. Represented geometrically, i.e., with $x$ and $y$ thought of as abscissa and ordinate, a continuous function appears as a connected curve, of which only one point corresponds to each value of $x$ between $a$ and $b$. This definition does not prescribe a common law for the individual parts, or as drawn without obeying any rule at all. It follows that such a function can be considered as completely determined for such an interval only if it given either graphically over the whole [extent of that] interval, or if it satisfies mathematical laws valid on its individual parts. As long as one has decided on the values of a function only for part of the interval, its continuation to the rest of the interval can be made entirely at will. (Dirichlet, 1837, ; translation from ; Birkhoff, 1973, p. 13)

There is a heated controversy among historians as to just how close Dirichlet got to the modern definition of function. Lakatos thinks he was still quite far away from the modern conception, and Bottazzini agrees. Their reasons are the following:

- At a jump discontinuity, Dirichlet says that a function has “two values”, which he denotes by $f(x+0)$ and $f(x-0)$.
- Dirichlet restricts his definition (the one given above) to continuous functions.

My own opinion is that Lakatos and Bottazzini are wrong, for the following reasons:

First, it does not seem to me that Dirichlet would have been at all surprised to see a definition in which $f$ had a unique value at a jump discontinuity and in which $f(x+0)$ and $f(x-0)$ were just notations for the right- and left-hand limits of $f$ at $x$. This is what those notations came to mean in the late 1800’s, and now they are usually written as $f(x+)$ and $f(x-)$. I don’t think Dirichlet was really confused on this point at all. In fact, I think his statement was just a way of saying that one could pick two reasonable values for $f$ at $x$. (Lakatos and Bottazzini would no doubt object to this, and insist that what I have written here is a back reading.)

In any case, compare this statement of Dirichlet’s to Fourier’s notion of what happened at a jump discontinuity: Fourier understood there to be a vertical line in the graph of the function at such a point. That is, Fourier thought of a function as a graph that could be drawn, in the eighteenth century tradition. So Dirichlet’s notion of what happens at a jump discontinuity is not at all foreign to us and is quite different from that of Fourier.

But of even more significance, I think, is the question of why Dirichlet restricted his definition to continuous functions:

One reason that Dirichlet did this is that his proof—based on the formulas for the Fourier coefficients—relies on integration. The only definition of the definite integral available at that point was Cauchy’s, which was stated in terms of continuous functions. (Even Cauchy’s proof of the existence of the definite integral was incomplete, however, since he did not have the concept of uniform continuity.) Dirichlet had succeeded in weakening this in his paper to accommodate a function that had a finite number of jump discontinuities, by taking the sum of the integrals over the subintervals on which the function was continuous. He believed that this line of reasoning could be extended to handle many more functions, and in his 1829 paper he suggested that it could handle functions that have an infinite number of discontinuities, so long as (in today's language) the set of these discontinuities is nowhere dense. In a later letter to Gauss, he suggested a condition that looks like what we would now call a set of measure 0. His example of the characteristic function of the rationals was produced to give an example of a function for which he did not see any way of defining an integral.

So Dirichlet was trying to make the class of allowable functions as large as he could, given the analytical tools that were available. In fact, he says as much in his 1829 paper. Riemann certainly understood this quite well—one of the main results of his paper was his definition of what we now call the Riemann integral. It was developed explicitly to further Dirichlet's goal of widening the class of functions to which the proofs of harmonic analysis could apply[14].

There certainly was an evolution in the understanding of what constituted a function from Dirichlet through Riemann to Weierstrass and Cantor, and then to Baire and Lebesgue. But I think Dirichlet's conception lies well within this path of development, and is fundamentally different from the notions of d'Alembert, Euler, and even Fourier[15].

**1872** Schwarz gave the modern proof that $P_r * f(\theta) \to f(\theta)$ when $f$ is continuous. (This was the paper in which Schwarz solved the Dirichlet problem for the disc.) This may have been the first proof that used what is now seen to be an approximate identity argument.

Subsequently, in 1885, Weierstrass proved what is now called the Weierstrass approximation theorem, by giving an approximate identity argument based on a kernel of the form $e^{-x^2}$. He was familiar with this kernel because essentially the same convolution provides a solution to the heat equation in mathematical physics.

**1876** du Bois Reymond gave an explicit construction of a continuous function whose Fourier series fails to converge at a particular point. He also constructed a continuous function whose Fourier series fails to converge at the points of a dense set.

This showed the futility of trying to prove pointwise convergence for continuous functions in general. It led to a number of investigations into ways of simplifying and understanding the kind of pathological examples that du Bois Reymond had produced. A particularly incisive analysis was made by Lebesgue in 1909, and this led ultimately to the Banach-Steinhaus theorem, which can be used to give an easy (although non-constructive) proof of du Bois Reymond's result.

Counterexamples were in the air at this time; one year earlier, du Bois Reymond had published Weierstrass's example of a continuous nowhere differentiable function[16].

**1880** Starting in this year, and possibly inspired by du Bois Reymond's counterexample, interest revived in dealing with series that diverged but for which there was a reasonable method of associating a sum. Such methods ultimately became known as *summability methods*. The first paper of this sort was by Frobenius in 1880—he showed that (using modern terminology)

[14] I have followed Hawkins (1970) and Dauben (1979) in this discussion.

[15] As another example: Fourier, whom Dirichlet held in high esteem, always used the term "discontinuity" in Euler's sense. So even though Fourier's work, more than any other's, demonstrated the need for a wider notion of function, he remains a true transitional figure, with one foot still in the past.

[16] The example actually dates from 1872 in a paper presented by Weierstrass to the Königl. Akademie der Wissenschaften. Weierstrass subsequently communicated his proof to du Bois Reymond, who published it with attribution in 1875.

Cesàro summability implies Abel summability. That is, if

$$s_k = \sum_{k=1}^{n} a_n$$

is the sequence of partial sums of a (possibly divergent) series, and if the sequence of averages

$$\frac{1}{n} \sum_{k=1}^{n} s_k$$

converges, say to a limit $S$, then also the power series $\sum_{n=1}^{\infty} a_n x^n$ has a limit as $x \uparrow 1$, and that limit is also $S$. This paper was followed by papers of Hölder (1882), Cesàro (1890), and Borel (1896).

## 8.3 The dawn of the twentieth century

**1900** Fejér proved that the (C,1) means of the Fourier series of a continuous function converge everywhere to that function. Fejér's theorem was a sensation, and led to an explosion of work in this area. In 1910 Hardy proved a Tauberian theorem for (C,1) summability—the first modern Tauberian theorem—and pointed out that his theorem gave a simple derivation of Dirichlet's theorem from Fejér's theorem. Immediately after that (also in 1910), Littlewood proved the considerably harder analogue of Hardy's theorem for Abelian summability. This led to the famous Hardy-Littlewood collaboration. In Littlewood's preface to Hardy's *Divergent Series*, he writes:

> The title holds curious echoes of the past, and of Hardy's past. Abel wrote in 1828: 'Divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever.' In the ensuing period of critical revision they were simply rejected. Then came a time when it was found that something after all could be done about them. This is now a matter of course, but in the early years of the century the subject, while in no way mystical or unrigorous, *was* regarded as sensational, and about the present title, now colourless, there hung an aroma of paradox and audacity.

We end by mentioning some of the later history concerning pointwise convergence of Fourier series:

**1915** Luzin hypothesized that the Fourier series of an $L^2$ function (and thus in particular, of a continuous function) converges almost everywhere.

**1923** In the other direction, Kolmogoroff constructed an $L^1$ function whose Fourier series diverges almost everywhere. In 1926 he constructed an $L^1$ function whose Fourier series diverges *everywhere*. This still left the Luzin conjecture unresolved, however.

**1966** Carleson proved the Luzin conjecture: the Fourier series of an $L^2$ function must converge pointwise almost everywhere. Hunt then proved this also holds for any $L^p$ function with $1 < p \leq \infty$.

## 8.4 A question

Now here is a question: Why was Fejér's theorem such a sensation? Looking at things from today's perspective, Poisson had given an essentially equivalent result 77 years previously, and it had subsequently been made perfectly rigorous by Schwarz.

The answer seems to involve the following elements:

- At the time that Poisson produced what is now called the Poisson kernel, the question that Fejér answered was not yet posed. It simply wasn't realized that he had answered a significant question.
- By the time of du Bois Reymond's counterexample, the idea of finding an interpretation for the sum of a possibly non-convergent series was regarded with great suspicion. No one believed at that time that anything useful could be done with such series.
- Later, when methods of summing divergent series began to be studied seriously, it seems that (C,1) summability of a series looked so much closer to ordinary summability that it seemed different in kind from what we now call Abel summability. It's not that Abel summability was unknown or forgotten—it was very well known. But it evidently didn't seem to be the same sort of thing as (C,1) summability, and was not yet referred to as a summability method.

  To put it another way, today we would say that the question is

  > "How can a function be recovered from its Fourier coefficients?"

  It may be however, that at the time, people thought that the question was

  > "How can a function be recovered from the partial sums of its Fourier series?"

  That question was clearly answered by Fejér.

  In both Fejér's announcement of 1900 and his full paper of 1904, he refers to Poisson's theorem and to Schwarz's paper. (In fact, Fejér was a student of Schwarz.) But in both cases, the point he makes is that his theorem, together with the 1880 Frobenius theorem, gives a proof of Poisson's theorem. He doesn't seem to regard Poisson's theorem as really addressing the same issue as his own theorem.

  In Borel's *Leçons sur les Séries Divergentes* of 1901, there is no mention of Fejér's theorem—presumably the book was written just as the theorem was announced. There is, however, an extensive discussion of Cesàro summability, and no mention whatsoever of Poisson's theorem or Abel's theorem. Borel at this time seems not to have thought of Poisson's work as an example of a summability method.

  Borel's book was revised in 1927 by Georges Bouligand, with Borel's approval. The second edition contained extensive material about Fourier series, and included both Fejér's theorem and Poisson's theorem. Nevertheless, the authors still saw them as intuitively very different: they referred to

  > . . . two principles of summation:
  >
  > 1. The principle of averages;
  > 2. The principle of convergence factors.

and called them “two quite distinct ideas”, which they then went on to show were really equivalent. (Borel, 1901, second edition, pages 92 and 88).

- So it may be—I am much less certain of this point, however—that Hardy was alone in 1910 in understanding that the Poisson integral was a method of summability, although even he did not use that term yet. In Hardy’s 1910 paper, he uses the term “summability” only to refer to Cesàro summability. But in Littlewood’s subsequent paper of that same year, he thanks Hardy for pointing out the connection of his Tauberian theorem for power series to Poisson’s theorem (as well as for suggesting the problem itself). And in 1913, a joint paper of Hardy and Littlewood lists several methods of associating a “sum” (their quotes) to a series, among which is what we now call Abel summability. Perhaps it was really the intensive work of Hardy and Littlewood in summability methods and Tauberian theorems that unified this whole area.

# References

Abel, Niels Henrik. 1826. *Untersuchung über die Reihe* $1+\frac{m}{1}x+\frac{m(m-1)}{1\cdot 2}x^2+\frac{m(m-1)(m-2)}{1\cdot 2\cdot 3}x^3+\dots$, Journal für die Reine und Angewandte Mathematik **1**, 311–339. This article was originally written in French, and was translated into German by A. L. Crelle, the editor of the Journal. The original article appears as “Recherches sur la série $1+\frac{m}{1}x+\frac{m(m-1)}{1\cdot 2}x^2+\frac{m(m-1)(m-2)}{1\cdot 2\cdot 3}x^3+\dots$” in Abel’s Œuvres Complètes, 1, 219–250.

———. 1881. *Œuvres Complètes*, Grondahl & Son, Christiania. 2 volumes. Edited by L. Sylow and S. Lie. Reprinted in 1973 by Johnson Reprint Corporation (New York). [QA3.A14].

Birkhoff, Garrett. 1973. *A Source Book in Classical Analysis*, Harvard University Press, Cambridge (Mass.) Edited by Garrett Birkhoff with the assistance of Uta Merzbach. [QA300.B54].

Borel, Émile. 1901. *Leçons sur les Séries Divergentes*, First, Gauthier-Villars, Paris. Second edition, revised with the assistance of Georges Bouligand, was published in 1928. [QA331.B72].

Bottazzini, U. (Umberto). 1986. *The higher calculus : a history of real and complex analysis from Euler to Weierstrass*, Springer, New York. Translated by Warren Van Egmond. [QA300.B67].

Dauben, Joseph Warren. 1979. *Georg Cantor; his mathematics and philosophy of the infinite*, Harvard University Press, Cambridge, Massachusetts. [QA248.D27].

Dirichlet, J. P. G. (Johann Peter Gustav) Lejeune. 1829. *Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données*, Journal für die Reine und Angewandte Mathematik **4**, 157–169. Werke, 1, 117–132.

———. 1837. *Über die Darstellung ganz willkürlicher Funktionen durch Sinus- und Cosinusreihen*, Repertorium der Physik **1**, 152–174. Werke, 1, 133–160.

———. 1863. *Démonstration d’un Théorème d’Abel*, Journal de Mathématiques Pures et Appliquées **Series 2, Volume 7**, 253–255. Note written by Liouville reporting a proof of Dirichlet, and published after Dirichlet’s death. Also in Dirchlet’s Werke, 2, 305–306.

———. 1889. *G. Lejeune Dirichlet’s werke*, G. Riemer, Berlin. 2 volumes. Edited by L. Kronecker. Author’s name actually appears as ‘Peter Gustav Lejeune Dirichlet’. [QA3.L42].

Fejér, Lipót. 1900. *Sur les fonctions bornées et intégrables.*, Comptes Rendus de l’Académie des Sciences (Paris) **131**, 984–987.

———. 1904. *Untersuchungen über Fouriersche Reihen*, Mathematische Annalen **58**, 51–69.

Gibson, George A. 1893. *On the History of the Fourier Series*, Proceedings of the Edinburgh Mathematical Society **11**, 137–166. This paper exhibits the understanding that was current after the early work of Cantor, and before the work of Fejér and Lebesgue.

Grattan-Guinness, Ivor. 1970. *The development of the foundations of mathematical analysis from Euler to Riemann*, MIT Press, Cambridge (Mass.) [QA300.G67].

———. 1972. *Joseph Fourier, 1768–1830; a survey of his life and work, based on a critical edition of his monograph on the propagation of heat, presented to the Institut de France in 1807*, MIT Press, Cambridge (Mass.) [QA29.F68G7].

Hardy, G. H. (Godfrey Harold) and J. E. (John Edensor) Littlewood. 1913. *The relations between Borel's and Cesàro's methods of summation*, Proc. London Math. Soc. **11**, 1–16. Hardy's Collected Papers, 6, 411–426.

Hardy, G. H. (Godfrey Harold). 1910. *Theorems relating to the summability and convergence of slowly oscillating series*, Proc. London Math. Soc. **8**, 301–320. Collected Papers, 6, 291–310.

———. 1949. *Divergent Series*, Oxford University Press, London. [QA295.H29].

———. 1966. *Collected Papers of G. H. Hardy; including joint papers with J. E. Littlewood and others; edited by a committee appointed by the London Mathematical Society*, Clarendon Press (Oxford University Press), London. Seven volumes. [QA3.H29].

Hawkins, Thomas. 1970. *Lebesgue's theory of integration: its origins and development*, University of Wisconsin Press, Madison, Wisconsin. Republished by Chelsea in 1975. [QA312.H34].

Hobson, Ernest William. 1927 (Volume I, 3d edition); 1926 (Volume II, 2nd edition). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, Cambridge University Press, Cambridge (England). Republished by Dover in 1957. [QA331.H6].

Jourdain, Philip E. B. 1913. *Note on Fourier's Influence on the Conception of Mathematics*, Proceedings of the Fifth International Congress of Mathematicians, (Cambridge, 22–28 August 1912), pp. 526–527. Edited by E. W. Hobson and A. E. H. Love. [QA1.I61].

Kline, Morris. 1972. *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York. [QA21.K516].

Lakatos, Imre. 1976. *Proofs and Refutations: the logic of mathematical discovery*, Cambridge University Press, Cambridge (England); New York. [QA8.4.L34].

Lebesgue, Henri. 1905. *Recherches sur la Convergence des Séries de Fourier*, Mathematische Annalen **61**, 251–280.

Littlewood, J. E. (John Edensor). 1910. *The converse of Abel's theorem on power series*, Proc. London Math. Soc. **9**, 438–448.

Luzin, Nicolai. 1998. *Function: Part I*, American Mathematical Monthly **105**, no. 1, 59–67. Translation by Abe Shenitzer of the first half of Luzin's article on *Function* in the first edition of *The Great Soviet Encyclopedia*, Vol. 59, pp. 314–334.

Riemann, Bernhard. 1868. *Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe*, Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen **13**, 87–132. Written in 1852; published after being found by Dedekind after Riemann's death. Werke, 227–264.

———. 1892. *Gesammelte Mathematische Werke*, B. G. Teubner, Leipzig. Edited by Heinrich Weber with the assistance of Richard Dedekind. Republished by Dover Publications (New York) in 1902 with a supplement edited by M. Noether and W. Wirtinger and an introduction by Hans Lewy. [QA3.R55].

Schwarz, Karl Herman Amandus. 1872. *Zur Integration der partiellen Differentialgleichung* $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$, Journal für die Reine und Angewandte Mathematik **74**, 218–253. Ges. Abh. 2, 175–210.

———. 1890. *Gesammelte Mathematische Abhandlungen*, J. Springer, Berlin. 2 volumes. [QA3.S38].

Shapiro, Harold S. 1969. *Smoothing and Approximation of Functions*, Van Nostrand Reinhold, New York. [QA221.S46].

Struik, Dirk Jan. 1969. *A Source Book in Mathematics, 1200–1800*, Harvard University Press, Cambridge (Mass.) [QA21.S88].

Truesdell, C. (Clifford). 1960. *The Rational Mechanics of Flexible or Elastic Bodies, 1638–1788*, Orell Füssli, Zurich. Volume 11, Part 2 of Leonhard Euler: Opera Omnia, Series 2. [Q113.E88].