Mathematical Methods

University of Cambridge Part IB Mathematical Tripos

David Skinner

Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge CB3 0WA
United Kingdom

d.b.skinner@damtp.cam.ac.uk
http://www.damtp.cam.ac.uk/people/dbs26/

Abstract: These are the lecture notes for the Mathematical Methods course given to students taking Part IB Maths in Cambridge during Michaelmas Term of 2014. The course aims to provide an introduction to Fourier Series and Fourier Transforms, Sturm–Liouville Theory, and PDEs such as the Laplace, heat and wave equations.

Preliminaries

Recommended Books

While these notes should contain the material we cover in the lectures, they're very far from a comprehensive treatment and undoubtedly reflect my idiosyncracies in choice of material. So it's a good idea to balance them with a proper textbook. Here are some of the ones I've found useful in preparing the course.

- Arfken, G. and Weber, H., *Mathematical Methods for Physicists*, Academic (2005).
  The single most suitable book for this course. Covers all the core material.
- Boas, M., *Mathematical Methods in the Physical Sciences*, Wiley (2005).
  Also appropriate for this course.
- Mathews, J. and Walker, R. *Mathematical Methods of Physics*, Benjamin Cummins (1970).
- Jeffreys, H. and Jeffreys B.*Methods of Mathematical Physics*, CUP 3rd edition (1999).
  A classic. To be found on the shelves of many generations of mathematical physicists.
- Körner, T. *Fourier Analysis*, Cambridge (1989).
  More advanced, but wonderful. Very engagingly written with a unique blend of mathematical rigour and historical anecdote. I enjoyed reading this a lot when preparing these notes.
- Renardy, M. and Rogers, R. *An Introduction to Partial Differential Equations*, Springer (2004).
  Again more advanced; contains lots of extra material going into further depth in the later parts of this course.

Acknowledgments

Nothing in these lecture notes is original to me. In particular, the notes are based on lecture notes I inherited from Prof R. Jozsa, which were in turn inherited from Prof. C. Caulfield. You can still find Prof. Jozsa's notes here. I've also borrowed heavily from sections of the books listed above. Any errors, major or minor, are of course mine. If you spot one please email me and point it out.

I am supported in part by the European Research Council under an FP7 Marie Curie Career Integration Grant.

1 Fourier Series

Many of the most important equations of mathematical physics are *linear*, including

Laplace's equation	$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\phi(x,y) = 0$
The heat (or diffusion) equation	$\left(\frac{\partial}{\partial t} - K\frac{\partial^2}{\partial x^2}\right)\phi(t,x) = 0$
The wave equation	$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)\phi(t,x) = 0$
Schrödinger's equation	$\left(\mathrm{i}\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - V(x)\right)\psi(t,x) = 0$
Maxwell's vacuum equations	$\nabla\cdot\mathbf{E} = 0 \qquad \nabla\cdot\mathbf{B} = 0$
	$\nabla\times\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} \qquad \nabla\times\mathbf{B} = \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}$

Linearity means that if we are given two solutions ϕ_1 and ϕ_2 of one of these equations – say the wave equation – then $\lambda_1\phi_1 + \lambda_2\phi_2$ is also a solution for arbitrary constants λ_1, λ_2.

With one possible exception, the real reason all these equations are linear is the same: they're approximations. The most common way for linear equations to arise is by slightly perturbing a general system. Whatever the complicated equations governing the dynamics of the underlying theory, if we just look to first order in the small perturbations then we'll find a linear equation essentially by definition[1]. For example, the wave equation will give a good description of ripples on the surface of a still pond, or light travelling through a pane of glass, but don't expect to use it to find out how big a splash you'll make when you bomb into the swimming pool, or if we shine a strong laser at the glass. Similarly, we'll learn how to use the heat equation to tell us about the average jiggling of the atoms in a metal bar when it's being gently warmed somewhere, but if we jiggle them too much then the metal bar will melt.

The possible exception is Schrödinger's equation in Quantum Mechanics. We know of many ways to generalize this equation, such as making it relativistic or passing to Quantum Field Theory, but in each case the analogue of Schrödinger's equation always remains exactly linear. No one knows if there is a fundamental reason for this (though it's certainly built into the principles of Quantum Mechanics at a deep level), or whether our experiments just haven't probed far enough.

In any case, learning to solve linear differential equations such as the above, and their generalizations to higher dimensions, is an important first step in understanding the dynamics of a very wide class of physical (and even biological) systems. Fourier's insight was to take linearity as the key: if we can find a class of simple solutions then we may be able to construct a more general one by taking linear combinations of these.

[1] Often with a source term.

1.1 Vectors

Let's begin by recalling a few facts about vectors that you met last year. A *vector space* over a field F (in this course we'll always take $F = \mathbb{R}$ or $F = \mathbb{C}$) is defined to be a set V together with the operation $+$ of addition, obeying

$$\begin{aligned} \text{commutativity} \qquad & \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \\ \text{associativity} \qquad & \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w} \\ \text{identity} \qquad & \exists!\ \mathbf{0} \in V \ s.t.\ \mathbf{0} + \mathbf{u} = \mathbf{u} \end{aligned}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and the operation of multiplication by a scalar $\lambda \in F$ that is

$$\begin{aligned} \text{distributive in } V \qquad & \lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v} \\ \text{distributive in } F \qquad & (\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}\ . \end{aligned}$$

It's often useful to give our vector space an *inner product*. This is a choice of map $(\ ,\ ) : V \times V \to F$ that obeys[2]

$$\begin{aligned} \text{conjugate symmetry} \qquad & (\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})^* \\ \text{linearity} \qquad & (\mathbf{u}, \lambda\mathbf{v}) = \lambda(\mathbf{u}, \mathbf{v}) \\ \text{additivity} \qquad & (\mathbf{u}, \mathbf{v} + \mathbf{w}) = (\mathbf{u}, \mathbf{v}) + (\mathbf{u}, \mathbf{w}) \\ \text{positive-definiteness} \qquad & (\mathbf{u}, \mathbf{u}) \geq 0 \text{ for all } \mathbf{u} \in V, \text{ with equality iff } \mathbf{u} = \mathbf{0}. \end{aligned}$$

The inner product gives us a notion of lengths and angles in our vector space. We define the length of a vector $\mathbf{u}$ to be the norm $\sqrt{(\mathbf{u}, \mathbf{u})}$ and then

$$\phi = \arccos\left(\frac{(\mathbf{u}, \mathbf{v})}{\sqrt{(\mathbf{u}, \mathbf{u})\,(\mathbf{v}, \mathbf{v})}}\right) \tag{1.1}$$

defines the angle between two vectors.

Note that if our vectors are real, then the property $(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})^*$ implies that $(\ ,\ )$ is symmetric in its arguments. In this case, the map $(\ ,\ ) : V \times V \to \mathbb{R}$ is bilinear. If $F = \mathbb{C}$ the maps is sometimes called *sesquilinear*.

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ form a *basis* of V if any element $\mathbf{u} \in V$ can be uniquely written as $\mathbf{u} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i$ for some scalars λ_i. The *dimension* of the vector space is the number of elements of any basis. A basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is *orthogonal wrt the inner product* if $(\mathbf{v}_i, \mathbf{v}_j)$ vanishes whenever $i \neq j$, the name coming from (1.1). The basis is *orthonormal* if also the length of each $\mathbf{v}_i$ is 1. If we're given an orthonormal basis, we can use the inner product to explicitly decompose a general into this basis. For example, if

$$\mathbf{u} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i\ , \tag{1.2}$$

[2]Beware! It's very common for some authors to define the inner product to be linear in the *first* entry, rather than the second as I have done here. I've chosen this way for maximal agreement with your Quantum Mechanics lectures. And because I'm a physicist.

then by orthonormality of the $\{\mathbf{v}_i\}$ basis,

$$(\mathbf{v}_j, \mathbf{u}) = \sum_{i=1}^{n} \lambda_i (\mathbf{v}_j, \mathbf{v}_i) = \lambda_j \,, \tag{1.3}$$

which tells us λ_j. For real vectors, λ_j is just the projection of $\mathbf{u}$ onto $\mathbf{v}_j$.

1.2 Spaces of functions as infinite dimensional vector spaces

Consider the set of complex valued functions on some domain Ω. Such a function f can be viewed as a map $f : \Omega \to \mathbb{C}$. The set of all such functions is naturally thought of as a vector space, where vector addition $+$ is just pointwise addition of the functions; that is, for $x \in \Omega$ we have

$$(f + g)(x) = f(x) + g(x) \tag{1.4}$$

where the addition on the rhs is just addition in $\mathbb{C}$. Likewise, we can multiply functions by scalars as

$$(\lambda f)(x) = \lambda \, f(x) \,, \tag{1.5}$$

where again the multiplication is just the usual multiplication in $\mathbb{C}$.

How about the inner product? One possible choice is to take

$$(f, g) \equiv \int_\Omega f(x)^* \, g(x) \, \mathrm{d}\mu \tag{1.6}$$

where $\mathrm{d}\mu$ is some choice of integration measure, and where the functions $f(x)$ and $g(x)$ are sufficiently well-behaved that the integral exists. The idea is that this is a simple generalization of the inner product between two finite dimensional vectors: if we think of the different point $x \in \Omega$ as labelling the different 'components' of our functions, then we multiply component of f and g together (after taking an appropriate complex conjugate) and then add them up (*i.e.,* integrate over Ω). The measure $\mathrm{d}\mu$ tells us how much weight to assign to each point of the domain.

As a simple example, if Ω is the interval $[a, b]$, then we may take the measure to be just $\mathrm{d}x$ so that

$$(f, g) = \int_a^b f(x)^* \, g(x) \, \mathrm{d}x \,. \tag{1.7}$$

As a second example, if Ω is the unit disc D_2 then we may take

$$(f, g) = \int_{r=0}^{1} \int_{\theta=0}^{2\pi} f(r, \theta)^* \, g(r, \theta) \, r \mathrm{d}r \, \mathrm{d}\theta \tag{1.8}$$

with measure $\mathrm{d}\mu = r \mathrm{d}r \, \mathrm{d}\theta$. Later in the course, we'll meet some other measures.

If the domain Ω is bounded, then we may sometimes wish to restrict the class of functions we consider by requiring they satisfy *boundary conditions*. Boundary conditions that preserve the vector space structure – in the sense that if f and g both satisfy the boundary conditions then so too does $\lambda_1 f + \lambda_2 g$ – are called *homogeneous*. For example, if $\Omega = [a, b]$ then the boundary conditions $f(a) = 0$, $f(a) = f(b)$ and $f(a) + 7f'(b) = 0$ are all homogeneous, whereas the boundary conditions $f(a) = 1$ or $f(a) + 3f'(b) = 7$ are not.

An important class of functions are *periodic* functions. In the first instance, these may be thought of as maps $f : \mathbb{R} \to \mathbb{C}$ that obey the condition $f(t+T) = f(t)$ for all t (and conventionally we take the smallest such T). The fixed constant T is called the *period*, while $1/T$ is the *frequency*. Note that f is fully specified once we give its values on $[0, T)$. The basic examples of periodic functions are just the trigonometric functions $\sin \omega t$ and $\cos \omega t$, which each have period $T = 2\pi/\omega$.

If we interpret the variable t as time, then T is the length of time is takes for our function to complete one whole oscillation, while $1/T$ is the number of oscillations per unit time. The constant ω is known as the *angular* frequency. It tells us the number of oscillations that fit in a 2π interval (useful when we're thinking in terms of a map from a circle). Sometimes, we may prefer to think of our function $f(x)$ as being periodic in space rather than time, for example $A \sin kx$. In this case, we call $\lambda \equiv 2\pi/k$ the *wavelength* as it tells us the spatial extent of one complete oscillation. $1/\lambda = k/2\pi$ is called the wavenumber and gives the number of waves per unit length, while the constant k is the angular wavenumber. In practice, we often do not distinguish between frequency/wavenumber and angular frequency/wavenumber, with the terminology being clear from the context.

In what follows below we'll often treat periodic functions as maps $f : S^1 \to \mathbb{C}$ from the circle. To simplify the notation we'll usually parameterize the circle with a coordinate $\theta \in [-\pi, \pi)$. So if you're given a periodic function $f(t)$ whose period is T, you must remember to set $\theta = 2\pi t/T$ and $\mathrm{d}\theta = 2\pi/T \, \mathrm{d}t$ to use the formulæ below!

1.3 Fourier series

The simplest and most frequently occurring complex-valued periodic function is the complex exponential $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$. This is periodic with period 2π, and so we can view it as a map $\exp : S^1 \to \mathbb{C}$ from the circle described by coordinates θ. By convention, we'll take $\theta \in [-\pi, \pi)$. Integer powers of these exponentials are orthogonal with respect to the inner product[3]

$$(\mathrm{e}^{\mathrm{i}m\theta}, \mathrm{e}^{\mathrm{i}n\theta}) = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}m\theta} \, \mathrm{e}^{\mathrm{i}n\theta} \, \mathrm{d}\theta = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(n-m)\theta} \, \mathrm{d}\theta = 2\pi \, \delta_{m,n} \tag{1.9}$$

where $m, n \in \mathbb{Z}$ and $\delta_{m,n}$ is the Kronecker δ-symbol defined by

$$\delta_{m,n} = \begin{cases} 1 & \text{when } n = m \\ 0 & \text{else.} \end{cases} \tag{1.10}$$

(To check the integral, note that whenever $m \neq n$ we're just integrating either sin or cos over a complete period. When $m = n$ we're just integrating 1.) Thus the set $\left\{ \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}n\theta} : n \in \mathbb{Z} \right\}$ form an orthonormal set of complex valued periodic functions.

Fourier's idea was to try to use this set as a basis in which to expand any such periodic function. Given an arbitrary function $f : S^1 \to \mathbb{C}$, we define the *Fourier coefficients* $\hat{f}_n$ of

[3]Recall that the inner product (,) on complex functions is *anti*linear in its first slot, which is the origin of the minus sign in the argument of the first exponential.

$f(\theta)$ by[4]

$$\hat{f}_n \equiv \frac{1}{2\pi}(\mathrm{e}^{\mathrm{i}n\theta}, f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta} f(\theta)\,\mathrm{d}\theta \tag{1.11}$$

for all $n \in \mathbb{Z}$, just as we did in (1.3). Fourier now claimed that

$$f(\theta) \stackrel{?}{=} \sum_{n\in\mathbb{Z}} \hat{f}_n\,\mathrm{e}^{\mathrm{i}n\theta} \tag{1.12}$$

thus expanding f in the basis $\{\mathrm{e}^{\mathrm{i}n\theta}/\sqrt{2\pi}\}$ by analogy with (1.2). This expansion is known as the *Fourier series* of $f(\theta)$. As we'll see later in the course, this idea and its generalizations has proven incredibly useful in any number of problems, from vibrating strings (violins and quantum gravity alike), to learning about the origin of galaxies from fluctuations in the early universe, to the electronic wizardry in your laptop, as well as to vast swathes of functional analysis. But to begin with, no-one believed him.

1.3.1 Reality conditions

As an aside, let me point out a common reformulation of the Fourier series that is relevant if $f(\theta) = f(\theta)^*$ so that f is real-valued (rather than $\mathbb{C}$-valued). In this case, the Fourier coefficients obey $\hat{f}_n^* = \hat{f}_{-n}$. This follows straight from their definition:

$$(\hat{f}_n)^* = \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\, f(\theta)\,\mathrm{d}\theta\right)^* = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{+\mathrm{i}n\theta}\, f(\theta)\,\mathrm{d}\theta = \hat{f}_{-n}\ , \tag{1.13}$$

using the reality of $f(\theta)$. We can use this property to reorganise the Fourier series in terms of standard trigonometric functions, because

$$\begin{aligned} f(\theta) &= \sum_{n\in\mathbb{Z}} \hat{f}_n\,\mathrm{e}^{\mathrm{i}n\theta} \\ &= \hat{f}_0 + \sum_{n=1}^{\infty} \hat{f}_n\,\mathrm{e}^{\mathrm{i}n\theta} + \sum_{n=-\infty}^{-1} \hat{f}_n\,\mathrm{e}^{\mathrm{i}n\theta} \\ &= \hat{f}_0 + \sum_{n=1}^{\infty} \left(\hat{f}_n\,\mathrm{e}^{\mathrm{i}n\theta} + \hat{f}_n^*\,\mathrm{e}^{-\mathrm{i}n\theta}\right) \\ &= \hat{f}_0 + \sum_{n=1}^{\infty} a_n \cos n\theta + b_n \sin n\theta \end{aligned} \tag{1.14}$$

where in going to the third line we relabelled $n \to -n$ in the last sum and used $\hat{f}_{-n} = \hat{f}_n^*$ for real f. In going to the final line we've used de Moivre's theorem $\mathrm{e}^{\mathrm{i}n\theta} = \cos n\theta + \mathrm{i}\sin n\theta$ and set $\hat{f}_n = (a_n - \mathrm{i}b_n)/2$ so that

$$a_n \equiv 2\,\mathrm{Re}\,\hat{f}_n \qquad \text{and} \qquad b_n \equiv -2\,\mathrm{Im}\,\hat{f}_n\,. \tag{1.15}$$

[4]The factor of $1/2\pi$ is for later convenience and is included by convention. Other common conventions replace this factor with 1 or $1/\sqrt{2\pi}$.

Note that the sum in our trigonometric Fourier series runs only over positive integers, but that we now have two sets of Fourier coefficients, a_n and b_n. From the definition of the $\hat{f}_n$s and (1.16) we see that

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos n\theta\, f(\theta)\,\mathrm{d}\theta \qquad \text{and} \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin n\theta\, f(\theta)\,\mathrm{d}\theta\,, \tag{1.16}$$

so a_n and b_n are the Fourier coefficients we'd find if we use $\sin n\theta$ and $\cos n\theta$ as our basis of periodic functions instead of the complex exponentials.

As we'll see later, the real form of the Fourier series is often useful when we are studying some real physical object (such as a vibrating violin string) and we wish to fix a solution of some linear differential equation to obey appropriate real boundary conditions. Boundary conditions may also mean that it is convenient to consider a function $g(\theta)$ defined not over a whole domain, but just on a half-interval, say $\theta \in [0, \pi]$. We can then construct a function defined throughout $[-\pi, \pi)$ by *extending* the given function $g(\theta)$ either as an even or odd function, so that

$$g(\theta) = \pm g(-\theta) \qquad \text{for } \theta \in [-\pi, 0)\,.$$

For example, if we have the (Dirichlet) boundary condition that $g(0) = 0$ (perhaps because a string is tied to a fixed wall there), then it is natural to extend $g(\theta)$ as an odd function, whereas for Neumann boundary conditions extension as an even function may be more appropriate. As above, if g is extended to be even then the real Fourier series will involve non-vanishing coefficients a_n and $\hat{f}_0$ (for $\cos n\theta$ and the constant). Conversely, if it is extended to be odd then the real Fourier series contains non-vanishing coefficients b_n of $\sin n\theta$ only.

However, it is worth pointing out that, even when we ultimately wish to obtain a real Fourier series, it is usually simpler to work with the complex exponential form during the intermediate steps of a calculation. We then impose the relevant boundary conditions and fix the real form only at the end. One reason for this is that the complex exponentials $\mathrm{e}^{\mathrm{i}n\theta}$ have the beautiful property that differentiation *wrt* θ becomes simply multiplication by in.

1.4 Fejér's theorem

Sadly, this section is non-examinable, at least for this course. You really shouldn't worry about that in Michaelmas.

What made people so reluctant to accept (1.12) was that Fourier claimed it would hold for an *arbitrary* periodic function $f(\theta)$, no matter whether it was smooth or had kinks, *e.g.* $f(\theta) = |\theta|$ for $\theta \in [-\pi, \pi)$ which has a kink at $\theta = 0$, or even functions with discontinuities, such as the *sawtooth function* $f(\theta) = \theta$ which is discontinuous at the point $-\pi = \pi$ on the circle. This claim caused a great deal of concern among mathematicians of the time. Laplace and Lagrange strongly doubted that (1.12) could be true, and blocked Fourier's attempts to publish his theory of heat flow based on using such series. In a rare moment of weakness, Cauchy incorrectly claimed to have a proof that (1.12) does make sense. Dirichlet did manage to prove the validity of the Fourier series for *continuous* functions with bounded continuous derivative, but these conditions are quite restrictive (for example,

neither of the functions above obey them). The restrictions are important too, because in 1876 du Bois-Reymond found a continuous function (without a continuous derivative) for which the series fails for certain values of θ, while Weierstrass constructed a continuous but nowhere differentiable function whose rôle in life (at least initially) seemed to be purely to pour scorn on Fourier.

The issue, of course, is whether the infinite sum (1.12) converges and, if it does, whether it actually converges to give $f(\theta)$. The first thing to realise is that there are many things we might wish to mean by 'converge'. For example, defining $S_n f$ to be the partial sum

$$S_n f \equiv \sum_{k=-n}^{n} \hat{f}_k \, \mathrm{e}^{\mathrm{i}n\theta} \tag{1.17}$$

where the coefficients $\hat{f}_k$ were defined in (1.11), we might ask simply that

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} |f(\theta) - S_n f(\theta)|^2 \, \mathrm{d}\theta = 0 \ . \tag{1.18}$$

If this holds, it tells us that as more and more terms are included in the sum, the deviation of the Fourier series from the true function tends to zero *on average*. But it still allows for arbitrarily large deviations in both directions. A stronger notion of convergence is *pointwise* convergence: we could ask that

$$\lim_{n\to\infty} |f(\theta) - S_n f(\theta)| = 0 \qquad \text{at every point } \theta \in S^1 \, , \tag{1.19}$$

so that the Fourier series converges to f everywhere on its domain. Stronger still is *uniform* convergence, where we ask in addition that the *rate* at which $S_n f$ converges to f is the same everywhere on S^1. (You'll learn much more about pointwise and uniform convergence in your Analysis II course.)

One of the main theorems in Fourier analysis was proved at the turn of the 20^{th} century by the Hungarian mathematician Fejér. Instead of asking whether $S_n f$ itself converges, Fejér asked whether, given all the Fourier coefficients $\hat{f}_n$, the original function $f(\theta)$ could be recovered. He showed that provided $f : S^1 \to \mathbb{C}$ is merely *continuous* then this can be done, with no further assumptions on the smoothness of f. He was nineteen.

Here's an outline of his proof. First, for any sequence $\{s_0, s_1, s_2, \ldots\}$, we construct a new sequence $\{\sigma_0, \sigma_1, \sigma_2, \ldots\}$ where

$$\sigma_n \equiv \frac{1}{n+1} \sum_{m=0}^{n} s_m \tag{1.20}$$

is just the arithmetic mean of the first $n+1$ terms in the original sequence. Césaro noticed that this new sequence may have better convergence properties than the original sequence. The standard example is to take $s_m = (-1)^m$ which clearly does not converge, while

$$|\sigma_n| = \left| \frac{1}{n+1} \sum_{m=0}^{n} s_n \right| = \frac{1}{n+1} \left| \sum_{m=0}^{n} s_m \right| \leq \frac{1}{n+1}$$

which converges to zero as $n \to \infty$.

Fejér realized that this idea could be applied to help the convergence of a Fourier series. From (1.17) and the definition of the Fourier coefficients we have

$$\begin{aligned} S_m f(\theta) &= \frac{1}{2\pi} \sum_{k=-m}^{m} \left[\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\phi} f(\phi)\, \mathrm{d}\phi \right] \mathrm{e}^{\mathrm{i}k\theta} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) \sum_{k=-m}^{m} \mathrm{e}^{\mathrm{i}k(\theta-\phi)}\, \mathrm{d}\phi \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi)\, D_m(\theta-\phi)\, \mathrm{d}\phi \end{aligned} \tag{1.21}$$

where $D_m(x) = \sum_{k=-m}^{m} \mathrm{e}^{\mathrm{i}kx}$ is known as the *Dirichlet kernel*. Fejér now applied Césaro's idea, setting

$$\sigma_n(f) = \frac{1}{n+1} \sum_{m=0}^{n} S_m f = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi)\, F_n(\theta-\phi)\, \mathrm{d}\phi \tag{1.22}$$

where the *Fejér kernel* $F_n(x) \equiv \frac{1}{n+1} \sum_{m=0}^{n} D_m(x)$. It's a straightforward exercise to show that

$$F_n(x) = \begin{cases} \dfrac{1}{n+1} \dfrac{\sin^2[(n+1)x/2]}{\sin^2[x/2]} & \text{for } x \neq 0 \\ n+1 & \text{when } x = 0\,, \end{cases} \tag{1.23}$$

but I'll skip the proof. To make progress, we first list some important properties of $F_n(x)$. These are *i)* $F_n(x) \geq 0$ everywhere, *ii)* $\frac{1}{2\pi} \int_{-\pi}^{\pi} F_n(\theta)\, \mathrm{d}\theta = 1$, and *iii)* $F_n(x) \to 0$ uniformly outside an arbitrarily small region $[-\delta, \delta]$ around $\theta = 0$. Property *i)* is obvious. Property *ii)* is best seen using the definition of F_n in terms of the Dirichlet kernel:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} F_n(\theta)\, \mathrm{d}\theta &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{1}{n+1} \sum_{m=0}^{n} \sum_{k=-m}^{m} \mathrm{e}^{\mathrm{i}k\theta} \right] \mathrm{d}\theta \\ &= \frac{1}{n+1} \sum_{m=0}^{n} \sum_{k=-m}^{m} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\theta}\, \mathrm{d}\theta \right] . \end{aligned}$$

The integral in square brackets vanishes whenever $k \neq 0$, and gives 1 when $k = 0$. For each value of m, the value $k = 0$ occurs exactly once, so the double sum gives $n + 1$ and property *ii)* is proved. Property *iii)* follows since for $\delta \leq |x| \leq \pi$

$$F_n(x) \leq \frac{1}{n+1} \frac{1}{\sin^2[x/2]} \leq \frac{1}{n+1} \frac{1}{\sin^2[\delta/2]} \to 0\,,$$

with convergence being uniform.

The idea of Fejér's proof is that since property *iii)* shows that $F_n(\theta-\phi)$ is concentrated on $\theta = \phi$ as $n \to \infty$, in this limit the integral on the right of (1.22) for $\sigma_n(f)$ receives an appreciable contribution only from a tiny neighbourhood of $\phi = \theta$. Since f is continuous, in this neighbourhood $f(\phi)$ may be approximated by $f(\theta)$, so

$$\sigma_n(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi)\, F_n(\theta-\phi)\, \mathrm{d}\phi \approx \frac{f(\theta)}{2\pi} \int_{-\pi}^{\pi} F_n(\theta-\phi)\, \mathrm{d}\phi = f(\theta)\,, \tag{1.24}$$

using property *ii*) of the Fejér kernel. Thus $|\sigma_n - f(\theta)| \to 0$ uniformly as $n \to \infty$ and so the σ_n converge uniformly on the original function. Once you're seen enough of Analysis II, you might like to fill in the 'epsilonics' to make the idea sketched above into a firm proof. It's actually possible to generalize the proof to allow $f : S^1 \to \mathbb{C}$ to be discontinuous at a finite number of isolated points $\{\theta_1, \theta_2, \ldots, \theta_r\} \in S^1$, provided $\int_{S^1} |f(\theta)|\, \mathrm{d}\theta$ exists[5]. Then $\sigma_n(f)$ converges to the original function at all points $\theta \in S^1$ where $f(\theta)$ *is* continuous.

Fejér assures us that we can recover any continuous function from its Fourier coefficients, but not that the partial Fourier series $S_n f$ itself actually converge when $n \to \infty$. In fact, one can prove that for continuous functions, $S_n f$ does indeed converge to $f(\theta)$ provided $\sum_{n=-\infty}^{\infty} |\hat{f}_n|$ converges. The proof is an application of the Weierstrass M test. Have a go at proving it if you know what this is!

1.5 Functions with discontinuities

When you flick on a light switch, the current passing through an average bulb passes very rapidly from zero to about 10^{18} electrons per second (a bit less than half an amp). Wifi networks transmit data via a signal that flickers between 0 and 1 up to a million times a second. In this section, we'll examine the behaviour of the Fourier coefficients of functions with sudden jumps. We'll always assume that our periodic function f has only a finite number of such discontinuities, say at isolated points $\{\theta_1, \theta_2, \ldots, \theta_r\} \in S^1$, and that it jumps by only a finite amount.

To get started, consider the *sawtooth function* defined by

$$f(\theta) = \theta \qquad \text{for } \theta \in [-\pi, \pi)\,. \tag{1.25}$$

This function is discontinuous at $\theta = -\pi$ on S^1, and gets its name because if we think of the function not as a map from a circle but as a map $f : \mathbb{R} \to \mathbb{R}$ that is periodic with period 2π, then the graph looks like the cutting edge of a saw, as in figure 1.5. Its Fourier coefficients $\hat{f}_n$ are

$$\hat{f}_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta \, \mathrm{d}\theta = 0 \tag{1.26}$$

when $n = 0$, and

$$\begin{aligned} \hat{f}_n &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\, \theta \, \mathrm{d}\theta = \frac{1}{2\pi\mathrm{i}} \left[-\frac{1}{n}\, \theta\, \mathrm{e}^{-\mathrm{i}n\theta} \right]_{-\pi}^{\pi} + \frac{1}{2\pi\mathrm{i}} \frac{1}{n} \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\, \mathrm{d}\theta \\ &= \frac{1}{\mathrm{i}n} (-1)^{n+1} \end{aligned} \tag{1.27}$$

when $n \neq 0$. The coefficients here decay rather slowly, like $1/n$. Despite this, as $n \to \infty$ the partial Fourier series

$$S_n f = \sum_{k=-n}^{n} \frac{(-1)^{k+1}}{\mathrm{i}k}\, \mathrm{e}^{\mathrm{i}k\theta} \tag{1.28}$$

[5] As a Riemann integral.

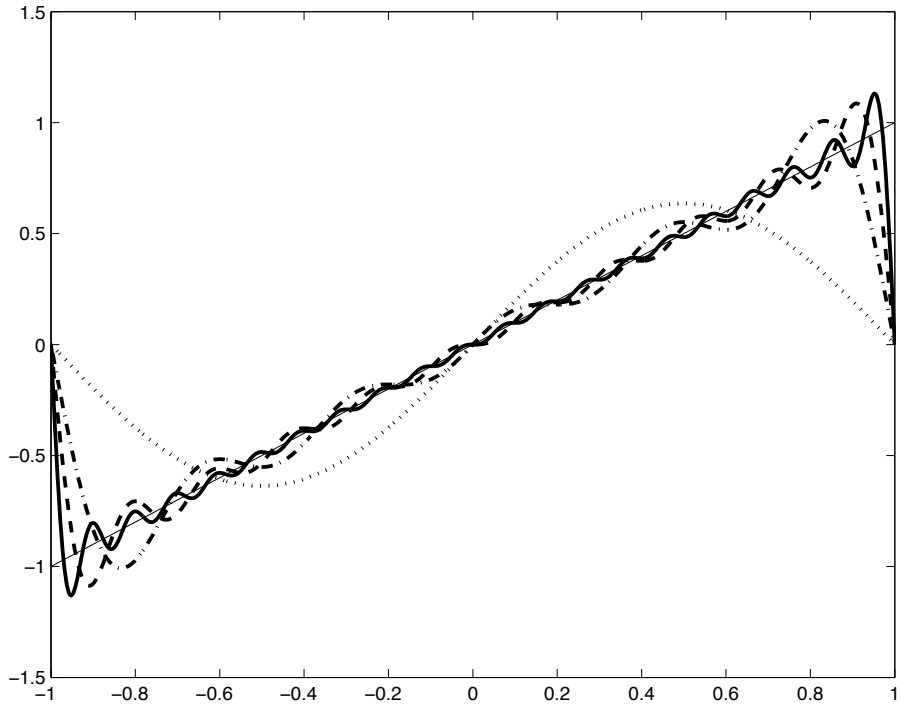

Figure 1. *Plots of the sawtooth function $f(\theta) = \theta$ (thin solid line) together with the partial Fourier sums $S_1 f(\theta)$ (dotted line), $S_5 f(\theta)$ (dot-dashed), $S_{10} f(\theta)$ (dashed) and $S_{20} f(\theta)$ (solid line). Both axes have been rescaled by π.*

do converge on the sawtooth function everywhere except at the discontinuity. Note that since the sawtooth function is real, we can represent it as a trigonometric Fourier series

$$f(\theta) = 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin n\theta \,. \tag{1.29}$$

The fact that $\hat{f}_0$ and all the coefficients of $\cos n\theta$ vanish is exactly as we would expect since $f(-\theta) = -f(\theta)$.

At the discontinuity $\theta = \pi$ the sawtooth function itself is discontinuous. Examining the Fourier series (1.29) we see that since $\sin k\pi = 0$ for any $k \in \mathbb{Z}$, $S_n f(\pi) = S_n f(-\pi) = 0$ for all n, so the Fourier series converges on zero at the location of the discontinuity. The significance of this is that zero is the *average* value of the sawtooth function on either side of the discontinuity.

We now want to show that the behaviour of the Fourier series for the sawtooth is typical for functions with discontinuities. That is, we want to show that if some function $g(\theta)$ jumps, say at $\theta = \pi$, then

$$S_n g(\pi) \to \frac{g(\pi^+) + g(\pi^-)}{2} \qquad \text{where} \qquad g(\pi^\pm) = \lim_{\epsilon \to 0} g(\pi \pm \epsilon) \,, \tag{1.30}$$

or in other words that its Fourier series converges to the average value on either side of the jump, just as we obtained for the sawtooth function. To do so, define

$$G(\theta) \equiv \begin{cases} g(\theta) + \dfrac{g(\pi^+) - g(\pi^-)}{2\pi} f(\theta) & \text{for } \theta \neq \pi \\ \dfrac{g(\pi^+) + g(\pi^-)}{2} & \text{at } \theta = \pi \ , \end{cases} \tag{1.31}$$

where $f(\theta)$ is exactly the sawtooth function from above. The point of this definition is that $G(\theta)$ is now continuous at π, because the discontinuity in our function $g(\theta)$ is balanced by the discontinuity of the sawtooth. The Fourier coefficients of G are linear combinations of those of g and those of the sawtooth. Since the Fourier series of the sawtooth converges, $S_n G$ will converge provided $S_n g$ does. But because G is continuous, its Fourier series will now converge to the original function G everywhere. In particular, at $\theta = \pi$ we have $S_n G(\pi) \to G(\pi) = (g(\pi^+) + g(\pi^-))/2$. However,

$$S_n G(\pi) = S_n g(\pi) + \frac{g(\pi^+) - g(\pi^-)}{2\pi} S_n f(\pi) = S_n g(\pi) \tag{1.32}$$

since the Fourier series $S_n f$ of the sawtooth vanishes at $\theta = \pi$. Thus as $n \to \infty$, $S_n g(\pi)$ converges to $(g(\pi^+) + g(\pi^-))/2$, as was to be shown.

1.6 Integration *vs.* differentiation

When we first learnt calculus, most people started with differentiation. The differential $\mathrm{d}f/\mathrm{d}x$ has an intuitively clear definition in terms of the difference in the values of f at nearby pointss. Better still, it is easy to apply this rule the some simple functions we were familiar with such as powers or trigonometric functions, and with practice we may have felt confident we could apply it to any given function. Integration, by contrast, is defined as the 'inverse of differentiation' and so (perhaps) seemed a little mysterious. If we want to know the indefinite integral of x for example, we first have to recognize that this is what we'd have obtained if we'd differentiated $x^2/2+c$. But you may have worried about how we would proceed if we'd been asked to integrate say $\ln(1-x)$, where we might not recognize its integral[6]. Indeed, many functions are simply *defined* to be the integral of some other function.

However, if we're not so much interested in an explicit closed-form expression for our function, but just knowing whether the integral or derivative exists, then matters are different. Vastly more functions can be integrated in principle than can be differentiated. This is because integration is a *smoothing* operation. For example, consider the step function

$$\Theta(x) \equiv \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \tag{1.33}$$

[6]In this case the integral is actually $\mathrm{Li}_2(x)$, known as a dilogarithm. According to the number theorist Don Zagier, it's the only mathematical function with a sense of humour.

which is discontinuous at the origin. The derivative of the step function fails to exist here[7] because the gradient of the function becomes infinite. However, it's easy to integrate

$$\int_{-\infty}^{x} \Theta(y)\,\mathrm{d}y = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0. \end{cases} \tag{1.34}$$

This functions is now continuous at the origin, though it has a kink there. Integrating further we'd create functions that are always zero on the negative x-axis, but which join at the origin in a smoother and smoother way. Conversely, differentiating these functions will lead to worse and worse behaviour at the origin. Note that none of these functions admit a Taylor series around $x = 0$.

The same is true for Fourier series. If we have a function $f : S^1 \to \mathbb{C}$ whose partial Taylor series $S_n f$ converge to f, then

$$\begin{aligned} S_n F &\equiv \int_{-\pi}^{\theta} S_n f(\phi)\,\mathrm{d}\phi \\ &= (\theta - \pi)\hat{f}_0 + \sum_{k=-n}^{-1} \hat{f}_k \frac{\mathrm{e}^{\mathrm{i}k\theta} - (-1)^k}{\mathrm{i}k} + \sum_{k=1}^{n} \hat{f}_k \frac{\mathrm{e}^{\mathrm{i}k\theta} - (-1)^k}{\mathrm{i}k} \end{aligned} \tag{1.35}$$

This new series is guaranteed converge, because the original one did by assumption and integration has suppressed each coefficient by a further power of k. Even if the original function had jump discontinuities, so that at some discrete points the Fourier series converged to the average value of f on either side of the discontinuity, we've seen that integration produces a continuous function for us, so the new series will converge to $F(\theta) = \int_{-\pi}^{\theta} f(\phi)\,\mathrm{d}\phi$ everywhere.

By contrast, if we differentiate a Fourier series term by term then we *multiply* each coefficient by $\mathrm{i}k$ and this makes convergence worse, perhaps fatally. For example, consider the *square wave* given by

$$f(\theta) = \begin{cases} -1 & \text{for } -\pi \leq \theta < 0 \\ +1 & \text{for } 0 < \theta < \pi \end{cases} \tag{1.36}$$

and shown in figure 1.6. You can check (*exercise!*) that this has Fourier series

$$f(\theta) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2n-1)\theta}{2n-1} \tag{1.37}$$

which converges to f everywhere except at the origin, where it converges to zero – the average of the values taken when the origin is approached from the left and from the right. If we formally differentiate term-by-term we obtain the series

$$f'(\theta) \stackrel{?!}{=} \frac{4}{\pi} \sum_{n=1}^{\infty} \cos(2n-1)\theta \tag{1.38}$$

[7] At least as a function. Later in the course we'll meet distributions.

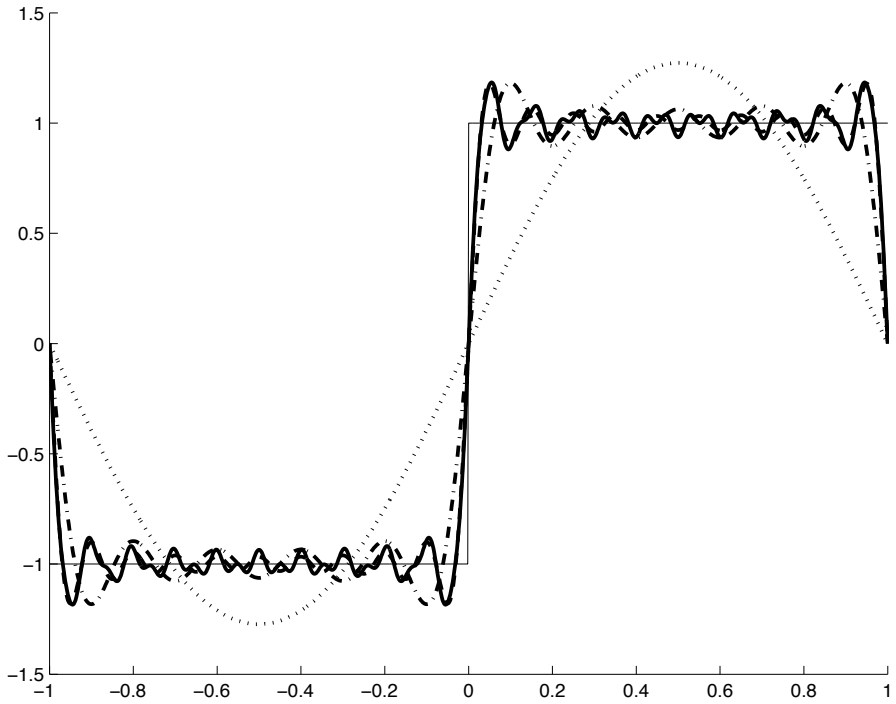

Figure 2. *Plots of the square wave $f(\theta)$ (thin solid line), together with its partial Fourier sums $S_1 f$ (dotted), $S_5 f$ (dot-dashed), $S_{10} f$ (dashed) and $S_{20} f$ (solid). The θ-axis has been rescaled by a factor of π compared to the values in the text.*

which is divergent. The problem, of course is that $f'(\theta)$ is *not defined* at the origin, so $f'(\theta)$ does not admit a Fourier series. You might think we could get around this by just defining $f'(0) \equiv c$ for some constant c. Perhaps you'd choose $c = 0$, which is the value of both $f'(0^+)$ and $f'(0^-)$? The problem is that this breaks the fundamental rule of calculus, that

$$f(\theta) = f(-\pi) + \int_{-\pi}^{\theta} f'(\phi)\, \mathrm{d}\phi\,,$$

because with any finite value of c the integral on the rhs vanishes. Such a rule would mean we never see f jump!

So when can we safely differentiate a Fourier series term by term? Clearly, it is not enough for $f(\theta)$ itself to have a Fourier series that converges. In fact, there is a theorem that if $f : S^1 \to \mathbb{C}$ is continuous *and* $\sum_{k\in\mathbb{Z}} |k||\hat{f}_k|$ converges, then f is differentiable and the partial sums

$$S_n(f') \equiv \sum_{k=-n}^{n} \mathrm{i}k\, \hat{f}_k\, \mathrm{e}^{\mathrm{i}k\theta}$$

converge uniformly to $f'(\theta)$ as $n \to \infty$. To see this, note that the conditions imply that both f and f' do indeed have Fourier series

$$f(\theta) = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} \hat{f}_n\, \mathrm{e}^{\mathrm{i}n\theta} \qquad \text{and} \qquad f'(\theta) = \frac{1}{2\pi} \sum_{n\in Z} \hat{\phi}_n\, \mathrm{e}^{\mathrm{i}n\theta} \tag{1.39}$$

with some Fourier coefficients $\hat{f}_n$ and $\hat{\phi}_n$. The Fourier series of f certainly converges if that of f' does. Furthmore, the Fourier coefficients of f' in (1.39) are, by definition,

$$\hat{\phi}_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{df}{d\theta}\,\mathrm{d}\theta = 0 \tag{1.40}$$

by periodicity of f, and

$$\hat{\phi}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\,\frac{df}{d\theta}\,\mathrm{d}\theta = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\,\frac{d}{d\theta}\mathrm{e}^{-\mathrm{i}n\theta}\mathrm{d}\theta = \mathrm{i}n\hat{f}_n \tag{1.41}$$

for $n \neq 0$, by integration by parts (the boundary term again cancelling by periodicity and continuity). But this is exactly what we would find by differentiating the Fourier series term by term.

1.7 The rate of convergence

In fact, the decay of the Fourier coefficients $\hat{f}_n k$ as $k \to \infty$ tells us quite generally about the order of differentiability (smoothness) of the function. The intuition here is that if a function has a very sharp feature – such as a sudden jump, or the cusp at $\theta = 0$ in $\sqrt{|\theta|}$ – then we will need to include a significant amount of very high frequency terms in its Fourier series. Last year, you learned that in relativity, the frequency of a wave is its energy in units of $\hbar$. Thus this same principle of sharp features needing high frequencies drives the construction of particle accelerators such as the LHC at CERN: if you want to things that happen at a very small scale, you need a very powerful microscope.

Let's now quantify these ideas. We will show that if we're given the Fourier coefficients $\hat{f}_k$ of some continuous function, we can read off the order of differentiability of the original f by seeing how quickly these coefficients fall with k.

Suppose that $f : S^1 \to \mathbb{C}$ together with its first $m-1$ derivatives are continuous, but that the m^{th} derivative $f^{(m)}$ jumps at some isolated points $\{\theta_1, \theta_2, \ldots, \theta_r\} \in S^1$. We further suppose that all these functions remain finite throughout. Now comes a trick. Integrate by parts to write the Fourier coefficient $\hat{f}_k$ (with $k \neq 0$) as

$$\begin{aligned}
\hat{f}_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta}\, f(\theta)\,\mathrm{d}\theta \\
&= \left[-f(\theta)\frac{\mathrm{e}^{-\mathrm{i}k\theta}}{2\pi\mathrm{i}k}\right]_{-\pi}^{\pi} + \frac{1}{2\pi\mathrm{i}k}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta}\, f'(\theta)\,\mathrm{d}\theta \\
&= \frac{1}{2\pi\mathrm{i}k}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta}\, f'(\theta)\,\mathrm{d}\theta\,,
\end{aligned} \tag{1.42}$$

where the boundary terms cancel since f is periodic and continuous everywhere on S^1.

If we perform this trick m times then the boundary terms always cancel, since f', f'' up to $f^{(m-1)}$ are continuous. Therefore

$$\hat{f}_k = \frac{1}{2\pi}\frac{1}{(\mathrm{i}k)^m}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta}\, f^{(m)}(\theta)\,\mathrm{d}\theta\,. \tag{1.43}$$

However, if we want to continue with this game, we must be careful because $f^{(m)}$ has discontinuities at $\{\theta_1, \theta_2, \ldots, \theta_r\}$ and its derivative cannot be defined there. To proceed, recall that even though it jumps, $f^{(m)}$ remains finite everywhere by assumption. So the integral (1.43) does not appreciably if we remove an infinitesimal neighbourhood of each of the discontinuities. For the remaining values of θ, $f^{(m+1)}$ is finite, so we can again integrate by parts. Putting this together, we have

$$\begin{aligned}\hat{f}_k &= \lim_{\epsilon\to 0} \frac{1}{2\pi}\frac{1}{(\mathrm{i}k)^m}\left[\int_{-\pi}^{\theta_1^-} + \int_{\theta_1^+}^{\theta_2^-} + \cdots + \int_{\theta_r^+}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta} f^{(m)}(\theta)\,\mathrm{d}\theta\right] \\ &= \lim_{\epsilon\to 0} \frac{1}{2\pi}\frac{1}{(\mathrm{i}k)^{m+1}}\left[\sum_{s=1}^{r}\left(f^{(m)}(\theta_s^+) - f^{(m)}(\theta_s^-)\right)\mathrm{e}^{-\mathrm{i}k\theta_s} + \int \mathrm{e}^{-\mathrm{i}k\theta} f^{(m+1)}(\theta)\,\mathrm{d}\theta\right]\end{aligned} \tag{1.44}$$

where $\theta_s^{\pm} = \theta_s \pm \epsilon$ with $\epsilon > 0$, and where the dangerous neigbourhoods $\theta_s^- < \theta < \theta_s^+$ (for $s = 1, \ldots, r$) are excluded from the final integral. The size $f^{(m)}(\theta_s^+) - f^{(m)}(\theta_s^-)$ of the jumps appear as a boundary term in the final line, terminating the process of integration by parts. By our assumptions, these jumps are finite and $f^{(m+1)}$ is finite everywhere except at the discontinuities of $f^{(m)}$, so the quantity in square brackets is finite. Thus, if we first meet a discontinuity after differentiating a function m times then the Fourier coefficient $\hat{f}_k$ falls as $O(1/k^{m+1})$.

1.8 Pragmatism

To the horror of the analysts, in this course we'll henceforth mostly gloss over these subtle issues of convergence. The emphasis instead will be on seeing how we can actually *use* Fourier series to solve various linear differential equations, very much in the spirit of Fourier's own approach. In partial defense of this position, I say again that in physics, the reason we're interested in these linear equations in the first place is typically because they provide an approximation to a deeper, more complicated underlying system. If we're having to go very far down the Fourier series to get reasonable convergence to our original function, it's likely because some this function has some sharp feature that is in any case unphysical.

As an example, the 'plucked string' function

$$f(\theta) = \frac{\pi}{2} - |\theta| \tag{1.45}$$

for $\theta \in [-\pi, \pi)$ has Fourier coefficients

$$\hat{f}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\left(\frac{\pi}{2} - |\theta|\right)\mathrm{d}\theta = \begin{cases} 0 & \text{when } k \text{ is even and } n \neq 0 \\ \dfrac{2}{\pi n^2} & \text{when } n \text{ is odd} \end{cases} \tag{1.46}$$

(*check this as an exercise!*). These fall off like $\sim 1/n^2$, in agreement with the general results of section 1.7, since f is continuous but has a discontinuous first derivative, and $S_n f$ converges to $f(\theta)$. If we keep just the first three non-vanishing terms then

$$S_n f = \frac{4}{\pi}\left(\cos\theta + \frac{1}{9}\cos 3\theta + \frac{1}{25}\cos 5\theta + \cdots\right)$$

gives us an approximation to the plucked string that is good to 1 part in 10. But to obtain very high accuracy, we need to keep many millions of terms. The reason such high frequency terms are needed is because we need waves with very rapid oscillations to reproduce the sharp kink in the plucked string. But a real string will not kink precisely; the string has some thickness and any attempt to sharpen the kink too far will eventually cause the material to break. In other words, our naive attempt to model the behaviour of our real string by a linear equation breaks down, and the Fourier analysis based on the idea of an infinite dimensional vector space is no longer appropriate.

However, there is a cautionary conclusion to this tail. If you look at figures 1.5 and 1.6 you will observe a persistent overshoot – around 18% of the true value – of the Fourier series near the discontinuity. These little blips are known as the *Gibbs-Wilbraham phenomenon* and are explored in more detail in the problems. What is happening is that because the sawtooth and square wave functions have discontinuities, while the Fourier series do converge *pointwise*[8], they do *not* do so *uniformly*: convergence is slower near the discontinuities. This phenomenon was first discovered by Wilbraham in 1848, but was forgotten as a mathematical curiosity. In the 1890s Michelson constructed a device which could compute Fourier series automatically. When fed the square wave, he noticed the overshoot and assumed the machine was faulty until Gibbs provided the explanation. During WW2, the new British invention of radar was crucial in locating and intercepting Lüftwaffe bombers. The radar pulse sent out was a sawtooth, and engineers were puzzled by the little blips near the edge of each tooth.

1.9 Parseval's identity

It's often interesting to know the integral of the square of a periodic function (or mod-square for a $\mathbb{C}$-valued function). For example, if a point x on a vibrating string has velocity $v(x)$ then the total kinetic energy of the string is

$$E = T(v, v) = T \int_0^L v^2(x)\, \mathrm{d}x$$

where T is the string's tension and L is the length of the string.

Parseval's identity allows us to express the norm-squared of a function – (v, v) in the above example – in terms of its Fourier series. Let's derive it in the case that the Fourier series converges to the original function $f(\theta)$ everywhere except perhaps at finitely many points (for instance, if f may jumps). We first consider the norm-squared $(S_n f, S_n f)$ of

[8] Recall that a sequence $\{s_0(\theta), s_1(\theta), \ldots, s_n(\theta), \ldots\}$ converges to $s(\theta)$ pointwise if for each θ and for each $\epsilon > 0$ there exists an integer N_0 such that $|s_N(\theta) - s(\theta)| < \epsilon$. However, the integer N_0 *can depend on* θ *as well as on* ϵ. Only if N_0 is independent of θ is the convergence uniform. You'll see much more of this in your Analysis II course.

the partial Fourier sums. Inserting their definition into the norm we have

$$\begin{aligned}(S_nf, S_nf) &= \int_{-\pi}^{\pi} \left[\sum_{j=-n}^{n} \hat{f}_j^* \, \mathrm{e}^{-\mathrm{i}j\theta}\right] \left[\sum_{k=-n}^{n} \hat{f}_k \, \mathrm{e}^{\mathrm{i}k\theta}\right] \mathrm{d}\theta \\ &= \sum_{j,k=-n}^{n} \left[\hat{f}_j^* \, f_k \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(k-j)\theta} \, \mathrm{d}\theta\right] \\ &= 2\pi \sum_{j,k=-n}^{n} \hat{f}_j^* f_k \, \delta_{k,j} = 2\pi \sum_{k=-n}^{n} |\hat{f}_k|^2 \,.\end{aligned} \tag{1.47}$$

and similarly

$$\begin{aligned}(S_nf, f) &= \int_{-\pi}^{\pi} \left[\sum_{k=-n}^{n} \hat{f}_k^* \, \mathrm{e}^{-\mathrm{i}k\theta}\right] f(\theta) \, \mathrm{d}\theta \\ &= \sum_{k=-n}^{n} \hat{f}_k^* \left[\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}k\theta} \, f(\theta) \, \mathrm{d}\theta\right] = 2\pi \sum_{k=-n}^{n} |\hat{f}_k|^2 \,.\end{aligned} \tag{1.48}$$

This shows that $(S_nf, S_nf) = (S_nf, f) = (f, S_nf)$ and therefore

$$\begin{aligned}\int_{-\pi}^{\pi} |S_nf(\theta) - f(\theta)|^2 \, \mathrm{d}\theta &= (S_nf, S_nf) + (f, f) - (S_nf, f) - (f, S_nf) \\ &= (f, f) - 2\pi \sum_{k=-n}^{n} |\hat{f}_k|^2 \,.\end{aligned} \tag{1.49}$$

Now, because the series S_nf converges to f everywhere except perhaps at some isolated points, the left hand side vanishes as $n \to \infty$, and so the right hand side must vanish too. So we obtain

$$\int_{-\pi}^{\pi} |f(\theta)|^2 \, \mathrm{d}\theta = 2\pi \sum_{k\in\mathbb{Z}} |\hat{f}_k|^2 \,. \tag{1.50}$$

This useful result is known as Parseval's identity.

Parseval's identity may be interpreted as an infinite dimensional version of Pythagoras' theorem. The Fourier coefficients are the coefficients of f in the orthonormal basis $\{\mathrm{e}^{\mathrm{i}n\theta}/\sqrt{2\pi} \,:\, n \in \mathbb{Z}\}$, so Parseval's formula says that the norm-squared of f is the sum of the (mod-)squares of its coefficients in this basis. The factor of 2π can be traced to our conventions in the defintion (1.11) of the Fourier coefficients. On the other hand, when we first introduced the inner product as the integral $(f, g) = \int_\Omega f^* g \, \mathrm{d}\mu$ in (1.6) we treated the value a function takes at each point x in its domain Ω as its components, with the norm-squared $(f, f) = \int |f|^2 \, \mathrm{d}\mu$ telling us to add the mod-square of these 'components' over Ω. We see that this way of thinking about the length of f agrees with its Fourier decomposition. More formally, if we view the Fourier series as a map from a function to the sequence $\{\hat{f}_n\}$, then viewing this sequence as the coefficients of an infinite dimensional vector, Parseval's identity tells us that this map is an *isometry* – meaning it preserves lengths.

Parseval's identity is good for generating rather pretty identities involving infinite sums of the type beloved by Euler. For example, consider again the sawtooth function $f(\theta) = \theta$ for $\theta \in [-\pi, \pi)$, whose Fourier coefficients we found in (1.27) to be

$$\hat{f}_n = \frac{\mathrm{i}}{n}(-1)^n \text{ for } n \in \mathbb{Z}_{>0} \qquad \text{while} \qquad \hat{f}_0 = 0\,,$$

with $\hat{f}_{-n} = (\hat{f}_n)^*$ since the sawtooth is real. while $\hat{f}_0 = 0$. Then Parseval's identity becomes

$$\frac{2\pi^3}{3} = \int_{-\pi}^{\pi} \theta^2 \,\mathrm{d}\theta = 2\pi \sum_{n\in\mathbb{Z}} |\hat{f}_n|^2 = 4\pi \sum_{n=1}^{\infty} \frac{1}{n^2}\,,$$

or in other words

$$\frac{\pi^2}{6} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots\,. \tag{1.51}$$

As an exercise, you might also like to show that

$$\frac{\pi^4}{90} = \sum_{n=1}^{\infty} \frac{1}{n^4}$$

using Parseval's identity for the *integral* of the sawtooth function.