The good, the bad and the ugly of kernels: why the Dirichlet kernel is not a good kernel

Peter Haggstrom
www.gotohaggstrom.com
mathsatbondibeach@gmail.com

June 14, 2025

1 Background

Even though Dirichlet's name is often associated with number theory, he also did fundamental work on the convergence of Fourier series. Dirichlet's rigorous insights into the many subtle issues surrounding Fourier theory laid the foundation for what students study today. In the early part of the 19^{th} century Fourier advanced the stunning, even outrageous, idea that an arbitrary function defined on $(-\pi, \pi)$ could be represented by an infinite trigonometric series of sines and cosines thus:

$$f(x) = a_0 + \sum_{k=1}^{\infty}[a_k \cos(kx) + b_k \sin(kx)] \tag{1}$$

He did this in his theory of heat published in 1822 although in the mid 1750s Daniel Bernoulli had also conjectured that shape of a vibrating string could be represented by a trigonometric series. Fourier's insights predated electrical and magnetic theory by several years and yet one of the widest applications of Fourier theory is in electrical engineering.

The core of Fourier theory is to establish the conditions under which (1) is true. It is a complex and highly subtle story requiring some sophisticated analysis. Applied users of Fourier theory will rarely descend into the depths of analytical detail devoted to rigorous convergence proofs. Indeed, in some undergraduate courses on Fourier theory, the Sampling Theorem is proved on a "faith" basis using distribution theory.

In what follows I have used Professor Elias Stein and Rami Shakarchi's book [Stein-Shakarchi] as a foundation for fleshing out the motivations, properties and uses of "good" kernels. The reason for this is simple - Elias Stein is the best communicator of the whole edifice of this part of analysis. I have left no stone unturned in terms of detail in the

proofs of various properties and while some students who are sufficienlty "in the zone" can gloss over the detail, others may well benefit from it. An example is the nuts and bolts of the basic Tauberian style proofs which are often ignored in undergraduate analysis courses.

2 Building blocks

Using properties of the sine and cosine functions such as (where k and m are integers):

$\int_{-\pi}^{\pi} \sin(kx)\sin(mx)dx = \begin{cases} 0 & \text{if } k \neq m \\ \\ \pi & \text{if } k = m = 0 \end{cases}$

$\int_{-\pi}^{\pi} \sin(kx)\cos(mx)dx = 0$

$\int_{-\pi}^{\pi} \cos(kx)\cos(mx)dx = \begin{cases} 0 & \text{if } k \neq m \\ 2\pi & \text{if } k = m = 0 \\ \pi & \text{if } k = m \neq 0 \end{cases}$

the coefficients of the Fourier series expansions could be recovered as:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx \tag{2}$$

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)dx \qquad k \geq 1 \tag{3}$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)dx \qquad k \geq 1 \tag{4}$$

If you have forgotten how to derive the basic sin and cosine formulas set out above just recall that:

$\int_{-\pi}^{\pi} \sin(kx)dx = \int_{-\pi}^{\pi} \cos(kx)dx = 0$ for $k = 1, 2, 3 \ldots$

You also need:

$\cos(kx)\ \cos(mx) = \frac{1}{2}\cos((k-m)x) + \cos((k+m)x)$;
$\sin(kx)\ \sin(mx) = \frac{1}{2}\cos((k-m)x) - \cos((k+m)x)$; and
$\sin(kx)\ \cos(mx) = \frac{1}{2}\sin((k-m)x) + \sin((k+m)x)$

The partial sums of the Fourier series of f can be expressed as follows:

$$f_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx + \sum_{k=1}^{n}\left[\left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)dt\right)\cos(kx) + \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)dt\right)\sin(kx)\right] \quad (5)$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx + \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\sum_{k=1}^{n}\cos(kt)\cos(kx) + \sin(kt)\sin(kx)\right)f(t)dt \quad (6)$$

The exchange of summation and integration is justified because the sums are finite. Hence we have:

$$f_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}\left(\frac{1}{2} + \sum_{k=1}^{n}\cos(k(t-x))\right)f(t)dt \quad (7)$$

The simplification of $\sum_{k=1}^{n}\cos(k(t-x)$ leads to the Dirichlet kernel, thus we need to find a nice closed expression for $\frac{1}{2} + \sum_{k=1}^{n}\cos(ku)$ and what better way to search for a closed form than to simply experiment with a couple of low order cases. Thus for n=1 we have to find a nice expression for $\frac{1}{2} + \cos u$. We know that $\cos u = \sin(u + \frac{\pi}{2})$ so in analogy with that why not investigate $\sin(u + \frac{u}{2})$ and see what emerges?

$$\begin{aligned}
\sin(u + \frac{u}{2}) &= \sin u\,\cos(\frac{u}{2}) + \sin(\frac{u}{2})\,\cos u \\
&= 2\sin(\frac{u}{2})\cos^2(\frac{u}{2}) + \cos u\,\sin(\frac{u}{2}) \\
&= \sin(\frac{u}{2})(2\cos^2(\frac{u}{2}) + \cos u) \\
&= \sin(\frac{u}{2})(\cos u + 1 + \cos u) \\
&= \sin(\frac{u}{2})(2\cos u + 1)
\end{aligned} \quad (8)$$

Hence we have that:

$$\frac{1}{2} + \cos u = \frac{\sin(u + \frac{u}{2})}{2\sin(\frac{u}{2})} \quad (9)$$

With this little building block we gamely extrapolate as follows:

$$\frac{1}{2} + \cos\, u + \cos\, 2u + \cdots + \cos\, nu = \frac{\sin((2n+1)\frac{u}{2})}{2\sin(\frac{u}{2})} \tag{10}$$

To prove that the formula is valid for all n we need to do is apply a standard induction to it. We have already established the base case of $n = 1$ since $\frac{1}{2} + \cos\, u = \frac{\sin(u+\frac{u}{2})}{2\sin(\frac{u}{2})} = \frac{\sin(3\frac{u}{2})}{2\sin(\frac{u}{2})}$. As usual we assume the formula holds for any n so that:

$$\frac{1}{2}+\cos\, u+\cos\, 2u+\cdots+\cos\, nu+\cos(n+1)u = \frac{\sin((2n+1)\frac{u}{2})}{2\sin(\frac{u}{2})}+\cos((n+1)u) = \frac{TOP}{2\sin(\frac{u}{2})} \tag{11}$$

$$\begin{aligned}
TOP &= \sin(nu + \frac{u}{2}) + 2\sin(\frac{u}{2})\, \cos((nu + \frac{u}{2}) + \frac{u}{2}) \\
&= \sin(\, nu)\, \cos(\frac{u}{2})+\cos(\, nu)\, \sin(\frac{u}{2})+2\sin(\frac{u}{2})\, \cos(nu+\frac{u}{2})\, \cos(\frac{u}{2})-2\sin(\frac{u}{2})\, \sin(nu+\frac{u}{2})\, \sin(\frac{u}{2}) \\
&= \sin(\, nu)\, \cos(\frac{u}{2})+\cos(\, nu)\, \sin(\frac{u}{2})+2\sin(\frac{u}{2})\, \cos(\frac{u}{2})\cos(\, nu)\, \cos(\frac{u}{2})-2\sin(\frac{u}{2})\, \cos(\frac{u}{2})\, \sin(\, nu)\, \sin(\frac{u}{2}) \\
&\quad - 2\sin^2(\frac{u}{2})\, s \in (\, nu)\, \cos(\frac{u}{2}) - 2\sin^2(\frac{u}{2})\, \cos(\, nu)\, \sin(\frac{u}{2}) \\
&= (1 - 2\sin^2(\frac{u}{2}))\, \sin(\, nu)\, \cos(\frac{u}{2}) + (1 - 2\sin^2(\frac{u}{2}))\, \cos(\, nu)\, \sin(\frac{u}{2}) \\
&\quad + \sin\, u\, \cos(\, nu)\, \cos(\frac{u}{2}) - \sin\, u\, \sin(\, nu)\, \sin(\frac{u}{2}) \\
&= \cos\, u\, \sin(\, nu)\, \cos(\frac{u}{2})+\cos\, u\, \cos(\, nu)\, \sin(\frac{u}{2})+\sin\, u\, \cos(\, nu)\, \cos(\frac{u}{2})-\sin\, u\, \sin(\, nu)\, \sin(\frac{u}{2}) \\
&= \cos\, u\, \sin(nu + \frac{u}{2}) + \sin\, u\, \cos(nu + \frac{u}{2}) = \sin(u + nu + \frac{u}{2}) \\
&= \sin((2n+3)\frac{u}{2})
\end{aligned} \tag{12}$$

Hence we do get $\frac{1}{2} + \cos\, u + \cos\, 2u + \cdots + \cos\, nu + \cos\,(n+1)u = \frac{\sin((2n+3)\frac{u}{2})}{2\sin(\frac{u}{2})}$.

Thus the formula is true for n+1.

If you find that derivation tedious you could start with:

$$\cos ku \sin(\frac{u}{2}) = \frac{1}{2}\{\sin((k+\frac{1}{2})u) - \sin((k-\frac{1}{2})u)\} \tag{13}$$

Then you get:

$$\sum_{k=1}^{n} \sin(\frac{u}{2}) \cos ku = \frac{1}{2}\sum_{k=1}^{n}\{\sin((k+\frac{1}{2})u) - \sin((k-\frac{1}{2})u)\}$$
$$= \frac{1}{2}\left\{(\sin(\frac{3u}{2}) - \sin(\frac{u}{2})) + (\sin(\frac{5u}{2}) - \sin(\frac{3u}{2})) + \cdots + (\sin((n+\frac{1}{2})u) - \sin((n-\frac{1}{2})u))\right\}$$
$$= \frac{1}{2}\left\{ -\sin(\frac{u}{2})) + (\sin((n+\frac{1}{2})u)\right\} \tag{14}$$

Hence on dividing the LHS of (14) by $\sin(\frac{u}{2})$ we have that:
$\cos u + \cos 2u + \cdots + \cos nu = \frac{1}{2}\left\{ -1 + \frac{\sin((n+\frac{1}{2})u)}{\sin(\frac{u}{2}))}\right\}$

Finally we have that $\frac{1}{2} + \cos u + \cos 2u + \cdots + \cos nu = \frac{\sin((n+\frac{1}{2})u)}{2\sin(\frac{u}{2}))}$

So going back to (7) we have:

$$f_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{\sin(\frac{(2n+1)(t-x)}{2})}{2\sin(\frac{t-x}{2})} f(t)dt \tag{15}$$

$$f_n(x) = \frac{1}{\pi}\int_{-\pi+x}^{\pi+x} \frac{\sin(\frac{(2n+1)(t-x)}{2})}{2\sin(\frac{t-x}{2})} f(t)dt \tag{16}$$

This works because $f(t+2\pi) = f(t)$ ie f is 2π periodic, as are sin and cos. The product of two 2π periodic functions is also 2π periodic since $f(x+2\pi)g(x+2\pi) = f(x)g(x)$.

Comment on integrals of 2π periodic functions

A point on a circle can be represented by $e^{i\theta}$ and is unique up to integer multiples of 2π. If F is a function "on the circle" then for each real θ we define $f(\theta) = F(e^{i\theta})$. Thus f is 2π periodic since $f(\theta) = f(\theta + 2\pi)$. All the qualities of f such as continuity, integrability and differentiablity apply on any interval of 2π.

There are some fundamental manipulations you can do with 2π periodic functions. If we assume that f is 2π periodic and is integrable on any finite interval [a,b] where a and

b are real, we have:

$$\int_a^b f(x)dx = \int_{a+2\pi}^{b+2\pi} f(x)dx = \int_{a-2\pi}^{b-2\pi} f(x)dx \tag{17}$$

Noting that $f(x) = f(x \pm 2\pi)$ because of the periodicity and making the substitution $u = x \pm 2\pi$ we see that (using $u = x + 2\pi$ as our substitution to illustrate):

$$\int_a^b f(x)dx \quad = \quad \int_a^b f(x \quad + \quad 2\pi)dx \quad = \quad \int_{a+2\pi}^{b+2\pi} f(u)du \tag{18}$$

The substitution $u = x - 2\pi$ leads to $\int_{a-2\pi}^{b-2\pi} f(x)dx$

The following relationships also prove useful:

$$\int_{-\pi}^{\pi} f(x+a)dx = \int_{-\pi}^{\pi} f(x)dx = \int_{-\pi+a}^{\pi+a} f(x)dx \tag{19}$$

The substitution $u = x+a$ gives $\int_{-\pi+a}^{\pi+a} f(x)dx$ while $\int_{-\pi}^{\pi} f(x+a)dx = \int_{-\pi}^{\pi} f(x+a)d(x+a)$ which is just $\int_{-\pi}^{\pi} f(z)dz$ since the variable integration z simply runs from $-\pi$ to π.

In order to evaluate (16) we make the following substitutions which apply when we split the integral in two parts: $t = x - 2u$ for $t \in [-\pi + x, x]$ and $t = x + 2u$ for $t \in [x, \pi + x]$

$$\begin{aligned} f_n(x) &= \frac{1}{\pi}\int_{-\pi+x}^{x} \frac{\sin(\frac{(2n+1)(t-x)}{2})}{2\sin(\frac{t-x}{2})} f(t)dt + \frac{1}{\pi}\int_{x}^{x+\pi} \frac{\sin(\frac{(2n+1)(t-x)}{2})}{2\sin(\frac{t-x}{2})} f(t)dt \\ &= \frac{1}{\pi}\int_{\frac{\pi}{2}}^{0} \frac{\sin((2n+1)(-u))}{2\sin(-u)} f(x-2u)(-2\,du) + \frac{1}{\pi}\int_{0}^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{2\sin\,u} f(x+2u)(2\,du) \\ &= \frac{1}{\pi}\int_{0}^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin\,u} f(x-2u)du + \frac{1}{\pi}\int_{0}^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin\,u} f(x+2u)du \end{aligned} \tag{20}$$

So finally we write $f_n(x)$ in terms of the Dirichlet kernel which is defined as:

$$\boxed{D_n(u) = \frac{\sin((2n+1)u)}{\sin\,u}} \tag{21}$$

Note that the Dirichlet kernel can also be defined as: $D_n(u) = \frac{\sin((n+\frac{1}{2})\frac{u}{2})}{\sin(\frac{u}{2})}$ and with a normalising factor such as: $D_n(u) = \frac{1}{2\pi}\frac{\sin((n+\frac{1}{2})\frac{u}{2})}{\sin(\frac{u}{2})}$

Thus (20) becomes:

$$f_n(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)f(x-2u)\,du + \frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)f(x+2u)\,du \tag{22}$$

That $\frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)du = \frac{\pi}{2}$ can be shown by using (10) and doing the straightforward integration. Thus from (10) we get:

$$1 + 2\,\cos\,u + 2\,\cos\,2u + \cdots + 2\,\cos\,nu = \frac{\sin((2n+1)\frac{u}{2})}{\sin(\frac{u}{2})} \tag{23}$$

Now letting $u = 2\,v$:

$$1 + 2\,\cos\,2v + 2\,\cos\,4v + \cdots + 2\,\cos\,2nv = \frac{\sin((2n+1)v}{\sin\,v} \tag{24}$$

Hence the relevant integral becomes:

$$\int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)v}{\sin\,v}dv = \int_0^{\frac{\pi}{2}} dv + \int_0^{\frac{\pi}{2}} 2\,\cos(2v)dv + \cdots + \int_0^{\frac{\pi}{2}} 2\,\cos(2nv)dv = \frac{\pi}{2} \tag{25}$$

Note that $\int_0^{\frac{\pi}{2}} 2\,\cos(2nv)dv = 0$ for $n \geq 1$ since $\cos(2nv)$ integrates to $\frac{1}{2n}\sin(2nv)$ which is zero at $v = \frac{\pi}{2}$ and 0. Trying to integrate $\int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)v}{\sin\,v}dv$ "cold" without the simple form on the RHS of (25) would end in despair.

Going back to (22) we can express it in the form used by Dirichlet:

$$f_n(x) = f_n^-(x) + f_n^+(x) \tag{26}$$

where $f_n^-(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)f(x-2u)\,du$ and $f_n^+(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)f(x+2u)\,du$

The aim is to prove that $f_n(x) \to f(x)$ as $n \to \infty$.

Observations about $f_n^+(x)$

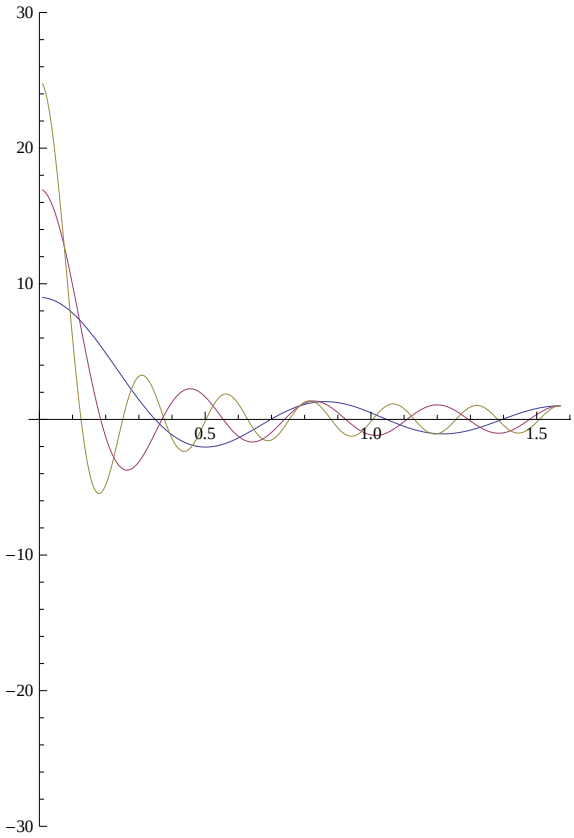
Looking at the definition of $f_n^+(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} D_n(u)f(x+2u)\,du$ it is clear that the value at x is actually independent of f(x) since the integral involves the variable u which runs over the interval $x \leq u \leq x+\pi$.

Some other properties of the Dirichlet kernel are as follows:

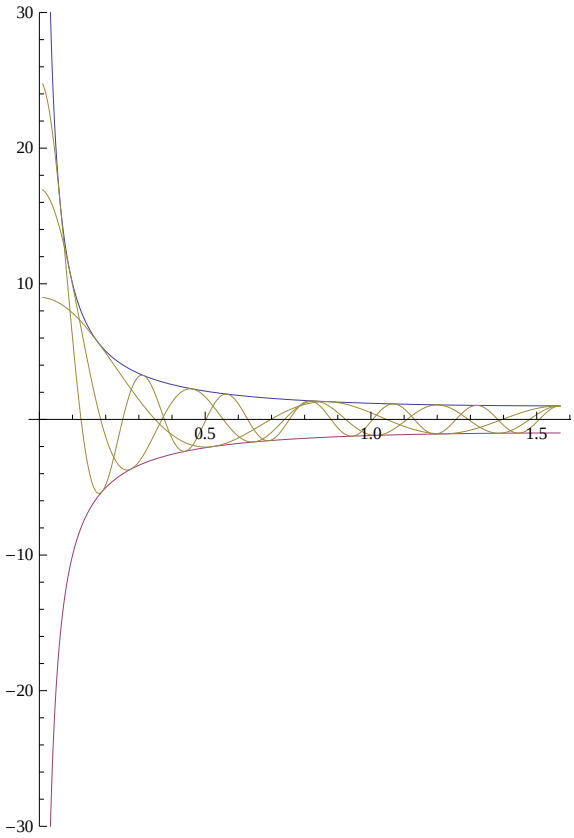
$$D_n(0) = D_n(1) = 2n+1 \tag{27}$$

From (24) we see that $D_n(0) = 1+2n$. Since $D_n(t) = \sum_{k=-n}^{n} e^{2\pi ikt} = 1+\sum_{k=1}^{n}(e^{2\pi ikt} + e^{-2\pi ikt}) = 1+2\sum_{k=1}^{n}\cos(2\pi kt)$ it follows that $D_n(1) = 1+2n$.

The effect of the Dirichlet kernel is that it isolates behaviour around zero. Because $D_n(u) = 0$ for the first time when $u = \frac{\pi}{2n+1}$ and the peak at $u = 0$ is $2n+1$ most of the area under the graph is under the first spike. Thus $\int_0^{\frac{\pi}{2n+1}} D_n(u)\,f(x+2u)\,du$ reprsents most of the area. The following graph for $n = 4, 8, 12$ shows how the graph of $D_n(u)$ evolves. It appears that as $u > \frac{\pi}{2n+1}$ the oscillations are damped down into an envelope with what appears to be a fairly constant amplitude.



The area under the first spike is $\int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du = 1.85159 + 3.1225 \times 10^{-17} i$ according to *Mathematica 8* with $n = 100$. The imaginary term arises from the algorithm used for the numerical integration which involves complex exponential terms. *Mathematica* gives the value of $\int_{\frac{\pi}{2n+1}}^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin u}\,du = -0.28115 - 3.1225 \times 10^{-17} i$ so that the sum of the two integrals is $\frac{\pi}{2}$ as derived analytically. Note that analytically we can say that $\int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du < \int_0^{\frac{\pi}{2n+1}} (2n+1)\,du = \pi$ and that $\int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du > \frac{1}{2}\frac{\pi}{2n+1}(2n+1) = \frac{\pi}{2}$ by taking the area of the inscribed right angle triangle with base $\frac{\pi}{2n+1}$ and height $2n+1$. Thus $\frac{\pi}{2} < \int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du < \pi$. Note that $\frac{\pm 1}{\sin u}$ forms two envelopes for the kernel as shown in the graph below:



If $0 \leq x < a \leq \frac{\pi}{2}$ then:

$$\Big|\int_x^a \frac{\sin((2n+1)u)}{\sin u}\,du\Big| \leq \int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du \tag{28}$$

We already know that $\frac{\pi}{2} < \int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u}\,du < \pi$ and that $\int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin u}\,du = \frac{\pi}{2}$. This means that $\int_{\frac{\pi}{2n+1}}^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin u}\,du < 0$ and so (28) holds. From a purely visual inspection it looks like the areas of the waves decrease in value and alternate in sign. In a general context we can see that where h(u) is a monotonically decreasing function, $\int_{2n\pi}^{(2n+1)\pi} h(u)\sin u\,du = \int_0^{\pi} h(u+2n\pi)\sin u\,du$ (just make the substitution $x = u - 2n\pi$ and also note the obvious 2π periodicity of $\sin u$). By observing that $\int_0^{\pi} \sin u\,du = 2$ we have that the integral lies between $2h(2n\pi)$ and $2h((2n+1)\pi)$ and because of the monoticity of h the integral must approach zero as $n \to \infty$. Also the alternating negativity of the signs can be seen from the fact that $\int_{(2n+1)\pi}^{(2n+2)\pi} h(u)\,\sin u\,du = -\int_0^{\pi} h(u+(2n+1)\pi)\,\sin u\,du$.

Because of the way $D_n(u)$ decays, when n is large the value of $\int_0^{\frac{\pi}{2}} D_n(u) f(x+2u)\,du$

will be dominated by f(x+2u) for $0 < u < \frac{\pi}{2n+1}$ and, as n is large, if f is continuous the value of f(x+2u) over this small interval will be pretty constant (just recall that continuity means that within a neighbourhood of u the values of f(x) will be arbitrarily close). This means that $\int_{\frac{\pi}{2n+1}}^{\frac{\pi}{2}} D_n(u)\, f(x+2u)\, du$ ought to be small. If we take the midpoint of the interval $0 < u < \frac{\pi}{2n+1}$ ie $u = \frac{\pi}{2(2n+1)}$ when n is large, the continuity of f ensures that $f(x+2u) = f(x+\frac{\pi}{2n+1})$ will be close to other values in this interval. Heuristically then the value of $\int_0^{\frac{\pi}{2}} D_n(u) f(x+2u)\, du$ could be approximated by $f(x+\frac{\pi}{2n+1}) \times$ area under first spike $\approx f(x+\frac{\pi}{2n+1})\,\frac{\pi}{2}$. Crudely then:

$$f_n^+(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{\sin((2n+1)u)}{\sin u}\, f(x+2u)\, du \approx \frac{1}{\pi}\int_0^{\frac{\pi}{2n+1}} \frac{\sin((2n+1)u)}{\sin u} f(x+2u)\, du$$
$$\approx \frac{1}{\pi}\, f(x+\frac{\pi}{2n+1})\,\frac{\pi}{2} = \frac{1}{2}\, f(x+\frac{\pi}{2n+1}) \quad (29)$$

Continuity is critical in the above analysis and if f is continuous at x from the right $f_n^+(x) \to \frac{1}{2}f(x)$. By identical reasoning, if f is continuous at x from the left $f_n^-(x) \to \frac{1}{2}f(x)$. Dirichlet's suggestible notation for these processes is $f(x+0) = \lim_{u\to x^+} f(u)$ and $f(x-0) = \lim_{u\to x^-} f(u)$

The details of showing the convergence of $f_n(x)$ to f(x) are relatively detailed and you can do no better for a straightforward yet rigorous explanation than by reading chapter 6 of [Bressoud].

One result that is fundamental to the original work on Fourier convergence is Riemann's Lemma which is as follows:

If g(u) is continuous on [a,b] where $0 < a < b \leq \frac{\pi}{2}$ then:

$$\lim_{M\to\infty} \int_a^b \sin(Mu)\, g(u)\, du = 0 \quad (30)$$

This is used in proving that $\lim_{n\to\infty} \int_a^{\pi/2} \frac{\sin((2n+1)u)}{\sin u}\, f(x+2u)\, du = 0$ where $0 < a < \frac{\pi}{2}$.

The proof involves showing that for any $\epsilon > 0$, $\exists M$ such that if $N \geq M$ then $|\int_a^b \sin(Nu)\, g(u)\, du| < \epsilon$. The usual approach is to perform a uniform partition of m equal subintervals of [a,b]

as follows:
$a = u_0 < u_1 < \cdots < u_n = b$ so that $u_k - u_{k-1} = \frac{b-a}{m}$

Because g is continuous on [a,b] it is uniformly continuous on [a,b] as well as any of its closed subintervals, so we can choose an m such that $|u - v| \leq \frac{b-a}{m} \Rightarrow |g(u) - g(v)| < \frac{\epsilon}{2(b-a)}$. This uniform continuity requirement is critical to estimating the size of the integral. Thus we have:

$$\begin{aligned}
\left| \int_a^b \sin(Mu)\, g(u)\, du \right| &= \left| \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \sin(Mu)\, g(u)\, du \right| = \left| \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \sin(Mu)\, [g(u_{k-1}) + g(u) - g(u_{k-1})]\, du \right| \\
&\leq \left| \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \sin(Mu)\, g(u_{k-1})\, du \right| + \left| \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \sin(Mu)\, [g(u) - g(u_{k-1})]\, du \right| \\
&\leq \sum_{k=1}^{m} \left| \int_{u_{k-1}}^{u_k} \sin(Mu)\, g(u_{k-1})\, du \right| + \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \left| \sin(Mu)\, [g(u) - g(u_{k-1})]\, du \right| \qquad (31)
\end{aligned}$$

Now continuity of g on [a,b] means that it is bounded ie $\exists B$ such that $|g(u)| \leq B$, $\forall u \in [a, b]$. Thus:

$$\begin{aligned}
\left| \int_a^b \sin(Mu)\, g(u)\, du \right| &\leq B \sum_{k=1}^{m} \left| \int_{u_{k-1}}^{u_k} \sin(Mu)\, du \right| + \sum_{k=1}^{m} \int_{u_{k-1}}^{u_k} \frac{\epsilon}{2(b-a)}\, du \\
&\text{noting the use of } |\sin(Mu)| \leq 1 \text{ in the second integral} \\
&= B \sum_{k=1}^{m} \frac{|-\cos(Mu_k) + \cos(Mu_{k-1})|}{M} + \sum_{k=1}^{m} \frac{\epsilon(u_k - u_{k-1})}{2(b-a)} \\
&\leq \frac{2Bm}{M} + \frac{\epsilon(b-a)}{2(b-a)} = \frac{2Bm}{M} + \frac{\epsilon}{2} \qquad (32)
\end{aligned}$$

Now w can choose M as large as we like to make $\frac{2Bm}{M} < \frac{\epsilon}{2}$ and so make the absolute value of the integral less than any arbitrary ϵ. Note here that m is a function of the choice of ϵ and B is simply a fixed global property for g on [a,b], but M is without constraint - we can make it as large as we like.

3 A more general discussion of kernels

With that background we can now move to a more general discussion of kernels and their properties and so see what makes a "good" kernel and why the Dirichlet kernel fails to be a "good" kernel. The conept of convolution is pivotal in what follows.

The convolution (this concept is explained in more detail later on) of two 2π periodic integrable functions f and g is written as:

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)g(x-y)dy \tag{33}$$

Because both f and g are 2π periodic if we let $u = x - y$ in (2) where x is treated as a constant, we get:

$$\begin{aligned}(f * g)(x) &= \frac{-1}{2\pi} \int_{x+\pi}^{x-\pi} f(x-u)g(u)du \\ &= \frac{1}{2\pi} \int_{x-\pi}^{x+\pi} f(x-u)g(u)du \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-u)g(u)du\end{aligned} \tag{34}$$

The last line is justified by the 2π periodicity of both f and g.

In this more general context we will see that if we have a family of "good" kernels $\{K_n\}_{n=1}^{\infty}$ and a function f which is integrable on the circle it can be shown that:

$$\lim_{n\to\infty} (f * K_n)(x) = f(x) \tag{35}$$

whenever f is continuous at x. If f is continuous everywhere the limit is uniform. There are several important applications of this principle but we have to develop some further concepts before delving into them. It is not immediately obvious why (35) would allow you to do anything useful since all it seems to do is say that if you convolve a function at a point with a special family of kernels and take the limit, you get the value of the function at the point. To get some understanding of the motivation for this definition you need to go back to some fundamental physical problems.

The classical problem of the solution of the steady state heat equation:

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0 \tag{36}$$

on the unit disc with boundary condition $u = f$ on the circle. The solution you get has the form:

$$u(r,\theta) = \sum_{n=-\infty}^{\infty} a_n r^{|n|} e^{in\theta} \tag{37}$$

If you cannot recall how to derive (37) all you need to do is to rewrite (36) as:

$$r^2\frac{\partial^2 u}{\partial r^2} + r\frac{\partial u}{\partial r} = -\frac{\partial^2 u}{\partial \theta^2} \tag{38}$$

Next you use the technique of separating variables which makes sense where you have essentially independent radial and angular coordinates. Thus you assume that $u(r,\theta) = f(r)g(\theta)$ and perform the relevant differentiation in (38) to get:

$$\frac{r^2 f''(r) + r f'}{f(r)} = -\frac{g''(\theta)}{g(\theta)} \tag{39}$$

Because the LHS of (39) is independent of θ but equals the RHS which is independent of r, they both must equal some constant. Owing to the fact that $g(\theta)$ is 2π periodic and we need bounded solutions, the constant $\lambda \geq 0$ and can be written as $\lambda = n^2$ where n is an integer. Thus we ultimately get $g(\theta) = Ae^{in\theta} + Be^{-in\theta}$ and $f(r) = r^{|n|}$ so that:

$$u_n(r,\theta) = r^{|n|} e^{in\theta} \tag{40}$$

The principle of superposition then leads to the general solution:

$$u(r,\theta) = \sum_{n=-\infty}^{n=\infty} a_n r^{|n|} e^{in\theta} \tag{41}$$

Here a_n is the n^{th} Fourier coefficient of f. It can be shown that if we take $u(r,\theta)$ as the convolution with the Poisson kernel some nice things happen. The Poisson kernel has this form for $0 \leq r < 1$:

$$P_r(\theta) = \frac{1-r^2}{1-2r\,cos\,\theta + r^2} \tag{42}$$

The hoped for convolution is this:

$$u(r,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)P_r(\theta-\phi)d\phi \tag{43}$$

The details of how you get to (35) from (21) will be spelt out below.

The limit in (35) is an important one and its proof is a straightforward application of the usual "(ϵ,δ)" approach. The proof goes like this. We take $\epsilon > 0$ and because f is continuous at x we can find a δ such that $|y| < \delta$ implies $|f(x-y) - f(x)| < \epsilon$. By assumption the K_n are good kernels (see (65) - (67) for the characteristics of a goof kernel) one of which properties is that $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(x)dx = 1$ ie the kernel is normalized to 1. We need to show that $\lim_{n\to\infty}(f * K_n)(x) = f(x)$ so we start with:

$$(f*K_n)(x) - f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)K_n(y)dy - f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)[f(x-y)-f(x)]dy \tag{44}$$

Therefore taking absolute values:

$$|(f * K_n)(x) - f(x)| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(y)[f(x-y) - f(x)]dy \right|$$

$$= \left| \frac{1}{2\pi} \int_{-\delta}^{\delta} K_n(y)[f(x-y)-f(x)]\,dy + \frac{1}{2\pi} \int_{-\pi}^{-\delta} K_n(y)[f(x-y)-f(x)]dy + \frac{1}{2\pi} \int_{\delta}^{\pi} K_n(y)[f(x-y)-f(x)]dy \right|$$

$$\leq \frac{1}{2\pi} \int_{-\delta}^{\delta} \Big| K_n(y)[f(x-y) - f(x)] \Big| \, dy + \frac{1}{2\pi} \int_{-\pi}^{-\delta} \Big| K_n(y)[f(x-y) - f(x)] \Big| \, dy +$$

$$\frac{1}{2\pi} \int_{\delta}^{\pi} \Big| K_n(y)[f(x-y) - f(x)] \Big| \, dy = L1 + L2 + L3 \quad (45)$$

To estimate L1 we need a property of good kernels set out in (66), namely, that $\exists M > 0$ such that $\forall n \geq 1$, $\int_{-\pi}^{\pi} |K_n(y)\,dy| \leq M$. Therefore:

$$L1 = \frac{1}{2\pi} \int_{-\delta}^{\delta} \Big| K_n(y)[f(x-y) - f(x)] \Big| \, dy \leq \frac{1}{2\pi} \int_{-\delta}^{\delta} \Big| K_n(y) \Big| \, \Big| [f(x-y) - f(x)] \Big| \, dy \leq \frac{\epsilon M}{2\pi} \quad (46)$$

Since f is continuous on $[-\pi, \pi]$ (and hence any closed sub-interval) it is bounded by some $B > 0$, ie $|f(x)| \leq B$, $\forall x \in [-\pi, \pi]$. We also need the third property of good kernels set out in (67), namely,that for every $\delta > 0$, $\int_{\delta \leq |y| \leq \pi} \big| K_n(y) \big| \, dy \to 0$ as $n \to \infty$, so that $\exists N_1$ such that $\int_{\delta \leq |y| \leq \pi} \big| K_n(y) \big| \, dy < \epsilon$, for all $n > N_1$
Thus:

$$L2{+}L3 = \frac{1}{2\pi} \int_{\delta \leq |y| \leq \pi} \big| K_n(y) \big| \, \big| [f(x{-}y){-}f(x)] \big| \, dy \leq \frac{1}{2\pi} \int_{\delta \leq |y| \leq \pi} \big| K_n(y) \big| \, (\big| f(x{-}y) \big| {+} \big| f(x) \big|) \, dy$$

$$\leq \frac{2B\epsilon}{2\pi} \quad (47)$$

Putting it all together we have that $|(f * K_n)(x) - f(x)| \leq \frac{M\epsilon}{2\pi} + \frac{2B\epsilon}{2\pi} < C\epsilon$ for some constant $C > 0$. So $(f * K_n)(x) \to f(x)$. If f is continuous everywhere it is uniformly continuous and δ can be chosen independently of ϵ.

4 The relationship between Abel means and convolutions

Recall from (35) how a convolution of a good kernel with a function gives the value of the function at a point. The fundamental fact is that Abel means can be represented as convolutions. Equally fundamental, the partial sums of the Fourier series of f convolved

with f gives the Dirichlet kernel (this is proved below). Once we have demonstrated the convergence properties of Abel means (and this requires some relatively subtle analysis) and then how they can be represented as convolutions, we effectively arrive at a solution to the steady state heat equation which has the right properties. It also becomes clearer that (35) is a non-trivial relationship. Welcome to hard core Fourier theory.

Because Fourier series can fail to converge at individual points and may even fail to converge at points of continuity, 19^{th} century mathematicians looked at the convergence properties of various types of means. G H Hardy's book "Divergent Series", AMS Chelsea Publishing 1991 [HardyDS] is all about investigating different types of means that yield consistent forms of convergence. By redefining convergence (a bit like redefining lateness so the trains run "on time"!) it is possible to get meaningful properties. Hence the relevance of *Cesàro* summability and Abel means.

First a definition:

A series of complex numbers $\sum_{k=0}^{\infty} c_k$ is said to be Abel summable to s if for every $0 \leq r < 1$ the series:

$$A(r) = \sum_{k=0}^{\infty} c_k \, r^k \tag{48}$$

converges and $\lim_{r \to 1} A(r) = s$

If a series converges to s then it is Abel summable to s. Thus ordinary convergence implies Abel summability. This and several other important propositions are exercises in Chapter 2 of [SteinShakarchi]. I have systematically gone through those exercises in the Appendix. They all involve fundamental techniques in analysis so it is worth following them through in detail.

It is shown in Chapter 6 of [SteinShakarchi] that:

$$u(r, \theta) = (f * P_r)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) P_r(\theta - \phi) d\phi \tag{49}$$

has the following properties:

(i) u has two continuous derivatives in the unit disc and satisfies $\Delta u = 0$.

(ii) If θ is any point of continuity of f, then

$$\lim_{n \to \infty} u(r, \theta) = f(\theta) \tag{50}$$

If f is continuous everywhere then the limit is uniform.

(iii) If f is continuous then $u(r,\theta)$ is the unique solution to the steady-state heat equation in the disc which satisfies (i) and (ii).

Thus the family of good kernels convolved with the function f acts like an identity in the limit. The process of convolution is developed in more detail below. We can show that the partial sums of the Fourier series can be represented as a convolution of the function f and the n^{th} Dirichlet kernel:

$$\begin{aligned} S_N(f)(x) &= \sum_{n=-N}^{N} \hat{f}(n)e^{inx} \\ &= \sum_{n=-N}^{N} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)e^{-iny}dy \right) e^{inx} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) \left(\sum_{n=-N}^{N} e^{in(x-y)} \right) dy \\ &= (f * D_N)(x) \end{aligned} \tag{51}$$

Note that the exchange of summation and integration above is legitimate because we are dealing with a finite sum.

Thus the sum is represented by the convolution of f with the Dirichlet kernel defined below.

"Good" kernels can be used to recover a given function by the use of convolutions. An extremely important result in Fourier Theory is the fact that the Fourier transform of a convolution is the product of the respective Fourier transforms ie:

$$\widehat{f * g}(n) = \hat{f}(n) * \hat{g}(n) \tag{52}$$

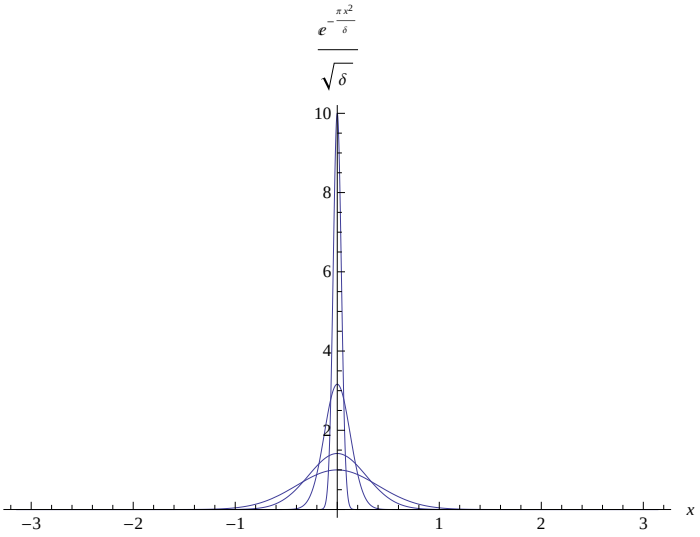
D_N is the N^{th} Dirichlet kernel given by $D_N(x) = \sum_{n=-N}^{N} e^{inx}$. If we let $\omega = e^{ix}$ then $D_N = \sum_{n=0}^{N} \omega^n + \sum_{n=-N}^{-1} \omega^n$ which are just two geometric series. The sums are respectively equal to $\frac{1-\omega^{N+1}}{1-\omega}$ and $\frac{\omega^{-N}-1}{1-\omega}$. This sum gives rise to the closed form of the Dirichlet kernel ie:

$$D_N(x) = \frac{1-\omega^{N+1}}{1-\omega} + \frac{\omega^{-N}-1}{1-\omega} = \frac{\omega^{-N}-\omega^{N+1}}{1-\omega} = \frac{\omega^{-(N+\frac{1}{2})}-\omega^{N+\frac{1}{2}}}{\omega^{\frac{-1}{2}}-\omega^{\frac{1}{2}}} = \frac{\sin(N+\frac{1}{2})x}{\sin\frac{1}{2}x} \tag{53}$$

Note that in (21) $D_N(u) = \frac{\sin((2N+1)u)}{\sin u}$ where $u = \frac{x}{2}$.

A good kernel enables the isolation of the behaviour of a function at the origin. The Dirac delta function provides a classic example of this behaviour. The diagram below shows a family of Gaussian kernels of the form:

$$K_\delta(x) = \frac{1}{\sqrt{\delta}} e^{\frac{-\pi x^2}{\delta}} \qquad \delta > 0 \tag{54}$$



The Gaussian kernel (and the Dirac function for that matter) are not mere mathematical abstractions invented for the delectation of analysts. In fact physics drove the development of the Dirac function in particular. In advanced physics textbooks such as that by John David Jackson, *Classical Electrodynamics* , Third Edition, John Wiley, 1999 [Jackson] there are derivations of the Maxwell equations using *microscopic* rather than macroscopic principles eg see section 6.6 of [Jackson]. If you follow the discussion in that book you will see that for dimensions large compared to 10^{-14}m the nuclei can be treated as point systems which give rise to the microscopic Maxwell equations:

$$\nabla \cdot \mathbf{b} = 0, \quad \nabla \times \mathbf{e} + \frac{\partial \mathbf{b}}{\partial t} = 0, \quad \nabla \cdot \mathbf{e} = \frac{\eta}{\epsilon_0}, \quad \nabla \times \mathbf{b} - \frac{1}{c^2}\frac{\partial \mathbf{e}}{\partial t} = \mu_0 \mathbf{j}$$

Here $\mathbf{e}$ and $\mathbf{b}$ are he microscopic electric and magnetic fields and η and $\mathbf{j}$ are the microscopic charge and current densities. A question arises as to what type of averaging of the microscopic fluctations is appropriate and the Jackson says that "at first glance one might think that averages over both space and time are necessary. But this is not true. Only a spatial averaging is necessary" [p.249 Jackson] Briefly the broad reason is that in any region of macroscopic interest there are just so many nuclei and electrons so that the spatial averaging "washes" away the time fluctuations of the microscopic fields which are essentially uncorrelated at the relevant distance ($10^{-8}m$).

The spatial average of $F(\mathbf{x}, t)$ with respect to some test function $f(\mathbf{x})$ is defined as:

$\langle F(\mathbf{x}, t) \rangle = \int F(\mathbf{x} - \mathbf{x}', t)\, f(\mathbf{x}')\, d^3x'$ where $f(\mathbf{x})$ is real and non-zero in some neighbourhood of $\mathbf{x} = \mathbf{0}$ and is normalised to 1 over all space. It is reasonable to expect that $f(\mathbf{x})$ is isotropic in space so that there are no directional biases in the spatial averages. Jackson gives two examples as follows:

$$f(\mathbf{x}) = \begin{cases} \frac{3}{4\pi R^3}, & r < R \\ 0, & r > R \end{cases}$$

and

$$f(\mathbf{x}) = (\pi R^2)^{-\frac{3}{2}} e^{-\frac{r^2}{R^2}}$$

The first example is an average of a spherical volume with radius R but it has a discontinuity at $r = R$. Jackson notes that this "leads to a fine-scale jitter on the averaged quantities as a single molecule or group of molecules moves in or out of the average volume" [Jackson, page 250]. This particular problem is eliminated by a Gaussian test function "provided its scale is large compared to atomic dimensions" [Jackson, p.250]. Luckily all that is needed is that the test function meets general continuity and smoothness properties that yield a rapidly converging Taylor series for $f(\mathbf{x})$ at the level of atomic dimensions. Thus the Gaussian plays a fundamental role in the rather intricate calculations presented by Jackson concerning this issue.

If we take $K_\delta(x)$ as our kernel defined on $(-\infty, \infty)$ we find that these Gaussian kernels satisfy the following three conditions:

$$\int_{-\infty}^{\infty} K_\delta(x)dx = 1 \tag{55}$$

That this is the case follows by a change of variable in $\int_{-\infty}^{\infty} e^{-\pi x^2}dx = 1$. If you cannot recall how to prove this see the article on completing the square in Gaussian integrals here: `http://www.gotohaggstrom.com/page2.html`

$$\int_{-\infty}^{\infty} |K_\delta(x)|dx \leq M \tag{56}$$

Since $\delta > 0$ and given (55) it is certainly the case that this integral is bounded by some number ie 1.

$$\text{For all } \eta > 0\ , \qquad \int_{|x|\ >\eta} |K_\delta(x)|dx \to 0 \qquad as\ \delta \to 0 \tag{57}$$

The change of variable $u = \frac{x}{\sqrt{\delta}}$ gives the integral $\int_{|u|>\frac{\eta}{\sqrt{\delta}}} e^{-\pi u^2}du$ which clearly involves the area under the long tails of the Gaussian and these go to zero as $\delta \to 0$ ie as $\frac{\eta}{\sqrt{\delta}} \to \infty$. More formally this can be seen as follows for $u > 1$:

$\int_{|u|>\frac{\eta}{\sqrt{\delta}}} e^{-\pi u^2}du = 2\int_{\frac{\eta}{\sqrt{\delta}}}^{\infty} e^{-\pi u^2}du < 2\int_{\frac{\eta}{\sqrt{\delta}}}^{\infty} e^{-\pi u}du = 2\left[\frac{-1}{\pi}e^{-\pi u}\right]_{\frac{\eta}{\sqrt{\delta}}}^{\infty}$ which $\to 0$ as $\delta \to 0$

Before looking at a more general proof of (35) it is worth trying a simple example to test the logic of (35). So let's start with $f(x) = (x+1)^2$ and see if by convolving f with the Gaussian kernel $K_\delta(x)$ defined above in (54) we can recover f(x) ie $(f * K_\delta)(x) = f(x)$.

$$(f * K_\delta)(x) = \frac{1}{\sqrt{\delta}}\int_{-\infty}^{\infty} (x-y+1)^2 e^{\frac{-\pi y^2}{\delta}}dy = I_1 + I_2 + I_3 \tag{58}$$

$$I_1 = \frac{1}{\sqrt{\delta}}\int_{-\infty}^{\infty} (x^2 - 2xy + y^2) e^{\frac{-\pi y^2}{\delta}}dy = J_1 + J_2 + J_3 \tag{59}$$

$$I_2 = \frac{2}{\sqrt{\delta}}\int_{\infty}^{\infty} (x-y) e^{\frac{-\pi y^2}{\delta}}dy \tag{60}$$

$$I_3 = \frac{1}{\sqrt{\delta}} \int_{\infty}^{\infty} e^{\frac{-\pi y^2}{\delta}} dy \tag{61}$$

Using (55) it is clear that $J_1 = x^2$. Using the fact that the integrand in J_2 is an odd function we have that $J_2 = 0$. In relation to J_3 we note that:

$$J_3 = \frac{1}{\sqrt{\delta}} \int_{-\infty}^{\infty} y^2 e^{\frac{-\pi y^2}{\delta}} dy = \frac{2}{\sqrt{\delta}} \int_{0}^{\infty} y^2 e^{\frac{-\pi y^2}{\delta}} dy = \frac{2}{\sqrt{\delta}} \int_{0}^{1} y^2 e^{\frac{-\pi y^2}{\delta}} dy + \frac{2}{\sqrt{\delta}} \int_{1}^{\infty} y^2 e^{\frac{-\pi y^2}{\delta}} dy$$
$$\leq \frac{2}{\sqrt{\delta}} \int_{0}^{1} y e^{\frac{-\pi y^2}{\delta}} dy + \frac{2}{\sqrt{\delta}} \int_{1}^{\infty} y^2 e^{\frac{-\pi y}{\delta}} dy \tag{62}$$

After making the substitution $v = \frac{-\pi y^2}{\delta}$, the first integral in (62) is equal to $\frac{\sqrt{\delta}}{\pi}(1 - e^{-\frac{\pi}{\delta}})$ which $\to 0$ as $\delta \to 0$.

To demonstrate that the last member of (62) goes to zero as $\delta \to 0$ we need to integrate by parts as follows:

Take any $M > 0$ as large as you like, then let $u = y^2$ so that $du = 2y\,dy$ and let $dv = e^{\frac{-\pi y}{\delta}} dy$ so that $v = \frac{-\delta}{\pi} e^{\frac{-\pi y}{\delta}}$. Then:

$$\frac{2}{\sqrt{\delta}} \int_{0}^{M} y^2 e^{\frac{-\pi y}{\delta}} dy = \frac{2}{\sqrt{\delta}} \left[\frac{-\delta y^2}{\pi} e^{\frac{-\pi y}{\delta}} \right]_0^M + \frac{4\sqrt{\delta}}{\pi} \int_{0}^{M} y e^{\frac{-\pi y}{\delta}} dy \tag{63}$$

Now $\frac{2}{\sqrt{\delta}} \left[\frac{-\delta y^2}{\pi} e^{\frac{-\pi y}{\delta}} \right]_0^M \to 0$ as $\delta \to 0$, noting that M is fixed. Integrating the last integral in (63) again by parts we get:

$$\frac{4\sqrt{\delta}}{\pi} \int_{0}^{M} y e^{\frac{-\pi y}{\delta}} dy = \frac{4\sqrt{\delta}}{\pi} \left[\frac{-\delta y}{\pi} e^{\frac{-\pi y}{\delta}} \right]_0^M + \frac{4\sqrt{\delta}}{\pi} \int_{0}^{M} \frac{\delta}{\pi} e^{\frac{-\pi y}{\delta}} dy \tag{64}$$

Again $\frac{4\sqrt{\delta}}{\pi} \left[\frac{-\delta y}{\pi} e^{\frac{-\pi y}{\delta}} \right]_0^M \to 0$ as $\delta \to 0$ and $\frac{4\sqrt{\delta}}{\pi} \int_0^M \frac{\delta}{\pi} e^{\frac{-\pi y}{\delta}} dy = \frac{4\delta^{\frac{5}{2}}}{\pi^3}[1 - e^{\frac{-\pi M}{\delta}}]$ which also $\to 0$ as $\delta \to 0$. Hence $J_3 = 0$. This means that $I_1 = x^2$.

It is now easily seen that $I_2 = \frac{2}{\sqrt{\delta}} \int_{-\infty}^{\infty} (x-y) e^{\frac{-\pi y^2}{\delta}} dy = \frac{2x}{\sqrt{\delta}} \int_{-\infty}^{\infty} e^{\frac{-\pi y^2}{\delta}} dy - \frac{2}{\sqrt{\delta}} \int_{-\infty}^{\infty} y e^{\frac{-\pi y^2}{\delta}} dy$ where the first integral equals $2x$ and the second is zero because the integrand is odd. Hence $I_2 = 2x$.

Due to (55) $I_3 = 1$ so that finally we have: $(f * K_\delta)(x) = x^2 + 2x + 1 = f(x)$ as advertised in (35).

4.1 Properties of a good kernel

Following Stein and Shakarchi, a family of kernels $K_n(x)_{n=1}^{\infty}$ on the circle (ie an interval of 2π) is said to be "good" if three conditions are satisfied:

(a) For all $n \geq 1$,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} K_n(x)\, dx = 1 \tag{65}$$

(b) There exists some $M > 0$ such that for all $n \geq 1$,

$$\int_{-\pi}^{\pi} |K_n(x)|\, dx \leq M \tag{66}$$

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

$$\int_{\delta \leq |x| \leq \pi} |K_n(x)|\, dx \ \to 0 \qquad \text{as } n \to \infty \tag{67}$$

Property (a) says that the kernels are normalised. Property (b) says that the integral of the kernels is uniformly bounded ie they don't get too big. Property (c) says that the "tails" of the kernels vanish in the limit - think of the tails of the classic Gaussian probability density.

Note that the "right" class of kernels depends on what type of convergence results one is interested in eg almost everywhere convergence, convergence in L^1 or L^∞ norms and what restrictions one wants to place on the functions under consideration.

Applying these three properties to the Dirichlet kernel, the question is whether it is a good kernel. If it were, (35) would allow us to conclude that a Fourier series of f converges to f(x) whenever f is continuous at x. At first blush the Dirichlet kernel looks like it might be a good kernel because it satisfies the first criterion for a good kernel. This is demonstrated as follows:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(x)\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{n=-N}^{n} e^{inx}\,dx = \frac{1}{2\pi}\sum_{n=-N}^{n}\int_{-\pi}^{\pi} e^{inx}\,dx = \frac{1}{2\pi}\sum_{n=-N}^{n}\left[\frac{e^{inx}}{in}\right]_{-\pi}^{\pi}$$
$$= \frac{1}{2\pi}\sum_{n=-N}^{n}\frac{2i\sin(n\pi)}{in} = \sum_{n=-N}^{n}\frac{\sin(n\pi)}{n\pi} = \lim_{n\to 0}\frac{\sin(n\pi)}{n\pi} = 1 \quad (68)$$

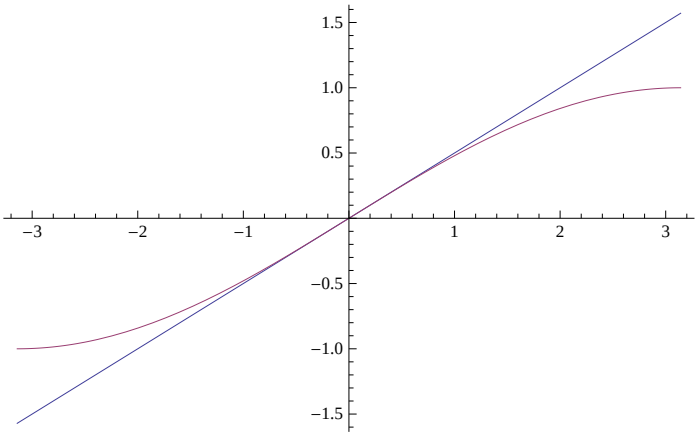
So far so good. Does the Dirichlet kernel satisfy the second property of a good kernel, namely, there exists some $M > 0$ such that for all $n \geq 1$, $\int_{-\pi}^{\pi} |K_n(x)|dx \leq M$? It is not immediately obvious that the Dirichlet kernel satisfies this second property and it takes some subtle analysis to demonstrate that it doesn't. This is Problem 2 in Chapter 2, page 66 of [SteinShakarchi].

Define

$$L_N = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(\theta)|\,dx \quad (69)$$

where $D_N(\theta) = \frac{\sin((N+\frac{1}{2})\theta)}{\sin(\frac{\theta}{2})}$. The aim is to show that $L_N \geq c\ln N$ for some constant $c > 0$ or, better still, that $L_N = \frac{4}{\pi^2}\ln N + \mathcal{O}(1)$.

For $\theta \in [-\pi, \pi]$, $|\sin(\frac{\theta}{2})| \leq |\frac{\theta}{2}|$ The following picture tells the story, but does it does not amount to a rigorous proof. If you know, for instance, that $\sin x = x(1-\frac{x^2}{\pi^2})(1-\frac{x^2}{2^2\pi^2})(1-\frac{x^2}{3^2\pi^2})\ldots$ then the inequality is obvious. Failing knowledge of the infinite product series for $\sin x$ you could fall back on Taylor's theorem with remainder. You would then get $\sin x = x + R_3(x)$ where $R_3(x) = \frac{x^3}{3!}\sin^{(3)}(\xi)$ and $-\pi \leq \xi \leq \pi$. The even powers vanish of course because $\sin^{(2n)}(0) = (-1)^k\sin(0) = 0$ for $k \geq 0$. Since $\sin^{(3)}(\xi) = -\cos(\xi)$ it follow that $\sin x \leq x$ and substitution of $x = \frac{\theta}{2}$ gives us the inequality we are after.



Therefore $\frac{1}{|\sin(\frac{\theta}{2})|} \geq \frac{2}{|\theta|}$ and hence:

$$L_N = \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_N(\theta)|d\theta \geq \frac{2}{2\pi}\int_{-\pi}^{\pi}\frac{|\sin((N+\frac{1}{2})\theta)|}{|\theta|}d\theta = \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{|\sin((N+\frac{1}{2})\theta)|}{|\theta|}d\theta \quad (70)$$

Let

$$I = \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{|\sin((N+\frac{1}{2})\theta)|}{|\theta|}d\theta \quad (71)$$

and make the substitution $u = (N+\frac{1}{2})\theta$ so that $du = (N+\frac{1}{2})d\theta$. Then (71) becomes:

$$\begin{aligned} I &= \frac{1}{\pi}\int_{-(N+\frac{1}{2})\pi}^{(N+\frac{1}{2})\pi}\frac{|\sin u|}{\frac{|u|}{N+\frac{1}{2}}}\frac{du}{N+\frac{1}{2}} = \frac{2}{\pi}\int_{0}^{(N+\frac{1}{2})\pi}\frac{|\sin u|}{|u|}\,du \\ &= \frac{2}{\pi}\left[\int_0^{\pi}\frac{|\sin u|}{|u|}\,du + \int_{\pi}^{N\pi}\frac{|\sin u|}{|u|}\,du + \int_{N\pi}^{N\pi+\frac{\pi}{2}}\frac{|\sin u|}{|u|}\,du\right] \\ &= \frac{2}{\pi}[I_0 + I_\pi + I_{N\pi}] \text{ where these symbols have the obvious meanings} \quad (72) \end{aligned}$$

We now proceed to estimate each of I_0, I_π and $I_{N\pi}$ as follows:

Since $\frac{\sin u}{u}$ is non-negative on [0,π]:

$$I_0 \quad = \quad \int_0^{\pi}\frac{\sin u}{u}\,du \qquad \geq \qquad \int_0^{\pi}\frac{\sin u}{\pi}\,du \quad = \quad \frac{2}{\pi} \text{ since } \frac{1}{u} \geq \frac{1}{\pi} \text{ on } (0,\pi] \quad (73)$$

To estimate I_π we need to split the integral up as follows:

$$I_\pi \quad = \quad \int_\pi^{N\pi} \frac{|\sin u|}{|u|}\, du \quad = \quad \sum_{k=1}^{N-1} \int_{k\pi}^{(k+1)\pi} \frac{|\sin u|}{|u|}\, du \quad (74)$$

Now in (74) for each k:

$$\int_{k\pi}^{(k+1)\pi} \frac{|\sin u|}{|u|}\, du \quad \geq \quad \frac{1}{(k+1)\pi} \int_{k\pi}^{(k+1)\pi} |\sin u|\, du \quad = \quad \frac{2}{(k+1)\pi} \quad (75)$$

since $\int_{k\pi}^{(k+1)\pi} |\sin u|\, du = 2$ for all integral $k \geq 0$.

Therefore,

$$I_0^{N\pi} \geq \sum_{k=0}^{N-1} \frac{2}{(k+1)\pi)} = \frac{2}{\pi} \sum_{k=0}^{N-1} \frac{1}{k+1} \geq \frac{2}{\pi}\, \ln N \quad (76)$$

In relation to the last inequality in (76), recall that $\int_1^N \frac{dx}{x} = \ln N$ and consider the rectangles formed by $(1,1), (2,1), (2,\frac{1}{2}), (3,\frac{1}{2})$, etc the areas of which are 1, $\frac{1}{2}$, $\frac{1}{3}$ and so on.

The final integral is:

$$I_{N\pi} = \int_{N\pi}^{N\pi+\frac{\pi}{2}} \frac{|\sin u|}{|u|}\, du \geq \frac{1}{(N+\frac{1}{2})\pi} \int_{N\pi}^{N\pi+\frac{\pi}{2}} |\sin u|\, du = \frac{1}{(N+\frac{1}{2})\pi} \int_0^{\frac{\pi}{2}} \sin u\, du = \frac{1}{(N+\frac{1}{2})\pi} \quad (77)$$

Putting the three integrals together from (70) and (72):

$$L_N \geq I = \frac{2}{\pi}\left[\frac{2}{\pi} + \frac{2}{\pi}\ln N + \frac{1}{(N+\frac{1}{2})\pi}\right] = \frac{4}{\pi^2}\ln N + \frac{4}{\pi^2} + \frac{2}{(N+\frac{1}{2})\pi^2} = \frac{4}{\pi^2}\ln N + \mathcal{O}(1) \quad (78)$$

since the sum of the other two terms is bounded.

The inequality in (78) demonstrates that the Dirichlet kernel fails to satisfy the second property of a good kernel (see (66)).

5 The $Fejér$ kernel is a good kernel

$Cesàro$ summability can be applied in the context of the $Fejér$ kernel which is defined as follows by reference to the n^{th} $Cesàro$ mean of the Fourier series ie:

$$\sigma_n(f)(x) = \frac{S_0(f)(x) + \cdots + S_{n-1}(f)(x)}{n} \tag{79}$$

Recall that $S_n(f)(x) = \sum_{k=-n}^{n} \hat{f}(n)e^{ikx}$ and that $S_n(f)(x) = (f * D_n)(x)$ from (51). The n^{th} $Fejér$ kernel is defined as:

$$F_n(x) = \frac{D_0(x) + \cdots + D_{n-1}(x)}{n} \tag{80}$$

With this definition:

$$\sigma_n(f)(x) = (f * F_n)(x) \tag{81}$$

To show that the $Fejér$ kernel is a good kernel we first need a closed form for F_n. Going back to (53) we hve that $D_k(x) = \frac{\omega^{-k} - \omega^{k+1}}{1-\omega}$ where $\omega = e^{ix}$ hence:

$$nF_n(x) = \sum_{k=0}^{n-1} \frac{\omega^{-k} - \omega^{k+1}}{1-\omega} = \frac{1}{1-\omega}\left\{\frac{1-\omega^{-n}}{1-\omega^{-1}} - \frac{\omega(1-\omega^n)}{1-\omega}\right\} = \frac{1}{1-\omega}\left\{\frac{\omega(1-\omega^{-n})}{\omega-1} - \frac{\omega(1-\omega^n)}{1-\omega}\right\}$$

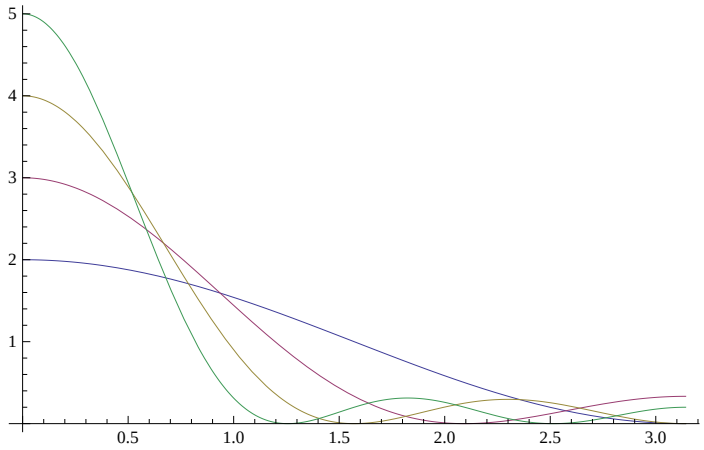
$$= \frac{\omega}{(1-\omega)^2}\left\{\omega^{-n} - 2 + \omega^n\right\} = \frac{\omega^{\frac{1}{2}}\omega^{\frac{1}{2}}}{(\omega^{\frac{1}{2}}\omega^{\frac{-1}{2}} - \omega^{\frac{1}{2}}\omega^{\frac{1}{2}})^2}\left(\omega^{\frac{-n}{2}} - \omega^{\frac{n}{2}}\right)^2 = \frac{1}{(\omega^{\frac{-1}{2}} - \omega^{\frac{1}{2}})^2}(\omega^{\frac{-n}{2}} - \omega^{\frac{n}{2}})^2$$

$$= \frac{1}{(-2i\sin(\frac{x}{2}))^2}(2i\sin(\frac{nx}{2}))^2 = \frac{\sin(\frac{nx}{2}))^2}{\sin(\frac{x}{2}))^2} \tag{82}$$

Therefore:

$$F_n(x) = \frac{\sin(\frac{nx}{2}))^2}{n\ \sin(\frac{x}{2}))^2} \tag{83}$$

The graph of $F_n(x)$ looks like this for $n = 2, 3, 4, 5$:



To show that $F_n(x)$ has the proper normalisation to be a good kernel we have to show that $\frac{1}{2\pi}\int_{-\pi}^{\pi} F_n(x)\,dx = 1$. But $\frac{1}{2\pi}\int_{-\pi}^{\pi} F_n(x)\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{D_0(x)+\cdots+D_{n-1}(x)}{n}\,dx$

From (68) we know that $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x)\,dx = 1$ so that $\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{D_0(x)+\cdots+D_{N-1}(x)}{n}\,dx = \frac{n}{n} = 1$.

The third requirement for a good kernel is that for every $\delta > 0$, $\int_{\delta\le|x|\le\pi}|F_n(x)|\,dx \to 0$ as $n \to \infty$. For $0 < \delta \le |x| \le \pi$, $|\sin x| \ge \frac{|x|}{2}$ and hence $\sin^2\frac{x}{2} \ge \frac{x^2}{4} \ge \frac{\delta^2}{4}$. Thus $F_n(x) \le \frac{4}{n\delta^2}$ so that $\int_{\delta\le|x|\le\pi}|F_n(x)|\,dx \to 0$ as $n \to \infty$. This establishes that the *Fejér* kernel is a good kernel.

6 APPENDIX OF FUNDAMENTAL ANALYTICAL RESULTS

6.1 A basic first result on convergence of averages

Suppose $x_n \to l$. Does it follow that the average $\frac{x_1+x_2+\cdots+x_n}{n} \to l$? That the answer is "yes" can be suggested by the observation that if $x_n = \mathcal{O}(\frac{1}{n})$, say, so that $x_n \to 0$, then it should follow that $\frac{x_1+x_2+\cdots+x_n}{n} \to 0$. By saying that $x_n = \mathcal{O}(\frac{1}{n})$ means that x_n is of *order* $\frac{1}{n}$ which is to say that there is an $A > 0$ such that $|x_n| \le \frac{A}{n}$

for all *large* n. Thus if $x_n = \mathcal{O}(\frac{1}{n})$ it follows that for all n beyond any such large N, $\left|\frac{x_1+x_2+\cdots+x_N+x_{N+1}+\cdots+x_n}{n}\right| \leq \left|\frac{x_1+x_2+\cdots+x_N}{n}\right| + \left|\frac{x_{N+1}+x_{N+2}+\cdots+x_n}{n}\right| \leq \frac{|x_1|+|x_2|+\cdots+|x_N|}{n} + \frac{|x_{N+1}|+|x_{N+2}|+\cdots+|x_n|}{n} \leq \frac{|x_1|+|x_2|+\cdots+|x_N|}{n} + \frac{(n-N)\frac{A}{n}}{n}$ which can be made arbitrarily small for n sufficiently large. Thus the sequence converges to 0.

This basic result is deceptively subtle in one respect possibly obscured by the following mechanical (N, ϵ) proof.

Let $x_n = y_n + l$. We then have to show that $\frac{y_1+y_2+\cdots+y_n}{n} \to 0$, if $y_n \to 0$ for then $x_n \to l$. By assumption that $y_n \to 0$ there exists an N_1 such that $|y_n| < \frac{\epsilon}{2}$ for all $n > N_1$. We now split the y_i as follows:

$$\frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{y_1 + y_2 + \cdots + y_{N_1}}{n} + \frac{y_{N_1+1} + y_{N_1+2} + \cdots + y_n}{n}$$

$$\text{so that} \left|\frac{y_1 + y_2 + \cdots + y_n}{n}\right| \leq \left|\frac{y_1 + y_2 + \cdots + y_{N_1}}{n}\right| + \left|\frac{y_{N_1+1} + y_{N_1+2} + \cdots + y_n}{n}\right|$$

$$\leq \frac{|y_1| + |y_2| + \cdots + |y_{N_1}|}{n} + \frac{|y_{N_1+1}| + |y_{N_1+2}| + \cdots + |y_n|}{n}$$

$$\leq \frac{|y_1| + |y_2| + \cdots + |y_{N_1}|}{n} + \frac{(n - N_1)\frac{\epsilon}{2}}{n} \quad (84)$$

This leaves $\frac{|y_1|+|y_2|+\cdots+|y_{N_1}|}{n}$ which is bounded by $\frac{N_1 B}{n}$ where $B = \max_{\{k=1,\ldots,N_1\}}\{|y_k|\}$. Now choose N_2 such that $\frac{N_1 B}{n} < \frac{\epsilon}{2}$ for $n > N_2$. Then for $n > \max\{N_1, N_2\}$: $\left|\frac{y_1+y_2+\cdots+y_n}{n}\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ which establishes the result.

There is a subtle, perhaps typically pedantic point, here which is alluded to by G H Hardy in "A Course of Pure Mathematics", Cambridge University Press, 2006 page 167 [HardyPM]. It is critical that N_1 and N_2 approach ∞ more slowly than n. Hardy is explicit on this point when he says that you divide the y_i into two sets $y_1, y_2, \ldots, y_p$ and $y_{p+1}, y_{p+2}, \ldots, y_n$ "where p is some function of n which tends to ∞ as $n \to \infty$ more slowly than n, so that $p \to \infty$ and $\frac{p}{n} \to 0$ eg we might suppose p to be the integral part of $\sqrt{n}$. In the step where the first part is shown to be bounded by $\frac{N_1 B}{n}$, for instance, it is essential in making this arbitrarily small that N_1 cannot approach ∞ at the same rate as n, for otherwise we would be left with something the order of B which may not be small.

Notation:
In what follows the "Big $\mathcal{O}$" notation $c_n = \mathcal{O}(\frac{1}{n})$ means that there exist an $A > 0$ such that for all sufficiently large n, $|c_n| \leq \frac{A}{n}$. Similarly the "Little o" notation $c_n = o(\frac{1}{n})$ means that $nc_n \to 0$ ie there exists an $A > 0$ such that for all sufficiently large n, $|nc_n| \leq \frac{A}{n}$

6.2 Convergence implies $Cesàro$ summability

If $\sum c_k \to s$ then $\sum c_k$ is also $Cesàro$ summable to s. Without loss of generality we can suppose that $s = 0$. We can do this for the following reason: suppose $\sum_{k=1}^{\infty} c_k \to s \neq 0$, then $s_n = \sum_{k=1}^{n} c_k \to s$ and hence the sequence $\{s_n - s\} \to 0$. This in turn means that $\frac{1}{n}\left(\sum_{k=1}^{n}(s_k - s)\right) \to (\sigma_n - \frac{s}{n}) = 0$ ie $\sigma_n \to \frac{s}{n} = 0$. In other words we may as well settle for $s = 0$ since that is easy.

We have to prove that $\sigma_n \to 0$ where $\sigma_n = \frac{s_1 + s_2 + \cdots + s_n}{n}$ and $s_n = c_1 + c_2 + \cdots + c_n$. Since $\sum_{k=1}^{\infty} c_k \to 0$ we have that $s_n \to 0$. Thus $\exists N$ such that $|s_n| < \epsilon$, $\forall n > N$.

Let $B = \max_{\{k=1,\ldots,N\}}\{|s_k|\}$

$$
\begin{aligned}
|\sigma_n| &= \left|\frac{s_1 + s_2 + \cdots + s_N + s_{N+1} + \cdots + s_n}{n}\right| \\
&\leq \frac{|s_1| + |s_2| + \cdots + |s_N|}{n} + \frac{|s_{N+1}| + |s_{N+2}| + \cdots + |s_n|}{n} \\
&\leq \frac{NB + (n-N)\epsilon}{n} = \frac{NB}{n} + (1 - \frac{N}{n})\epsilon < \frac{NB}{n} + \epsilon < \epsilon + \epsilon = 2\epsilon \quad (85)
\end{aligned}
$$

Once again in (64) is has been implicitly assumed that N increases more slowly than n. Thus $\sigma_n \to 0$ and so $\sum c_k$ is $Cesàro$ summable to 0.

6.3 Convergence implies Abel summability ie Abel summability is stronger than ordinary or $Cesàro$ summability

This is exercise 13 in Chapter 2, page 62 of [SteinShakarchi]. We need to show that if $\sum_{k=1}^{\infty} c_k$ converges to a finite limit s then the series is Abel summable to s. For the reasons given in 5.2 it is enough to prove the theorem when $s = 0$. In what follows, for convenience, I won't bother assuming the series members are complex since nothing of importance is lost by simply assuming that the numbers are reals. So on the assumption that the series converges to 0 let $s_n = c_1 + c_2 + \cdots + c_n$.

The broad idea is to get an expression for $\sum_{k=1}^{n} c_k\, r^k$ in terms of sums of s_n and r^n because we know that since $0 \leq r < 1$, $r^n \to 0$ and - this is a critical observation - the s_n are bounded since the series converges to zero.

We start with:

$$\sum_{k=1}^{n} c_k\, r^k = c_1 r + c_2\, r^2 + c_3\, r^3 + \cdots + c_n\, r^n \tag{86}$$

We then see what we can make out of this:

$$\sum_{k=1}^{n} s_k\, r^k = s_1 r + s_2\, r^2 + s_3\, r^3 + \cdots + s_n\, r^n = c_1\, r + (c_1+c_2) r^2 + (c_1+c_2+c_3) r^3 + \cdots + (c_1+c_2+\cdots+c_n) r^n$$

$$= c_1\, r + c_2\, r^2 + c_3\, r^3 + \cdots + c_n\, r^n + c_1\, r^2 + (c_1+c_2) r^3 + (c_1+c_2+c_3) r^4 + \cdots + (c_1+c_2+\cdots+c_{n-1}) r^n$$

$$= \sum_{k=1}^{n} c_k\, r^k + r\{c_1\, r + (c_1+c_2) r^2 + (c_1+c_2+c_3) r^3 + \cdots + (c_1+c_2+\cdots+c_{n-1}) r^{n-1}\}$$

$$= \sum_{k=1}^{n} c_k\, r^k + r \sum_{k=1}^{n-1} s_k\, r^k \;\; = \sum_{k=1}^{n} c_k\, r^k + r \sum_{k=1}^{n} s_k\, r^k - s_n r^{n+1} \tag{87}$$

Thus from (87) we get that:

$$\sum_{k=1}^{n} c_k\, r^k = (1-r) \sum_{k=1}^{n} s_k\, r^k + s_n r^{n+1} \tag{88}$$

Now because $\sum_{k=1}^{\infty} c_k$ converges to 0, the s_n also converge to 0 (they are also bounded, of course). Hence, for any fixed $0 \leq r < 1$, $s_n\, r^{n+1} \to 0$ as $n \to \infty$. This leaves us to estimate as $r \to 1$:

$$(1-r) \sum_{k=1}^{n} s_k\, r^k \tag{89}$$

Now for n sufficiently large, $|s_k| \leq 1 - r$ for all $k \geq n$. Let $B = max\{|s_1|, |s_2| \ldots |s_{n-1}|\}$ Then:

$$(1-r)\left|\sum_{k=1}^{\infty} s_k\, r^k\right| = (1-r)\left|\sum_{k=1}^{n-1} s_k r^k + \sum_{k=n}^{\infty} s_k r^k\right| \le (1-r)\left|\sum_{k=1}^{n-1} s_k r^k\right| + (1-r)\left|\sum_{k=n}^{\infty} s_k r^k\right|$$

$$\le (1-r)B\,\frac{r(1-r^{n-1})}{1-r} + (1-r)^2\frac{r^n}{1-r}$$

$$= Br(1-r^{n-1}) + (1-r)r^n$$

which $\to 0$ as $r \to 1$, noting that n is fixed so that $1 - r^{n-1} \to 0$ as $r \to 1$ (90)

Thus we have shown that the series is Abel summable to zero. The converse of what has been shown is not neccessarily true since $c_n = (-1)^n$ is Abel summable to $\frac{1}{2}$ but the alternating series $1 - 1 + 1 - 1 + \ldots$ does not converge. That c_n is Abel summable follows from the fact that the n^{th} partial sum of $\sum_{k=0}^{\infty}(-1)^k r^k$ is dominated by the convergent geometric series $\sum_{k=0}^{n} r^k$. Since $A(r) = 1 - r + r^2 - r^3 + r^4 - \ldots$, we have that $rA(r) = r - r^2 + r^3 - r^4 + \ldots$. Hence $(1 + r)A(r) = 1$ and so $A(r) = \frac{1}{1+r}$ which has limit $\frac{1}{2}$ as $r \to 1$.

6.4 *Cesàro* summability implies Abel summability

There is an analogous result for *Cesàro* summability, namely, that if a series $\sum_{k=1}^{\infty} c_k$ is *Cesàro* summable to σ then it is Abel summable to σ. The concept of *Cesàro* summability is based on the behaviour of the n^{th} *Cesàro* mean which is defined as:

$$\sigma_n = \frac{s_1 + s_2 + \cdots + s_n}{n} \tag{91}$$

The s_i are the partial sums of the series of complex or real numbers $c_1 + c_2 + c_3 + \cdots = \sum_{k=1}^{\infty} c_k$. That is, $s_n = \sum_{k=1}^{n} c_k$. If σ_n converges to a limit σ as $n \to \infty$ then the series $\sum_{k=1}^{\infty} c_k$ is said to be *Cesàro* summable to σ.

To prove that *Cesàro* summability of $\sum_{k=1}^{\infty} c_k$ implies Abel summability of $\sum_{k=1}^{\infty} c_k$ we have to develop a relationship between $\sum_{k=1}^{\infty} c_k\, r^k$ and $\sum_{k=1}^{\infty} k\sigma_k r^k$. One way to get a relationship is to simply expand $\sum_{k=1}^{n} k\sigma_k\, r^k$ and see if there is any structural appearance of $\sum_{k=1}^{\infty} c_k\, r^k$. Thus:

$$\sum_{k=1}^{\infty} k\sigma_k\, r^k = \sigma_1\, r + 2\sigma_2\, r^2 + 3\sigma_3\, r^3 + 4\sigma_4\, r^4 + \cdots = s_1\, r + (s_1+s_2)r^2 + (s_1+s_2+s_3)r^3 + (s_1+s_2+s_3+s_4)r^4 + \ldots$$

$$= c_1\, r + (2c_1 + c_2)r^2 + (3c_1 + 2c_2 + c_3)r^3 + (4c_1 + 3c_2 + 2c_3 + c_4)r^4 + \ldots \tag{92}$$

Now to see a useful structure it is useful to write out the last line of (92) as a series of collected terms and then look down the diagonals of the representation as follows:

$$\begin{aligned}\sum_{k=1}^{\infty} k\sigma_k\, r^k = {} & c_1\, r + 2c_1 r^2 + 3c_1 r^3 + 4c_1 r^4 + 5c_1 r^5 + \dots \\ & + c_2 r^2 + 2c_2 r^3 + 3c_2 r^4 + 4c_2 r^5 + \dots \\ & + c_3 r^3 + 2c_3 r^4 + 3c_3 r^5 + \dots \\ & + c_4 r^4 + 2c_3 r^5 + 3c_4 r^6 + \dots \end{aligned} \tag{93}$$

Looking down the diagonals we see the following structure:

$$\begin{aligned} & c_1\, r + c_2 r^2 + c_3 r^3 + c_4 r^4 + \dots \\ & + 2c_1 r^2 + 2c_2 r^3 + 2c_3 r^4 + 2c_4 r^5 + \dots \\ & + 3c_1 r^3 + 3c_2 r^4 + 3c_3 r^5 + \dots \\ & + 4c_1 r^4 + 4c_2 r^5 + 4c_3 r^6 + \dots \end{aligned} \tag{94}$$

Thus (92) can be rewritten as:

$$\sum_{k=1}^{\infty} k\sigma_k\, r^k = \sum_{k=1}^{\infty} c_k r^k + 2r \sum_{k=1}^{\infty} c_k r^k + 3r^2 \sum_{k=1}^{\infty} c_k r^k + 4r^3 \sum_{k=1}^{\infty} c_k r^k + \dots \tag{95}$$

Now the trick here is to realise that (95) is formally equal to:

$$\sum_{k=1}^{\infty} k\sigma_k\, r^k = \frac{1}{(1-r)^2} \sum_{k=1}^{\infty} c_k r^k \tag{96}$$

To see this just do the long division: $\frac{1}{1-2r+r^2} = 1 + 2r + 3r^2 + 4r^3 + \dots$

Thus we get the relationship we were after:

$$\sum_{k=1}^{\infty} c_k r^k = (1-r)^2 \sum_{k=1}^{\infty} k\sigma_k\, r^k \tag{97}$$

We assume as in section 6.2 that $\sigma \to 0$ for similar reasons. We split the infinite sum in (97) into two sums as follows:

$$(1-r)^2 \sum_{k=1}^{\infty} k\sigma_k \, r^k = (1-r)^2 \sum_{k=1}^{N} k\sigma_k \, r^k + (1-r)^2 \sum_{k=N+1}^{\infty} k\sigma_k \, r^k = L1 + L2 \tag{98}$$

We want to show that $\left| \sum_{k=1}^{n} c_k r^k \right| \to 0$ as $n \to \infty$.

The N is chosen this way: we know that since $\sigma \to 0$, $\forall \epsilon > 0$, $\exists N$ such that $|\sigma_k| < \epsilon$, $\forall k > N$. This will be used when estimating L2. We start with estimating L1 as follows:

$$\begin{aligned}
|L1| = (1-r)^2 \left| \sum_{k=1}^{N} k\sigma_k \, r^k \right| = (1-r)^2 \left| \sum_{k=1}^{N} (s_1 + s_2 + \cdots + s_k) \, r^k \right| &\leq (1-r)^2 \sum_{k=1}^{N} (|s_1| + |s_2| + \cdots + |s_k|) \, r^k \\
&= (1-r)^2 \sum_{k=1}^{N} (|c_1| + |c_1 + c_2| + \cdots + |c_1 + c_2 + \cdots + c_k|) \, r^k \\
&\leq (1-r)^2 \sum_{k=1}^{N} (|c_1| + |c_1| + |c_2| + \cdots + |c_1| + |c_2| + \cdots + |c_k|) \, r^k \\
&= (1-r)^2 \sum_{k=1}^{N} (k|c_1| + (k-1)|c_2| + \cdots + |c_k|) \, r^k \\
&\leq (1-r)^2 \sum_{k=1}^{N} N^2 c \, r^k \leq (1-r)^2 N^3 c
\end{aligned} \tag{99}$$

Here $c = max_{j=1,2,\ldots N}\{|c_j|\}$. Since N and c are fixed, $(1-r)^2 N^3 c \to 0$ as $r \to 1$. Thus $L1 \to 0$ as $r \to 1$.

Showing that $L2 = (1-r)^2 \sum_{k=N+1}^{\infty} k\sigma_k r^k \to 0$ is trickier. We need a preliminary result which essentially boils down to:

$$\lim_{x \to \infty} x e^{-x} = 0 \tag{100}$$

This limit is proved in calculus and analysis courses and is a very important limit. To prove it one can assume that $t > 1$ and let β be any positive rational exponent. Clearly then $t^\beta > 1$ (just think of a binomial expansion with a positive rational exponent). Since $t^\beta > 1$ it follows that $t^{\beta-1} > \frac{1}{t}$.
Now $\ln x = \int_1^x \frac{dt}{t} < \int_1^x t^{\beta-1} dt = \frac{x^\beta - 1}{\beta} < \frac{x^\beta}{\beta}$.

If $\alpha > 0$ we can choose a smaller $\beta > 0$ such that:

$$0 < \frac{\ln x}{x^\alpha} < \frac{x^{\beta-\alpha}}{\beta} \tag{101}$$

Now $\frac{x^{\beta-\alpha}}{\beta}$ tends to 0 as $x \to \infty$ because $\beta < \alpha$. Thus $x^{-\alpha} \ln x \to 0$.

In effect, (100) says that e^x tends to ∞ more rapidly than any power of x. To see this, because $x^{-\alpha} \ln x \to 0$ as $x \to \infty$ where $\alpha > 0$, let $\alpha = \frac{1}{\beta}$ from which it follows that $x^{-\alpha\beta} (\ln x)^\beta = x^{-1} (\ln x)^\beta \to 0$. If we let $x = e^y$ we see that $e^{-y} y^\beta \to 0$. Since $e^{\gamma y} \to \infty$ if $\gamma > 0$ and $e^{\gamma y} \to 0$ if $\gamma < 0$, we see that $(e^{-y} y^\beta)^\gamma = e^{-\gamma y} y^{\beta\gamma} \to 0$. In other words the result holds for any power of y.

To estimate L2 we consider $(1-r)^2 \sum_{k=N+1}^M k\sigma_k r^k$ as $M \to \infty$. Thus:

$$|L2| = (1-r)^2 \left| \sum_{k=N+1}^{M} k\sigma_k r^k \right| \leq (1-r)^2 \sum_{k=N+1}^{M} k|\sigma_k| r^k < (1-r)^2 \sum_{k=N+1}^{M} k\epsilon\, r^k \leq (1-r)^2 \epsilon \int_N^M x e^{(\ln r)x} dx \tag{102}$$

Note that for all $k > N$, $|\sigma_k| < \epsilon$ hence $k\,|\sigma_k| < k\,\epsilon$

Integrating by parts we get:

$$\epsilon\,(1-r)^2 \int_N^M x e^{(\ln r)x} dx = \epsilon\,(1-r)^2 \left\{ \left[\frac{x e^{(\ln r)x}}{\ln r} \right]_N^M - \frac{1}{\ln r} \int_N^M e^{(\ln r)x} dx \right\} =$$
$$\epsilon\,(1-r)^2 \left\{ \frac{M e^{(\ln r)M} - N e^{(\ln r)N}}{\ln r} - \frac{1}{(\ln r)^2} [e^{(\ln r)M} - e^{(\ln r)N}] \right\} \tag{103}$$

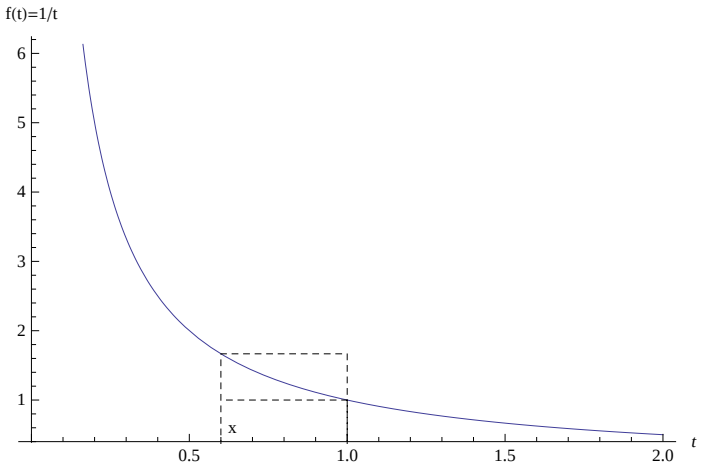
First fix r, noting that N is already fixed, and also note that since $\ln r < 0$ for $0 < r < 1$, then as $M \to \infty$, $M e^{(\ln r)M} \to 0$ using (100) and the comments relating to it. Accordingly, as $M \to \infty$ (103) becomes:

$$\epsilon\,(1-r)^2\left\{\frac{-Ne^{(\ln r)N}}{\ln r}+\frac{e^{(\ln r)N}}{(\ln r)^2}\right\}=\epsilon(1-r)^2\,e^{(\ln r)N}\left\{\frac{1-N\ln r}{(\ln r)^2}\right\}<\epsilon\,e^{(\ln r)N}\,(1-N\ln r)\left\{\frac{1-r}{\ln r}\right\}^2 \tag{104}$$

Clearly $e^{(\ln r)N}\to 1$ and $(1-N\ln r)\to 1$ as $r\to 1^-$. The behaviour of $\left\{\frac{1-r}{\ln r}\right\}^2$ as $r\to 1^-$ can be established by using L'Hôpital's rule or a direct method. Since both the numerator and denominator in $\left\{\frac{1-r}{\ln r}\right\}^2$ approach zero as $r\to 1^-$ the limit of $\frac{1-r}{\ln r}$ is -r ie $\left(\frac{\frac{d(1-r)}{dr}}{\frac{d(\ln r)}{dr}}\right)$ which approaches -1, hence the required limit is its square which is 1.
Alternatively, we can use the definition of $\ln x$ as follows:

For $0<x<1$, $\ln(1-x)=-\int_{1-x}^{1}\frac{dt}{t}$ and using the diagram below it is clear that:

$$x\le -\ln(1-x)\le\frac{x}{1-x} \tag{105}$$



Substituting $x=1-r$ in (84) gives:

$$1-r<-\ln r<\frac{1-r}{r}\Rightarrow 1<\frac{-lnr}{1-r}<\frac{1}{r}\ \ \therefore\frac{-\ln r}{1-r}\to 1\text{ as }r\to 1 \tag{106}$$

Thus $\left\{\frac{1-r}{\ln r}\right\}^2 \to 1$ as $r \to 1^-$

Finally, our estimate of L2 boils down to this following on from (102)-(104):

$$|L2| \quad < \quad \epsilon\, e^{(\ln r)N}\,(1 \;-\; N\ln r)\left\{\frac{1-r}{\ln r}\right\}^2 \quad \to \quad \epsilon \;\times\; 1 \;\times\; 1 \;\times\; 1 \quad = \quad \epsilon \qquad (107)$$

Thus $L2 \to 0$ as $r \to 1^-$ and we have established that if the series is *Cesàro* summable to 0 then it is also Abel summable to 0.

Thus what we have got to is this:

$$\boxed{\text{convergent} \Rightarrow \textit{Cesàro} \text{ summable} \Rightarrow \text{Abel summable}} \qquad (108)$$

None of these implications can be reversed. However, using so-called "Tauberian" theoems we can find conditions on the rate of decay of the c_k which allow the implications to be reversed. This is what Exercise 14 of Chapter 2 of [SteinShakarchi] is about.

6.5 Applying Tauberian conditions to reverse the implications

When dealing with the convergence of sequences and functions there is a concept of "regularity" which means that the method of summation (ie averaging) ensures that every convergent series converges to its ordinary sum. We have just seen that the *Cesàro* and Abelian methods of summation are regular since $\sum_{k=1}^{\infty} c_k = s$ implies both $s_n = \frac{s_1+s_2+\cdots+s_n}{n} \to s$ and $f(x) = \sum_{k=1}^{\infty} c_k x^k \to s$.

An Abelian type of theorem is essentially one which asserts that, if a sequence or function behaves in a regular fashion, the some average of the sequence or function will also behave regularly. The converses of Abelian theorems are usually false (and as you work through the proofs below you will see why this is so) but if some method of summation were reversible it would only be summing already convergent series and hence be of no interest.

What Tauber did was to place conditions on the rate of decay of the c_k in order to achieve non-trivial reversibility of the the implications in (108). The simplest rate of decay which Tauber chose was $c_k = o(\frac{1}{n})$ ie $nc_n \to 0$. G H Hardy's book "Divergernt

Series" [HardyDS] contains a detailed exploration of all the issues and is well worth reading. Unfortunately Hardy had a snobbish and pedantic style which annoyed applied mathematicians during his life so you have to make allowances for that.

The generalisation of the concepts of convergence is as follows. If $s_n \to s$ then we can equivalently say that:

$$\sum a_k r^k \to s \ , \ (1-r)\sum s_k r^k \to s, \ \ y\sum s_k e^{-ky} \to s, \ \ \frac{1}{x}\sum s_k e^{(\frac{-k}{x})} \to s \qquad (109)$$

See p.283 of [HardyDS].

We first show that if $\sum c_n$ is *Cesàro* summable to σ and $c_n = o(\frac{1}{n})$, ie $nc_n \to 0$, then $\sum c_n$ converges to σ. This is Problem 14(a) page 62 in [SteinShakarchi].

Since $\sum c_n$ is *Cesàro* summable to σ, $\exists N_1$ such that $|\sigma_n - \sigma| < \frac{\epsilon}{3}$, $\forall n > N_1$

Now

$$|s_n - \sigma| = |s_n - \sigma_n + \sigma_n - \sigma| \leq |s_n - \sigma_n| + |\sigma_n - \sigma| \quad (110)$$

We need to "massage" $s_n - \sigma_n$ order to get something useful involving the c_k. One way to do this is as follows:

$$\begin{aligned}
|s_n - \sigma_n| &= \left| s_n - \frac{(s_1 + s_2 + \cdots + s_n)}{n} \right| = \left| \frac{(n-1)s_n - (s_1 + s_2 + \cdots + s_{n-1})}{n} \right| = \\
&\left| \frac{(n-1)(c_1 + c_2 + \cdots + c_n) - \big\{c_1 + (c_1 + c_2) + \cdots + (c_1 + c_2 + \cdots + c_{n-1})\big\}}{n} \right| \\
&= \left| \frac{(n-1)(c_1 + c_2 + \cdots + c_n) - \big\{(n-1)c_1 + (n-2)c_2) + (n-3)c_3 \cdots + 2c_{n-2} + c_{n-1})\big\}}{n} \right| \\
&= \left| \frac{(n-1)c_n + (n-2)c_{n-1} + (n-3)c_{n-2} + \cdots + (N_2 - 1)c_{N_2} + \cdots + 2c_3 + c_2}{n} \right| \\
&\leq \frac{(|c_2| + 2|c_3| + \cdots + (N_2 - 1)|c_{N_2}|)}{n} + \frac{(N_2|c_{N_2+1}| + \cdots + (n-1)|c_n|)}{n} \\
&\leq (N_2 - 1)\frac{(|c_2| + |c_3| + \cdots + |c_{N_2}|)}{n} + (n-1)\frac{(|c_{N_2+1}| + \cdots + |c_n|)}{n} = L1 + L2 \quad (111)
\end{aligned}$$

Now because $c_n = o(\frac{1}{n})$, for any $\epsilon > 0$ there exists an $A > 0$ and N_2 such that $|c_n| \leq \frac{A}{n}$ for sufficiently large n. Thus we can find an N_2 such that $|c_n| < \frac{\epsilon}{3n}$ for all $n > N_2$ and hence L2 is estimated as follows

$$L2 \;=\; (n\,-\,1)\frac{(|c_{N_2+1}| + \cdots + |c_n|)}{n} \;\leq\; \frac{(n-1)(n-N_2)\frac{\epsilon}{3n}}{n} \;<\; \frac{n^2\epsilon}{3n^2} \;<\; \frac{\epsilon}{3} \quad (112)$$

Let $c = \max_{\{k=2,\ldots,N_2\}}\{|c_k|\}$.

$$L1 = (N_2 - 1)\frac{(|c_2| + |c_3| + \cdots + |c_{N_2}|)}{n} \leq \frac{(N_2-1)^2\, c}{n} < \frac{\epsilon}{3}$$
$$\text{since } \exists N_3 \text{ such that } \frac{(N_2-1)^2 c}{n} < \frac{\epsilon}{3} \;\; \forall n > N_3 \quad (113)$$

So choosing $n > \max\{N_1, N_2, N_3\}$ we have that:

$$|s_n - \sigma| \leq |s_n - \sigma_n| + |\sigma_n - \sigma| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \quad (114)$$

Hence $s_n \to \sigma$ as required. Note that without the condition $c_n = o(\frac{1}{n})$ L2 would not necessarily be small, for example, if the c_n were simply bounded L2 would be of order n.

Problem 14(b) on page 62 [SteinShakarchi] deals with imposing conditions on Abel summability to ensure convergence. The classic Tauberian result is this: If $\sum c_n$ is Abel summable to s and $c_n = o(\frac{1}{n})$ ie $nc_n \to 0$ then $\sum c_n \to s$.

Recall that $\sum c_n$ being Abel summable to s means that $A(r) = \sum_{k=0}^{\infty} c_k r^k$ converges for all r, $0 \leq r < 1$ and $\lim_{r\to 1} A(r) = s$.

Let $S_n = \sum_{k=0}^{n} c_k$ and as discussed previously, we can take the limit s to be zero without any loss of generality. Then we need to show that $S_n \to 0$ as $n \to \infty$:

Now

$$|S_n| = |S_n - A(r) + A(r)| \leq |S_n - A(r)| + |A(r)| \quad (115)$$

$$\left|S_n - A(r)\right| = \left|\sum_{k=0}^{n} c_k - \sum_{k=0}^{\infty} c_k\, r^k\right| = \left|\sum_{k=0}^{n} c_k(1-r^k) - \sum_{k=n+1}^{\infty} c_k\, r^k\right|$$
$$\leq \sum_{k=0}^{n} |c_k|\,(1-r^k) + \sum_{k=n+1}^{\infty} |c_k|\, r^k \quad (116)$$

Now if we let $r = 1 - \frac{1}{n}$ then $r \to 1$ as $n \to \infty$. We can estimate $1 - r^k$ as follows:

$$1 - r^k = (1-r)(1 + r + r^2 + \cdots + r^{k-1}) \leq k(1-r) = \frac{k}{n} \quad (117)$$

Now because of the Tauberian condition that $k\, c_k \to 0$, for any $\epsilon > 0$, we can find an N such that $k\,|c_k| < \frac{\epsilon}{2}$ for all $k > N$.

Thus:

$$\sum_{k=0}^{n} |c_k|\,(1-r^k) = \sum_{k=0}^{N} |c_k|\,(1-r^k) + \sum_{k=N+1}^{n} |c_k|\,(1-r^k) < \sum_{k=0}^{N} \frac{k|c_k|}{n} + \sum_{k=N+1}^{n} \frac{k|c_k|}{n}$$
$$< \sum_{k=0}^{N} \frac{k|c_k|}{n} + \frac{n-N}{n}\,\frac{\epsilon}{2} < \sum_{k=0}^{N} \frac{k|c_k|}{n} + \frac{\epsilon}{2} \quad (118)$$

But $\sum_{k=0}^{N} \frac{k|c_k|}{n}$ can be made less than $\frac{\epsilon}{2}$ for n sufficiently large since the $k|c_k|$ are bounded for $k = 0, 1, \ldots N$. Thus $\sum_{k=0}^{n} |c_k|\,(1-r^k) < \epsilon$.

Using the Tauberian condition in the final sum in (116) we see that:

$$\sum_{k=n+1}^{\infty} |c_k|\, r^k \leq \frac{\epsilon}{2n} \sum_{k=n+1}^{\infty} r^k = \frac{\epsilon}{2n}\,\frac{r^{n+1}}{1-r} < \frac{\epsilon}{2n(1-r)} < \frac{\epsilon}{2} \quad (119)$$

since $k|c_k| < \frac{\epsilon}{2}$ for all $k > n$ implies $|c_k| < \frac{\epsilon}{2k} < \frac{\epsilon}{2n}$ and $1 - r = \frac{1}{n}$.

Thus from (116) we can see that for sufficiently large n , $\left|S_n - A(r)\right| < \epsilon$ and we know that $|A(r)| < \frac{\epsilon}{2}$ by virtue of the hypothesis of Abel summability to 0. Thus using (115), $S_n \to 0$ ie $\sum c_n \to 0$

It is worth noting that without the Tauberian condition $\sum_{k=0}^{n} \frac{k|c_k|}{n}$ is not necessarily small. Hardy and Littlewood developed theorems which placed various conditions on the c_k to generalise the basic Tauberian condition. For instance, one theorem runs like this:

If $\sum c_n$ is a series of positive terms such that as $n \to \infty$, $\lambda_n = c_1 + c_2 + \cdots + c_n \to \infty$, and $\frac{c_n}{\lambda_n} \to 0$, and $\sum a_n e^{-\lambda_n x} \to s$ as $x \to 0$ and $a_n = \mathcal{O}(\frac{\lambda_n - \lambda_{n-1}}{\lambda_n})$ then $\sum a_n$ is convergent to s.

BIBLIOGRAPHY

[Bressoud] David Bressoud, "A Radical Approach to Real Analysis", Second Edition, The Mathematical Association of America, 2007

[HardyPM] G H Hardy "A Course of Pure Mathematics", Cambridge University Press, 2006

[HardyDS] G H Hardy "Divergent Series" , AMS Chelsea Publishing, 1991

[Jackson] John David Jackson, "Classical Electrodynamics", Wiley, Third Edition, 1999

[SteinShakarchi] Elias M Stein, Rami Shakarchi, "Fourier Analysis: An Introduction", Princeton University Press, 2003

7 History

Created 16 September 2012
24 May 2019: updated graph for extended interval $[-\pi, \pi]$ rather than just $[0, \pi]$
14 June 2025 - I noticed a bad cut and paste from a previous step in an inequality in section 6.1 which should have read: Thus if $x_n = \mathcal{O}(\frac{1}{n})$ it follows that for all n beyond any such large N, $\left|\frac{x_1+x_2+\cdots+x_N+x_{N+1}+\cdots+x_n}{n}\right| \leq \left|\frac{x_1+x_2+\cdots+x_N}{n}\right| + \left|\frac{x_{N+1}+x_{N+2}+\cdots+x_n}{n}\right| \leq \frac{|x_1|+|x_2|+\cdots+|x_N|}{n} + \frac{|x_{N+1}|+|x_{N+2}|+\cdots+|x_n|}{n} \leq \frac{|x_1|+|x_2|+\cdots+|x_N|}{n} + \frac{(n-N)\frac{A}{n}}{n}$ which can be made arbitrarily small for n sufficiently large. Thus the sequence converges to 0.