Fourier analysis

Fourier analysis started even before Fourier, but in the context of Fourier's work on the *Analytic Theory of Heat*, it began with a claim regarding expansions of functions as trigonometrical series, what we now call Fourier series. Such expansions form the framework for writing down solutions to the heat equation in terms of initial temperature distributions.

[refer to Enrique A. Gonzalez-Velasco The American Mathematical Monthly, Vol. 99, No. 5 (May, 1992), pp. 427-441 doi:10.2307/2325087]

We will not follow the historic development of Fourier series. Instead, we will begin with a finite version of the Fourier transform, the so-called *discrete Fourier transform*, then proceed to discuss Fourier series expansions of periodic functions and Fourier transforms of functions defined on the whole real line as limiting cases. As a note regarding terminology, the word *discrete* when applied to Fourier transforms sometimes applies to the finite case, but it also sometimes applies to the case of sequences defined on the integers. Here we will always use the term *discrete* in reference to the finite case only.

2.1 Discrete Fourier Transform

The discrete Fourier transform (DFT) is the $N \times N$ matrix $F = F_N$ such that $F_{jk} = \frac{1}{\sqrt{N}} e^{2\pi i(j-1)(k-1)}$ with $i = \sqrt{-1}$. Here are the matrices for the cases $N = 2, 3, 4$:

$$F_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \qquad F_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{8\pi i/3} \end{pmatrix}; \qquad F_4 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

The matrix F_N is unitary, that is, the $\mathbb{C}^N$ inner product of the j-th and k-th columns $\mathbf{c}_j$ and $\mathbf{c}_k$ of F_N is

$$\langle \mathbf{c}_j, \mathbf{c}_k \rangle = \frac{1}{N} \sum_{\ell=0}^{N-1} e^{2\pi i j\ell/N} e^{-2\pi i k\ell/N} = \frac{1}{N} \sum_{\ell=0}^{N-1} e^{2\pi i(j-k)\ell/N} = \delta_{jk} \tag{2.1}$$

2.1.1 Complex exponentials

Points (x, y) in the plane can be expressed in polar coordinates by $(r\cos\theta,\, r\sin\theta)$ where $r = \sqrt{x^2+y^2}$ and $\theta = \arctan(y/x)$. In other words, θ is the angle that z makes with the positive x-axis. Points on the unit circle then have the form $(\cos\theta,\, \sin\theta)$. Regarding (x, y) as a complex number $z = x + iy$, $i = \sqrt{-1}$, we can then write a point on the unit circle as $\cos\theta + i\sin\theta$. The Taylor series expansion $\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ can be extended to complex z to define $\mathrm{e}^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$. In particular,

$$\begin{aligned} \mathrm{e}^{i\theta} &= \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} = 1 + i\theta - \frac{\theta^2}{2} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \cdots \\ &= \big(1 - \frac{\theta^2}{2} + \frac{\theta^4}{4!} + \cdots\big) + i\big(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots\big) \\ &= \cos\theta + i\sin\theta \end{aligned}$$

by referring to the Taylor series expansions of $\sin\theta$ and $\cos\theta$. We can also justify the *law of exponents*

$$\mathrm{e}^{i\theta}\,\mathrm{e}^{i\phi} = \mathrm{e}^{i(\theta+\phi)}$$

by invoking the sum-angle identities

$$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$$
$$\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$$

so that

$$\begin{aligned}\cos(\theta+\phi) + i\sin(\theta+\phi) &= \cos\theta\cos\phi - \sin\theta\sin\phi + i(\sin\theta\cos\phi + \cos\theta\sin\phi)\\ &= (\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi).\end{aligned}$$

This tells us that we can multiply complex numbers in their polar forms $z = r\mathrm{e}^{i\theta}$ and $w = s\mathrm{e}^{i\phi}$ by $zw = rs\mathrm{e}^{i(\theta+\phi)}$. Since multiplication by $\mathrm{e}^{i\theta}$ adds an angle θ in polar coordinates we can regard multiplication by $\mathrm{e}^{i\theta}$ as *rotation through an angle* θ. Now we are in a position to be more precise about the orthogonality of the complex exponentials.

2.1.2 Unitarity of F_N

The last identity was proved at the end of Chapter 1 but here is another look from a slightly different point of view.

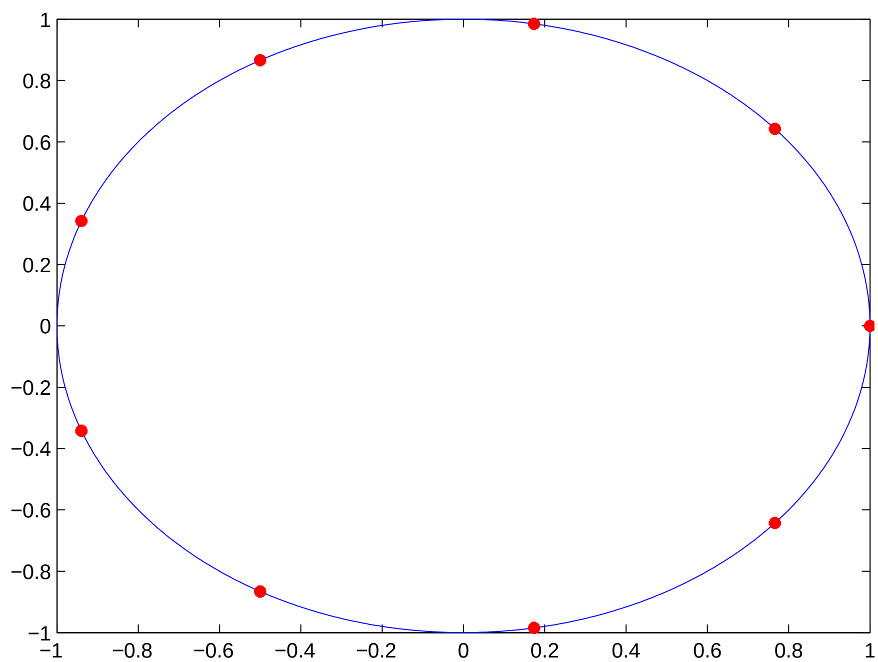

Fig. 2.1. Ninth roots of unity.

The exponential sum in (2.1) is invariant under rotation by $e^{2\pi ij/N}$, in other words,

$$e^{2\pi ij/N}\sum_{\ell=0}^{N-1} e^{2\pi ij\ell/N} = \sum_{\ell=0}^{N-1} e^{2\pi ij/N}\,e^{2\pi ij\ell/N} = \sum_{\ell=0}^{N-1} e^{2\pi ij(\ell+1)/N} = \sum_{\ell=1}^{N} e^{2\pi ij\ell/N} = \sum_{\ell=0}^{N-1} e^{2\pi ij\ell/N}$$

where the last sneaky fact can be visualized in Figure 2.1: the last sum above starts with $(1,0)$ and adds up the coordinates of the dots as one proceeds around the circle. The penultimate sum also adds the same set of numbers, it just starts at $e^{2\pi ij/N}$ and goes all the way around to $e^{2\pi iN/N} = 1 + 0i$. Seen in another way, multiplication by $e^{2\pi ij/N}$ rotates each of the roots of unity j times. So the whole set of roots is preserved under this multiplication and, when one sums them up, one gets the same result as if one had not rotated them. Thus, if $S = \sum_{\ell=0}^{N-1} e^{2\pi ij\ell/N}$ then $S = e^{2\pi ij/N} S$ which is a rotation of the complex number S through $2\pi j/N$. If $0 < j < N$ this can only happen if $S = 0$ so this proves that the sum of the N-roots of unity is zero.

2.2 Fast Fourier transform

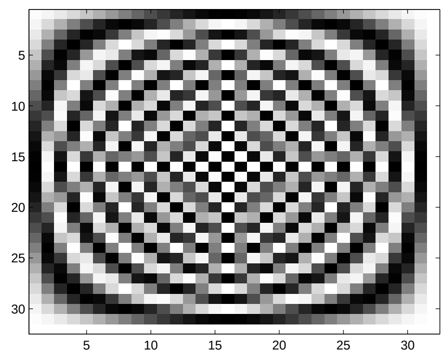

Fig. 2.2. Real part of DFT matrix for $N = 32$

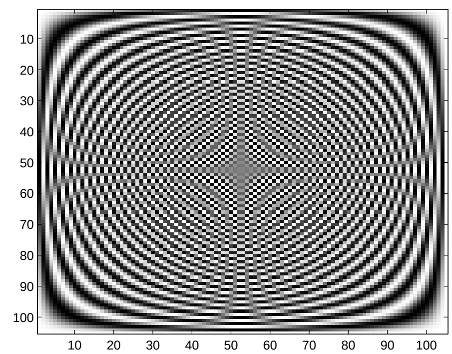

Fig. 2.3. Real part of DFT matrix for $N = 105$

The Fast Fourier Transform refers to an algorithm for implementing the DFT matrix. Ordinary matrix multiplication requires N^2 multiplications along with several additions, but it is the multiplications that are the computationally intensive part in floating point arithmetic. When $N = 2^L$ is a power of two, the computational complexity of the FFT is only $O(N \log N)$. The algorithm for implementing the FFT was discovered by John Tukey and James Cooley. There are now many variations of the basic algorithm for different specialized purposes. In the Cooley-Tukey algorithm a vector $\mathbf{x} = (x_0, x_1, \dots x_{2^L-1})$ is first split into its *even* part $\mathbf{x}_{\mathrm{e}} = (x_0, x_2, \dots x_{2^L-2})$ and *odd* part $\mathbf{x}_{\mathrm{o}} = (x_1, x_3, \dots x_{2^L-1})$ which are now both elements of $\mathbb{C}^{N/2}$ and their DFTs can be computed using the $\frac{N}{2} \times \frac{N}{2}$ dimensional Fourier matrix. Letting $\mathbf{X}$, $\mathbf{X}_{\mathrm{e}}$ and $\mathbf{X}_{\mathrm{o}}$ denote the respective Fourier transforms one then has, setting $M = N/2$,

$$
\begin{aligned}
\mathbf{X}_k &= \sum_{\ell=0}^{(N/2)-1} x_{2\ell}\, \mathrm{e}^{-2\pi i(2\ell)k/N} + x_{2\ell+1}\, \mathrm{e}^{-2\pi i(2\ell+1)k/N} \\
&= \sum_{\ell=0}^{(M)-1} x_{\mathrm{e},\ell}\, \mathrm{e}^{-2\pi i\ell k/M} + \mathrm{e}^{-\pi ik/M} x_{\mathrm{o},\ell}\, \mathrm{e}^{-2\pi i\ell k/M} \\
&= X_{\mathrm{e},k \bmod N/2} + \mathrm{e}^{-2\pi i(k \bmod N/2)/N} X_{\mathrm{o},k \bmod N/2}
\end{aligned}
$$

in which $k \bmod N/2 = k$ if $k < N/2$ and $k \bmod N/2 = k - (N/2)$ if $N/2 \le k < N$. It is important for this purpose that we index the entries of $\mathbf{x}$ by the index set $\{0, 1, \dots, N-1\}$. In this way we have expressed the computation of the $N \times N$ DFT matrix as two DFTs of half the size – which is possible when N is even – followed by multiplication of the even components by the identity and of the odd components by the diagonal matrix $\operatorname{diag} \mathrm{e}^{-2\pi i(k \bmod N/2)/N}$. In effect one replaces the N^2 multiplications required for computing the DFT by $2(N/2)^2 + 4N$ multiplications to compute the half size DFTs plus diagonal terms. If N is a power of two one can proceed to replace the $(N/2)^2$ multiplications by $2(N/4)^2$ and so on until one gets down to 2×2 matrices. In the end one has to compute $2^L = N$ 2×2 DFTs, together with a series of diagonal matrices.

Exercise 2.2.1. Write the 16×16 DFT matrix explicitly as a product of block diagonal matrices and discuss how to generalize this to arbitrary powers of 2.

2.3 Fourier series

The discrete Fourier transform corresponds to finite, uniformly spaced samples of the periodic functions $e^{2\pi int}$ where the samples are taken at the points $t = k/N$ and n runs from $n = 0$ to $n = N - 1$. Because of periodicity, we get the same values for $n = N + 1$ that we do for $n = 1$. When we let N tend to infinity we obtain the functions $e^{2\pi int}$ as limits, taken in the sense of uniform limits of periodic interpolants of the DFT vectors on the real line as illustrated in Figure 2.4.

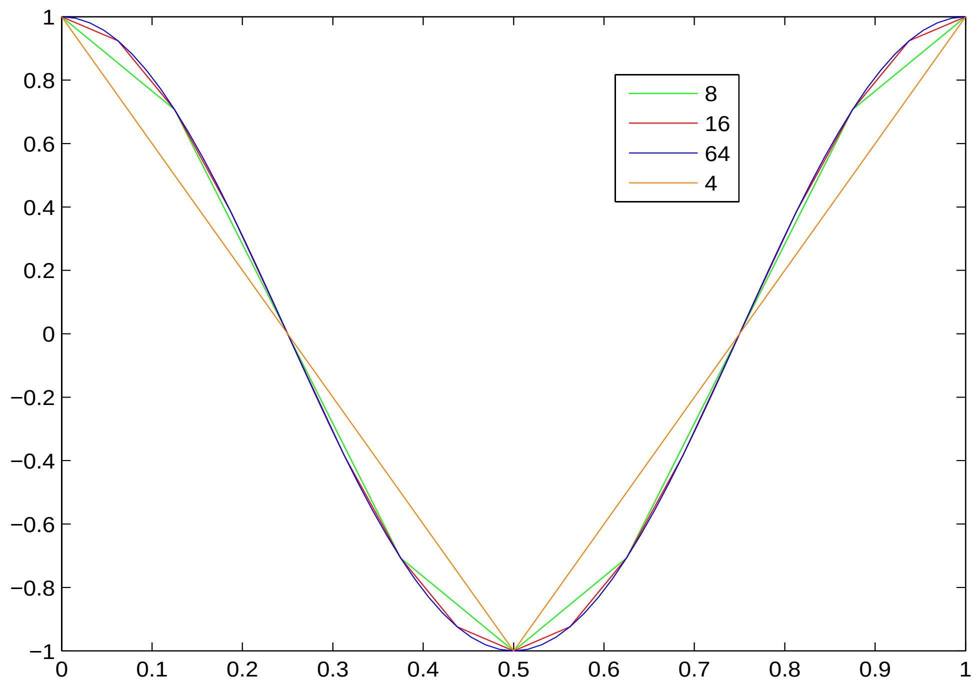

Fig. 2.4. Linear interpolants converging to the real part of $e^{2\pi it}$ on $[0, 1)$.

The orthogonality property is preserved in the sense that, if $E_{N,j}$ denotes the N-th column of the unitary Fourier matrix, $E_{N,j} = \frac{1}{\sqrt{N}}(1, e^{2\pi ij/N}, \ldots, e^{2\pi ij(N-1)/N})$ then

$$\lim_{N\to\infty} \langle E_{N,j},\, E_{N,k}\rangle_{\mathbb{C}^N} = \int_0^1 e^{2\pi ijt} e^{-2\pi ikt}\, dt = \delta_{jk}. \tag{2.2}$$

This limit interprets the finite dimensional inner products as uniform Riemann sum approximations of the limiting case.

What is crucial for us is that periodic functions, expressed in a suitable way as limits of interpolants of their sample values, can also be written as *infinite series* of exponentials.

To make sense of this, let $x(t)$ be a 1-periodic, complex-valued function of t, and let $\mathbf{x}_N$ denote the sample sequence $\mathbf{x}_N = (x(0), x(1/N), \ldots, x((N-1)/N))$ obtained by sampling $x(t)$ along the partition points k/N. In this sense, the inner product of $\mathbf{x}_N$ with the j-th row $\mathbf{f}_j^T$ of F_N, suitably normalized, becomes a Riemann approximation of the j-th *Fourier coefficient* $\hat{x}[j]$ of $x(t)$. That is,

$$\frac{1}{\sqrt{N}}\langle \mathbf{x}_N,\, \mathbf{f}_j\rangle = \frac{1}{N}\sum_{\ell=0}^{N} x(\ell/N) e^{-2\pi ij\ell/N} \approx \int_0^1 x(t)\, e^{-2\pi ijt}\, dt \equiv \hat{x}[j]$$

On the other hand, the reconstruction of $\mathbf{x}_N$ from its DFT becomes an approximation of the Fourier expansion of $x(t)$ at the given sample points. That is,

$$x(k/N) = \mathbf{x}_N(k) = (F_N^*(F_N\mathbf{x}_N))(k) = \sum_{j=0}^{N-1} \widehat{\mathbf{x}_N}(j) e^{2\pi ij(k/N)} \approx \sum_{j=-\infty}^{\infty} \hat{x}[j] e^{2\pi ij(k/N)} = \sum_{j=-\infty}^{\infty} \hat{x}[j] e^{2\pi ijt}\Big|_{t=k/N}.$$

In this we we can say that the DFT of the samples of $x(t)$ and its inversion, for large N, provides an approximation of the Fourier coefficients of $x(t)$ and its Fourier series representation. Thus, given a one-periodic function on the real line, we define

$$S(x)(t) = \sum_{k=-\infty}^{\infty} \hat{x}[k]\mathrm{e}^{2\pi ikt}. \tag{2.3}$$

The question of when and in what sense this series *converges to* x will be considered momentarily.

2.3.1 Centering the approximation

One point that could use some clarification is that of how the Fourier basis elements are ordered. The integers have a natural ordering by magnitude and sign, but $\mathbb{Z}_N$, the group of integers modulo N, does not have any natural ordering. In order to make the DFT matrix look more like the *bi-infinite matrix* $\mathrm{e}^{2\pi ikt}$ where $k \in \mathbb{Z}$ one can use the so-called *centered* DFT matrix (CDFT). The rows and columns of CDFT are obtained by replacing j (row number) and k (column number) by $j-(N-1)/2$ and $k-N/2$, respectively, noting that $j-(N-1)/2 \equiv (j+(N+1)/2) \mod N$. In other words, all we have done is rearrange the rows and columns of the DFT matrix. Observe that when $N=3$ we have:

$$F_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 & 1 \\ 1 & \mathrm{e}^{2\pi i/3} & \mathrm{e}^{4\pi i/3} \\ 1 & \mathrm{e}^{4\pi i/3} & \mathrm{e}^{8\pi i/3} \end{pmatrix} \mapsto CF_N = \frac{1}{\sqrt{3}}\begin{pmatrix} \mathrm{e}^{8\pi i/3} & 1 & \mathrm{e}^{4\pi i/3} \\ 1 & 1 & 1 \\ \mathrm{e}^{4\pi i/3} & 1 & \mathrm{e}^{2\pi i/3} \end{pmatrix} = \frac{1}{\sqrt{3}}\begin{pmatrix} \mathrm{e}^{-4\pi i/3} & 1 & \mathrm{e}^{4\pi i/3} \\ 1 & 1 & 1 \\ \mathrm{e}^{-2\pi i/3} & 1 & \mathrm{e}^{2\pi i/3} \end{pmatrix}$$

The reason for taking N to be odd is that there is no *middle* column when N is even. The reason for shifting the rows as well as the columns is so that the inverse CDFT will also be centered.

2.4 Convergence of Fourier series and $L^2(\mathbb{T})$

Fourier got into hot water with Laplace and Lagrange because of his implicit claim that any function could be expanded in a Fourier series. This claim turns out to be true, at least for the right class of functions, but its proof is far from simple and the whole question of what the statement means was not really resolved until the early 1960s – around the same time as the FFT algorithm was discovered – when Carleson proved that the Fourier series of a function $f \in L^2(\mathbb{T})$ converges to f at *almost every point* where the terms *almost every* and *converge* are given precise meanings that only a true mathematician can appreciate.

2.5 Convergence of Fourier series

In this section we will consider a simpler property, namely that if $f \in L^2(\mathbb{T})$ then the Fourier series

$$S(f)(t) = \sum_{k=-\infty}^{\infty} \hat{f}[k]\mathrm{e}^{2\pi ikt}$$

converges to f in L^2-norm, that is,

$$\lim_{M,N\to\infty} \sum_{k=-M}^{N} \|f - \hat{f}[k]\mathrm{e}^{2\pi ikt}\|_{L^2(\mathbb{T})} = 0.$$

2.5.1 Trigonometric polynomials

A trigonometric polynomial has the form $p(t) = \sum_{k=-M}^{N} c_k \mathrm{e}^{2\pi ikt}$ where $c_k \in \mathbb{Z}$. The terminology comes from the fact that the values of the *Laurent polynomial* $p(z) = \sum_{k=-M}^{N} c_k z^k$ of the complex variable z is being restricted to the unit circle $|z|=1$ whose elements are parameterized by $z = \mathrm{e}^{2\pi it}$, $0 \le t < 1$. Because the functions $\mathrm{e}^{2\pi ikt}$ form an orthonormal family with respect to the inner product $\langle p, q\rangle$ on $L^2(\mathbb{T})$ (see 2.2), it follows that for $p(t) = \sum_{k=-M}^{N} c_k \mathrm{e}^{2\pi ikt}$, the coefficient c_k is given by $c_k = \langle p, \mathrm{e}^{2\pi ikt}\rangle$. The trigonometric polynomials form a linear subspace of $L^2(\mathbb{T})$. In addition, the orthogonality of the functions $\mathrm{e}^{2\pi ikt}$ gives us the following.

Lemma 2.5.1. *If $p(t) = \sum_{k=-M}^{N} c_k e^{2\pi ikt}$ is a trigonometric polynomial then $p(t)$ is equal to its Fourier series, that is, $c_k = \langle p, e^{2\pi ikt}\rangle = \hat{p}[k]$.*

Proof. Because of the orthogonality of the functions $e^{2\pi ikt}$ it follows from linearity of the inner product that

$$\hat{p}[k] \equiv \langle p, e^{2\pi ikt}\rangle = \sum_{k=-M}^{N} c_\ell \langle e^{2\pi i\ell t}, e^{2\pi ikt}\rangle = \sum_{k=-M}^{N} c_\ell \delta_{k\ell} = c_k.$$

Since the exponentials are orthogonal, they are also linearly independent so the Fourier coefficients of p are unique. Therefore $p(t) = \sum_{k=-M}^{N} \hat{p}[k]e^{2\pi ikt}$ which was to be shown.

Proposition 2.5.2. *(Baby Parseval) Suppose that f and g are trigonometric polynomials. Then*

$$\langle f,\, g\rangle_{L^2(\mathbb{T})} = \langle \{\hat{f}[k],\}, \{\hat{g}[k]\}\rangle_{\ell^2(\mathbb{Z})}.$$

We can assume that f, g are trigonometric polynomials of degree at most N. Then

$$\begin{aligned}\langle f,\, g\rangle_{L^2(\mathbb{T})} &= \int_0^1 f(t)\bar{g}(t)\,dt = \int_0^1 \sum_{k=-N}^{N} \hat{f}[k]e^{2\pi ikt} \sum_{\ell=-N}^{N} \bar{\hat{g}}[\ell]e^{-2\pi i\ell t}\,dt \\ &= \sum_{k,\ell=-N}^{N} \hat{f}[k]\bar{\hat{g}}[\ell] \int_0^1 e^{2\pi i(k-\ell)t}\,dt \\ &= \sum_{k,\ell=-N}^{N} \hat{f}[k]\bar{\hat{g}}[\ell]\delta_{k,\ell} \\ &= \sum_{k=-N}^{N} \hat{f}[k]\bar{\hat{g}}[k] = \langle \{\hat{f}[k],\}, \{\hat{g}[k]\}\rangle_{\ell^2(\mathbb{Z})}.\end{aligned}$$

Decay of Fourier coefficients and Fourier transforms

Smoothness of a periodic function is reflected by decay of its Fourier coefficients. Suppose that $f \in L^2(\mathbb{T})$ is continuously differentiable. Then f' is continuous, hence in $L^2(\mathbb{T})$ and its k-th Fourier coefficient is

$$\begin{aligned}\widehat{f'}[k] = \int_0^1 f'(t)\,e^{-2\pi ikt}\,dt &= [f(t)e^{-2\pi ikt}|_0^1 - \frac{1}{-2\pi ik}\int_0^1 f(t)\,e^{-2\pi ikt}\,dt \\ &= \frac{1}{2\pi ik}\hat{f}[k].\end{aligned}$$

Remark 2.5.3. By reiterating the integration by parts, it is possible to show that if $f \in C(\mathbb{T})$ has p continuous derivatives then the Fourier coefficients of f satisfy $|\hat{f}[k]| \le C/|k|^p$.

2.5.2 Convolution and Fourier partial sums

The *circular convolution* of a pair of one-periodic functions f, g is defined by

$$f * g(t) = \int_{\mathbb{T}} f(t-s)\,g(s)\,ds = \int_{0\le s<t} f(t-s)\,g(s)\,ds + \int_{t\le s<1} f(1+t-s)\,g(s)\,ds.$$

It is called circular convolution because the points s, t are regarded as points on the one-dimensional torus, so their values are taken modulo one.

Exercise 2.5.4. Show that circular convolution is commutative, that is, $f * g = g * f$.

Define the *Dirichlet kernel* $D_N(t) = \sum_{k=-N}^{N} \mathrm{e}^{2\pi ikt}$. Then D_N is real-valued and we can define the N-th partial sum of the Fourier series of f by

$$\begin{aligned} S_N f(t) = f * D_N(t) &= \int f(s) D_N(t-s)\, ds \\ &= \sum_{k=-N}^{N} \int_0^1 f(s)\mathrm{e}^{2\pi ik(t-s)}\, ds = \sum_{k=-N}^{N} \mathrm{e}^{2\pi ikt} \int_0^1 f(s)\mathrm{e}^{-2\pi iks}\, ds \\ &= \sum_{k=-N}^{N} \hat{f}[k]\mathrm{e}^{2\pi ikt}. \end{aligned}$$

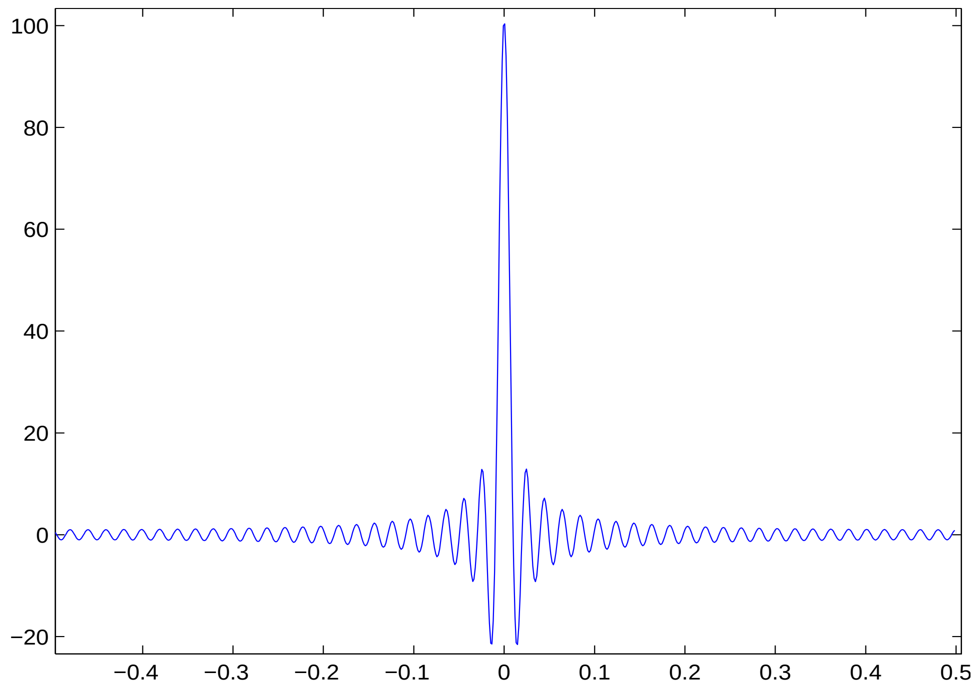

Fig. 2.5. Dirichlet function D_N for $N = 50$.

2.5.3 Convergence of partial sums for nice *f*

Up to this point we have been beating around the bush a little. In analysis, at some point one has to get a little dirty. Now is the time for that. We want to show that if f is well-behaved, say continuously differentiable, then $S_N f(t) \to f(t)$ uniformly at every point. To do so we need to use properties of $D_N(t)$. The main ideas of the argument will be provided, with some of the details of the estimates left to the reader.

Proposition 2.5.5. *For all* N, $\int_0^1 D_N(s)\, ds = 1$. *Fix* $\delta > 0$. *Then for every* $\epsilon > 0$ *there exists* M *such that, whenever* $N > M$ *we have* $|\int_{\delta<s<1-\delta} D_N(s)| < \epsilon$ *whenever* $\delta < |s| < 1$.

Proof. The first statement follows since $\int_0^1 D_N(s)\, ds = \sum_{k=-N}^{N} \int_0^1 \mathrm{e}^{2\pi iks}\, ds = \sum_{k=-N}^{N} \delta_{k0} = 1$. Next, by the geometric series formula,

$$\begin{aligned} D_N(x) = \sum_{k=-N}^{N} \mathrm{e}^{2\pi ikx} &= \mathrm{e}^{-2\pi iNx}\sum_{k=0}^{2N}\mathrm{e}^{2\pi ikx} \\ &= \mathrm{e}^{-2\pi iNx}\frac{1-\mathrm{e}^{2\pi i(2N+1)x}}{1-\mathrm{e}^{2\pi ix}} \\ &= \frac{\mathrm{e}^{\pi ix}}{\mathrm{e}^{\pi ix}}\frac{\mathrm{e}^{-(2N+1)\pi ix}-\mathrm{e}^{(2N+1)\pi ix}}{\mathrm{e}^{-\pi ix}-\mathrm{e}^{\pi ix}} = \frac{\sin(2N+1)\pi x}{\sin\pi x}. \end{aligned}$$

This confirms that $D_N(x) \to 2N+1$ as $x \to 0$. The estimate on the integral is a little trickier. On $[0,1)$, D_N has zeros when $(2N+1)\pi t = k\pi$ or when $t = k/(2N+1)$. The denominator has little change between consecutive pairs of zeros. In particular,

$$\begin{aligned} \int_{k/(2N+1)}^{(k+2)/(2N+1)} \frac{\sin(2N+1)\pi t}{\sin\pi t}\,dt &= \int_{k/(2N+1)}^{(k+2)/(2N+1)} \left(\frac{\sin(2N+1)\pi t}{\sin\pi t} - \frac{\sin(2N+1)\pi t}{\sin\pi k/(2N+1)}\right)dt \\ &\quad + \int_{k/(2N+1)}^{(k+2)/(2N+1)} \frac{\sin(2N+1)\pi t}{\sin\pi k/(2N+1)}\,dt. \end{aligned}$$

The denominator in the second integral is a nonzero constant and the numerator is integrated over one period so the second integral is zero. The first integral can be rewritten

$$\int_{k/(2N+1)}^{(k+2)/(2N+1)} \sin(2N+1)\pi t\left(\frac{\sin\pi k/(2N+1)-\sin\pi t}{\sin\pi t\sin\pi k/(2N+1)}\right)dt$$

Using the fact that $\sin x \geq 2x/\pi$ when $0 \leq x \leq \pi/2$ and the fact that the derivative of $\sin x$ is bounded, the integrand is at most a fixed multiple of N/k^2 while the interval of integration essentially has length one, so this integral is at most a multiple of $1/k^2$. The idea then is to take $\delta \sim 1/(N\epsilon)$. When adding these integrals from $k \sim N\delta$ up to, say, $k \sim N-1$ one is effectively summing $1/k^2$ for these k to get a sum bounded by a multiple of $1/k = 1/(N\delta) = \epsilon$. This proves the second statement.

Exercise 2.5.6. Give a rigorous account of the estimates just outlined above.

Theorem 2.5.7. *If $f \in C^1(\mathbb{T})$ then $D_N * f \to f$ uniformly as $N \to \infty$.*

Proof. Using periodicity one can express circular convolution as an integral over $[-1/2, 1/2$. Then

$$\begin{aligned} \Big|D_N * f(x) - f(x)\Big| &= \Big|\int_{-1/2}^{1/2} f(x-t)D_N(t) - f(x)\int_{-1/2}^{1/2} D_N(t)\,dt\Big| \\ &\leq \Big|\int_{|t|<\delta}(f(x-t)-f(x))D_N(t)\,dt\Big| + \Big|\int_{1/2\geq|t|>\delta}(f(x-t)-f(x))D_N(t)\,dt\Big|. \end{aligned}$$

For $f \in C^1(\mathbb{T})$ one has

$$\int_{|t|<\delta}(f(x-t)-f(x))D_N(t)\,dt\Big| \leq \sup_{|t|<\delta}|f(x-t)-f(x)|\int_{|t|<\delta}|D_N| \leq C\delta\log N\sup|f'(x)|$$

while

$$\Big|\int_{1/2\geq|t|>\delta}(f(x-t)-f(x))D_N(t)\,dt\Big| \leq C\frac{\log N}{N\delta}\sup|f(x)| < \epsilon\sup|f(x)|$$

where δ was chosen so that $\int_{|t|>\delta}\big|D_N\big| < \epsilon$. As above, we could take $\delta \sim 1/(N\epsilon)$ which is small for N large enough. Together these estimates show that, for any $\epsilon > 0$ fixed, $|D_N * f(x) - f(x)| < \epsilon$ if N is large enough. Thus, $D_N f \to f$ uniformly.

Exercise 2.5.8. Verify that if $\delta \sim \log N/(N\epsilon)$ then $\int_{1/2\geq|t|>\delta}|D_N(t)|\,dt < \epsilon$ for N large enough.

This leads to Parseval's theorem for teenagers.

Theorem 2.5.9. *(Parseval for teenagers) Suppose that f and g are in $C^1(\mathbb{T})$. Then*

$$\langle f,\, g\rangle_{L^2(\mathbb{T})} = \langle \{\hat{f}[k],\}, \{\hat{g}[k]\}\rangle_{\ell^2(\mathbb{Z})}. \tag{2.4}$$

Proof. By polarization it suffices to show that if $f \in C^1(\mathbb{T})$ then

$$\|f\|_2^2 = \sum_{k\in\mathbb{Z}} |\hat{f}[k]|^2.$$

By Bessel's inequality it is enough to show that, for every $\epsilon > 0$,

$$\|f\|_2 \le \|\{\hat{f}[k]\}\|_{\ell^2(\mathbb{Z})} + \epsilon.$$

But if $f \in C^1(\mathbb{T})$ then for N large enough we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} |\hat{f}[k]|^2 \ge \sum_{k=-N}^{N} |\hat{f}[k]|^2 &= \|f * D_N\|_2^2 \\ &= \|f - (f - f * D_N)\|_2^2 \ge \big(\|f\|_2 - \|f * D_N - f\|_2\big)^2 \\ &\ge \big(\|f\|_2 - \epsilon\big)^2\end{aligned}$$

since $\|g\|_2 \le \sup|g(t)|$ and by Theorem 2.5.7. Taking square roots gives the desired estimate. This proves Parseval for teenagers.

What we are really after is Parseval's theorem for grown ups.

Theorem 2.5.10. *(Parseval for grown ups) Suppose that f and g are in $L^2(\mathbb{T})$. Then*

$$\langle f,\, g\rangle_{L^2(\mathbb{T})} = \langle \{\hat{f}[k],\}, \{\hat{g}[k]\}\rangle_{\ell^2(\mathbb{Z})}. \tag{2.5}$$

First we should observe that if $f \in L^2(\mathbb{T})$ then $f \in L^1(\mathbb{T})$ so its coefficients $\hat{f}[k]$ are well-defined because

$$|\hat{f}[k]| = \Big|\int_0^1 f(t)\, \mathrm{e}^{2\pi i k t}\, dt\Big| \le \int_0^1 \Big|f(t)\, \mathrm{e}^{2\pi i k t}\Big|\, dt = \int_0^1 \Big|f(t)\Big|\, dt.$$

We cannot show the grown up version in its entirety because we need to use the following result, whose proof requires fundamentals of *measure theory* that go beyond the scope of this text.

Proposition 2.5.11. *$C^1(\mathbb{T})$ is dense in $L^2(\mathbb{T})$.*

Nevertheless, assuming this proposition we can argue as follows. Fix $\epsilon > 0$ and choose $g \in C^1(\mathbb{T})$ such that $\|f - g\|_2 < \epsilon/2$. By Bessel's inequality, $\sum_{k=-\infty}^{\infty} |\hat{f}[k] - \hat{g}[k]|^2 < \epsilon^2/4$. Then

$$\begin{aligned}\|\{\hat{f}[k]\}\|_{\ell^2(\mathbb{Z})} &\ge \|\{\hat{g}[k]\}\|_{\ell^2(\mathbb{Z})} - \|\{\hat{f}[k] - \hat{g}[k]\}\|_{\ell^2(\mathbb{Z})} \\ &= \|g\|_{\ell^2(\mathbb{Z})} - \|\{\hat{f}[k] - \hat{g}[k]\}\|_{\ell^2(\mathbb{Z})} \\ &\ge \|g\|_{L^2(\mathbb{T})} - \epsilon/2 \\ &= \|f - (f-g)\|_{L^2(\mathbb{T})} - \epsilon/2 \\ &\ge \|f\|_{L^2(\mathbb{T})} - \|(f-g)\|_{L^2(\mathbb{T})} - \epsilon/2 \ge \|f\|_{L^2(\mathbb{T})} - \epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary Plancherel's theorem for grown-ups follows from this and Bessel's inequality.

Remark 2.5.12. The essential property of $L^2(\mathbb{T})$ that we have used here is that $L^2(\mathbb{T})$ is complete: any Cauchy sequence in $L^2(\mathbb{T})$ converges to a limit in $L^2(\mathbb{T})$. Because of Bessel's inequality the trigonometric polynomials $S_N f(t) = \sum_{k=-N}^{N} \hat{f}[k]\, \mathrm{e}^{2\pi i k t}$ form a Cauchy sequence in $L^2(\mathbb{T})$ and so they have a limit in $L^2(\mathbb{T})$. The next step is to guarantee that the limit is f and not some other function. This is where we made use of pointwise convergence on C^1 functions.

As a corollary we have

Corollary 2.5.13. *Trigonometric polynomials are dense in* $L^2(\mathbb{T})$. *In particular, if* $f \in L^2(\mathbb{T})$ *then* $\|D_N * f - f\|_2 \to 0$ *as* $N \to \infty$.

Proof. By Plancherel's formula we have

$$\|f - D_N * f\|_2^2 = \sum_{|k|>N} |\hat{f}[k]|^2 \to 0$$

since $\sum_{k=-\infty}^{\infty} |\hat{f}[k]|^2 < \infty$.

Exercise 2.5.14. Write a matlab script for computing the partial sums $D_N * f$ for sample vectors with an arbitrary number of terms/samples as input parameters. Compute the partial sum estimates for sample sequences of the following, thought of as restrictions to $[0,1]$ of periodic functions. (1) $|\sin 2\pi t|$, (2) $T(t) = (2 - 2|t - 1/2|)$, (2) $\mathbb{1}_{[1/4,3/4]}$.

2.6 Fourier transforms and $L^2(\mathbb{R})$

Fourier series arise as limits of discrete Fourier representations of sample approximations. Here one takes N uniformly spaced samples on $[0,1)$ and lets $N \to \infty$. Any periodic function of period P is a dilation of a periodic function of period one. In this case, one can write the Fourier series of a periodic function of period P as

$$f = \frac{1}{P} \sum_{k=-\infty}^{\infty} \langle f, \mathrm{e}^{2\pi ikt/P} \rangle \, \mathrm{e}^{2\pi ikt/P}$$

with convergence in $L^2([-P/2, P/2])$. Now suppose that f is any square integrable function that vanishes outside of some bounded interval. Then, for any fixed value of t we can think of the sum on the right as a Riemann approximation of

$$f(t) = \int_{-\infty}^{\infty} \langle f(s), \mathrm{e}^{2\pi is\xi} \rangle \, \mathrm{e}^{2\pi it\xi} \tag{2.6}$$

evaluated at the sample points $\xi = k/P$. In the limit as $P \to \infty$ one obtains the coefficient mapping

$$f \mapsto \int_{-\infty}^{\infty} f(t) \, \mathrm{e}^{-2\pi it\xi} \, dt = \hat{f}(\xi)$$

called the *Fourier transform* of f. Equation 2.6 is called the Fourier inversion formula. The limiting argument outlined above is the sort of reasoning that Fourier had in mind when defining the Fourier transform and the Fourier representation of a function defined on all of $\mathbb{R}$. However, the sense in which one can interpret this formula involves serious mathematical issues such as the nontrivial question of when the integral actually makes sense.

2.6.1 Properties of Fourier transforms

In what follows suppose that $f(t)$ is absolutely integrable on $\mathbb{R}$ and define its Fourier transform

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(t) \, \mathrm{e}^{-2\pi it\xi} \, dt.$$

Exercise 2.6.1. Show that $\hat{f}$ is a continuous function of ξ.

Exercise 2.6.2. Show that if $T_a f(t) = f(t-a)$ then $\widehat{T_a f}(\xi) = \mathrm{e}^{-2\pi ia\xi} \hat{f}(\xi)$.

Exercise 2.6.3. Show that if $M_b f(t) = \mathrm{e}^{2\pi ibt} f(t)$ then $\widehat{M_b f}(\xi) = \hat{f}(\xi - b)$.

Exercise 2.6.4. Show that if $D_\lambda f(t) = \frac{1}{\lambda} f\left(\frac{t}{\lambda}\right)$ then $\widehat{D_\lambda f}(\xi) = \hat{f}(\lambda\xi)$.

Exercise 2.6.5. Show that if $\tilde{f}(t) = f(-t)$ then $\widehat{\tilde{f}}(\xi) = \overline{\hat{f}(\xi)}$.

Exercise 2.6.6. Show that if f is absolutely integrable and has a derivative $\frac{d}{dt}f$ that is also absolutely integrable then $\widehat{\frac{d}{dt}f}(\xi) = -2\pi i\xi \hat{f}(\xi)$.

Exercise 2.6.7. The convolution of two absolutely integrable functions f and g is defined as

$$f * g(t) = \int_{-\infty}^{\infty} f(t-s)\, g(s)\, ds.$$

Show that $f * g(t) = g * f(t)$, that $\|f * g\|_{L^1} \le \|f\|_{L^1}\|g\|_{L^1}$ (where $\|f\|_{L^1} = \int |f|$) and show that $\widehat{f * g}(\xi) = \hat{f}(\xi)\, \hat{g}(\xi)$.

Exercise 2.6.8. Explain why, if $f \in L^1(\mathbb{R})$, one must have $|\hat{f}(\xi)| \to 0$ as $|\xi| \to \infty$.

Exercise 2.6.9. Consider the integral function $G(z) = \int_{-\infty}^{\infty} \mathrm{e}^{-t^2+itz}\, dt$. Find an ODE governing G and find its solution with initial value. Use this to compute the Fourier transform of $f(t) = \mathrm{e}^{-\pi t^2}$.

Exercise 2.6.10. Use symmetry and integration by parts to compute the Fourier transform of the function $f(t) = \mathrm{e}^{-|t|}$.

2.6.2 Continuous functions and $L^2(\mathbb{R})$

We would like to prove a version of Plancherel's theorem for $L^2(\mathbb{R})$. We will outline the argument leaving the details as a list of exercises. We denote by $C_c(\mathbb{R})$ the space of continuous functions that vanish when $|x|$ is large enough.

Exercise 2.6.11. Suppose that $f \in C_c(\mathbb{R})$. Let $G(t) = \mathrm{e}^{-\pi t^2}$. Show that $\int G = 1$. Also show that the Fourier inversion formula (2.6) is valid for $f * G_s$ where $G_s(t) = (G(t/s)/s$ where $s > 0$.

Exercise 2.6.12. Suppose that $f \in C_c(\mathbb{R})$. Let $G(t) = \mathrm{e}^{-\pi t^2}$. Show that $\int G = 1$. Also show that the Fourier inversion formula (2.6) is valid for $f * G_s$ where $G_s(t) = (G(t/s)/s$ where $s > 0$.

Exercise 2.6.13. Prove the Fourier uniqueness theorem for $f \in C_c(\mathbb{R})$. That is, if $\hat{f}(\xi) = 0$ for all ξ then $f \equiv 0$.

Exercise 2.6.14. (Baby Plancherel) Prove that if $f \in C_c(\mathbb{R})$ then

$$\int_{-\infty}^{\infty} |G_s * f(t)|^2\, dt = \int_{-\infty}^{\infty} |\widehat{G}(s\xi)\hat{f}(\xi))|^2\, d\xi$$

Exercise 2.6.15. Prove that if $f \in C_c(\mathbb{R})$ then $f * G_s \to f$ uniformly.

Exercise 2.6.16. (Plancherel for teenagers) Prove that if $f \in C_c(\mathbb{R})$ then

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt = \int_{-\infty}^{\infty} |\hat{f}(\xi))|^2\, d\xi$$

Theorem 2.6.17. *Continuous, compactly supported functions are dense in $L^2(\mathbb{R})$. That is, if $f \in L^2(\mathbb{R})$ and $\epsilon > 0$ is fixed then there is a $\varphi \in C_c(\mathbb{R})$ such that $\|f - \varphi\|_2 < \epsilon$.*

Exercise 2.6.18. (Plancherel for grownups) Prove that if $f \in L^2(\mathbb{R})$ then

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt = \int_{-\infty}^{\infty} |\hat{f}(\xi))|^2\, d\xi$$

Your proof should involve a reasonable and valid definition of $\hat{f}$ for any $f \in L^2(\mathbb{R})$.

Exercise 2.6.19. (Parseval for grownups) Prove that if $f \in L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R})$ then

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(t)\, \bar{g}(t)\, dt = \int_{-\infty}^{\infty} \hat{f}(\xi)\, \overline{\hat{g}(\xi)}\, d\xi.$$

Your proof should involve a reasonable and valid definition of $\hat{f}$ for any $f \in L^2(\mathbb{R})$.

Exercise 2.6.20. Suppose that φ is any continuous, compactly supported function in $\mathbb{R}$ such that $\int \varphi = 1$. Prove that if f is any function in $L^2(\mathbb{R})$ then $\|f * \varphi_s - f\|_2 \to 0$ as $s \to 0$ where $\varphi_s(x) = \varphi(x/s)/s$ for $s > 0$.

2.6.3 Convergence at jumps and Gibbs phenomenon

If f is an otherwise continuous periodic function with an isolated jump discontinuity then its Fourier series converges to the average height at the jump. This is an old theorem that is illustrated in the Figures below. On the other hand, the series does not converge in a particularly nice way because the partial sums always exhibit a phenomenon called *Gibbs overshoot*. Figure 2.6 illustrates a 4096-point sampled version of the function $\mathbb{1}_{[1/2,1)}$ regarded as a periodic function restricted to $[0,1)$. Figure 2.7 illustrates the Gibbs phenomenon for an approximate reconstruction from 1the first 28 DFT terms. Figure 2.8 illustrates the propogation of Gibbs phenomenon when one uses more terms (128 terms in blue and 256 terms in red). The jumps get localized closer to the jump as one takes more terms, but the *magnitude* of the jump does not get any smaller. This does not cause too much trouble for *mean squared error* estimates because the area under the overshoot gets smaller as the overshoot region narrows when taking more terms. However, from the standpoint of perceptual information this error does create a nuisance and points to one source of error when applying Fourier methods for compression of images that contain sharp edges.

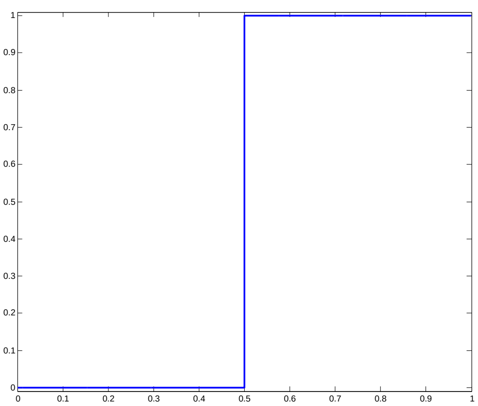

Fig. 2.6. A function having a simple jump discontinuity

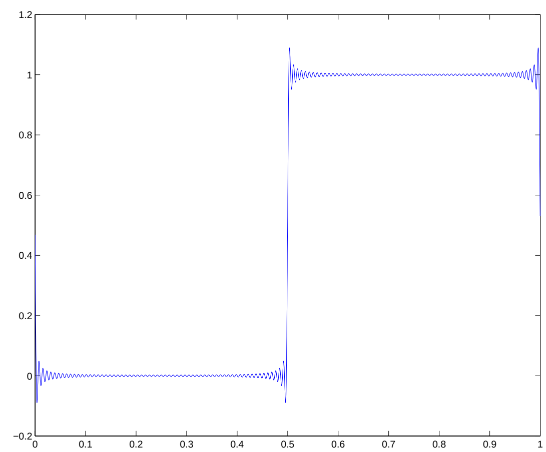

Fig. 2.7. Illustration of Gibbs overshoot for a simple jump.

Multidimensional Fourier transforms

Multidimensional versions of all of the Fourier transforms (finite, discrete/periodic and real-variable) exist in higher dimensions. Some of the properties such as Plancherel's theorem generalize in straightforward, if not trivial ways to several variables. Other properties such as pointwise convergence of Fourier series are far more complicated in several variables because, for example, the definition of a partial sum with respect to pairs of integers is not straightforward. Here we will just stick to definitions and a couple of brief observations.

Fourier transforms of functions on $\mathbb{R}^N$

Suppose that $f = f(t_1, \dots, t_n$ is a function defined and integrable over $\mathbb{R}^n$. Then one can define its Fourier transform $\hat{f}(\xi) = \hat{f}(\xi_1, \dots, \xi_n)$

$$\hat{f}(\xi) = \int \cdots \int f(t_1, \dots, t_n)\, \mathrm{e}^{-2\pi i(t_1\xi_1 + \cdots + t_n\xi_n)}\, dt_1 \cdots dt_n = \int_{\mathbb{R}^n} f(t)\, \mathrm{e}^{-2\pi i t\cdot\xi}\, dt$$

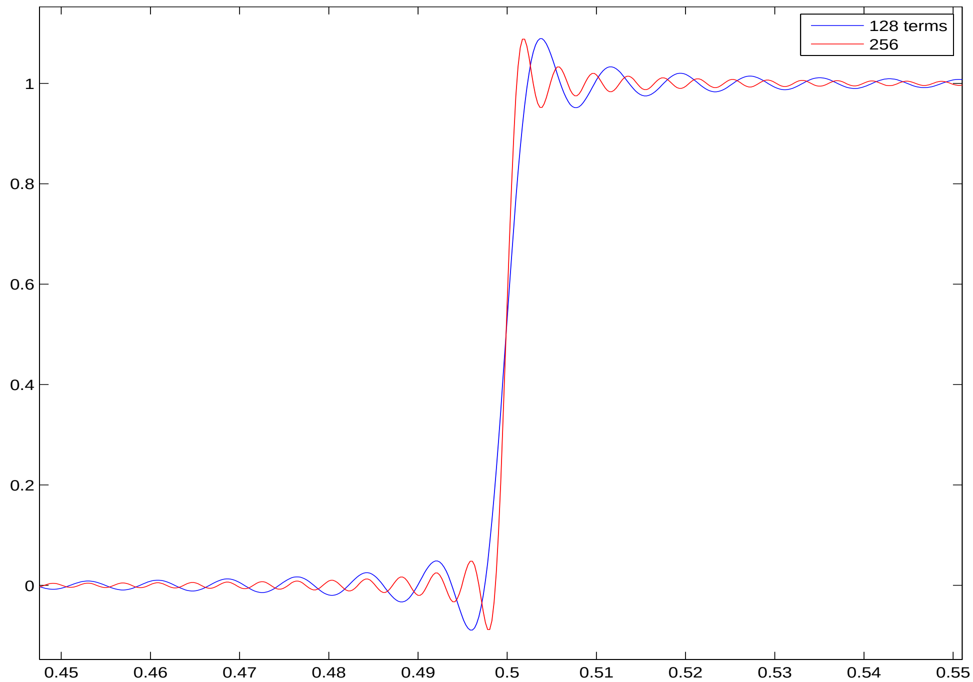

Fig. 2.8. Zoom in on 128 and 256 point partial sums neat $t = 1/2$

The inverse Fourier transform of $g(\xi)$ is

$$g^{\vee}(t) = \int_{\mathbb{R}^n} g(\xi)\, \mathrm{e}^{2\pi i t\cdot\xi}\, d\xi.$$

The Plancherel and Parseval theorems for $\mathbb{R}^n$ are the same as those for $\mathbb{R}$ except with $\mathbb{R}$ replaced by $\mathbb{R}^n$.

Exercise 2.6.21. Show that if $f \in L^1(\mathbb{R}^n$ has the form $f(t) = f_1(t_1)f_2(t_2)\cdots f_n(t_n)$. Show that $\hat{f}(\xi) = \hat{f}_1(t_1)\cdots\hat{f}_n(t_n)$. Compute the Fourier transform of $f(t_1, t_2) = \mathbb{1}_{[-1/2,1/2]}(t_1)\mathbb{1}_{[-1/2,1/2]}(t_2)$

Exercise 2.6.22. Use polar coordinates to compute the Fourier transform in $\mathbb{R}^2$ of $\mathbb{1}_{|t|\leq 1}$ where $\{|t| \leq 1\}$ is the closed unit disc of radius one centered at the origin.

Fourier series functions on $\mathbb{T}^N$ and sequences in $\mathbb{Z}^N$

A $\mathbf{v}$-periodic function of n variables $\mathbf{t} = (t_1, \dots, t_n)$ is a function satisfying $f(\mathbf{t} + v_i) = \mathbf{t}$ for all $t \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, \dots, v_n)$. When $\mathbf{v} = (1, 1, \dots, 1)$ if f is $(1, 1, \dots, 1)$-periodic then f can be thought of as a function on $[0, 1) \times \cdots \times [0, 1) = [0, 1)^n = \mathbb{T}^n$. A $(1, 1, \dots, 1)$-periodic function has Fourier coefficients

$$\hat{f}[k] = \int_0^1 \cdots \int_0^1 f(t_1, \dots, t_n)\, \mathrm{e}^{-2\pi i t_1 k_1 + \cdots t_n k_n}\, dt_1 \cdots dt_n = \int_{\mathbb{T}^n} f(t)\, \mathrm{e}^{-2\pi i t\cdot k}\, dt$$

where $k = (k_1, \dots, k_n) \in \mathbb{Z}^n$. On the other hand, if $\mathbf{c} = \{c_k\}$ is a sequence defined for $k = (k_1, \dots, k_n) \in \mathbb{Z}^n$ then

$$\mathcal{F}(\mathbf{c})(\xi) = \sum_{k_1} = 1^n \cdots \sum_{k_n=-\infty}^{\infty} c_{(k_1,\dots,k_n)}\, \mathrm{e}^{2-\pi i k_1\xi_1 + \cdots + k_n\xi_n} = \sum_{k\in\mathbb{Z}^n} c_k\, \mathrm{e}^{-2\pi i k\cdot\xi}$$

and the inverse Fourier transform of a sequence $\mathbf{d} = \{d_k\}$ is $\mathcal{F}^{-1}(\mathbf{d})(t) = \sum_k d_k\, \mathrm{e}^{2\pi i k\cdot t}$.

Fourier transforms of vectors $\mathbb{C}^N$

In the same way we can define a DFT for sequence arrays. We consider here just the two-dimensional case since this will be the one important for image processing. If $\mathbf{x} = \{(x_{j,k}\}_{j,k=1}^{M,N}$ then

$$\mathcal{F}(\mathbf{x})(\ell, m) = \frac{1}{\sqrt{MN}} \sum_{j=1}^{M} \sum_{k=1}^{N} x_{j,k} \mathrm{e}^{-2\pi i(j\ell/M)+(km/N)}.$$

The inverse 2-D DFT is defined similarly with '$-$' replaced by '$+$' in the exponential.

2-D FFT

There is a two-dimensional FFT algorithm just as with the 1-D case. The most basic version simply performs a one-dimensional transform of each row and then uses the resulting row outputs as inputs into a column transform. If $K = MN$ then each row transform is $(N \log N$ so to transform all of the rows requires $MN \log N$ operations. Similarly, to transform all of the resulting columns requires $MN \log M$ operations. Importantly, one has to add these numbers of operations to get the total operation count which, since $\log a + \log b = \log ab$, is $O(K \log K)$. There are variations of this approach that are more suitable for dealing with high dimensional data or very large matrix inputs that require *out-of-core* storage of intermediate values.

2.7 Separation of variables and the heat equation

Fourier employed developments in trigonometric series to solve the heat equation. Consider a circular ring that is perfectly insulated in the sense that no heat can escape outside the ring, but so that heat can flow around the ring at a rate depending on conductive properties of the ring. Parameterize the cross sections of the ring by values $x \in [0, 1)$ taken modulo one so that $1 = 0$. Then the temperature $u(x, t)$ at the x-th cross section of the ring at time $t > 0$ can be regarded as a periodic function of x. By empirical methods one can show as Fourier first did that the temperature obeys the heat equation

$$\frac{\partial u}{\partial t} = -\kappa^2 \frac{\partial^2 u}{\partial x}^2. \tag{2.7}$$

One solves the heat equation using the method of *separation of variables.* That is, one first finds solutions of the form $\psi(x, t) = \phi(x)T(t)$. By (2.7) it must be the case that

$$\frac{1}{T}\frac{dT}{dt} = \kappa^2 \frac{\partial^2 \phi}{\partial \phi^2}.$$

Since one side depends only on t and the other only on x both sides must equal the same constant λ^2. Then $T(t) = \mathrm{e}^{-\lambda^2\kappa^2}t$ and $\varphi(x) = \mathrm{e}^{i\kappa\lambda x}$ or really its real part. Since φ has to be one-periodic it follows that $\kappa\lambda$ is an integer multiple of 2π so $\varphi(x) = \varphi_n(x) = \mathrm{e}^{2\pi i n x}$ and then $T(t) = T_n(t) = \mathrm{e}^{-(2n\pi)^2 t}$ and $\psi_n = \mathrm{e}^{2\pi i n x - 4\pi^2 n^2 t}$. Next one observes that any superposition of such separated solutions is a solution so we can express u as

$$u(x, t) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{2\pi i n x - 4\pi^2 n^2 t}.$$

If we observe that u should be real-valued then the c_n have to work out so that we can write

$$u(x, t) = a_0 + \sum_{n=1}^{\infty} \big(a_n \cos 2\pi n x + b_n \sin 2\pi n x\big) \mathrm{e}^{-4\pi^2 n^2 t}.$$

At time zero one then has the expansion

$$u(x, 0) = a_0 + \sum_{n=1}^{\infty} \big(a_n \cos 2\pi n x + b_n \sin 2\pi n x\big).$$

Fourier's claim was that such an expansion could be used to represent an arbitrary initial cross-sectional temperature distribution.

2.8 Multipliers

Convolution operators arise in all sorts of ways in harmonic analysis. One of the classic ways in which they arise is as differentiation which, in the language of distributions, is convolution with the derivative of a Dirac delta mass. Integration by parts allows one to represent differentiation in the time domain on $\mathbb{R}$ in terms of multiplication in the frequency domain. That is, assuming that f' *vanishes at infinity* in a suitable sense,

$$\mathcal{F}(f')(\xi) = \int f'(t)\, \mathrm{e}^{-2\pi i t\xi}\, dt = \int f(t)\, \frac{d}{dt}\mathrm{e}^{-2\pi i t\xi}\, dt = -2\pi i\xi \hat{f}(\xi). \tag{2.8}$$

In this way, if $P(d/dx)$ is a polynomial $a_0 + a_1(d/dx) + \cdots + a_N(d/dx)^N$ in d/dx then

$$P(d/dx)f)^\wedge(\xi) = P(-2\pi i\xi)\hat{f}(\xi).$$

This observation allows one to use the Fourier transform to solve an ODE of the form $P(d/dx)f = g$ by

$$g \mapsto \hat{g} \mapsto \hat{g}/P(-2\pi i\xi) \mapsto (\hat{g}/P(-2\pi i\xi))^\vee = f.$$

In general one can solve the convolution equation $f * Q = g$ by setting $f = (\hat{q}/\hat{Q})^\vee)$. The operator $M_Q f = f * Q$ is called a convolution operator or *multiplier* since its action is given by multiplication in the Fourier domain.

2.8.1 Deconvolution and applications

Often convolution operators are associated with real physical measurements that generate errors in a predictable way. In particular, measurements usually involve a sort of averaging process in which inputs at different spatial or temporal locations relative to a fixed location contribute with specific weights. These weights are encoded in what is known in optics as a *point spread function* so named for image processing applications such astronomical observations because it determines the image generated when a single intense point is placed at the origin. Figure 2.9 represents a numeric Gaussian point spread function in a 64×64 matrix. Figure 2.12 is a cartoon stick figure image to which the convolution by the Gaussian point spread function will be applied. Here the convolution is accomplished numerically by multiplying the corresponding DFTs. The blurred stick figure is shown in Figure **??**. One then deconvolves by dividing by an approximation of the inverse multiplier. On the Fourier side, one divides by the DFT of the PSF plus some small epsilon that is input for stability purposes. This epsilon prevents one from dividing by zero, which would cause serious errors. On the other hand, one still sees a lot of aliasing in the deconvolved image. This combination of aliasing and instability can cause major limitations of the deconvolution method, even when one has a very good estimate of the PSF. Later we will consider methods that are better adapted spatially.

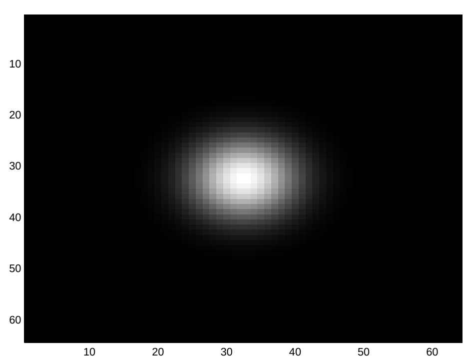

Fig. 2.9. A Gaussian point spread function

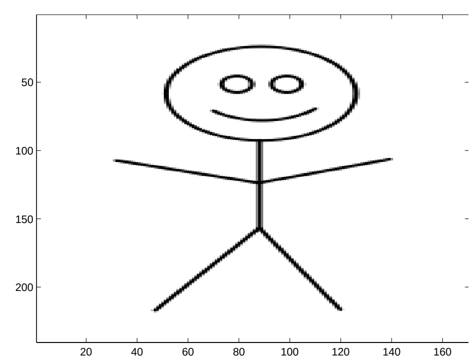

Fig. 2.10. Original stick figure image

The algorithm just outlined uses two-variable Fourier transforms, in this case the matlab `fft2` algorithm. One begins with an image X and a point spread function P, also regarded as an image matrix. The blurring is then accomplished by the convolution

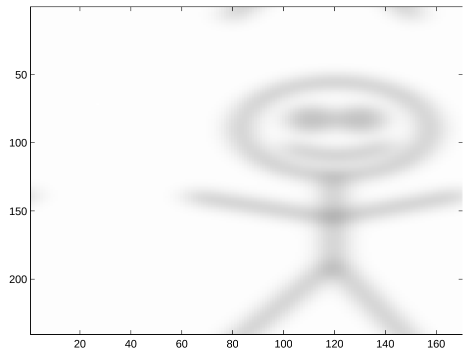

Fig. 2.11. Blurred stick figure

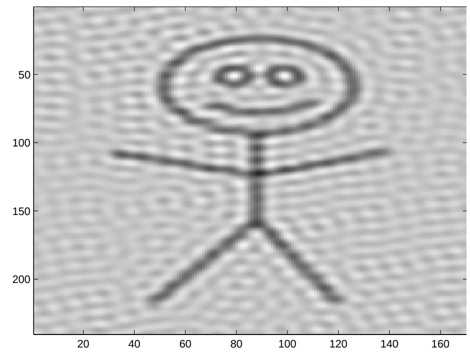

Fig. 2.12. Deconvolved blurred stick figure

$$X \mapsto \texttt{ifft2}\,(\texttt{fft2}\,X .* \texttt{fft2}\,P)$$

where `ifft2` is the two-variable inverse DFT and ".∗" denotes coordinatewise multiplication. The deconvolution is accomplished in a similar way. Instead of multiplying in the Fourier domain by $\texttt{fft2}\,P$ one divides by $\texttt{fft2}\,P + \epsilon$ where ϵ is chosen to avoid a divide by zero error.

Exercise 2.8.1. Reproduce a similar illustration of blurring and deconvolution using a point spread function and an image of your choice. See if you can improve the deblurring by varying the type of regularization (addition of ϵ in our case. Make sure that you use the same size FFT on the image and the PSF so that the pointwise multiplication of DFTs is defined.

2.9 Bandlimited functions on $\mathbb{R}$

2.9.1 Why are signals bandlimited

When we say that a signal is bandlimited we mean that its frequency spectrum is contained in a compact interval, or that it contains no component oscillating above a fixed frequency. There are good reasons for these assumptions. In the case of a transmitted signal, the transmission medium often kills oscillations beyond a given frequency. In the case of mammalian auditory systems, the system does not respond to energy in frequencies beyond a certain band (20 kHz for most humans) and so one might as well treat the signal as if it were bandlimited. Of course there are errors that result from doing so. We will not go into the analysis of such errors here systematically. But in the notes and exercises for this chapter we will consider some consequences of the bandlimiting hypothesis.

2.9.2 The Paley-Wiener space

A function $f \in L^1(\mathbb{R})$ is said to be Ω-bandlimited if its Fourier transform $\hat{f}$ vanishes outside of $[-\Omega/2, \Omega/2]$ (some authors use $[-\Omega, \Omega]$ instead – our convention refers to the length of the interval on which $\hat{f}$ lives). Many analytic properties of the space of bandlimited functions were developed by Paley and Wiener [**?**] and so we refer to the subspace of $L^2(\mathbb{R})$ of those f bandlimited to $[-\Omega/2, \Omega/2]$ as the *Paley-Wiener space* PW_Ω. When $\Omega = 1$ we just refer to the space as PW. On the other hand, if $E \subset \mathbb{R}$ is closed then PW_E refers to those $f \in L^2(\mathbb{R})$ such that $\hat{f} = 0$ outside of E. Before proceeding we want to recall some basic properties of the Fourier transform. First, suppose $f \in L^2(\mathbb{R})$ is bandlimited to an interval $[\alpha, \beta]$, that is, $\hat{f}(\xi) = 0$ if $\xi \notin [\alpha, \beta]$. Set $\lambda = (\beta - \alpha)$ and $\mu = (\alpha + \beta)/(2\lambda)$ and let $f_{\lambda\mu} = \frac{1}{\sqrt{\lambda}} \mathrm{e}^{-2\pi i \mu t} f\left(\frac{t}{\lambda}\right)$. Then

$$\begin{aligned}
\widehat{f_{\lambda\mu}}(\xi) &= \frac{1}{\sqrt{\lambda}} \int \mathrm{e}^{-2\pi i \mu t} f\Big(\frac{t}{\lambda}\Big) \mathrm{e}^{-2\pi i t \xi}\, dt \\
&= \sqrt{\lambda} \int \mathrm{e}^{-2\pi i \mu \lambda u} f(u) \mathrm{e}^{-2\pi i \lambda u \xi}\, dt \\
&= \sqrt{\lambda} \int \mathrm{e}^{-2\pi i \mu \lambda u} f(u) \mathrm{e}^{-2\pi i u \lambda (\xi + \mu)}\, dt = \sqrt{\lambda} \hat{f}(\lambda(\xi + \mu).
\end{aligned}$$

Now if $\lambda(\xi+\mu)$ is in the support $[\alpha,\beta]$ of $\hat{f}$ then $\alpha \le \lambda(\xi+\mu) \le \beta$ or $\alpha/\lambda - \mu \le \xi \le \beta/\lambda - \mu$. Since $\mu = (\alpha+\beta)/(2\lambda)$ this says that $\alpha/\lambda-(\alpha+\beta)/(2\lambda) \le \xi \le \beta/\lambda-(\alpha+\beta)/(2\lambda)$ or $(\alpha-\beta)/(2\lambda) \le \xi \le (\beta-\alpha)/(2\lambda)$ and since $\lambda = (\beta-\alpha)$ this says that $-1/2 \le \xi \le 1/2$. This says that the mapping $f \mapsto f_{\lambda\mu}$ is a unitary isomorphism from $\mathrm{PW}_{[\alpha,\beta]}$ onto PW. Thus, up to modulations and dilations, $\mathrm{PW}_{[\alpha,\beta]}$ is really just PW and if we want to understand the former we just need to understand the latter.

Exercise 2.9.1. Show that the zeros of a function $f \in \mathrm{PW}$ must be isolated.

2.9.3 The classical sampling theorem of Shannon, Whittaker, ...

Suppose that $f \in \mathrm{PW}$. Consider the inverse Fourier transform of $\hat{f}$ which is defined as

$$\begin{aligned} f(x) &= \int_{-1/2}^{1/2} \hat{f}(\xi)\mathrm{e}^{2\pi i x\xi}\, dx \\ &= \int_{-1/2}^{1/2}\int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{-2\pi i t\xi}\, dt\mathrm{e}^{2\pi i x\xi}\, dx \\ &= \int_{-\infty}^{\infty} f(t)\int_{-1/2}^{1/2} \mathrm{e}^{-2\pi i(t-x)\xi}\, dx\, dt \\ &= \int_{-\infty}^{\infty} f(t)\frac{\sin\pi(t-x)}{\pi(t-x)}\, dx\, dt \end{aligned}$$

Exercise 2.9.2. Show that $\int_{-1/2}^{1/2} \mathrm{e}^{2\pi i t\xi}\, dt = \sin\pi t/(\pi t)$.

Recall that the functions $\mathrm{e}^{2\pi i k\xi}$ form an orthonormal basis for $L^2([-1/2,1/2])$ as k ranges over the integers. Suppose now that $f \in \mathrm{PW}$. Then $\hat{f} \in L^2([-1/2,1/2])$ and we can write $\hat{f}$ on $[-1/2,1/2]$ in terms of its Fourier series

$$\hat{f}(\xi) = \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{2\pi i k\xi}$$

where $c_k = \int_{-1/2}^{1/2} \hat{f}(\eta)\mathrm{e}^{-2\pi i k\eta}$. By the Fourier inversion formula as we just saw, $c_k = f(-k)$. Therefore we can write $\hat{f}(\xi) = \sum_{k=-\infty}^{\infty} f(k)\mathrm{e}^{-2\pi i k\xi}$ with convergence in $L^2([-1/2,1/2])$. Taking the inverse Fourier transform of both sides we then see, from the calculation above, that

$$f(x) = \sum_{k=-\infty}^{\infty} f(k)\frac{\sin\pi(x-k)}{\pi(x-k)}. \tag{2.9}$$

By Plancherel's theorem the series converges in $L^2(\mathbb{R})$. Equation 2.9 is known as the classical sampling formula. The result has been known in one form or another for hundreds of years, but it is referred to, most commonly, as *Shannon's sampling formula* because Claude Shannon 1949 [**?**] interpreted the formula in a useful way from the point of signal processing and information theory. Similar mathematical treatments and, to a lesser degree, physical interpretations, had been carried out earlier by V. A. Kotelnikov in 1933 [**?**], by E.T. Whittaker in 1915 [**?**] and by Gabor in 1946 [**?**].

Corollary 2.9.3. *The functions* $s_k(x) = \frac{\sin\pi(x-k)}{\pi(x-k)}$ *form an orthonormal basis for PW.*

The function $s(x) = \sin\pi x/(\pi x)$ is commonly referred to as the sinc function (pronounced "sink") and sometimes written as $\mathrm{sinc}\,(x)$. Shannon's theorem is useful because it says that any bandlimiting function can be "sinc interpolated" from its integer samples $\{f(k)\}$. By employing a dilation change of variables as above, Shannon's formula can be extended to the PW_Ω spaces, namely, if $f \in \mathrm{PW}_\Omega$ then we can write

$$f(x) = \frac{1}{\Omega}\sum_{k=-\infty}^{\infty} f\Big(\frac{k}{\Omega}\Big)\mathrm{sinc}\,(\Omega x - k) \tag{2.10}$$

with convergence in $L^2(\mathbb{R})$. If $\Omega > 1$ then we can still recover a signal in PW_Ω from regularly spaced samples, but the price to pay for allowing more bandwidth is that the samples have to be taken at a higher *rate* of Ω samples per unit time. The factor Ω is called the *Nyquist rate* (pron. [ny:kvist]), after Harry Nyquist [**?**] who determined in 1927 that the number of independent pulses that could be put through a telegraph channel per unit time is limited to twice the bandwidth of the channel. Shannon's theorem says, in essence, that the Nyquist rate can actually be achieved.

The primary limitation of the sampling theorem is that the interpolating sinc functions have poor localization properties. Specifically, $\operatorname{sinc} x$ only decays like $1/x$ for large values of x. As a consequence, the value of $f(x)$ is influenced by the samples $f(k)$ for a large number of values of k.

Exercise 2.9.4. Prove that the function $\operatorname{sinc} x \notin L^1(\mathbb{R})$, that is, $\int_{-A}^{A} |\operatorname{sinc} t|\, dt$ diverges as $A \to \infty$.

2.9.4 Bandlimited wavelets

The space PW of functions bandlimited to $[-1/2, 1/2]$ is a proper subspace of PW_2, the space of square-integrable functions bandlimited to $[-1, 1]$. In other words, Fourier transforms of elements of PW are otherwise arbitrary elements of $L^2([-1/2, 1/2])$ while the Fourier transforms of elements of PW_2 are otherwise arbitrary elements of $L^2([-1, 1])$. Imposing the condition that $f \in \mathrm{PW}_2 \ominus \mathrm{PW}$ – the orthogonal complement of PW inside PW_2 – is the same as imposing the condition that $\hat{f}$ is supported in $[-1, 1]$ and is orthogonal to all functions in L^2 that are supported in $[-1/2, 1/2]$. By the following exercise, it follows that $\hat{f}(\xi) = 0$ for $\xi \in [-1/2, 1/2]$.

Exercise 2.9.5. Show that if $[a, b] \subset [c, d]$ then $f \in L^2([c, d])$ is orthogonal to all elements of $L^2([c, d])$ that are supported in $[a, b]$ if and only if $f = 0$ on $[a, b]$.

Denote by Q the orthogonal complement of PW *inside* of PW_2, that is, $Q = \mathrm{PW}_2 \ominus \mathrm{PW}$.

Corollary 2.9.6. $Q = \mathrm{PW}_{[-1,-1/2]} \oplus \mathrm{PW}_{[1/2,1]}$

Now we recall that if $f \in \mathrm{PW}_{[\alpha,\beta]}$ then $f_{\lambda\mu} \in \mathrm{PW}$. Conversely, if we take $\nu = (\alpha + \beta)/2$, $\lambda = (\beta - \alpha)$ and $g(t) = \sqrt{\lambda} \mathrm{e}^{2\pi i \nu t} f(\lambda t)$ then $\hat{g}$ is supported in $[\alpha, \beta]$. For us, $\nu = 3/4$ and $\lambda = 1/2$ so any element of $\mathrm{PW}_{[1/2,1]}$ has the form $\mathrm{e}^{3\pi i t/2} f(t/2)$ and any element of $\mathrm{PW}_{[-1,-1/2]}$ has the form $\mathrm{e}^{-3\pi i t/2} h(t/2)$ for some f, h in PW.

This gives us one relationship between PW and the side spaces. We are more interested in a description that allows us to relate sample sequences across scales. One would like to think of Q as the *other half* of PW inside PW_2. Since PW is generated by integer shifts of the sinc function and PW_2 by half integer shifts of the dilated sinc function one might even conjecture that this other half might be generated by integer shifts of some other function $\psi \in Q \subset \mathrm{PW}_2$. That is, is there a function $\psi \in Q$ such that any $g \in Q$ can be expressed as $g = \sum_k d_k \psi(x - k)$? Just as the sinc function is the inverse Fourier transform of $\mathbb{1}_{[-1/2,1/2]}$ such a ψ should be the inverse function of some fundamental function that lives on $[-1, -1/2] \cup [1/2, 1]$.

Consider then the function $w_0(\xi) = \operatorname{sgn}(\xi) \mathrm{e}^{\pi i \xi} \mathbb{1}_{[1/2,1]}(|\xi|)$ and set $w_k(\xi) = \mathrm{e}^{-2\pi i k \xi} w_0(\xi)$. These functions live inside $[-1, -1/2] \cup [1/2, 1]$ and are orthonormal since

$$
\begin{aligned}
\langle w_k(\xi),\, w_\ell(\xi)\rangle &= \int \operatorname{sgn}(\xi)\mathrm{e}^{\pi i(1-2k)\xi}\mathbb{1}_{[1/2,1]}(|\xi|)\operatorname{sgn}(\xi)\overline{\mathrm{e}^{\pi i(1-2\ell)\xi}\mathbb{1}_{[1/2,1]}(|\xi|)}\, d\xi \\
&= \int_{-1}^{-1/2} + \int_{1/2}^{1} \mathrm{e}^{-2\pi i(k-\ell)\xi} d\xi \\
&= \int_{1/2}^{1} \big(\mathrm{e}^{-2\pi i(k-\ell)\xi} + \mathrm{e}^{2\pi i(k-\ell)\xi}\big) d\xi \\
&= \int_{1/2}^{1} 2\cos 2\pi(k-\ell)\xi\, d\xi
\end{aligned}
$$

by replacing $\xi \mapsto -\xi$ in the integral over $[-1, -1/2)$. The last integral is equal to one if $k = \ell$, while

$$
2\int_{1/2}^{1} \cos 2\pi(k-\ell)\xi d\xi = \frac{1}{\pi(k-\ell)} \big[\sin 2\pi(k-\ell)\big|_{1/2}^{1} = 0
$$

showing that the w_k are orthonormal.

Next we claim that any element of Q can be expressed as a linear combination of shifts of $\psi = w^\vee$. This is done by completing the steps of the following exercise.

Exercise 2.9.7. Explain why the Fourier transform of any element of $f \in Q$ can be written $\hat{f}(\xi) = \sum_{k\in\mathbb{Z}} \mathrm{e}^{4\pi ikx}\big(a_k \mathbb{1}_{[-1,-1/2]}(\xi)+b_k \mathbb{1}_{[1/2,1]}(\xi)\big)$. Show that if $\langle \hat{f},\, w_e ll\rangle = 0$ for all $\ell \in \mathbb{Z}$ then $\int_{1/2}^{1} b_k \mathrm{e}^{\pi i(4k+2\ell-1)\xi} - a_k \mathrm{e}^{-\pi i(4k+2\ell-1)\xi} = 0$ for all k,ℓ. Show that these equations imply that $a_k = b_k$ and $a_k = -b_k$ for all k.

Exercise 2.9.8. Show that the inverse Fourier transform is $\psi(t) = \eta(t+1/2)$ where $\eta(t) = (\cos 2\pi t - \cos \pi t)/(\pi i t)$ is plotted in Figure 2.13.

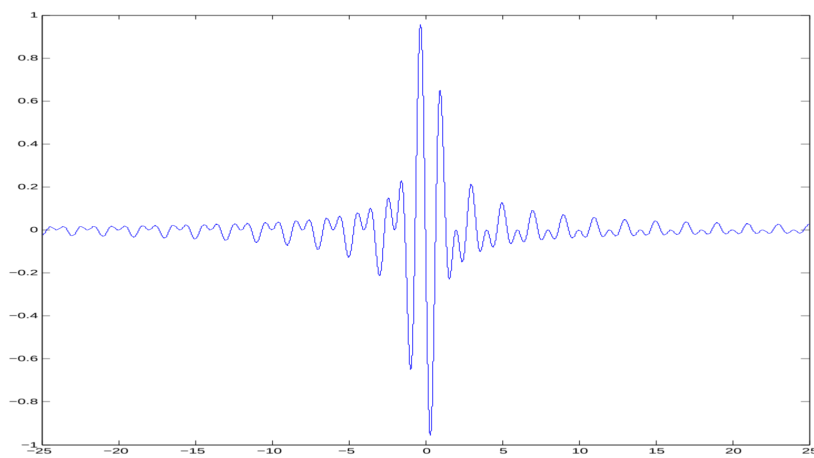

Fig. 2.13. Bandlimited wavelet function η

A scaling property of the sinc function

Since the functions $s_k = \mathrm{sinc}\,(x-k)$ form an ONB for PW, one can write the orthogonal projection onto PW as

$$f \mapsto \sum_k \langle f,\, s_k\rangle s_k.$$

The sampling theorem tells us that, when $f \in PW$, $\langle f,\, s_k\rangle = f(k)$. On the other hand, when $f \in \mathrm{PW}_2$, $f(x) = (1/2)\sum_\ell f(\ell/2)s(2x-\ell)$. Therefore,

$$\langle f,\, s_k\rangle = \frac{1}{2}\sum_\ell f\Big(\frac{\ell}{2}\Big)\langle s(2x-\ell),\, s(x-k)\rangle$$

and one can express the projection of PW_2 onto PW in terms of the coefficients $\langle s(2x-\ell),\, s_k(x)\rangle$.

It would be nice to get a discrete equation in terms of samples alone. This is where we use the fact that (i) PW $\subset \mathrm{PW}_2$ and (ii) the functions $\sqrt{2}\,\mathrm{sinc}\,(2x-k)$ form an orthonormal basis for PW_2. By Parseval's theorem,

$$\langle s(x),\, s(2x-k)\rangle = \Big\langle \hat{s},\, \frac{1}{2}\mathrm{e}^{-\pi ik\xi}\hat{s}\Big(\frac{\xi}{2}\Big)\Big\rangle = \frac{1}{2}\int_{-1/2}^{1/2} \mathrm{e}^{-\pi ik\xi}\, d\xi = \frac{\sin\frac{\pi k}{2}}{\pi k}$$

and, in general,

$$\langle s(2x-\ell),\, s(x-k)\rangle = \frac{(-1)^k \sin\pi\frac{\ell}{2}}{\pi(\ell-2k)}$$

with the understanding that this quantity becomes 1/2 when $\ell = 2k$. Thus, if $f\in \mathrm{PW}_2$ then

$$\langle f,\, s_k\rangle = \sum_{\ell=-\infty}^{\infty} f\Big(\frac{\ell}{2}\Big) S_{k\ell}; \qquad S_{k\ell} = \frac{(-1)^k \sin\pi\frac{\ell}{2}}{\pi(\ell-2k)}.$$

We can also turn this around and compute the sample values of $s(x)$ as an element of PW_2. Since the functions $\{\sqrt{2}\, s(2x-k)\}$ form an ONB for PW_2 we can write

$$s(x) = 2\sum_{k=-\infty}^{\infty} \langle s(x),\, s(2x-k)\rangle s(2x-k) \tag{2.11}$$

and conclude that

$$s(x) = \sum_{k=-\infty}^{\infty} \frac{\sin\frac{\pi k}{2}}{\pi k}\, s(2x-k). \tag{2.12}$$

Of course, we already knew this to be the case from the sampling formula (2.10) with $\Omega = 2$ and f the sinc function.

Exercise 2.9.9. Use Matlab to plot the partial sums on the right hand for k up to around 100 on the interval $[-100, 100]$ and comment on the accuracy of the partial sums.

Exercise 2.9.10. Let $h_k = \frac{\sin\pi k/2}{\pi k}$ and define $w(x) = \sum_k (-1)^k h_{1-k} s(2x-k)$. Prove that $w(x)$ is orthogonal s_k for all k. Consequently, show that any element of $\mathrm{PW}_{[-1,-1/2]}\oplus \mathrm{PW}_{[1/2,1]}$ can be expressed as a sum of shifts of $w(x)$.

Exercise 2.9.11. Compute *numeric* sinc functions by using a centered DFT matrix. Use sizes of $N = 128, 256, \dots$ for the DFT and sinc functions of size $N/4, N/8, \dots$. Plot your results and include comments.

A bandlimited wavelet basis for $L^2(\mathbb{R})$

One can also dilate the functions $w_k = \mathrm{e}^{-2\pi i k\xi}\, w_0(\xi)$ with $w_0(\xi) = \mathrm{sgn}\,(\xi)\,\mathrm{e}^{\pi i\xi}\mathbb{1}_{[1/2,1]}(\xi)$ by setting

$$w_{jk}(\xi) = 2^{-j/2}\mathrm{e}^{-2\pi i 2^{-j}k\xi} w_0(2^{-j}\xi).$$

For given j the functions are supported on and orthogonal over $2^{j-1}\le|\xi|\le w^j$. Altogether, the functions w_{jk} form a complete orthonormal basis for $L^2(\mathbb{R})$.

Exercise 2.9.12. Show the the inverse Fourier transform of w_{jk} is $\psi_{jk}(x) = 2^{j/2}\psi(2^j x-k)$ and explain why these functions form an orthonormal basis for $L^2(\mathbb{R}$. Explain why $\psi_{jk}(t) = 2^{j/2}\psi(2^j t-k)$ is the inverse Fourier transform of w_{jk} and, hence, the functions ψ_{jk} also form an ONB for $L^2(\mathbb{R})$. These functions are called *bandlimited wavelets.*

2.10 Better Bandlimited wavelets

The bandlimited wavelet ψ is a shift by one-half of $(\cos 2\pi t - \cos\pi t)/(\pi t)$ so it decays no better than $1/\pi t$ at ∞. So the bandlimited wavelets just defined have the same drawback as the sinc function, namely, poor spatial/temporal localization.

2.10.1 Bell functions

Suppose that we can find a real-valued function b on $[0, \infty)$ such that

$$\sum_{j=-\infty}^{\infty} b^2\left(2^j \xi\right) \equiv 1$$

on $(0, \infty)$. In fact, $\mathbb{1}_{[1,2)}$ is such a function. However, it will be helpful to have a function that has better time localization than the sinc function ($\mathbb{1}_{[1,2)}$ is the Fourier transform of a modulated sinc function).

We begin by defining a *bell function* $b(x)$ as follows. First, let $\phi(x)$ be a nonnegative, symmetric function supported in $[-1, 1]$ having integral $\pi/2$, and define $\theta(x) = \int_{-\infty}^{x} \phi(t)\, dt$. Then $\theta(x) - \pi/4$ is a nondecreasing antisymmetric function having lower and upper bounds $\pm\pi/4$. One can stretch θ by setting $\theta_\epsilon(x) = \theta(x/\epsilon)$. Set $s_\epsilon(x) = \sin\theta_\epsilon(x)$ and $c_\epsilon(x) = \cos\theta_\epsilon(x)$. Then $s_\epsilon(x) = 0$ if $x < -\epsilon$, $s_\epsilon(x) = 1$ if $x > \epsilon$, $s_\epsilon(0) = 1/\sqrt{2}$ and $s_\epsilon(-x) = c_\epsilon(x)$ as is easily checked by properties of sine and cosine. Now set $b(\xi) = s_{1/3}(\xi - 1)c_{2/3}(\xi - 2)$. We can replace the function $\mathbb{1}_{[1,2)}$ used above in defining w_0 by the function b, that is, now set $\omega_0 = \mathrm{sgn}(\xi)\mathrm{e}^{\pi i \xi} b(2|\xi|)$ and set $\omega_{jk} = 2^{-j/2}\mathrm{e}^{-2\pi i 2^{-j} k\xi}\omega(2^{-j}\xi)$.

One has the following theorem, whose proof will not be given here, but relies on special properties of the functions $s_{1/3}$ and $c_{2/3}$. It should be remarked that the particular wavelet bases constructed here are part of a general theory of *local trigonometric bases*, see *e.g.* [**?**, **?**, **?**] for more details. Discrete counterparts of such bases were first introduced by H. Malvar [**?**] in the context of subband coding for speech processing.

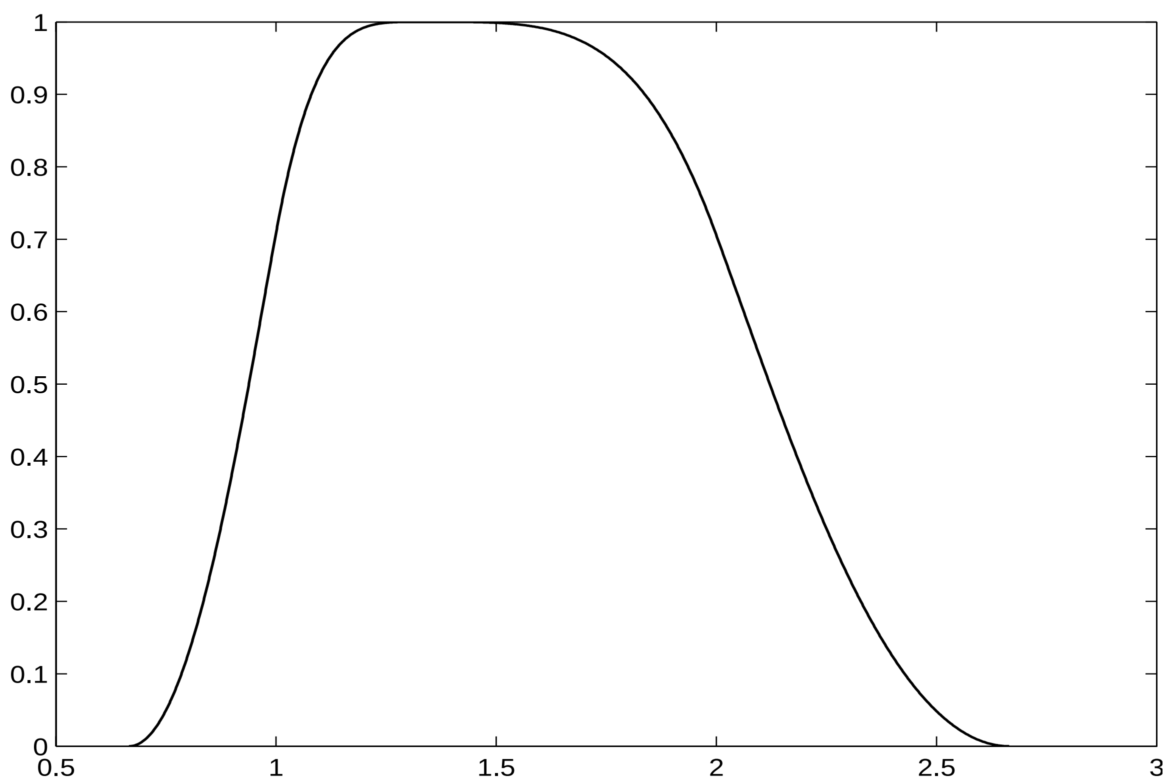

Fig. 2.14. The bell function $b(x)$

Theorem 2.10.1. *With ω_0 defined as above, let $\psi(x)$ be the inverse Fourier transform of ω_0. The functions $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ form an orthonormal basis for $L^2(\mathbb{R})$.*

Importantly, if the function $\phi(x)$ is N times differentiable then $\psi(x)$ will decay at infinity like $1/|x|^N$ or faster. In particular ψ will have much better localization properties than the sinc function. The family ψ_{jk} is an orthonormal basis for $L^2(\mathbb{R})$ called a *wavelet basis*. This type of basis will be considered in detail in Chapter **??**.

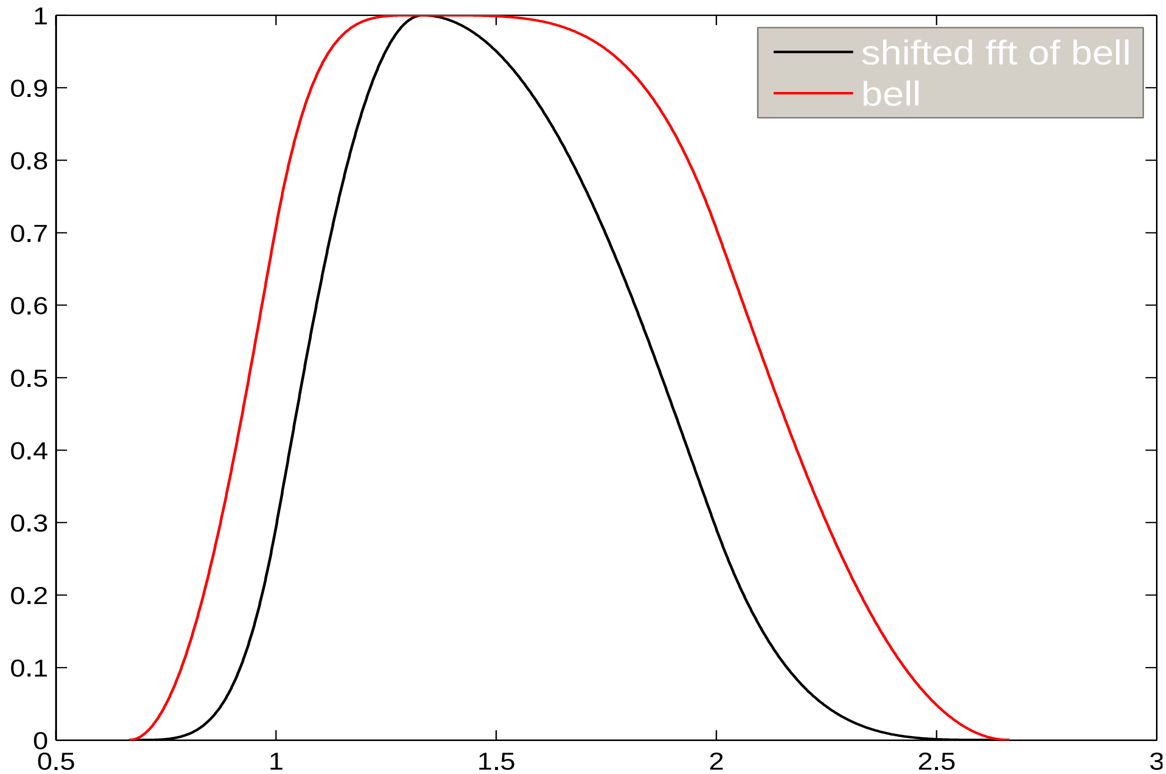

Fig. 2.15. Discrete $b(x)$ and its shifted DFT, illustrating time-frequency localization

Exercise 2.10.2. What is so special about the parameters 1/3 and 2/3 used in the construction of the bell function $b(x)$?

Exercise 2.10.3. (Hard) Prove rigorously that $\sum_{j=1}^{\infty} b^2(2^j\xi) = 1$.

2.11 Time and bandlimited functions

The sinc function is our primary example of a bandlimited function. And any other function in PW is a convolution with the sinc function. But other functions can have better decay than the sinc function: the bell function $b(x)$ discussed above is such an example.

There are no functions that are simultaneously time and bandlimited. This follows from the theory of entire functions but there is an even more basic approach.

Consider now the operators $Q_T f(x) = f\mathbb{1}_{[-T/2,T/2]}$ and $P_\Omega f(x) = (\hat{f}\mathbb{1}_{[-\Omega/2,\Omega/2]})^\vee = (Q_\Omega \hat{f})^\vee$. These operators are *Fourier duals.*

Exercise 2.11.1. Formulate an analogue for these operators in the case of the DFT for N-point signals.

Exercise 2.11.2. Formulate an analogue for any pair of multiplication operators that differ by a Hermitian operator. What special properties do Fourier transform pairs have that other pairs f, Uf where U is unitary, might not?

2.12 Appendix: Fourier transform based signal compression

Signal compression and or denoising can be done with respect to any basis as will be discussed further in Chapter 5. Here we discuss the particular issue of Fourier compression/denoising. The problem is to reconstruct a given signal from its largest Fourier coefficients. The signal that we consider in this example

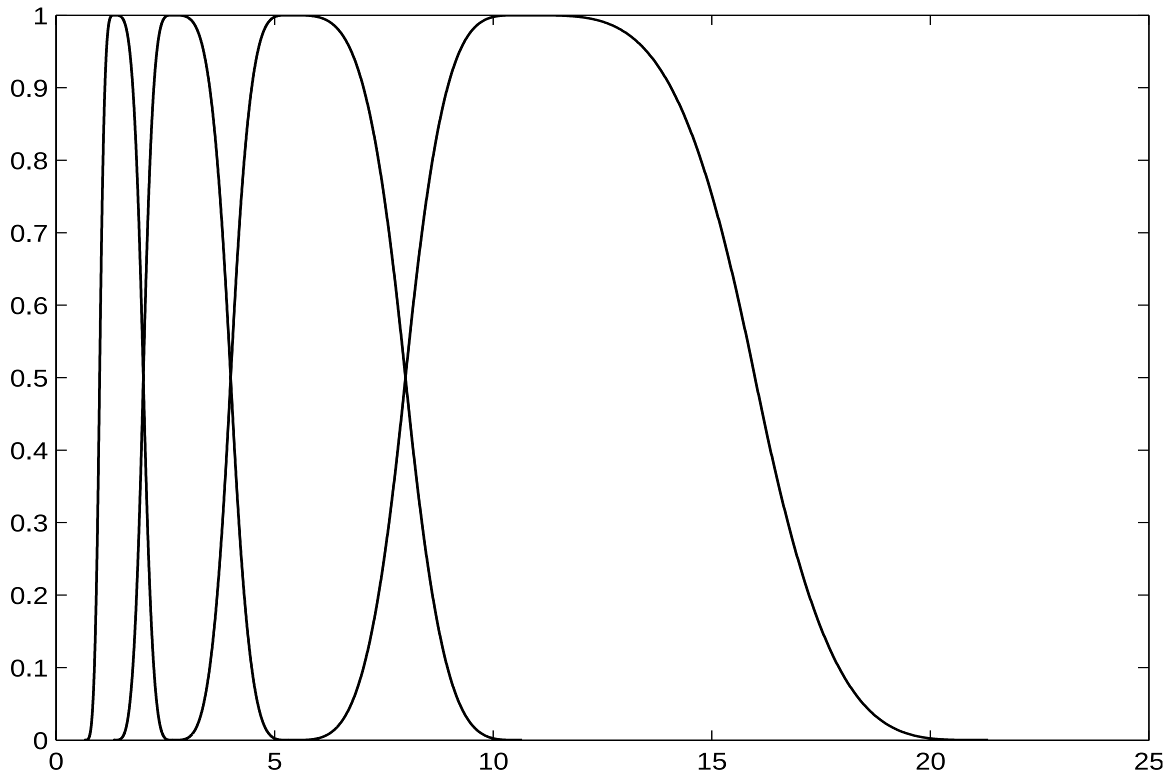

Fig. 2.16. Illustration that $\sum b^2(2^j\xi) = 1$. Note symmetry of contiguous terms with respect to $y = 1/2$.

is the word "Bueller" pronounced by Ben Stein who plays the role of Ferris Bueller's teacher in the movie "Ferris Bueller's Day Off." The original sound is contained in the file `buellershort.wav`. In order to convert this to a manipulable file you need to use matlab's `wavread` utility:

```
>>[x,fs]=wavread('buellershort.wav');
```

The vector x should have size $8192 = 2^{13}$, slightly less than one second of audio sampled at the rate $fs = 11025$. One can also use the `wavwrite` utility to output audio signals in windows wave format. One has to specify the signal and the sample rate as in the following example.

```
>>wavwrite(nx,11025,'noisybueller.wav');
```

; The utility for reconstruction from a given percentage of the coefficients is called `fouriercompress.m` and the script follows.

```
 function y=fouriercompress(X,P);


% fouriercompress reconstructs an approximation of
% of a one-dimensional signal from P percent of its largest
% FFT coefficients
% X should be a 1xN row vector


X=double(X);                        % floating point
subplot(1,2,1);
plot(X);
axis([1 length(X) min(X) max(X)]);
title('original data');
M=length(X);
```

```
p=(100-P)/(100);


    tic                                   % start clock
    % start clock
    fX = fft(X);
     disp(['-----------']);
    disp('Forward FFT');
    toc                                   % stop clock
    disp(['-----------']);

    tic
    fcsort = sort(abs(fX(:)));            % sort fourier coeff by magnitudes
    fcerr = cumsum(fcsort.^2);            % sum of squares
    fcerr = flipud(fcerr);                % decreasing order
    fthresh = fcsort(floor(p*M));         % specify threshold
    cf_X = fX .* (abs(fX) > fthresh);     % keep large
     disp(['-----------']);
    disp('Sorting/thresholding');
    toc
    disp(['-----------']);

    tic
    icf_X = ifft(cf_X);
     disp(['-----------']);
    disp('Inverse FFT');
    toc
    disp(['-----------']);
    y=real(icf_X);
    subplot(1,2,2)
    plot(icf_X);
    axis([1 length(X) min(X) max(X)]);
    title('reconstruction from large Fourier coefficients');
```

A 400 sample chunk of the original Bueller signal is plotted at left in Figure 2.17 and the corresponding chunk of the reconstructed signal is on the right. From this chunk it looks as though one has eliminated a high amplitude term but one has to consider that the eliminated term does not extend over the full duration of the signal so does not show up as a large Fourier coefficient. A similar procedure can be used for removing noise. It should be mentioned that it is difficult to analyze performance of audio processing by looking at pictures. The noisy bueller signal is contained in the file `noisybueller.wav`

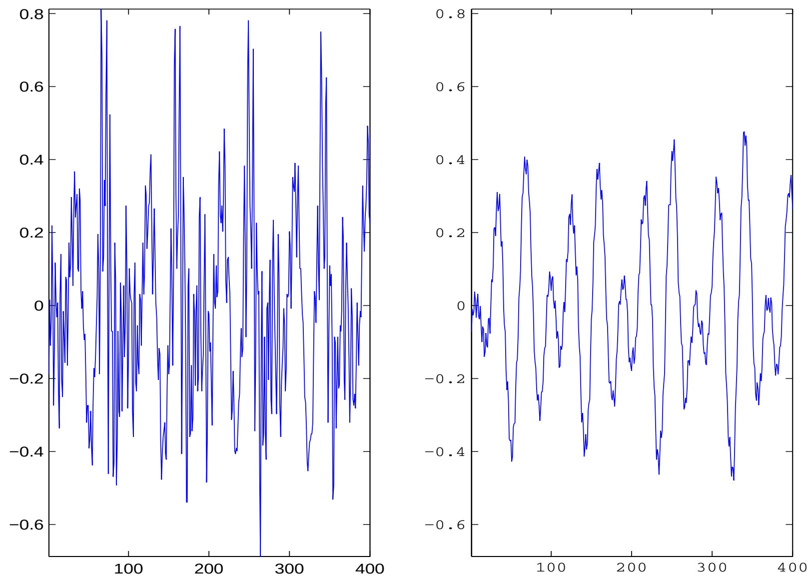

Fig. 2.17. 400 sample chunk of original bueller signal (left) and corresponding chunk of reconstruction from top 10 percent of Fourier terms (right)

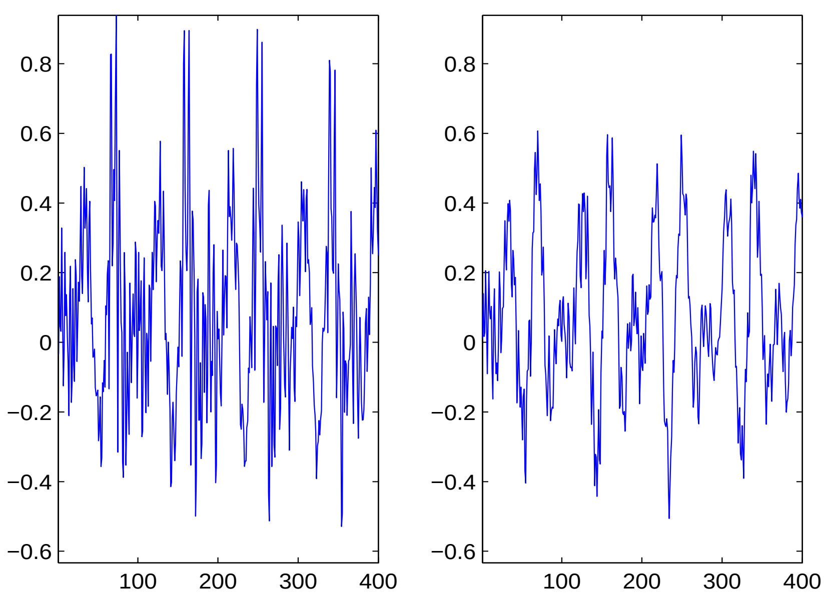

Fig. 2.18. 400 sample chunk of original noisy bueller signal (left) and corresponding chunk of reconstruction from top 10 percent of Fourier terms (right)