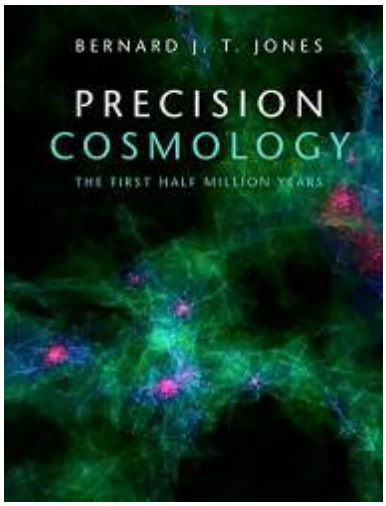
Fourier Analysis

A Supplement to “Precision Cosmology”

Bernard Jones



The Fourier representation of functions plays an important part on all branches of physics and engineering, as well as in other fields like finance. There are numerous forms of the Fourier representation and here the focus is on the relationships between these forms.

This is one of a set of Supplementary Notes and Chapters to “Precision Cosmology”.
Some of these Supplements might have been a chapter in the book itself, but were regarded either as being somewhat more specialised than the material elsewhere in the book, or somewhat tangential to the main subject matter.
The are mostly early drafts and have not been fully proof-read.
Please send comments on errors or ambiguities to “PrecisionCosmology(at)gmail.com”.

Contents

1 Fourier Analysis

1.1 Fourier representation of random fields

It has become customary to represent functions in general and random fields in particular as the superposition of sinusoidal oscillations. In random function theory this provides an alternate view of functions that offers considerable insight into the nature of the underlying random processes and provides methods of characterising random functions. However, in the final analysis, what we most frequently measure and use in our theories are distribution functions and correlation functions of the various fields concerned.

Short biography: **Joseph Fourier (1768-1830)** Orphaned by the time he was 10 years of age, Fourier attended the École Royale Militaire of Auxerre in 1780 and by 1783 showed considerable distinction in mathematics. He was an active supporter of the French Revolution, though not of the period of Terror, and was himself imprisoned in 1794. Following the death of Robespierre, he was freed and attended the École Normale in Paris where he was taught by Lagrange, Laplace and Monge. In 1797 he succeeded Lagrange at the École Polytechnique in the chair of Analysis and Mechanics. Between 1804 and by 1807 he worked on his memoir, " On the Propagation of Heat in Solid Bodies". The work was heavily criticised by Lagrange and Laplace who complained that functions simply could not be expanded in series of trignometric functions. Nonetheless, he submitted this memoir for a prize in 1811. Lagrange and Laplace were members of the prize committee who, somewhat grudgingly, awarded Fourier the prize. Fourier was the first person to draw attention to the so-called "Greenhouse Effect" and might well be regarded as the founder of the study of planetary atmospheres (Pierrehumbert, 2004).

Fourier analysis is named after Joseph Fourier who introduced the technique in his book *Theorie Analytique de la Chaleur* Fourier (1822). The work in fact dates from around 1807 when Fourier published a memoir on the subject, but was not available in English until Cambridge University Press published a translation by Alexander Freeman in 1878. Originally the technique was used to analytically solve partial differential equations, but it has survived and thrived in the computer age with the advent of the Discrete Fourier Transform ('DFT') and the Fast Fourier Transform ("FFT").[1]

The seminal modern work on the Fourier transform is perhaps that of Bracewell (1991).

[1] As with all histories there are those who are readily recognised for their contributions and those who are not. What is perhaps most remarkable is that Gauss had written a paper as early as 1805 or 1806 on the orbit of the asteroid Juno which had been discovered by Harding in 1804. In that paper he used what we would now refer to as the Fourier representation of the orbit. Remarkably, Gauss' algorithm for computation was what we now refer to as the "Fast Fourier Transform". This work was only published in Gauss' collected works (Gauss, 1866), and remained unknown until Goldstine (1977, pp. 249-253) drew attention to it.

1.2 Functions on a finite interval: Fourier series

There are several variants and approaches to representing a function $f(x)$ as a combination of sine and cosine functions of different frequencies. There is the *Fourier series* representation of a function $f(\theta)$ on the interval $\theta \in [-\pi, \pi]$:

$$f(\theta) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty}(A_n \cos n\theta + B_n \sin n\theta) \tag{1.1}$$

$$A_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\ \cos n\theta\, d\theta,\ n = 0, 1, \ldots \tag{1.2}$$

$$B_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)\ \sin n\theta\, d\theta,\ n = 1, 2, \ldots \tag{1.3}$$

This was the essence of the pioneering work of Gauss[1] and Fourier and what Lagrange and Laplace had objected to. The extension of such representations outside of the interval $[-\pi, \pi]$ is periodic.

1.2.1 Real Fourier series

By redefining a variable x

$$x = \frac{\theta L}{2\pi} \tag{1.4}$$

we get a representation of $f(x)$ on the interval $[-L/2, L/2]$:

$$f(x) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty}\left(A_n \cos\frac{2\pi nx}{L} + B_n \sin\frac{2\pi nx}{L}\right) \tag{1.5}$$

$$A_n = \frac{2}{L}\int_{-L/2}^{L/2} f(x)\ \cos\frac{2\pi nx}{L}\, dx, \qquad B_n = \frac{2}{L}\int_{-L/2}^{L/2} f(x)\ \sin\frac{2\pi nx}{L}\, dx, \tag{1.6}$$

Here n runs through the values $n = 0, 1, \ldots \infty$ if we define $B_0 = 0$. The functions $f(\theta)$ and $f(x)$ defined through these series expansions are periodic if extended outside of their respective domains $[0, 2\pi]$ and $[-L/2, L/2]$. The mean square value of the function $f(x)$ is given by

$$\langle f(x)^2\rangle = \frac{1}{L}\int_{-L/2}^{L/2} f(x)^2 dx = A_0^2 + \frac{1}{2}\sum_{n=1}^{\infty}(A_n^2 + B_n^2) \tag{1.7}$$

This is sometimes referred to as *Rayleigh's theorem* for the Fourier series representation of a function $f(x)$ defined on an interval $[-L/2, L/2]$. It is a Fourier series variant of *Parseval's theorem* that applies to Fourier transforms. We shall be discussing a number of variants of the Fourier representation of functions, each with its own version of the Parseval theorem.

Because $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$ we can write (1.5) as

$$f(x) = \sum_{n=-\infty}^{\infty}\left(a_n \cos\frac{2\pi nx}{L} + b_n \sin\frac{2\pi nx}{L}\right) \tag{1.8}$$

where $A_n = a_n + a_{-n}$ and $B_n = b_n - b_{-n}$. We then have $2a_{\pm n} = A_n$ and $2b_{\pm n} = (-1)^n B_n$, thus eliminating the factors of 2 that multiply the integrals in (1.6).

1.2.2 Complex Fourier series

We can rewrite either representation (1.5) or (1.8) in an the equivalent complex form

Complex Fourier series representation

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp\left[\frac{2\pi i n x}{L}\right], \quad c_n = \frac{1}{L}\int_{-L/2}^{L/2} f(x) \exp\left[-\frac{2\pi i n x}{L}\right] \tag{1.9}$$

where the relationship between the c_n and the a_n, b_n of equation (1.8) is

$$c_n = (a_n - ib_n)/2,\ n > 0, \qquad c_n = (a_n + ib_n)/2,\ n < 0 \tag{1.10}$$

As with (1.6) the Fourier representation of $f(x)$ on the interval $[-L/2, L/2]$ is periodic if extended outside of that interval. It should also be noted that the coefficients c_n will be complex numbers and can be written as an amplitude $C_n = |c_n|$ and a phase angle $\phi_n = \arg c_n$. With that, the Fourier series is

$$f(x) = \sum_{n=-\infty}^{\infty} R_n \exp\left(\frac{2\pi i n x}{L} + \phi_n\right) \tag{1.11}$$

The R_n and ϕ_n are respectively referred to as the *amplitude and phase of the n^{th} Fourier component*.

1.2.3 Parseval's theorem

From here it is but a short step to the important *Parseval theorem*[2]

$$\frac{1}{L}\int_{-L/2}^{L/2} |f(x)|^2\, dx = \sum_{n=-\infty}^{\infty} |c_n|^2 \tag{1.13}$$

This illustrates one of the important features of data representation by orthogonal functions: the sum on the right hand side is a sum of the squared moduli of the individual amplitudes a_n, and involves no cross-terms involving cross-products $a_m^* a_n$. In this sense the components in the Fourier representation make independent contributions to the integral of $|f(x)|^2$. This in turn means that we could truncate the sum on the right hand side

[2] A more accurate statement of Parseval's theorem concerns two periodic (possibly complex valued) functions, $A(x)$ and $B(x)$, on the real line. If $A(x)$ and $B(x)$ are Riemann integrable with Fourier series representations $A(x) = \Sigma_{-\infty}^{\infty} a_n e^{inx}$ and $B(x) = \Sigma_{-\infty}^{\infty} b_n e^{inx}$ then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} A(x)B(x)^* dx = \sum_{-\infty}^{\infty} a_n b_n^*. \tag{1.12}$$

where * denoted complex conjugation. Equation (1.13) is a special case of this with $A(x) = B(x)$.

after a given number of terms without influencing the contribution of the lower-n terms in representing the function.

Short biography: **Marc-Antoine Parseval 1755-1836** Parseval was a minor royalist noble at the time of the French Revolution and was imprisoned in 1792, later to be freed with the rise of Napoleon. However, as a consequence of his publishing critiques of Napoleon's regime, he had to escape from France in order to avoid a warrant for his arrest. He later returned to Paris. Parseval was not a particularly distinguished mathematician: he wrote only five papers and was never elected to the Académie de Sciences.

1.3 Functions on the real line: Fourier transforms

We can now take the limit as $L \to \infty$ of (1.9). To achieve this it is convenient to make the transformation of variables:

$$\omega = \frac{2\pi n}{L}, \qquad g(\omega) = 2\pi L\, c_n \tag{1.14}$$

In the limit $n \to \infty$, this replaces the discrete n-labels on the c_n with the continuous variable ω, and the discrete set of coefficients a_n with the continuous function $g(\omega)$. As the size L of the interval increases during this limiting process, a fixed value of y corresponds to a value of n such that n/L is fixed. The variable ω is referred to as the *angular frequency*.[3]

Going through the mathematics[4] we get

$$f(x) = \sum_{n=\infty}^{\infty} c_n e^{2\pi i n x/L} \;\to\; \int_{-\infty}^{\infty} \underbrace{\frac{2\pi}{L} g(\omega)}_{c_n} \underbrace{e^{i\omega x}}_{e^{2\pi i n x/L}} \underbrace{\frac{L}{2\pi} d\omega}_{dn} = \int_{-\infty}^{\infty} g(\omega)\, e^{i\omega x}\, d\omega \tag{1.15}$$

[3] The angular frequency ω is measured in radians per second, this is the SI unit irrespective of whether we refer to the time or the space domain. In the time domain ω is associated with a period of oscillation $T = 2\pi/\omega$ seconds. If we associate a so-called *ordinary frequency* $\nu = \omega/2\pi$ with this, then ν is conventionally measured in *Hertz*. In engineering we often see the symbol f used for ν. In the space domain we identify a wavelength $\lambda = 2\pi/\omega$ with a mode of spatial frequency ω.

[4] The tricky part of the limiting process is to convert the sum in the first of equations (1.9) into an integral. Schematically, the argument involves defining a frequency $\omega = (2\pi/L)n$ and replacing the a_n by $(2\pi/L)\, g(\omega)$. From this definition of ω, the interval of frequency $d\omega$ corresponds to an interval dn of n given by $d\omega = (2\pi/L)dn$. In the sum, the value of dn is $dn = 1$ and so with this we have

See also Jackson (1998, §2.9), but note that Jackson uses a different normalisation for $g(\omega)$. This is, in essence, the definition of the Riemann integral appearing on the right hand side.

The formal proof requires two conditions on $f(x)$ (due to Dirichlet). Firstly $f(x)$ must be piecewise smooth and have only a finite number of discontinuities and maxima on all finite intervals. Further, $\int_{-\infty}^{\infty} |f(x)|\, dx$ should converge. Neither of these conditions is true for a homogeneous stochastic (random) function and so, for such a function $f(x)$, there is no prior guarantee that there exists any such Riemann integrable function $g(\omega)$. The technical way around this is to say that there exists a density $dG(\omega)$ such that the right hand side, $\int_{-\infty}^{\infty} e^{i\omega x} dG(\omega)$ exists in the Riemann-Stieltjes sense. For simplicity we can think that $dG(\omega) = g(\omega)d\omega$. See Batchelor (1982, §2.5) in the context of turbulent flow, and Martinez and Saar (2001, §7.2.2) and Adler (1981) in the context of statistics for more about this.

Fourier transform on $\mathbb{R}$:

$$f(x) = \int_{-\infty}^{\infty} g(\omega)\, e^{i\omega x} d\omega, \qquad g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x)\, e^{-i\omega x} dx \tag{1.16}$$

$g(\omega)$ is referred to as the Fourier transform of $f(x)$, while $f(x)$ is the inverse Fourier transform of $g(\omega)$. The direction is defined by the sign in the exponential. Note how the normalisation factor $1/2\pi$ has appeared in the transform from real space $f(x)$ to frequency space $g(\omega)$. We shall have a lot to say about that later.

The same limiting process leads to the version of Parseval's theorem (1.13) for the Fourier representation of a function on the entire real line:

Parseval's theorem on the real line

$$\int_{-\infty}^{\infty} |\, f(x)\,|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\, g(\omega)\,|^2\, d\omega, \qquad |\, g(\omega)\,|^2 = g^*(\omega)\, g(\omega) \tag{1.17}$$

The function $g(\omega)$ is generally a complex valued function which we can write $g(\omega)$ as the sum of its real and imaginary parts:

$$g(\omega) = g_R(\omega) + i g_I(\omega); \quad g_R(\omega) = Re[g(\omega)],\ g_I(\omega) = Im[g(\omega)] \tag{1.18}$$

The Fourier transform $g(\omega)$ has the following symmetries:

$f(x)$ real	$g_R(\omega)$ is even $g_I(\omega)$ is odd
$f(x)$ even	$g(\omega)$ is even
$f(x)$ odd	$g(\omega)$ is odd

These must be checked for normalisation!

There are extensive tables of Fourier transforms of functions $f(x)$. A small but interesting sample is provided in Table 1.1.

1.4 Normalisations factors, notation and other conventions

1.4.1 The 2π normalisation factors

As stressed by Bracewell (1991) in his opening chapter there is no unique normalisation for the Fourier transform.

The normalisation factors of $1/2\pi$ appearing in (1.16) and (1.129) arise from the route we have taken from the Fourier series to the Fourier transform. In particular, the normalisation of $g(\omega)$ in its definition in equation (1.14) is somewhat arbitrary and becomes a matter of convention. Had we taken a slightly different normalisation, $h(\omega) = \sqrt{2\pi} L\, a_n$, we would

Table 1.1 Some Fourier transform pairs

$f(x)$	$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\,e^{-i\omega x}\,dx$	$f(x)$	$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\,e^{-i\omega x}\,dx$
$\delta(x)$	$\frac{1}{2\pi}$	$\cos\Omega x$	$\frac{1}{2}(\delta(\omega+\Omega)+\delta(\omega-\Omega))$
$\delta(x-x_0)$	$\frac{1}{2\pi}e^{-i\omega x_0}$	$\sin\Omega x$	$\frac{i}{2}(\delta(\omega+\Omega)-\delta(\omega-\Omega))$
1	$\delta(x)$	$f(x)\sin\Omega x$	$\frac{i}{4\pi}[F(\omega+\Omega)-F(\omega-\Omega)]$ [a]
$e^{-\alpha\lvert x\rvert},\ \alpha>0)$	$\frac{\alpha}{\pi}\frac{1}{\alpha^2+\omega^2}$	$\mathrm{rect}\left(\frac{x}{L}\right)$ [b]	$\frac{L}{2\pi}\,\mathrm{sinc}\left(\frac{\omega L}{2}\right)$ [c]
$e^{-\alpha x^2},\ \alpha>0$	$\frac{1}{2\alpha\sqrt{\pi}}e^{-\omega^2/4\alpha}$	$\frac{d^n f}{dx^n}$	$\frac{(i\omega)^n}{2\pi}F(\omega)$

[a] $F(\omega)$ is Fourier transform of $f(x)$ as indicated in the column headers.
[b] $\mathrm{rect}(x) = 1$ if $|x| \le 1/2$, 0 otherwise.
[c] $\mathrm{sinc}(x) = (\sin\pi x)/\pi x$.

have reached a Fourier transform that had a more symmetrical appearance:

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(\omega)\,e^{i\omega x}d\omega, \qquad h(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)\,e^{-i\omega x}dx \tag{1.19}$$

In fact, with arbitrary normalisation we would always end up with forward and backward transforms of the form

$$f(x) = A\int_{-\infty}^{\infty} h(\omega)\,e^{i\omega x}d\omega, \qquad h(\omega) = B\int_{-\infty}^{\infty} f(x)\,e^{-i\omega x}dx, \quad AB = \frac{1}{2\pi} \tag{1.20}$$

for the transform and its inverse. But no matter which normalisation we choose we must always have $AB = 1/2\pi$. In general, the usual choice is perhaps $A = 1/2\pi$, though in cosmology there is a prevalence of $A = 1$ and B $= 1/2\pi$.

In engineering there is a tendency to work in terms of the 'ordinary frequency' $\nu = \omega/2\pi$, measured in Hertz (Hz, cycles per second) in the time domain rather than angular frequency ω measured in radians per second. With $B = 1$, the Fourier transform (1.20) becomes symmetric when expressed in terms of ν and we can write the Fourier transform pair $f(x) \leftrightarrow s(\nu)$ as:

$$s(\nu) = \int_{-\infty}^{\infty} f(x)\,e^{-2\pi i\nu x}dx, \qquad f(x) = \int_{-\infty}^{\infty} s(\nu)\,e^{2\pi i\nu x}d\nu \tag{1.21}$$

The relationship between $s(\nu)$ and $h(\omega)$ is

$$h(\omega) = s(\nu)\Big|_{\nu=\omega/2\pi}, \quad \omega = 2\pi\nu \tag{1.22}$$

This can be traced back to the choice of variables (1.14) used in deriving the Fourier transform on the real line. In this case the definition of ω in (1.14) would be replaced by $\nu = n/L$. This variant has the perceived advantage that the Fourier transform of the delta function $f(x) = \delta x$ is $s(\nu) = 1$.

The other issue concerns the sign of the frequency used in equations (1.16) and (1.129). If in (1.16) we regard $f(x)$ as the 'real world' function of our perception and $g(\omega)$ as the 'mathematical' transform space, we see that the transformation from the real world to the mathematical world, $f(x) \rightarrow g(\omega)$ involves an integration with $e^{-i\omega x}$ while the reverse transformation $g(\omega) \rightarrow f(x)$ involves $e^{i\omega x}$. Because our Fourier transform involves all frequencies $-\infty < \omega < \infty$ we could make the transforms look different by making the swap $\omega \rightarrow -\omega$. In the case of the symmetric form (1.19) this swap makes no difference except with regard to the physical issue of which of $f(x)$ and $g(\omega)$ is regarded as the 'real world'.[8]

1.4.2 The Fourier transform in cosmology

The Fourier transform is used in cosmology, particularly in the discussion of linear, small amplitude, perturbations, in order to split the random density field into independent sinusoidal modes. Each mode has a wavelength and the evolution of a mode depends directly on the physics that applies to that scale. The issue of normalisation is however ever present.

Pre-empting our discussion of the Fourier transform in 3-dimensions the safest approach is to introduce an arbitrary normalisation factor V such that the Fourier transform, $F(\mathbf{k})$, of a field $f(\mathbf{x})$, and its inverse are given by

$$f(\mathbf{x}) = \frac{V}{(2\pi)^3}\int_V F(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k}, \quad F(\mathbf{k}) = \frac{1}{V}\int_V f(\mathbf{x})\, e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}. \tag{1.23}$$

The volume V is notionally the volume of the universe, or of a large enough representative sample. The product of the pre-multiplying factors is $(2\pi)^3$, as it should be in the 3-dimensional case, whatever the value of V.[9] See also section 1.15.1, equations (1.132) and (1.133).

The convention adopted in this book is $V = (2\pi)^3$, so that the real space value of $f(\mathbf{x})$ is simply the sum over Fourier components $e^{i\mathbf{k}\cdot\mathbf{x}}$ weighted by their amplitudes $dF(\mathbf{k}) = F(\mathbf{k})d\mathbf{k}$.

1.4.3 Why Fourier transforms?

The Fourier transform is by no means the only transform available: we could in principle use any set of orthogonal functions to describe a random field. So we should wonder under what circumstances the Fourier transform is the appropriate tool for analysis of random fields.

The Fourier transform, in $\mathbb{R}^3$, is tied to the notion of rectangular Cartesian coordinates, while the discrete version, which acts on a Cartesian grid in a finite box, imposes additional conditions of periodicity on the box boundaries. This makes it less suitable for describing data on a sphere, or large scale data on a curved space-time manifold where the deviations

[8] Jackson (1998) and Arfken and Weber (2005) both use the definition (1.19) but Arfken reverses the sign of the frequency, $\omega \rightarrow -\omega$. Press et al. (2007, §12.0) use (1.21), but with the sign of the frequency reversed.

[9] As Kolb and Turner (1990, §9.2.2, footnote 7, p.325) suggest, we can ignore the factors of V since they play no crucial role.

Table 1.2 Some properties of Fourier transforms

[a]	Function $f(x)$	Fourier transform $F(\omega)$ [b]	constants
Linearity	$f(x) = af_1(x) + bf_2(x)$	$F(\omega) = aF_1(\omega) + bF_2(\omega)$	a, b
Scaling	$f(\alpha x)$	$\frac{1}{\|\alpha\|} F\left(\frac{\omega}{\alpha}\right)$	α
Shifting	$f(x - x_0)$	$F(\omega)\, e^{-\omega x_0}$	x_0
Multiplication	$f_1(x) f_2(x)$	$\int_{-\infty}^{\infty} F_1(\sigma) F_2(\omega - \sigma)\, d\sigma$	
Convolution	$f(x) = \int_{-\infty}^{\infty} f_1(\xi) f_2(x - \xi)\, d\xi$	$F(\omega) = F_1(\omega) F_2(\omega)$	

[a] $f(x) \Leftrightarrow F(\omega)$, $f_1(x) \Leftrightarrow F_1(\omega)$, $f_2(x) \Leftrightarrow F_2(\omega)$ are Fourier transform pairs

[b] The convention used here is $f(x) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega$ and $F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$

from Euclidean geometry may become important. Indeed, it is essential that the orthogonal functions used to represent data somehow reflect the boundary conditions on the data. Data on a sphere is periodic in the azimuthal angle, and so our chosen functions should reflect that.

Even in a Euclidean geometry with a Cartesian coordinate system, the Fourier transform is difficult to apply to a random field defined on a random set of points without re-sampling the data onto a discrete grid.

Check these for normalisation

Nonetheless, Fourier transforms have a number of very desirable properties some of which are summarised in table 1.2.

1.5 The Dirac delta-function

1.5.1 A definition

We will occasion to use a function called *the Dirac delta-function*, $\delta(x)$ that is zero everywhere except at $x = 0$ and whose integral is unity:

$$\delta x = 0, \quad x \neq 0, \tag{1.24}$$

$$\int_{\mathcal{I}} \delta(x) dx = 1, \quad x = 0 \in \mathcal{I} \tag{1.25}$$

Technically speaking, $\delta(x)$ is a generalised function, rather than a a function that is defined on an interval. It has the important property that

$$\int_{\mathcal{I}} f(y)\, \delta(x - y)\, dy = f(x), \quad \forall x \in \mathcal{I} \tag{1.26}$$

1.5.2 Dirac delta-function and the Fourier transform

For practical purposes, the δ-function is usually viewed as the limit of a Gaussian function of unit area when the width, or variance, tends to zero. An alternate model that relates the δ-function to the Fourier transform, is to write

$$\delta(x) = \lim_{w\to\infty} \frac{\sin wx}{\pi x} \tag{1.27}$$

As $w \to \infty$ the function $\sin wx/x \to w$ and has width $2\pi/w$: it gets more spike-like. The integral of $\sin wx/x$ over the entire real line is 1, whatever the value of w. We notice from this definition[12] (1.27) that

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega x} d\omega \tag{1.29}$$

The δ-function is the Fourier transform of unity (*i.e.* $f(x) = 1$).

Similarly we can calculate that the δ-function $\delta(\omega)$ in ω-space is simply

$$\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x} dx \tag{1.30}$$

It is important to note that the functions (1.29) and (1.30) may look like a Fourier transform pair, but they are not: they are merely definitions of δ-functions in x-space and in ω-space. The $1/2\pi$ normalisations are required to make the area under the *delta*-function unity in whichever space is being used.

1.5.3 Dirac delta-function in 3-dimensions

Here we have the analogous definitions

$$\delta(\mathbf{x}) = 0, \quad \mathbf{x} \neq 0, \tag{1.31}$$

$$\int_{\mathcal{V}} \delta(\mathbf{x})\, d\mathbf{x} = 1, \quad \mathbf{x} \in \mathcal{V} \tag{1.32}$$

with the property

$$\int_{\mathcal{V}} f(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}) d\mathbf{x} = f(\mathbf{y}), \quad \forall \mathbf{x} \in \mathcal{V} \tag{1.33}$$

The integral representation of $\delta(\mathbf{x})$ is

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} e^{i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k} \tag{1.34}$$

where the integral is over all space. As before this is not a Fourier transform and so the $(2\pi)^3$ normalisation is not a matter of choice: it is required for the integral of the δ-function to be unity.

[12] It is simple to evaluate the integral

$$f_w(x) = \frac{1}{(2\pi)^{1/2}}\int_{-w}^{+w} 1 \times e^{iwx} dw = (2\pi)^{1/2}\frac{\sin wx}{x}. \tag{1.28}$$

Then, on comparing with (1.27), we see that $\lim_{w\to\infty} f_w(x) = \delta(x)$.

1.6 Spatial Random Functions: some preliminaries

One of the key areas of application for Fourier Transforms is in the prescription of random fields. Simply put, a random field is a set of random numbers defined on a point set. The point set might be regular, *i.e.* a grid, or it may be a set of points that are themselves randomly distributed. The points might occupy a $1, 2, \ldots, N$-dimensional space. In one dimension we frequently think of the points of the set being distributed in time. In that sense the points are part of a time series of data samples. In more than one dimension we generally think of the points being distributed spatially. The values are generally taken to be simple scalar values, though we can equally well consider the distribution of vectors and tensors, as is commonplace in turbulence theory.

1.6.1 Random Fields

One main issue is to pin down what we mean by a function taking on random values. In the simplest case, the values are simply taken independently from a given statistical distribution. At a more complicated level, the value assigned to a point might depend on other points in its neighbourhood. To make progress it is convenient to fix attention on the simplest case: when the values of the function are taken from an isotropic zero-mean Gaussian distribution, allowing for possible pair-wise correlations among the points.

The advantage of considering an isotropic Gaussian random field is that the statistical structure of the field is set by the Gaussian distribution function, no other information is needed.

Consider a scalar field $\phi(\mathbf{r})$. The probability $P(\phi_1, \phi_2, r)$ that the values of the function at two points, $\mathbf{r}_1, \mathbf{r}_2$ separated by a distance $r = |\,\mathbf{r}_1 - \mathbf{r}_2\,|$ are $\phi(\mathbf{r}_1), \phi(\mathbf{r}_2)$ take on values ϕ_1, ϕ_2 is

$$P_2(\phi_1, \phi_2, r) = \frac{1}{2\pi\sigma^2\sqrt{1 - C^2(r)}} \exp\left\{-\frac{\phi_1^2 + \phi_2^2 - 2C(r)\phi_1\phi_2}{2\sigma^2[1 - C^2(r)]}\right\} \tag{1.35}$$

This describes a statistically isotropic Gaussian random field with 2-point correlation function $C(r)$.

The visual appearance of the structure is the field is governed entirely by the correlation function, $C(r)$ specified in the underlying distribution. In an isotropic field the correlation function will only a a function of the pairwise separation of point pairs.

The relevance to Fourier analysis is simply that, for such a field, the Fourier transform of the correlation function is guaranteed to exist. The guarantee comes from the important Wiener-Khinchine theorem that we will discuss below. The Fourier transform of the correlation function is called the *Power Spectrum* of the random field.

In 2- or 3-dimensions the correlation function of a density field tells us about shapes of the density run around the highest peaks and deepest troughs of the density map. The Fourier description of the same field breaks the description up into independent components that combine to generate structure on different length scales. Those components are sine and cosine waves of different frequencies and their Fourier amplitudes. In practical

terms the Fourier representation plays an important role in processing the image of the field. A more extensive discussion is found in the Supplement on *Random Fields*.

For simplicity in what follows we shall restrict attention to a one dimensional random field $X(t)$, where t is a parameter along the line. The value taken on at the point t will be denoted by $x(t)$. The probability density that $X(t)$ takes the value $x(t)$ at t by $f(x;t)$ and the probability that $X(t)$ values $x_1(t_1)$ at t_1 and $x_2(t_2)$ at t_2 by $f(x_1, x_2; t_1, t_2)$.

1.6.2 Stationary Random Processes

An important class of random process are the *Stationary Random Processes*.

Definition 1 A random process $X(t)$ is said to be *stationary in the strict sense* if its statistics are not affected by a shift in the time origin:

$x(t)$ and $x(t+\epsilon)$ have the same statistics for all ϵ

It follows from this that $f(x;t) = f(x;t+\epsilon)$ and hence

$$f(x;t) = f(x) \tag{1.36}$$

the probability density is independent of t. Note that this concerns only the 1-point probability density function, it makes no statement about the higher point functions.

Consequently, for a stationary random process, the expectation value of the random field must be the same everywhere:

$$\mathbb{E}[x(t)] = \mu = \text{constant} \tag{1.37}$$

Likewise, the covariance of the values at two points separated by a distance $\tau = |t_1 - t_2|$ must be independent of position along the line and depend only on τ:

$$\mathbb{E}[x(t)x(t+\tau)] = R(\tau) = R(-\tau) \tag{1.38}$$

The function $R(\tau)$ is the autocorrelation function of the process $X(t)$. This is discussed further below (equ. 1.42).

1.6.3 Mean, autocorrelation and autocovariance

We can express the mean of the random process $X(t)$ directly in terms of the underlying probability density $f(x)$:

$$\langle x(t)\rangle = \mu(t) = \mathbb{E}[x(t)] = \int_{-\infty}^{+\infty} x f(x;t)\, dx \tag{1.39}$$

The integral is simply the first moment of x taken over its distribution function $f(x)$. There is another way of expressing this, which is to take an average value over the real line on which $x(t)$ is defined:

$$\bar{x} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)dt \tag{1.40}$$

In making this integral we are saying that the mean value, $\langle x(t)\rangle$, given in terms of the distribution of $X(t)$ is the same as the average value, $\bar{x}$, of any single realisation $x(t)$ taken over the real line.

Likewise, we can write the *Autocorrelation Function* $R(t_1, t_2)$ of the random process $X(t)$ as the joint moment of the random variables $X(t_1)$ and $X(t_2)$:

$$R(t_1, t_2) = \langle x(t_1)x(t_2)\rangle = \mathbf{E}[x(t_1)x(t_2)] \tag{1.41}$$

$$= \int_{-\infty}^{+\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \tag{1.42}$$

The *autocovariance* of $x(t_1)$ and $x(t_2)$ is

$$C(t_1, t_2) = \mathbb{E}\left[(x(t_1) - \mu(t_1))(x(t_2) - \mu(t_2))\right] \tag{1.43}$$

$$= R(t_1, t_2) - \mu(t_1)\mu(t_2) \tag{1.44}$$

Note that the variance of $x(t)$ is then

$$\mathrm{Var}\,(x) = \sigma^2_{x(t)} = C(t, t) = R(t, t) - \mu^2(t). \tag{1.45}$$

1.7 Some mathematical preliminaries

1.7.1 Dealing with stochastic functions

In this subsection we shall retain a modicum of mathematical correctness in order to show where things comes from and what formal restrictions exist on the various measures we use to characterise random functions. For more details see Martinez and Saar (2001, §7.2.2) and Papoulis and Pillai (2002, §9-3 and §11-4). [13]

In keeping with the previous discussion (§1.3) we consider a random function $f(x)$ on the real line that can be expressed on any finite interval $[-L/2, L/2]$ as the limit[14] of a sum of discrete Fourier components of complex amplitude g_k and frequency $\omega_k = 2\pi k/L$:

$$f(x) = \lim_{n\to\infty} \sum_{k=1}^{n} g_k\, e^{i\omega_k x} \stackrel{??}{=} \int_{-\infty}^{\infty} g(\omega) e^{i\omega x} d\omega, \quad -L/2 < x < L/2. \tag{1.47}$$

As the number of modes n increases, the frequency difference $\omega_k - \omega_{k-1}$ between neighbouring components decreases and, it might be argued, improves the approximation to the integral on the right. This is certainly the case for piecewise continuous and differentiable functions $f(x)$ (see Carleson, 1966, for a strict statement and proofs).

However, the more mathematically minded reader might justifiably be concerned that

[13] The definition of the Fourier transform used by Papoulis and Pillai (2002) is $A = 1/2\pi, B = 1$ in equations (1.20). Martinez and Saar (2001) use $A = 1, B = 1/2\pi$ but with frequency reversed: $\omega \to -\omega$.

[14] Technically, since we are dealing with stochastic processes, this is the *limit in the mean*, generally denoted by l.i.m.. For a sequence of random variables $a_i, a_2, \ldots$ this is defined by the equation

$$\text{l.i.m.}\, a_n = a \quad \Longleftrightarrow \quad \lim_{n\to\infty} \mathbb{E}\left[|\, a_n - a\,|^2\right] = 0. \tag{1.46}$$

This is also referred to as *convergence in the mean square sense*.

we are dealing with random functions, $f(x)$ and $g(\omega)$, that are not smooth and probably not differentiable anywhere. Realisations of such functions may not even be integrable in the strict Riemann sense.

They would be correct, we need to use a more sophisticated notion of integration: the Fourier-Stieltjes stochastic integral.[15]. When the function $f(x)$ in (1.47) is a stochastic function, it may be dense with discontinuities and thus not differentiable. We have to avoid the temptation to assert that the there exists a function $g(\omega)$ that when integrated as in (1.47) yields $f(x)$.

1.7.2 The Fourier-Stieltjes integral

Short biography: **Thomas Stieltjes (1856-1894** Thomas Stieltjes was born in Holland and died in France at the age of 38. He had worked as an assistant to the director of the Leiden observatory, until 1883 when he gave up that position to become a mathematician at the University of Delft. The following year he applied for a position in Groningen, but was turned down by the education department because he lacked the necessary diplomas. Stieltjes' friend and fellow mathematician Charles Hermite solved this problem by getting the University of Leiden to grant Stieltjes an honorary doctorate (Lorentz gave his support for the award). In 1885 he and his family went to Paris and shortly thereafter he completed a PhD thesis at the École Normale Superieur under the stewardship of Hermite and Darboux. In 1889 he took up a professorship in Toulouse, France.

Because many texts on Fourier Analysis use the notation of and work with Fourier-Stieltjes integration, it is worth briefly presenting it here.

It has long been known (see Wiener, 1930; Cramér, 1946) that the limit on the right hand side of (1.47) can be expressed as a Fourier-Stieltjes integral

$$f(x) = \int_{-\infty}^{\infty} e^{i\omega x} dG(\omega) \tag{1.48}$$

where the increments $dG(\omega)$ are complex stochastic functions having mean zero for any ω, *i.e.* $\mathbb{E}[dG(\omega)] = 0$. Moreover, the values of $dG(\omega)$ in non-overlapping ω-intervals are statistically independent (Yaglom, 1952, 1974). In other words, $G(\omega)$ has orthogonal increments $dG(\omega)$.

The increments non-overlapping increments $dG(\omega)$ and $dG(\omega')$ are statistically independent means that we can write

$$\mathbb{E}[dF(\omega)dF(\omega')^*] = 0, \quad \omega \neq \omega' \tag{1.49}$$

Hence there is a function $\mathcal{P}(\omega)$ such that

$$\mathbb{E}[dF(\omega)dF^*(\omega')] = \mathcal{P}(\omega)\delta(\omega - \omega')d\omega d\omega' \tag{1.50}$$

where $\delta(\omega)$ is the Dirac delta function. The function $\mathcal{P}(\omega)$ is called the *Power Spectrum*.

[15] It might be argued, by non-mathematicians, that any realisation of a random function is only ever sampled at a finite number of frequencies, so there is no need to ever go to the limit in equation (1.47). We nevertheless would like to be reassured that the answer does not change as we go to ever higher resolutions. That is why, when developing the theory of realisations of random functions, mathematicians consider the conditions under which that limit can be guaranteed to exist.

The analogue of (1.48) for the Fourier-Stieltjes representation of a density field fluctuations of amplitude $\delta\rho(\mathbf{x})$, in 3-dimensions is to write

$$\delta\rho(\mathbf{x}) = \int e^{i\mathbf{k}\cdot\mathbf{r}} dZ(\mathbf{k}) \tag{1.51}$$

for some function $Z(\mathbf{k})$. In the cosmological literature we generally write $dZ(\mathbf{k}) = \delta_{\mathbf{k}} d^3\mathbf{k}$.

Having said all that, whenever we see $dG(\omega)$ we can think of $dG(\omega = g(\omega)d\omega$ whether or not there is reason to be concerned that it may not be true[4]. In practice we generally assume that the field exists in some finite box having periodic boundary conditions, and that the functions $f(x)$ and $g(\omega)$ are defined on a regular grid. We never in fact let the box size go to infinity or let the grid elements become infinitely small. This is the concept that has implicitly been used in the most discussions of cosmological perturbations.

1.7.3 Wiener-Khinchine theorem

Let us now return to our random function $X(t)$ defined on the real line, and its realisations $x(t)$. We can, without loss of generality, take the random function $x(t)$ to be a zero-mean complex-valued random function. We have seen (equation 1.42) that the covariance of $x(t)$ is $R(t,t') = \mathbb{E}[x(t)\,x(t')]$ and that the process $X(t)$ is *stationary* if $R(t,t')$ is a function only of the distance between the two points t,t': $R(t,t') = R(|\,t-t'\,|)$.

We shall write the auto-correlation function of the random function $x(t)$ as

$$R(\tau) = \mathbb{E}[x(t)\,x^*(t+\tau)] \tag{1.52}$$

In practice, this expectation is calculated using the underlying probability density $f(x)$ of the random process $X(t)$.

It can be shown that there exists a *spectral function* $\mathscr{P}(\omega)$ such that

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}\, d\mathscr{P}(\omega) \tag{1.53}$$

The statement of the existence of such a function $\mathscr{P}(\omega)$ for a stationary random process $X(t)$ is the *Wiener-Khinchine theorem*.

The theorem is non-trivial since does it not require the existence of the Fourier transform of the process $f(x)$ (and indeed the Fourier transform does not exist if $f(x)$ is not square-integrable or absolutely integrable). However, if $\mathscr{P}(\omega)$ is itself absolutely continuous, then there exists a function $\mathcal{P}(\omega) = \mathscr{P}'(\omega)$ which we call the *Power Spectral density*, or simply *the Power Spectrum*. We shall henceforth write

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}\, \mathcal{P}(\omega)\, d\omega \tag{1.54}$$

with its inverse transform

$$\mathcal{P}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau} d\tau. \tag{1.55}$$

From equation (1.52) we see that $R(0) = \mathbb{E}[x(t)^2] \geq 0$, so from (1.54)

$$\int_{-\infty}^{\infty} \mathcal{P}(\omega)\, d\omega \geq 0. \tag{1.56}$$

We can in fact go one step better than this and show that $\mathcal{P}(\omega) \geq 0$ for all ω. The power spectrum $\mathcal{P}\omega$ is never negative-valued. That implies that not all functions $R(\tau)$ can be covariance functions.

1.7.4 The convolution theorem

An important way of combining two functions $f(x)$ and $g(x)$ is the process of *convolution*. Formally the convolution of $f(x)$ and $g(x)$ is the function

$$h(x) = \int_{-\infty}^{\infty} g(y) f(x-y) dy \tag{1.57}$$

So at every point y, the function $h(x)$ is the weighted sum of the values of f about each point x, where the value of the weighting factor g depends on the distance from the point x. In practical terms we might think of $f(x)$ being smoothed or blurred by $g(y)$.

Using our convention for the Fourier transform, the convolution theorem asserts that

The Convolution Theorem

$$H(\omega) = 2\pi F(\omega) G(\omega) \tag{1.58}$$

The Fourier transform of the convolution of two functions is the product of their Fourier transforms.

The factor 2π appears here because of our convention using the normalisation $A = 1, B = 1/2\pi$ in equation (1.20).[16]

The convolution of two function $f(x)$ and $g(x)$ is often denoted by

$$f \star g = \int_{-\infty}^{\infty} g(y) f(x-y) dy \tag{1.60}$$

or sometimes the more explicit $(f \star g)(x)$ to emphasise the dependence. In this notation we have

$$f \star g = g \star f \tag{1.61}$$

$$f \star (g \star h) = (f \star g) \star h \tag{1.62}$$

$$f \star (g + h) = f \star g + f \star h. \tag{1.63}$$

If we denote the operation of taking the Fourier transform by $\mathcal{F}(f) = F(\omega)$ we can express the convolution theorem as

$$\mathcal{F}(f \star g) = \mathcal{F}(f)\mathcal{F}(g) = F(\omega)G(\omega) \tag{1.64}$$

[16] The proof is very simple:

$$\begin{aligned} H(\omega) &= B \int_{-\infty}^{\infty} h(x)\, e^{-i\omega x} dx = B \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, g(y) f(x-y)\, e^{-i\omega x} \\ &= B \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dz\, dy\, g(y) f(z) e^{-i\omega(y+z)} = \frac{1}{B} \left(B \int_{-\infty}^{\infty} g(y)\, e^{-i\omega y} dy \right) \left(B \int_{-\infty}^{\infty} f(z)\, e^{-1\omega z} dz \right) \\ &= \frac{1}{B} F(\omega) G(\omega) \end{aligned} \tag{1.59}$$

(The transformation $z = x - y$ has $dz = dx$ at fixed y). Our convention is $B = 1/2\pi$.

In this notation we have such niceties as

$$\mathcal{F}\{\mathcal{F}\{f(x)\}\} = f(-x), \quad \mathcal{F}\{\mathcal{F}\{\mathcal{F}\{\mathcal{F}\{f(x)\}\}\}\} = f(x). \tag{1.65}$$

1.8 The power spectrum and correlation function

If $R(\tau)$ is such that $\int_{-\infty}^{\infty} |R(\tau)|\, d\tau < \infty$, then there exists a function $\mathcal{P}(\omega)$ such that

Relationship between auto-correlation function and the Power Spectrum:

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}\mathcal{P}(\omega)\, d\omega, \quad \mathcal{P}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}R(\tau)\, d\tau \tag{1.66}$$

The auto-correlation function, $R(\tau)$, and the power spectrum, $\mathcal{P}(\omega)$, are a Fourier transform pair.

The function $\mathcal{P}(\omega)$ is called the *Spectral Density* of the random process $f(x)$ and is also referred to as the *Power Spectrum*.[17]

It is clear from the first of equations (1.66) that

$$R(-\tau) = R^*(\tau) \quad \Rightarrow \quad \mathcal{P}(\omega) \text{ is a real-valued function of } \omega \tag{1.67}$$

We note also that if $f(x)$ is a real valued function, then

$$R(-\tau) = R(\tau) \quad \text{if } f(x) \text{ is real-valued} \tag{1.68}$$

It can further be shown that $P(k)$ is non-negative:

$$R(0) \geq 0, \quad \text{and} \quad \mathcal{P}(\omega) \geq 0, \quad \forall\omega. \tag{1.69}$$

This shows that not all functions $R(\tau)$ as defined by (1.66) can be auto-correlation functions: $R(\tau)$ must have a non-negative Fourier transform.

1.8.1 Real-valued stationary random fields

If the random process $f(x)$ is a real-valued function, as is most often the case, then $R(\tau)$ is real and $R(\tau) = R(-\tau)$. [18] The Fourier transforms then simplify to

$$\mathcal{P}(k) = \frac{2}{\pi}\int_0^{\infty} R(\tau)\cos\omega\tau\, d\tau, \quad R(\tau) = 2\int_0^{\infty} \mathcal{P}(\omega)\cos\omega\tau\, d\omega \tag{1.70}$$

In this way we can rid ourselves of the negative frequencies that would otherwise appear in our expressions.

We notice that if, as an example, we consider the zero-mean random velocity field $u(x)$, then by (1.70) the variance of the velocity field $\mathbb{E}[u(x)u(x)] = \overline{u}^2$ is given by

$$\tfrac{1}{2}\overline{u}^2 = \int_0^{\infty} S(\omega)d\omega. \tag{1.71}$$

[17] The relationships (1.66) and their normalisation factors are standard and do not depend on the which definition one takes for the normalising factors in the Fourier transform.

[18] This is not the case, for example, in quantum mechanics.

Here $S(\omega)$ denotes the power spectrum of the the random field $u(x)$. The power spectrum, integrated over all frequencies, is the mean kinetic energy. We can interpret the quantity $S(\omega)d\omega$ as the contribution to the mean kinetic energy from Fourier modes in the frequency range $[\omega, \omega + d\omega]$. This goes some way to rationalising the choice of normalisation of the power spectrum as used in fluid dynamics.

1.9 Discrete Fourier Transform (DFT)

Here we will start from the description of data in a finite box, where the data has been sampled on a discrete rectangular grid. This corresponds rather closely to the situation in image data samples and in numerical N-Body simulations. There is a discussion of the "DFT" in Press et al. (2007, §12.1) which uses the same conventions as used here [19].

There are two important limits which we have already touched on. We can take the limit as the number of grid points gets larger while keeping the box the same size, this gives us an increasingly high resolution view of the data in the box. Alternatively, we can keep our grid cells fixed and let the size of the box increase, adding grid cells as necessary.

The preceding transformations are applicable when the analytic form of the function $f(x)$ is known, or is, in principle, knowable. If the values of $f(x)$ are defined at a set of discrete points these may be less than useful.

1.9.1 Definition of the discrete Fourier transform

We start by considering a function $f(x)$ defined on a set of uniformly spaced points $x_0, x_1, \ldots x_{N-1}$ where it has values $f(x_0), f(x_1), \ldots, f(x_{N-1})$. The constant interval between successive points will be written as $\Delta x = x_{i+1} - x_i$, so the i^{th} point is at $x_i = x_0 + i\Delta x,\ i = 1, 2, \ldots, N-1$.

The discrete Fourier representation of this function $f(x)$ is defined by the equation

$$F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)\ \exp\left[-\frac{2\pi i u x}{N}\right], \quad u = 0, 1, \ldots, N-1 \tag{1.72}$$

It is straightforward to verify that this can be inverted so as to express $f(x)$ in terms of $F(u)$:

$$f(x) = \sum_{u=0}^{N-1} F(u)\ \exp\left[\frac{2\pi i u x}{N}\right], \quad x = 0, 1, \ldots, N-1 \tag{1.73}$$

These bear a superficial resemblance to the complex Fourier series (1.9) on an interval and the Fourier transform (1.16) on the real line. In particular we note the parallel with the convention we have adopted for the normalising factors of the continuous Fourier transform (equations 1.16): the normalising factor is $1/N$ in the forward transform, $f(x) \to F(u)$ and unity in the backward transform $F(u) \to f(x)$.

[19] Note however that Press et al. (2007, §21.0) use the "$B = 1$" convention for the continuous Fourier transform, *cf.* equations (1.20) and (1.21). However, because of their choice of frequency sign, it is simple to interchange the roles of which variables are real-space and which are Fourier-space.

1.9.2 Some properties of the discrete Fourier transform

This discrete Fourier transform has equivalents to the properties of the other Fourier transforms. The discrete equivalent of Parseval theorem (see §1.2.3) becomes

$$\sum_{n=0}^{N-1} |F(u)|^2 = \frac{1}{N}\sum_{n=0}^{N-1} |f(n)|^2 \tag{1.74}$$

It is easy to verify that transforming from $f \rightarrow F$ using (1.72) and then transforming the result $F \rightarrow g$ using (1.73) results in $g = f$.[20] The Fourier transform and its inverse are periodic:

$$F(x) = F(x+N), \quad F(u) = F(u+N) \tag{1.76}$$

as can be verified by direct substitution.[21] The Fourier transform also exhibits what is referred to as *conjugate symmetry*:

$$F(u) = F^*(-u), \quad |F(u)| = |F(-u)| \tag{1.78}$$

This periodicity and the symmetry in the modulus of the Fourier transform is very important: it tells us that the terms $|F([N/2]+1)|, \ldots |F(N-1)|$ are the same as the terms $|F([N/2]-1)|, \ldots, |F(0)|$. If N is even, there are $[N/2]$ independent amplitudes and $[N/2]$ independent phases.

1.9.3 Convolution theorem

The all-important convolution theorem is also available if we define the convolution [22] by

$$h(x) = f \star g = \sum_{n=0}^{N-1} f(n)\, g(x-n), \quad x = 0, 1, \ldots, N-1. \tag{1.80}$$

However, this equation is not quite as self-evident as it looks. It might be that the function $f(x)$ is defined on a set of R points, while $g(x)$ is defined on a set of S points. So what then is the appropriate value to be assigned to N in this equation? The constraint on N

[20] Proof of this requires the identity

$$\sum_{x=0}^{N-1} e^{2\pi irs/N} e^{-2\pi isx/N} = \begin{cases} N, & r = s \\ 0, & \text{otherwise} \end{cases} \tag{1.75}$$

[21] This works as follows:

$$F(u+N) = \frac{1}{N}\sum_{n=0}^{N-1} f(n) \exp\left[-\frac{2\pi i(u+N)n}{N}\right] = \frac{1}{N}\sum_{n=0}^{N-1} f(n) \exp\left[-\frac{2\pi iun}{N}\right] e^{-2\pi in} = F(u) \tag{1.77}$$

The proof of (1.78) follows on similar lines.

[22] It should be remarked that the convolution described by the equation

$$k(x) = f \circ g = \sum_{n=0}^{N-1} f(n)\, g(x+n), \quad x = 0, 1, \ldots, N-1. \tag{1.79}$$

is referred to as the *correlation* between f and g.

arises from the requirement that the Fourier transform of $h(x)$ be periodic with distinct, non-overlapping periods.

Whatever the value of N, provided it is larger than both of R and S, we can pad the f and g values with zeros to make them both arrays of size N. These padded sequences are referred to as *extended sequences*. Naïvely, one might think that to choose N equal to the larger of R and S would suffice, but then the periods of the convolution h would overlap. That is referred to as *wrap-around error*. The answer is that

$$N \geq R + S - 1 \tag{1.81}$$

The bigger the difference between N and $R + S - 1$ the more separated will be the cycles.

The convolution theorem is easily derived:

$$H(u) = \mathcal{F}(f \star g) = F(u)G(u) \tag{1.82}$$

1.9.4 Matrix formulation of the discrete Fourier transform

If we define

$$W = e^{2\pi i/N} \tag{1.83}$$

then the forward discrete Fourier transform (1.72) can be written as

$$F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)W^{-ux}, \quad u = 0, \ldots, N-1. \tag{1.84}$$

We note that $W^N = 1$, so W is the N^{th} root of unity. Accordingly we sometimes emphasise this fact by writing it as W_N.

If we define the vectors $\mathbf{f} = \{f(x)\}$ and $\mathbf{F} = \{F(u)\}$ we can write this as a matrix product

$$\mathbf{F} = N^{-1}\,\mathbf{F}\,\mathbf{f} \tag{1.85}$$

$$\begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ \vdots \\ F_{N-1} \end{pmatrix} = N^{-1} \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & W^{-1} & W^{-2} & \ldots & W^{-(N-1)} \\ 1 & W^{-2} & W^{-4} & \ldots & W^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W^{-(N-1)} & W^{-2(N-1)} & \ldots & W^{-(N-1)^2} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{pmatrix} \tag{1.86}$$

We see that the j,k entry of the matrix $\mathbf{F}$ is just W^{jk}, where the values of the indices j and k range over $0, \ldots, N-1$. The fact that W_N is an N^{th} root of unity plays an important role. Noticing, for example, that $W^{-j(k+N)} = W^{-jk}$ we can recover the periodicity properties of the Discrete Fourier transform.[23]

[23] The matrix if W's in equation (1.86) is an example of a *Vandermonde Matrix*. Such matrices have the generic form

$$\mathbf{V} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_0 & x_1 & \ldots & x_{n-1} \\ \vdots & \vdots & \ldots & \vdots \\ x_0^{n-1} & x_1^{n-1} & \ldots & x_{n-1}^{n-1} \end{pmatrix} = V(x_0, \ldots, x_{n-1}) \tag{1.87}$$

We can more easily appreciate this if, for example, we set $N = 4$ for which the matrix $\mathbf{F}$ is simply

$$\mathbf{F}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & W_4 & W_4^2 & W_4^3 \\ 1 & W_4^2 & W_4^4 & W_4^6 \\ 1 & W_4^3 & W_4^6 & W_4^9 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & W & W^2 & W^3 \\ 1 & W^2 & 1 & W^2 \\ 1 & W^3 & W^2 & W \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \tag{1.88}$$

where in the second matrix we have dropped the subscript "4" on W_4 for clarity. This follows from the fact that $W_N^{pq} = W_M^{pq \bmod N}$, W_N being an N^{th} root of unity. We shall return to this below.

For large matrices where the value of N can be factored into a product of two numbers, such matrices can be considerably simplified. This leads to a variety of algorithms that reduce the number of operations required in computing the matrix product that is implicit in the discrete Fourier transform.

1.10 The Fast Fourier Transform: FFT

The problem with the deceptively simple-looking equations (1.72) and (1.73) is that they are computationally inefficient. In equation (1.72) there are N additions, and each of these N additions requires another N operations. While for modest N-values this would not be a problem with modern computers, for very large N, or in large 3-dimensional problems this can be the limiting factor for the size of data sets that can be handled.[24]

1.10.1 FFT: Danielson, Lanczos, Cooley, Tukey

Over 200 years ago, in 1805 and two years before Fourier's work, Gauss solved this problem (Heideman et al., 1985) but it was not until the work of Danielson and Lanczos (1942), independently rediscovering what Gauss had done, that any further progress was made. The Danielson-Lanczos theorem lies at the heart of the Fast Fourier Transform (see (Press et al., 2007, §12.2)).

What Danielson and Lanczos did was to show that a discrete Fourier transform of $2N$ points could be reduced to two N-point transforms, one for the odd-numbered points and the other for the even-numbered points. So, if N were a power of 2, the process could be recursively reduced to a set of trivial 1-point transforms. The 1-point transform is simply one complex multiplication and addition, so this saves a huge amount of computational effort. The number of computations for an N-point transform is $O(N \log_2 N)$.

The argument is simple, for N divisible by 2 we can write the Fourier transform (1.72)

See Golub and van Loan (1996, §4.6, 4.6.4) for a discussion of such matrices and a written-out example of the FFT.

[24] For $N = 64$ points the FFT speed-up is a factor of ~ 10, while for 1024 points the speed-up is a factor of ~ 100 and for 32000 points it is over ~ 1000.

as the sum of an odd-point and an even-point transform:

$$F(u) = \frac{1}{N}\sum_{n=0}^{N/2-1} W^{u(2n)} f_{2n} + \frac{1}{N}\sum_{n=0}^{N/2-1} W^{u(2n+1)} f_{2n+1}, \quad W = e^{2\pi i u/N} \tag{1.89}$$

$$= \frac{1}{2}\left(\frac{2}{N}\sum_{n=0}^{N/2-1} (W^2)^{un} f_{2n} + W^u \frac{2}{N}\sum_{n=0}^{N/2-1} (W^2)^{un} f_{2n+1}\right) \tag{1.90}$$

$$\tag{1.91}$$

We now need to notice that

$$F^{Even}(u) = \frac{2}{N}\sum_{n=0}^{N/2-1} (W^2)^{un} f_{2n}, \quad u = 0, 1, \ldots, N/2-1 \tag{1.92}$$

is the $N/2$-point Fourier transform of the even-numbered ordinates f_{2n}. $n = 0, \ldots, N/2-1$. Likewise

$$F^{Odd}(u) = \frac{2}{N}\sum_{n=0}^{N/2-1} (W^2)^{un} f_{2n+1}, \quad u = 0, 1, \ldots, N/2-1 \tag{1.93}$$

is the $N/2$-point Fourier transform of the even-numbered ordinates f_{2n+1}. $n = 0, \ldots, N/2-1$. It is important to notice that $F^{Even}(u)$ and $F^{Odd}(u)$ are of dimension $N/2$, and so do not themselves cover the $N/2$ points $u = N/2, dots, N-1$. With these definitions and $W_N^{u+N/2} = -W_4^u$, we arrive at

Danielson-Lanczos Lemma

$$F(u) = \frac{1}{2}\left(F^{Even}(u) + W^u F^{Odd}(u)\right), \quad u = 0, \ldots, N-1 \tag{1.94}$$

$$F(u+N/2) = \frac{1}{2}\left(F^{Even}(u) - W^u F^{Odd}(u)\right), \quad u = 0, \ldots, N-1 \tag{1.95}$$

These last lines are the Danielson-Lanczos Lemma.

Two other papers appeared in 1958 presenting different approaches to computing the Fourier sum: Good (1958) and Goertzel (1958).[25]

The FFT revolution came in 1965 with the publication of the seminal paper of Cooley and Tukey (1965). The existence of the paper is generally attributed to Richard Garwin, who at the time was a physicist the IBM Watson Laboratory at Columbia University.[26] Garwin had visited Tukey at Princeton in 1963 and discussed the computation of Fourier transforms, and then went on to ask Cooley of the IBM Watson Research Center at Yorktown Heights to write a computer program implementing the scheme described by Tukey.

[25] The Goertzel algorithm is subject to numerical instabilities due to round-off errors, but it is the fastest algorithm for calculating just a few Fourier coefficients.

[26] Garwin was certainly the 'facilitator' of the FFT. Garwin, on instructions from Edward Teller, had previously led the design of the first, proof of concept, Hydrogen bomb in 1952 (named 'Mike'). His involvement with nuclear weapons continued into the 1960's when he was working on methods of verifying the SALT agreement on nuclear testing: he needed the FFT for the seismic detectors. The detectors were built and installed around the Soviet Union and could locate an explosion to within 15-20 kilometres.

Cooley wrote the draft of the paper and, reportedly, it was Tukey who added in the reference to Good's paper.[27]

The importance of the algorithm was recognised and was put into the public domain. [28]

1.10.2 FFT: a simple example using 4 points

There are many explanations of how the FFT works ranging from chapters in books and lecture notes to entire books. The classic text is Brigham (1988, *The Faster Fourier Transform and Its Applications*). Among the best shorter reads are (Gentleman and Sande, 1966, *Fast Fourier Transfroms - for Fun and Profit*) written shortly after the publication of the Cooley-Tukey algorithm, and of course Press et al. (2007).

If we look at the $N = 4$ example in equation (1.88) we can see how the FFT works. It turns out that we can factorise $\mathbf{F}_4$ as follows:

$$\mathbf{F}_4 = \underbrace{\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & W^1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -W^1 \end{pmatrix}}_{} \underbrace{\begin{pmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^0 \\ 0 & 0 & W^0 & W^2 \end{pmatrix}}_{Block diagonal} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{Permutation} \tag{1.96}$$

$$= \begin{pmatrix} \mathbf{I}_2 & \mathbf{W}_2 \\ \mathbf{I}_2 & -\mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \mathbf{F}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{pmatrix} \underbrace{\begin{pmatrix} & \\ & \end{pmatrix}}_{Permutation}, \quad \mathbf{F}_2 = \begin{pmatrix} 1 & W^0 \\ W^0 & W^2 \end{pmatrix}, \quad \mathbf{W}_2 = \begin{pmatrix} 1 & 0 \\ 0 & W^1 \end{pmatrix} \tag{1.97}$$

We can now see how this works on an input vector $\mathbf{f} = (f_0, f_1, f_2, f_3)$. The permutation matrix does the following:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} f_0 \\ f_2 \\ f_1 \\ f_3 \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{even} \\ \mathbf{f}^{odd} \end{pmatrix} \tag{1.98}$$

It rearranges the elements of the input data into the even-numbered elements of the original input vector, followed by the odd-numbered elements. This is precisely the split required for application of the Danielson-Lanczos theorem (equations 1.94 and 1.95). If we now perform the next matrix multiplication we have

$$\begin{pmatrix} \mathbf{F}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{pmatrix} \begin{pmatrix} \mathbf{f}^{even} \\ \mathbf{f}^{odd} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_2\mathbf{f}^{even} \\ \mathbf{F}_2\mathbf{f}^{odd} \end{pmatrix} \tag{1.99}$$

Finally, the last matrix multiplication gives

$$\mathbf{F}_4\mathbf{f} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{W}_2 \\ \mathbf{I}_2 & -\mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \mathbf{F}_2\mathbf{f}^{even} \\ \mathbf{F}_2\mathbf{f}^{odd} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_2\mathbf{f}^{even} + \mathbf{W}_2\mathbf{F}_2\mathbf{f}^{odd} \\ \mathbf{F}_2\mathbf{f}^{even} - \mathbf{W}_2\mathbf{F}_2\mathbf{f}^{odd} \end{pmatrix} \tag{1.100}$$

[27] The Cooley and Tukey (1965) paper cites only two references: the work of Danielson and Lanczos (1942) was not cited and was presumably not known to either of the authors.

[28] There is an extensive interview with James Cooley about the origins of the FFT algorithm, and in particular about the patent issue, available from the IEEE Global History network at
`http://www.ieeeghn.org/wiki/index.php/Oral-History:James_W._Cooley`
(James W. Cooley: an oral history conducted in 1997 by Andrew Goldstein, IEEE History Center, New Brunswick, NJ, USA)

The factors on the right hand side are the factors that appear in the Danielson-Lanczos Lemma (1.94, 1.95). The question of where the factorisation (1.96) came from is discussed in detail in van Loan (1987, Ch.1.).

The key feature here is the reduction of $\mathbf{F}_4$ to $\mathbf{F}_2$. The process is clearly recursive and so this factorisation of the original matrix performs the required reduction. Of course, if we were to replace the subscript "4" by "512" we would have a reduction of $\mathbf{F}_{512}$ to $\mathbf{F}_{256}$. One of the factors in gaining a speed advantage is the sparsity of the matrices involved in the computation: they have many zeroes (there are 48 elements in the three matrices in (1.96) of which 28 are zero). Conversely, one of the tricky parts is an efficient implementation of the permutation of the odd an even members of the data array at each stage of the recursion. Cooley and Tukey came up with an ingenious and efficient way of doing this involving representation of the indices on the data vectors by binary numbers and then reversing the bits to get the resorted indices (see Press et al. (2007, §12.2)).

1.10.3 FFT: source code for programming

There are many implementations of the Fast Fourier Transform, notably Press et al. (2007, *Numerical Recipes*). Another popular source is the"FFTW05" of Frigo and Johnson (2005), which according to the web site [29] "FFTW is a C subroutine library for computing the discrete Fourier transform (DFT) in one or more dimensions, of arbitrary input size, and of both real and complex data (as well as of even/odd data, i.e. the discrete cosine/sine transforms or DCT/DST)".

1.11 The Fourier transform in 2 dimensions

The generalisation of the Discrete Fourier Transform, ("DFT"), to two, three and higher dimensions is relatively straightforward, but involves some complications that are best illustrated in the simpler 2-D case.[30]

1.11.1 Variants of Discrete Fourier transform in 2 dimensions

Consider a function $f(x,y)$ defined on an $M\times N$ grid on which the points are simply labelled by their indices $m = 0,1,\ldots,M-1$ and $n = 0,1,\ldots,N-1$. The space of the discrete Fourier transform of $f(m,n)$ is also discrete and occupies an $M \times N$ space on which we shall label the grid points by (u,v), $u = 0,1,\ldots,M-1$ and $v = 0,1,\ldots,N-1$. We can refer to the Fourier space representation as the (u,v)-*plane representation*. The 2-D Fourier transform

[29] The FFTW package was developed at MIT by Matteo Frigo and Steven G. Johnson. The web site is at `http://www.fftw.org/`
FFTW is distributed freely under the GNU GPL license.

[30] This section owes much to the discussion of image compression in the venerable, but rather outdated, book by Rosenfeld and Kak (1982, §§2.2, 5.3). The book is renowned for its clarity of exposition of the image processing techniques that were available at that time.

from the (m,n) plane to the (u,v) plane is defined as

$$F(u,v) = \frac{1}{MN}\sum_{0}^{M-1}\sum_{0}^{N-1} f(m,n)\exp\left[-2\pi i\left(\frac{mu}{M}+\frac{nv}{N}\right)\right], \quad u = 0,\ldots M-1,\ v = 0\ldots,N-1 \tag{1.101}$$

which can be shown to have has inverse

$$f(m,n) = \sum_{0}^{M-1}\sum_{0}^{N-1} F(u,v)\exp\left[2\pi i\left(\frac{mu}{M}+\frac{nv}{N}\right)\right], \quad m = 0,\ldots M-1,\ n = 0\ldots,N-1 \tag{1.102}$$

These are the direct generalisations of the 1‘-dimensional transforms (1.72) and (1.73).

We have the identity

$$\sum_{0}^{M-1}\sum_{0}^{N-1} |\, f(m,n)\,|^2 = MN\sum_{0}^{M-1}\sum_{0}^{N-1} |\, F(u,v)\,|^2 \tag{1.103}$$

We can rearrange the terms in (1.101) so that it appears as

$$F(u,v) = \frac{1}{M}\sum_{0}^{M-1}\frac{1}{N}\left[f(m,n)\exp\left(-2\pi i\frac{n}{N}v\right)\right]\exp\left(-2\pi i\frac{m}{M}u\right) \tag{1.104}$$

which will be recognised as the 1D u-direction Fourier transform of the 1D Fourier transform of $f(m,n)$ in the v-direction. Thinking of $f(m,n)$ as a matrix, this equation is the Fourier transform of the rows, in which each row is replaced by its Fourier transform, followed by the Fourier transform of the updated columns. In other words, computing the 2D Fourier transform of a 2D function $f(m,n)$ can be executed simply as two successive fast Fourier transforms, one in either direction.

By virtue of the periodicity of the exponentials $f(m,n)$ and $F(u,v)$ exhibit the following symmetries:

$$\begin{array}{ll} f(-m,n) = f(M-m,n) & F(u,-v) = (u,N-v) \\ f(m,-n) = f(m,N-n) & F(-u,v) = (M-u,v) \\ f(-m,-n) = f(M-m,N-n) & F(-u,-v) = (M-u,N-v) \end{array} \tag{1.105}$$

This illustrates the periodic extension of the functions defined by their discrete Fourier transform.

1.11.2 Centring the origin of the (u,v)-plane

As a consequence of these symmetries, we can, if we wish, put the origin of the coordinates (u,v)-plane at the point $(M/2,N/2)$ in the centre of the grid. To achieve this we need to redefine the real-space function $f(m,n)$ as

$$\hat{f}(m,n) = (-1)^{m+n}\, f(m,n) \tag{1.106}$$

Taking the Fourier transform of $\hat{f}(m,n)$ yields a function $\hat{F}(u,v)$ that, in the light of the previous symmetries, is easily shown to be related to the original $F(u,v)$ by

$$\hat{F}(u,v) = F(u-\frac{M}{2},v-\frac{N}{2}). \tag{1.107}$$

In other words $F(u,v)$ and $\hat{F}(u,v)$ display the same information, but $F(0,0)$ is located at the centre of the (u,v) grid when we plot $\hat{F}(u,v)$.

1.12 Other (real-valued) Fourier-like transforms

The discrete Fourier transform takes on complex values in both the real (m,n) space and in the Fourier (u,v) space. There are two important transforms analogous to the Fourier transform that use on real valued functions and real valued transforms: the Cosine transform and the Hartley transform. Technically these are mappings $\mathbb{R}^N \to \mathbb{R}^N$. Here we continue with our previous notation. We denote a real-valued function defined on a 1-dimensional grid on which the points are labelled with index $n = 0, 1, \ldots N$ and its transform as $f(n)$ and $F(u)$ respectively. In 2-dimensions, we analogously denote the transform pair by $f(m,n)$ and $F(u,v)$.

1.12.1 Discrete Cosine transform in 1 and 2 dimensions

The one-dimensional cosine transform of a function $f(m)$ is defined as

$$F(u) = c(u)\frac{2}{N}\sum_{m=0}^{N-1} f(m)\cos\left(m+\frac{1}{2}\right)\frac{\pi u}{N}, \qquad c(u) = \begin{cases} 1/\sqrt{2}, & u = 0 \\ 1, & u = 1, 2. \ldots, N-1 \end{cases} \tag{1.108}$$

and the inverse transform is

$$f(m) = \sum_{u=0}^{N-1} c(u)\,F(u)\cos\left(m+\frac{1}{2}\right)\frac{\pi u}{N}. \tag{1.109}$$

The 2-D version is very similar:

$$F(u,v) = 4c(u,v)\frac{1}{MN}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n)\cos\left(m+\frac{1}{2}\right)\frac{\pi u}{M}\ \cos\left(n+\frac{1}{2}\right)\frac{\pi v}{N} \tag{1.110}$$

$$c(u,v) = \begin{cases} 1/2, & u = v = 0 \\ 1, & u = 1, 2, \ldots, M-1 \\ 1, & v = 1, 2, \ldots, N-1 \end{cases} \tag{1.111}$$

The inverse is

$$f(m,n) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c(u,v)\,F(u,v)\cos\left(m+\frac{1}{2}\right)\frac{\pi u}{M}\ \cos\left(n+\frac{1}{2}\right)\frac{\pi v}{N} \tag{1.112}$$

$$m = 0, 1, \ldots, M-1 \tag{1.113}$$

$$n = 0, 1, \ldots, N-1 \tag{1.114}$$

There are, of course, fast versions of the Cosine transform.

1.12.2 Fourier or cosine transform - what's the difference?

The cosine transform is manifestly related to the Fourier transform. If we consider our function $f(m)$ defined on M points and create a new function $\bar{f}(m)$ that spans $2M$ points by zero-padding:

$$\bar{f}(m) = \begin{cases} f(m), & m = 0, 1, \ldots, M-1, \\ 0, & m = N, N+1, \ldots, 2N-1 \end{cases} \tag{1.115}$$

then we can write the cosine transform as

$$F(u) = 2c(u)\frac{1}{N}\mathcal{R}\left[\exp\left(i\frac{\pi u}{2N}\right)\sum_{m=0}^{2N-1}\bar{f}(m)\exp\left(i\frac{2\pi um}{2N}\right)\right], \quad u = 0, 1, \ldots, N-1. \tag{1.116}$$

Here, $\mathcal{R}(z)$ denotes the real part of the complex number z. This is just a $2N$-point Fourier transform and it provides one way of performing a fast cosine transform.

So it is reasonable to ask whether it makes any difference to use one rather than the other. Perhaps surprisingly, the answer is "yes, the cosine transform encodes real space data more efficiently than the Fourier transform" and that fact is why the cosine transform is used as a way of representing images in the JPEG image compression standard.

The way the JPEG compression works is that the image is divided into square blocks that are a power of 2 in size. Each block is transformed using the cosine transform and the coefficients that make the least contribution to the variance of the image block are set to zero, simply on the grounds that they do not make much contribution to the perceived structure of the reconstructed image in the locality of the block. It turns out that the mean square difference between the original image and the reconstructed image is can be as big as 30%, with the cosine transform offering the better rendition of the original.[31]

1.12.3 Discrete Hartley transform in 1 and 2 dimensions

The discrete Hartley transform is also Fourier-like in that it uses sines and cosines to represent data. In one dimension, the Hartley transform $H(u)$ of $f(m)$ is

$$H(u) = \frac{1}{N}\sum_{m=0}^{M-1} f(m)\left[\cos\left(\frac{2\pi mu}{M}\right) + \sin\left(\frac{2\pi mu}{M}\right)\right], \quad u = 0, 1, \ldots, M-1. \tag{1.117}$$

$$f(m) = \sum_{u=0}^{M-1} H(u)\left[\cos\left(\frac{2\pi mu}{M}\right) + \sin\left(\frac{2\pi mu}{M}\right)\right], \quad m = 0, 1, \ldots, M-1. \tag{1.118}$$

At first sight this simple sum of a sine and cosine might look a little strange, but, in effect, we have exchanged $e^{ix} = \cos x + i \sin x$ with a function that is generally labelled

$$\operatorname{cas} x = \cos x + \sin x = \sqrt{2}\cos(x - \pi/4). \tag{1.119}$$

[31] The larger the block the smaller the mean square difference, but the cosine transform out-performs the Fourier transform for all block sizes. Much of the difference lies in the block boundaries which are perceived as a 'blockiness' in the reconstructed image. Using some particular wavelet transforms of the entire image gives better results even than the cosine transform, by removing the hiatus at the block boundaries, and does so with considerably less computational cost.

The latter equality connects the Hartley transform directly to the discrete cosine transform. The transform is periodic with period M, as is easily verified by substituting $m \to M + m$, and it has the property that $H(-u) = H(M - u)$. This latter relationship allows us to view the Hartley transform on the interval $-M/2 < u \leq M/2$ with the origin near the centre of the $u - axis$, and with a nicer symmetry about this origin.

There is a simple relationship between the coefficients of the Fourier transform, $F(u)$, and Hartley transform, $H(u)$, of $f(m)$:[32]

$$\mathcal{R}F(u) = \tfrac{1}{2}(H(M-u) + H(u)) = \tfrac{1}{2}(H(-u) + H(u)) \tag{1.120}$$

$$\mathcal{I}F(u) = \tfrac{1}{2}(H(M-u) - H(u)) = \tfrac{1}{2}(H(-u) - H(u)) \tag{1.121}$$

$$H(u) = \mathcal{R}F(u) - \mathcal{I}F(u) = \mathcal{R}[(1+i)F(u)] \tag{1.122}$$

since $H(M-u) = H(-u)$ if we extend the frequency space to negative frequencies $-M/2 < u \leq M/2$, just as was done in the Fourier transform. In other words, we define the negative m values of $H(u)$ by

$$H(-u) = H(M-u), \quad -N/2 \leq u < 0. \tag{1.123}$$

We also note from (1.120) and (1.121) that $H(-u) \pm H(u)$ define the even and odd parity parts of the transform $H(u)$.

1.12.4 Hartley versus Fourier

So we see that the Hartley transform simply rearranges the Fourier coefficients and stores the data as real numbers. We reduce M complex numbers of the Fourier domain to M real numbers without loss of information: the Fourier transform can be recovered directly from the Hartley transform. The Hartley transform removes the redundancy that is inherent in the Fourier transform.

It is easy to show that the Power spectrum can be calculated from the Hartley coefficients via

$$P(u) = \tfrac{1}{2}[H(u)^2 + H(N-u)^2] = \tfrac{1}{2}[H(u)^2 + H(-u)^2] \tag{1.124}$$

where the second equality simply uses the extension of the domain of the transform to negative frequencies.

The Hartley transform is often seen in the engineering world, and, in particular, in the representation of time series. Moreover, from a computational point of view the Hartley transform uses fewer operations than the Fourier transform and can be done in-place. So why is the Hartley transform not used more often in the world of physics?

The answer is possibly that the Hartley transform can be seen as a computational device, rather than as a device that provides a physically meaningful or insightful representation of a phenomenon. Fourier transforms are, for example, meaningful within the domains of

[32] The notation used here is that the real and imaginary parts of the complex number $z = a + ib$ are denoted by $a = \mathcal{R}z$ and $b = \mathcal{I}z$.

wave propagation, quantum mechanics, or vibrations, but the Hartley transform conveys no deeper insight into these phenomena.[33]

1.12.5 Examples of Hartley transform

The following are trivial examples of what a Hartley transform looks like by considering the Hartley transform of a function defined on the real line:

$$H(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)(\cos ux + \sin ux)dx \tag{1.125}$$

Using this we can derive the following Hartley transform pair:

$$f(x) = \text{rect}(x) \quad \leftrightarrow \quad H(u) = \frac{\sin \pi a}{\pi a} \tag{1.126}$$

1.12.6 Hartley transforms in 2-dimensions

There are two variants of the discrete Hartley transform in two dimensions:

$$H(u,v) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\text{cas}\left(\frac{2\pi mu}{M}\right)\text{cas}\left(\frac{2\pi nu}{N}\right) \tag{1.127}$$

$$H(u,v) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\text{cas}\left(\frac{2\pi mu}{M} + \frac{2\pi nu}{N}\right) \tag{1.128}$$

where the function $\text{cas}\, x$ is as defined in (1.119). The first of these is a "separable transform": it can be resolved as a product of two 1-d transforms and so allows the use of the fast Hartley transform for its evaluation. The second has all the symmetries analogous to the 1-d transform and has a simple convolution theorem, but it is not separable and so cannot benefit from the standard fast mode of evaluation.

1.13 Discrete z-transform in 1 and 2 dimensions

1.14 The Fourier transform in 3 dimensions

The analogous equations in 3 spatial dimensions define the Fourier transform $g(\mathbf{k})$ of a function $f(\mathbf{r})$ in $\mathbb{R}^3$ as

[33] The same could not be said of the wavelet transform which adds the feature of multi-resolution (scaling) to the transform in a way that is notat all evident in the domain of the Fourier transform.

Fourier transform on $\mathbb{R}^3$:

$$f(\mathbf{r}) = \int_{\mathbf{k}} g(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3\mathbf{k}, \quad g(\mathbf{k}) = \frac{1}{(2\pi)^3} \int_{\mathbf{x}} f(\mathbf{r})\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, d^3\mathbf{r} \tag{1.129}$$

Here we have adopted the standard notation that a 3-space wave-vector is denoted by the symbol $\mathbf{k}$. The symbol ω used in the 1-dimensional version (1.16) is generally thought of as a temporal frequency, though there is no hard and fast rule about this.

The multi-dimensional Fourier transform is said the be *separable* since its kernel $K(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{x}}$ has the property that it can be written as the product

$$K(x_1, x_2, x_3, k_1, k_2, k_3) = K_1(x_1, k_1)K_2(x_2, k_2)K_3(x_3, k_3). \tag{1.130}$$

where (x_1, x_2, x_3) are the components of $\mathbf{r}$ in a real-space Cartesian coordinate system, and (k_1, k_2, k_3) are the components of the vector $\mathbf{k}$ in a Cartesian coordinate system in Fourier space. Consequently, the 3-dimensional Fourier transform and its inverse (1.129) can be evaluated by performing the integrals in each direction, one after another. The kernel $K(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{x}}$ is also *symmetric* since each of the separated components, K_1, K_2, K_3 has the same functional form. Separability and symmetry are of particular importance in numerical work when using the discrete Fourier transform in multiple dimensions.

1.15 The Fourier transform in cosmology

One of the most important measures of the inhomogeneity of the distribution of cosmic material is the variance of the density in sample volumes of differing size. In other words, how "rough" the distribution is when view on different scales. If the universe is indeed homogeneous on the largest scales, then the measure of the roughness should tend to zero as the averaging scale gets larger.

Towards this end, it is convenient to decompose the density inhomogeneities into Fourier components. However, for a homogeneous stationary stochastic function $f(\mathbf{x})$ defined on an infinite volume, the function $g(\mathbf{k})$ in equation (1.129) may not be defined, in which case the notion of handling the individual Fourier amplitudes is not strictly admissible.[34]

1.15.1 Describing a field of random density fluctuations

The standard way around such 'technical problems' in cosmology is to consider a box of finite volume V which is far larger than the length scale any of the phenomena of interest without considering what happens when $V \to \infty$: the answer is always correct as long as the box is finite.

[34] A formal mathematical requirement for the existence of the Fourier transform is that $\int |f(\mathbf{x})| d\mathbf{x}$ should be bounded. The spectral analysis of turbulent fluid flow has a long history of facing up to this technicality by working directly with the statistically orthogonal increments $dG(\mathbf{k})$ which can be thought of as $g(\mathbf{k})\, d\mathbf{k}$ when $g(\mathbf{k})$ exists (Batchelor, 1982). See Section 1.7 where this is described in more detail.

The basic premise, as described by Kolb and Turner (1990, §9.2), Coles and Lucchin (2002, Ch. 13) and others, is that the density fluctuation field $\delta(\mathbf{x})$ in a finite cubic box of side L and volume V can be decomposed into a sum over discrete Fourier components:

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_{\mathbf{k}} \delta^*_{\mathbf{k}}\, e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{1.131}$$

Formally, this is the 3-dimensional analogue of equation (1.9) for the Fourier series representation of a function defined on an interval. Here, the function $\delta(\mathbf{x})$ is a continuous, random function, while the values of $\mathbf{k}$ are a (possibly infinite) discrete set of wave-numbers. The Fourier amplitudes $\delta_{\mathbf{k}}$ are then given by[35]

$$\delta_{\mathbf{k}} = \frac{1}{V}\int_V \delta(\mathbf{x})\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x} \tag{1.132}$$

Since the field $\delta(\mathbf{x})$ is real, $\delta^*_{\mathbf{k}} = \delta_{-\mathbf{k}}$. Equation (1.132), in the limit of large volumes, has an inverse that can be written[36]

$$\delta(\mathbf{x}) = \frac{V}{(2\pi)^3}\int_V \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \tag{1.133}$$

We have encountered these equations when discussing the normalisation of the Fourier transforms (see equations 1.23).

We notice that both the forward and backward transforms (1.131) and (1.132) are linear functions of the $\delta(\mathbf{k})$ and $\delta(\mathbf{x})$. This means that we can consider a small amplitude density fluctuation field $\delta(\mathbf{x})$ as being made up of independent (Fourier) components having different frequencies. We can subject these components independently to a variety of physical processes, and provided they remain small, we can then re-assemble the results into a new field that is the consequence of the physical processes acting on $\delta(\mathbf{r})$.

1.15.2 Further normalisation issues

This is in part repeated in several places

Matarrese et al. (1997, Appendix A) summarise several important possibilities for deploying the 2π factors in the definition of the Fourier transform, the power spectrum and the bispectrum. In terms of the preceding definitions they write

$$\delta_{\mathbf{k}} = \frac{1}{V}\int \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \tag{1.134}$$

$$\langle \delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\rangle = D\,\mathcal{P}(k)\,\delta_D(\mathbf{k}_1+\mathbf{k}_2) \tag{1.135}$$

$$\langle \delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\delta_{\mathbf{k}_3}\rangle = E\; B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\,\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) \tag{1.136}$$

The constants V, D and E are all independent. We follow Bertschinger (1992) in putting V $= (2\pi)^3$ and $D = E = 1$.

[35] Kolb and Turner (1990) comment in a footnote that "The reader may wish to follow the advice of one of the authors to ignore the various factors of V; they play no crucial role."

[36] This lacks rigour, but it is in fact the limit of (1.131) as the grid becomes finer and $\delta_{\mathbf{k}} \to \delta(\mathbf{k})$. Following the advice of the previous footnote we could then simply set $V = 1$ and everything would look fine.

1.15.3 Density fluctuations on a grid

There are several approaches to this, aspects of which are discussed in Bertschinger (1992, §5.8, which we follow closely here).[37]

The standard way around this in cosmology is to consider a box of side L and volume V which is far larger than the length scale any of the phenomena of interest and divide it up onto a regular grid of N^3 cells. We then use the discrete Fourier transform representation of a density field on an $N \times N \times N$ grid.

Representing cosmological density fields and their Fourier counterparts on a lattice is important for computational purposes. When we sample the density field on a cubic lattice we shall denote the sample values by $\delta(\mathbf{x_j})$ to denote the value of the function $\delta(\mathbf{x})$ at the grid point identified by the grid label (j_1, j_2, j_3) in each of the three coordinate directions and whose coordinate position is $\mathbf{x_j}$.[38] We shall suppose that the cube is divided into a lattice of N^3 sub-cubes, so the grid spacing is $\Delta x = L/N$, which corresponds to a grid spacing $\Delta k = 2\pi/L$ in the Fourier $\mathbf{k}$-space. The normalisation V_L is then $V_L = N^3$ (compare with equation 1.72). The Fourier transform pair for a lattice spacing of $\Delta k = L/N$, in our conventions, is

$$\delta_{\mathbf{k}} = \frac{1}{N^3} \sum_{\mathbf{j}} e^{i\mathbf{k}\cdot\mathbf{x_j}}\ \delta(\mathbf{x_j}), \quad \delta(\mathbf{x_j}) = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x_j}}\ \delta_{\mathbf{k}} \tag{1.137}$$

The subscripts on the sums is understood to range over the lattice points in real space and in Fourier space.

The field $\delta(\mathbf{k})$ is the limit of the grid-sampled values $\delta_{\mathbf{k}}$ as the grid size is reduced to zero. From (1.132) this is

$$\delta(\mathbf{k}) = \lim_{\Delta k \to 0} \left(\frac{\Delta x}{2\pi}\right)^3 n^3 \delta_{\mathbf{k}} = \lim_{\Delta k \to 0} (\Delta k)^{-3} \delta_{\mathbf{k}} \tag{1.138}$$

and we see that in this formulation the $\delta_{\mathbf{k}}$ are dimensionless numbers. We now arrive at the relationship between the variance of the random process $\delta_{\mathbf{k}}$ and the power spectrum defined on a grid having spacing $\Delta k = 2\pi/L$:

$$\langle |\, \delta_{\mathbf{k}}\, |^2 \rangle \simeq \mathcal{P}(k)(\Delta k)^3 \tag{1.139}$$

with equality in the limit $\Delta k \to 0$. Here $\mathcal{P}(k)$ is the power spectral density and $\langle |\, \delta_{\mathbf{k}}\, |^2 \rangle$ is the power per Fourier mode. In other words, $\mathcal{P}(k)(\Delta k)^3$ is the contribution to the variance of $\delta(\mathbf{x})$ from waves in a volume $(\Delta k)^3$ at wave-vector $\mathbf{k}$.

[37] The excellent Bertschinger (1992) article is, unfortunately, not readily accessible. I have, accordingly, taken the liberty of adapting the contents of section §5.8 of that article for the present discussion.

[38] This notation may strike the reader as somewhat arcane, but that is what is used in practice. We could label the points simply by their $\mathbf{j} = (j_1, j_2, j_3)$ values, but we prefer to keep an explicit reference to the actual physical coordinates $\mathbf{x_j}$.

References

Adler, R.J. 1981. *The geometry of Random Fields*. John Wiley & Son.

Arfken, G.B., and Weber, H.J. 2005. *Mathematical Methods for Physicists*. 6th. edn. Elsevier Academic Press.

Batchelor, G.K. 1982. *The Theory of Homogeneous Turbulence*. Cambridge Science Classics. Cambridge University Press. First published 1953.

Bertschinger, E. 1992. Large-Scale Structures and Motions: Linear Theory and Statistics. Page 65 of: Martinez, V. J., Portilla, M., and Saez, D. (eds), *New Insights into the Universe*. Lecture Notes in Physics, Berlin Springer Verlag, vol. 408.

Bracewell, R.N. 1991. *The Fourier Transform & Its Applications*. 3rd edition edn. McGraw-Hill Higher Education.

Brigham, E.O. 1988. *Fast Fourier Transform and Its Applications*. Prentice Hall.

Carleson, Lennart. 1966. On convergence and growth of partial sums of Fourier series. *Acta Math.*, **116**, 135–157.

Coles, P., and Lucchin, F. 2002. *Cosmology: The Origin and Evolution of Cosmic Structure*. Wiley-Blackwell.

Cooley, J.W., and Tukey, J.W. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Computation*, **19**, 297–301.

Cramér, H. 1946. *Mathematical methods of statistics*. Princeton Univ. Press.

Danielson, G.C., and Lanczos, C. 1942. Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Institute*, **233**, 365–380, 435–452.

Fourier, J. 1822. *Theorie Analytique de la Chaleur*. Firman Didot.

Frigo, M., and Johnson, S.G. 2005. The Design and Implementation of FFTW3. *Proceedings of the IEEE*, **93**(2), 216–231. Special issue on "Program Generation, Optimization, and Platform Adaptation".

Gauss, C.F. 1866. Nachlass; Theoria Interpolationis methodo nova tractata. Pages 265–303 of: *Carl Fredrich Gauss, Werke, Band 3*. Königlichen Gesellschaft der Wissenschaften.

Gentleman, W.M., and Sande, G. 1966. Fast Fourier Transform for Fun and Profit. Pages 563–678 of: *AFIPS Proc., 1966 Fall Joint Computer Conf. Washington, DC: Spartan, 1966*. AFIPS Proceedings, vol. 29.

Goertzel, G. 1958. An algorithm for the evaluation of nite trigonometric series. *Am. Math. Monthly*, **65**, 34–35.

Goldstine, H.H. 1977. *A History of Numerical Analysis from the 16th through the 19th century*. Springer-Verlag.

Golub, G. H., and van Loan, C. F. 1996. *Matrix Computations*. 3rd edn. Johns Hopkins University Press.

Good, I.J. 1958. The interaction algorithm and practical Fourier analysis. *J. Roy. Stat. Soc*, **20**, 361–372.

Heideman, M.T., Johnson, D.H., and Burrus, C.S. 1985. Gauss and the history of the Fast Fourier Transform. *Archive for History of Exact Sciences*, **34**, 265–267.

Jackson, J.D. 1998. *Classical Electrodynamics*. 3rd. edn. John Wiley & Sons.

Kolb, E. W., and Turner, M. S. 1990. *The early universe.*

Martinez, V.J., and Saar, E. 2001. *Statistics of the Galaxy Distribution*. Chapman and Hall/CRC.

Matarrese, S., Verde, L., and Heavens, A. F. 1997. Large-scale bias in the Universe: bispectrum method. MNRAS, **290**(Oct.), 651–662.

Papoulis, A., and Pillai, S.U. 2002. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Higher Education. 4th. edition.

Pierrehumbert, R.T. 2004. Warming the world. *Nature*, **432**, 677–677.

Press, W.H., Teukolsky, S., Vetterling, W.T., and Flannery, B.P. 2007. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press; 3rd. edition.

Rosenfeld, A., and Kak, A.C. 1982. *Digital Picture Processing, Volume 1*. 2nd edn. Academic Press.

van Loan, C.F. 1987. *Computational Frameworks for the Fast Fourier Transform*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics.

Wiener, N. 1930. Generalized harmonic analysis. *Acta Mathematica*, **55**, 117–258.

Yaglom, A.N. 1952. Introduction to the theory of stationary random functions. *Uspekhi Mat. Nauk*, **7**, 3–168.

Yaglom, A.N. 1974. *Introduction to the Theory of Stationary Random Functions* . Dover Publications Inc.