Linear Algebra for Theoretical Neuroscience (Part 4)

Ken Miller

Current versions of all parts of this work can be found at http://www.neurotheory.columbia.edu/~ken/math-notes. Please feel free to link to this site.

6 The Fourier Transform

6.1 Introductory Remarks: Motivations for Studying the Fourier Transform

You've probably encountered the Fourier transform somewhere before. The basic idea, you will recall, is that any arbitrary function can be decomposed into a weighted sum of sines and cosines. Why is this interesting? Let me outline the basic ideas here. You're not expected to understand every detail here, but only to get the flavor. We'll work things out in detail in the rest of this chapter.

The main idea is that one can think of expressing a function in terms of sines and cosines as expressing that function in a different basis: just as we write $\mathbf{f} = \sum_j f_j \mathbf{e}_j$ to express a vector in a given basis $\mathbf{e}_i$, so we write something like $f(t) = \sum_j \left[f_j^c \cos_j(t) + f_j^s \sin_j(t) \right]$ to express the function $f(t)$ as a weighted sum over some set of cosine and sin functions (I've left vague, for a moment, exactly what we mean by the j^{th} cosine or sin in this sum, but roughly you can think of j as representing frequency). Why is a particular basis useful? We've found that a useful basis is one that diagonalizes a matrix that is important in the problem we're studying. That turns out to be why the Fourier transform is important – it's a change of basis that diagonalizes a whole class of matrices (or more generally, *linear operators*, which are to functions what matrices are to vectors – more on this later[1]) that come up extremely frequently in neurobiological and other scientific problems.

In particular, the Fourier transform is going to allow us to solve our two example problems in certain simple, important cases:

- **Development in a set of synapses:** We considered the equation

$$\tau \frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w} \tag{6.1}$$

  Consider a one-dimensional array of input neurons. Suppose that the correlations between two inputs only depend on the separation between them: $C_{ij} = c(i - j)$. This is plausible, *e.g.* in the retina, if we express distance in terms of retinal ganglion cell spacings (which get bigger in terms of degrees of visual space with increasing eccentricity), correlations between the spontaneous activities of two retinal ganglion cells in the dark fall off roughly as a function of the distance between the two cells. Then, as we'll see, we can solve our equation with the

[1] So, you couldn't wait? OK, very briefly: a matrix maps a vector to a vector; "linear operator" is the name for the equivalent operation that maps functions to functions. Examples of linear operators are the derivative operator: $\frac{d}{dt}$ operates on $f(t)$ to give a new function, $\frac{df}{dt}$; or an integral operator: $g(t) = \int k(t-t')f(t')$ represents the operation of "convolution with $k(t)$" acting on $f(t)$ to give a new function $g(t)$. As we'll see, if you discretize – discretely sample or bin the t axis – then functions become vectors and linear operators become matrices.

Fourier transform – this will turn out to be the transformation to the basis that diagonalizes $\mathbf{C}$. Thus, the Fourier transform will allow us to understand, in this case, how the structure of the matrix $\mathbf{C}$ determines the structure of the principal eigenvectors and thus of the receptive fields that develop.

The Fourier transform will also solve this problem in the case of a two-dimensional (or three-dimensional) array of input neurons, provided the correlation between two inputs depends only on the separation between them.

- **Activity in a network of neurons:** We considered the equation

$$\tau \frac{d\mathbf{b}}{dt} = -(\mathbf{1} - \mathbf{B})\mathbf{b} + \mathbf{h} \tag{6.2}$$

Again, consider a one-dimensional network of neurons. Suppose connections are just dependent on the separation between two neurons – each neuron excites those near it, say – so that $B_{ij} = b(i - j)$ for some function b. This is a lot less plausible for connectivity than for correlations, but it might make sense if we imagine what we are calling a "neuron" is really a set or assembly of neurons. At any rate, if connectivity is separation-dependent, then again, $\mathbf{B}$ will be diagonalized by the Fourier transform, and thus the Fourier transform will solve our problem. Again, this will also work in two or more dimensions, provided connectivity between two neurons depends only on the separation between them.

More generally, the Fourier transform will allow us to solve the large class of problems represented by *convolutions*:

Definition 6.1 *The* **convolution** *of a function $f(t)$ with a function $g(t)$ is defined by*

$$f \circ g(t) = \int dt' f(t - t')g(t') \tag{6.3}$$

By letting $p = t - t'$, one can show that equivalently this is $f \circ g(t) = \int dp\, g(t - p)f(p) = \int dt' g(t - t')f(t')$, that is, the convolution is symmetric in g and f. (These formulae work in arbitrary dimensions, that is, t and t' can be one-dimensional or many-dimensional (though both must have the same dimensions), so long as you interpret $\int dt'$ to be an integral over all of the dimensions of t'.)

We can also think of the discrete version of a convolution, and the Fourier transform will also solve those. Suppose we only sample the t axis discretely, say at evenly spaced points t_i with spacing Δt. Then the equation for a convolution becomes

$$f \circ g(t_i) = \sum_j f(t_i - t_j)g(t_j)\Delta t \tag{6.4}$$

(Note that as $\Delta t \to 0$, this becomes Eq. 6.3). Think of the value of a function at t_i as the i^{th} component of a vector, *e.g.* $g(t_j) \to g_j$, and similarly think of $f(t_i, t_j)$ as the (i, j) element of a matrix, $f(t_i, t_j)\Delta t \to F_{ij}$ (where we've incorporated the constant Δt into our definition of the matrix). Then the equation for this discrete convolution becomes

$$(\mathbf{f} \circ \mathbf{g})_i = \sum_j F_{ij} g_j \tag{6.5}$$

where the value of F_{ij} only depends on the separation of its components: $F_{ij} = f(t_i - t_j)\Delta t$, which in turn only depends on $i - j$. But this is just the equation we have discussed in our two examples above: when the matrix $\mathbf{C}$ (first example) or $\mathbf{B}$ (second example) only depends on the separation of its components, then the equations that arise in our simple examples involve a discrete convolution. Thus, convolutions, considered more generally to include the discrete as well as the continuous form, include the two examples we discussed above; the Fourier transform will solve all such convolutions.

Convolutions arise in many cases. We discussed two examples above, here are three more:

- You have an image $I(x)$. You want to smooth it by applying a Gaussian filter to each point: replace each intensity value $I(x)$ by the Gaussian-weighted average of the intensities around x. Letting $G(x)$ be the Gaussian function you smooth with, the smoothed image is $I \circ G(x) = \int dx' G(x - x')I(x')$.

- You model a cell as having a linear response $r(t)$ (representing the rate of the neuron's firing) to its stimulus $s(t)$. However, there is some temporal integration: the neuron's present response is some weighted average of the stimulus over the last 100-150 msec. Then the neuron's response is given by $r(t) = \int dt' L(t - t')s(t')$ where $L(t)$ tells the weighting of stimuli that occured t in the past.

  For example, an LGN cell can be reasonably approximated as having a response $r(t) = \int dx dt' K(x)L(t - t')s(x, t')$ where $s(x, t)$ is the luminance at point x at time t, and $K(x)$ describes the spatial center-surround structure of the receptive field. This is an independent temporal convolution for each pixel or spatial point x. A more accurate description is $r(t) = \int dx dt' K(x)L(x, t - t')s(x, t')$; here the temporal kernel or weighting function L can be different for each point x, to express the fact that different spatial points can take different temporal averages; in particular, the surround integrates more slowly than the center.

- Suppose each activation of a synapse at time t' leads to opening of a conductance with a time course $g(t - t')$. Let the activity of the presynaptic cell be given by $\rho(t)$ – very roughly, ρ is positive when the cell spikes and zero otherwise (we'll see how to define ρ more precisely later). Then the total conductance at time t is $\int dt' g(t - t')\rho(t')$

All of these examples and more will be greatly simplified by use of the Fourier transform.

The Fourier transform is also important for a practical, computational reason: there is a very fast algorithm, the Fast Fourier transform or FFT, that allows the transformation to the Fourier basis to be done much faster than by matrix multiplication. For a transformation in an N-dimensional vector space, the FFT requires on the order of $N \log N$ operations, whereas matrix multiplication requires on the order of N^2 operations. The Fourier transformation matrix has a fair amount of redundancy, which is exploited to achieve this fast algorithm. The computational speed of the FFT makes the Fourier transform even more computationally useful than it might be otherwise. We won't go into the FFT in these notes, but its existence is something you should be aware of (to use the FFT, you can apply any standard packaged routine, such as the ones in *Numerical Recipes* or the ones available in *Matlab*).

6.2 Introducing the Fourier Transform: The Fourier transform for a function on a finite domain

We all know what the Fourier transform means: any (reasonable) function can be expressed as a sum of sines and cosines. In this section, we develop this basic idea, showing you (or reminding you of) what the Fourier transform is and how it works in what is hopefully a familiar context. In subsequent sections, we'll show what this all has to do with vectors, matrices, and unitary transformations, but for this section we'll leave that aside.

6.2.1 The Fourier transform and its inverse

Let's explicitly write down the idea that any function can be expressed as a sum of sines and cosines. Consider a function $f(t)$ defined over a finite interval T, that is, defined for $-T/2 < t \leq T/2$ (we'll take away this restriction to finite intervals in later sections). Then the function can be expressed as a sum over all the sines and cosines that have an integral number of complete cycles in the interval T:

$$f(t) = \sum_{k=0}^{\infty} \left[f_k^c \cos\left(\frac{2\pi k t}{T}\right) + f_k^s \sin\left(\frac{2\pi k t}{T}\right) \right] \tag{6.6}$$

$$= f_0^c + \sum_{k=1}^{\infty} \left[f_k^c \cos\left(\frac{2\pi k t}{T}\right) + f_k^s \sin\left(\frac{2\pi k t}{T}\right) \right] \tag{6.7}$$

(To go from the first line to the second, we have noted that $\cos(0) = 1$ and $\sin(0) = 0$). The terms for a given k in Eq. 6.7 are a cosine and a sin that each have exactly k complete cycles in the interval T.[2] Thus, these terms have wavelength T/k: as t increases by T/k, the argument of the corresponding cos and sin increase by 2π, representing one complete cycle.

It is helpful (and a good way to avoid mistakes or missing factors) to keep track of the units. Let the units of any quantity x be given by $[x]$, so in particular t has units $[t]$, *e.g.* time. Since the arguments of the sin and cosine must be dimensionless, and $2\pi t/T$ is dimensionless, then k is a dimensionless number. The frequency is given by k/T, which has units $1/[t]$; *e.g.* if t is time and $T = 10$ sec, then the k^{th} sin or cosine has temporal frequency $0.1k$ Hz (where 1 Hz=1/sec=1 cycle/sec). Since sin and cos are dimensionless, f_k^c and f_k^s have the same units as $f(t)$: $[f(t)] = [f_k^c] = [f_k^s]$.

Perhaps you are accustomed to computing the power spectrum of a function in Matlab or other software, that is, determining the power at each frequency in the function. The power at frequency k/T is proportional to $|f_k^c|^2 + |f_k^s|^2$ (see further discussion below); so f_k^c and f_k^s translate fairly directly into quantities that you may be used to measuring.

Note that this Fourier expansion assumes the function is periodic: because every element of the right side of Eq. 6.7 is periodic with period T, so too $f(t)$ as defined by this expansion is periodic, that is, $f(t+T) = f(t)$. Nonetheless one can use the Fourier expansion to represent arbitrary functions on a finite interval. If the underlying function is continuous but not periodic (*i.e.*, $\lim_{t\to -T/2} f(t) \neq f(T/2)$), the Fourier reconstruction will show a discontinuity at this point – in the limit of an infinite number of terms in the expansion, it will get $f(t)$ right for $-T/2 < t < T/2$, while the reconstructed $f(T/2)$ will split the difference between the actual $f(T/2)$ and $\lim_{t\to -T/2} f(t)$ (this is in general how the Fourier reconstruction treats a discontinuity in the reconstructed function).

In real life, we must always deal with finite samples of a function (*e.g.* a neuron's voltage or spike sequence as a function of time, sampled over some finite time), and in most cases the function is not periodic over this finite length (if at all). The "imaginary discontinuity" – a very high-frequency change – imposed on a finitely-sampled non-periodic function by the Fourier reconstruction can interfere with estimates of the function's true underlying frequency components. There are various methods for dealing with this real-life situation, typically involving various combinations of "windowing" the function in clever ways (taking a finite snippet of a function means multiplying the

[2] Recall that cos and sin are 2π-periodic, that is, they go through a complete cycle every time their argument progresses through 2π, and so go through k complete cycles if their argument progresses through $2\pi k$ for integer k: $\cos(\theta + 2\pi k) = \cos(\theta)$, $\sin(\theta + 2\pi k) = \sin(\theta)$, for any integer k. So by restricting k to integers and writing the argument as $\frac{2\pi k t}{T}$, we ensure that there are an integral (k) number of complete cycles in T: $\cos\left(\frac{2\pi k(t+T)}{T}\right) = \cos\left(\frac{2\pi k t}{T} + 2\pi k\right) = \cos\left(\frac{2\pi k t}{T}\right)$.

true function by a window function that is 1 over the sampled region, 0 outside; instead one can use window functions that taper more gradually to zero) and "zerofilling" – embedding the finite sample in a larger interval, with the values of the function in the expanded region set to zero. For a good introduction to these issues and the various methods of dealing with real-world problems, see the book **Numerical Recipes**.

We can reexpress the above Fourier expansion in terms of complex exponentials; this is the more standard form that one works with, because it is computationally so much more convenient. Noting that $\cos\theta = \frac{1}{2}\left(e^{\imath\theta} + e^{-\imath\theta}\right)$ and $\sin\theta = \frac{1}{2\imath}\left(e^{\imath\theta} - e^{-\imath\theta}\right) = \frac{-\imath}{2}\left(e^{\imath\theta} - e^{-\imath\theta}\right)$, we rewrite Eq. 6.7 as:

$$\begin{aligned} f(t) &= f_0^c + \sum_{k=1}^{\infty}\left[\frac{f_k^c}{2}\left(e^{\imath\frac{2\pi kt}{T}} + e^{-\imath\frac{2\pi kt}{T}}\right) - \frac{\imath f_k^s}{2}\left(e^{\imath\frac{2\pi kt}{T}} - e^{-\imath\frac{2\pi kt}{T}}\right)\right] && (6.8)\\ &= f_0^c + \sum_{k=1}^{\infty}\left[\frac{1}{2}(f_k^c - \imath f_k^s)e^{\imath\frac{2\pi kt}{T}} + \frac{1}{2}(f_k^c + \imath f_k^s)e^{-\imath\frac{2\pi kt}{T}}\right] && (6.9)\\ &= \frac{1}{r_1}\sum_{k=-\infty}^{\infty} f_k e^{\imath\frac{2\pi kt}{T}} && (6.10)\end{aligned}$$

where

$$f_k = \begin{cases} \frac{r_1}{2}(f_k^c - \imath f_k^s) & k \geq 1 \\ r_1 f_0^c & k = 0 \\ \frac{r_1}{2}(f_{-k}^c + \imath f_{-k}^s) & k \leq -1 \end{cases} \qquad (6.11)$$

and r_1 is an arbitrary normalizing constant that we include for later convenience. We again note units: from Eq. 6.10, $[f_k]/[r_1] = [f(t)]$. We can reverse Eq. 6.11 to find

$$\begin{aligned} f_k^c &= \begin{cases} (f_k + f_{-k})/r_1 & k \neq 0 \\ f_k/r_1 & k = 0 \end{cases} && (6.12)\\ f_k^s &= \imath(f_k - f_{-k})/r_1 && (6.13)\end{aligned}$$

Eq. 6.10 is the usual form for writing the expansion of a function $f(t)$ defined on an interval T as a sum of sines and cosines. While it is equivalent to Eq. 6.7, it is computationally far simpler to work with.

Equations 6.10 or 6.7 are actually the *inverse* Fourier transform: they tell, given the coefficients f_k, how to reconstitute the function $f(t)$. The Fourier transform tells, given the function $f(t)$, how to determine the coefficients f_k – the amplitudes of the sines and cosines that add together to give $f(t)$. The Fourier transform is given as follows:

$$f_k = \frac{1}{r_2}\int_{-T/2}^{T/2} dt\ f(t)e^{-\imath\frac{2\pi kt}{T}} \qquad (6.14)$$

where r_2 is another to-be-determined constant. Note that units are given by $[f_k] = [f(t)][t]/[r_2]$; combining this with $[f_k] = [r_1][f(t)]$, we find that $[r_1][r_2] = [t]$.

We can verify that Eq. 6.14 is the correct expression, and determine what r_1 and r_2 must be, by substituting Eq. 6.10 into Eq. 6.14:

$$f_k = \frac{1}{r_2}\int_{-T/2}^{T/2} dt\ f(t)e^{-\imath\frac{2\pi kt}{T}} \tag{6.15}$$

$$= \frac{1}{r_2}\int_{-T/2}^{T/2} dt\ \left[\frac{1}{r_1}\sum_{l=-\infty}^{\infty} f_l e^{\imath\frac{2\pi lt}{T}}\right] e^{-\imath\frac{2\pi kt}{T}} \tag{6.16}$$

$$= \frac{1}{r_1 r_2}\sum_{l=-\infty}^{\infty} f_l \int_{-T/2}^{T/2} dt\ e^{\imath\frac{2\pi t(l-k)}{T}} \tag{6.17}$$

$$= \frac{1}{r_1 r_2}\sum_{l=-\infty}^{\infty} f_l T\delta_{lk} \tag{6.18}$$

$$= \frac{T}{r_1 r_2} f_k. \tag{6.19}$$

Here we have used the result

$$\int_{-T/2}^{T/2} dt\ e^{\imath\frac{2\pi t(l-k)}{T}} = T\delta_{lk} \ \text{ for } l \text{ and } k \text{ integers} \tag{6.20}$$

which is easily proven by doing the integral (explicitly done in the Appendix, Section 6.9.2). Geometrically, this result can be understood as follows: the integral takes $e^{\imath\frac{2\pi t(l-k)}{T}}$ through $l-k$ complete cycles. Refer back to Fig. 4.1, and think of the integral as a sum: the integral is summing up the various $e^{\imath\frac{2\pi t(l-k)}{T}}$ vectors, weighted by dt, as they circle round the complex plane $l-k$ times. If $l-k$ is nonzero, all these different vectors have to cancel out: for any given vector in the sum, there is an equal and opposite vector that has an equal weight. Thus, the only way the integral can be nonzero is if $l=k$, in which case $e^{\imath\frac{2\pi t(l-k)}{T}} = 1$ and $\int_{-T/2}^{T/2} dt\ 1 = T$.

Equations 6.15-6.19 give $f_k = \frac{T}{r_1 r_2} f_k$. Thus, the requirement on r_1 and r_2 is that $r_1 r_2 = T$ (which satisfies our previous finding on units: $[r_1][r_2] = [t]$). Any choice of r_1 and r_2 satisfying this will do. It turns out to be convenient[3] to choose $r_1 = r_2 = \sqrt{T}$, so let's adopt that. Note that this gives units $[f_k] = [t]^{1/2}[f(t)]$.

With this choice, we can now summarize:

Definition 6.2 *The* **Fourier transform** *of a function $f(t)$ defined on a finite interval $-T/2 < t \le T/2$ is given by*

$$f_k = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} dt\ f(t)e^{-\imath\frac{2\pi kt}{T}} \tag{6.21}$$

The **inverse Fourier transform** *in this case is given by*

$$f(t) = \frac{1}{\sqrt{T}}\sum_{k=-\infty}^{\infty} f_k e^{\imath\frac{2\pi kt}{T}} \tag{6.22}$$

The coefficients of the sin and cosine functions in the expansion of $f(t)$ in terms of sin's and cosines

[3]It will make the transform to the Fourier basis unitary, that is, from an orthonormal basis to an orthonormal basis; otherwise the Fourier basis vectors will be mutually orthogonal but not correctly normalized. This will make more sense later on.

(Eq. 6.7) are then given by

$$f_k^c = \begin{cases} (f_k + f_{-k})/\sqrt{T} & k \neq 0 \\ f_k/\sqrt{T} & k = 0 \end{cases} \tag{6.23}$$

$$f_k^s = \imath(f_k - f_{-k})/\sqrt{T} \tag{6.24}$$

6.2.2 More About the Fourier transform, and power spectra

Let's consider two other aspects of the Fourier transform. First, there are some simple correspondences between the structure of a function $f(t)$ and the structure of its Fourier transform.

Problem 6.1 *Show that the following is true:*

Fact 6.1 *$f(t)$ is real, or equivalently $f(t) = f(t)^*$, if and only if its Fourier transform coefficients satisfy $f_k = f_{-k}^*$.*

To show the "if", use Eq. 6.22, compute $f(t)^*$, and use $f_k = f_{-k}^*$ to show that $f(t)^* = f(t)$. (Hint: if $g(k)$ is some function of k, $\sum_{k=-\infty}^{\infty} g(k) = \sum_{k=-\infty}^{\infty} g(-k)$ – you're just summing the same terms in a different order.)

To show the "only if", do the reverse: use Eq. 6.22 to write $f(t)^*$, assume this is equal to $f(t)$, and see what this implies for the f_k. You will need to use the fact that if $\sum_{k=-\infty}^{\infty} a_k e^{\imath\frac{2\pi kt}{T}} = \sum_{k=-\infty}^{\infty} b_k e^{\imath\frac{2\pi kt}{T}}$, then $a_k = b_k$ for all k.[4]

Fact 6.1 also can be seen from Eq. 6.7: if $f(t)$ is real, then all the f_k^c and f_k^s must also be real (since they multiply real sin and cos functions in Eq. 6.7). From Eqs. 6.23-6.24, this in turn is true if and only if $(f_k + f_{-k})$ is real while $(f_k - f_{-k})$ is purely imaginary, which is enough to prove that $f_k = f_{-k}^*$.[5] Note that $f_k = f_{-k}^*$ implies that $f_0 = f_0^*$, that is, f_0 is real for a real function $f(t)$; this should be no surprise, since f_0 is just the "DC component" or integral of the function: $f_0 = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} f(t)dt$.

Another correspondence between the structure of a function and that of its Fourier transform comes from considering whether a function is even or odd (or neither). An even function $f(t)$ is defined as one for which $f(t) = f(-t)$; an odd function has $f(t) = -f(-t)$. The cosine is an even function, the sin is an odd function. If $f(t)$ is even, then only the even (cosine) terms can contribute to the expansion in Eq. 6.7 – the coefficients f_k^s of the odd (sin) terms must all be zero.[6] Similarly, if $f(t)$ is odd, then only the odd (sin) terms can contribute, and all of the cosine coefficients f_k^c must be zero. Expressing this in terms of the f_k using Eqs. 6.23-6.24, we have:

Fact 6.2 *If $f(t)$ is even, meaning $f(t) = f(-t)$, then its Fourier transform is also even: $f_k = f_{-k}$. If $f(t)$ is odd, meaning $f(t) = -f(-t)$, then its Fourier transform is also odd: $f_k = -f_{-k}$.*

[4]To see this, apply (1/T)$\int_{-T/2}^{T/2} dt e^{-i\frac{2\pi mt}{T}}$ to each sum; using Eq. 6.20, this will pick out the m^{th} coefficient. Thus, if the sums are equal, each coefficient is equal.

[5]Write $f_k + f_{-k} = 2r_a$ and $f_k - f_{-k} = 2\imath r_b$ where r_a and r_b are real. Adding and subtracting these equations yields $f_k = r_a + ir_b$, $f_{-k} = r_a - ir_b$.

[6]Proof: a sum of two nonzero even functions is even (show this); a sum of two nonzero odd functions is odd; a sum of a nonzero even and a nonzero odd function is neither even nor odd. So the sum of the sin terms is an odd function, the sum of the cosine terms is an even function, and you can't build an even function by adding any nonzero sin terms to the cosine terms.

Second, let's gain more clarity about the relationship between the f_k and the power at a given frequency. We've already stated that the power at a frequency k/T is proportional to $|f_k^c|^2 + |f_k^s|^2$. Using Eqs. 6.23-6.24, this in turn is proportional to $(|f_k|^2 + |f_{-k}|^2)$ (or just $|f_k|^2$ for $k = 0$). So, for $k \neq 0$ one must count both the positive and negative frequency components to determine the power. For a real function, $|f_k|^2 = |f_{-k}|^2$, so for $k \neq 0$ the power at frequency k/T is proportional to twice the power at the positive-frequency component. Thus, if one only looks only at non-negative frequencies for a real function, the power at frequencies $k/T > 0$ is proportional to $2|f_k|^2$, while the power at $k = 0$ is proportional to $|f_0|^2$ (no factor of 2).

To gain more insight into why power sums across frequencies, let's define the *power* in $f(t)$ to be $P_f = \frac{1}{T}\int_{-T/2}^{T/2} |f(t)|^2 dt$. This is defined by analogy to many situations in physics, where the energy in a wave is proportional to the square of its amplitude, and the power is the average energy delivered per unit time.

Problem 6.2 *Using Eq. 6.22, show that*

$$|f(t)|^2 = \frac{1}{T}\sum_{k,l=-\infty}^{\infty} f_k^* f_l e^{\imath\frac{2\pi(l-k)t}{T}} \tag{6.25}$$

Then, using Eq. 6.20, show that

$$P_f = \frac{1}{T}\int_{-T/2}^{T/2} |f(t)|^2 dt = \frac{1}{T}\sum_{k=-\infty}^{\infty} |f_k|^2 \tag{6.26}$$

This result — that the sum or integral of the absolute square of a function in real space is equal to the sum or integral of the absolute square in Fourier space — is known as Parseval's Theorem*.*

Thus, defining the power as the average of the absolute square of a function, we find that this power is just $1/T$ time the sum of the absolute squares of the Fourier components. The power in each frequency component adds independently to give the total power. This relationship is what makes it natural to represent the power at a signed frequency k as being proportional to $|f_k|^2$ (or at an unsigned frequency $k \neq 0$ as being proportional to $|f_k|^2 + |f_{-k}^2|$).

We will see later that Parseval's Theorem is a natural consequence of the fact that the Fourier transform can be regarded as a unitary coordinate transformation – the sum of the absolute square of the components of a vector is the length of the vector, and this is preserved under unitary transformations. It is probably not yet clear to you what I am talking about – what integrals of functions have to do with lengths of vectors, or what unitary transformations of vectors have to do with Fourier transformations of functions – but it will become clear.

6.2.3 The convolution theorem

The convolution theorem will show that going to the Fourier domain greatly simplifies a convolution – the convolution becomes a simple, frequency-by-frequency multiplication. To see this, we'll consider the convolution $f \circ g(t)$ of two T-periodic functions, $f(t)$ and $g(t)$:

$$f \circ g(t) = \int_{-T/2}^{T/2} dt'\ f(t - t')g(t') \quad \text{for } -T/2 < t \leq T/2 \tag{6.27}$$

Note that the argument $t - t'$ of $f(t - t')$ can range from $-T$ to T, so for the convolution to make sense for functions defined on the finite range $-T/2$ to $T/2$, we have to extend the functions by considering them to be periodic – as is implied by their expression as a Fourier series.

To simplify notation, let's just call the convolution $c(t)$: $c(t) = f \circ g(t)$. So, let's take its Fourier transform:

$$\begin{aligned} c_k &= \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dt\ c(t) e^{-\imath \frac{2\pi k t}{T}} && (6.28) \\ &= \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dt\ \left[\int_{-T/2}^{T/2} dt'\ f(t-t') g(t') \right] e^{-\imath \frac{2\pi k t}{T}} && (6.29) \\ &= \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dt\ \int_{-T/2}^{T/2} dt'\ f(t-t') g(t') e^{-\imath \frac{2\pi k (t-t')}{T}} e^{-\imath \frac{2\pi k t'}{T}} && (6.30) \end{aligned}$$

We let $p = t - t'$, and continue:

$$c_k = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dt'\ g(t') e^{-\imath \frac{2\pi k t'}{T}} \int_{-T/2+t'}^{T/2+t'} dp\ f(p) e^{-\imath \frac{2\pi k p}{T}} \quad (6.31)$$

Now, consider any periodic function $h(p+T) = h(p)$. Because it is periodic, the integral of h over a segment of length T is the same for any such segment – it doesn't matter where the segment is centered. In particular,[7] $\int_{-T/2+t'}^{T/2+t'} dp\ h(p) = \int_{-T/2}^{T/2} dp\ h(p)$. So since both $f(p)$ and the complex exponential (and hence their product) are T-periodic, we can continue:

$$\begin{aligned} c_k &= \int_{-T/2}^{T/2} dt'\ g(t') e^{-\imath \frac{2\pi k t'}{T}} \left[\frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dp\ f(p) e^{-\imath \frac{2\pi k p}{T}} \right] && (6.34) \\ &= f_k \int_{-T/2}^{T/2} dt'\ g(t') e^{-\imath \frac{2\pi k t'}{T}} && (6.35) \\ &= \sqrt{T} f_k g_k && (6.36) \end{aligned}$$

This is the convolution theorem:

Theorem 6.1 Convolution theorem: *The Fourier transform $(f \circ g)_k$ of the convolution $f \circ g(t)$ of two functions $f(t)$ and $g(t)$ is given by:*

$$(f \circ g)_k = \sqrt{T} f_k g_k \quad (6.37)$$

Note that the factor $\sqrt{T}$ depends on our choice of the normalization of the Fourier transform and its inverse; for example, had we chosen $r_1 = T$, $r_2 = 1$, it would disappear. Thus the main point is that the Fourier transform of the convolution is just (up to a normalization factor) the product of the Fourier transforms of the two functions; "convolution in real space becomes multiplication in Fourier space".

---

[7] This can be seen formally as follows:

$$\int_{T/2}^{T/2+t'} dp\ h(p) = \int_{T/2-T}^{T/2+t'-T} dp\ h(p) = \int_{-T/2}^{-T/2+t'} dp\ h(p) \quad (6.32)$$

and therefore

$$\int_{-T/2+t'}^{T/2+t'} dp\ h(p) = \int_{-T/2+t'}^{T/2} dp\ h(p) + \int_{T/2}^{T/2+t'} dp\ h(p) = \int_{-T/2}^{-T/2+t'} dp\ h(p) + \int_{-T/2+t'}^{T/2} dp\ h(p) = \int_{-T/2}^{T/2} dp\ h(p). \quad (6.33)$$

Problem 6.3 *1. By exactly the same method, prove the correlation theorem, which gives the Fourier transform of the cross-correlation between two functions (think spike trains):*

***Theorem 6.2* Correlation theorem:** *Define the cross-correlation $C_{f,g}(t)$ of two T-periodic functions $f(t)$, $g(t)$ as*

$$C_{f,g}(t) = \int_{-T/2}^{T/2} dt' \; f(t+t')g(t') \quad \text{for } -T/2 < t \leq T/2 \tag{6.38}$$

Then its Fourier transform is

$$(C_{f,g})_k = \sqrt{T} f_k g_{-k} \tag{6.39}$$

2. Now prove the same thing by (1) defining $h(t) = g(-t)$; (2) showing that this implies $h_k = g_{-k}$; (3) showing $C_{f,g}(t) = f \circ h(t)$; (4) applying the convolution theorem to $f \circ h$.

6.3 Why the Fourier transform is like a vector change of basis

What do we mean by an orthonormal basis for a vector space? It is a set of vectors $\mathbf{e}_i$ that satisfy the following properties:

- Orthonormality: $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$.
- Completeness: Any vector $\mathbf{v}$ can be written $\mathbf{v} = \sum_i v_i \mathbf{e}_i$ where $v_i = \mathbf{e}_i \cdot \mathbf{v}$.

Suppose we think of the values $f(t)$ of the function f for each point t as being like the components v_i of a vector $\mathbf{v}$ for each dimension i. For example, if we only sample $f(t)$ at some discrete set of points t_i separated by Δt — so there are $N = T/\Delta t$ points — then the function $f(t)$ becomes an N-dimensional vector with components $f_i = f(t_i)$. Indeed in the real (digital) world, we usually end up dealing only with discretely sampled versions of functions — *e.g.*, we sample continuous voltage traces with an A/D converter to produce a discrete string of voltages, one for each time-sample point; we represent a function on a computer as a discrete array of sample points when we perform calculations on it, such as computing its power spectrum. It makes sense to continue to think of the $f(t)$ for different t's as the "components" of the "vector" represented by $f(t)$, even when we take the limit $\Delta t \to 0$, $N \to \infty$ and go back to the case of a continuous function. Another way to think of this is that there is one component for each dimension or degree of freedom of a vector – each axis along which the vector can independently vary. In a sense, each point t provides an independent dimension for a function $f(t)$ – the function can assume a different value at each t – so we can think of it as a sort of vector with a continuously infinite set of dimensions indexed by t.[8]

To keep this correspondence of functions and vectors in mind, I'll adopt a slight change of notation: just as we write vectors $\mathbf{v}$ and their components v_i, so I will write functions $\mathbf{f}$ with components $f(t)$.

Then, just as we define the dot product for vectors as $\mathbf{v} \cdot \mathbf{q} = \sum_i v_i^* q_i$, so we can define the dot product between functions as:

[8] Of course, if your function space is restricted to be locally continuous, it is not quite true that the function can assume an independent value at each t; but if derivatives can be arbitrarily large, then the values in arbitrarily small intervals dt can vary independently ... In general, the more constraints you put on your function space (continuity, existence of derivatives, frequency cutoff, ...), the fewer dimensions it will have, and accordingly in some sense the lower the dimension of the set of basis functions required to span the space (*i.e.* to serve as a basis for constructing all the functions in the space). But this is a deep subject about which I know next to nothing so I'll stop there.

Definition 6.3 *The dot product of two T-periodic functions* $\mathbf{f}$ *and* $\mathbf{g}$ *is defined by*

$$\mathbf{f} \cdot \mathbf{g} = \int_{-T/2}^{T/2} dt\ f^*(t) g(t) \tag{6.40}$$

Then the fact that any T-periodic function can be represented by its Fourier expansion can be interpreted as follows. Take the T-periodic functions as our "vector space" (the careful mathematician must define the space of functions more carefully, *e.g.* putting restrictions like being continuous or finite-valued, to restrict the space to the functions the Fourier expansion can describe, and to ensure they will behave like a vector space, *e.g.* addition of two of them will stay within the space). Let's define:

Definition 6.4 *The* **Fourier basis functions** *for the set of T-periodic functions are the functions* $\mathbf{e}_k$ *with components* $e_k(t) = \frac{1}{\sqrt{T}} e^{\imath \frac{2\pi k t}{T}}$, *for* k *an integer,* $-\infty < k < \infty$.

These functions are orthonormal: using Eqs. 6.20 and 6.40,

$$\mathbf{e}_k \cdot \mathbf{e}_l = \frac{1}{T} \int_{-T/2}^{T/2} dt\ e^{\imath \frac{2\pi (l-k) t}{T}} = \delta_{lk}. \tag{6.41}$$

They are also complete: the Fourier transform says that any function $\mathbf{f}$ in our space can be written

$$f(t) = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{\infty} f_k e^{\imath \frac{2\pi k t}{T}} = \sum_{k=-\infty}^{\infty} f_k \frac{e^{\imath \frac{2\pi k t}{T}}}{\sqrt{T}} = \sum_k f_k e_k(t) \tag{6.42}$$

or $\mathbf{f} = \sum_k f_k \mathbf{e}_k$. Furthermore the components f_k can be found from the appropriate dot product:

$$f_k = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} dt\ f(t) e^{-\imath \frac{2\pi k t}{T}} = \int_{-T/2}^{T/2} dt\ \frac{e^{-\imath \frac{2\pi k t}{T}}}{\sqrt{T}} f(t) = \mathbf{e}_k \cdot \mathbf{f} \tag{6.43}$$

So, the Fourier basis looks just like an orthonormal basis for our function space, making the Fourier transform just a unitary transformation to this basis. Which raises the question: what basis were we transforming *from*?

In dealing with vectors, the "current basis" – a set of basis vectors described in their own coordinate system – is the set of vectors with one coordinate equal to 1 and all other coordinates equal to zero. As we've suggested, a vector's i^{th} coordinate, v_i, becomes a function's t^{th} value, $f(t)$. The function that has one "coordinate" – say its value at $t = t_0$ – nonzero, and all other coordinates zero, is the Dirac delta function $\delta(t - t_0)$. This function is described in detail in the Appendix, Section 6.9.1. Briefly, one can think of it as the limit, as $\Delta t \to 0$, of a function that is equal to $\frac{1}{\Delta t}$ on the region of width Δt centered at t_0, and 0 elsewhere. This function is real, it is zero for $t \neq t_0$, and it integrates to 1: $\int_{-T/2}^{T/2} dt\ \delta(t - t_0) = 1$ for $-T/2 < t < T/2$. Furthermore, just as, for any vector v and basis vector $\mathbf{e}_k$, $\mathbf{e}_k \cdot \mathbf{v} = v_k$ where v_k is the component of $\mathbf{v}$ in the $\mathbf{e}_k$ direction, so for any function $\mathbf{f}$, its integral with $\delta(t - t_0)$ picks out $f(t_0)$: $\int_{-T/2}^{T/2} dt\ \delta(t - t_0) f(t) = f(t_0)$ for $-T/2 < t < T/2$ (see Appendix for more details). Just as we think of the i^{th} basis vector $\mathbf{e}_i$ as having j^{th} component $(\mathbf{e}_i)_j$, so we can think of the function $\delta(t - t_0)$ as the t_0^{th} basis function $\mathbf{e}_{t_0}$, whose t^{th} component is $e_{t_0}(t) = \delta(t - t_0)$.

Thus, we assert that the basis in which our functions were originally defined — call it the "real-space basis", since it is traditional to refer to the original functions as living in "real space" and their Fourier transforms as living in "fourier space" — is the space of Dirac delta functions:

Definition 6.5 *The* **real-space basis functions** *for the set of T-periodic functions are the functions* $\mathbf{e}_{t_0}$ *defined by* $e_{t_0}(t) = \delta(t - t_0)$ *for* $-T/2 < t_0 < T/2$.

Note that these functions have a continuous index: there is a continuous infinity of possible values t_0, so there is a continuous infinity of real-space basis functions. In contrast, there is a discrete infinity of Fourier-space basis vectors – they are indexed by the discretely infinite set of all possible integers. Don't get too tangled up thinking about this.

Well, let's see if this definition makes sense. First, are they orthonormal?:

$$\mathbf{e}_{t_0} \cdot \mathbf{e}_{t_1} = \int_{-T/2}^{T/2} dt\ \delta^*(t - t_0)\delta(t - t_1) = \delta(t_0 - t_1) \tag{6.44}$$

Hmmmm ... This seems almost right – it's a delta function – but it's a different kind of delta function. Instead of the Kronecker delta, $\delta_{t_0 t_1}$, we have the Dirac delta, $\delta(t_0 - t_1)$. It turns out this is the right – or consistent – way to define what it means for a continuously-indexed set of functions to be orthonormal. Kronecker delta functions have two discrete indices, and make sense for dealing with sums and discrete spaces; the Dirac delta function has one continuous index, and makes sense for dealing with integrals and continuous spaces. One way to see this is that the Kronecker delta, δ_{kl}, gives the coordinates of the identity operator, $\mathbf{1}$, which is defined by the fact that $\mathbf{1}\mathbf{v} = \mathbf{v}$ or $\sum_j \mathbf{1}_{ij} v_j = v_i$ for any vector $\mathbf{v}$. The continuous version of the latter equation is $\int_{-T/2}^{T/2} dt\, \mathbf{1}(t - t')f(t') = f(t)$ for any function $\mathbf{f}$, which requires that $\mathbf{1}(t - t') = \delta(t - t')$. That is, the Dirac delta is the identity operator on functions, just as the Kronecker delta is the identity operator on vectors.

Second, are they complete? Yes, any function $\mathbf{f}$ can be expanded in them:

$$f(t) = \int_{-T/2}^{T/2} dt_0\ f(t)\delta(t - t_0) = \int_{-T/2}^{T/2} dt_0\ f(t_0)\delta(t - t_0) = \int_{-T/2}^{T/2} dt_0\ f(t_0)e_{t_0}(t) \tag{6.45}$$

or

$$\mathbf{f} = \int_{-T/2}^{T/2} dt_0\ f(t_0)\mathbf{e}_{t_0} \tag{6.46}$$

This is the continuous form of the discrete expression $\mathbf{v} = \sum_i v_i \mathbf{e}_i$; we have to integrate rather than sum over the continuous index t_0. (To perhaps belabor the point: the equivalent of the i^{th} component v_i is the t_0^{th} component $f(t_0)$, while the equivalent of the i^{th} basis vector $\mathbf{e}_i$ is the t_0^{th} basis function $\mathbf{e}_{t_0}$). Furthermore, the component $f(t_0)$ is found from the appropriate dot product:

$$f(t_0) = \mathbf{e}_{t_0} \cdot \mathbf{f} = \int_{-T/2}^{T/2} dt\ e_{t_0}^*(t)f(t) = \int_{-T/2}^{T/2} dt\ \delta(t - t_0)f(t) \tag{6.47}$$

Finally, the convolution theorem can be understood as saying that the Fourier basis diagonalizes the convolution operator – it is the basis of eigenfunctions of the convolution operator. We generalize the concept of a matrix $\mathbf{M}$ that takes a vector $\mathbf{v}$ to another vector $\mathbf{M}\mathbf{v}$, to the concept of a linear operator $\mathcal{O}$ that takes a function $\mathbf{f}$ to another function $\mathcal{O}\mathbf{f}$ (the "linear" part means $\mathcal{O}(a\mathbf{f} + b\mathbf{g}) = a\mathcal{O}\mathbf{f} + b\mathcal{O}\mathbf{g}$). For example, the convolution operator $\mathbf{g}\circ$ is linear and takes $\mathbf{f}$ to $\mathbf{g} \circ \mathbf{f}$ (defined in Eq. 6.27). The eigenvectors $\mathbf{e}_i$ of a matrix $\mathbf{M}$ satisfy $\mathbf{M}\mathbf{e}_i = \lambda_i \mathbf{e}_i$; similarly the eigenfunctions $\mathbf{e}_x$ of a linear operator $\mathcal{O}$ satisfy $\mathcal{O}\mathbf{e}_x = \lambda_x \mathbf{e}_x$.

Problem 6.4 *Show that the Fourier basis functions* $\mathbf{e}_k$ *are eigenfunctions of the convolution operator. That is, show that*

$$(\mathbf{g} \circ \mathbf{e}_k)(t) = \int_{-T/2}^{T/2} dt' \; g(t-t') \frac{e^{\imath \frac{2\pi k t'}{T}}}{\sqrt{T}} \tag{6.48}$$

$$= g_k e^{\imath \frac{2\pi k t}{T}} = \sqrt{T} g_k e_k(t) \tag{6.49}$$

or $\mathbf{g} \circ \mathbf{e}_k = \sqrt{T} g_k \mathbf{e}_k$. *(Hint: put the factor* $1 = e^{-\imath \frac{2\pi k t}{T}} e^{\imath \frac{2\pi k t}{T}}$ *into the integral, and combine the first exponential with the* $e^{\imath \frac{2\pi k t'}{T}}$ *term.) Thus the Fourier basis functions* $\mathbf{e}_k$ *are eigenfunctions of the convolution operator* $g\circ$*, with eigenvalue* λ_k *just given by* $\sqrt{T} g_k$*, or* $\sqrt{T}$ *times the* k^{th} *Fourier component of the function* g *defining the convolution.*

When we transform to the eigenvector basis $\mathbf{e}_i$ of a matrix $\mathbf{M}$, the matrix becomes diagonal, with diagonal entries given by its eigenvalues λ_i. In this basis, the operation of the matrix $\mathbf{M}$ on a vector $\mathbf{v}$ is just component-wise multiplication: if $\mathbf{v} = \sum_i v_i \mathbf{e}_i$, then $\mathbf{M}\mathbf{v} = \sum_i \lambda_i v_i \mathbf{e}_i$, or $(\mathbf{M}\mathbf{v})_i = \lambda_i \mathbf{v}_i$. Compare this to expressing $\mathbf{M}\mathbf{v}$ in any other coordinate system, say $\mathbf{e}_a$: then $(\mathbf{M}\mathbf{v})_a = \sum_b M_{ab} v_b$, that is, to compute the a^{th} component of $\mathbf{M}\mathbf{v}$, one must sum over *all* components of $\mathbf{v}$.

Precisely the same thing happens in the case of the convolution operator $\mathbf{g}\circ$: the Fourier basis $\mathbf{e}_k$ is its eigenfunction basis, with corresponding eigenvalues $\lambda_k = \sqrt{T} g_k$. In this basis, the operation of the convolution on a function $\mathbf{f}$ is just component-wise multiplication: if $\mathbf{f} = \sum_k f_k \mathbf{e}_k$, then $\mathbf{g} \circ \mathbf{f} = \sum_k \lambda_k f_k \mathbf{e}_k = \sqrt{T} \sum_k g_k f_k \mathbf{e}_k$ or $(\mathbf{g} \circ \mathbf{f})_k = \sqrt{T} g_k f_k$. Compare this to expressing $\mathbf{f}$ in some other coordinate system, say the real-space coordinate system: there $(\mathbf{g} \circ \mathbf{f})(t) = \int_{-T/2}^{T/2} dt' \; g(t-t') f(t')$, that is, to compute the t component of $(\mathbf{g} \circ \mathbf{f})$, one must sum over *all* components of $\mathbf{f}$.

In summary: the Fourier transform can be understood as a unitary transform of a function $\mathbf{f}$ from the real-space basis, where the function's components are $f(t_0)$, to the Fourier basis, where the function's components are f_k. We are describing the same function, but in a different coordinate system. The Fourier transform is useful because it transforms to the eigenvector basis of a wide class of operators, including convolutions and derivatives. One can generalize everything we've established so far – the properties of Hermitian and Unitary and Normal operators, that Normal operators have a complete basis of eigenvectors and Hermitian operators have real eigenvalues, etc. (there are some mathematical exceptions, in infinite-dimensional spaces, but none we ever have to worry about in real life). It all goes right through, and the intuitions we derived from N-dimensional spaces go all the way through to continuously-infinite-dimensional spaces. This should not be too surprising, since in the real digital world we always deal with functions as finite-dimensional vectors, and this seems to work just fine.

From here, one can go in two directions: one can follow the real world and discretely sample t, and so consider Fourier transforms of finite-dimensional vectors; or one can let $T \to \infty$, and consider Fourier transforms of functions on an infinite domain. For finite-dimensional vectors, the Fourier basis also becomes finite-dimensional. For functions on an infinite domain in real space, the Fourier coordinates also live on a continuous infinite domain – k goes continuously from $-\infty$ to ∞. So in those two cases Fourier space and real space look alike, unlike in the case we've considered. Otherwise, everything looks more or less identical to what we have found here. The convolution theorem, the relationships between the structure in Fourier coordinates vs. in real coordinates, etc., all go through unchanged except for appropriate changes between sums and integrals for discrete vs. continuous indexes and some possible changes in normalization factors.

We'll go through these two cases in a moment. But first, let's better understand why the Fourier transform is so useful.

6.4 Why the Fourier Transform is So Useful: Diagonalizing Translation-Invariant Operators

There is a very useful theorem about matrices and linear operators that goes like this:

Theorem 6.3 *Suppose two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ each have a complete basis of eigenvectors. Then they have a common basis of eigenvectors if and only if they commute: that is, if and only if $\mathbf{M}_1\mathbf{M}_2 = \mathbf{M}_2\mathbf{M}_1$. Similarly, two linear operators $\mathcal{O}_1$ and $\mathcal{O}_2$ with complete bases of eigenfunctions have a common basis of eigenfunctions if and only if they commute, $\mathcal{O}_1\mathcal{O}_2 = \mathcal{O}_2\mathcal{O}_1$.*

Sketch of a proof for matrices: If $\mathbf{M}_1$ and $\mathbf{M}_2$ have a common basis of eigenvectors, then $M_1 = \mathbf{U}\mathbf{D}_1\mathbf{U}^\dagger$ and $M_2 = \mathbf{U}\mathbf{D}_2\mathbf{U}^\dagger$ for diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ and the *same* unitary matrix $\mathbf{U}$.[9] Using the facts that diagonal matrices always commute with each other and that $\mathbf{U}^\dagger\mathbf{U} = \mathbf{1}$, we find $\mathbf{M}_1\mathbf{M}_2 = \mathbf{U}\mathbf{D}_1\mathbf{U}^\dagger\mathbf{U}\mathbf{D}_2\mathbf{U}^\dagger = \mathbf{U}\mathbf{D}_1\mathbf{D}_2\mathbf{U}^\dagger = \mathbf{U}\mathbf{D}_2\mathbf{D}_1\mathbf{U}^\dagger = \mathbf{U}\mathbf{D}_2\mathbf{U}^\dagger\mathbf{U}\mathbf{D}_1\mathbf{U}^\dagger = \mathbf{M}_2\mathbf{M}_1$. Conversely, if $\mathbf{M}_1\mathbf{M}_2 = \mathbf{M}_2\mathbf{M}_1$ and if $\mathbf{e}_i^1$ is a basis of eigenvectors of $\mathbf{M}_1$ with eigenvalues λ_i^1, then $\mathbf{M}_2\mathbf{M}_1\mathbf{e}_i^1 = \lambda_i^1\mathbf{M}_2\mathbf{e}_i^1$, but also $\mathbf{M}_2\mathbf{M}_1\mathbf{e}_i^1 = \mathbf{M}_1\mathbf{M}_2\mathbf{e}_i^1$, so $\mathbf{M}_1(\mathbf{M}_2\mathbf{e}_i^1) = \lambda_i^1(\mathbf{M}_2\mathbf{e}_i^1)$, that is, $\mathbf{M}_2\mathbf{e}_i^1$ is also an eigenvector of $\mathbf{M}_1$ with eigenvalue λ_i. If there is only one eigenvector of $\mathbf{M}_1$ with eigenvalue λ_i^1, then this implies that $\mathbf{M}_2\mathbf{e}_i^1 \propto \mathbf{e}_i^1$ and $\mathbf{e}_i^1$ is also an eigenvector of $\mathbf{M}_2$ (but most likely with a different eigenvalue). If there are multiple eigenvectors of $\mathbf{M}_1$ with eigenvalue λ_i, then the proof gets a little trickier – when there's multiple eigenvectors with the same eigenvalue, then any linear combination of them is also an eigenvector with the same eigenvalue, so all we know in that case is that $\mathbf{M}_2\mathbf{e}_i^1$ is such a linear combination. But the bottom line in that case is that one can always pick the right linear combinations as the basis so that $\mathbf{M}_1$ and $\mathbf{M}_2$ share the same eigenvectors.

Now here's an example of why this theorem is useful. Consider the operator $\mathcal{T}_a$ on functions that translates them by a: $\mathcal{T}_a f(t) = f(t+a)$.[10] If an operator commutes with the translation operator, it means that the operator has the same form no matter where in the function it is applied. Most linear filters you will encounter are translation-invariant (meaning that they commute with the translation operator), for example when you smooth an image you apply the same smoothing function everywhere on the image, you don't smooth by different amounts at different points on the image. So the translation-invariant operators are a very important class of linear operators. Some examples of translation-invariant operators include:

- The derivative operator $\frac{d}{dt}$: $\mathcal{T}_a \frac{d}{dt} f(t) = \mathcal{T}_a f'(t) = f'(t+a) = \frac{d}{dt} f(t+a) = \frac{d}{dt}\mathcal{T}_a f(t)$.
- The convolution operator $\mathbf{g}\cdot$: $\mathcal{T}_a(\mathbf{g}\cdot\mathbf{f})(t) = (\mathbf{g}\cdot\mathbf{f})(t+a) = \int_{-T/2}^{T/2} dt'\, g(t+a-t')f(t') = \int_{-T/2-a}^{T/2-a} dt'\, g(t-t')f(t'+a) = \int_{-T/2}^{T/2} dt'\, g(t-t')\mathcal{T}_a f(t') = (\mathbf{g}\cdot\mathcal{T}_a\mathbf{f})(t)$.

(To understand the elimination of the $-a$'s in the integral limits in the second-to-last step, see footnote 7.) The theorem tells us that, if we can find the eigenfunctions of the translation operator, we will be able to diagonalize any translation-invariant operator – including any derivative or any convolution – because they will share a common basis of eigenfunctions with the translation operator.

What are the eigenfunctions of the translation operator?

[9] We will later see that for a matrix $\mathbf{M}$ that has a complete but non-orthonormal basis of eigenvectors, the expression for $\mathbf{M}$ in terms of a diagonal matrix $\mathbf{D}$ is $\mathbf{M} = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$ where the columns of the invertible matrix $\mathbf{C}$ are the eigenvectors. The proof goes through equally well with this expression.

[10] I haven't explained what the adjoint of an operator, as opposed to a matrix, is, but the adjoint of $\mathcal{T}_a$ is $\mathcal{T}_{-a}$, and these two commute with one another, so $\mathcal{T}_a$ is a normal operator (in particular, since $\mathcal{T}_{-a}$ is the inverse of $\mathcal{T}_a$, $\mathcal{T}_a$ is unitary) and so has a complete orthonormal basis of eigenfunctions. Exercise 6.4 will show the matrix version of $\mathcal{T}_a$ in the discrete case (*i.e.*dealing with vectors rather than functions), and there these relationships will be clear.

Theorem 6.4 *The eigenfunctions of the translation operator are the Fourier basis functions,* $e_k(t) = \frac{1}{\sqrt{T}} e^{\imath \frac{2\pi k t}{T}}$.

The proof is simple: $\mathcal{T}_a e_k(t) = e_k(t+a) = \frac{1}{\sqrt{T}} e^{\imath \frac{2\pi k(t+a)}{T}} = e^{\imath \frac{2\pi k a}{T}} \frac{1}{\sqrt{T}} e^{\imath \frac{2\pi k t}{T}} = e^{\imath \frac{2\pi k a}{T}} e_k(t)$. That is, if we translate one of the Fourier basis functions by a, we get back the same Fourier basis function, multiplied by the complex number $e^{\imath \frac{2\pi k a}{T}}$ (this is the corresponding eigenvalue of the translation operator). Furthermore, for most a, the eigenvalues $e^{\imath \frac{2\pi k a}{T}}$ for different k's are distinct, meaning that this is the *only* eigenvector basis of the translation operator (eigenvectors are ambiguous only when two or more eigenvectors share the same eigenvalue).

Thus: *the reason the Fourier transform is so useful is that it diagonalizes all translation-invariant operators.* The Fourier transform is the transform to the basis of eigenvectors of the translation operator, and these are also eigenvectors of any translation-invariant operator. By transforming to the Fourier basis, in one fell swoop any such operator is diagonalized. This is the main reason why the Fourier transform is so powerful and so commonly used. (Another reason is that a very fast computational implementation, the fast Fourier transform, exists.)

6.5 The Discrete Fourier Tranform

Suppose we take a T-periodic function $f(t)$ and sample it discretely at N points separated by Δt, $N\Delta t = T$. Thus, in place of $f(t)$ defined on $-T/2 \leq t \leq T/2$, we consider the vector $\mathbf{v}$ with components $v_j = f(j\Delta t)$, $j = [-N/2]^+, [-N/2]^+ + 1, \ldots, [N/2]^- - 1, [N/2]^-$. Here, we define $[-N/2]^+$ to be the smallest integer greater than or equal to $-N/2$, and $[N/2]^-$ to be the largest integer that is strictly less than $N/2$ (*e.g.* if $N = 30$, the components go from -15 to 14, while if $N = 31$, the components go from -15 to 15). Letting $\tilde{v}_k$ be the k^{th} component of the Fourier transform of $\mathbf{v}$, we discretize Eq. 6.21 to find:

$$\tilde{v}_k = \frac{1}{\sqrt{N\Delta t}} \sum_{j=[-N/2]^+}^{[N/2]^-} \Delta t \, v_j e^{-\imath \frac{2\pi k j \Delta t}{N \Delta t}} \tag{6.50}$$

$$= \sqrt{\frac{\Delta t}{N}} \sum_{j=[-N/2]^+}^{[N/2]^-} v_j e^{-\imath \frac{2\pi k j}{N}} \tag{6.51}$$

What about the inverse transform? Notice that $\tilde{v}_k$ is N-periodic, $\tilde{v}_{k+N} = \tilde{v}_k$. The appropriate generalization of Eq. 6.22, which reduces to that equation in the limit $N \to \infty$, $\Delta t \to 0$, $N\Delta t = T$, is

$$v_j = \frac{1}{\sqrt{N\Delta t}} \sum_{k=[-N/2]^+}^{[N/2]^-} \tilde{v}_k e^{\imath \frac{2\pi k j}{N}} \tag{6.52}$$

Since both v_k and $\tilde{v}_k$ are N-periodic, it is common to express the Fourier transform and its inverse in terms of vectors with components running from 0 to $N-1$ rather than from $[-N/2]^+$ to

$[N/2]^-$. For example, using the periodicity of the $\tilde{\mathbf{v}}_k$, we can rewrite Eq. 6.52 as:

$$v_j = \frac{1}{\sqrt{N\Delta t}} \sum_{k=[-N/2]^+}^{[N/2]^-} e^{(2\pi\imath jk/N)} \tilde{v}_k \tag{6.53}$$

$$= \frac{1}{\sqrt{N\Delta t}} \left[\sum_{k=0}^{[N/2]^-} e^{(2\pi\imath jk/N)} \tilde{v}_k + \sum_{k=[-N/2]^+}^{-1} e^{(2\pi\imath jk/N)} \tilde{v}_{k+N} \right] \tag{6.54}$$

$$= \frac{1}{\sqrt{N\Delta t}} \left[\sum_{k=0}^{[N/2]^-} e^{(2\pi\imath jk/N)} \tilde{v}_k + \sum_{k=[N/2]^+}^{N-1} e^{(2\pi\imath jk/N)} \tilde{v}_k \right] \tag{6.55}$$

$$= \frac{1}{\sqrt{N\Delta t}} \sum_{k=0}^{N-1} e^{(2\pi\imath jk/N)} \tilde{v}_k \tag{6.56}$$

Thus, we can equally well think of frequency $(N-1)$ as frequency -1, frequency $(N-2)$ as frequency -2, etc. The same change can be performed for Eq. 6.51.

(Recall what is meant by a negative frequency: $e^{(2\pi\imath jk/N)} = \cos(2\pi jk/N) + \imath \sin(2\pi jk/N)$, while substituting $-k$ for k gives $e^{(2\pi\imath j(-k)/N)} = \cos(2\pi jk/N) - \imath\sin(2\pi jk/N) = \cos(2\pi jk/N) + \imath\sin([2\pi jk/N] + \pi)$ (recall that $\cos(x) = \cos(-x)$, $\sin(x) = -\sin(-x)$, and $\sin(x+\pi) = -\sin(x)$). So a negative frequency $-k$ is just like the positive frequency k except that the imaginary part – the sinusoid – is phase shifted by 180^o (that is, by π radians) relative to the real part.)

Problem 6.5 *It may seem odd that a high frequency like $(N-1)$ is the same as a low frequency like -1. To convince yourself of this, draw the real and imaginary parts of $e^{(2\pi\imath jk/N)}$ for $k = N-1$ and $k = -1$, as a function of integers j from 0 to $N-1$, for some small N, say $N = 4$. You'll find that the values coincide at each integer j (as they must given the periodicity with period N of the exponential, for integer j), though the values would be wildly different for j in between the integers. Because we're only looking discretely, at integer j's, we can't tell the difference between the two – they are identical in our discrete, N-dimensional world.*

The factors of Δt in Eqs. 6.51-6.56 are annoying, but we've seen that all that matters is the *product* of the factors in front of the forward and inverse equations, rather than their separate values alone, and Δt cancels out of that product. So let's eliminate Δt, and define the discrete Fourier transform as follows:

Definition 6.6 *Given an N-dimensional vector* $\mathbf{v}$*, with components v_k in the current basis: the* **Fourier transform** *of* $\mathbf{v}$ *is the vector* $\tilde{\mathbf{v}}$ *defined by*

$$\tilde{v}_j = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{(-2\pi\imath jk/N)} v_k \tag{6.57}$$

Note, jk is an integer: it is the product of the values of j and k, the two indices in question. The **inverse Fourier transform** *is defined by*

$$v_j = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{(2\pi\imath jk/N)} \tilde{v}_k \tag{6.58}$$

To show that this is the right expression for the inverse transform, we need to first establish a fundamental relationship for complex exponentials:

$$\sum_{k=0}^{N-1} e^{2\pi \imath jk/N} = \begin{cases} N & \text{if } j \text{ is an integral multiple of N} \\ 0 & \text{otherwise} \end{cases} \tag{6.59}$$

This is a discrete analog of Eq. 6.20 for the continuous case. There is a proof and a full discussion in the Appendix, see Eq. 6.151.

Exercise 6.1 *Write down the proof of Eq. 6.59. Here's a brief outline: For j an integral multiple of N – $j = pN$ for some integer p – this is easy: $e^{2\pi\imath pk} = 1$ for all integer k, and there are N terms, so the sum gives N. For j not an integral multiple of N, the task is to show that the sum gives 0. Geometrically, note that the complex exponentials in the sum represent a sequence of unit vectors in the complex plane (see Fig. 4.1 with an angle $2\pi j/N$ between successive vectors. The vectors go around the unit circle counterclockwise j times as k goes from 0 to N. So, rotating each vector through an angle $2\pi j/N$ just takes each vector to the next one around the circle; this also takes the last one to the first one. Therefore, rotating the whole set of vectors by this angle leaves the set unchanged, and therefore must leave the sum over the vectors unchanged. But rotating the whole set of vectors by this angle must also rotate the sum of the vectors by this angle (this sum is just a complex number, hence it is also a vector in the complex plane). Unless this angle represents a complete rotation — that is, unless j is an integral multiple of N — you can't rotate a nonzero vector by this angle and get the same vector back. So the sum must be zero. To express this mathematically, note that rotating a complex number by $2\pi j/N$ means multiplying it by $\exp(\imath 2\pi j/N)$. Write $s(j)$ for the sum, multiply each vector in the sum by $\exp(\imath 2\pi j/N)$, show that this leaves the sum unchanged, but show that this also yields $\exp(\imath 2\pi j/N)s(j)$. So $\exp(\imath 2\pi j/N)s(j) = s(j)$, so either $\exp(\imath 2\pi j/N) = 1$ or $s(j) = 0$.*

Problem 6.6 *Understanding the Fourier transform as a change of basis:*

- *Show that the Fourier transform can be rewritten as the vector equation,*

$$\tilde{\mathbf{v}} = \mathbf{U}^{\mathrm{FT}}\mathbf{v} \tag{6.60}$$

  where $\mathbf{U}^{\mathrm{FT}}$ is the matrix with components $U^{\mathrm{FT}}_{jk} = \frac{1}{\sqrt{N}} e^{-2\pi\imath(jk)/N}$.

- *Rewrite the matrix $\mathbf{U}^{\mathrm{FT}}$ as*

$$\mathbf{U}^{\mathrm{FT}} = \begin{pmatrix} \tilde{\mathbf{e}}_0^\dagger \\ \tilde{\mathbf{e}}_1^\dagger \\ \cdots \\ \tilde{\mathbf{e}}_{N-1}^\dagger \end{pmatrix} \tag{6.61}$$

  where $\tilde{\mathbf{e}}_j^\dagger$ is the row vector with components $(\tilde{\mathbf{e}}_j^\dagger)_k = \frac{1}{\sqrt{N}} e^{-2\pi\imath jk/N}$; that is,

$$\tilde{\mathbf{e}}_j^\dagger = \frac{1}{\sqrt{N}}(1, e^{-2\pi\imath j/N}, e^{-2\pi\imath(2j)/N}, \ldots, e^{-2\pi\imath(N-1)j)/N}) \tag{6.62}$$

- *Show that the $\tilde{\mathbf{e}}_j$ are orthonormal, $\tilde{\mathbf{e}}_j^\dagger \tilde{\mathbf{e}}_k = \delta_{jk}$, by using Eq. 6.59. (Don't forget to take the complex conjugate in the elements of $\tilde{\mathbf{e}}_j^\dagger$ relative to those of $\tilde{\mathbf{e}}_k$.) Therefore, $\mathbf{U}^{\mathrm{FT}}$ is a unitary matrix, $\mathbf{U}^{\mathrm{FT}}(\mathbf{U}^{\mathrm{FT}})^\dagger = \mathbf{1}$.*

- *Thus, establish that the Fourier transform is a transformation to the orthonormal basis of vectors* $\tilde{\mathbf{e}}_j$ *defined, in the pre-transform basis, by* $(\tilde{\mathbf{e}}_j)_k = e^{2\pi\imath jk/N}$, *or*

$$\tilde{\mathbf{e}}_j = \frac{1}{\sqrt{N}}(1, e^{2\pi\imath j/N}, e^{2\pi\imath(2j)/N}, \ldots, e^{2\pi\imath(N-1)j/N})^{\mathrm{T}} \tag{6.63}$$

(See exercise 4.11).

- *Rewrite this new basis in terms of* cos *and* sin*:* $\tilde{\mathbf{e}}_j = \tilde{\mathbf{e}}_j^{\mathrm{RE}} + i\tilde{\mathbf{e}}_j^{\mathrm{IM}}$ *where*

$$\tilde{\mathbf{e}}_j^{\mathrm{RE}} = \frac{1}{\sqrt{N}}(1, \cos(2\pi j/N), \cos(2\pi(2j)/N), \ldots, \cos(2\pi(N-1)j/N))^{\mathrm{T}} \tag{6.64}$$

and

$$\tilde{\mathbf{e}}_j^{\mathrm{IM}} = \frac{1}{\sqrt{N}}(0, \sin(2\pi j/N), \sin(2\pi(2j)/N), \ldots, \sin(2\pi(N-1)j/N))^{\mathrm{T}} \tag{6.65}$$

- *Show that* $\tilde{\mathbf{e}}_j^{\mathrm{RE}}$ *is a cosine vector that oscillates, in going from the* 0^{th} *to the* $(N-1)^{th}$ *elements, through* j *cycles; while* $\tilde{\mathbf{e}}_j^{\mathrm{IM}}$ *is a sin vector that oscillates through* j *cycles. Thus, the Fourier transform is a transformation to a basis of cos and sin vectors of every integral frequency from 0 to (N-1)* (*here, "frequency" is "cycles/vector"; noting that, in any given basis vector, the last cycle is actually not quite completed, but would be completed if the vector were given one more element, namely* $(\tilde{\mathbf{e}}_j)_N = (\tilde{\mathbf{e}}_j)_0$).

Problem 6.7 *The inverse Fourier transform:*

- *Apply the matrix* $\mathbf{U}^{\mathrm{FT}\dagger}$ *to Eq. 6.60, to show that the inverse Fourier transform is given by*

$$\mathbf{v} = \mathbf{U}^{\mathrm{FT}\dagger}\tilde{\mathbf{v}} \tag{6.66}$$

(Note, since $\mathbf{U}^{\mathrm{FT}}$ *is unitary, you know that* $\mathbf{U}^{\mathrm{FT}\dagger} = (\mathbf{U}^{\mathrm{FT}})^{-1}$*; so nothing fancy is required here, just multiply the matrices.) Now show that, in components, this is*

$$v_j = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} e^{2\pi\imath jk/N}\tilde{v}_k \tag{6.67}$$

(*recall that for any matrix* $\mathbf{M}$, $(\mathbf{M}^\dagger)_{jk} = M^*_{kj}$).

- *From Eq. 6.67, show that*

$$\mathbf{U}^{\mathrm{FT}\dagger} = \begin{pmatrix} \tilde{\mathbf{e}}_0^{\mathrm{T}} \\ \tilde{\mathbf{e}}_1^{\mathrm{T}} \\ \cdots \\ \tilde{\mathbf{e}}_{N-1}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} (\tilde{\mathbf{e}}_0^*)^\dagger \\ (\tilde{\mathbf{e}}_1^*)^\dagger \\ \cdots \\ (\tilde{\mathbf{e}}_{N-1}^*)^\dagger \end{pmatrix} \tag{6.68}$$

where the $\tilde{\mathbf{e}}_j$ *are the Fourier basis vectors (Eq. 6.63).*

- *By taking the adjoint of Eq. 6.61, show that we can also write* $\mathbf{U}^{\mathrm{FT}\dagger} = (\ \tilde{\mathbf{e}}_0\ \tilde{\mathbf{e}}_1\ \ldots\ \tilde{\mathbf{e}}_{N-1}\ )$. *(Thus, the unitary matrix* $\mathbf{U}^{\mathrm{FT}\dagger}$ *is symmetric – its columns are the same as its rows, so it is equal to its transpose – but it is* not *Hermitian, that is, it is not equal to its adjoint. The same is true of its adjoint, the unitary matrix* $\mathbf{U}^{\mathrm{FT}}$*.)*

Problem 6.8 *Let's look at some examples:*

- *Let N=2.*
  - *Write down the two Fourier basis vectors,* $\tilde{\mathbf{e}}_0$ *and* $\tilde{\mathbf{e}}_1$*. Draw these two vectors in the coordinate system of the original basis.*[11]
  - *Interpret the real and imaginary parts of* $\tilde{\mathbf{e}}_0$ *and* $\tilde{\mathbf{e}}_1$ *graphically in terms of* sin *and* cos*, by using the depiction of vectors shown in figure 6.1. To do this, graph the relevant* sin *or* cos *function as a continuous function on the interval from 0 to 2 (hand-sketch is fine); and then show the elements of the corresponding real or imaginary part of* $\tilde{\mathbf{e}}_0$ *or* $\tilde{\mathbf{e}}_1$*, as in Fig. 6.1, as the values of this function at 0, 1, and 2.*
  - *Write down the matrix* $\mathbf{U}^{\mathrm{FT}}$*.*
  - *Write down* $\tilde{\mathbf{v}}$*, the Fourier transform of the vector* $\mathbf{v} = (v_0, v_1)^{\mathrm{T}}$*.*
- *Let N=4.*
  - *Write down the four Fourier basis vectors,* $\tilde{\mathbf{e}}_0, \ldots, \tilde{\mathbf{e}}_3$*.*
  - *Interpret the real and imaginary parts of these vectors in terms of sin and cos, as in the the* $N = 2$ *case, but now over the interval from 0 to 4.*
  - *Write down the matrix* $\mathbf{U}^{\mathrm{FT}}$*.*
  - *Write down* $\tilde{\mathbf{v}}$*, the Fourier transform of the vector* $\mathbf{v} = (v_0, v_1, v_2, v_3)^{\mathrm{T}}$*.*

You should come away with two senses: (1) the Fourier transform is just another change of basis, albeit a special one; and (2) the Fourier basis vectors in N dimensions are discrete approximations to the set of sin and cos functions that have from 0 to $(N-1)$ *cycles.*

Exercise 6.2 *Let* $\tilde{\mathbf{e}}_j$ *be the Fourier basis, and let* $x = e^{2\pi\imath/N}$*. Note that* $\tilde{\mathbf{e}}_j = \frac{1}{\sqrt{N}}(1, x^j, x^{2j}, \ldots, x^{(N-1)j})^{\mathrm{T}}$*. This series, if continued, would be periodic — that is,* $x^{Nj} = 1$*,* $x^{(N+1)j} = x^j$*,* $x^{(N+2)j} = x^{2j}$*, … — because for any k,* $x^{(k+N)j} = x^{kj}$*.*

Exercise 6.2 demonstrates that the N-dimensional Fourier transform basis vectors $\tilde{\mathbf{e}}_j$ can naturally be thought of as infinite periodic vectors, with period N, rather than as a finite vector of length N. This makes it natural to think of any N-dimensional vector $\mathbf{v} = \sum_{j=0}^{(N-1)} v_j \tilde{\mathbf{e}}_j$ as an infinite periodic vector with period N — it is a linear combination of the basis vectors $\tilde{\mathbf{e}}_j$, each of which is periodic.

Let $y = x^* = e^{-2\pi\imath/N}$. Then exercise 6.2 also means that we can write the Fourier transform matrix, $\mathbf{U}^{\mathrm{FT}}$, as

$$\mathbf{U}^{\mathrm{FT}} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & y & y^2 & \ldots & y^{(N-2)} & y^{(N-1)} \\ 1 & y^2 & y^4 & \ldots & y^{2(N-2)} & y^{2(N-1)} \\ & & & \ldots & & \\ 1 & y^{(N-2)} & y^{2(N-2)} & \ldots & y^{(N-2)(N-2)} & y^{(N-1)(N-2)} \\ 1 & y^{(N-1)} & y^{2(N-1)} & \ldots & y^{(N-2)(N-1)} & y^{(N-1)(N-1)} \end{pmatrix} \tag{6.69}$$

[11] You may be confused by the fact that the new bases are not obtained by a simple rotation from the old bases. If so, consider renaming these basis vectors, so that what was $\tilde{\mathbf{e}}_0$ is now $\tilde{\mathbf{e}}_1$, and what was $\tilde{\mathbf{e}}_1$ is now $\tilde{\mathbf{e}}_0$. With this renaming, the 2-D Fourier transform is just a rotation of bases by $-45^{\cdot}$. Thus, the Fourier transform, with basis vectors named as given in the text, amounts to a rotation of bases through $-45^{\cdot}$, plus a mirror reflection or exchange of bases (mirror reflection about the initial $(1,0)$ axis). This is just a matter of naming; by reordering the Fourier Transform basis vectors, the transformation to that basis would just be a rotation.

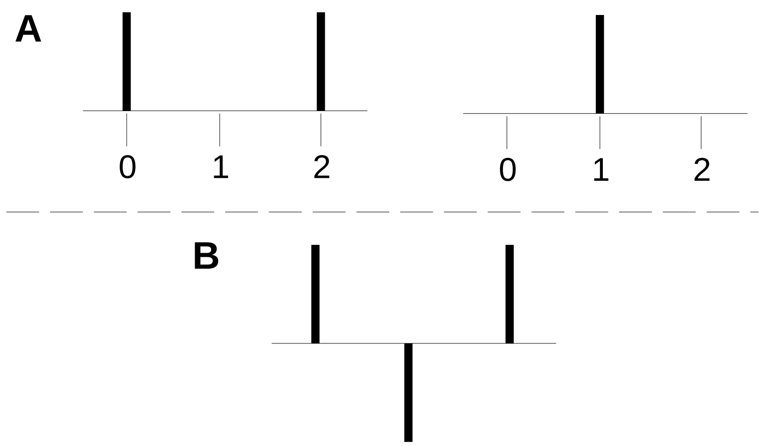

Figure 6.1: **Another Way of Depicting Vectors**
We can think of the two-dimensional vector $\mathbf{v} = (v_0, v_1)^{\mathrm{T}}$ as being a periodic function of the integers i: $v_{i+2} = v_i$ for all i; in particular, $v_2 = v_0$. We can then depict this vector on the real line, by showing its values at each integer, as shown. In A, the two usual basis vectors are depicted: left, $(1, 0)$; right, $(0, 1)$. In B, the vector $(1, -1)$ is shown.

Due to the periodicity of y^k ($y^N = 1$, so $y^{(k+N)j} = y^{kj}$), there are only N different numbers in this matrix. For example, $y^{2(N-1)} = y^{-2} = y^{N-2}$. So we can rewrite this matrix, for example, as

$$\mathbf{U}^{\rm FT} = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & y & y^2 & \ldots & y^{(N-2)} & y^{(N-1)} \\ 1 & y^2 & y^4 & \ldots & y^{(N-4)} & y^{(N-2)} \\ & & \ldots & & & \\ 1 & y^{(N-2)} & y^{(N-4)} & \ldots & y^4 & y^2 \\ 1 & y^{(N-1)} & y^{(N-2)} & \ldots & y^2 & y \end{pmatrix} \tag{6.70}$$

This redundancy in the Fourier transform matrix – the fact that its N^2 entries include only N different numbers, which are N powers of a single number – provides the basis for the fast Fourier transform, a method of computing $\mathbf{U}^{\rm FT}\mathbf{v}$ in order $N \log N$ rather than order N^2 multiplications.

6.5.1 The convolution theorem for discrete Fourier Transforms

The power of the Fourier transform is its power to diagonalize translation-invariant operators, which in the discrete case amounts to diagonalizing convolutions:

Definition 6.7 *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be two N-dimensional vectors, which we shall think of as extended periodically:* $v_{j+N} = v_j$*,* $w_{j+N} = w_j$*, for any integer* j*. The* **convolution** *of* $\mathbf{v}$ *with* $\mathbf{w}$ *is defined as the vector* $(\mathbf{v} * \mathbf{w})$ *with components*

$$(\mathbf{v} * \mathbf{w})_j = \sum_{k=0}^{N-1} v_{j-k} w_k \tag{6.71}$$

As we've noted, convolutions are operations that come up constantly. They arise whenever you are applying the same operation to a vector without regard for where in the vector you are: for example, replacing each element by a weighted average of the element and its two nearest neighbors. Such operations are called *translation-invariant* — if you translate the vector by p, taking $w_j \mapsto w_{j+p}$, the operation is not changed. The importance of the Fourier transform is: *the Fourier transform is the transform that diagonalizes a convolution*, and more generally that diagonalizes a translation-invariant operator. As we've seen, in the continuous case, this also means that the Fourier transform diagonalizes derivative operators (*e.g.* $\frac{d}{dt}$).

Problem 6.9 *Show that* $\mathbf{v}*\mathbf{w} = \mathbf{w}*\mathbf{v}$*. To do this: in Eq. 6.71, substitute* $p = j-k$ *to find* $(\mathbf{v}*\mathbf{w})_j = \sum_{p=j-(N-1)}^{j} w_{j-p} v_p$*. Now use the periodicity of the two vectors to show that* $\sum_{p=j-(N-1)}^{j} w_{j-p} v_p = \sum_{p=0}^{N-1} w_{j-p} v_p$ *(Show that the two sums are summing the same terms, just in different orders). This last term, in turn, is* $\mathbf{w} * \mathbf{v}$*.*

Problem 6.10 *Show that we can write the convolution in terms of a matrix multiplication,* $\mathbf{v}*\mathbf{w} = \mathbf{V}\mathbf{w}$*, where* $\mathbf{V}$ *is the matrix*

$$\mathbf{V} = \begin{pmatrix} v_0 & v_{N-1} & \ldots & v_2 & v_1 \\ v_1 & v_0 & \ldots & v_3 & v_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ v_{N-2} & v_{N-3} & \ldots & v_0 & v_{N-1} \\ v_{N-1} & v_{N-2} & \ldots & v_1 & v_0 \end{pmatrix} \tag{6.72}$$

(recall that $v_{-1} = v_{N-1}$*, and more generally,* $v_j = v_{j+N}$*). A matrix of the form given in Eq. 6.72 is also known as a* circulant *matrix.*

Now we will show that convolutions are solved by the Fourier transform, and learn how to do Fourier transforms at the same time. Consider the convolution vector, $\mathbf{v} * \mathbf{w}$, with components $(\mathbf{v} * \mathbf{w})_j = \sum_{k=0}^{N-1} v_{j-k} w_k$. Let $\tilde{\mathbf{v}}$ be the Fourier transform of $\mathbf{v}$, $\tilde{\mathbf{w}}$ be the Fourier tranform of $\mathbf{w}$, and $\widetilde{\mathbf{v} * \mathbf{w}}$ be the Fourier transform of $(\mathbf{v} * \mathbf{w})$. We will show that, in Fourier space, the convolution becomes $(\widetilde{\mathbf{v} * \mathbf{w}})_k = \sqrt{N} \tilde{v}_k \tilde{w}_k$. That is, the Fourier tranform of the vector $\mathbf{v} * \mathbf{w}$ is the vector whose k^{th} element is $\sqrt{N} \tilde{v}_k \tilde{w}_k$.

Thus: Convolution in real space means, in Fourier space, just multiplying together, frequency by frequency, the components of two vectors. That is, in Fourier space, operating with $\mathbf{v}*$ or with $\mathbf{V}$ on $\mathbf{w}$ is just multiplication of the k^{th} component of $\tilde{\mathbf{w}}$ by $\tilde{v}_k$ (and then rescaling of the entire Fourier-space vector by $\sqrt{N}$). Intuitively, $\mathbf{v}$ is selecting certain frequencies of $\mathbf{w}$: if $\mathbf{v}$ is dominated by low frequencies, the convolution will amplify the low frequencies of $\mathbf{w}$ relative to the high frequencies. The different frequency components of $\tilde{\mathbf{w}}$ are not mixed together by the convolution; each frequency component is just multiplied by a different factor. This means that, under the Fourier transform, $\mathbf{V}$ becomes the diagonal matrix $\tilde{\mathbf{V}}$ with diagonal elements $\tilde{V}_{ii} = \sqrt{N} \tilde{v}_i$. Thus, the Fourier basis vectors $\tilde{\mathbf{e}}_i$ diagonalize the convolution, and the corresponding eigenvalues are proportional to the corresponding frequency components of $\mathbf{v}$, $\tilde{v}_i$.

Problem 6.11 *We will show that* $(\widetilde{\mathbf{v} * \mathbf{w}})_k = \sqrt{N} \tilde{v}_k \tilde{w}_k$. *To show this, we will apply the following general method for executing a Fourier transform:*

1. *Replace each real-space component by its expression in terms of Fourier-space components (Eq. 6.67).*

2. *Collect all the terms corresponding to the original real-space summation(s); these will only involve exponentials, and the sum(s) over them will turn into delta function(s) as in Eqs. 6.59.*

3. *Use each delta function to execute one of the remaining sums.*

4. *Apply the Fourier tranform operator to each side of the equation, or just read off the Fourier transform from the equation that you have at this point.*

Let's see how each of these steps works in practice:

1. *Equation 6.67 gives* $v_j = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} e^{2\pi \imath jp/N} \tilde{v}_p$, *and similarly for* $\mathbf{w}$, *where the dummy variable* p *has been used rather than* k *as in Eq. 6.67. Substituting for* v_{j-k} *and* w_k *in the definition of* $(\mathbf{v} * \mathbf{w})_j$, *using the dummy variable* q *in the expression for* w_k, *gives:*

$$(\mathbf{v} * \mathbf{w})_j = \sum_{k=0}^{N-1} v_{j-k} w_k = (1/N) \sum_{k=0}^{N-1} \left(\sum_{p=0}^{N-1} e^{2\pi \imath (j-k)p/N} \tilde{v}_p \right) \left(\sum_{q=0}^{N-1} e^{2\pi \imath kq/N} \tilde{w}_q \right) \tag{6.73}$$

2. *You now collect all the terms that depend on the original summation variable,* k*: these are*

$$\sum_{k=0}^{N-1} e^{2\pi \imath (j-k)p/N} e^{2\pi \imath kq/N} = e^{2\pi \imath jp/N} \sum_{k=0}^{N-1} e^{-2\pi \imath k(p-q)/N} \tag{6.74}$$

But, from Eq. 6.59,

$$\sum_{k=0}^{N-1} e^{-2\pi \imath k(p-q)/N} = N \delta_{((p-q) \bmod N)0} \tag{6.75}$$

Thus, you obtain

$$(\mathbf{v} * \mathbf{w})_j = \sum_{k=0}^{N-1} v_{j-k} w_k = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} e^{2\pi\imath jp/N} \tilde{v}_p \tilde{w}_q \delta_{((p-q) \bmod N)0} \tag{6.76}$$

NOTE: You could not have collected terms to do the sum over p or over q, because these contain v_p or w_q. Only the sum over k was a pure sum over exponentials, of the form given in Eq. 6.59.

3. *Use the delta function to do the summation over q to obtain*

$$(\mathbf{v} * \mathbf{w})_j = \sum_{k=0}^{N-1} v_{j-k} w_k = \sum_{p=0}^{N-1} e^{2\pi\imath jp/N} \tilde{v}_p \tilde{w}_p \tag{6.77}$$

4. (a) *By comparing this to Eq. 6.67, you can read off that the Fourier transform of $\mathbf{v} * \mathbf{w}$ is $(\widetilde{\mathbf{v} * \mathbf{w}})_p = \sqrt{N} \tilde{v}_p \tilde{w}_p$*

   (b) *Alternatively, take the Fourier transform of each side by applying Eq. 6.57. Each side of Eq. 6.77 expresses the j^{th} component of a vector in real space; so, to find the k^{th} component in Fourier space, use Eq. 6.57 (with $j \leftrightarrow k$) and Eq. 6.77 to find:*

$$(\widetilde{\mathbf{v} * \mathbf{w}})_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{-2\pi\imath kj/N} (\mathbf{v} * \mathbf{w})_j = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{-2\pi\imath kj/N} \sum_{p=0}^{N-1} e^{2\pi\imath jp/N} \tilde{v}_p \tilde{w}_p \tag{6.78}$$

The sum over j gives $\sum_{j=0}^{N-1} e^{-2\pi\imath j(k-p)/N} = N\delta_{((k-p) \bmod N)0}$. Using this to do the p summation leaves

$$(\widetilde{\mathbf{v} * \mathbf{w}})_k = \sqrt{N} \tilde{v}_k \tilde{w}_k \tag{6.79}$$

Congratulations! You've not only proven an important theorem, the convolution theorem, but you've also used a very general method for transforming equations to Fourier space: replace each component of the equation by its expression in terms of Fourier-space components, collect sums over exponentials to give delta functions, use these to reduce the remaining sums, and perhaps take the Fourier transform of the final results. When the dust clears, you have an equation in Fourier-space.

The above was a brute-force method of doing the calculation, but it's the general way to do such calculations so it was good to learn it. We now give a more elegant proof that gives more insight into why the Fourier basis diagonalizes a convolution:

Exercise 6.3 *We show that the Fourier basis vectors are eigenvectors of any convolution, as follows. We write $\tilde{\mathbf{e}}_j = \frac{1}{\sqrt{N}}(1, x^j, x^{2j}, \ldots, x^{(N-1)j})^{\mathrm{T}}$, as in problem 6.2. To compute $\mathbf{V}\tilde{\mathbf{e}}_j$, we write out its elements as follows:*

$$\mathbf{V}\tilde{\mathbf{e}}_j = \begin{pmatrix} (\mathbf{V}\tilde{\mathbf{e}}_j)_0 \\ (\mathbf{V}\tilde{\mathbf{e}}_j)_1 \\ \ldots \\ (\mathbf{V}\tilde{\mathbf{e}}_j)_{N-1} \end{pmatrix} = \frac{1}{\sqrt{N}} \begin{pmatrix} v_0 + x^j v_{N-1} + x^{2j} v_{N-2} + \ldots + x^{(N-1)j} v_1 \\ v_1 + x^j v_0 + x^{2j} v_{N-1} + \ldots + x^{(N-1)j} v_2 \\ \ldots \\ v_{N-1} + x^j v_{N-2} + x^{2j} v_{N-3} + \ldots + x^{(N-1)j} v_0 \end{pmatrix} \tag{6.80}$$

Now, using the fact that $x^{Nj} = 1$, show that

- $(\mathbf{V}\tilde{\mathbf{e}}_j)_1 = x^j(\mathbf{V}\tilde{\mathbf{e}}_j)_0$;
- $(\mathbf{V}\tilde{\mathbf{e}}_j)_2 = x^j(\mathbf{V}\tilde{\mathbf{e}}_j)_1 = x^{2j}(\mathbf{V}\tilde{\mathbf{e}}_j)_0$;
- $(\mathbf{V}\tilde{\mathbf{e}}_j)_3 = x^j(\mathbf{V}\tilde{\mathbf{e}}_j)_2 = x^{3j}(\mathbf{V}\tilde{\mathbf{e}}_j)_0$;
- ....

Thus, show that $\mathbf{V}\tilde{\mathbf{e}}_j = (\mathbf{V}\tilde{\mathbf{e}}_j)_0\ (1, x^j, x^{2j}, \ldots, x^{(N-1)j})^{\mathrm{T}} = (\mathbf{V}\tilde{\mathbf{e}}_j)_0\ \sqrt{N}\ \tilde{\mathbf{e}}_j$

Finally, let's rewrite this. Let $\lambda_j = \sqrt{N}(\mathbf{V}\tilde{\mathbf{e}}_j)_0 = (v_0 + x^j v_{N-1} + x^{2j} v_{N-2} + \ldots + x^{(N-1)j} v_1)$. *Multiply through by* $1 = x^{-Nj}$ *to convert this to* $\lambda_j = (v_0 + x^{-j} v_1 + x^{-2j} v_2 + \ldots + x^{-(N-1)j} v_{N-1})$. *Show that* $\lambda_j = \sqrt{N}\tilde{v}_j$*, where* $\tilde{v}_j$ *is the* j^{th} *component of the Fourier transform of* $\mathbf{v}$ *(look back at Eqs. 6.61-6.62, and recall the definition in problem 6.2,* $x = e^{2\pi\imath/N}$*). Thus,* $\mathbf{V}\tilde{\mathbf{e}}_j = \lambda_j\tilde{\mathbf{e}}_j = \sqrt{N}\tilde{v}_j\tilde{\mathbf{e}}_j$*. In summary,* $\tilde{\mathbf{e}}_j$ *is an eigenvector of* $\mathbf{V}$ *with eigenvalue* $\lambda_j = \sqrt{N}\tilde{v}_j$*. (Compare this result to the result of Problem 6.11, and make sure you understand the relationship between the two, e.g., how to derive the result of Problem 6.11 directly from the present result).*

In words: each row of the convolution matrix $\mathbf{V}$ *is just a translation-by-1 of the previous row (look at Eq. 6.72 to understand what this means). This, along with the fact that everything is assumed periodic, means that the operation of each successive row of* $\mathbf{V}$ *on a vector* $\mathbf{w}$ *is equivalent to the first row of* $\mathbf{V}$ *acting on a translated version of* $\mathbf{w}$ *(e.g., operation of the second row is like the first row acting on a translated-by-(*-1*) version of* $\mathbf{w}$*; etc.). The Fourier basis vectors* $\tilde{\mathbf{e}}_j$ *are precisely those vectors that are eigenvectors under the operation of translation: under translation-by-1,* $\tilde{\mathbf{e}}_j \mapsto x^j\tilde{\mathbf{e}}_j$*. So, a convolution acting on* $\tilde{\mathbf{e}}_j$ *gives a result proportional to* $(1, x^j, x^{2j}, \ldots, x^{N-1})$*, that is, proportional to* $\tilde{\mathbf{e}}_j$*.*

Exercise 6.4 *For those interested, here's more about translation-invariance. Let the left-translation operator* $\mathbf{L}$ *be defined by its action on a vector:* $\mathbf{L}(v_0\ v_1\ \ldots\ v_{N-1})^{\mathrm{T}} = (v_1\ v_2\ \ldots\ v_{N-1}\ v_0)^{\mathrm{T}}$ *(recall, we're taking all vectors to be periodic, so* $v_N = v_0$*). Define the right-translation operator similarly:* $\mathbf{R}(v_0\ v_1\ \ldots\ v_{N-1})^{\mathrm{T}} = (v_{N-1}\ v_0\ \ldots\ v_{N-3}\ v_{N-2})^{\mathrm{T}}$*. As a matrix,* $\mathbf{L}$ *has components*

$$\mathbf{L} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \tag{6.81}$$

while $\mathbf{R} = \mathbf{L}^{\mathrm{T}}$ *is*

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \tag{6.82}$$

It's easy to show $\mathbf{RL} = \mathbf{LR} = \mathbf{1}$*; since* $\mathbf{R} = \mathbf{L}^{\mathrm{T}}$*, this means that* $\mathbf{R}$ *and* $\mathbf{L}$ *are unitary. Show that any convolution matrix* $\mathbf{V}$ *commutes with the translation operators:* $\mathbf{RV} = \mathbf{VR}$*,* $\mathbf{LV} = \mathbf{VL}$*. Therefore, it is invariant under a translation:* $\mathbf{V} \mapsto \mathbf{RVR}^{-1} = \mathbf{VRR}^{-1} = \mathbf{V1} = \mathbf{V}$*, and similarly for* $\mathbf{L}$*.*

Show that the Fourier basis vectors $\tilde{\mathbf{e}}_j$ *are eigenvectors of the translation operators:* $\mathbf{L}\tilde{\mathbf{e}}_j = x^j\tilde{\mathbf{e}}_j$*,* $\mathbf{R}\tilde{\mathbf{e}}_j = x^{-j}\tilde{\mathbf{e}}_j$*, with* x *as defined in problem 6.2. In fact, for either* $\mathbf{L}$ *or* $\mathbf{R}$*, the Fourier basis vectors are a complete orthonormal basis of eigenvectors (complete, because there are* N *of them), with distinct eigenvalues. The Fourier basis vectors are the only eigenvectors of either translation* *operator.*[12] *Note that this is quite a special property: for example, the ordinary basis,* $(\mathbf{e}_i)_j = \delta_{ij}$*, does not translate into multiples of itself; rather,* $\mathbf{Re}_0 = \mathbf{e}_1$*,* $\mathbf{Re}_1 = \mathbf{e}_2$*,* $\mathbf{Le}_0 = \mathbf{e}_{N-1}$*,* $\mathbf{Le}_1 = \mathbf{e}_0$*, etc.*

Whenever two normal matrices commute, they have a common basis of eigenvectors. In this case, this means that, because $\mathbf{V}$ *commutes with the translation operators, and because the Fourier basis is the unique basis of eigenvectors for these operators, the Fourier basis also forms the eigenvectors of* $\mathbf{V}$*. To see this in this case, note that* $\mathbf{V}\tilde{\mathbf{e}}_j$ *is an eigenvector of* $\mathbf{L}$ *or* $\mathbf{R}$*, with the same eigenvalue as* $\tilde{\mathbf{e}}_j$*:* $\mathbf{LV}\tilde{\mathbf{e}}_j = \mathbf{VL}\tilde{\mathbf{e}}_j = x^j\mathbf{V}\tilde{\mathbf{e}}_j$*, and similarly for* $\mathbf{R}$*. But this means that* $\mathbf{V}\tilde{\mathbf{e}}_j \propto \tilde{\mathbf{e}}_j$*: that is,* $\tilde{\mathbf{e}}_j$ *is an eigenvector of* $\mathbf{V}$*.*

In summary, we've come to understand the discrete Fourier transform as a change of basis to a special set of basis vectors. The Fourier basis is the basis of cos and sin vectors of every integral frequency from 0 to (N-1). This basis is special because it is the basis that diagonalizes convolutions, and more generally diagonalizes any translation-invariant operator. This is because, as shown in Ex. 6.4, the Fourier basis vectors are the eigenvectors of the translation operator: they are precisely the vectors that return a multiple of themselves under translation.

6.6 The Fourier Transform for Functions on an Infinite Domain

We again begin from Eqs. 6.21-6.22, but now we are going to take the limit $T \to \infty$, to the case of a function on an infinite domain. We use our freedom to juggle the factors in front of the Fourier transform and its inverse – only the product of these factors matters – to replace Eqs. 6.21-6.22 with the equations

$$f_k = \int_{-T/2}^{T/2} dt\ f(t)e^{-\imath\frac{2\pi kt}{T}} \tag{6.83}$$

$$f(t) = \frac{1}{T}\sum_{k=-\infty}^{\infty} f_k e^{\imath\frac{2\pi kt}{T}} \tag{6.84}$$

We replace the frequency k with $m = 2\pi k/T$; the interval between frequencies m is $\Delta m = 2\pi/T$. We let $\tilde{f}(m)$ be the m^{th} frequency component of the Fourier transform of $f(t)$. Equations 6.21-6.22 become:

$$\tilde{f}(m) = \int_{-T/2}^{T/2} dt\ f(t)e^{-\imath mt} \tag{6.85}$$

$$f(t) = \frac{1}{T\Delta m}\sum_{m=-\infty}^{\infty} \Delta m \tilde{f}(m)e^{\imath mt} \tag{6.86}$$

$$= \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} \Delta m \tilde{f}(m)e^{\imath mt} \tag{6.87}$$

where the sum over m is in steps of Δm. We take the limit $T \to \infty$, $\Delta m \to 0$ to obtain

$$\tilde{f}(m) = \int_{-\infty}^{\infty} dt\ f(t)e^{-\imath mt} \tag{6.88}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(m)e^{\imath mt} \tag{6.89}$$

[12]The N eigenvectors of a matrix are unique, except that arbitrary linear combinations of eigenvectors sharing a common eigenvalue are also eigenvectors. Here, all eigenvalues are distinct.

Finally, if we wish, we can use our freedom to rearrange the factors out front to make them more symmetrical, to arrive at:

Definition 6.8 *The* **Fourier transform** *of a function $f(t)$ defined on the infinite interval is given by*

$$\tilde{f}(m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\ f(t) e^{-\imath mt} \tag{6.90}$$

The **inverse Fourier transform** *in this case is given by*

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dm\ \tilde{f}(m) e^{\imath mt} \tag{6.91}$$

In this case, the sum-over-exponentials that you need to prove that this is the correct inverse is

$$\int_{-\infty}^{\infty} dt\ e^{\imath kt} = 2\pi\delta(k) \tag{6.92}$$

where the right side is the Dirac delta function. The derivation of this formula is given in the Appendix (see Eq. 6.150).

Problem 6.12 *Prove that Eq. 6.91 for the inverse Fourier transform is indeed the right formula, by substituting Eq. 6.90 for $\tilde{f}(m)$ into the right side of Eq. 6.91 (note that the dummy integration variable in Eq. 6.90 must be called something other than t, e.g. t', since t is already in use in Eq. 6.91), collecting the appropriate exponentials, converting them into delta functions using Eq. 6.92, and so showing that the right side of Eq. 6.91 indeed gives $f(t)$.*

Problem 6.13 *Prove the convolution theorem for functions on an infinite interval: take the Fourier transform $\tilde{c}(m)$ of $c(t) = \int_{-\infty}^{\infty} dt'\ g(t-t')f(t')$ and show that the result is proportional to $\tilde{g}(m)\tilde{f}(m)$. You can follow the proofs in either section 6.2.3 or section 6.5.1 – or better yet, do it twice, once following each section!*

Problem 6.14 *Consider the equation $h(x) = a(x) \int dx'\ g(x-x')f(x')$ (limits of $-\infty$ to ∞ are assumed where limits are not stated). Show that its Fourier transform is $\tilde{h}(m) \propto \int dp\ \tilde{a}(m-p)\tilde{g}(p)\tilde{f}(p)$. You can do this the brute force way, by writing down the expression for $\tilde{h}(m)$; using Eq. 6.91 to express each of the functions $a(x)$, $g(x-x')$, and $f(x')$ in terms of their Fourier transforms (be sure to use a different dummy integration variable for each one, i.e. one's frequency variable might be called k, another's l, another's m; and be sure to have the appropriate factor – x, $x-x'$, or x' for the three different functions – in the exponential; so for example you would write $g(x-x') = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dl\, \tilde{g}(l) e^{il(x-x')}$); converting sums of exponentials to delta functions; and letting the dust clear. Do that, it's good practice. But also note, you can do it more directly, as follows: just as a convolution in real space is a product in Fourier space, so a product in real space (of the two functions $a(x)$ and $b(x) = \int dx'\ g(x-x')f(x')$) is a convolution in Fourier space. Use this fact, along with the convolution theorem which gives the Fourier transform of $b(x)$.*

6.7 The Fourier Transform in Multiple Dimensions

When working in multiple dimensions, the Fourier transform can be applied independently to each dimension. The result of doing so, however, can be represented very compactly in formulae that look just like those we've already developed, except that products are replaced with dot products, and some factors of d, the number of dimensions, are inserted. Suppose $\mathbf{x}$ is a d-dimensional vector, and $f(\mathbf{x})$ is a scalar function of $\mathbf{x}$. Then:

- **On an infinite domain:**

$$\tilde{f}(\mathbf{k}) = \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{x}\ f(\mathbf{x})e^{-\imath\mathbf{k}\cdot\mathbf{x}} \tag{6.93}$$

$$f(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \tilde{f}(\mathbf{k})e^{\imath\mathbf{k}\cdot\mathbf{x}} \tag{6.94}$$

with the corresponding sum of exponentials

$$\int d^d\mathbf{x}e^{\imath\mathbf{k}\cdot\mathbf{x}} = (2\pi)^d\delta^d(\mathbf{k}) \tag{6.95}$$

Here, $\delta^d(\mathbf{k})$ is the d-dimensional Dirac delta function, defined by $\int d^d\mathbf{k}\ \delta^d(\mathbf{k}) = 1$, $\delta^d(\mathbf{k}) = 0$ for $\mathbf{k} \neq 0$ (or alternatively, defined as the product of one-dimensional delta functions, one for each dimension of $\mathbf{k}$).

Note that $\left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{x}\ e^{-\imath\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x})$ can be rewritten as

$$\left[\frac{1}{\sqrt{2\pi}}\int dx_0 e^{-\imath k_0 x_0}\right]\left[\frac{1}{\sqrt{2\pi}}\int dx_1 e^{-\imath k_1 x_1}\right][\ldots]\left[\frac{1}{\sqrt{2\pi}}\int dx_{d-1} e^{-\imath k_{d-1} x_{d-1}}\right] f(\mathbf{x}) \tag{6.96}$$

This is why the independent application of the Fourier transform along each dimension can be written so compactly in the form of Eq. 6.93.

- **On a finite domain:**

$$f_{\mathbf{k}} = \frac{1}{T^{d/2}} \int d^d\mathbf{x}\ f(\mathbf{x})e^{-\imath\frac{2\pi\mathbf{k}\cdot\mathbf{x}}{T}} \tag{6.97}$$

$$f(\mathbf{x}) = \frac{1}{T^{d/2}} \sum_{\mathbf{k}} f_{\mathbf{k}} e^{\imath\frac{2\pi\mathbf{k}\cdot\mathbf{x}}{T}} \tag{6.98}$$

Here, the integral is over the hypercube of length T on a side centered at the origin, and the $\sum_{\mathbf{k}}$ is a sum in which each component of $\mathbf{k}$ goes from $-\infty$ to ∞. The corresponding sum of exponentials is

$$\int d^d\mathbf{x}e^{\imath\frac{2\pi\mathbf{x}\cdot\mathbf{k}}{T}} = T^d\delta_{\mathbf{k}0} \tag{6.99}$$

where the delta function $\delta_{\mathbf{k}0}$ is zero for $\mathbf{k} \neq 0$ and 1 for $\mathbf{k} = 0$.

- **For a discrete vector:** Suppose we discretize $f(\mathbf{x})$ by sampling $\mathbf{x}$ on a grid of length N on a side, where $\mathbf{x}$ has d dimensions. We can represent sample points as $\mathbf{x}_{\mathbf{j}}$ where $\mathbf{j}$ is a d-dimensional set of integers, telling the location on the grid of the sample point; each component of $\mathbf{j}$ runs from 0 to $N-1$. Thus we can represent the discretized f as a vector with a multi-dimensional index: $v_{\mathbf{j}} = f(\mathbf{x}_{\mathbf{j}})$. Then the Fourier transform is

$$\tilde{v}_{\mathbf{j}} = \frac{1}{N^{d/2}} \sum_{\mathbf{k}} e^{-i\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{N}} v_{\mathbf{k}} \tag{6.100}$$

$$v_{\mathbf{k}} = \frac{1}{N^{d/2}} \sum_{\mathbf{j}} e^{i\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{N}} \tilde{v}_{\mathbf{j}} \tag{6.101}$$

The corresponding sum of exponentials is

$$\sum_{\mathbf{k}} e^{i\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{N}} = \begin{cases} N^d & \text{if every component of } \mathbf{j} \text{ is an integral multiple of N} \\ 0 & \text{otherwise} \end{cases} \tag{6.102}$$

Problem 6.15 *Prove the convolution theorem on an infinite domain in d dimensions: the Fourier transform of* $c(\mathbf{x}) = \int d^dy\, f(\mathbf{x}-\mathbf{y})g(\mathbf{y})$ *is* $\tilde{c}(\mathbf{k}) = (2\pi)^{d/2}\tilde{f}(\mathbf{k})\tilde{g}(\mathbf{k})$.

Problem 6.16 *Let's return to our toy neural activity problem. Consider our equation for the activity in a network of neurons,* $\tau\frac{d\mathbf{b}}{dt} = -(\mathbf{1}-\mathbf{B})\mathbf{b}+\mathbf{h}$. *Let's make the transition from discrete vectors to continuous variables: in place of* b_i *we'll have* $b(x)$, *in place of* B_{ij} *we'll have* $B(x,y)$, *etc. (Recall: we can get from the continuous equation to the discrete one by discretely sampling the continuous variables* x *at points* x_i *and* y *at points* y_i*; then we befine* $b_i = b(x_i)$, $B_{ij} = B(x_i, y_j)$, *etc. Here we are just reversing the process, going from a discrete equation to a continuous one.) For simplicity we'll work in one dimension, i.e.* x *and* y *are one-dimensional variables (although the mathematics is identical in arbitrary dimensions except that factors of* $\sqrt{2\pi}$ *become* $(2\pi)^{d/2}$ *and products* kx *become dot products* $\mathbf{k}\cdot\mathbf{x}$*). Let's assume that interactions are translation invariant, which means that* $B(x,y) = B(x-y)$ *– the interactions only depend on the separation between two points, and don't otherwise depend on the particular points involved. Let's work on an infinite domain. So the continuous version of the equation becomes*

$$\tau\frac{d}{dt}b(x) = -b(x) + \int dy\, B(x-y)b(y) + h(x) \tag{6.103}$$

Take the Fourier transform of this equation (you can use the convolution theorem) and show that the result is

$$\begin{aligned}\tau\frac{d}{dt}\tilde{b}(k) &= -\tilde{b}(k) + \sqrt{2\pi}\tilde{B}(k)\tilde{b}(k) + \tilde{h}(k) &(6.104)\\ &= -\left(1-\sqrt{2\pi}\tilde{B}(k)\right)\tilde{b}(k) + \tilde{h}(k) &(6.105)\end{aligned}$$

or

$$\frac{d}{dt}\tilde{b}(k) \equiv \lambda(k)\tilde{b}(k) + \tilde{h}(k)/\tau \tag{6.106}$$

where we have defined

$$\lambda(k) \equiv -\left(1-\sqrt{2\pi}\tilde{B}(k)\right)/\tau \tag{6.107}$$

The equation has been diagonalized – the development of $\tilde{b}(k)$ *is independent of that of all other* $\tilde{b}(l)$ *for* $l \neq k$, *and is governed by a simple one-dimensional equation.*

Let's assume that $h(x)$ *is time independent. Then recall from Section 0.5 that the solution to Eq. 6.106 is*

$$\begin{aligned}\tilde{b}(k,t) &= \tilde{b}(k,0)e^{t\lambda(k)} - \frac{\tilde{h}(k)}{\tau\lambda(k)}\left(1-e^{t\lambda(k)}\right), \qquad \lambda(k)\neq 0 &(6.108)\\ \tilde{b}(k,t) &= \tilde{b}(k,0) + \tilde{h}(k)t/\tau, \qquad \lambda(k) = 0 &(6.109)\end{aligned}$$

The solution in real space is then

$$b(x,t) = \int dk\, \tilde{b}(k,t)\frac{e^{ikx}}{\sqrt{2\pi}} \tag{6.110}$$

The functions $\frac{e^{ikx}}{\sqrt{2\pi}}$ *are the normalized*[13] *eigenfunctions of the operator "*$-(\mathbf{1}-\mathbf{B})/\tau$*", which has become the integral operator* $-\frac{1}{\tau}\int dy\,[\delta(x-y)-B(x-y)]\cdot$. *Show that* $\frac{e^{ikx}}{\sqrt{2\pi}}$ *is indeed an eigenfunction,*

[13]The normalization is $\int dx\, \frac{e^{-ikx}}{\sqrt{2\pi}}\frac{e^{ik'x}}{\sqrt{2\pi}} = \delta(k-k')$; this is the continuous analogue of the discrete eigenvector normalization $\mathbf{e}_k \cdot \mathbf{e}_{k'} = \sum_i (\mathbf{e}^*_k)_i(\mathbf{e}_{k'})_i = \delta_{kk'}$.

with eigenvalue $\lambda(k)$ (note, when we say that $f(x)$ is an eigenfunction of some integral operator $\int dy\, K(x-y)\cdot$, with eigenvalue λ, we mean that $\int dy\, K(x-y)f(y) = \lambda f(x)$, analogously to the vector expression $\sum_j K_{ij}v_j = \lambda v_i$, which is the components version of $\mathbf{Kv} = \lambda\mathbf{v}$).

Suppose we write $\mathbf{e}_k$ for the k^{th} eigenfunction, which has components (values at x) $e^{ikx}/\sqrt{2\pi}$, analogously to writing $\mathbf{e}_j$ for the j^{th} eigenvector with components $(\mathbf{e}_j)_i$. Similarly, let's write $\mathbf{b}(t)$ for the function with components $b(x,t)$. We'll write $b_k(t)$ for $b(k,t)$, which is the component of $\mathbf{b}(t)$ along the k^{th} eigenfunction. And we'll similarly write h_k for $h(k)$, and write λ_k for $\lambda(k)$, which is the eigenvalue corresponding to the k^{th} eigenfunction. Assuming $\lambda_k \neq 0$ for all k, show that we can write the solution as

$$\mathbf{b}(t) = \int dk\, \mathbf{e}_k \left[b_k(0)e^{t\lambda_k} - \frac{h_k}{\tau\lambda_k}\left(1 - e^{t\lambda_k}\right)\right] \tag{6.111}$$

Except for the use of $\int dk$ rather than $\sum_i$, this is exactly the form of Eq. 3.56, the solution to an inhomogeneous equation in the discrete case.

That's all you need to write down for the problem, but let's think about the result: what does this tell us? First of all, the eigenfunctions, in this case of translation-invariant connectivity, are the Fourier modes: oscillations e^{ikx} for all k's. The size of the k^{th} eigenfunction in the solution $b(x,t)$ is determined by its coefficient $\tilde{b}(k,t)$, which in turn is determined by Eqs. 6.108-6.109. (Of course, $b(x,t)$ is real, which means that $\tilde{b}(-k,t) = \tilde{b}^*(k,t)$, which means that these solutions come in pairs that add to real cosine and sine oscillations: $\tilde{b}(k,t)e^{ikx} + \tilde{b}^*(k,t)e^{-ikx} = 2\mathrm{RE}\,\tilde{b}(k,t)\cos(kx) - 2\mathrm{IM}\,\tilde{b}(k,t)\sin(kx)$). So the independently growing solutions are mixtures of spatial sine and cosine waves of activity with spatial period $2\pi/k$.

Second of all, if $\lambda(k) > 0$ for some k – meaning $\tilde{B}(k) > 1/\sqrt{2\pi}$ – then the oscillation with the corresponding k grows exponentially without bound. We say then that the dynamics are unstable. If $\lambda(k) = 0$ for some k, the oscillation with the corresponding k grows linearly without bound. If some $\lambda(k) = 0$ and no $\lambda(k) > 0$, we say the dynamics are "marginally stable".

If $\lambda(k) < 0$ for all k – meaning $\tilde{B}(k) < 1/\sqrt{2\pi}$ for all k – then for every k, the amplitude $b(k,t)$ evolves exponentially to its fixed point: $\lim_{t\to\infty} \tilde{b}(k,t) = -\frac{\tilde{h}(k)}{\tau\lambda(k)}$. In this case we say the dynamics are stable. Each mode is determined by its corresponding input $\tilde{h}(k)$, amplified by the factor $\frac{1}{\tau\lambda(k)}$. Thus, modes with the largest $\lambda(k)$, meaning the ones with the largest $\tilde{B}(k)$ (the ones with $\tilde{B}(k)$ closest to $1/\sqrt{2\pi}$) will be most amplified.

Intuitively, the size of $\tilde{B}(k)$ tells the size, in the interaction function $B(x-y)$, of an oscillation of wavelength $2\pi/k$. Such an oscillation represents excitation at some distances (where $B(x-y)$ is positive) and inhibition at other distances (where $B(x-y)$ is negative). This leads to activity patterns that oscillate with a similar spatial period between being excited and being inhibited. The largest such oscillation – the k for which $\tilde{B}(k)$ is maximal – will grow the fastest. If this largest $\tilde{B}(k)$ is small enough that the dynamics are stable, it is this mode that will be most amplified relative to its input $\tilde{h}(k)$.

Exercise 6.5 *Show that if we work in d dimensions – so that x, y, k, etc. become d-dimensional vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{k}$, etc. – Eq. 6.106 remains identical except that the factor of $\sqrt{2\pi}$ in the expression for $\lambda(\mathbf{k})$ is replaced by $(2\pi)^{d/2}$. More generally, convince yourself that all of Problem 6.16 is identical in d dimensions if (1) $\sqrt{2\pi}$ is replaced everywhere with $(2\pi)^{d/2}$; (2) factors like kx are replaced by dot products, $\mathbf{k}\cdot\mathbf{x}$; and (3) integrals are interpreted as d-dimensional integrals and delta functions are interpreted as d-dimensional delta functions.*

Exercise 6.6 *Let's also return to our toy problem involving development of ocular dominance. We begin with the equation $\tau\frac{d}{dt}\mathbf{w} = \mathbf{Cw}$, for the development of the inputs w_i to a single postsynaptic* *cell. We assume that* $\mathbf{w}$ *includes two types of inputs – w^L (left eye) and w^R (right eye). The correlations are of four types: C^{IJ} for $I, J \in \{L, R\}$ represents the correlation between inputs of type I and those of type J Thus our equation becomes*

$$\tau \frac{d}{dt}\begin{pmatrix} \mathbf{w}^L \\ \mathbf{w}^R \end{pmatrix} = \begin{pmatrix} C^{LL} & C^{LR} \\ C^{RL} & C^{RR} \end{pmatrix}\begin{pmatrix} \mathbf{w}^L \\ \mathbf{w}^R \end{pmatrix} \tag{6.112}$$

We take the continuum limit, working on an infinite domain, and let's work in two dimensions; so we have $w^L(\mathbf{x})$ in place of w_i^L and $C^{LL}(\mathbf{x}, \mathbf{y})$ in place of C_{ij}^{LL}, etc., where $\mathbf{x}$ and $\mathbf{y}$ are two-dimensional retinotopic positions of the inputs to the cell being studied. We assume translation-invariance, so that $C^{LL}(\mathbf{x}, \mathbf{y}) = C^{LL}(\mathbf{x} - \mathbf{y})$. Thus we arrive at the equations

$$\tau \frac{d}{dt}\begin{pmatrix} w^L(\mathbf{x}) \\ w^R(\mathbf{x}) \end{pmatrix} = \int d^2\mathbf{y} \begin{pmatrix} C^{LL}(\mathbf{x}-\mathbf{y}) & C^{LR}(\mathbf{x}-\mathbf{y}) \\ C^{RL}(\mathbf{x}-\mathbf{y}) & C^{RR}(\mathbf{x}-\mathbf{y}) \end{pmatrix}\begin{pmatrix} w^L(\mathbf{y}) \\ w^R(\mathbf{y}) \end{pmatrix} \tag{6.113}$$

Assume that the two eyes are identical – or more formally that the equations are symmetric under interchange of the two eyes, that is, unchanged if we replace R with L and vice versa. This implies that $C^{LL} = C^{RR}$, $C^{LR} = C^{RL}$; so let's give these new names, $C^{Same} = C^{LL} = C^{RR}$, $C^{Opp} = C^{LR} = C^{RL}$ (where "Same" means same-eye and "Opp" means opposite-eye). Show that in this case the equation can be "diagonalized" by going to the coordinates $w^S(\mathbf{x}) = (w^L(\mathbf{x}) + w^R(\mathbf{x}))/\sqrt{2}$, $w^D(\mathbf{x}) = w^L(\mathbf{x}) - w^R(\mathbf{x})/\sqrt{2}$, so that the resulting equation is

$$\tau \frac{d}{dt}\begin{pmatrix} w^S(\mathbf{x}) \\ w^D(\mathbf{x}) \end{pmatrix} = \int d^2\mathbf{y} \begin{pmatrix} C^{Same}(\mathbf{x}-\mathbf{y}) + C^{Opp}(\mathbf{x}-\mathbf{y}) & 0 \\ 0 & C^{Same}(\mathbf{x}-\mathbf{y}) - C^{Opp}(\mathbf{x}-\mathbf{y}) \end{pmatrix}\begin{pmatrix} w^S(\mathbf{y}) \\ w^D(\mathbf{y}) \end{pmatrix} \tag{6.114}$$

Thus, just as in the cases we studied before, the sum of the two eyes' inputs grows independently of the difference between the two eyes' inputs (this follows quite generally from assuming a symmetry under interchange of the two eyes). We have "diagonalized" the left/right part of the equation by going to sum and difference coordinates; now, as should be becoming familiar, we can diagonalize the spatial part of the equation by taking a Fourier transform. Show (you can use the 2-D convolution theorem) that the result is

$$\tau \frac{d}{dt}\begin{pmatrix} \widetilde{w^S}(\mathbf{k}) \\ \widetilde{w^D}(\mathbf{k}) \end{pmatrix} = (2\pi)\begin{pmatrix} \widetilde{C^{Same}}(\mathbf{k}) + \widetilde{C^{Opp}}(\mathbf{k}) & 0 \\ 0 & \widetilde{C^{Same}}(\mathbf{k}) - \widetilde{C^{Opp}}(\mathbf{k}) \end{pmatrix}\begin{pmatrix} \widetilde{w^S}(\mathbf{k}) \\ \widetilde{w^D}(\mathbf{k}) \end{pmatrix} \tag{6.115}$$

Let's define $\lambda^D(\mathbf{k}) = (2\pi)\left[\widetilde{C^{Same}}(\mathbf{k}) - \widetilde{C^{Opp}}(\mathbf{k})\right]$, $\lambda^S(k) = (2\pi)\left[\widetilde{C^{Same}}(\mathbf{k}) + \widetilde{C^{Opp}}(\mathbf{k})\right]$. Then the solution is

$$\begin{pmatrix} \widetilde{w^S}(k,t) \\ \widetilde{w^D}(k,t) \end{pmatrix} = \begin{pmatrix} e^{t\lambda^S(k)} & 0 \\ 0 & e^{t\lambda^D(k)} \end{pmatrix}\begin{pmatrix} \widetilde{w^S}(k,0) \\ \widetilde{w^D}(k,0) \end{pmatrix} \tag{6.116}$$

Of course back in real space we have the solutions

$$w^S(\mathbf{x},t) = \int d^2\mathbf{k}\, \widetilde{w^S}(\mathbf{k},t)\frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{2\pi} \tag{6.117}$$

$$w^D(\mathbf{x},t) = \int d^2\mathbf{k}\, \widetilde{w^D}(\mathbf{k},t)\frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{2\pi} \tag{6.118}$$

So, what does this mean? To develop ocular dominance, we want the inputs to our postsynaptic cell to become all from one eye – all left or all right. That means that throughout the input space, $w^D(x)$ should have the same sign – positive everywhere if the left eye dominates, negative everywhere *if the right eye dominates. That in turn means that, for ocular dominance to develop, the* $\mathbf{k} = 0$ *mode should dominate* w^D *(that is, it should be the fastest-growing eigenfunction of* w^D*) – for* $k \neq 0$, w^D *oscillates in sign with frequency* $2\pi/|k|$*, meaning that it oscillates between regions of the receptive field dominated by the left eye and regions dominated by the right eye.*[14] *So, for ocular dominance to develop, we need* $\lambda^D(\mathbf{k})$ *to be peaked at* $\mathbf{k} = 0$ *so that the* $\mathbf{k} = 0$ *mode is the fastest-growing mode of* w^D*. One scenario under which this will happen is if* $C^{Same}(\mathbf{x}) - C^{Opp}(\mathbf{x}) \geq 0$ *for all* $\mathbf{x}$ *–that is, if at all separations* $\mathbf{x}$*, an input is better correlated with another input of its own eye than it is with another input of the opposite eye. This seems reasonable – we might expect correlations to fall monotonically with distance* $|x|$*, but to always be greater within than between eyes until both within- and between-eye correlations go to zero. That will cause* w^D *to develop a pattern consistent with ocular dominance.*

Additionally, for ocular dominance to develop, we would like w^D *to grow faster than* w^S*. If* w^S *grows faster than* w^D*, both eye's inputs will be growing (recall that* $w^L \propto w^S + w^D$*,* $w^R \propto w^S - w^D$*), although the difference between them* w^D *will also be growing; instead we would like one eye's inputs to grow and the other eye's inputs to shrink, which requires that* w^D *grow faster than* w^S*. Suppose we restrict attention to the* $k = 0$ *modes. Then* $\lambda^S(0) = (2\pi) \int d^2\mathbf{x} \left[C^{Same}(\mathbf{x}) + C^{Opp}(\mathbf{x})\right]$*, while* $\lambda^D(0) = (2\pi) \int d^2\mathbf{x} \left[C^{Same}(\mathbf{x}) - C^{Opp}(\mathbf{x})\right]$*. So for the difference to grow faster than the sum, we need* $\int d^2\mathbf{x}\, C^{Opp}(\mathbf{x}) < 0$ *– that is, the two eyes should be anticorrelated (*C^{Opp} *should be negative over a significant range). However, it is common to consider that there are additional constraints on* w^S *that limit its growth – for example, there is some upper limit on total synaptic strength; and we sometimes assume that the sum* $\int d^2\mathbf{x}\, w^S(\mathbf{x})$ *is fixed in order to capture the biological idea that there is a competition between the eyes, so that when one eye's synapses grow the other eye's synapses must shrink. Such additional constraints can suppress the growth of the sum and allow ocular dominance to develop even with positive correlations between the eyes (note, we expect vision to cause the two eyes to be positively correlated, since they tend to see the same scenes; but in many species ocular dominance develops before the onset of vision).*

One can make the model a bit more realistic in several ways. In the version above, the post-synaptic cell receives input from a set of inputs stretching to infinity in the 2-D plane. One can instead make the inputs localized by multiplying the integral in Eq. 6.113 by an "arbor function", $A(\mathbf{x})$*, peaked at* $\mathbf{x} = 0$ *and falling off to zero over the range of allowed connectivity. Adding the arbor function breaks the translation-invariance of the equations – changing position leads to a different interaction because there is a different value of* $A(\mathbf{x})$ *– so the Fourier modes no longer spatially diagonalize the equation. But what we have learned in the "infinite arbor" case, above, remains informative – eigenfunctions of* w^D *that have a characteristic wavelength of oscillation* $\mathbf{k}$ *grow at a rate roughly proportional to* $\lambda^D(k)$*, although they are localized eigenfunctions rather than Fourier modes. One can also restrict weights to stay within some allowed range,* $0 \leq w^L(\mathbf{x}) \leq w_{\max} A(\mathbf{x})$ *and similarily for* $w^R(\mathbf{x})$*. This makes the equation nonlinear, but it remains linear when all the weights are far from these boundary values, so the linear analysis allows us to decide which modes grow the fastest early on, before weights reach the limiting values, and these fastest-growing modes end up largely determining the structure of the final receptive field that develops.*

6.8 Using the Fourier Transform: Solving the Diffusion Equation

A nice example of the use of the Fourier transform is also a classic problem that everyone who uses mathematics should know how to do: solving the diffusion equation. We're going to just go

[14] Recall that, because w^D is real, $\widetilde{w^D}(\mathbf{k}, t) = \widetilde{w^D}*(-\mathbf{k}, t)$, so we can obtain real functions by combining $\widetilde{w^D}(\mathbf{k}, t)e^{i\mathbf{k}\cdot\mathbf{x}} + w^D*(\mathbf{k}, t)e^{-i\mathbf{k}\cdot\mathbf{x}} = 2\text{RE}\ \widetilde{w^D}(\mathbf{k}, t)\cos(\mathbf{k}\cdot\mathbf{x}) - 2\text{IM}\ \widetilde{w^D}(\mathbf{k}, t)\sin(\mathbf{k}\cdot\mathbf{x})$.

through the math, but it won't do you any good to just read it – you need to go through it with a pencil and paper, replicating each of the steps to your own satisfaction.

The diffusion equation, or heat equation, describes the diffusion in time of the concentration of a substance, or the spread in time of the temperature from a heat source. The equation is[15]

$$\frac{\partial c(\mathbf{r},t)}{\partial t} = D\nabla^2 c(\mathbf{r},t) \tag{6.122}$$

Here, $c(\mathbf{r},t)$ is the concentration, or the temperature, at position $\mathbf{r}$ and time t; D is the diffusion constant, or the "thermal diffusivity" when heat is considered. The operator ∇^2 is the sum of the second spatial derivatives in however many dimensions one is considering; for example, in three dimensions, $\mathbf{r} = (x,y,z)^{\mathrm{T}}$, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

We assume the substance is diffusing (or the heat is spreading) in an infinite domain, and that we are given the initial state $c(\mathbf{r},0)$, and we want to compute how the concentration changes with time thereafter. To do this, we make use of the fact that the Fourier transform diagonalizes the differential operator ∇^2 in Eq. 6.122. We express c in terms of its spatial Fourier transform:

$$c(\mathbf{r},t) = \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \tilde{c}(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.123}$$

Applying $\frac{\partial}{\partial t}$ to the right side of Eq. 6.123 just converts $\tilde{c}(\mathbf{k},t)$ to $\frac{\partial \tilde{c}(\mathbf{k},t)}{\partial t}$, since $\tilde{c}(\mathbf{k},t)$ is the only thing on the right side that depends on t. Applying ∇^2 to the right side of Eq. 6.123 just converts $e^{i\mathbf{k}\cdot\mathbf{r}}$ to $\nabla^2 e^{i\mathbf{k}\cdot\mathbf{r}} = -|\mathbf{k}|^2 e^{i\mathbf{k}\cdot\mathbf{r}}$, since $e^{i\mathbf{k}\cdot\mathbf{r}}$ is the only thing on the right side that depends on spatial position $\mathbf{r}$. Thus, Eq. 6.122 applied to Eq. 6.123 yields

$$\left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \frac{\partial \tilde{c}(\mathbf{k},t)}{\partial t} e^{i\mathbf{k}\cdot\mathbf{r}} = D\left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ (-|\mathbf{k}|^2)\tilde{c}(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6.124}$$

These two integral equations can only be equal if the coefficients of each exponential $e^{i\mathbf{k}\cdot\mathbf{r}}$ are identical,[16] so we obtain the equation

$$\frac{\partial \tilde{c}(\mathbf{k},t)}{\partial t} = -D|\mathbf{k}|^2\tilde{c}(\mathbf{k},t) \tag{6.125}$$

[15]For those of you who remember the divergence and the gradient: Eq. 6.122 typically arises as follows. Consider the case of diffusion. First, there is a current that follows the gradient of the concentration, that is, substances flow from regions of higher concentration to regions of lower concentration. We express this as

$$\mathbf{j} = -k_1\nabla c \tag{6.119}$$

where $\mathbf{j}$ is the current, k_1 is a constant, ∇ is the gradient operator, *e.g.* in 3 dimensions $\nabla = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}$, and $\hat{x}$, $\hat{y}$, and $\hat{z}$ are the unit vectors in the x, y, and z directions respectively. Second, there is a conservation of substance, so that the concentration at a point changes in time according to the net flow of substance into or out of the point, which is captured by the divergence of the current, $\nabla\cdot\mathbf{j} = \frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z}$. The corresponding equation is

$$\frac{\partial c}{\partial t} = -k_2\nabla\cdot\mathbf{j} \tag{6.120}$$

Putting these two equations together yields

$$\frac{\partial c}{\partial t} = D\nabla^2 c \tag{6.121}$$

where $D = k_1k_2$.

[16]To see this, one can apply $\int d^d\mathbf{r}\ e^{-i\mathbf{k}'\cdot\mathbf{r}}$ to both sides of Eq. 6.124; the result of the $\mathbf{r}$ integral is $(2\pi)^d\delta^d(k-k')$, which allows us to do the $\mathbf{k}$ integral, obtaining Eq. 6.125.

One ultimately learns to just read off an equation like Eq. 6.125 from an equation like Eq. 6.122: each application of a spatial derivative pulls down a factor if $\imath\mathbf{k}$ in Fourier space, so the ∇^2 in real space becomes $-|\mathbf{k}|^2$ in Fourier space.

Equation 6.125 is just an ordinary differential equation for $\tilde{c}(\mathbf{k}, t)$ – the Fourier transform has indeed diagonalized the equation, rendering the development of each Fourier mode $\tilde{c}(\mathbf{k}, t)$ independent of that of all the others – and so we can write down the solution in Fourier space:

$$\tilde{c}(\mathbf{k}, t) = \tilde{c}(\mathbf{k}, 0)e^{-D|\mathbf{k}|^2 t} \tag{6.126}$$

Finally, we need to work out the solution in real space; to do this, we'll need to recall that the initial condition in Fourier space, $c(\mathbf{k}, 0)$, can be expressed as the Fourier transform of the initial condition in real space, $c(\mathbf{r}', 0)$:

$$\begin{aligned}
c(\mathbf{r}, t) &= \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \tilde{c}(\mathbf{k}, t)e^{\imath\mathbf{k}\cdot\mathbf{r}} && (6.127)\\
&= \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \tilde{c}(\mathbf{k}, 0)e^{-D|\mathbf{k}|^2 t}e^{\imath\mathbf{k}\cdot\mathbf{r}} && (6.128)\\
&= \left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{k}\ \left[\left(\frac{1}{2\pi}\right)^{d/2} \int d^d\mathbf{r}'\ c(\mathbf{r}', 0)e^{-\imath\mathbf{k}\cdot\mathbf{r}'}\right] e^{-D|\mathbf{k}|^2 t}e^{\imath\mathbf{k}\cdot\mathbf{r}} && (6.129)\\
&= \left(\frac{1}{2\pi}\right)^{d} \int d^d\mathbf{r}'\ \left[\int d^d\mathbf{k}\ e^{\imath\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-D|\mathbf{k}|^2 t}\right] c(\mathbf{r}', 0) && (6.130)\\
&= \int d^d\mathbf{r}'\ G(\mathbf{r} - \mathbf{r}', t)c(\mathbf{r}', 0) && (6.131)
\end{aligned}$$

where the *Green's function*, $G(\mathbf{r} - \mathbf{r}', t)$, is given by

$$G(\mathbf{r} - \mathbf{r}', t) = \left(\frac{1}{2\pi}\right)^{d} \int d^d\mathbf{k}\ e^{\imath\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-D|\mathbf{k}|^2 t} \tag{6.132}$$

That is, the concentration as a function of time is determined as the convolution of the Green's function with the initial distribution of concentration. Intuitively, the Green's function tells the distribution over time of an initial localized pulse of concentration: if the initial condition were a single pulse at position $\mathbf{r}'$, $c(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}')$, then the solution is $c(\mathbf{r}, t) = G(\mathbf{r} - \mathbf{r}', t)$. The actual initial condition is a weighted average of delta pulses, each weighted by the value of the initial concentration at the given point: $c(\mathbf{r}, 0) = \int d^d\mathbf{r}'\ \delta(\mathbf{r} - \mathbf{r}')c(\mathbf{r}', 0)$; and each of these pulses develops in time independently according to the Green's function, so that the final solution is a weighted sum over each pulse of its independent outcome, $G(\mathbf{r} - \mathbf{r}', t)$, weighted by the size of the pulse, $c(\mathbf{r}', 0)$: $c(\mathbf{r}, t) = \int d^d\mathbf{r}'\ G(\mathbf{r} - \mathbf{r}', t)c(\mathbf{r}', 0)$.

So the last step remaining is to solve for the Green's function $G(\mathbf{r}, t)$ in real space. But the expression for G is just the inverse Fourier transform of a two-dimensional spherical Gaussian, where by spherical we mean the variance is the same in all dimensions. We can solve this by letting

$p = \sqrt{Dt}\mathbf{k}$ and completing the square:

$$G(\mathbf{r},t) = \left(\frac{1}{2\pi}\right)^d \int d^d\mathbf{k}\; e^{i\mathbf{k}\cdot\mathbf{r}} e^{-D|\mathbf{k}|^2 t} \tag{6.133}$$

$$= \left(\frac{1}{2\pi\sqrt{Dt}}\right)^d \int d^d\mathbf{p}\; e^{i\mathbf{p}\cdot\mathbf{r}/\sqrt{Dt}} e^{-|\mathbf{p}|^2} \tag{6.134}$$

$$= \left(\frac{1}{2\pi\sqrt{Dt}}\right)^d \int d^d\mathbf{p}\; e^{-(\mathbf{p}-i\mathbf{r}/2\sqrt{Dt})^2} e^{-|\mathbf{r}|^2/4Dt} \tag{6.135}$$

$$= \left(\frac{1}{2\pi\sqrt{Dt}}\right)^d e^{-|\mathbf{r}|^2/4Dt} \int d^d\mathbf{p}\; e^{-(\mathbf{p}-i\mathbf{r}/2\sqrt{Dt})^2} \tag{6.136}$$

Finally to perform the last integral, we can change variables to $\mathbf{q} = \mathbf{p} - i\mathbf{r}/2\sqrt{Dt}$, so the integral becomes an integral of $e^{-|\mathbf{q}|^2}$ where the limits of integration along the i^{th} component of $\mathbf{q}$ go from $-\infty - ir_i/2\sqrt{Dt}$ to $\infty - ir_i/2\sqrt{Dt}$; here we have to know that, because the Gaussian $e^{-|\mathbf{q}|^2}$ is an analytic function, $\int_{-\infty}^{\infty} dq\, e^{-\mathbf{q}^2} = \int_{-\infty-ic}^{\infty-ic} dq\, e^{-\mathbf{q}^2}$ for any c; so the integral is simply $\int_{-\infty}^{\infty} d^d\mathbf{q}\, e^{-\mathbf{q}^2} = \pi^{d/2}$, yielding

$$G(\mathbf{r},t) = \left(\frac{1}{4\pi Dt}\right)^{d/2} e^{-|\mathbf{r}|^2/4Dt} \tag{6.137}$$

As $t \to 0$, this goes to a delta function, $G(\mathbf{r},0) = \delta^d(\mathbf{r})$, as it should since $G(\mathbf{r},t)$ represents the response at time t to a delta-pulse at time 0 (see Eq. 6.140 and surrounding text in the Appendix as to why this gives a delta function). For finite t, an initial delta peak of substance spreads as a Gaussian, with width $\sigma = \sqrt{2Dt}$ growing, and peak height shrinking, as the square-root of the time. For $t \to \infty$, the Green's function goes to zero everywhere: an initial delta pulse of substance diffuses away to infinity and nothing is left.

6.9 Appendix: Delta Functions and the Sums over Complex Exponentials That Realize Them

6.9.1 The Dirac delta function

In Chapter 2, we introduced the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1 & i-j=0 \\ 0 & i-j\neq 0 \end{cases} \tag{6.138}$$

We can abstract two additional key features of the Kronecker delta:

- It sums to 1 for any sum that passes through $i-j=0$: $\sum_{i=j-n}^{j+m} \delta_{ij} = 1$ for any nonnegative integers n, m.
- In any sum that passes through $i-j=0$, it pulls out one component of a vector: $\sum_{i=j-n}^{j+m} \delta_{ij} v_j = v_i$ for any nonnegative integers n, m.

When we deal with continuous functions rather than vectors, it is very convenient to have an analogue of the Kronecker delta. This was realized by the physicist Paul Dirac, and the resulting function is called the Dirac delta function, $\delta(x)$, a function of a continuous variable x. It satisfies the following properties, analogous to those of the Kronecker delta:

- It is zero wherever $x \neq 0$: $\delta(x) = 0$ for $x \neq 0$.

- It integrates to 1 over any region that includes $x = 0$: $\int_{-\epsilon_1}^{\epsilon_2} dx\ \delta(x) = 1$ for any $\epsilon_1 > 0$, $\epsilon_2 > 0$.
- In an integral over any region that includes its argument, it pulls out one value of a function: $\int_{-\epsilon_1}^{\epsilon_2} dy\ f(x-y)\delta(y) = f(x)$ for any $\epsilon_1 > 0$, $\epsilon_2 > 0$.

How can $\delta(x)$ be nonzero only at a single point, $x = 0$, and yet integrate to something finite? Clearly, the value at $x = 0$ must be infinite, and just the "right" infinity to integrate to 1. We can define $\delta(x)$ as a limit of a sequence of functions that each integrate to 1, where the limiting value for any nonzero x is 0. For example:

- Let $f_\gamma(x)$ be a pulse of width γ, height $1/\gamma$ (this will always integrate to 1):

$$f_\gamma(x) = \begin{cases} \frac{1}{\gamma} & \frac{-\gamma}{2} \leq x \leq \frac{\gamma}{2} \\ 0 & \text{otherwise} \end{cases} \tag{6.139}$$

  Then, $\lim_{\gamma\to 0} f_\gamma(x) = \delta(x)$.

- Let $g_\sigma(x)$ be a Gaussian of width σ normalized to integrate to 1:

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{6.140}$$

  Then $\lim_{\sigma\to 0} g_\sigma(x) = \delta(x)$.

With these definitions, the validity of the first two conditions named above – $\delta(x)$ is nonzero whenever $x \neq 0$, and integrates to 1 – should be clear. What about the third property? Consider $\int dy\ \ h(x-y) f_\gamma(y)$ for some function $h(x)$. Since f_γ is constant over a finite region and zero elsewhere, this integral is equal to $\frac{1}{\gamma}\int_{\frac{-\gamma}{2}}^{\frac{\gamma}{2}} dy\ \ h(x-y)$. But as $\gamma \to 0$, the value of $h(x-y)$ becomes constant in the interval of the integral: $h(x-y) = h(x) - h'(x)y + (1/2)h''(x)y^2 + \ldots$, and $|y| \leq \gamma/2$ is going to zero. We can write $h(x-y) = h(x) + O(\gamma)$, where $O(\gamma)$ indicates terms that are linear or higher order in γ. So as $\gamma \to 0$, $\int_{\frac{-\gamma}{2}}^{\frac{\gamma}{2}} dy\ h(x-y) \to \gamma\,(h(x) + O(\gamma))$. Hence, $\lim_{\gamma\to 0}\int dy\ h(x-y) f_\gamma(y) = \lim_{\gamma\to 0} \frac{1}{\gamma}\gamma\,(h(x) + O(\gamma)) = h(x)$. But also $\lim_{\gamma\to 0}\int dy\ h(x-y) f_\gamma(y) = \int dy\ h(x-y)\delta(y)$,[17] giving the third property.

6.9.2 Sums Over Complex Exponentials

Most of our manipulations of Fourier transforms depend on various sums or integrals (which we'll just call 'sums') over complex exponentials that yield Kronecker deltas or Dirac delta functions. All of these have the same geometrical interpretation: if the argument of the complex exponentials in the sum is not always 0 (or an integral multiple of 2π), the sum goes over complex exponentials that point in a set of directions evenly distributed around the unit circle in the complex plane (Fig. 4.1); hence the vectors pointing in different directions cancel out, and the sum gives 0. When the argument of the complex exponential is 0 (or an integral multiple of 2π), then the complex exponential is 1, and so the sum just yields the result of substituting 1 for the complex exponential.

We can begin with the result of Eq. 6.20:

[17] We should be careful in interchanging the limit and the integral, since the integral is itself a limit, $\int dy\ \ldots = \lim_{\Delta y\to 0}\sum \Delta y \ldots$, and one must be sure that the result does not depend on the order of taking the two limits. But to really be careful we should also be careful about defining the delta function, and this leads to a long and perhaps not-so-interesting story ... So just trust me that following one's intuitions here gives the right answer.

Complex Exponential Sum 6.1

$$\int_{-T/2}^{T/2} dt\ e^{\imath\frac{2\pi t(l-k)}{T}} = T\delta_{lk} \quad \textit{for } l \textit{ and } k \textit{ integers} \tag{6.141}$$

This is easily proved by performing the integral:

$$l = k: \qquad \int_{-T/2}^{T/2} dt\ e^{\imath\frac{2\pi t(l-k)}{T}} = \int_{-T/2}^{T/2} dt = T \tag{6.142}$$

$$l \neq k: \qquad \int_{-T/2}^{T/2} dt\ e^{\imath\frac{2\pi t(l-k)}{T}} = \left.\frac{e^{\imath\frac{2\pi t(l-k)}{T}}}{2\pi\imath(l-k)/T}\right|_{-T/2}^{T/2} = \frac{T}{2\pi\imath(l-k)}\left(e^{\imath\pi(l-k)} - e^{-\imath\pi(l-k)}\right) \tag{6.143}$$

$$= \frac{T}{2\pi\imath(l-k)} 2\imath \sin\left[\pi(l-k)\right] = 0 \tag{6.144}$$

(note, $\sin m\pi = 0$ for any integer m).

Now, apply $\frac{1}{T}\sum_{k=-\infty}^{\infty}$ to both sides of Eq. 6.141, to obtain

$$\int_{-T/2}^{T/2} dt\ \frac{1}{T}\sum_{k=-\infty}^{\infty} e^{-\imath\frac{2\pi t(k-l)}{T}} = 1 \quad \text{for } l \text{ and } k \text{ integers} \tag{6.145}$$

We can simplify by setting $p = k - l$, obtaining

$$\int_{-T/2}^{T/2} dt\ \frac{1}{T}\sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}} = 1 \tag{6.146}$$

where p is summed over integral values. Let $f(t) = \frac{1}{T}\sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}}$. Then Eq. 6.146 becomes $\int_{-T/2}^{T/2} dt\ f(t) = 1$. We will show that $f(t) = 0$ for $t \neq nT$ with integral n. These two conditions together imply that, on the region $-T/2 \leq t \leq T/2$, $f(t) = \delta(t)$. Finally it is obvious from the 2π-periodicity of the complex exponential that $f(t)$ is periodic with period T: $f(t+nT) = f(t)$ for any integer n. This gives the final result that $f(t) = \sum_{n=-\infty}^{\infty}\delta(t-nT)$, that is,

Complex Exponential Sum 6.2

$$\frac{1}{T}\sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}} = \sum_{n=-\infty}^{\infty}\delta(t-nT) \tag{6.147}$$

where the p and n sums both extend over integer values.

It remains to show that $f(t) = 0$ for $t \neq nT$ with integral n. Let $s(t) = Tf(t) = \sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}}$. This represents a sum over an infinite sequence of vectors in the complex plane with successive vectors separated by an angle $-2\pi t/T$ (*e.g.* see Fig. 4.1 and draw a subset of this sequence until you can visualize the sequence of vectors going round and round the unit circle of the complex plane). So rotating all the vectors by $-2\pi t/T$ will take each vector into an adjacent vector; this will leave the entire set, and thus the sum, unchanged. Yet the sum is itself a vector in the complex plane, which will also be rotated by $-2\pi t/T$; if this rotation brings the sum back to itself, then either $-2\pi t/T$ must represent an integral number of complete rotations around the complex plane, or the sum must be zero. Mathematically, this rotation is achieved by multiplying by $e^{-2\pi\imath t/T}$:

$e^{-2\pi\imath t/T}s(t) = \sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi t(p+1)}{T}} = \sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}} = s(t)$. From $e^{-2\pi\imath t/T}s(t) = s(t)$, it follows that either $e^{-2\pi\imath t/T} = 1$ or $s(t) = 0$. But $e^{-2\pi\imath t/T} = 1$ if and only if $t = nT$ for some integer n, so if $t \neq nT$, then $s(t) = 0$ and $f(t) = 0$.

Now we can rearrange Eq. 6.147 as follows: let $m_p = 2\pi p/T$, so that we have an infinite sequence of m's separated by $\Delta m = 2\pi/T$. Then we can write

$$\frac{1}{T}\sum_{p=-\infty}^{\infty} e^{-\imath\frac{2\pi tp}{T}} = \frac{1}{T\Delta m}\sum_{p=-\infty}^{\infty} e^{-\imath m_p t}\Delta m = \frac{1}{2\pi}\sum_{p=-\infty}^{\infty} e^{-\imath m_p t}\Delta m \tag{6.148}$$

Combining this with Eq. 6.147 gives

$$\frac{1}{2\pi}\sum_{p=-\infty}^{\infty} e^{-\imath m_p t}\Delta m = \sum_{n=-\infty}^{\infty} \delta\left(t - \frac{2\pi n}{\Delta m}\right) \tag{6.149}$$

Finally, taking the limit as $\Delta m \to 0$ (which is $T \to \infty$), we find

Complex Exponential Sum 6.3

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dm\ e^{-imt} = \delta(t) \tag{6.150}$$

Finally, we establish the simple discrete sum (Eq. 6.59 in the main text):

Complex Exponential Sum 6.4

$$\sum_{k=0}^{N-1} e^{2\pi\imath jk/N} = N\delta_{(j \bmod N)0} \tag{6.151}$$

In this case, the complex exponentials in the sum represent a sequence of unit vectors in the complex plane with an angle $2\pi j/N$ between successive vectors, beginning on the real axis, and ending $2\pi j/N$ before the real axis after having made j revolutions around the unit circle (again, look at Fig. 4.1 to see what this means; draw these vectors for $N = 4$, say, and $j = 1$ or $j = 3$). We use the same argument used above: if we rotate all the vectors by $2\pi j/N$, we will end up with the same set of vectors: each vector will rotate into the next in the sequence, except the last vector will rotate into the initial one. Thus, the sum of the rotated vectors must be the same as the sum of the original vectors. But rotating each vector by $2\pi j/N$ rotates their sum (regarded as a vector in the complex plane) by $2\pi j/N$: if this is not an integral number of complete revolutions, then if the sum were nonzero, it would be changed by the rotation. Since the sum must be zero if $2\pi j/N$ does not represent an integral number of complete rotations, that is, if j is not an integral multiple of N. Thus, the sum is proportional to $\delta_{(j \bmod N)0}$. When j is an integral multiple of N, then the complex exponential is 1, and there are N terms in the sum, so the result is N.

As above, we can make this argument in equations as follows. We call the sum $s(j) = \sum_{k=0}^{N-1} e^{2\pi\imath jk/N}$. Multiply $s(j)$ by $e^{2\pi\imath j/N}$; this represents rotating each vector in the complex plane by $2\pi j/N$. This gives $\sum_{k=0}^{N-1} e^{2\pi\imath j(k+1)/N}$, which we can rewrite as $\sum_{k=1}^{N} e^{2\pi\imath jk/N}$. But $e^{2\pi\imath j(0)/N} = e^{2\pi\imath j(N)/N}$, and therefore we can rewrite this as $\sum_{k=0}^{N-1} e^{2\pi\imath jk/N}$; but this is $s(j)$ again. Thus, we've shown that $e^{2\pi\imath j/N}s(j) = s(j)$; if $e^{2\pi\imath j/N} \neq 1$ (that is, if j is not an integral multiple of N), then this implies that $s(j) = 0$.