Digital Filters and Z Transforms

The contents of this chapter are:

There are more *Mathematica* commands in this chapter than in previous ones, although the details of the commands are not vital to the exposition. We continue using *Mathematica*'s notation: **E** is 2.71828... , I is the square root of minus 1, etc.

■ Digital Filters

In general a filter takes an input **x** and produces an output **y**:

$$x \rightarrow y$$

Usually a filter is specified in terms of some frequency response, say $C[\omega_j]$, which we apply to a time series x_k. As already mentioned, we can apply the effects of the filter in either the time domain or the frequency domain.

In the time domain, we just convolve x_k with the inverse Fourier transform of $C[\omega_j]$. In the frequency domain we multiply $C[\omega_j] = C_j$ and the Fourier transform of x_k. Of course, from the Convolution Theorem we know these two approaches are equivalent:

$$x_k * c_k = \left(\frac{1}{n} \sum_{j=0}^{n-1} X_j \, C_j \, E^{2\pi I j k / n} \right) \frac{1}{\Delta}$$

Recall that we defined a linear filter in the *Convolution* chapter as one in which the output is proportional to the magnitude of the input, and a causal filter is one in which the output is zero for all times earlier than the beginning of the input. The most general linear causal filter can be written as:

$$Y_i = \sum_{m=0}^{M} d_m x_{i-m} + \sum_{n=1}^{N} e_n Y_{i-n}$$

The **M + 1** coefficients d_m and **N** coefficients e_n are fixed and define the filter response. The filter produces each new output value from the current and **M** previous input values, and from its own **N** previous output values.

If **N** is greater than zero, so the output depends on earlier outputs, then the filter includes a "feedback loop."

Some terminology surrounds this filter, useful for looking in an index of some book to find out what you need to know. If the second term in the above equation is zero, i.e. **N** = 0, the filter is said to be "nonrecursive." Yet more terminology: such a filter is also said to have a "finite impulse response (FIR)". The meanings of the terms will become clearer soon.

If the second term in the above equation is non-zero, i.e. **N** > 0, the filter is "recursive", having an "infinite impulse response (IIR)". Another term used with these filters is "autoregressive moving average (ARMA)".

Consider filtering in the frequency domain. The Fourier transform has enfolded all of the values of the time series into *each* value of the transform. Thus, we can think of such filter as being acausal since the outputs at any given time depends on all the inputs at all times. Put another way, real time filtering in general has to be done in the time domain; we cannot take the Fourier transform of the time series until we know all the values.

The filter's frequency response can be shown to be:

$$C[\omega_j] = \sum_{m=0}^{M} d_m E^{-I m \omega_j \Delta} \Bigg/ \left(1 - \sum_{n=1}^{N} e_n E^{-I n \omega_j \Delta} \right)$$

Thus we can find the frequency response once we know d_m and e_n. However, the inverse problem of finding d_m and e_n from a desired frequency response is not so easy, and entire books have been devoted to this topic. In the next sections we discuss one such technique.

■ The Z Transform

We begin by writing the discrete Fourier transform of a time series in yet another way:

$$\mathbf{F}[\omega_j] \;=\; \left(\sum_{k=0}^{n-1} \mathbf{f}_k \, \mathbf{E}^{-I\,\omega_j\,k\,\Delta} \right) \Delta$$

Define the complex variable:

$$\mathbf{z} \;\equiv\; \mathbf{E}^{-I\,\omega_j\,\Delta}$$

Then the Fourier transform becomes:

$$\mathbf{F}[\omega_j] \;=\; \left(\sum_{k=0}^{n-1} \mathbf{f}_k \, \mathbf{z}^k \right) \Delta$$

Thus the Fourier transform becomes just a polynomial in **z**; we call this the "Z transform" of f_k. We shall often write this as **F[z]**. For the remainder of this chapter, we shall also revert to our usual convention that Δ is one.

As we shall see, this simple transform means that many problems in filter design can be solved just by using simple polynomial arithmetic.We shall also discover that the whole arsenal of techniques of complex analysis is also available.

You will probably not be surprised (or particularly happy) to know that some sources define **z** to be the positive exponential. There are also forms where the summation is in powers of z^{-n}.

If f_k is a step function:

$$\mathbf{f}_k \;=\; 1 \;/;\; k \,\geq\, 0$$
$$\mathbf{f}_k \;=\; 0 \;/;\; k \,<\, 0$$

then the Z transform is just:

$$\mathbf{F}[\mathbf{z}] \;=\; 1 \,+\, \mathbf{z} \,+\, \mathbf{z}^2 \,+\, \ldots$$

In the limit as the length of the time series, **n,** goes to infinity, this is just:

$$\mathbf{Limit}[\,\mathbf{f}[\mathbf{z}]\,,\,\mathbf{n} \to \infty\,] \;=\; \frac{1}{1\,-\,\mathbf{z}}$$

■ Inverse Filters

A filter **c** takes an input **x** and produces an output **y**:

$$x \xrightarrow{c} y$$

The inverse filter c^{-1} undoes the effect of **c**:

$$y \xrightarrow{c^{-1}} x$$

This is a problem you would need to solve if the "filter" is some instrumental effect that you wanted to eliminate to determine the "true" signal **x**.

With the Z transform this is easy. The effect of the filter is just:

$$Y[z] = C[z] \, X[z]$$

so:

$$X[z] \;=\; \frac{1}{C[z]} \, Y[z] \;=\; \left(\frac{1}{C[z]} \, C[z] \right) X[z]$$

So, the Z transform of the inverse filter is just the reciprocal of the Z transform of the filter.

For example, imagine that **c** is a "two term filter:"

$$c = \{1, \, 1/2\}$$

The output is just the convolution with **x**:

$$y_i \;=\; \frac{1}{2} \, x_{i-1} \;+\; x_i$$

The Z transform of the filter is:

$$C[z] \;=\; 1 \;+\; \frac{z}{2}$$

If the input is:

$$x = \{1, 1\}$$
$$X[z] = 1 + z$$

then the output is just:

$$Y[z] = C[z] X[z] = 1 + \frac{3z}{2} + \frac{z^2}{2}$$

The inverse filter is just:

$$C^{-1}[z] = \frac{1}{C[z]} = \frac{1}{1 + \frac{z}{2}} = 1 - \frac{z}{2} + \frac{z^2}{4} - \frac{z^3}{8} + \ldots$$

A moment's thought will convince you that the coefficients in the above series:

$$\{1, -1/2, 1/4, -1/8, \ldots\}$$

is just the impulse response of the inverse filter. In any case, the list of coefficients is a complete specification of the Z transform.

You should notice that if we have a filter with just a few terms, such as **1 + z/2**, the inverse filter has a very large number of significantly non-zero terms. This is just a consequence of the mathematics of reciprocals of polynomials.

We close this section with another example:

```
c = {1, 2}
C[z] = 1 + 2 z
```

The inverse is:

$$C^{-1}[z] = \frac{1}{C[z]} = \frac{1}{1 + 2z} = 1 - 2z + 4z^2 - 8z^3 + \ldots$$

with coefficients:

$$\{1, -2, 4, -8, \ldots\}$$

But the series does not converge! Thus **c** has a non-realisable inverse. A moment's reflection may convince you that a non-realisable inverse is closely associated with the situation that the original filter depends on all inputs, both from the past and from the future, i.e. when the filter is acausal. Thus, we can not undo the effects of such a filter.

■ Causal Filters

Now we can look a little more closely at the conditions for a causal filter, which is the main type considered here. Generalising from the previous section, a simple two term filter:

$$C[z] = \alpha + \beta z$$

has a non-realisable inverse if:

$$\left| \frac{\beta}{\alpha} \right| \geq 1$$

Thus for a stable causal filter:

$$\left| \frac{\alpha}{\beta} \right| > 1$$
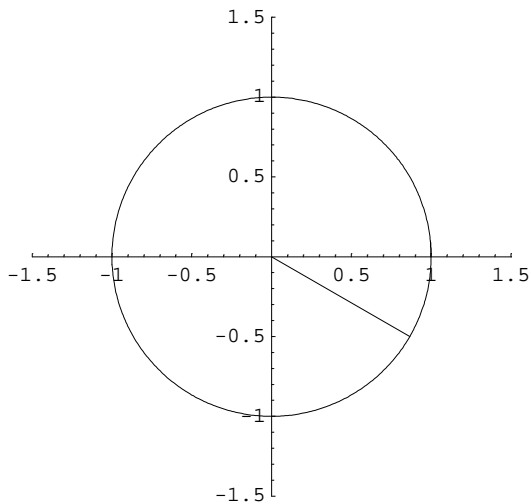
The inverse filter:

$$C^{-1}[z] = \frac{1}{\alpha + \beta z}$$

has a pole at:

$$z_r = -\frac{\alpha}{\beta}$$

This means that for a stable causal filter $|z_R| > 1$.

Now, **z** can be represented as a unit circle in the complex plane:

In the above figure we have also shown a line of unit length making an angle of -30 degrees (-.16667 π radians) with the **Re[z]** axis. Since for $\Delta = 1$:

$$z \equiv E^{-I \omega_j}$$

the point where the line meets the circle corresponds to:

$$\omega_j \ = \ +.16667 \, \pi \, \text{sec}^{-1}$$

For a causal filter **C[z]**, then, the root z_R must lie outside the unit circle.

Imagine you have a complicated filter with 100 terms:

$$c \ = \ \{c_0, \ c_1, \ c_2, \ \ldots, \ c_{99}\}$$

The Z transform is:

$$C[z] \ = \ \sum_{n=0}^{99} c_n \, z^n$$

Since convolution is commutative and associative, we can factor this into 99 two-term filters:

$$C[z] \ = \ (z \ - \ r_1) \, (z \ - \ r_2) \ \ldots \ (z \ - \ r_{99}) \, * \, c_{99}$$

The c_{99} factor is just the normalisation or *gain* of the filter. Thus if **C[z]** is causal each and every root r_i must lie outside the unit circle.

■ A Narrow Band Filter

Here we will design and then implement a narrow band filter.

$$C[z] = \frac{1}{z - a}$$

$$a \ = \ (1 \ + \ \epsilon) \, E^{-I \, \omega_0}$$

We think of ϵ as being small and real, and ω_0 as real. This filter produces complex output for real input; we will fix this in the next section. We will also show in the **Another Implementation** section that the filter is causal and recursive.

There is a singularity in **C[z]** at a point $(1 + \epsilon)$ from the origin and making an angle of -ω_0 with the real axis:
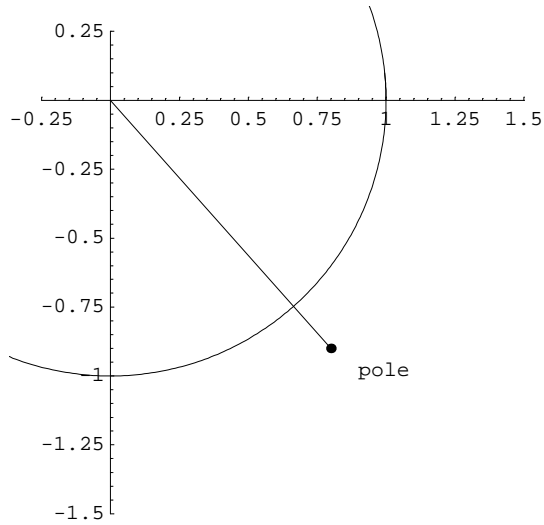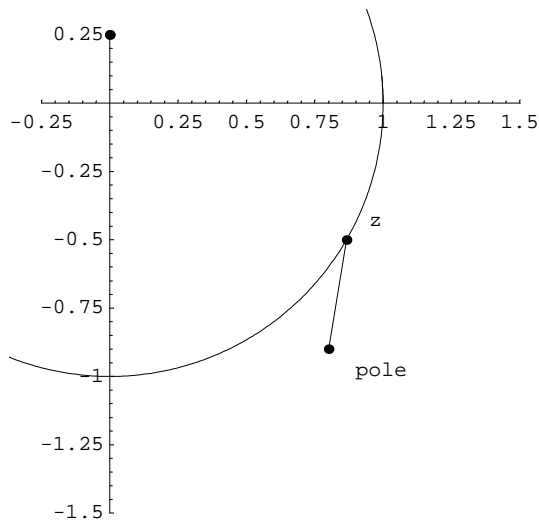
For the illustrated pole at **{0.8, -0.9}**, the values are:

$$\epsilon = \textbf{Sqrt[Plus @@ Power[\{0.8, -0.9\}, 2]]} - 1 = 0.204159$$

$$\omega_0 = -\textbf{ArcTan @@ \{0.8, -0.9\}} = 0.26870\,\pi$$

For any value of z, represented as a point on the unit circle, the value of the denominator of **C[z]**, **z - a**, is the "phasor" pointing from the pole to **z**:

[If you have never heard of a "phasor" before, you can think of it as an analog of a vector having magnitude and direction in the complex plane.]

The minimum length of the phasor is ϵ, which occurs when the angular frequency of **z** is ω_0. Note that since the filter **C[z]** is the reciprocal of the phasor, this is the value of **z** for which the output of the filter is a maximum.

We represent the phasor for arbitrary **z** as:

$$\mathbf{p} = \mathbf{b}\,\mathbf{E}^{\mathbf{I}\,\theta}$$

and the filter is **1/p**. Thus the maximum amplitude of the filter is **1/ε** and its phase is **-θ**.

It should be fairly clear from the above figure that as the pole moves closer to unit circle, the values of the phasor which are small occur for a more restricted ranges of z than if the pole is further away. Thus ε is also a measure of the width of the peak. The symmetry of the figure also shows that the peak in the filter is also symmetric. Thus the above is a design for a narrow band filter.

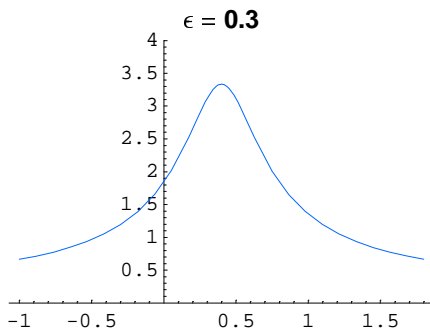A nearly trivial amount of algebra and the Law of Cosines shows that the magnitude of **1/p** is:

$$\mathbf{value[\omega\_]} := \frac{1}{\sqrt{(\epsilon + 1)^2 - 2\,\mathbf{Cos}[\omega - \omega_0]\,(\epsilon + 1) + 1}}$$

For $\epsilon = 0.1$ and $\omega_0 = 0.4$, this looks like:
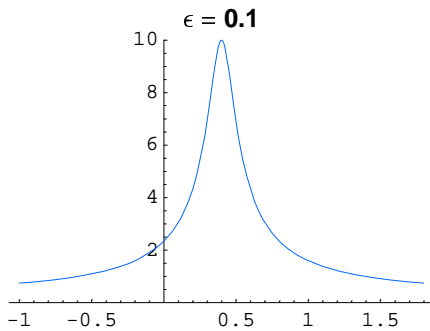
A filter with this type of frequency response we shall call a *narrow band* filter

Increasing ε to 0.3, which moves the pole further away from the unit circle, widens the peak:

Usually a "filter" has a maximum gain of one; higher maximums indicate an amplifier. We ignore the question of normalisation; for this filter normalisation would just be multiplying the filter by ε.

One measure of the width of the peak is the values of the frequency for which the amplitude is one-half of the maximum. These are:

$$\texttt{Solve}\left[\texttt{value[}\omega\texttt{]} \;==\; \frac{1}{2\,\epsilon}, \; \omega\right]$$

$$\left\{\left\{\omega \rightarrow -\texttt{ArcCos}\left[\frac{2 + 2\,\epsilon - 3\,\epsilon^2}{2\,(1 + \epsilon)}\right] + \omega_0\right\}, \; \left\{\omega \rightarrow \texttt{ArcCos}\left[\frac{2 + 2\,\epsilon - 3\,\epsilon^2}{2\,(1 + \epsilon)}\right] + \omega_0\right\}\right\}$$

The above form tells us the frequencies for the half-maximum given a value of ϵ. Usually when we design a filter we are given the frequencies for, say, the half-width at half maximum and need to determine the value of ϵ; although a bit of a mess, the solution is:

$$\texttt{Solve}\left[\texttt{value[}\omega\texttt{]} \;==\; \frac{1}{2\,\epsilon}, \; \epsilon\right]$$

$$\left\{\left\{\epsilon \rightarrow \frac{1}{3}\left(1 - \texttt{Cos}[\omega - \omega_0] - \sqrt{7 - 8\,\texttt{Cos}[\omega - \omega_0] + \texttt{Cos}[\omega - \omega_0]^2}\right)\right\},\right.$$

$$\left.\left\{\epsilon \rightarrow \frac{1}{3}\left(1 - \texttt{Cos}[\omega - \omega_0] + \sqrt{7 - 8\,\texttt{Cos}[\omega - \omega_0] + \texttt{Cos}[\omega - \omega_0]^2}\right)\right\}\right\}$$

In *Mathematica* we can evaluate this fairly simply. Say $\omega = 1.1$ and $\omega_0 = 1$. Then the value of ϵ is:

$$\texttt{\% /. } \{\omega \rightarrow \texttt{1.1}, \; \omega_0 \rightarrow \texttt{1}\}$$

$$\{\{e \rightarrow 0.0594003\}, \; \{e \rightarrow -0.0560697\}\}$$

Clearly, the second solution is for a pole inside the unit circle and is not the one we want.

Finally, we mentioned above that this filter produces complex output for real input. We can see this simply by noting that:

$$\texttt{C[z]} \;=\;$$
$$\frac{1}{\texttt{z} - \texttt{a}} \;=\; -\frac{1}{\texttt{a}}\left(\frac{1}{1 - \texttt{z}/\texttt{a}}\right) \;=\; -\frac{1}{\texttt{a}}\left(1 + \frac{\texttt{z}}{\texttt{a}} + \left(\frac{\texttt{z}}{\texttt{a}}\right)^2 + \left(\frac{\texttt{z}}{\texttt{a}}\right)^3 + \ldots\right)$$

Note that **a** is complex unless it happens to lie on the real axis. Thus the output of the filter is necessarily complex.

■ Real Output

Our example filter is:

$$C[z] = \frac{1}{z - a}$$
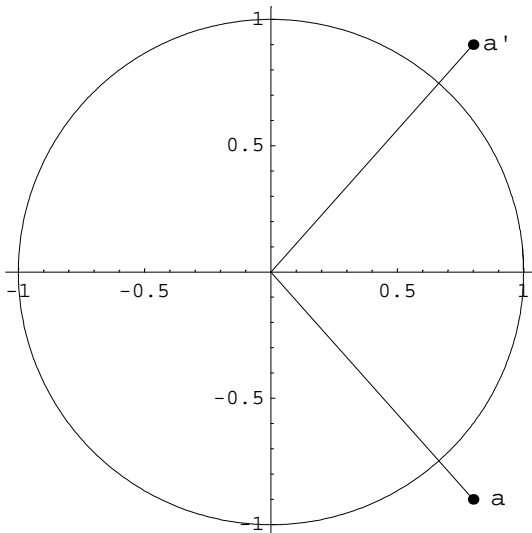
$$\mathbf{a} = \mathbf{(1 + \epsilon)} \; \mathbf{E^{-I\,\omega_0}}$$

Imagine a second filter:

$$\mathbf{C'[z]} \; = \; \frac{1}{\mathbf{z - a'}}$$

where **a'** is just the complex conjugate of **a**:

$$\mathbf{a' = Conjugate[a] = (1 + \epsilon) \; E^{+I\,\omega_0}}$$

The pole of this second filter makes an angle of $+\omega_0$ with the real axis.

The product of these two filters, **C[z] C'[z]**, will have maxima at $-\omega_0$ and $+\omega_0$, and the output of the combination will be real if its input is real. This generalises to any filter.

■ Another Implementation

Earlier we pointed out that a filter such as:

$$\mathbf{C[z]} \; = \; \frac{1}{\mathbf{z - a}}$$

expands into an infinite polynomial in **z/a** which converges fairly slowly. Of course, the inverse of the above filter is only **z - a**, a first order polynomial. In this section we discuss a technique to avoid having

to deal with very long, possibly infinite, series. Also, as promised, we shall show that our narrow band filter is causal and recursive.

For an arbitrary filter **C[z]** with a realisable inverse $C^{-1}[z] \equiv D[z]$, assume that the inverse is a polynomial of order L_d:

$$\frac{1}{C[z]} = D[z] = \sum_{k=0}^{L_d} d_k \, z^k$$

Since this is the inverse filter, it takes an output time series y_k and reproduces the input time series x_k. We write this as the convolution of **d** and **y**:

$$x_i = \sum_{k=0}^{L_d} d_k \, y_{i-k}$$

Now a trick! We rewrite the above as:

$$x_i = \sum_{k=1}^{L_d} d_k \, y_{i-k} + d_0 \, y_i$$

and rearrange:

$$d_0 \, y_i = x_i - \sum_{k=1}^{L_d} d_k \, y_{i-k}$$

Note that this filter is obviously causal: the output at time **i** depends on the input at time **i** and on the outputs at earlier times. Since the output does depend on the earlier outputs, the filter is also recursive.

Of course, the above way of representing the filter **C[z]** is equivalent to:

$$Y[z] = C[z] \, X[x]$$

But if the original filter **C[z]** has a large radius of convergence, then the inverse will have a smaller radius of convergence, i.e. L_d will be less than then the effective length of **C[z]**; for our narrow band filter L_d is one.
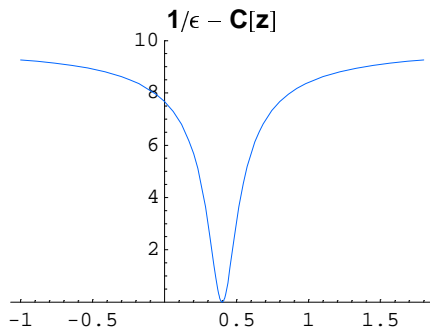
■ Other Filters

We have been dealing almost exclusively with the narrow band filter:

$$C[z] = \frac{1}{z - a}$$

$$a = (1 + \epsilon) \, E^{-I \, \omega_0}$$
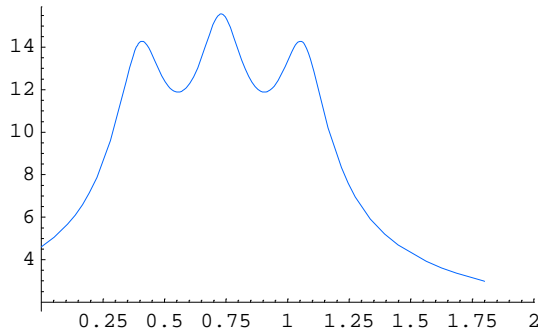
where the maximum amplitude of $1/\epsilon$ occurs when $\omega = \omega_0$. Then $1/e$ minus $C[z]$ is called a *notch* filter:

Imagine you have two narrow band filters with different values of w0 just far enough apart that the half-widths at half maxima are roughly the same. Then we have a band-pass filter. For example, here is a filter with three poles:
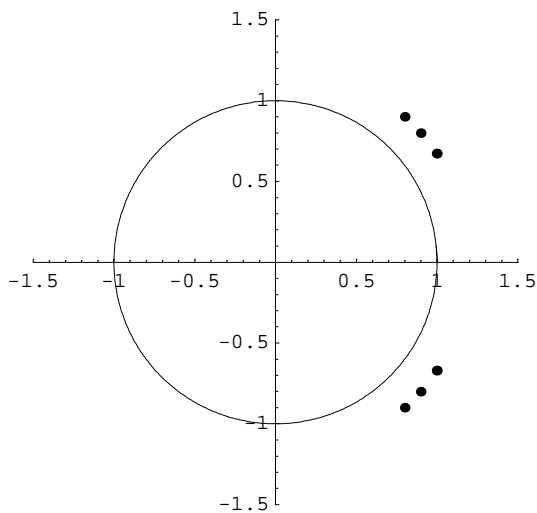
whose output is:

Adding additional narrow band filters at increasing values of w0 will increase the width of the band. Note that in general such a filter will not have a perfectly flat gain in the region of frequencies it passes. Getting such a nice shape is somewhat beyond the level of this course.

If the succession of narrow band filters includes the real axis, we have a low pass filter. Can you see how to design a high pass filter?

Finally, consider the following filter:

$$C[z] = \frac{z - a^*}{1 - az}$$

In terms of our definitions of **z** and **a** this can be re-written as:

$$e^{-i\omega} \left(\frac{1 - (1 + \epsilon)\, e^{i\,(\omega + \omega_0)}}{1 - (1 + \epsilon)\, e^{-i\,(\omega - \omega_0)}} \right)$$

The phase of this filter will obviously be pretty complex, but its gain is *one* for any frequency ω.

■ Power Spectra

So far when doing filtering the frequency domain we have only considered the *amplitude* of the frequency components. Often it is the *power*, not the amplitude that is of interest. However, we know that the power $\mathcal{P}$ is related to the amplitude ψ by:

$$\mathcal{P} \propto |\psi|^2$$

Thus, within questions of overall normalisation and units, if the frequency spectrum is $C[\omega]$ the power spectrum is $|C[w]|^2$. This relation will become more complicated later when we consider stochastic or noise processes of infinite duration.

Earlier we discussed the output of a narrow band filter, whose values are:

$$\text{value}[\omega\_] := \frac{1}{\sqrt{(\epsilon + 1)^2 - 2\,\text{Cos}[\omega - \omega_0]\,(\epsilon + 1) + 1}}$$

The power is just the square of this:

$$\text{power}[\omega\_] := \frac{1}{(\epsilon + 1)^2 - 2\,\text{Cos}[\omega - \omega_0]\,(\epsilon + 1) + 1}$$

The maximum value of this equation is $1/\epsilon^2$. The values of ϵ for which the power is one half the maximum value is given by:

$$\text{Solve}\left[\text{power}[\omega] == \frac{1}{2\,\epsilon^2},\ \epsilon\right]$$

$$\left\{\left\{\epsilon \to 1 - \text{Cos}[\omega - \omega_0] - \sqrt{3 - 4\,\text{Cos}[\omega - \omega_0] + \text{Cos}[\omega - \omega_0]^2}\right\},\right.$$
$$\left.\left\{\epsilon \to 1 - \text{Cos}[\omega - \omega_0] + \sqrt{3 - 4\,\text{Cos}[\omega - \omega_0] + \text{Cos}[\omega - \omega_0]^2}\right\}\right\}$$

If, say, the full width at half maximum (FWHM) in the power spectrum is 2 *radians/sec* (so the half width at half maximum is 1 radian/sec) and the central frequency is 1.5×10^3 *radians/sec*, we can make *Mathematica* give us the values of of ϵ with:

$$\% \ /.\ \{\omega \to 1501,\ \omega_0 \to 1500\}$$

$$\{\{\epsilon \to -0.603654\},\ \{\epsilon \to 1.52305\}\}$$

The first solution above is inside the unit circle and ignored.

■ Authors

This document is Copyright © Richard C. Bailey and David M. Harrison, 1998, 1999. It was prepared using *Mathematica,* which produced a PostScript file; *Adobe Distiller* was then used to prepare the PDF file. This is version 1.4 of the document, date (m/d/y) 03/03/99.