Many more special zTs and properties can be derived but this is enough for now. We will return to the zT when we study signal processing systems. Systems are often defined by complex recursions, and the zT will enable us to convert these into simple algebraic equations.

EXERCISES

4.11.1 Write a graphical program that allows one to designate a point in the z-plane and then draws the corresponding signal.

4.11.2 Plot the z transform of $\delta_{n,m}$ for various m.

4.11.3 Prove the linearity of the zT.

4.11.4 Express $\mathrm{zT}(\alpha^n x_n)$ in terms of $x(z) = \mathrm{zT}(x_n)$.

4.11.5 What is the z transform of the following digital signals? What is the ROC?

1. $\delta_{n,2}$
2. u_{n+2}
3. $a^n u(n)$
4. $a^n u(-n-1)$
5. $\frac{1}{2}^n u_n + \frac{3}{2}^n u_{-n}$

4.11.6 What digital signals have the following z transforms?

1. z^{-2}
2. z^{+2}
3. $\frac{1}{1-2z^{-1}}$ ROC $|z| > |2|$

4.11.7 Prove the following properties of the zT:

1. linearity
2. time shift $\mathrm{zT}s_{n-k} = z^{-k}S(z)$
3. time reversal $\mathrm{zT}s_{-n} = S(\frac{1}{z})$
4. conjugation $\mathrm{zT}s_n^* = S^*(z^*)$
5. rescaling $\mathrm{zT}(\alpha^n s_n) = S(\frac{z}{\alpha})$
6. z differentiation $\mathrm{zT}(ns_n) = -z\frac{d}{dz}S(z)$

4.12 The Other Meaning of Frequency

We have discussed two quite different representations of functions, the Taylor expansion and the Fourier (or z) transform. There is a third, perhaps less widely known representation that we shall often require in our signal processing work. Like the Fourier transform, this representation is based on frequency, but it uses a fundamentally different way of thinking about the concept of frequency. The two usages coincide for simple sinusoids with a single constant frequency, but differ for more complex signals.

Let us recall the examples with which we introduced the STFT in Section 4.6. There we presented a pure sinusoid of frequency f_1, which abruptly changed frequency at $t = 0$ to become a pure sine of frequency f_2. Intuition tells us that we should have been able to recover an *instantaneous frequency*, defined at every point in time, that would take the value f_1 for negative times, and f_2 for positive times. It was only with difficulty that we managed to convince you that the Fourier transform cannot supply such a frequency value, and that the uncertainty theorem leads us to deny the existence of the very concept of instantaneous frequency. Now we are going to produce just such a concept.

The basic idea is to express the signal in the following way:

$$s(t) = A(t) \cos\Big(\Phi(t)\Big) \tag{4.65}$$

for some $A(t)$ and $\Phi(t)$. This is related to what is known as the *analytic representation* of a signal, but we will call it simply the *instantaneous representation*. The function $A(t)$ is known as the *instantaneous amplitude* of the signal, and the $\Phi(t)$ is the *instantaneous angle*. Often we separate the angle into a linear part and the deviation from linearity

$$s(t) = A(t) \cos\Big(\omega t + \phi(t)\Big) \tag{4.66}$$

where the frequency ω is called the *carrier frequency*, and the residual $\phi(t)$ the *instantaneous phase*.

The *instantaneous frequency* is the derivative of the instantaneous angle

$$2\pi f(t) = \frac{d\Phi(t)}{dt} = \omega + \frac{d\phi(t)}{dt} \tag{4.67}$$

which for a pure sinusoid is exactly the frequency. *This* frequency, unlike the frequencies in the spectrum, is a single function of time, in other words, a signal. This suggests a new world view regarding frequency; rather than understanding signals in a time interval as being made up of many frequencies, we claim that signals are fundamentally sinusoids with well-defined instantaneous amplitude and frequency. One would expect the distribution of different frequencies in the spectrum to be obtained by integration over the time interval of the instantaneous frequency. This is sometimes the case.

Consider, for example, a signal that consists of a sinusoid of frequency f_1 for one second, and then a sinusoid of nearby frequency f_2 for the next second. The instantaneous frequency will be f_1 and then jump to f_2; while the spectrum, calculated over two seconds, will contain two spectral lines at f_1 and f_2. Similarly a sinusoid of slowly increasing instantaneous frequency will have a spectrum that is flat between the initial and final frequencies.

This new definition of frequency seems quite useful for signals that we usually consider to have a single frequency at a time; however, the instantaneous representation of equation (4.65) turns out to very general. A constant DC signal can be represented (using $\omega = 0$), but it is easy to see that a constant plus a sinusoid can't. It turns out (as usual, we will not dwell upon the mathematical details) that all DC-less signals can be represented. This leads to an apparent conflict with the Fourier picture. Consider a signal composed of the *sum* of the two sinusoids with close frequencies f_1 and f_2; what does the instantaneous representation do, jump back and forth between them? No, this is exactly a *beat* signal (discussed in exercise 2.3.3) with instantaneous frequency a constant $\frac{1}{2}(f_1 + f_2)$, and sinusoidally varying amplitude is with frequency $\frac{1}{2}|f_1 - f_2|$. Such a signal is depicted in Figure 4.13. The main frequency that we see in this figure (or hear when listening to such a combined tone) is the instantaneous frequency, and after that the effect of $A(t)$, *not* the Fourier components.

We will see in Chapter 18 that the instantaneous representation is particularly useful for the description of communications signals, where it is the basis of *modulation*. Communications signals commonly carry informa-

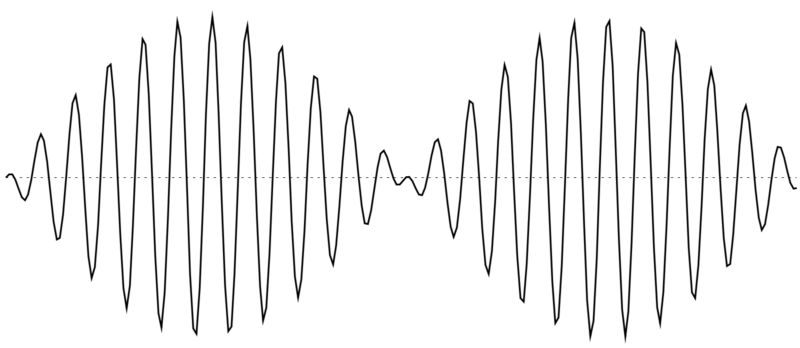

Figure 4.13: The beat signal depicted here is the sum of two sinusoids of relatively close frequencies. The frequencies we see (and hear) are the average and half-difference frequencies, *not* the Fourier components.

tion by varying (modulating) the instantaneous amplitude, phase, and/or frequency of a sinusoidal 'carrier'. The carrier frequency is the frequency one 'tunes in' with the receiver frequency adjustment, while the terms *AM* (**A**mplitude **M**odulation) and *FM* (**F**requency **M**odulation) are familiar to all radio listeners.

Let us assume for the moment that the instantaneous representation exists; that is, for any *reasonable* signal $s(t)$ without a DC component, we assume that one can find carrier frequency, amplitude, and phase signals, such that equation (4.65) holds. The question that remains is *how* to find them. The answering of this question is made possible through the use of a mathematical operator known as the Hilbert transform.

The Hilbert transform of a real signal $x(t)$ is a real signal $y(t) = \mathcal{H}x(t)$ obtained by shifting the phases of all the frequency components in the spectrum of $x(t)$ by $90°$. Let's understand why such an operator is so remarkable. Assume $x(t)$ to be a simple sinusoid.

$$x(t) = A\cos(\omega t)$$

Obtaining the $90°$ shifted version

$$y(t) = \mathcal{H}x(t) = A\cos\left(\omega t - \frac{\pi}{2}\right) = A\sin(\omega t)$$

is actually a simple matter, once one notices that

$$y(t) = A\cos\left(\omega\left(t - \frac{\pi}{2\omega}\right)\right) = x\left(t - \frac{\pi}{2\omega}\right)$$

which corresponds to a time delay. So to perform the Hilbert transform of a pure sine one must merely delay the signal for a time corresponding to one quarter of a period. For digital sinusoids of period L samples, we need to use the operator $\text{z}^{-\frac{L}{4}}$, which can be implemented using a FIFO of length $L/4$.

However, this delaying tactic will not work for a signal made up of more that one frequency component, e.g., when

$$x(t) = A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t)$$

we have

$$y(t) = \mathcal{H}x(t) = A_1\sin(\omega t) + A_2\sin(\omega t)$$

which does *not* equal $x(t - \tau)$ for any time delay τ.

Hence the Hilbert transform, which shifts all frequency components by a quarter period, independent of frequency, is a nontrivial operator. One way of implementing it is by performing a Fourier transform of the signal, individually shifting all the phases, and then performing an inverse Fourier transform. We will see an alternative implementation (as a *filter*) in Section 7.3.

Now let us return to the instantaneous representation

$$x(t) = A(t)\cos\Big(\omega t + \phi(t)\Big) \tag{4.68}$$

of a signal, which we now call $x(t)$. Since the Hilbert transform instantaneously shifts all $A\cos(\omega t)$ to $A\sin(\omega t)$, we can explicitly express $y(t)$.

$$y(t) = \mathcal{H}\,x(t) \;=\; A(t)\sin\Big(\omega t + \phi(t)\Big) \tag{4.69}$$

We can now find the instantaneous amplitude by using

$$A(t) = \sqrt{x^2(t) + y^2(t)} \tag{4.70}$$

the instantaneous phase via the (four-quadrant) arctangent

$$\phi(t) = \tan^{-1}\frac{y(t)}{x(t)} - \omega t \tag{4.71}$$

and the instantaneous frequency by differentiating the latter.

$$\omega(t) = \frac{d\phi(t)}{dt} \tag{4.72}$$

The recovery of amplitude, phase, or frequency components from the original signal is called *demodulation* in communications signal processing.

We have discovered a method of constructing the instantaneous representation of any signal $x(t)$. This method can be carried out in practice for digital signals, assuming that we have a numeric method for calculating the Hilbert transform of an arbitrary signal. The instantaneous frequency similarly requires a numeric method for differentiating an arbitrary signal. Like the Hilbert transform we will see later that differentiation can be implemented as a filter. This type of application of numerical algorithms is what DSP is all about.