WHAT THEY NEVER TOLD YOU ABOUT DIGITAL POWER SPECTRA
spectral leakage. . . windowing. . . power spectra from autocorrelation

1. BRIEF REVIEW: TWO WAYS TO CALCULATE THE POWER SPECTRUM

We have the conceptual descriptions of the two methods:

(1) Take the FT of the time-varying signal; multiply the result by its complex conjugate. This is the power spectrum. We call this the FT×CC method.

(2) Take the autocorrelation function of the time-varying signal; take its cosine FT. This is the exactly identical power spectrum. We call this the AC→FT method.

And we have the formal mathematical description. Here, $E(t)$ is the input voltage, which is a function of time, and $P(\nu)$ is the power spectrum, which is a function of frequency. We have:

(1) FT×CC method: First, take the FT of the input voltage:

$$E(\nu) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} E(t) e^{[2\pi i]\nu t} dt \; ; \tag{1a}$$

and then multiply the result by its complex conjugate:

$$P(\nu) = E(\nu) \times [E(\nu)]^* \; . \tag{1b}$$

(2) AC→FT method: First, get the autocorrelation function of the input voltage:

$$A(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} E(t) E(t+\tau) dt \; . \tag{2a}$$

and then take the cosine FT—which is equivalent to the standard complex transform but keeping only the real part:

$$P(\nu) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} A(\tau) e^{[2\pi i]\nu t} d\tau. \tag{2b}$$

2. EFFECTS OF T BEING FINITE: INTEGRAL TRANSFORMS

Equations (1) and (2) are fine as they stand, but when we sample a signal T is finite. As explained in *FT's with FT's*, this limits our spectral resolution. But it also has another effect:

"spectral leakage".

To understand this, consider the analytic solution of equations (1) for a monochromatic cosine wave of frequency ν_s, that is $E(t) = \cos(2\pi\nu_s t)$. For $T \to \infty$, $P(\nu) = 0$ except for $\nu = \nu_s$; this is commonly expressed by the term "Dirac delta function". But more generally, we have as the analytic solution

$$P(\nu) = \left(\frac{\sin(2\pi(\nu - \nu_s)T)}{2\pi(\nu - \nu_s)T}\right)^2 . \tag{3}$$

It is simpler to write $\delta\nu = (\nu - \nu_s)$ and write

$$P(\nu) = \left(\frac{\sin(2\pi\delta\nu T)}{2\pi\delta\nu T}\right)^2 . \tag{4}$$

and we see that this is just the famous $\frac{\sin x}{x}$ type behavior (but squared because it's the power spectrum). This quantitatively expresses the degradation in frequency resolution when you sample for a finite time; roughly, the spectral resolution is the width of this function, which is $\sim \frac{1}{2T}$.

But there's *more to it!* Not only is the infinitesmally-narrow delta function replaced by a function of width $\sim \frac{1}{2T}$ but also the function $\frac{\sin x}{x}$ has "sidelobes": it spreads power out over a huge range of frequencies, albeit—after squaring, anyway—at a low level. The spread-out power gets weaker as you move away from the signal, that is as $\delta\nu$ increases—but it doesn't decrease monotonically. Rather, the sidelobes mean that there the spread-out power goes to zero periodically for $\delta\nu = \frac{1}{2T}$. In colloquial terms, the power of the monochromatic sine wave "leaks" away in the power spectrum to all frequencies. Not surprisingly, this is called "spectral leakage". It's covered in chapter 12.7 of *Numerical Recipes.*

You can reduce spectral leakage by using a "window function" $W(t)$ that is not "square". What on Earth does this mean? In Equation (1a), when you use a finite T you are integrating over a "window" of total length $2T$, as follows:

$$E(\nu) = \frac{1}{2T} \int_{-T}^{+T} W(t) E(t) e^{[2\pi i]\nu t} dt . \tag{5}$$

In equation (1a), the window function $W(t)$ is unity; when we let $W(t) = 1$ for the time window of length $2T$, then this is called a "square window function". A square window function produces "sharp edges" in the sense that the function $E(t)$ has its full amplitude for all t within the window and then cuts off precipitiously at the ends of the window. *It's this sudden cutoff—the sharp edge—that produces spectral leakage.*

We can reduce spectral leakage by making $W(t)$ a function that produces a *gradual* cutoff so as to eliminate the sharp edge. We make $W(t)$ a function that gradually falls to zero at the window edges $-T$ and T, such as

$$W(t) = \frac{1}{2}\left(1 + \cos\frac{\pi t}{T}\right) \ . \tag{6}$$

This drastically reduces the sidelobes. This comes at a price, though: the spectral resolution gets worse—by about a factor of two. The reason is that the window function $W(t)$ goes to zero at the ends of the time window $2T$, which effectivly makes the window shorter. *Note:* We also use the term "weighting function" for $W(t)$.

3. EFFECTS OF T BEING FINITE: DISCRETE TRANSFORMS

3.1. The Weighting Function in Discrete Form.

All of the above discussion on integral transforms carries over directly into discrete transforms. At this point you should refer to equation (5) in *FT's with FT's* and its accompanying discussion. To refresh your memory, we'll recapitulate the essentials: we sample at intervals t_{smpl}, which is the same thing as using a sampling frequency $\nu_{smpl} = \frac{1}{t_{smpl}}$. We let $f_N \equiv \frac{\nu_{smpl}}{2}$. We sample $2J$ points, so $T = Jt_{smpl}$. We get a total bandwidth B (for real sampling, we get a J-point spectrum with $B = f_N$; for complex sampling, we get a $2J$-point spectrum with $B = 2f_N$, which includes both positive f_N and negative f_N) and the frequency spacing between the points on the power spectrum is $\Delta\nu = \frac{f_N}{J} = \frac{\nu_{smpl}}{2J}$ (this applies for both real and complex sampling). The time of any sample is jt_{smpl}, where $j = -J \rightarrow J-1$.

The frequency of any computed point on the power spectrum is $\nu = k\Delta\nu = \frac{k}{J}f_N$. For real sampling, $k = 0 \rightarrow J-1$; for complex sampling, $k = -J \rightarrow J-1$). To shorten the notation we write k instead of $k\Delta\nu$ and j instead of jt_{smpl} and obtain equation (5) in *FT's with FT's*:

$$E(k) = \frac{1}{2J}\sum_{j=-J}^{J-1} W(j)E(j)e^{[\pi i]\frac{kj}{J}} \ . \tag{7}$$

Here, $W(j)$ is the window function.

If we use a square window function, then $W(j) = 1$ and the above equation (4) applies. The digital form of equation (4) is

$$P(k) = \left(\frac{\sin(\pi\delta k)}{\pi\delta k}\right)^2 \ , \tag{8}$$

where we write $\delta\nu = \delta k\Delta\nu$ and realize that δk is *almost never an integer* because it almost never happens that a pure sine wave happens to have a frequency that lies precisely at one of the computed spectral points, whose frequencies are $k\Delta\nu$. In the rare case that δk *is* an integer, then the pure sine wave lies exactly on one of the computed spectral points and *all of the remaining computed spectral points happen to fall on the zeros of the sidelobes!* So in this special case you don't see any spectral leakage! But more generally, of course, you do.

If we use a Hanning window function, then $W(j)$ is the weighting function in time, which translates from equation 6 as

$$W(j) = \frac{1}{2}\left(1 + \cos\frac{\pi j}{J}\right) \ . \tag{9}$$

This above equation is for the case $j = -J \rightarrow (J-1)$. If $j = 0 \rightarrow (2J-1)$ it's

$$W(j) = \frac{1}{2}\left(1 - \cos\frac{\pi j}{J}\right) \ . \tag{10}$$

3.2. The PRICES YOU PAY for Non-Square Weighting—and Why It's Worth Paying.

In digital sampling, you pay a price in two ways for using a non-square window. First is the degradation in spectral resolution, which we discussed above. The second is sensitivity. When you use a non-square window, you are giving some samples less weight than others. It's as if you took fewer samples. And taking fewer samples means you get less signal/noise.

If you use a non-square weighting function you can reduce this degradation in sensitivity by including some groups of points more than once in your Fourier transforms. For example, if you take 256-point Fourier transforms, then with a square window you would transform the first group of 256 points; then "move over" by 256 points to a completely independent group and transform them; and repeat this process *ad infinitum.* However, with a non-square window you can gain back your sensitivity by transforming the first group of 256 points; then "move over" by, say, only 128 points so that the next group is *not* a completely independent group and transform this new group; and repeat this process *ad infinitum.* You pay for this reduced degradation in sensitivity by spending more computer time. *Numerical Recipes* discusses the trade-offs.

If there's not a sharp spike in the power spectrum then spectral leakage doesn't matter much. So unless there's interference, which usually *is* in the form of sharp spikes, you can avoid all of the above compromises by using a square window.

But: (**This is important!**) If you've used a square window and later on wish you hadn't—tough luck. You can't recover the original. This means interference—which is a problem for observations from Campbell Hall—can totally ruin your spectrum, even though it's only a single sharp spike. Then again, if you haven't used a square window, and later on find you don't have interference and wish you had—tough luck. You can't recover the original then, either. This means you've lost sensitivity and you can't regain it. So...

In radio astronomy, the most popular windowing/smoothing function is the "Hanning" function. That's the function we've written in equations 6, 9, and 10. *Numerical Recipes* doesn't think much of Hanning weighting. But if you look at their Figure 12.7.2, you see that it really cuts the leakage! With our interference situation, this is not a bad choice, and it's the only one we provide. Because we're computer-time limited, we don't include any point more than once in the Fourier transform.

4. THE AC→FT METHOD IN REAL (i.e., DIGITAL) LIFE

4.1. Digital Calculation of the Autocorrelation Function.

The way to digitally calculate the autocorrelation function $A(\tau)$ of the time-dependent function $E(t)$ is to carry its definition, which is by an analytic integral, to numerical summation in the standard way. We begin shifting the origin of the time axis so that all times are positive, which is the way we usually think of our samples, so we write

$$A(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_0^{+2T} E(t)E(t+\tau)dt \; . \tag{11}$$

Now let's translate this into a digital sum. We have $2N$ discrete samples separated uniformly in time by $\Delta t = t_{smpl} = \frac{1}{\nu_{smpl}}$. Recall your elementary calculus in which an integral was defined by cutting up the x-axis into tiny bits and taking a sum over the bits. Here we are integrating over time, so it makes sense to make the "tiny bit" equal to the sample time t_{smpl}. In terms of sample number n, we can write $t = nt_{smpl}$ and $dt = t_{smpl}$, so $E(t) = E(nt_{smpl})$ and, for convenience, we just write $E(n)$. Similarly, we will calculate $A(\tau)$ only for discrete values of τ whose increment is also t_{smpl}; we write $\tau = jt_{smpl}$ and, as for E, we write $A(j)$ instead of $A(jt_{smpl})$. With all this, the direct analog of equation (3) in summation form is

$$A(j) = \lim_{N \to \infty} \frac{1}{2N} \sum_{n=0}^{2N-1} E(n)E(n+j) \; . \tag{12}$$

This looks innocent enough, but it misses a fundamental fact of the real world: we don't live forever, so we can't let $N \to \infty$. No problem; we just get rid of the $\lim_{N \to \infty}$ and write the even simpler form

$$A(j) = \frac{1}{2N} \sum_{n=0}^{2N-1} E(n)E(n+j) \; . \tag{13}$$

But wait! We have just $2N$ samples of E—that is, $E(n)$ is defined only for $n = 0 \rightarrow 2N - 1$. So in the sum, whenever $n + j > 2N - 1$, we're in trouble—we have no samples to put in the sum! In other words, when N is finite, you have the problem of "end effects". What to do?

Now's the time to go back and review *FT's with FT's*. In §4 and Figure 2 of that handout we stressed that the summation form of the FT implicitly, and necessarily, assumes that both the input and output arrays are *periodic* outside the fundamental window of length $2N$—and the period is just $2N$. So it's obvious what to do: when $n + j > 2N - 1$, you use $E(n + j - 2N)$.

Similarly, $A(j)$ is periodic with period $2N$. Let's call $J = N$. Thus, $A(j)$ is defined for the interval $j = 0 \rightarrow 2J - 1$. And, of course, this periodicity makes it easy to generate values for $j < 0$.

4.2. CALCULATING THE FOURIER TRANSFORM OF THE AUTOCORRELATION FUNCTION

We do this just as discussed in *FT's with FT's*. In that handout, we rewrite its equation (5) here using the variables appropriate here, that is...

$$P(k) = \frac{1}{2J} \sum_{j=-J}^{J-1} A(j) e^{[\pi i]\frac{kj}{J}} \ . \tag{14}$$

Here, as in *FT's with FT's*, the frequency $\nu = \frac{k\nu_{smpl}}{2J}$, and for convenience we write $P(k)$ instead of $P(\frac{k\nu_{smpl}}{2J})$.

Now let's notice that, with a suitable change of variables in our above equation (2a), you can easily determine that $A(\tau) = A(-\tau)$: the autocorrelation function is *symmetric* in τ. This means that the imaginary portion of its FT is automatically zero. So, in taking the FT, you don't even have to specify that we want just the real part of the result!

Finally, re-read §7 of *FT's with FT's* to see how to handle the negative frequencies. To summarize that discussion, you add $P(k)$ for the negative and positive frequencies to get the correct result. This means that you can just take twice the result for, say, the positive frequencies. But because the zero-frequency result doesn't have a negative counterpart, you *don't* multiply it by two.

5. WEIGHTING FUNCTIONS IN THE FT×CC VERSUS THE AC→FT METHOD

There's a fundamental difference between applying weighting functions in the two methods of getting power spectra. In the FT×CC method, you apply the weighting function $W(t)$ to the *voltage*, that is *before* the Fourier transform; and then you "detect" the signal by squaring (really, by multiplying by its complex conjugate). So the weighting function is also "squared". Alternatively, in the AC→FT method, you apply $W(t)$ to the *correlation function*, which is equivalent to the *detected* voltage; the "squaring" has already taken place, so the weighting function does *not* get "squared". These applications of $W(t)$ *are not equivalent* and, furthermore, *cannot be made to be equivalent.* In the final analysis—which is too much to discuss here, but the essence is that $W(t) < 1$ so "squaring" it means that it gets smaller—this means that leakage is always much smaller with the FT×CC method.