LECTURE 15

DIGITAL FILTERS - II

When digital filters first arose they were viewed merely as a variant of the classical analog filters; people did not see them as essentially new and different. This is exactly the same mistake that was made endlessly by people in the early days of computers. I was told repeatedly, until I was sick of hearing it, that computers were nothing more than large, fast desk calculators. "Anything you can do by a machine you can do by hand.", so they said. This simply ignores the speed, accuracy, reliability, and lower costs of the machines vs. humans. Typically a single order of magnitude change (a factor of 10) produces fundamentally new effects, and computers are many, many times faster than hand computations. Those who claimed that there was no essential difference never made any significant contributions to the development of computers. Those who did make significant contributions viewed computers as something new to be studied on their own merits and not as merely more of the same old desk calculators, perhaps souped up a bit.

This is a common, endlessly made, mistake; people always want to think that something new is just like the past - they like to be comfortable in their minds as well as their bodies - and hence they prevent themselves from making any significant contribution to the new field being created under their noses. Not that everything that is said to be new really is new, and it is hard to decide in some cases when something is new, yet the all too common reaction of, "It's nothing new." is stupid. When something is claimed to be new, do not be too hasty to think it is just the past slightly improved - _it may be a great opportunity for you to do significant things_. But again it may be nothing new.

The earliest digital filter I used, in the early days of primitive computers, was one that smoothed first by 3's and then by 5's. Looking at the formula for smoothing, the smoothing by 3's has the transfer function

$$H(\omega) = \{\sin(3/2)\omega/3\sin(\omega/2)\}$$

which is easy to draw, Figure 15-1. The smoothing by 5's is the same except that the 3/2 becomes a 5/2 and is again easy to draw. Figure 15-1. One filter followed by the other is obviously their product, (each multiplies the input eigenfunction by the transfer function at that frequency), and you see that there will be three zeros in the interval, and the terminal value will be 1/15. An examination will show that the upper half of the frequencies were fairly well removed by this very simple program for computing a running sum of 3 numbers, followed by a running sum of 5 - as is common in computing practice the divisors were left to the very

1

end where they were allowed for by one multiplication, by 1/15.

Now you may wonder how, <u>in all its detail</u>, a <u>digital filter</u> removes frequencies from a stream of numbers – and even students who have had courses in digital filters may not be at all clear as to how the miracle happens. Hence I propose, before going further, to design a very simple digital filter and show you the inner working on actual numbers.

I propose to design a simple filter with just two coefficients, and hence I can meet exactly two conditions on the transfer function. When doing theory we use the angular frequency ω, but in practice we use rotations f, and the relationship is

$$f = \omega/2\pi \qquad (-1/2 < f < 1/2)$$

Let the first condition on the digital filter be that at f = 1/6 the transfer function is exactly 1, (this frequency is to get through the filter unaltered), and the second condition at f = 1/3 it is to be zero (this frequency is to be stopped completely). My simple filter has the form, with the two coefficients a and b,

$$y_n = au_{n-1} + bu_n + au_{n+1}$$

Substituting in the eigenfunction $\exp\{2\pi ifn\}$ we will get the transfer function, and using n = 0 for convenience,

$$H(f) = b + 2a \cos 2\pi f$$

For f = 1/6 $H(1/6) = 1 = b + 2a(1/2) = b + a$

For f = 1/3 $H(1/3) = 0 = b + 2a(-1/2) = b - a$

The solution is

$$a = b = 1/2$$

and the smoothing filter is simply

$$y_n = (1/2)[u_{n-1} + u_n + u_{n+1}]$$

In words, the output of the filter is the sum of three consecutive inputs divided by 2, and the output is opposite the middle input value. [It is the earlier smoothing by 3's except for the coefficient 1/2.]

Now to produce some sample data for the input to the filter. At the frequency f = 1/6 we use a cosine at that frequency taking the values of the cosine at equal spaced values n = 0, 1, ... , while the second column of data we use the second frequency f = 1/3, and finally on the third column is the sum of the two other columns and is a signal composed of the two frequencies in equal amounts.

n	1/6	1/3	sum
0	1	1	2
1	1/2	-1/2	0
2	-1/2	-1/2	-1
3	-1	1	0
4	-1/2	-1/2	-1
5	1/2	-1/2	0
6	1	1	2
7	1/2	-1/2	0
8	-1/2	-1/2	-1
...	...	...	...

Let us run the data through the filter. We compute, according to the filter formula, the sum of three consecutive numbers in a column and then divide their sum by 2. Doing this on the first column you will see that each time the filter is shifted down one line it reproduces the input function (with a multiplier of 1). Try the filter on the second column and you will find that every output is exactly 0, the input function multiplied by its eigenvalue 0. The third column, which is the sum of the first two columns, should pass the first and stop the second frequency, and you get out exactly the first column. You can try the 0 frequency input and you should get exactly 3/2 for every value, if you try f = 1/4 you should get the input multiplied by 1/2 (the value of the transfer function at f = 1/2).

You have just seen a digital filter in action. <u>The filter decomposes the input signal into all its frequencies, multiplies each frequency by its corresponding eigenvalue, (the transfer function), and then adds all the terms together to give the output</u>. The simple linear formula of the filter does all this!

We now return to the problem of designing a filter. What we often want ideally is a transfer function that has a sharp cutoff between the frequencies it passes exactly (with eigenvalues 1), and those which it stops (with eigenvalues 0). As you know, a Fourier series can represent such a discontinuous function, but it will take an infinite number of terms. However, we have only a modest number available if we want a practical filter; $2k + 1$ terms in the smoothing filter gives only $k + 1$ free coefficients, and hence only $k + 1$ arbitrary conditions can be met by the corresponding sum of cosines.

If we simply expand the desired transfer function into a sum of cosines and then truncate it we will get a least squares approximation to the transfer function. But at a discontinuity the least squares fit is not what you probably think it is.

To understand what we will see at a discontinuity we must investigate the <u>Gibbs' phenomena</u>. We first recall a theorem: If a series of continuous functions converges uniformly in a closed interval then the limit function is continuous. But the limit

function we want to approximate is not continuous, it has a jump (discontinuity) between the pass and stop bands of frequencies. No matter how many terms in the series we take, since there cannot be a uniform convergence, we can expect(?) to see a significant overshoot in the neighborhood of the singularity. As we take more terms the size of the overshoot will _not_ approach 0.

Another story. Michelson, of Michelson-Morley fame, built an analog machine to find the coefficients of a Fourier series out to 75 terms. The machine could also, because of the duality of the function and the coefficients, go from the coefficients back to the function. When Michelson did this he observed an overshoot and asked the local mathematicians why it happened. They all said that it was his equipment - and yet he was well known as a very careful experimenter. Only Gibbs, of Yale, listened and looked into the matter. The simplest direct approach is to expand a standard discontinuity, say the function

$$f(t) = 1 \text{ for } x > 0 \text{ and } -1 \text{ for } x < 0$$

into a Fourier series of a finite number of terms, rearrange things, and then find the location of the first maximum and finally the corresponding height of the function there. One finds, Figure 15-2, an overshoot of 0.08949, or 8.949% overshoot, in the limit as the number of terms in the Fourier series approaches infinity. Many people had the opportunity to discover (really rediscover) the Gibbs' phenomena, and it was Gibbs who made the effort. It is another example of what I maintain, there are opportunities all around and few people reach for them. As Pasteur said, "Luck favors the prepared mind." This time the person who was prepared to listen and help a first class scientist in his troubles got the fame.

I remarked that it was rediscovered. Yes. In the 1850's the contradiction in Cauchy's textbooks, that (1) a convergent series of continuous functions converged to a continuous function (it was so stated in his book!), and (2) the Fourier expansion of a discontinuous function (also in his book) flatly contradicted each other. Some people looked into the matter and found that they needed the concept of _uniform convergence_. Yes, the overshoot of the Gibbs' phenomena occurs for any series of continuous functions, not just to the Fourier series, and was known to some people, but it had not diffused into common usage. For the general set of orthogonal functions the amount of overshoot depends on where in the interval the discontinuity occurs, which differs from the Fourier functions where the amount of the overshoot is independent of where the discontinuity occurs.

We need to remind you of another feature of the Fourier series. If the function exists (for practical purposes) then the coefficients fall off like $1/n$. If the function is continuous, Figure 15-3, (the two extreme end values must be the same) and the derivative exists then the coefficients fall off like $1/n^2$; if the first derivative is continuous and the second derivative exists then they fall off like $1/n^3$; if the second derivative is continuous and the third derivative exists then $1/n^3$. etc. Thus

4

the rate of convergence is directly observable from the function along the real line - which is not true for the Taylor series whose convergence is controlled by singularities which may lie in the complex plane.

Now we return to our design of a smoothing digital filter using the Fourier series to get the leading terms. We see that the least squares fit has trouble at any discontinuity - there is a nasty overshoot in the transfer function for any finite number of terms, no matter how far out we go.

To remove this overshoot we first examine the Lanczos' window, also called a "box car", or a "rectangular" window. Lanczos reasoned that if he averaged the output function over an interval of the length of a period of the highest frequency present, then this averaging would greatly reduce the ripples. To see this in detail we take the Fourier series expansion truncated at the N-th harmonic, and integrate about a point t in a symmetric interval of length 1/N of the whole interval. Set up the integral for the averaging,

$$(N/2\pi) \, \text{INT}[x-\pi/N, \; x+\pi/N; \; g(s) \; ds]$$

$$= (N/2\pi) \, \text{INT}[\{a_0/2 + \text{SUM}[1,N;\{a_k \cos ks + b_k \sin ks]\}ds]$$

We now do the integrations,

$$= a_0/2 + (N/2\pi) \text{SUM}[1,N;(a_k/k)\sin ks - (b_k/k)\cos ks] \, | \, x+\pi/N, x-\pi/N$$

apply a little trigonometry for the difference of sines and cosines from the two limits,

$$= a_0/2 + \text{SUM}[1,N;\{a_k \cos kx + b_k \sin kx] [\{\sin (\pi k/N)/(\pi k/N)\}]$$

and you come out with the original coefficients multiplied by the so called <u>sigma factors</u>

$$\sigma(N,k) = \sin(\pi k/N)/(\pi k/N)$$

An examination of these numbers as a function of k, (N being fixed and is the number of terms you are keeping in the Fourier series), you will find that at k = 0 the sigma factor is 1, and the sigma factors fall off until at k = N they are 0. Thus they are another example of a window. The effecct of the Lanczos' window is to reduce the overshoot to about 0.01189, (by about a factor of 7), and the first minimum to 0.00473, (by about a factor of 10), which is a significant but not complee reduction of the Gibbs' phenomenon.

But back to my adventures in the matter. I knew, as you do, that at the discontinuity the truncated Fourier expansion takes on the mid-value of the two limits, one from each side. Thinking about the finite, discrete case, I reasoned that instead of taking all 1 values in the pass band and 0 values in the stop band, I should take 1/2 at the transition value. Lo, and behold, the transfer function becomes

$$[(\sin Nx)/2N\sin x/2]\cos x/2$$

and now has an extra factor (back in the rotational notation)

$$\cos(\pi f)$$

and the N + 1 in the sine term goes to N as well as the denominator N + 1 going to N. Clearly this transfer function is nicer than the Lanczos' as a low pass filter since it vanishes at the Nyquist frequency, and further dampens all the higher frequencies. I looked around in books on trigonometric series and found it in only one, Zygmund's two volume work where it was called the modified series. The extra "being prepared" did not necessarily pay off this time in a great result, but having found it myself I naturally reasoned that using even more modification of the coefficients of the Fourier series, (how much and where remained to be found), I might do even better. In short, I saw more clearly what "windows" were, and was slowly led to a closer examination of their possibilities.

A still third approach to the important the Gibbs' phenomena is via the problem of combining Fourier series. Let g(x) be, (and we are using the neutral variable x for a good reason)

$$g(x) = SUM[-\infty,\infty; \; c_k \exp\{ikx\}]$$

and another function be

$$h(x) = SUM[-\infty,\infty; \; d_m \exp\{imx\}]$$

The sum and difference of g(x) and h(x) are clearly the corresponding series with the sum or difference of the coefficients.

The product is another matter. Evidently we will have again a sum of exponentials, and setting n = k + m we will have the coefficients as indicated

$$g(x)h(x) = SUM[n=-\infty,\infty; \; SUM[k=-\infty,\infty; \; c_k d_{n-k}] \exp\{inx\}]$$

The coefficient of exp{inx}, which is a sum of terms, is called the convolution of the original arrays of coefficients.

In the case where there are only a few nonzero coefficients in the c_k coefficient array, for example, say symmetrically placed about 0, we will have for the coefficient

$$SUM[-K,K; \; c_k d_{n-k}]$$

and this we recognize as the original definition of a digital filter! Thus a filter is the convolution of one array by another, and that in turn is merely the multiplication of the corresponding functions! Multiplication on one side is convolution on the other side of the equation.

As an example of the use of this observation, suppose, as

often occurs, that there is potentially an infinite array of data, but that we can record only a finite number of them, (for example, turning on or off a telescope while looking at the stars). This function u_n is being looked at through the rectangular window of all 0's outside a range of $(2N + 1)$ 1's - the value 1 where we observe and the value 0 where we do not observe. When we try to compute the Fourier expansion of the original array from the observed data we must convolve the original coefficients by the coefficients of the window array,

$$\exp\{-iNx\} + \exp\{-i(N-1)x\} + \ldots + \exp(0) + \ldots + \exp\{iNx\}$$

Generally we want a window of unit area, so we need, finally, to divide by $(2N + 1)$. The array is a geometric progression with the starting value of $\exp\{-iNx\}$, and constant ratio of $\exp\{ix\}$,

$$\exp\{-iNx\}[1 - \exp\{i(2N+1)x\}]/[1 - \exp\{ix\}](2N+1) =$$

$$= \sin\{(N + 1/2)x\}/(2N+1)\sin(x/2)$$

At $x = 0$ this takes on the value 1, and otherwise oscillates rapidly due to the sine function in the numerator, and decays slowly due to the increase of the sine in the denominator (the range in x is $(-\pi,\pi)$. Thus we have the typical diffraction pattern of optics.

In the continuous case, before sampling, the situation is much the same but the rectangular window we look through has the transform of the general form (ignoring all details)

$$(\sin x)/x$$

and the convolution of a step function (a discontinuity) with it will, upon inspection, be the Gibbs' phenomena. Figure 15-2. Thus we see the Gibbs' phenomena overshoot in another light.

Some rather difficult trigonometric manipulation will directly convince you that whether we sample the function and then limit the range of observations, or limit the range and then sample, we will end up with the same result; theory will tell you the same thing.

The simple modification of the discrete Lanczos' window by changing only the outer two coefficients from 1 to 1/2 produced a much better window. The Lanczos' window with its sigma factors modified all the coefficients, but its shape had a corner at the ends, and this means, due to periodicity, that there are two discontinuity in the first derivative of the window shape - hence slow convergence. If we reason that using weights on the coefficients of the raw Fourier series of the form of <u>a raised cosine</u>

$$w_k = (1 + \cos \pi k/N)/2$$

then we will have something like the Lanczos' window but now there will be greater smoothness, hence more rapid convergence.

Writing this out in the exponential form we find that the weights on the exponentials are

$$1/4, \ 1/2, \ 1/4$$

This is the von Hann window - smoothing in the domain of the data with these weights is equivalent to windowing (multiplying) in the frequency domain. Actually I had rediscovered the von Hann window in the early days of our work in power spectra, and later John Tukey found that von Hann had used it long, long before in connection with economics. An examination of what it does to the signal shows that it tails off rapidly, but has some side lobes through which other parts of the spectrum "leak in".

We were at times dealing with a spectrum that had a strong line in it, and when looking elsewhere in the spectrum through the von Hann window its side lobes might let in a lot of power. The Hamming window was devised to make the maximum side lobe a minimum. The cost is that there is much more total leakage in the mean square sense, but a single strong line is kept under control. If you call the von Hann window a "raised cosine" with weights

$$1/4, \ \ 1/2, \ \ \ 1/4$$

the the Hamming window is a "raised cosine on a platform" with weights

$$0.23, \ \ \ 0.54, \ 0.23$$

(Figure 15-4). Actually the weights depend on N, the length of data, but so slightly that these constants are regularly used for all cases. The Hamming window has a mysterious, hence popular, aura about it with its peculiar coefficients, but it was designed to do a particular job and is not a universal solution to all problems. Most of the time the von Hann window is preferable. There are in the literature possibly 100 various windows, each having some special merit, and none having all the advantages you may want.

To make you a true insider in this matter I must tell you yet another story. I used to tease John Tukey that you are famous only when your name was spelled with a lower case letter such as watt, ampere, volt, fourier (sometimes), and such. When Tukey first wrote up his work on Power Spectra, he phoned me from Princeton and asked if he could use my name on the Hamming window. After some protesting on the matter, I agreed with his request. The book came out with the name "hamming"! There I am!

It must be your friends, in some sense, who make you famous by quoting and citing you, and it pays, so I claim, to be helpful to others as they try to do their work. They may in time give you credit for the work, which is better than trying to claim it yourself. Cooperation is essential in these days of complex projects; the day of the individual worker is dying fast. Team work is more and more essential, and hence learning to work in a
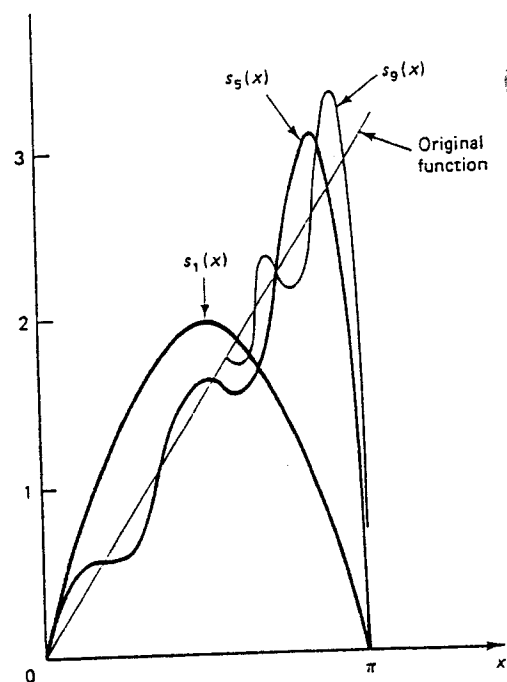
team, indeed possibly seeking out places where you can help others, is a good idea. In any case the fun of working with good people on imporant problems is more pleasure than the resulting fame. And the choice of imporatnt problems means that generally management will be willing to supply all the assistance you need.

In my many years of doing computing at Bell Telephone Laboratories I was very careful never to write up a result that involved any of the physics of the situation lest I get a reputation for "stealing other's ideas". Instead I let them write up the results, and if they wanted me to be a co-author, fine! Teamwork implies a very careful consideration for others and their contributions, and they may see their contributions in a differnt light than you do!
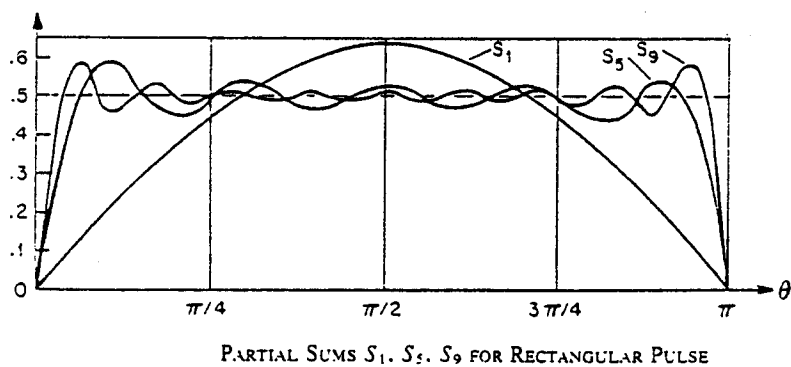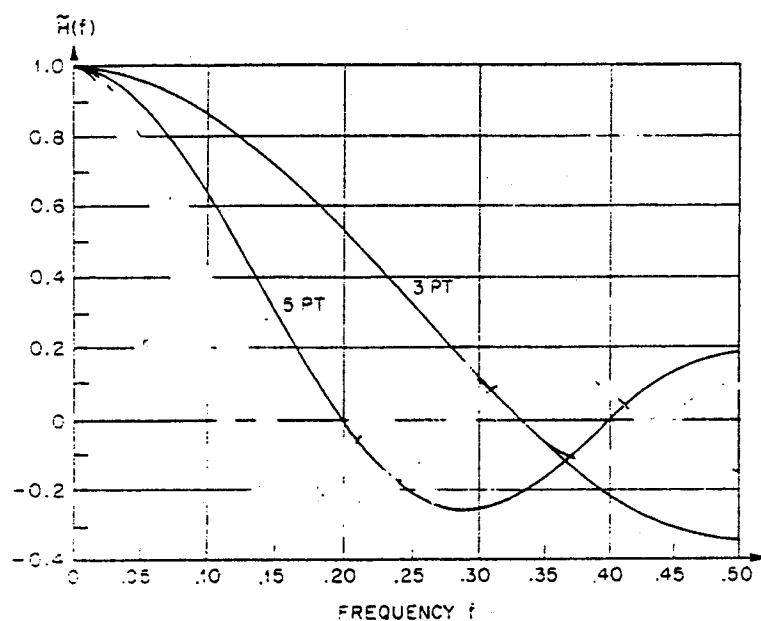
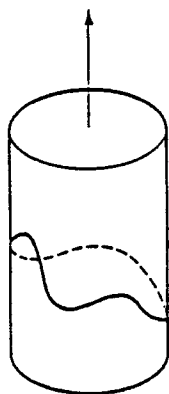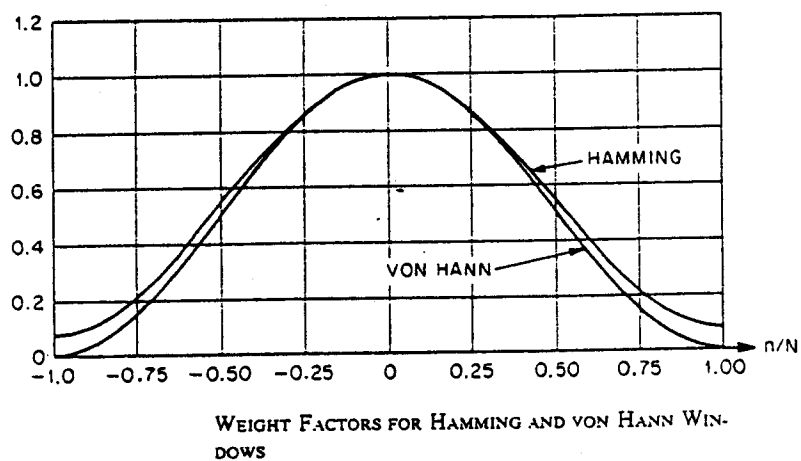PARTIAL SUMS S_1, S_5, S_9 FOR RECTANGULAR PULSE

Figure 15-2

Figure 15-3

WEIGHT FACTORS FOR HAMMING AND VON HANN WINDOWS

Figure 15-4