LECTURE 17

DIGITAL FILTERS -IV

We now turn to <u>recursive filters</u> which have the form

$$y_n = \text{SUM}[j=0,K;\ c_j u_{n-j}] + \text{SUM}[j=1,K;\ d_j y_{n-j}]$$

From this formula it will be seen that we have values on only one side of the current value n, and that we use both old and the current signal values, u_n, and old values of the outputs, y_n. This is classical, and arises from the fact that we are often processing a signal in real time and do not have access to future values of the signal.

But considering basics, we see that if we did have "future values" then a two sided prediction would probably be much more accurate. We would then, in computing the y_n values, face a system of simultaneous linear equations - nothing to be feared in these days of cheap computing. We will set aside this observation, noting only that often these days we record the signal on a tape or other media, and later process it in the lab - and therefore we have the future available now. Again, in picture processing, a recursive digital filter that used only data from one side of the point being processed would be foolish since it would not to use some of the available, relevant information.

The next thing we see is that the use of old output as new input means that we have <u>feedback</u> - and that automatically means questions of <u>stability</u>. It is a condition that we must watch at all times in the design of a recursive filter; it will restrict what we can do. Stability here means that the effects of the initial conditions do not dominate the results.

Being a linear system we see that whatever pure frequency we put into the filter when in the steady state, <u>only</u> that frequency can emerge, though it may be phase shifted. The transients, however, can have other frequencies which arise from the solution of the homogeneous difference equation. The fact is that <u>we are solving a difference equation with constant coefficients with the u_n terms forming the "forcing function"</u> - <u>that is exactly what a recursive filter is, and nothing else</u>.

We therefore assume for the steady state (which ignores the transients)

$$x_n = A_I \exp\{i\omega t\} \qquad y_n = A_O \exp\{i\omega t\}$$

(with the A's possibly complex to allow for the phase shift), and this leads, on solving for the ratio of A_O/A_I, to the transfer function

$$A_O/A_I = SUM[j=0,K;\ c_j exp\{-ij\omega\}/\{1 - SUM[j=1,K; d_j exp\{ij\omega\}\}$$

This is a <u>rational function</u> in the complex variable exp{iωt} = z rather than, as before with non-recursive filters, a polynomial in z. There is a theory of Fourier series representation of a function; there is not as yet a theory of the representation of a function as the ratio of two Fourier series (though I see no reason why there cannot be such a theory). Hence the design methods are at present not systematic, (as Kaiser did for the non-recursive filter design theory), but rather a collection of trick methods. Thus we have Butterworth, two types of Chebyshev (depending on having the equal ripples in the pass or the stop band, and elliptic filters (whose name comes from the fact that elliptic functions are used) that are equal ripple in both.

I will only talk about the topic of feedback. To make the problem of feedback graphic I will tell you a story about myself. One time long ago I was host of a series of six, one half hour, TV programs about computers and computing, and it was made mainly in San Francisco. I found myself out there frequently, and I got in the habit of staying always in the same room in the same hotel - it is nice to be familiar with the details of your room when you are tired late at night or when you may have to get up in the middle of the night - hence the desire for the same room.

Well, the plumber had put nice, large diameter pipes in the shower, Figure 17-1. As a result in the morning when I started my shower it was too cold, so I turned up the hot water knob, still too cool, so more, still too cool, and more, and then when it was the right temperature I got in. But of course it got hotter and hotter as the water that was admitted earlier finally got up the pipe and I had to get out, and try again to find a suitable adjustment of the knob. The <u>delay</u> in the hot water getting to me was the trouble. I found myself, in spite of many experiences, in the same classic hunting situation of instability. You can either view my response as being too strong, (I was too violent in my actions), or else the detection of the signal was too much delayed, (I was too hasty in getting into the tub). Same effect in the long run! Instability! I never really got to accept the large delay I had to cope with, hence I daily a minor trouble first thing in the morning! In this graphic example you see the essence of instability.

I will not go on to the design of recursive digital filters here, only note that I had effectively developed the theory myself in coping with corrector formulas for numerically solving ordinary differential equations. The form of the corrector in a predictor-corrector method is

$$y_{n+1} = SUM[j=0,K;\ c_j y_{n-j}] + SUM[j=0,K;\ d_j y'_{n-j}]$$

We see that the u_j of the recursive filter are now the derivatives y_n' of the output and come from the differential equation. In the standard nonrecursive filter there no feedback paths - the y_n that are computed do not appear later in the right hand side. In the differential equation formula they appear both in this feedback path and also through the derivative terms they form another, usually nonlinear, feedback path. Hence stability is a more difficult topic for differential equations than it is for recursive filters.

These recursive filters are often called "infinite impulse response filters" (IIR) because a single disturbance will echo around the feedback loop, which even if the filter is stable will die out only like a geometric progression. Being me, of course I asked myself if all recursive filters had to have this property, and soon found a counter example. True, it is not the kind of filter you would normally design, but it showed that their claim was superficial. If you will only ask yourself, "Is what I am being told really true?" it is amazing how much you can find that is, or borders on, being false, even in a well developed field!

In Lecture 26 I will take up the problem of dealing with the expert. Here you see a simple example of what happens all too often. The experts were told something in class when they were students first learning things, and at the time they did not question it. It becomes an accepted fact, which they repeat and never really examine to see if what they are saying is true or not, especially in their current situation.

Let me now turn to another story. A lady in the Math Dept. at Bell Telephone Laboratories was square dancing with a physicist one weekend at a party, and on Monday morning in the hallway she casually mentioned to me a problem he had. He was measuring the number of counts in a radioactive experiment at each of, as I remember, 256 energy level. It is called "the spectrum of the process". His problem was that he needed the derivative of the data.

Well, you know that: (a) the number of nuclear counts at a given energy is bound to produce a ragged curve, and (b) differentiating this to get the local slope is going to be a very difficult thing to do. The more I thought about her casual remark the more I felt that he needed real guidance - meaning me! I looked him up in the Bell Telephone Laboratories phone book and explained my interest and how I got it. He immediately wanted to come up to my office, but I was obdurate and insisted on meeting in his laboratory. He tried using his office, and I stuck to the lab. Why? Because I wanted to size up his abilities and decide if I thought his problem was worth my time and effort, since it promised to be a tough nut to crack. He passed the lab test with flying colors - he was clearly a very competent experimenter. He was at about the limit of what he could do - a week's run to get the data and a lot of shielding was around the radio-active source, hence not much we could do to get better data. Furthermore, I was soon convinced that although I knew little about the details, his experiment was important to physics as well as to Bell Telephone Laboratories. So I took on the problem. Moral: To the extent that you can choose, then work on problems that you think will be important.

Obviously it was a smoothing problem, and Kaiser was just teaching me the facts, so what better to do than take the experimentalist to Kaiser and get Kaiser to design the appropriate differentiating filter? Trouble immediately! Kaiser had always thought of a signal as a function of time, and the square of the area under the curve as the energy, but here the energy was the independent variable! I had repeated trouble with Kaiser over this point until I bluntly said, "All right, his energy is time and the measurements, the counts, is the voltage." Only then could Kaiser do it. The curse of the expert with their limited view of what they can do. I remind you that Kaiser is a very able man, yet his expertise, as so often happens to the expert, limited his view. Will you in your turn do better? I am hoping that such stories as this one will help you avoid that pitfall.

As I earlier observed, it is usually the signal that is in the lower part of the Nyquist interval of the spectrum and the noise is pretty well across the whole of the Nyquist interval, so we needed to find the cutoff edge between the meaningful physicist's signal and the flat white noise. How to find it? First, I extracted from the physicist the theoretical model he had in his mind, which was a lot of narrow spectral lines of gaussian shape on top of a broad gaussian shape, (I suspected Cauchy shapes, but did not argue with him as the difference would be too small, given the kind of data we had). So we modeled it, and he created some synthetic data from the model. A quick spectral analysis, via an FFT, gave the signal confined to the lowest 1/20 of the Nyquist interval. Second, we processed a run of his experimental data and found the same location for the edge! What luck! For once theory and practice agreed! We would be able to remove about 95% of the noise. Kaiser finally wrote for him a program that would, given the cutoff edge position wherever the experimenter chose to put it, design the corresponding filter. The program: (1) designed the corresponding differentiating filter, (2) wrote the program to compute the smoothed output, and then (3) processed the data through this filter without any interference from the physicist.

I later caught the physicist adjusting the cutoff edge for different parts of the energy data on the same run, and had to remind him that there was such a thing as "degrees of freedom", and that what he was doing was not honest data processing. I had much more trouble, once things were going well, to persuade him that to get the most out of his expensive data, he should actually work in the square roots of the counts as they had equal variances. But he finally saw the light and did so. He and Kaiser wrote a classic paper in the area, as it opened the door on a new range of things that could be done.

My contribution? Mainly, first identifying the problem, next getting the right people together, then monitoring Kaiser to keep him straight on the fact that filtering need not have exclusively to do with time signals, and finally, reminding them of what they knew from statistics, (or should have known and probably didn't).

It seems to me from my experience that this role is increasingly needed as people get to be more highly specialized and narrower and narrower in their knowledge. Someone has to keep the larger view and see to it that things are done honestly. I think that I came by this role from long a long education in the hands of John Tukey, plus a good basic grounding in the form of the universal tool of Science, namely Mathematics. I will talk in Lecture 23 about the nature of Mathematics.

Most signal processing is indeed done on time signals. But most digital filters will probably be designed for small, special purpose studies, not necessarily signals in time. This is where I ask for your future attention. Suppose when you are in charge of things at the top, you are interested in some data that shows past records of relative expenses of manpower to equipment. It is bound to be noisy data, but you would like to understand, in a basic sense, what is going on in the organization - what long term trends are happening - so slowly that people hardly sense them as they happen, but which never-the-less are fundamental to understand if you are to manage well. You will need a digital filter to smooth the data to get a glimpse of the trend, <u>if</u> it exists. You don't want to find a trend when it does not exist, but if it does you want to know pretty much what it has been, so you can project what it is likely to be in the near future. Indeed, you might want to observe, if the data will support it, any change in the slope of the trend. Some signals, such as the ratio of fire power to tonnage of the corresponding ship, need not involve time at all, but will tell you something about the current state of the Navy. You can, of course, also study the relationship as a function of time.

I suggest strongly, that at the top of your career you will be able to use a lot of low level digital filtering of signals, whether in time or not, so you will be better able to manage things. Hence, I claim, that you will probably design many more filters for such odd jobs than you will for radar data reduction and such standard things. It is usually in the new applications of knowledge where you can expect to find the greatest gains.

Let me supply some warnings against the misuse of intellectual tools, and I will talk Lecture 27 on topics closer to statistics than I have time for now. Fourier analysis implies linearity of the underlying model. You can use it on slightly nonlinear situations, but often elaborate Fourier analyses have failed because the underlying phenomena was too nonlinear. I have seen millions of dollars go down that drain when it was fairly obvious to the outsider that the nonlinearities would vitiate all the linear analysis that they could do using the Fourier function approach. When this was pointed out to them, their reply seemed to be that they did not know what else to do, so they persisted in doing the wrong thing! I am not exaggerating here.

How about nonlinear filters? The possibilities are endless, and must, of course, depend on the particular problem you have on hand. I will take up only one, the running median filter. Given a set of data you compute the running median, and that is the output. Consider how it will work in practice. First, you see that it will tend to smooth out any local noise - the median will be near the average, which is the straight line least squares fit used for local smoothing. But at a discontinuity, Figure 17-2, say we picture a flat level curve and then a drop to another flat curve, what will the filter do? With an odd number of terms in the median filter, you see that the output will stay up until you have more than half of the points on the lower level, where upon it will jump to the lower level. It will follow the discontinuity fairly well, and will not try to smooth it out completely! For some situations that is the kind of filtering you want. Remove the noise locally, but do not lose the sudden changes in the state of the system being studied.

I repeat, Fourier analysis is linear, and there exist many nonlinear filters, but the theory is not well developed beyond the running median. Kalmann filters are another example of the use of partially nonlinear filters, the nonlinear part being in the "adapting" itself to the signal.

One final basic observation that I made as I tried to learn digital filters. One day in examining a book on Fourier integrals, I found that there is a theorem which states that the variability of the function times the variability of its transform must exceed a certain constant. I said to myself, "What else is this than the famous uncertainty principle of Quantum Mechanics?" Yes, every linear theory must have an uncertainty principle involving conjugate variables. Once you adopt the linear approach, and QM claims absolute additivity of the eigenstates, then you must find an uncertainty principle. Linear time invariance leads automatically to the eigenfunctions $e^{i\omega t}$. They lead immediately to the Fourier integral, and Fourier integrals have the uncertainty principal. It is as if you put on blue tinted glasses; everywhere you look you must see things with a bluish tint! You are therefore not sure that the famous uncertainty principle of QM is really there or not; it may be only the effect of the assumed linearity. More than most people want to believe, what we see depends on how we approach the problem! Too often we see what we want to see, and therefore you need to consciously adopt a scientific attitude of doubting your own beliefs.

To illustrate this I will repeat the Eddington story of the fishermen. They used a net for fishing, and when they examined the size of the fish they had caught they decided that there is a minimum size to the fish in the sea.

In closing, if you do not, now and then, doubt accepted rules it is unlikely that you will be a leader into new areas; if you doubt too much you will be paralyzed and will do nothing. When to doubt, when to examine the basics, when to think for yourself, and when to go on and accept things as they are, is a matter of style, and I can give no simple formula on how to decide. You must learn from your own study of life. Big advances usually come from significant changes in the underlying beliefs of a field. As our state of knowledge advances the balances between aspects of doing research change. Similarly, when you are young then serendipity has probably a long time to pay off, but when you are old it has little time and you should concentrate more on what is at hand.

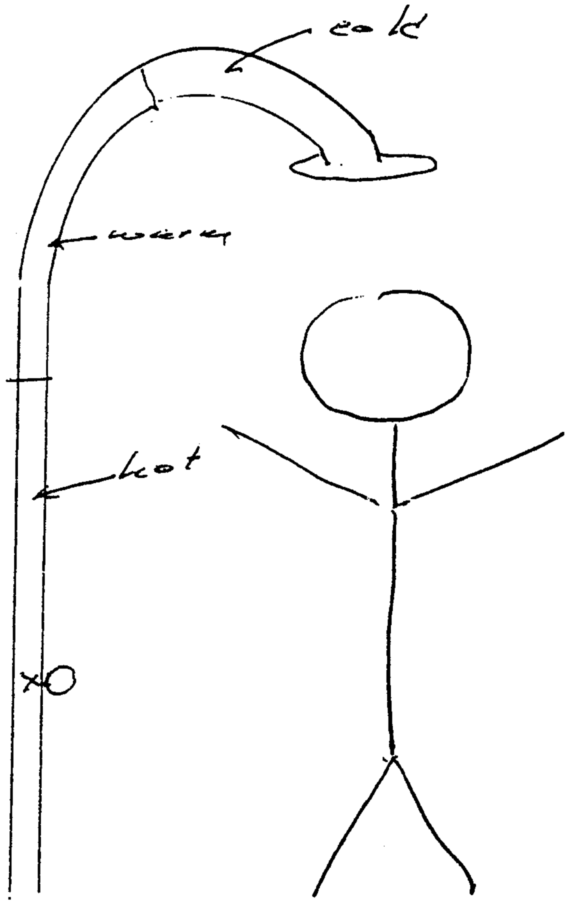

Figure 17-1

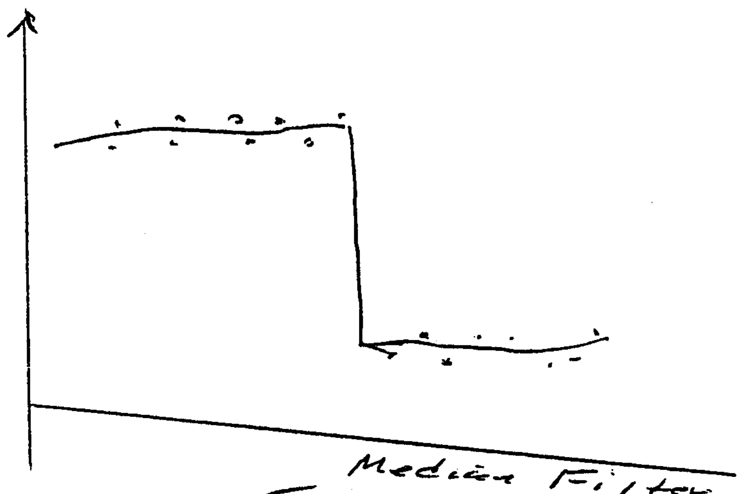

Figure 17-2