LECTURE 20

SIMULATION-III

I will continue the general trend of the last lecture, but center on the old expression "garbage in, garbage out", often abbreviated GIGO. The idea is that if you put ill-determined numbers and equations (garbage) in then you can only get ill-determined results (garbage) out. By implication the converse is tacitly assumed, if what goes in is accurate then what comes out must be accurate. I shall show that both of these assumptions can be false.

Because many simulations still involve differential equations we begin by considering the simplest first order differential equations of the form

$$y' = f(x,y)$$

You recall that a direction field is simply drawing at each point in the x-y plane a line element with the slope given by the differential equation. Figure 20-1. For example, the differential equation

$$y' = x^2 + y^2 \qquad y(0) = 1$$

has the indicated direction field. Figure 20-2. On each of the concentric circles

$$x^2 + y^2 = k$$

the slope is always the same, the slope depending on the value of k. These are called isoclines.

Looking at the following picture, Figure 20-3, the direction field of another differential equation, on the left you see a diverging direction field, and this means that small changes in the initial starting values, or small errors in the computing, will soon produce large differences in the values in the middle of the trajectory. But on the right hand side the direction field is converging, meaning that large differences in the middle will lead to small differences on the right end. In this single example you see both that small errors can become large ones, and that large ones can become small ones, and furthermore, that small errors can become large and then again become small. Hence the accuracy of the solution depends on where you are talking about it, not any absolute accuracy over all. The function behind all this is

$$y(x) = \exp(-x^2)$$

whose differential equation is, upon differentiating,

$$y'(x) = -2x \exp\{-x^2\} = -2xy(x).$$

Probably in your mind, you have drawn a "tube" about the "true, exact solution" of the equation, and seen that the tube expands first and then contracts. This is fine in two dimensions, but when I have a system of n such differential equations, 28 in the Navy intercept problem mentioned earlier, then these tubes about the true solutions are not exactly what you might think they were. The four circle figure in two dimensions, leading to the n-dimensional paradox by ten dimensions, Lecture 9, shows how tricky such imagining may become. This is simply another way of looking at what I said in earlier Lectures about stable and unstable problems; but this time I am being more specific to the extent that I am using differential equations to illustrate matters.

How do we numerically solve a differential equation? Starting with only one first order ordinarily differential equation of first degree, we imagine the direction field. Our problem is that from the initial value, which we are given, we want to get to the next nearby point. If we take the local slope from the differential equation and move a small step forward along the tangent line then we will make a only small error. Figure 20-4. Using that point we go to the next point, but as you see from the Figure we gradually depart from the true curve because we are always using "the slope that was", and not a typical slope in the interval. To avoid this we "predict" a value, use that value to evaluate the slope there, (use the diferential equation), and then use the average slope of the both ends to estimate the average slope to use, Figure 20-5. Then using this average slope we move the step forward again, this time using a "corrector" formula. If the predicted and corrected values are "close" then we assume that we are accurate enough, but if they are far apart then we we must shorten the step size. If the difference is too small then we should increase the step size. Thus the traditional "predictor-corrector" methods have built into them an automatic mechanism for checking the step-by-step error - but this step-by-step error is, of course, not the <u>whole accumulated error</u> by any means! The accumulated error clearly depends on the convergence or divergence of the direction field.

We used simple straight lines for both predicting and correcting. It is much more economical, and accurate, to use higher degree polynomials, and typically this means about fourth degree polynomials, (Milne, Adams-Bashforth, Hamming, etc.). Thus we must use several old values of the function and derivative to predict the next value, and then using this in the differential equation we get an estimated new slope, and with this slope plus using old values of the function and slope, we correct the value. A moment's thought and you see that the corrector is just a recursive digital filter where the input data are the derivatives, and the output values are the positions. Stability and all that we discussed there are relevant. As mentioned before, there is the extra feedback through the differential equation's predicted value that goes into the corrected slope. But both are simply solving a difference equation - recursive digital filters are simply this formula and nothing more. They are not just transfer functions as your course in digital filters might have made you think; plainly and simply, you are computing numbers coming from a difference equation. There is a difference however. In the filter you are strictly processing by a linear formula, but because in the differential equation there is the nonlinearity that arises from the evaluation of the derivative terms, it is not exactly the same as a digital filter.

If you have n differential equations then you are dealing with a vector with n components; you predict each component forward, evaluate each of the n derivatives, correct each predicted value, and finally take the step, or reject it if the error is too large in a sense you think fairly measures the local error. You tend to think about small errors as a "tube" surrounding the actual computed trajectory, but again you need to remember the four circle paradox that in a high dimension the "tubes" are not at all like you wish they were.

Now let me note a significant difference between the two approaches, numerical analysis and filter theory. The classical methods of numerical analysis, and still about the only one you will find in the accepted texts, use polynomials to approximate functions, but the recursive filter used frequencies as the basis for evaluating the formula! This is a different thing entirely!

To see this difference suppose we are to build a simulator for humans landing on Mars. The classical formulas will concentrate on the trajectory shape in terms of local polynomials, and the path will have small discontinuities in the acceleration as we move from interval to interval. In the frequency approach we will concentrate on getting the frequencies right and let the actual positions be what happen. Ideally the trajectories are the same; practically they can be quite different.

Which solution do you want? The more you think about it the more you realize that the pilot in the trainer will want to get the "feel" of the landing vehicle, and this seems to mean that the frequency response of the simulator should feel right to the pilot. If the position is a bit off, then the feedback control during landing can compensate for this, but if it feels wrong in the actual flight then the pilot is going to be bothered by the new experience which was not in the simulator. It has always seemed to me that the simulator should prepare the pilots for the actual experience as best we can (we cannot fake out for long the lower gravity of Mars), so that they will feel comfortable when the real event occurs, having experienced it many times in the trainer. But the fact is that we know far too little of what the pilot "feels" (senses). Does the pilot feel only the Fourier real frequencies, or maybe they also feel the decaying Laplace complex frequencies, (or should we use wavelets?). Do different pilots feel the same kinds of things? We need to know more than we apparently now do about this important design criterion.

The above is the standard conflict between the mathematician's and engineer's approaches. Each has a different aim in solving the differential equations (and in many other problems), and hence they get different results out of their calculations. If you are involved in a simulation then you see that there can be highly concealed matters that are important in practice, but that the mathematicians are unaware of and they will deny that the effects matter. But looking at the two trajectories I have crudely drawn, Figure 20-6, the top curve is accurate in position but the corners will give a very different "feel" than reality will, and the second curve will be more wrong in position but more right in "feel". Again, you see why I believe that the person with the insight into the problem must get deep inside the solution methods and not accept traditional methods of solution.

I now turn to another story about the early days of Nike guided missile testing. At this point they were field testing at White Sands what was called "the telephone pole tests". They were simply firings where the missile was to follow a preassigned trajectory, and at the last moment explode so that the whole would not come down outside the range and do great damage, rather that the parts would more gently fall to the ground in the range and supposedly do less harm. The object of the tests was to get realistic measurements of drag, lift, and other properties as functions of altitude and velocity, for purposes of settling the details of the design as well as for improving the design.

I found my friend back at the Labs wandering around the halls looking quite unhappy. Why? Because the first two of some six test shots have broken up in mid-flight and no one knew why. The delay meant that the data to be gathered to enable us to go to the next stage of design was not available and hence the whole project was in serious trouble. I observed to him that if he would give me the differential equations describing the flight I would put a girl on the job of hand calculating the solution, (big computers were not readily available in the late 40's). In about a week they delivered seven first order equations, and the girl was ready to start. But what are the starting conditions just before the trouble arose? (I did not in those days have the computing capacity to do the whole trajectory rapidly.) They didn't know! The telemetered data was not clear just before the failure. I was not surprised, and it did not bother me much. So we used the guessed altitude, slope, velocity, angle of attack, etc. one for each of the seven variables of the trajectory; one condition for each equation. Thus I had garbage in. But I had earlier realized the nature of the field trials being simulated was such that small deviations from the proposed trajectory would be corrected automatically by the guidance system! I was dealing with a strongly convergent direction field.

We found that both pitch and yaw were stable but that as each one settled down it threw more energy into the other; thus there was not only the traditional stability oscillations in pitch and yaw, but due to the rotation of the missile about its long axis there was a periodic transfer of increasing energy between them. Once the computer curves for even a short length of the trajectory were shown everyone realized immediately that they had forgotten the cross connection stability, and they knew how to correct it. Now that we had the solution they could then also read the hashed up telemetered data from the trials and check that the period of the transfer of energy was just about correct - meaning that they had supplied the correct differential equations to be computed. I had little to do except to keep the girl on the desk calculator honest and on the job. My real contribution was: (1) the realization that we could simulate what had happened, which is now routine in all accidents but was novel then, and (2) the recognition that there was a convergent direction field so that the initial conditions need not be known accurately.

My reason for telling you the story is to show you that GIGO need not be right. Another example comes from my earliest Los Alamos experience on bomb simulation. I gradually came to realize that behind the computation was fairly inaccurate data for computing <u>the equation of state</u>, which relates pressure to density, (and temperature which I will ignore for the moment). Data from high pressure labs, from estimates from earthquakes, from estimates from the density of the cores of stars, and finally from the asymptotic theory of infinite pressures were plotted as a set of points on a very large piece of graph paper, Figure 20-7. Then large French curves were used to draw a curve connecting the thinly scattered points. We then read this curve to 3 1/3 decimal places, meaning we guessed at a 5 or a 0 in the fourth place. We used those numbers to subtabulate a five digit table, and at places in the table to six digit numbers, which were then the official data for the actual computations we ran. I was at that time, as I earlier said, sort of a janitor of computing, and my job was to keep things going to free the physicists to do their job.

At the end of the war I stayed on at Los Alamos an extra six months, and one of the reasons was that I wanted to know how it was that such inaccurate data could have led to such accurate predictions for the final design. With, at last, time to think for long periods, I found the answer. In the middle of the computations we were using effectively second differences; the first differences gave the forces on each shell on one side, and the differences from the adjacent shells on the two sides gave the resultant force moving the shell. We had to take thin shells, hence we were differencing numbers which were very close to each other and hence the need for many digits in the numbers. But further examination showed that as the device goes off, any one shell went up the curve and possibly at least partly down again, so that any local error in the equation of state was approximately averaged out over its history. What was important to get from the equation of state was the curvature, and as already noted even that had only to be on the average correct. Hence garbage in, but accurate results out never-the-less!

These examples show what was loosely stated before; if there is feedback in the problem for the numbers used, then they need not necessarily be accurately known. Just as in H. S. Black's great insight of how to build feedback amplifiers, Figure 20-8, so long as the gain is very high only the one resistor in the feedback loop need be accurately chosen, all the other parts could be of low accuracy. From the Figure 20-8 you have the equation

$$\text{input} \qquad\qquad \text{output}$$

$$[y + (1/10)x](-10^9) = x$$

$$10^9 y = [-x - 10^8 x]$$

$$x = -10y/[1 + 10^{-8}]$$

We see that almost all the uncertainty is in the one resistor of size 1/10, and the gain of the amplifier, (-10^{-9}), need not be accurate. Thus the feedback of H. S. Black allows us to accurately build things out of mostly inaccurate parts.

You see now why I cannot give you a nice, neat formula for all situations; it must depend on how the particular quantities go through the whole of the computation; the whole computation must be understood as a whole Do the inaccurate numbers go through a feedback situation where their errors will be compensated for, or are they vitally out in the open with no feedback protection? The word "vitally" because it is vital to the computation, if they are not in some feedback position, to get them accurate.

Now this fact, once understood, impacts design! Good design protects you from the need for too many highly accurate components in the system. But such design principles are still, to this date, ill-understood and need to be researched extensively. Not that good designers do not understand this intuitively, merely that it is not easily incorporated into the design methods that you were taught in school. Good minds are still needed inspite of all the computing tools we have developed. But the best mind will be the one who gets the principle into the design methods taught so that it will be automatically available for lesser minds!

I now look at another example, and the principle that enabled me to get a solution to an important problem. I was given the differential equation

$$y'' = \sinh y - kx, \quad (0.1 < k < 10), \quad y(0) = 0, \quad y(\infty) \sim \ln 2kx$$

You see immediately that the condition at infinity is really the right hand side of the differential equation equated to 0. Figure 20-9.

But consider the stability. If the y at any fairly far out point gets a bit too large, then the sinh y is much too large, the second derivative is then very positive, and the curve shoots off to plus infinity. Similarly, if the y is too small the curve shoots off to minus infinity. And it does not matter which way you go, left to right, or right to left. In the past I had used the obvious trick when facing a divergent direction field of simply integrating in the opposite direction and you get an accurate solution. But in the above problem you are, as it were, walking the crest of a sand dune, and once both feet are one side of the crest you are bound to slip down.

You can probably believe that while I could find a decent power series expansion, and an even a better non-power series approximate expansion around the origin, still I would be in trouble as I got fairly well along the solution curve, especially for large k. All the analysis I, or my friends, could produce was inadequate. So I went to the proposers and first objected to the condition at infinity, but it turned out that the distance was being measured in molecular layers, and (in those days) any realistic transistor would have effectively an infinity number of layers. I objected then to the equation itself; how could it represent reality? They won again, so I had to retreat to my office and think.

It was an important problem in the design and understanding of the transistors then being developed. I had always claimed that if the problem was important and properly posed then I could get some kind of a solution. Therefore, I must find the solution; I had no escape if I were to hold on to my pride.

It took some days of mulling it over before I realized that the very instability was the clue to the method to use. I would track a piece of the solution, using the differential analyzer I had at that time, and if the solution shot up then I was a bit too high in my guess at the corresponding slope, and if it shot down I was a bit too low. Thus piece by piece I walked the crest of the dune, and each time the solution slipped on one side or the other I knew what to do to get back on the track. Yes, having some pride in your ability to deliver what is needed is a great help in getting important results under difficult conditions. It would have been so easy to dismiss the problem as insoluble, wrongly posed, or any other excuse you wanted to tell yourself, but I still believe that important problems properly posed can be used to extract some useful knowledge that is needed. A number of space charge problems I have computed showed the same difficult instability in either direction.

I need to introduce for the next story the idea of a Rorschach test which was popular in my youth. A blob of ink is put on a piece of paper, it is squeezed on a fold, and when it is opened you have a symmetric blot with essentially a random shape. A sequence of these blots is shown to the subject and they are asked to report on what they see. Their answers were used to analyse the "personality" of the person. Obviously what a person reports is a figment of their imagination since the blot is essentially a random shape. It is like watching the clouds in the sky and discussing what the shapes they resemble; it is your imagination and not reality that you are discussing, and as such it is, to some extent, revealing things about yourself and not about the clouds. I believe the ink blot method is no longer in use.

Now to the next story. A psychologist friend at Bell Telephone Laboratories once built a machine with about 12 switches and a red and a green light. You set the switches, pushed a button, and either you got a red or a green light. After the first person tried it twenty times they wrote a theory of how to make the green light come on. The theory was given to the next victim and they had their twenty tries and wrote their theory, and so on endlessly. The stated purpose of the test was to study how theories evolved.

But my friend, being the kind of person he was, had connected the lights to a random source! One day he observed to me that no person in all the tests (and they were all high class Bell Telephone Laboratories scientists) ever said that there was no message. I promptly observed to him that not one of them was either a statistician or an information theorist, the two classes of people who are intimately familiar with randomness. A check revealed that I was right!

This is a sad commentary on your education. You are lovingly taught how one theory was displaced by another, but you are seldom taught to replace a nice theory with nothing but randomness! And this is what was needed; the ability to say that the theory you just read is no good and that there was no definite pattern in the data, only randomness.

I must dwell on this point. Statisticians regularly ask themselves, "Is what I am seeing really there, or is it merely random noise?" They have tests to try to answer these questions. Their answer is not a yes or no, but only with some confidence a "yes" or "no". A 90% confidence limit means that typically in ten trys you will make the wrong decision about once, <u>if all the other hypotheses are correct!</u>. Either you will chose when there is nothing there, (Type 1 error), or you will reject when there is something there, (Type 2 error). Much more data is needed to get to the 95% confidence limit, and these days data can often be very expensive to gather. Getting more data is also time consuming so that the decision is further delayed - a favorite trick of people in charge who don't want to bear the responsibility of their position - "Get more data", they say.

Now I suggest to you quite seriously, many simulations are nothing more than Rorschach tests. I quote a distinguished practioneer of management decision theory, Jay Forrester, "From the behavior of the system, doubts will arise that will call for a review of the original assumptions. From the process of working back and forth between assumptions about the parts and the observed behavior of the whole, we improve our understanding of the structure and dynamics of the system. This book is the result of several cycles of re-examination and revision by the author".

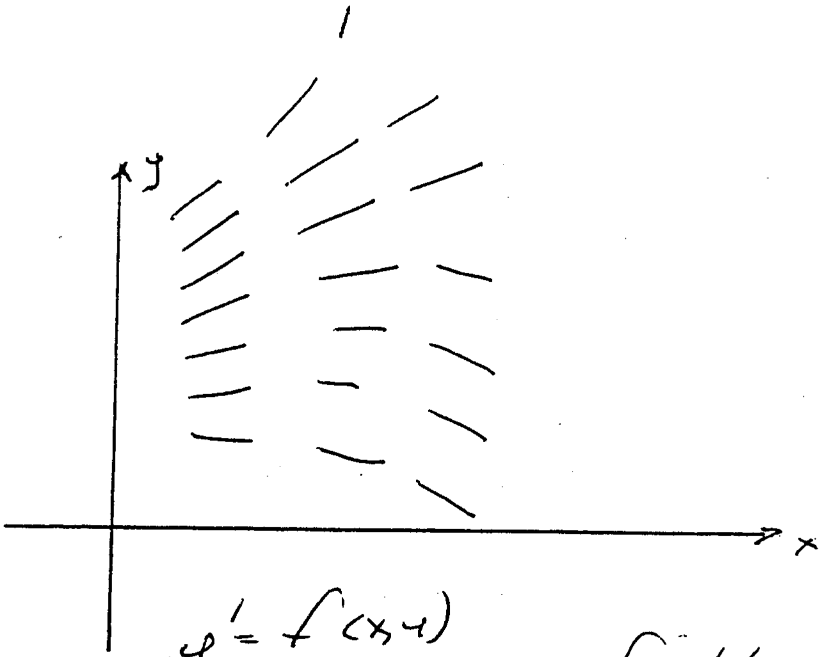
How is the outsider to distinguish this from a Rorschach test? Did he merely find what he wanted to find, or did he get at "reality"? Regrettably, many, many simulations have a large element of this adjusting things to get what they want to get.

It is so easy a path to follow. It is for this reason that traditional Science has a large number of safeguards, which these days are often simply ignored.

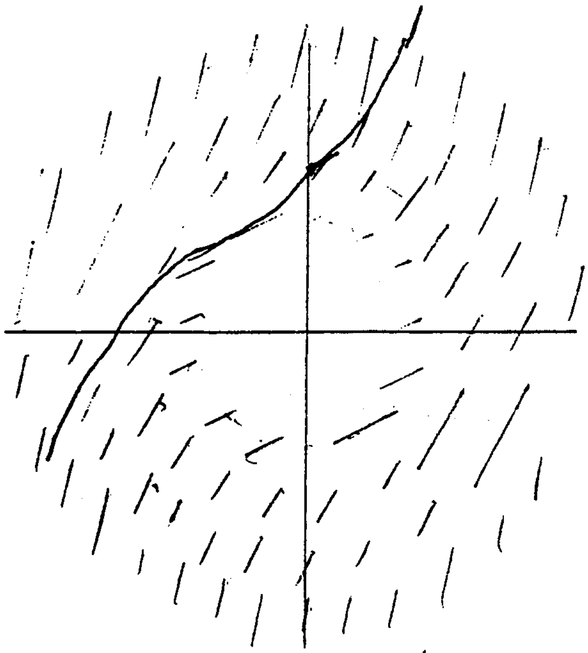
Do you think you can do things safely, that you know better? Consider the famous double blind experiments that are usual in medical practice. The doctors first found that if the patients thought they were getting the new treatment then they responded with better health, and those who thought they were part of the control group felt they were not getting it and did not improve. The doctors then randomized the treatment and gave some patients a placebo so that the patient could not respond and fool the doctors this way. But to their horror, the doctors also found that the doctors, knowing who got the treatment and who did not, also found improvement where they expected to and not where they did not. As a last resort, the doctors have widely accepted the double blind experiment - until all the data are in neither the patients nor the doctors know who gets the treatment and who does not. Then the statistician opens the sealed envelop and the analysis is carried out. The doctors wanting to be honest found that they could not be! Are you so much better in doing a simulation that you can be trusted not to find what you want to find? Self-delusion is a very common trait of humans.

I started in Lecture 19 with the problem of why anyone should believe in a simulation that has been done. You now see the problem more clearly. It is not easy to answer unless you have taken a lot more precautions than are usually done. Remember also you are probably going to be on the receiving end of many simulations to decide many questions that will arise in your highly technical future; there is no other way than simulations to answer the question "What if ... ?" In Lecture 18 I observed that decisions must be made and not postponed forever if the organization is not to flounder and drift endlessly - and I am supposing that you are going to be among those who must make the choices. Simulation is essential to answer the "What if ... ?", but it is full of dangers, and is not to be trusted just because a large machine and much time has been used to get the nicely printed pages, or colorful pictures on the oscilloscope. If you are the one to make the final decision then in a real sense you are responsible. Committee decisions, which tend to diffuse responsibility, are seldom the best in practice - most of the time they represent a compromise which has none of the virtues of any path and they tend to end in mediocrity. Experience has taught me that generally a decisive boss is better than a waffling one - you know where you stand and can get on with the work that needs to be done!

The "What if ... ?" will arise often in your futures, hence the need for you to master the concepts and possibilities of simulations, and be ready to question the ressults and to dig into the details when necessary.



$y' = f(x,y)$
direction field
Figure 20-1



$y' = x^2 + y^2 \quad y(0) = 1$
Direction field
Figure 20-2

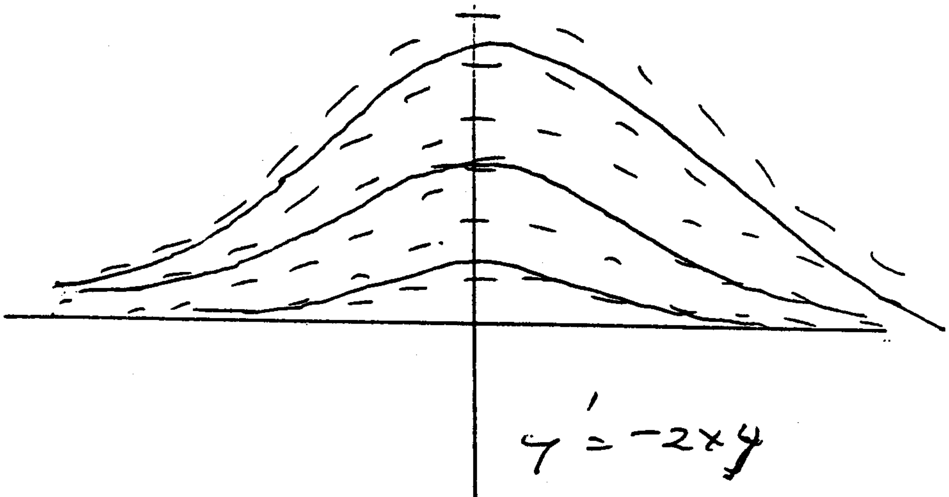


Figure 20-3

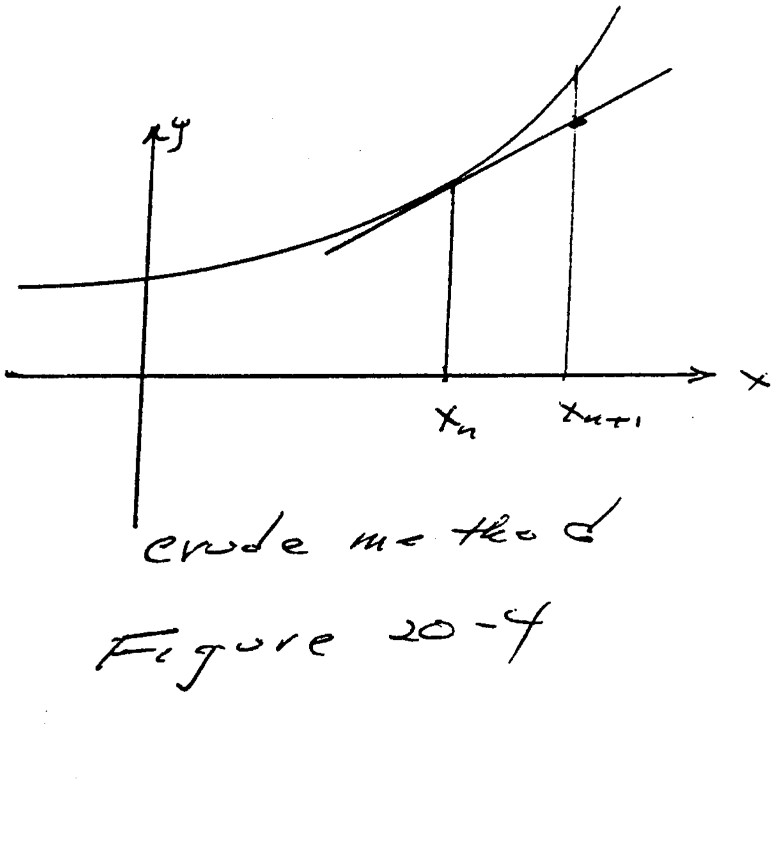
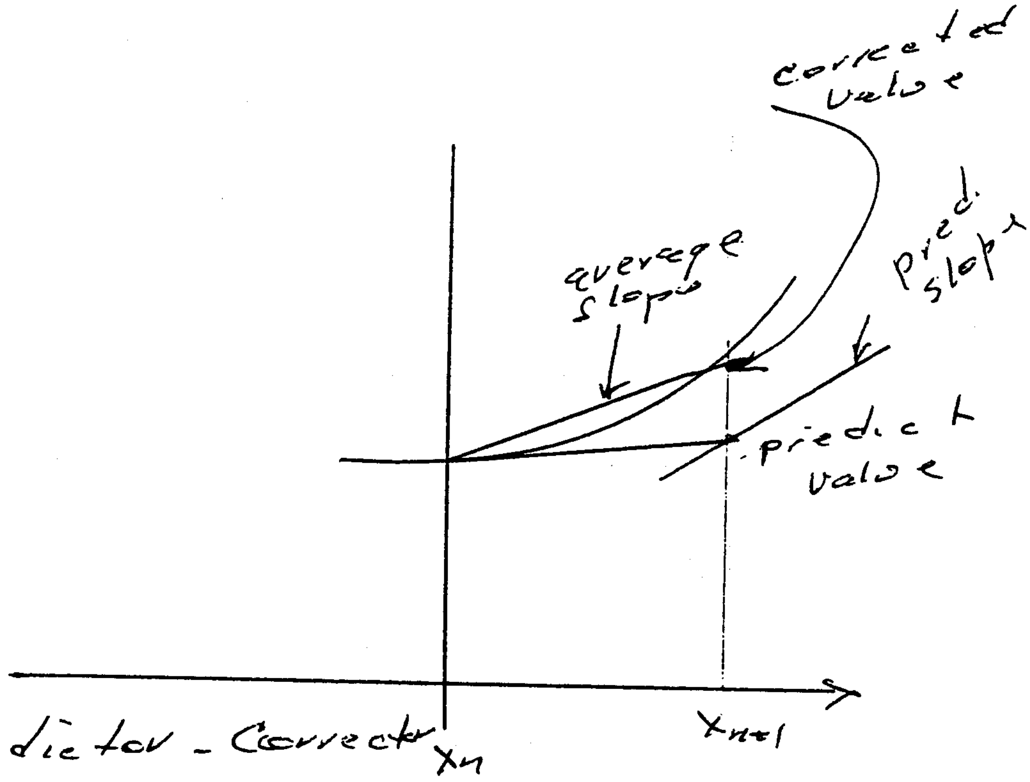


Figure 20-4



Predictor - Corrector

Figure 20-5

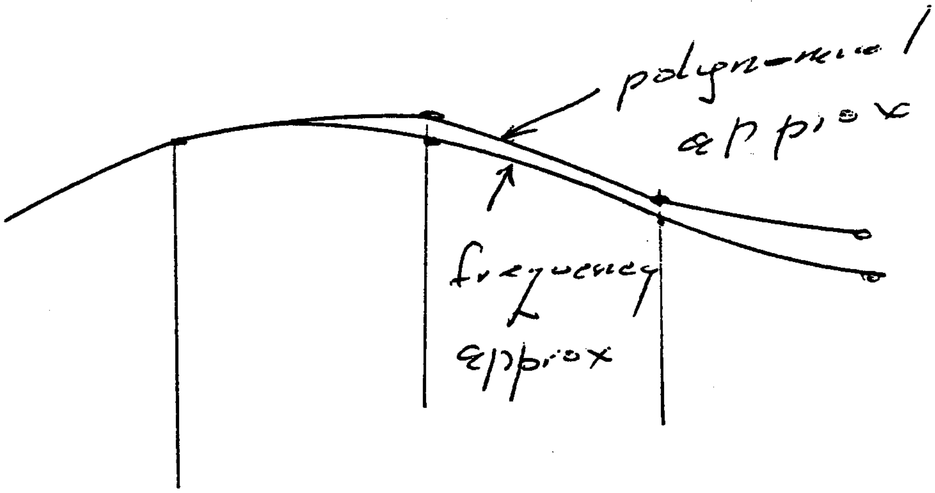
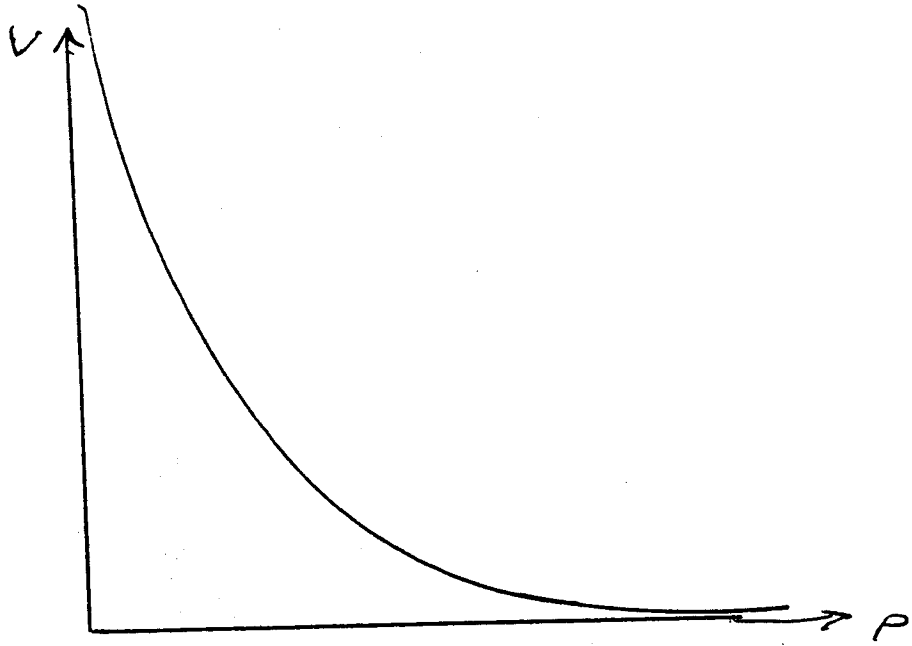
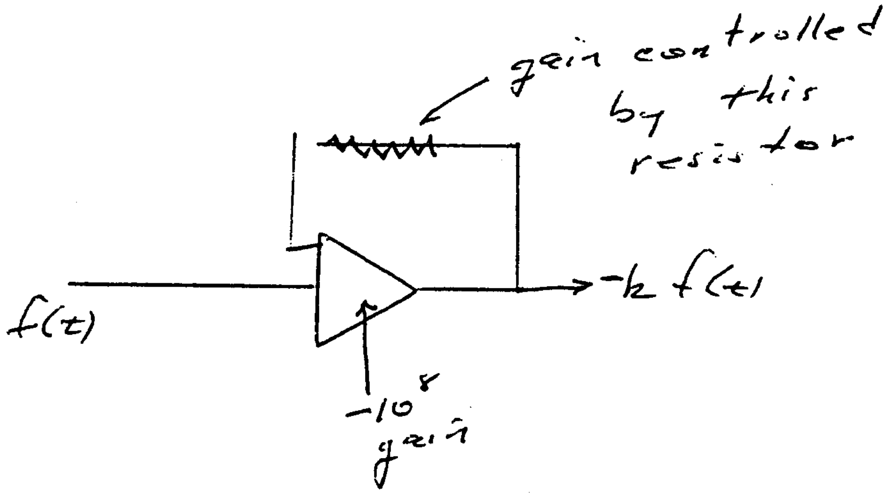


Figure 20-6



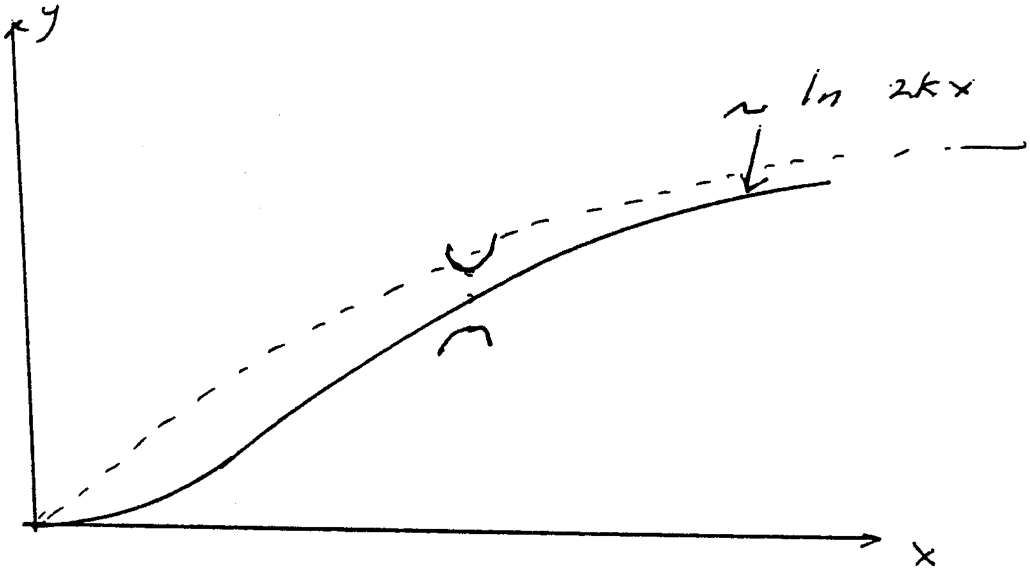
Equation of State

Figure 20-7



H. S. Black's feedback

Figure 20-8



$$\varphi'' = \sinh \varphi - kx$$

Figure 20-9