Subject: LECTURES ON SAMPLED DATA SYSTEMS ANALYSIS
From: R. W. Sittler
To: V. A. Nedzel

Memorandum No. 2M-0671, Massachusetts Institute of Technology, Lincoln Laboratory
The research reported in this document was supported jointly by the Department of the Army, the Department of the Navy, and the Department of the Air Force under Air Force Contract No. AF 19(122)-458.

Signed R W. Sittler
Author

Date: 22 August 1957
Rec'd 22 August 1957

RWS:jhw
Reissued February 1958

Edited by Paul Mennen
First edition 03-Nov-03

Forward

Although it appears that Sittler put these notes in the public domain, the original included hand drawn figures and at least in my copy were not always clear. I republished these notes (with Sittler's permission) to make them more readable and accessible - both because of their historic significance and because I don't think any modern author has covered the same topics as eloquently. Sittler's feat is even more remarkable considering that, unlike today's authors, he did not have decades of research in digital signal processing from which to draw.

Sittler chose z to represent a unit delay and z^{-1} to represent a unit advance - the opposite of current convention. Dealing with either convention is not difficult so don't let this dissuade you from exploring Sittler's notes and problem sets. They are an excellent introduction to sampled-data system theory. Apart from z vs. z^{-1}, Sittler's notation and terminology align well with more recent texts. (Actually geo-physicists and economists, like Sittler, still use z for a unit delay – a choice you may even decide is the wiser.)

*It's possible I have introduced some errors in transcribing this work. If you discover any such errors, no matter how trivial, please let me know so future readers don't have to puzzle over them. You can always find me at **paul@mennen.org**.*

Thanks,
Paul Mennen

Contents

1. Orientation and definitions

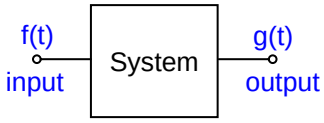
1.1 Systems analysis

By a system we usually mean a complex mechanism that has been built up by interconnecting simpler components. These components generally have a simple behavior; but when they are interconnected they can influence each other so that the overall system has an interestingly complex behavior.

A system can be a physical mechanism like an automobile automatic transmission (mechanical) system, or a television receiver (electrical, mechanical) system. Or it can be an abstract system that has been mentally constructed to make the solution of a mathematical problem easier to obtain and appreciate. In this latter case, instead of a system that operates on physical quantities, such as shaft rotations or electrical voltage, we may have one that operates on non-physical quantities such as probabilities.

Some of these system quantities will be under external control; others will be observable, but not controllable. We are normally interested in the behavior of system quantities as functions of time when we have some of them under external control.

It will be sufficient in the following to limit ourselves to two system quantities, one that we control and the other that we observe. The first we call the input to the system and designate it as a function of time by f(t). The second we call the output or response and designate it by g(t). The system provides the connecting mechanism between these quantities. We represent the system by a box between two lines showing that we apply f(t) at the input point.



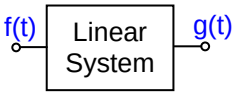
The problem of systems analysis is to describe mathematically how the system produces g(t) from f(t). Often the output is the unknown, which we desire to find given the input. However, we may also seek to find the input that produces a given output or to find the system that will convert a given input to a given output.

Often we speak of a system as a filter because we conceive of its job as being the removal from the input, f(t), of unwanted noise. For example, f(t) may be the sum of a desired function $f_d(t)$ and an undesired function $f_u(t)$. Then we would like to discover the filter system which converts $f(t)=f_d(t)+f_u(t)$ to $g(t)=f_d(t)$.

1.2 Linear systems

Only in special circumstances do our systems analysis problems become mathematically tractable. To make any progress at all, we must restrict the class of systems that we will consider, so we treat only linear systems.

A linear system is one having a superposition property as follows: Suppose we first apply $f_1(t)$ as the input and find that $g_1(t)$ is the response. Now what would happen if we had used $af_1(t) + bf_2(t)$ as the input where a and b are constants? In general, if the system is not linear, the response to this combination or superposition of inputs could be quite arbitrary. But a linear system has the defining property that the superposition of inputs $af_1(t) + bf_2(t)$ produces an output which is a corresponding superposition $ag_1(t) + bg_2(t)$.



Input f(t)	*Output g(t)*
$f_1(t)$	$g_1(t)$
$f_2(t)$	$g_2(t)$
$af_1(t)+bf_2(t)$	$ag_1(t)+bg_2(t)$

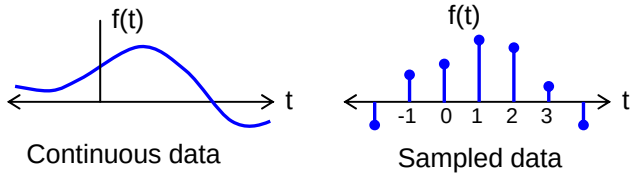
(a and b are constants)

By definition this property must hold for any $f_1(t)$, $f_2(t)$, a and b.

1.3 Sampled-data systems

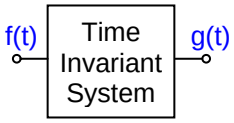
We will further restrict ourselves by considering linear systems that don't operate continuously in time but only at discrete instants. The system quantities change abruptly from value to value as the system is indexed in time. One can visualize the system as operating in jerks rather than smoothly. For this class of systems f(t) and g(t) are defined only at common discrete times, which we may take to be t = 0, ±1, ±2, ±3, ... To make the notation clearer, we will write f(n) and g(n) or f_n and g_n instead of the original f(t), g(t) where n takes on all integral values.

Since the input and output are described in terms of discrete samples these systems are called sampled-data systems. The mathematical analysis of such systems proceeds in essentially the same way as for continuous data systems, but the details are simpler and the physical significance of the mathematics is easier to appreciate. Also there are important applications for sampled data systems analysis.



1.4 Time-invariant systems

There is still another restriction we must make to simplify our considerations; that is, we treat only time-invariant systems. A system is time-invariant if its behavior does not depend on when the input is applied. More precisely, if we know that $g_1(t)$ is the response to $f_1(t)$, then the defining property of a time-invariant system is that $g_1(t+\tau)$ must be the response to $f_1(t+\tau)$ where τ is any constant. We insist that this property be true for any choice of f(t) and τ.



Input f(t)	*Output g(t)*
$f_1(t)$	$g_1(t)$
$f_1(t+\tau)$	$g_1(t+\tau)$

For sampled-data systems, of course, we restrict t and τ to integral values.

A linear system need not be time-invariant and vice-versa. In the following, we will consider only linear, time-invariant, sampled-data systems.

1.5 Problems

1.5.1) A system relates its input f(t) and output g(t) according to the formula

$$g(t) = \int_0^t e^{-u} f(t-u)\, du$$

Is this system linear, time-invariant? What is the effect of changing the system to raise the upper limit of the integral to ∞ ?

1.5.2) Suppose f(t) and g(t) are related by the differential equation

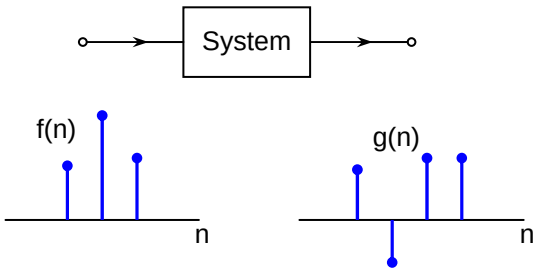
$$\frac{d^2 g(t)}{dt^2} + (ae^{-t} + 1)\frac{dg(t)}{dt} + bg^2(t) = f(t)$$

What is the character of the system if a = b = 0 ?
If a = 0, b = 1 ?
If a = 1, b = 0 ?
If a = b = 1 ?

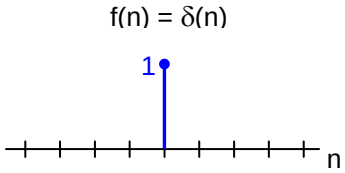
2. Impulse response and superposition

2.1 Impulse response

From now on we discuss only linear, time-invariant, sampled-data systems. The input f(n), and output g(n), are sequences of values defined for positive and negative integral values of n. We think of n as measuring units of time.



The simplest possible input consists of a unit sample at n=0 and zero sample values elsewhere for f(n). We will call this special input function δ(n).



The output function, g(n) for such an input is of such fundamental importance that we give it a special name and reserve a special notation for it. This function is called the impulse response (or unit-sample response) and is denoted by h(n).

The question may arise as to whether h(n) depends only on the system mechanism or whether, in fact, various outputs can occur with a unit-sample input depending on the past history of the system. But this behavior is implicitly excluded by our definition of system linearity, for if we apply an identically zero input to any linear system,

$$f(n) = 0 = 0 \cdot 0$$

the output is likewise identically zero.

$$g(n) = 0 \cdot (response\ \ to\ 0\ input) = 0$$

But, if the system can respond to δ(n) with possible outputs $h_1(n)$, $h_2(n)$, then it must respond to δ(n)–δ(n) with the output $h_1(n)–h_2(n)$. From what we have shown above, $h_1(n)=h_2(n)$.

We summarize by saying that for our system

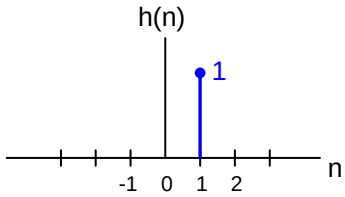
1. If f(n)=δ(n), then g(n)=h(n)

2. If f(n)=0, then g(n)=0

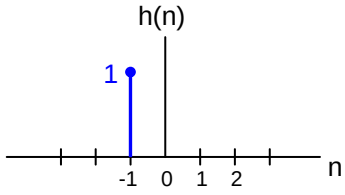
and that the form of h(n) depends on the construction of the system.

2.2 Physical realizability

We illustrate our remarks by finding the impulse response of two different systems. System #1 operates so that any input value is simply delayed one unit of time before it appears at the output. It apparently has the impulse response



System #2 operates in a similar way except that the output occurs one unit of time before the corresponding input value. The impulse response is



Both of these systems are linear. (Prove this.) But the second represents a perfect predictor; its output can occur before the input causing it is applied. A purely physical system could not operate in this way.

We call systems in which the impulse is zero for negative n physically realizable systems since physical examples of such systems can be constructed. However, we won't restrict our considerations to such systems since there are many examples of non-physical problems that lead to physically unrealizable systems.

2.3 Superposition

If we know h(n) for a system, then we can determine the output for an arbitrary input. To do this we must make use of the properties of linearity and time-invariance.

First of all we note that we can write any input as the sum of unit samples each of which has been shifted in time by a certain number of units and has been multiplied by a certain constant.

$$f(n) = f(0)\delta(n) + f(1)\delta(n-1) + f(2)\delta(n-2) + \ldots + f(-1)\delta(n+1) + f(-2)\delta(n+2) + \ldots$$

$$= \sum_{k=-\infty}^{+\infty} f(k)\,\delta(n-k)$$

By definition the response to δ(n) is the impulse response, h(n). Using time invariance, the response to δ(n–k) is h(n–k). Using linearity the response to f(k) δ(n–k) is f(k) h(n–k). Again the response to a sum of inputs is the sum of responses to the separate inputs. Therefore, our output must be

$$g(n) = \sum_{k=-\infty}^{+\infty} f(k)\,h(n-k)$$

(The extension of our definition of linearity to infinite as well as finite sums is needed to rigorously justify this conclusion.) This summation is called the superposition summation. It completely solves the problem of determining g(n) from f(n) when the impulse response, h(n), is known.

The particular summation process used here is often referred to as the convolution of f(n) and h(n) to g(n) and is indicated by the notation

$$g(n) = f(n) * h(n)$$

The convolution summation can be written in a new form by introducing the change of variable, k=n–m. Since the range of k is from –∞ to +∞ for any n, the range of m must he the same.

$$g(n) = \sum_{m=-\infty}^{+\infty} f(n-m)\,h(m)$$

$$g(n) = \sum_{k=-\infty}^{+\infty} h(k)\;f(n-k)$$

(We have substituted k for m as the dummy variable.) We see that this last form is the same as the original except that the roles of h(n) and f(n) are reversed. Hence we write:

$$g(n) = \sum_{k=-\infty}^{+\infty} f(k)h(n-k) = \sum_{k=-\infty}^{+\infty} h(k)\,f(n-k)$$

or

$$g(n) = f(n) * h(n) = h(n) * f(n)$$

2.4 Superposition (continued)

The actual superposition summation may be performed algebraically, or graphically by two methods. The best method to use depends on the circumstances of the problem. As an example

$$f(n) = \begin{cases} 0 & n < 0 \\ (1/2)^n & n \geq 0 \end{cases}$$

$$h(n) = \begin{cases} 0 & n < 0 \\ (1/3)^n & n \geq 0 \end{cases}$$

Proceeding algebraically,

$$g(n) = \sum_{k=-\infty}^{+\infty} f(k)\,h(n-k)$$

$$= \begin{cases} \sum_{k=0}^{n} (\tfrac{1}{2})^k (\tfrac{1}{3})^{n-k} & n \geq 0 \\ 0 & n < 0 \end{cases}$$

The sum is

$$\sum_{k=0}^{n} (\tfrac{1}{2})^k (\tfrac{1}{3})^{n-k} = (\tfrac{1}{3})^n \sum_{k=0}^{n} (\tfrac{3}{2})^k$$

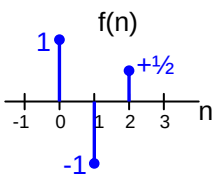
$$= (\tfrac{1}{3})^n \frac{1-(\tfrac{3}{2})^{n+1}}{1-\tfrac{3}{2}} = 3(\tfrac{1}{2})^n - 2(\tfrac{1}{3})^n$$

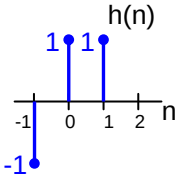
Therefore

$$g(n) = \begin{cases} 0 & n < 0 \\ 3(\tfrac{1}{2})^n - 2(\tfrac{1}{3})^n & n \geq 0 \end{cases}$$

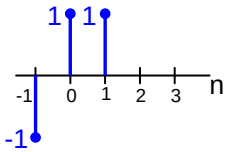
The alternate summation for g(n), of course, gives the same result. (Try it.)

To illustrate the graphical methods we take a second example.

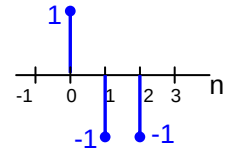




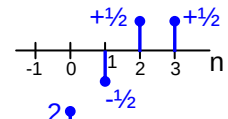
The first way to find g(n) is to graphically form the superposition summation as we did in its derivation. We plot the separate responses to single samples at n=0, ±1, ±2, ... of values f(0), f(±l), ... and add the results. Our example gives:



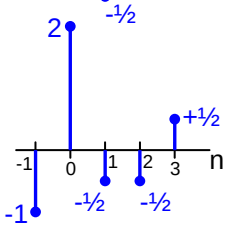
response to f(0)δ(n)



response to f(1)δ(n-1)

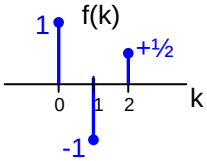


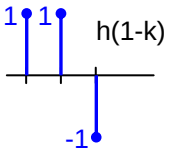
response to f(2)δ(n-2)



g(n)

In the second method we fix n in the sum $\sum_{k=-\infty}^{+\infty} f(k)h(n-k)$ and interpret it graphically. To form this sum we first plot f(k). Then beneath f(k) we plot h(n–k) which requires us to reflect the values of h(k) about the origin and then to slide them n units to the right. For our example with n=1 these plots are





Then we multiply corresponding samples above and below and add all these products. Our example gives $g(1) = (1 - 1 - \frac{1}{2}) = -\frac{1}{2}$. To find other values of g(n) we slide the reflected h(k) by differing amounts.

We describe this process as convolving h(k) with f(k). Our previous two forms of writing the superposition summation suggest that we may interchange the role of f(k) and h(k) in this process. (Check this with the example.)

2.5 Problems

2.5.1) The input, f(t), and output, g(t), of a continuous data system are governed by the differential equation:

$$\frac{dg(t)}{dt} + g(t) = f(t)$$

Verify that the system is linear and time invariant if we suitably restrict the solution, g(t), by omitting any part of the solution of the homogeneous equation. Verify that the particular solution is given by the superposition integral (analogous to superposition summation).

$$g(t) = \int_{-\infty}^{t} f(\tau) e^{-(t-\tau)} d\tau$$

2.5.2) Repeat the above considerations for a sampled-data system governed by the difference equation

$$g(n) - \tfrac{1}{2}\, g(n-1) = f(n)$$

Verify that a particular solution is

$$g(n) = \sum_{k=-\infty}^{n} f(k) \left(\tfrac{1}{2}\right)^{n-k}$$

What is the impulse response, h(n)? Is the system physically realizable?

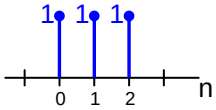
2.5.3) Find (algebraically) the response of a system with impulse response

$$h(n) = \begin{cases} \left(-\tfrac{1}{2}\right)^n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

to an input of the form

$$f(n) = \begin{cases} n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

2.5.4) Find graphically (two methods) the response of a system whose input and impulse response both have the form

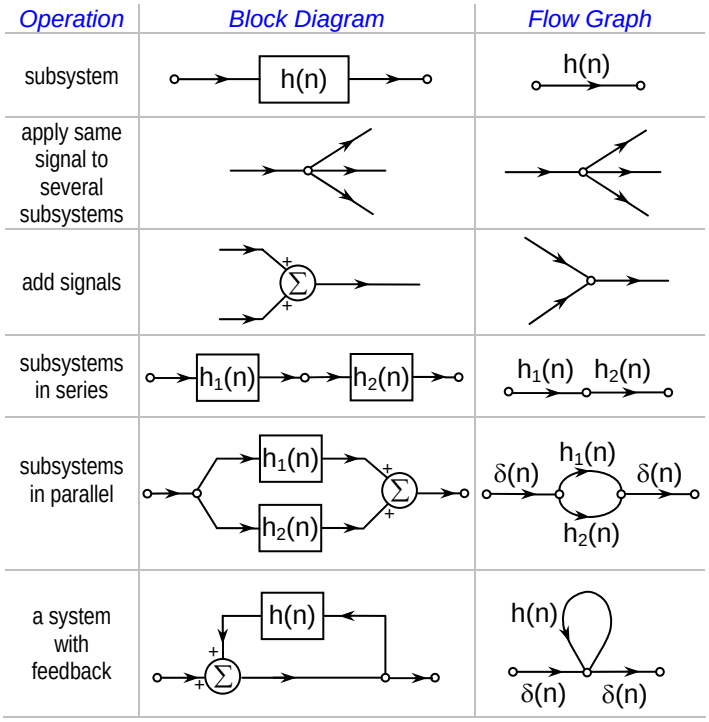


3. Complex systems

3.1 Block diagrams and flow graphs

Several systems may be connected together to form a larger system. In this process some of the subsystem outputs become inputs to other subsystems. We often specify the way the subsystems are interconnected by a schematic diagram. There are two standard and equivalent diagram notations. One leads to the so-called block diagram of the system; the other results in the system flow graph.

We have been using the block diagram notation for a system, that is, a box between two lines designating that the input is applied to one end and the output received at the other (according to the arrows).



Inside the box we indicate the system impulse response. In the flow graph notation we omit the box and simply indicate by a line between two terminal points, an arrow, and the system impulse response, the way in which the system operates. (Refer to the figure.)

The easiest way to interpret the diagrams is to imagine that the input and output functions are telephone signals. The arrows on our diagrams show the directions that these signals flow, and the impulse responses show what operations are to be performed on the signals during their flow through the branches of the communication network. This analogy is so good that we will constantly use the terminology suggested by it in discussing systems problems regardless of their source.

When we combine several subsystems, it's often necessary to indicate that the same function (signal) is to be applied to several lines. This operation is indicated in both notations by leading the signal to a terminal from which we draw several diverging lines. The terminal serves to transmit along all outgoing lines the signal it receives on its incoming line.

There is a similar need for an indication of the addition of two or more signals to form a sum, which is then transmitted to the rest of the network. In block diagram notation this addition is depicted by leading incoming lines to a circle (summing box) where their addition is specified by + signs on these lines. The sum is transmitted on the outgoing line. If we wish to subtract any of the input signals we may indicate this by using – signs. In the flow graph notation an addition is simply indicated by bringing the lines into a terminal. Subtractions cannot be indicated in this way; a merging of lines always indicates addition. A subtraction can be obtained only by first passing the signal to be subtracted through a subsystem which changes its sign ($h(n)=-\delta(n)$) and then adding it.

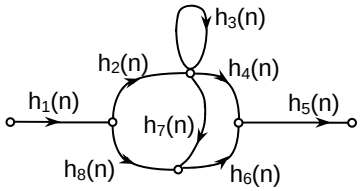
When the output of one subsystem becomes the input to a second we say that the subsystems are in series and indicate this in both notations by placing the subsystems in tandem.

When the same input is applied to each of two subsystems and their outputs are added we say that the subsystems are in parallel and indicate this with the aid of the notations already mentioned.

It's possible that the input to a subsystem be derived in part from its output (perhaps after this output has been processed by other subsystems). When this happens, the system is said to contain feedback. It's the presence of feedback that generally makes systems problems interesting since its use will enable us to obtain relatively complex behavior from systems built by connecting a few simple subsystems.

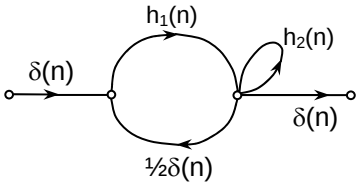
In the following discussions we will generally use the flow-graph notation because of its conciseness. A flow graph, then, will consist of a number of branches and tie points. Each branch indicates a subsystem with the

arrow indicating the direction of signal flow, and the h(n) function, the impulse response. The tie points, called nodes, operate in such a way that the signals on all entering lines are added, and the sum is transmitted along each of the outgoing lines. The graph generally has an incoming and outgoing branch, which serves to introduce the over-all system input and tap off the over-all system output.



It's important to note that if one simply wants to transport a signal unchanged from one node to another, a subsystem (h(n)=δ(n)) is required. Each branch must have an impulse response specified even if this response is trivial. The reason for this is that in eliminating the box as a notation for the subsystem, the flow graph has no geometrical way of distinguishing between a simple connecting line and a subsystem for which the input and output are not identical.

As an example suppose a system is composed of two major subsystems with impulse responses $h_1(n)$ and $h_2(n)$ respectively. The input to the h_1 subsystem is the sum of the over-all system input and one-half of the over-all system output. The input to the h_2 subsystem is the sum of the outputs of the h_1 and h_2 subsystems. The output of the over-all system is the same as the input to the h_2 subsystem. These specifications allow us to construct the flow graph.



The flow graph diagram for a system is not unique. There may be several possible diagrams, each differing in some detail, but all describing the same system. Often, however, one of these will be obviously the simplest and most reasonable.

3.2 Over-all system properties

The system formed by connecting linear, time-invariant subsystems is itself linear and time-invariant (provided that the system is quiescent until excited). We can easily convince ourselves of this fact. Suppose that the overall system input and output are f(n) and g(n). Besides these functions we consider all of the subsystem inputs and outputs,

$$f_1(n), g_1(n), f_2(n), g_2(n), \cdots$$

Some of the above functions are formed by adding together some of the other functions at the system tie points. By assumption the subsystem operations which derive $g_1(n)$ from $f_1(n)$, $g_2(n)$ from $f_2(n)$, etc., are linear and time-invariant. Now consider the set of functions,

$$f(n+m), g(n+m), f_1(n+m), g_1(n+m),$$
$$f_2(n+m), g_2(n+m), \cdots$$

where n is any integer. Because of the time-invariance property of the subsystems (and tie-point additions) these functions form a consistent set of functions related by the system operations. In other words, g(n+m) is a response to f(n+m).

The proof of linearity is similar. Suppose we have for the input $f_a(n)$ the set

$$f_a(n), g_a(n), f_{1a}(n), g_{1a}(n), f_{2a}(n), g_{2a}(n), \cdots$$

and for the input $f_{obs}(n)$ the set

$$f_b(n), g_b(n), f_{1b}(n), g_{1b}(n), f_{2b}(n), g_{2b}(n), \cdots$$

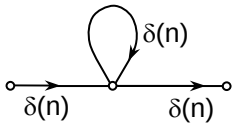
Then by the linear behavior of the subsystems (and tie-point additions) the following set is also consistent.

$$af_a + bf_b, \quad ag_a + bg_b, \quad af_{1a} + bf_{1b}, \quad ag_{2a} + bg_{2b}, \quad \cdots$$

This implies that ag_a+bg_b is an over-all response to af_a+bf_b.

In each of the above proofs we must be assured that the responses are, in fact, unique. We can insure this by agreeing to consider only an over-all system whose output is identically zero whenever the input is identically zero. By the argument given in section 2.1, non-uniqueness of the response would then lead to a contradiction.

It's important to note that this additional assumption is a real restriction since the quiescent properties of the individual subsystems as shown in section 2.1 do not suffice to insure this property for the over-all system. For example, the following system for a zero input



has a possible output g(n)=A, where A is any constant. The quiescent properties of individual subsystems are

not violated, however, since a subsystem output is nowhere obtained without an input. Hence we would agree to consider the system only on condition that A=0.

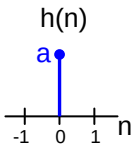
In general, we will consider only systems all of whose node signals are identically zero if the input signal is identically zero. This assumption is consistent with the quiescent properties of the subsystems and suffices to make the over-all system linear.

Our over-all linear system, now, has an additional useful property: the superposition property of response to several inputs applied to the same input node also holds if the inputs are applied to different nodes of the graph. That is, even if f_a and f_b are applied separately to different nodes yielding g and g_b respectively, at the output (one given node), then $a_{ga}+b_{gb}$ is the output for af_a and bf_b applied simultaneously at the two input nodes. The proof is the same as above for the superposition of inputs at the same input node. This properly justifies our concentration on the simple one input - one output system, for if we know how to handle problems of this type, the application of the superposition property for multiple inputs immediately supplies the answers to the general case.

3.3 System representation

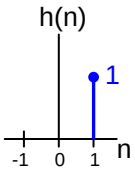
In this section we show that any linear time-invariant system can be represented by interconnecting subsystems of three elementary types. Or in other words, given the over-all system impulse response, we desire to construct a system having this response by using simple building blocks.

The first subsystem type we need is the constant gain. The output of this subsystem is defined to be a constant multiple, a, of the input and has the impulse response



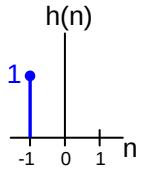
$h(n) = a\,\delta(n)$

The second type of subsystem is the unit delay. The present output value is always equal to the immediately preceding input value. The impulse response is



$h(n) = \delta(n-1)$

The third type of subsystem is the unit advance in which the signal is advanced one unit of time rather than delayed. The impulse response is



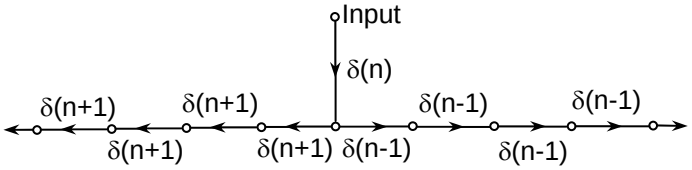
$h(n) = \delta(n+1)$

Systems requiring unit advance subsystems in their construction are evidently not physically realizable.

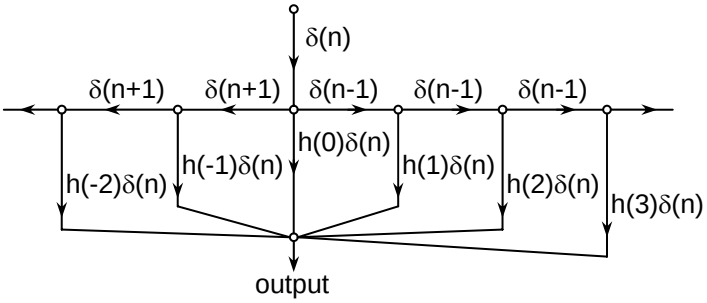
To construct a flow graph of a system using these subsystems we expand its impulse response in a series of δ functions:

$$h(n) = h(0)\delta(n) + h(1)\delta(n-1) + h(2)\delta(n-2) + \cdots$$
$$+ h(-1)\delta(n+1) + h(-2)\delta(n+2) + \cdots$$

Now consider a graph in which an infinite number of unit delays have been placed in series, an infinite number of unit advances have been placed in series, and the over-all system input is fed to both of these arrangements.



The above expanded form for the impulse response means that we think of h(n) as the sum of a set of single sample values each of which occurs at the proper time relative to the input unit sample. h(2) δ(n–2), for example, means that we require a sample of value h(2) to occur after two units delay. But by tapping off the signal at various nodes of the above graph we can get unit samples (for a unit sample input) with any advance or delay. These only need to be passed through subsystems with gains, h(n), n = 0, ±1, ±2, . . . and added to provide an output signal which is the required impulse response.

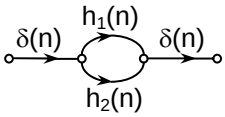


This representation, leading as it often does to an infinite number of branches, is not a practical one for the study of the system. We have used it simply to establish that a representation of any system using the simple subsystem elements exists. In fact, by using feedback in connecting the elementary subsystems we can, for most practical cases, find a representation with a finite number of branches. Just how this is done will become clear later.

3.4 Over-all system response

In this section we want to investigate the problem of finding the impulse response of a system when only the impulse responses of the subsystems composing it are specified.

For a system composed of two subsystems in parallel the problem is easy

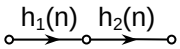


The response here is always the sum of the separate responses of the two subsystems to the same applied input. Use of this fact with the special case of the input f(n)=δ(n) we immediately find

$$h(n) = h_1(n) + h_2(n)$$

A generalization to any number of paralleled subsystems is immediate.

If the two subsystems are arranged in series then the

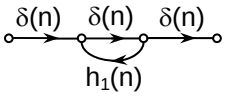


output of the first is always the input to the second, and again by treating the case f(n)=δ(n) the first subsystem responds with $h_1(n)$ and the output of the second, which is the impulse response we seek, is given by the superposition summation (or convolution)

$$h(n) = \sum_{k=-\infty}^{+\infty} h_1(k)h_2(n-k) = h_1(n) * h_2(n)$$

An important result follows from interchanging $h_1(n)$ and $h_2(n)$. As we have seen before in section 2.3 the value of the sum is unchanged, and therefore, the result of passing a signal through a series of systems does not depend on the order in which these are arranged. Again we could generalize this result to an arrangement of any number of series subsystems.

When the system contains feedback, the problem becomes much more difficult. For example the system



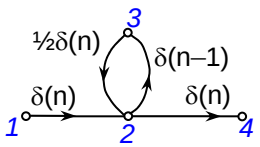
with the input δ(n) and output h(n), leads to the following equation for h(n).

$$h(n) = \delta(n) + \sum_{k=-\infty}^{+\infty} h(k)h_1(n-k)$$

(Verify this.) The equation is in reality an infinite set of simultaneous algebraic equations for the values of h(n), with one equation for each value of n and with δ(n) and $h_1(n)$ as known functions. The solution of this set of equations is cumbersome to obtain by direct methods. At this point we must admit that our analysis methods are inadequate and wait until we have introduced more powerful methods for handling feedback systems.

There is an important case, however, where in spite of feedback, the system impulse response or response to any input may be found sample-by-sample. This case arises when the system graph is composed of nothing but the gain and delay elements mentioned in the preceding section. It's then easy to trace through the system operation, one time unit at a time, computing each node signal and obtaining a step-by-step record of output sample values.

The best way to explain this process is to give an example.



We apply δ(n) to node 1 and find h(n) at node 4. The signal at nodes 2 and 4 is the same. The signal at node 3 is the signal at node 2 after a unit delay. This node 3 signal is then multiplied by ½ and added to the node 1 signal to form the node 2 signal.

$$h(n) = 0 \qquad for\ n < 0$$

Then we find the following values at times n ≥ 0.

n	node 1	node 2	node 3	node 4
0	1	1	0	1
1	0	½	1	½
2	0	¼	½	¼
3	0	⅛	¼	⅛
4	0	1/16	⅛	1/16
...	...	...	...	...

It should be clear that in this example we can deduce

$$h(n) = \begin{cases} 0 & n < 0 \\ (1/2)^n & n \geq 0 \end{cases}$$

The method is chiefly useful, however, for finding the first few non-zero values of h(n).

Note that we have just given an example of a system composed of a finite number of the elementary subsystems (using feedback) but whose impulse response contains an infinite number of non-zero samples. The realization of this system by the method of the previous section, then, would require an infinite number of subsystem branches.

3.5 Problems

3.5.1) A system is composed of three major subsystems with impulse responses $h_1(n)$, $h_2(n)$, $h_3(n)$. The over-all system input is added to the h_3 subsystem output to form the h_1 subsystem input. The h_2 output is subtracted from the h_1 output to form the h_2 input. The over-all system output is the same as the h_1 input. The h_3 input is the same as the h_2 input. Draw a flow graph for this system.

3.5.2) Draw a flow graph for a system whose input and output, f(n) and g(n), are related by the difference equation

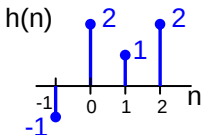
$$g(n) + \tfrac{1}{3}\, g(n-1) = f(n)$$

3.5.3) Draw a flow graph for a system specified by the equations

$$g(n) + \tfrac{1}{2}\, g(n-1) + e(n) = f(n)$$
$$e(n) + \tfrac{1}{3}\, g(n-2) = f(n-1)$$

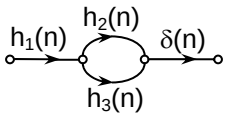
f(n) is the input, g(n) the output, and e(n) the signal at an inner node of the graph. Make sure that the graph contains a node at which e(n) appears. Hint: Solve the first equation for g(n) and the second for e(n), and interpret.

3.5.4) A system has the impulse response



Find a flow graph composed only of gains, unit delays, and unit advances which gives, this response.

3.5.5) Find the over-all h(n) for the system:

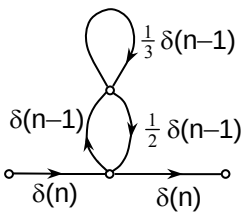


where:

$$h_1(n) = \begin{cases} 0 & n < 0 \\ (1/3)^n & n \geq 0 \end{cases} \qquad h_2(n) = \begin{cases} 0 & n < 0 \\ 1 & n \geq 0 \end{cases}$$

$$h_3(n) = \begin{cases} 0 & n < 1 \\ -4/3 & n = 1 \\ -1 & n > 1 \end{cases}$$

3.5.6) Find (by following through the system operation) the first five non-zero samples of the impulse response for the system



4. Transforms and transfer functions

4.1 Signal transforms

To extend the domain of systems problems that we can handle with facility, we must introduce additional methods for describing signals.

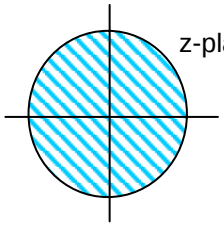
As a first case, suppose we have an input signal, f(n), which has only zero samples before n=0. Then, using the complex variable z, we form the Maclaurin series

$$f(0) + f(1)z + f(2)z^2 + f(3)z^3 + \ldots$$

If we suitably restrict the behavior of f(n) this series will represent some function, F(z), of the complex variable, z, as long as z lies within the circle of convergence of the series. In that case, F(z) will be an analytic function of z in the neighborhood of z=0. The restriction is needed to insure that the radius of convergence is not zero. In nearly all cases of practical interest, F(z) will have only isolated singularities and, therefore, we can analytically continue F(z) to extend its definition throughout the z-plane. We will frequently perform this continuation by noting that the given series is the Maclaurin expansion of a known function in closed form.

$$F(z) = f(0) + f(1)z + f(2)z^2 + f(3)z^3 + \ldots$$

for z in circle of convergence



The restriction to be placed on f(n) is essentially a limitation on its rate of growth as $n \to \infty$ so that by picking z with $|z|$ sufficiently small (but not zero) the series will converge. It's sufficient to restrict f(n) to a geometrical rate of increase. To be precise, we assume the existence of a finite positive number M such that

$$\lim_{n \to \infty} \frac{|f(n)|}{M^n} = 0$$

Then if we take z with $|z| < 1/M$, each of the terms of our series (for large enough n) will be less in absolute value than the corresponding term of a convergent geometrical series. The F(z) series then converges.

Note that f(n) is not required to decrease in magnitude with increasing n or even to remain bounded. It simply must not increase too fast. A high rate of increase means a small circle of convergence, but a restriction to geometrical rates insures that the circle does not vanish to a single point, z=0.

For an f(n) specified as above, then, we can find a function F(z). Conversely, if F(z) is given, by expanding it about z=0, f(n) may be found as the coefficients of the expansion. F(z) is called the transform of f(n). f(n) is the inverse transform of F(z). f(n) is the description of the signal in the time domain; F(z) is its description in the transform domain. The transform domain is also called the frequency domain although for sampled data systems the connection with the concept of frequency is not immediate.

As an example, take

$$f(n) = \begin{cases} 0 & n < 0 \\ A^n & n \geq 0 \end{cases}$$

where A is a constant. Then

$$F(z) = 1 + Az + A^2z^2 + A^3z^3 + \cdots$$

We recognize this immediately as a geometric series with the sum

$$F(z) = \frac{1}{1 - Az}$$

which represents F(z) in the entire z-plane. Having been given F(z), f(n) can he found by expansion by long division

$$\begin{array}{r} 1 + Az + A^2z^2 + \cdots \\ 1 - Az \;\overline{)\,1} \\ \underline{1 - Az} \\ Az \\ \underline{Az - A^2z^2} \\ A^2z^2 \\ \cdots \end{array}$$

As a second example, suppose

$$F(z)=\frac{1+z^2}{1-\frac{1}{2}z-\frac{1}{2}z^2}$$

This fraction can again be developed by long division. To readily deduce f(n), however, we proceed by a partial fraction expansion.

$$F(z)=\frac{1+z^2}{1-\frac{1}{2}z-\frac{1}{2}z^2}=-2+\frac{3-z}{(1-z)(1+\frac{1}{2}z)}$$

$$=2+\frac{\frac{4}{3}}{1-z}+\frac{\frac{5}{3}}{1+\frac{1}{2}z}$$

$$-2+\tfrac{4}{3}(1+z+z^2+\cdots)+\tfrac{5}{3}(1-\tfrac{1}{2}z+\tfrac{1}{4}z^2+\cdots)$$

From the coefficients we see that

$$f(n)=\begin{cases}0 & n<0\\ 1 & n=0\\ \frac{4}{3}+\frac{5}{3}(-\frac{1}{2})^n & n>0\end{cases}$$

For a third example, take

$$f(n)=\begin{cases}0 & n<0\\ n & n\geq 0\end{cases}$$

Then

$$F(z)=z+2z^2+3z^3+4z^4+\cdots$$

$$=z(1+2z+3z^2+4z^3+\cdots)$$

$$=z\tfrac{d}{dz}(z+z^2+z^3+z^4+\cdots)$$

$$=z\tfrac{d}{dz}z(1+z+z^2+z^3+\cdots)$$

$$=z\tfrac{d}{dz}\left(\frac{z}{1-z}\right)$$

$$F(z)=\frac{z}{(1-z)^2}$$

(The termwise differentiation is justified by the uniform convergence of the series within its circle of convergence.)

In most systems problems the signal transforms are indeed as above rational functions of z. The techniques of the examples are, therefore, constantly useful. (Find the poles and zeros of the above transforms and the regions of convergence of the Maclaurin expansions.

4.2 Signal transforms (cont.)

To treat a second case of signal transforms assume that f(n) has only zero samples for positive values of n. Then we form the series, which is to represent the transform

$$F(z)=f(0)+f(-1)z^{-1}+f(-2)z^{-2}+f(-3)z^{-3}+\ldots$$

As before we restrict f(n) to have no more than geometrical growth as n→ –∞. Then the series for z converges outside of some circle about z=0. Our expansion of F(z) is now the Taylor expansion about the point z=∞. Remarks and techniques similar to those of the previous section also hold here.

However, simply stating F(z), leads to ambiguity unless we specify which of the two types of f(n) the transform represent. We can expand

$$F(z)=\frac{1}{1-z}=1+z+z^2+z^3+\cdots$$

or

$$F(z)=\frac{1}{1-z}=\frac{-z^{-1}}{1-z^{-1}}=-z^{-1}-z^{-2}-z^{-3}-\cdots$$

The first expansion about z = 0 gives

$$f(n)=\begin{cases}0 & n<0\\ 1 & n\geq 0\end{cases}$$

The second expansion about z = –∞ gives

$$f(n)=\begin{cases}0 & n\geq 0\\ -1 & n<0\end{cases}$$

We must state which of the f(n) we mean to represent. The problem context will always make the selection clear.

Now consider case three where f(n) has non-zero samples for both positive and negative values of n.

Again we represent f(n) by a power series which gives F(z) in its region of convergence.

$$F(z) = f(0) + f(1)z + f(2)z^2 + f(3)z^3 + \ldots$$
$$+ f(-1)z^{-1} + f(-2)z^{-2} + f(-3)z^{-3} + \ldots$$

This series is a Laurent series, which has an annular region of convergence, the region of overlap common to the separate regions of convergence of the ascending and descending parts of the series. That is, we must again limit the rate of growth of f(n) as n→∞ and n→ –∞ so that such a common region exists.

Again the problem arises of how to expand a given F(z) in closed form. Again we must know something about f(n) to be able to select the proper expansion. Each of these expansions will converge in a different annulus. It will be clear from the problem which to select. Three cases are common.

First, f(n) may have only a finite number of its non-zero samples for negative n. Secondly, f(n) may have only a finite number of such samples for positive n. And thirdly, f(n) may be bounded as both n→∞ and n→ –∞ and tend to zero in at least one of these cases with geometrical rapidity. That is, there exists a positive number, M, with M<1 such that at least one of the following is true.

$$\lim_{n\to\infty} \frac{|f(n)|}{M^n} = 0$$

$$\lim_{n\to-\infty} \frac{|f(n)|}{M^n} = 0$$

For the first type of f(n), the region of convergence for the F(z) expansion is inside a circle about the origin except that the center, z = 0, is omitted. In the second instance, the region of convergence is everywhere outside a circle about the origin except for z = ∞. In the third instance, the region is an annular one about the origin, either having the circle of unit radius as one of its boundaries or containing this circle in its interior.

Consider the example

$$F(z) = \frac{1}{3z - 2z^2 - 1}$$

We specify that f(n) is bounded as n→∞ and tends to zero as n→ –∞. f(n) cannot be found by division until we make clear by a partial fraction expansion which are to be the ascending and which are to be the descending portions of the series.

We obtain

$$F(z) = \frac{1}{(1-z)(2z-1)} = \frac{1}{1-z} + \frac{2}{2z-1}$$

According to our specification of f(n) we must arrange this expression as

$$F(z) = \frac{1}{1-z} + \frac{z^{-1}}{1-\frac{1}{2}z^{-1}}$$

$$= (1 + z + z^2 + \cdots) + (z^{-1} + \tfrac{1}{2}z^{-2} + \tfrac{1}{4}z^{-3} + \cdots)$$

so that

$$f(n) = \begin{cases} 1 & n \geq 0 \\ 2^{n+1} & n < 0 \end{cases}$$

(What would the answer be if we had specified f(n) to be zero for n>0?)

4.3 Transfer functions, response problem

We have shown how to represent signals such as the input, f(n), and output, g(n), of a system as transforms.

$$F(z) = \sum_{n=-\infty}^{+\infty} f(n)\, z^n$$

$$G(z) = \sum_{n=-\infty}^{+\infty} g(n)\, z^n$$

The impulse response, h(n) of a system may similarly be represented.

$$H(z) = \sum_{n=-\infty}^{+\infty} h(n)\, z^n$$

The transform, H(z), in this case is given the special name, transfer function.

The simplest examples of such transfer functions arise by transforming the impulse responses of the three elementary subsystems (gain, unit delay, unit advance). For the simple gain

$$h(n) = A\delta(n) \qquad H(z) = A$$

A unit delay gives

$$h(n) = \delta(n-1) \qquad H(z) = z$$

The unit advance yields

$$h(n) = \delta(n+1) \qquad H(z) = z^{-1}$$

Now that the system input, output, and impulse response have been converted to the transform domain, we discuss the response problem in this domain; that is, F(z) and H(z) are given and G(z) is to be found. This problem is solved by converting the superposition integral to the transform domain.

$$g(n) = \sum_{k=-\infty}^{+\infty} f(k)\, h(n-k)$$

Multiplying by z^n and summing over n we find after some manipulation

$$\sum_{n=-\infty}^{+\infty} g(n)z^n = \sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f(k)z^k \cdot h(n-k)z^{n-k}$$

$$= \sum_{k=-\infty}^{+\infty} f(k)z^k \cdot \sum_{n=-\infty}^{+\infty} h(n)z^n$$

$$G(z) = F(z) \cdot H(z)$$

For this result to be correct the power series expansions of F(z) and H(z), which are to represent f(n) and h(n), must have a common region of convergence. Then G(z) must be developed in a series convergent in this same region if we are to find g(n). For example, if f(n) = 0 for n>0 and the system is physically realizable, then g(n) = 0 for n<0 and all the expansions refer to a circle of convergence about the origin. (The interchange of order of summation above is justified by the absolute convergence of the series.)

The foregoing result shows that the solution of the response problem in the transform domain is simple since it requires only a multiplication of the input transform by the transfer function to find the output transform. Therefore an indirect procedure of going from f(n) and h(n) to F(z) and H(z), then computing G(z)=F(z) H(z) and finally finding g(n) may actually be easier than a direct time operation.

As a first set of examples, consider any f(n) and the h(n) responses of three elementary subsystems. For the simple gain, H(z) = A,

$$G(z) = A\,F(z)$$

which is obviously correct, for when F(z) is expanded, each of its coefficients is to be multiplied by A. For the unit delay, H(z) = z,

$$G(z) = z\,F(z)$$

and each power of z in the F(z) expansion is to be raised by unity so that g(n) = f(n–1). The effect of the unit advance is similarly obvious.

As another example, suppose

$$f(n) = \begin{cases} 0 & n<0 \\ (1/2)^n & n \geq 0 \end{cases}$$

$$h(n) = \begin{cases} 0 & n<0 \\ (1/3)^n & n \geq 0 \end{cases}$$

Then we compute

$$F(z) = 1 + \tfrac{1}{2}z + \tfrac{1}{4}z^2 + \cdots = \frac{1}{1-\frac{1}{2}z}$$

$$H(z) = 1 + \tfrac{1}{3}z + \tfrac{1}{9}z^2 + \cdots = \frac{1}{1-\frac{1}{3}z}$$

The output transform is then easily found and interpreted.

$$G(z) = F(z) \cdot H(z)$$

$$\frac{1}{(1-\frac{1}{2}z)(1-\frac{1}{3}z)} = \frac{3}{1-\frac{1}{2}z} - \frac{2}{1-\frac{1}{3}z}$$

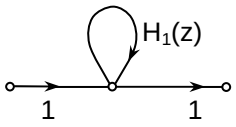
$$= 3(1 + \tfrac{1}{2}z + \tfrac{1}{4}z^2 + \cdots) - 2(1 + \tfrac{1}{3}z + \tfrac{1}{9}z^2 + \cdots)$$

$$g(n) = \begin{cases} 0 & n<0 \\ 3(1/2)^n - 2(1/3)^n & n \geq 0 \end{cases}$$

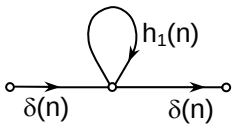
Because of the simplicity of the transform domain formula G=FH as opposed to that of the time domain g=f*h, problems other than the direct response problem can be easily handled. So we may specify the output and transfer function and be required to find the input that will produce this output (F=G/H), or the transfer function that will convert a given input to a given output may be desired (H=G/F).

4.4 Transforms and flow graphs

Since the impulse response, h(n), or transfer function, H(z), may equally well represent the characteristics of a system (it being understood the way in which H(z) is to be expanded), we may draw all of our system flow graphs with specified branch transfer functions. For example,

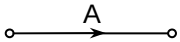


in place of

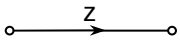


In what follows we will always use the transfer function notation.

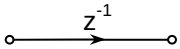
For graphs composed of the three elementary subsystems, all branches will be of the following types.



gain of A

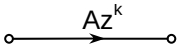


unit delay

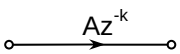


unit advance

To save space we will often combine some of these branches to provide



gain of A and k units of delay



gain of A and k units of advance

(Verify that the branch transfer functions do represent the indicated operations.)

4.5 Problems

4.5.1) Find the transforms of the following signals:

a.) $$f(n)=\begin{cases}0 & n<0\\ 1-2(-1/5)^n & n\ge 0\end{cases}$$

b.) $$f(n)=\begin{cases}3^n & n<0\\ 2 & n=0\\ 0 & n>0\end{cases}$$

c.) $$f(n)=\begin{cases}4^n & n<0\\ (1/4)^n & n\ge 0\end{cases}$$

d.) $$f(n)=\begin{cases}0 & n<0\\ 1/n! & n\ge 0\end{cases}$$

e.) $$f(n)=\begin{cases}0 & n<0\\ n^2 & n\ge 0\end{cases}$$

f.) $$f(n)=\begin{cases}0 & n<0\\ n2^n & n\ge 0\end{cases}$$

4.5.2) Find the time functions represented by the following transforms:

a.) $F(z)=\dfrac{1}{1+z^2}$ $(f(n)=0 \ \ for \ \ n<0)$

b.) $F(z)=\dfrac{1-z}{1-\frac{5}{4}z+\frac{1}{4}z^2}$ $(f(n)=0 \ for \ n<0)$

c.) $F(z)=\dfrac{z}{1-2z+z^2}$ $(f(n)=0 \ for \ n>0)$

d.) $F(z)=\ln(1-z)$ $(f(n)=0 \ \ for \ n<0)$

e.) $F(z)=\sin(z)$

f.) $F(z)=\dfrac{1+z}{-2+5z-2z^2}$ $(f(n)\to 0$ $as \ n\to\infty$ $and \ n\to-\infty)$

4.5.3) Why is it impossible to represent the following time function by a transform?

$$f(n)=2^n \qquad (for\,all\,n)$$

4.5.4) Find the correct missing input, output, or impulse response (f(n), g(n), h(n)) when the following are known. Use transform methods.

a.) $$f(n)=\begin{cases}0 & n<0\\ n & n\geq 0\end{cases}$$

$$h(n)=\begin{cases}0 & n<0\\ (1/2)^n & n\geq 0\end{cases}$$

b.) $$f(n)=\begin{cases}3^n & n<0\\ (1/3)^n & n\geq 0\end{cases}$$

$$h(n)=\delta(n)-\tfrac{1}{3}\delta(n-1)$$

c.) $$h(n)=\begin{cases}0 & n<0\\ 1 & n\geq 0\end{cases}$$

$$g(n)=\begin{cases}0 & n<0\\ (1/2)^n & n\geq 0\end{cases}$$

d.) $$f(n)=\begin{cases}0 & n<0\\ (1/2)^n & n\geq 0\end{cases}$$

$$g(n)=\delta(n)$$

4.5.5) Can the physical realizability of a system be decided by inspecting its transfer function, H(z)?

5. Complex systems in the transform domain

5.1 Flow graphs without feedback

Now we return to the problem of finding the over-all system behavior when only the behavior of its component subsystems is known. We will find that working in the transform domain will simplify what we already know and, moreover, will permit us to solve feedback problems which are difficult to treat in the time domain.

The problem treated in this chapter is the finding of the over-all system transfer function when the subsystem transfer functions are known. First we consider systems with no feedback.

Systems without feedback are built up by connecting the subsystems in such a way that the signal cannot recirculate to any node it has once visited. If you can leave any node and follow the branches in the direction of the arrows back to that node, then the system has feedback. A node with this property is said to have feedback around it.

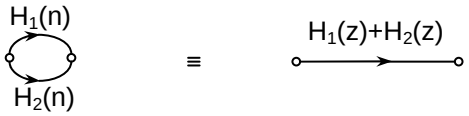
The transform, F(z), of the sum of two signals, $f_1(n) + f_2(n)$, is the sum of the transforms of the separate signals.

$$F(z) = \sum_{k=-\infty}^{+\infty} \left[f_1(k) + f_2(k) \right] z^k$$

$$= \sum_{k=-\infty}^{+\infty} f_1(k) z^k + \sum_{k=-\infty}^{+\infty} f_2(k) z^k$$

$$= F_1(z) + F_2(z)$$

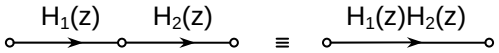
We may apply this result immediately to find the transfer function of a system composed of two (or more) paralleled subsystems.



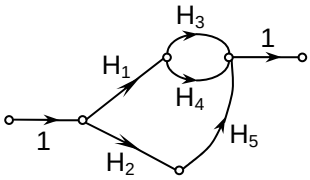
(We have omitted explicitly showing any input, output branches.)

For the two series connected subsystems, $H_1(z)$ and $H_2(z)$, and the input F(z), the output of the first subsystem is $F(z)H_1(z)$. The output of the second is then $F(z)H_1(z) \cdot H_2(z) = G(z)$. The over-all transfer function is then

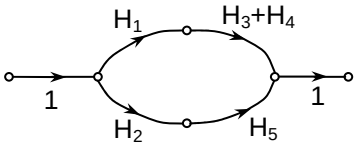
$$H(z) = \frac{G(z)}{F(z)} = H_1(z) H_2(z)$$

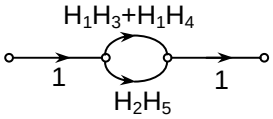


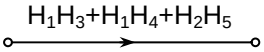
By using the above rules any graph without feedback can be reduced to a single branch whose transfer function is that of the over-all system. As an example of the reduction of series-parallel systems consider the graph,



This system may be reduced in the following steps:

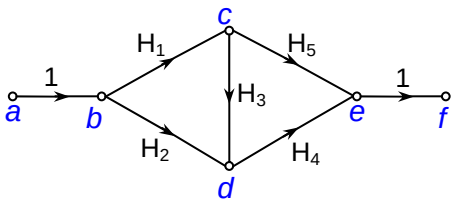






A more powerful way to find the over-all transfer function, a way that will suffice to reduce any system without feedback, is to apply a unit sample at the over-all system input. The signal appearing at each node is then computed (transform form), working stepwise from input to output. The output transform is the required transfer function. The stepwise process will always succeed since, without feedback, the signal at any node can be expressed entirely in terms of signals already found towards the input end of the graph.

As an example consider the graph,

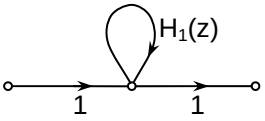


The transform of the input at node *a* is 1. At *b* we also have 1. At *c* the signal is H_1. At *d* the signal is $H_2+H_1H_3$. At *e* the signal is $H_1H_5+H_2H_4+H_1H_3H_4$. This is the signal that appears at the output node *f* and is the over-all system transfer function.

Another way of obtaining the above result is to find all possible ways of tracing from input to output of the graph. Each way adds to the transfer function a term consisting of the product of subsystem transfer functions encountered along the pathway.

5.2 Flow graphs with feedback

The key to the solution of the feedback graph reduction problem is found in the following fundamental example:

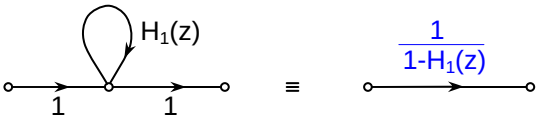


The graph has essentially one node with feedback around it. $H_1(z)$ is called the loop transfer function. The output, G(z), is the sum of the input, F(z), and the signal which is fed in by the loop. This latter signal is derived from the output by passing it around the loop. Therefore:

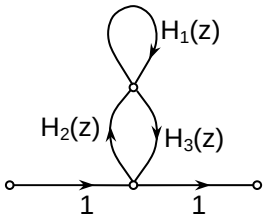
$$G(z) = F(z) + G(z)H_1(z)$$

The over-all transfer function, H(z)=G(z)/F(z), can now be found by dividing both sides by F(z) and rearranging:

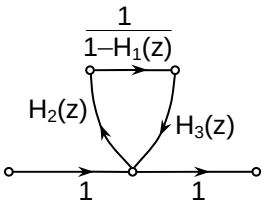
$$H(z) = \frac{G(z)}{F(z)} = \frac{1}{1 - H_1(z)}$$



As an illustration of the utility of this fundamental reduction, consider the following graph.



The loop on top may be connected in the circuit with H_2 and H_3 as a single branch with transfer function $1/(1-H_1)$.



Applying the same procedure, the remaining loop is eliminated.

$$H = \frac{1}{1 - \dfrac{H_2H_3}{1-H_1}} = \frac{1-H_1}{1-H_1-H_2H_3}$$

The formula $H = 1/(1–H_1)$ may be interpreted by making the expansion

$$H(z) = \frac{1}{1-H_1(z)} = 1 + H_1(z) + H_1^{\,2}(z) + H_1^{\,3}(z) + \cdots$$

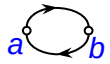
The resulting series has the significance of representing the impulse in terms of a sum of signals. The first is the direct transmission of the unit sample to the output. The second is the signal that emerges after circulating once about the loop. The third is due to the twice circulating signal, and so on. This interpretation shouldn't be taken too seriously, however, for the expansion is not always mathematically valid; the condition for convergence, $|H_1(z)| < 1$, may not be obtained in the region of the of the z-plane where the power series expansion of $H_1(z)$ is to represent $h_1(n)$.

5.3 General reduction procedure

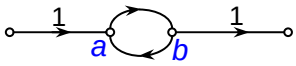
The reduction of flow graphs with no feedback and the reduction of the single loop feedback are together a sufficient arsenal with which to attack the general reduction problem. A stepwise procedure will now be given by which any graph with a finite number of branches can be progressively simplified until it's

reduced to a single branch with a determined transfer function.

As a preliminary step it's necessary to indicate on the graph the input and output at separate nodes that are distinct from the internal nodes of the graph. This may always be accomplished by adding an extra branch with unity transfer function to apply the input to the specified node and a similar branch to tap the signal for the output. For example, if in the graph



node *b* is to receive the output signal and node *a* is to be the node to which the input is applied, then we explicitly write

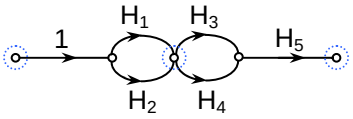


The structure of the graph may frequently be such that these input and output connections are already supplied for us. The appearance of non-unity transfer functions on one or both of these connections is not detrimental. The configuration of the graph is the thing of interest.

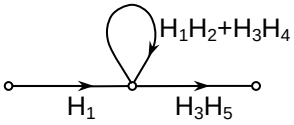
The first step in reducing a graph is to select certain nodes of the graph called residual nodes. There are two kinds of residual nodes. The input node and output node are residual nodes. In addition we select some or all of the nodes of the graph having feedback around them as residual nodes. This selection is made so that if these residual nodes are cut out of the graph, the feedback of the graph is destroyed. By judiciously choosing these internal residual nodes we can keep their number to a minimum. It's well to start the selection with nodes having a large number of feedback paths through them. We circle the residual nodes to distinguish them from the remainder.

The second step is to find the transfer functions that govern the flow of signals between each of the residual nodes (two directions) and the transfer functions of the loop flows around each residual node. These transfer functions can always be found and indicated by single branches since the residual nodes have been selected so that no feedback path can be completed before being interrupted by a residual node. The resulting reduction problems, then, are simply those of systems without feedback.

To illustrate the procedure up to this point, consider the following graph.



We select three residual nodes as shown. Besides the input and output nodes, the center node is chosen since severing all of its incoming and outgoing lines will cut all feedback paths. (We could have chosen the two remaining nodes instead. but the fewer residual nodes, the better.) Now the signal flow from residual node to residual node has no feedback and each such flow has a transfer function, which we find and put on a simplified diagram showing only residual nodes.

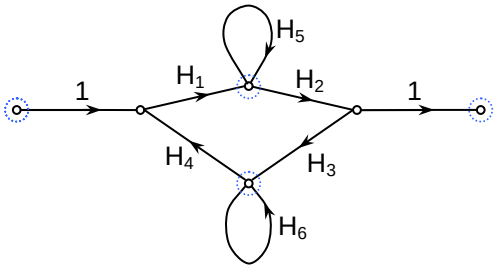


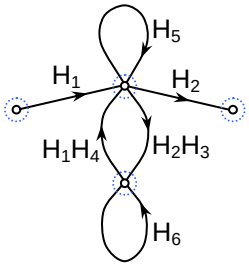
In finding each of these transfer functions we may think of each node in turn as a unit sample transmitting station, and for each such choice consider each node in turn as the receiving station. In this process all residual nodes not transmitting or receiving are dead and, in fact, block the passage of the signal. They may be cut out of the graph until their time comes.

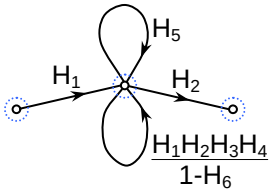
After this process has been completed the simplified graph generally will contain one or more residual nodes with a loop (as in the above example). In fact, each internal residual node will have such a loop since it has been chosen to have this property. If no internal node is necessary the graph is fully reduced by the above procedure and contains no feedback.

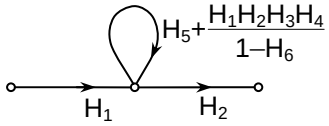
The third step consists of replacing one of the loops by a direct branch according to the result described in the previous section. This procedure introduces another node but it removes the feedback at this point. Now if we repeat step 1 we find that one less internal residual node is required to break the feedback. Continuing through the same steps in sequence we remove the internal residual nodes one by one until the graph is entirely reduced.

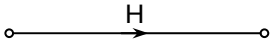
(Verify the results of the following example.)











$$H = \frac{H_1 H_2}{1 - H_5 - \dfrac{H_1 H_2 H_3 H_4}{1 - H_6}} = \frac{H_1 H_2 (1 - H_6)}{(1 - H_5)(1 - H_6) - H_1 H_2 H_3 H_4}$$

If the flow graph contains an infinite number of branches there is no general method for reducing it. The problem of finding the over-all transfer function is then difficult and may be explicitly solved only if there is some regularity in the pattern of the graph and the subsystem transfer functions that can be exploited. Each problem must be handled on its own merits.

It goes without saying that any of the above reductions which combine transfer functions are only valid so long as these transfer functions all have some common z-plane region in which their power series expansions represent the respective impulse responses. Normally we find that the physical background of the problem insures that this requirement is fulfilled. The two most prevalent cases arise when h(n)=0 for n sufficiently negative and when h(n)→0 as n→∞ and n→ –∞. The common region in the first case is in a circle about the origin (perhaps omitting the center point) and in the second case is an annulus containing the unit circle (boundary). The input and output signals to be used will match these properties.

5.4 Flow graphs with rational H(z)

We have seen that to find the over-all transfer function the only required operations are addition (H_1+H_2), multiplication (H_1H_2), and division ($1/(1–H_1)$). If all subsystem transfer functions are rational functions of z (ratios of finite polynomials in z), these operations lead only to new rational functions and hence if the graph has a finite number of branches, H(z), the over-all function is rational. In particular, any system composed of a finite number of gains, delays, and advances, leads to a rational H(z). If the input transform is likewise rational, then the response problem is solved by interpreting the rational output transform in the time domain. Most practical systems problems will be of this nature.

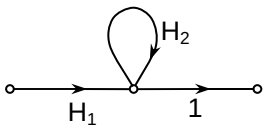
It can be shown conversely that any rational H(z) can be realized from a system in which the subsystems are composed of a finite number of simple gains, delays, or advances. We can write H(z) in the form:

$$H(z) = \frac{p_{-k} z^{-k} + p_{-k+1} z^{-k+1} + \cdots + p_{m-1} z^{m-1} + p_m z^m}{1 - (q_1 z + q_2 z^2 + \cdots + q_n z^n)}$$

$$= \frac{H_1}{1 - H_2}$$

$$= H_1 \cdot \frac{1}{1 - H_2}$$

The graph has the form:



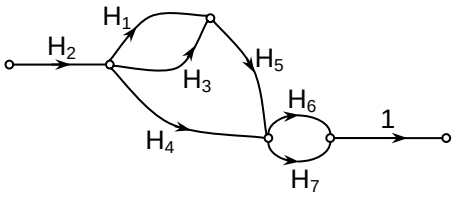
where H_1 and H_2 are composed of series-parallel combinations of gains, delays, and advances. This realization is not unique.

For these reasons we will find that rational H(z) and graphs with gains, delays, and advances will engage most of our future attention.

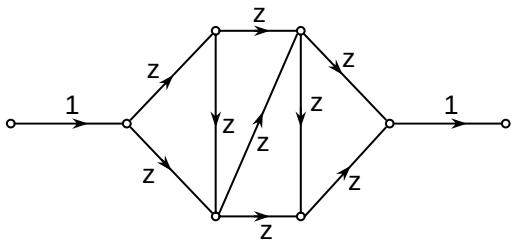
5.5 Problems

5.5.1) Reduce the following flow graphs:

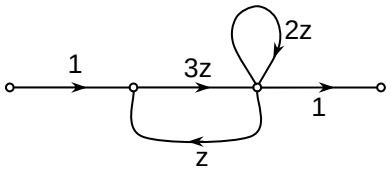
a.)



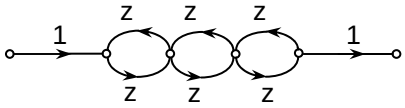
b.)



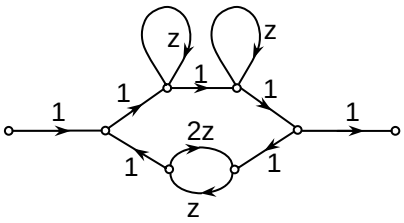
c.)



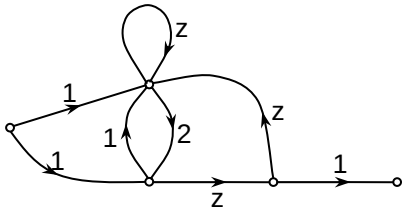
d.)



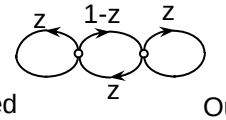
e.)



f.)



g.)



Input applied to left node.

Output applied to right node.

5.5.2) Find a system composed of only gains and delays which has the following transfer function:

$$H(z) = \frac{z^2 + 2z + 3}{z^2 + z + 2}$$

5.5.3) In the system of problem 5.5.l.c the following input is applied.

$$f(n) = \begin{cases} 0 & n < 0 \\ 1 & n \geq 0 \end{cases}$$

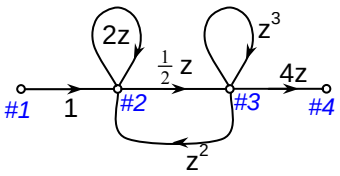
What is the response?

6. Difference equation and matrix descriptions

6.1 System difference equations

In this and the next section we will be concerned only with systems composed of subsystems whose transfer functions have the form, Az^k, where A is a constant and k is an integer, positive, negative or zero. If there are a finite number of branches in the system flow graph, then as we have seen, the over-all system transfer function will be a rational function of z.

The operation Az^k on a signal f(n) produces a delay of k units of time and a multiplication by A to give Af(n–k). This means that it's always possible to give a mathematical description of a system composed of these subsystems by writing down a system of difference equations. Consider this example:



The signal at node *#1* is f(n), the input. The signal at *#4* is g(n), the output. We designate the signals at *#2* and *#3* as $e_2(n)$ and $e_3(n)$.

An equation can be written for each node except the input, indicating how the signals that enter these nodes are derived from the other node signals.

$$e_2(n) = f(n) + 2e_2(n-1) + e_3(n-2)$$
$$e_3(n) = \tfrac{1}{2}e_2(n-1) + e_3(n-3)$$
$$g(n) = 4e_3(n-1)$$

If f(n) is given, then we have the right number of equations to solve for the unknowns. In this case there are three equations to be solved for g(n), $e_2(n)$, $e_3(n)$. Only a particular solution is sought - one that reduces to zero when f(n) is identically zero.

Corresponding to the above set of equations in the time domain is a similar set expressing the same relationships in the transform domain.

$$E_2(z) = F(z) + 2zE_2(z) + z^2E_3(z)$$
$$E_3(z) = \tfrac{1}{2}zE_2(z) + z^3E_3(z)$$
$$G(z) = 4zE_3(z)$$

We could find H(z)=G(z)/F(z) by solving these equations for G(z) with F(z)=1. It's evident that the flow graph reduction procedure is essentially a method for the solution of a set of simultaneous linear algebraic equations, the transform domain equations. (In this conclusion we are not limited to subsystems of the assumed restricted form.) The feedback removal, residual node by residual node, is nothing more than a stepwise elimination of variables in these equations.

For many problems a graphical reduction is easier than a formal solution of the equations and vice-versa. In general, we will use the graph whenever there is an exploitable pattern in it. That is, the utility of graphical reduction is highest when we can take advantage of some peculiarity of its structure to make the reduction easy. Such a peculiarity would be difficult to perceive in the mathematical equations themselves. On the other hand, if there are no exploitable peculiarities a more direct use of the equations is often preferable. This latter situation arises if almost all of the possible interconnections between nodes of the graph exist.

Although one might solve the response problem entirely in terms of the system of difference equations in the time domain, system analysts view this alternative with disfavor. The same answers would be obtained, but with more algebra. Experience has also shown that a full perspective on system problems can only be gained by a familiarity with transform methods. So we will be more interested in converting a given system of difference equations to the transform domain and picturing them by a flow graph than the other way around.

Frequently our knowledge of the physics and logic of a systems mechanism will supply us with a set of difference equations. We will then want to construct a flow graph for the system and supply its branches with the appropriate subsystem transfer functions.

Each equation of the set will represent the signal at a node of the graph in terms of signals at other nodes. There will be one equation for each node of the graph except the input. The input f(n) enters the equations as a forcing term. The output g(n), and intermediate node signals, $e_1(n)$, $e_2(n)$ etc., enter as dependent variables.

This shows that to form the graph we should write the equations in the form:

$$e_1(n) = \alpha_1(e_1, e_2, \cdots, g, f)$$
$$e_2(n) = \alpha_2(e_1, e_2, \cdots, g, f)$$
$$\cdots$$
$$g(n) = \alpha(e_1, e_2, \cdots, g, f)$$

where the α functions represent linear combinations of the indicated signals (with any delays or advances).

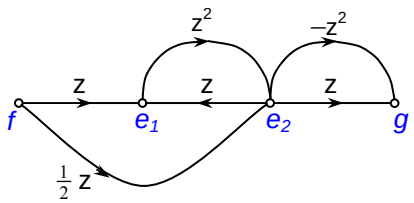
As an example consider the set of equations:

$$0 = e_1(n) - e_2(n-1) - f(n-1)$$
$$0 = e_1(n-1) - e_2(n+1) + \tfrac{1}{2} f(n)$$
$$0 = g(n+1) - e_2(n) + e_2(n-1)$$

We solve the first for e_1, the second for e_2, and the third for g. (Notice the change of time index in the second and third equations.)

$$e_1(n) = e_2(n-1) + f(n-1)$$
$$e_2(n) = e_1(n-2) + \tfrac{1}{2} f(n-1)$$
$$g(n) = e_2(n-1) - e_2(n-2)$$

This set can be converted directly to a graph.



By solving the first equation for e_2 and the second for e_1 a different graph is obtained. (Find it.)

Any finite set of equations that has a determined solution can be properly arranged and converted to a graph. For our rearrangement we start with a variable that appears in the fewest equations and write one of these equations with the variable alone on the left. We strike this equation and its variable from consideration, choose a second variable and equation similarly and continue. The second variable will always be found in the remaining equations since the alternative is that both it and the first variable appear only in the first selected equation. But this is impossible because then the remaining equations would over-determine the remaining variables. Hence, the process of selection can continue unti1 the entire set of equations has been properly arranged and graphed.

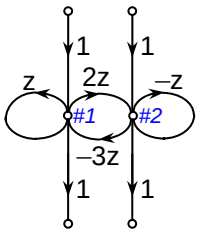
In most instances, systems of equations can be converted to graphs in an off-hand way without a conscious attempt at rearrangement. In fact, the graph can usually be written down directly from a verbal statement of the system relationships.

Graphs can also be drawn for infinite systems of equations. How to do this is usually obvious in each specific case.

6.2 Matrix description

Now we further specialize our considerations to systems built up of subsystems whose transfer functions are of the form, Az. The effect of each branch of the graph is then to delay by a unit time and multiply by a constant any signal that passes through it. We want to show: *(1)* how the behavior of such systems may be described by means of operations with matrices and *(2)* how we may convert to matrix form a systems problem which has been described in terms of difference equations.

In this section we will treat the first of the above questions. Suppose we are given the system graph



We indicate two possible input connections at nodes #1 and #2 and also tap the signals at these nodes for two possible outputs.

Then there are a total of four node-to-node transfer functions, $H_{ij}(z)$, i=1, 2; j=1, 2. We denote the node #1, #2 inputs by $f_1(n)$, $f_2(n)$ and outputs, (the node signals themselves) by $e_1(n)$, $e_2(n)$.

The difference equations describing this system are:

$$e_1(n) = e_1(n-1) - 3e_2(n-1) + f_1(n)$$
$$e_2(n) = 2e_1(n-1) - e_2(n-1) + f_2(n)$$

We can write these equations in matrix form:

$$\begin{bmatrix} e_1(n) & e_2(n) \end{bmatrix} =$$

$$\begin{bmatrix} e_1(n-1) & e_2(n-1) \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix} + \begin{bmatrix} f_1(n) & f_2(n) \end{bmatrix}$$

Suppose the system is quiescent ($e_1(n)=e_2(n)=f_1(n)=f_2(n)=0$; $n<0$) until n=0 when $f_1(0)$ and $f_2(0)$ are applied. Then assume that $f_1(n)=0$, $f_2(n)=0$ for $n>0$. For this kind of input the $f_1(0)$ and $f_2(0)$ impose the initial conditions:

$$e_1(0) = f_1(0)$$
$$e_2(0) = f_2(0)$$

and $e_1(n)$ and $e_2(n)$ are determined for $n>0$ from the stepwise application of the equation:

$$\begin{bmatrix} e_1(n) & e_2(n) \end{bmatrix} = \begin{bmatrix} e_1(n-1) & e_2(n-1) \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix}$$

Let's write this equation in the abbreviated form:

$$\underline{e(n)} = \underline{e(n-1)}\,\underline{A}$$

where $\underline{A}$ is the square matrix with i, j coefficients equal to the i, j branch gain factors. By applying this formula by stages for n = 1, 2, 3, . . . we obtain:

$$\underline{e(1)} = \underline{e(0)}\,\underline{A}$$
$$\underline{e(2)} = \underline{e(1)}\,\underline{A} = \underline{e(0)}\,\underline{A}^2$$
$$\underline{e(3)} = \underline{e(2)}\,\underline{A} = \underline{e(0)}\,\underline{A}^3$$
$$\cdots$$

and in general

$$\underline{e(n)} = \underline{e(0)}\,\underline{A}^n \qquad n \geq 0$$

(We have defined $\underline{A}^0 = \underline{1}$, the unit matrix, to have this equation hold for n=0.)

To interpret this result take one of the $e_i(0)$ equal to one and the other zero and focus attention on the $e_j(n)$ output for a particular j. We see that this output is just given by the i,j element of the matrix $\underline{A}^n$. Since the assumed initial conditions are just those that would be attained by applying a unit sample at node i at t=0 with no other inputs, the $\underline{A}^n$ matrix is the matrix of node-to-node impulse responses, $h_{ij}(n)$.

$$\underline{A}^n = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ \\ h_{21}(n) & h_{22}(n) \end{bmatrix}$$

The matrix, $\underline{H(z)}$, of node-to-node transfer functions, $H_{ij}(z)$, is the transform of the $\underline{A}^n$ matrix sequence:

$$\underline{H(z)} = \sum_{n=0}^{\infty} z^n \underline{A}^n = \begin{bmatrix} \sum z^n h_{11}(n) & \sum z^n h_{12}(n) \\ \\ \sum z^n h_{21}(n) & \sum z^n h_{22}(n) \end{bmatrix}$$

By transforming the equation $\underline{e(n)} = \underline{e(0)}\,\underline{A}^n$ we obtain:

$$\sum_{n=0}^{\infty} z^n \underline{e(n)} = \sum_{n=0}^{\infty} \underline{e(0)}\, z^n \underline{A}^n$$

$$\underline{E(z)} = \underline{e(0)}\,\underline{H(z)}$$

where $\underline{E(z)}$ is the matrix of node signal transforms.

The $\underline{H(z)}$ matrix may be determined by multiplying its defining equation by the matrix $(\underline{1} - z\,\underline{A})$ to obtain:

$$(\underline{1} - z\underline{A})\,\underline{H(z)} = \sum_{n=0}^{\infty} z^n \underline{A}^n - \sum_{n=0}^{\infty} z^{n+1} \underline{A}^{n+1}$$

$$= \underline{1}$$

So that,

$$\underline{H(z)} = (\underline{1} - z\underline{A})^{-1}$$

Since $\underline{A}$ is known, $\underline{H(z)}$ may be obtained by inverting the $(\underline{1} - z\underline{A})$ matrix. The $H_{ij}(z)$ have a common denominator found by evaluating the determinant $|\underline{1} - z\underline{A}|$. Numerators can be found from the determinant cofactors, but we will show an easier method.

Once the denominator is known we can take this as the transform of an input. Then if this input is applied to node i, the resulting output transforms, E(z), at node j will be the numerators of the transfer functions $H_{ij}(z)$. Let

$$H_{ij}(z) = \frac{N_{ij}(z)}{D(z)}$$

where $D(z) = |\underline{1} - z\underline{A}|$. Then if $F_i(z)=D(z)$ we find

$$E_j(z) = F_i(z) H_{ij}(z) = N_{ij}(z)$$

Since for a graph with a finite number of states D(z) and $N_{ij}(z)$ are finite order polynomials in z, $N_{ij}(z)$ can be obtained by following through the graph operation directly by stepwise computation of node signals.

Continuing the above example to find $H_{11}(z)$:

$$(\underline{1} - z\underline{A}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - z \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix} = \begin{bmatrix} 1-z & -2z \\ 3z & 1+z \end{bmatrix}$$

$$D(z) = |\underline{1} - z\underline{A}| = 1 + 5z^2$$

Applying to the quiescent system at node 1 the signals 1,0,5 at t=0,1,2 we obtain

- At t=0: $e_1(0)=1$, $e_2(0)=0$.
- At t=1: $e_1(1)=1$, $e_2(1)=2$.
- At t=2: $e_1(2)=0$, $e_2(2)=0$.
- For t>2 no input is applied and the system remains dead.

The results show that:

$$N_{11}(z) = 1 + z; \qquad N_{12}(z) = 2z$$

and

$$H_{11}(z) = \frac{1+z}{1+5z^2}; \qquad H_{12}(z) = \frac{2z}{1+5z^2}$$

The same results would have been obtained by flow-graph reduction (check this).

The above method for determining a node-to-node transfer function is often less laborious than the flow-graph reduction method. It can be used, however, only if the graph has a finite number of nodes, and the branch transfer functions have the form assumed in this section.

6.3 Matrix description (cont.)

We now consider the problem of converting a given set of difference equations to the special form needed for the matrix description of the system. This is best explained with an example. Consider these equations:

$$\begin{aligned} 0 &= e_1(n+1) + e_2(n) - g(n+1) + f(n) \\ 0 &= g(n) - e_1(n-1) + e_2(n-2) + f(n) \\ 0 &= e_2(n) + e_1(n-1) \end{aligned}$$

Here f(n) is the input, g(n) is the output and $e_1(n)$ and $e_2(n)$ are intermediate node signals.

First, in each equation shift the term or terms with the highest time indexes to the left side (except for any term involving the input which is to remain on the right):

$$\begin{aligned} g(n+1) - e_1(n+1) &= e_2(n) + f(n) \\ g(n) &= e_1(n-1) - e_2(n-2) - f(n) \\ e_2(n) &= -e_1(n-1) \end{aligned}$$

Next, if necessary, we change the index to make n the index of all terms on the left.

$$\begin{aligned} g(n) - e_1(n) &= e_2(n-1) + f(n-1) \\ g(n) &= e_1(n-1) - e_2(n-2) - f(n) \\ e_2(n) &= -e_1(n-1) \end{aligned}$$

Now we inspect the terms on the right to see whether the index of all terms is n–1 (except terms involving the input). If any lower index appears, say n–k in e_i (n–k), k>1, then we replace e_1(n–k) with a new signal $r_{i,k-2}$(n–1), and we invent a sequence of signals, r_{ij}(n), that determines $r_{i,k-2}$(n–1) from e_i(n–k) by equations of the desired form (indices n on left, indices n–1 on right).

$$\begin{aligned} r_{i0}(n) &= e_1(n-1) \\ r_{i1}(n) &= r_{i0}(n-1) \\ r_{i2}(n) &= r_{i1}(n-1) \\ &\cdots \\ r_{i,k-2}(n) &= r_{i,k-3}(n-1) \end{aligned}$$

These equations determine

$$r_{i,k-2}(n-1) = r_{i,k-3}(n-2) = \cdots = r_{i0}(n-k+1) = e_i(n-k)$$

In our example we need to put

$$r_{20}(n) = e_2(n-1)$$

as an extra equation and substitute

$$e_2(n-2) = r_{20}(n-1)$$

in the second equation of the original set to obtain the set

$$\begin{aligned} g(n) - e_1(n) &= e_2(n-1) + f(n-1) \\ g(n) &= e_1(n-1) - r_{20}(n-1) - f(n) \\ e_2(n) &= -e_1(n-1) \\ r_{20}(n) &= e_2(n-1) \end{aligned}$$

Then we solve these equations for the variables appearing on the left. By subtracting the first equation above from the second, we obtain an equation for e_1(n). The other equations may be written as before to obtain

$$\begin{aligned} e_1(n) &= e_1(n-1) - e_2(n-1) - r_{20}(n-1) - f(n-1) - f(n) \\ g(n) &= e_1(n-1) - r_{20}(n-1) - f(n) \\ e_2(n) &= -e_1(n-1) \\ r_{20}(n) &= e_2(n-1) \end{aligned}$$

Equations A

In more difficult cases some systematic method such as the Gauss-Jordan procedure would be used to solve for the left side variables. (It's conceivable that after using this procedure some of the left side terms disappear, in which case we can reduce the system order by eliminating some states.). Now we can write equations A in the matrix form:

$$\begin{bmatrix} e_1(n) & g(n) & e_2(n) & r_{20}(n) \end{bmatrix} =$$

$$\begin{bmatrix} e_1(n-1) & g(n-1) & e_2(n-1) & r_{20}(n-1) \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ -1 & -1 & 0 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} -f(n-1)-f(n) & -f(n) & 0 & 0 \end{bmatrix}$$

The input –f(n–1) –f(n) is to be applied to the e_1 node, –f(n) to the j node, and no input to the e_2 and r_{20} nodes.

If we want the impulse response of the original system then we set f(n)=δ(n). Then since f(n)=f(n–1)=0 for n<0,

$$e_1(n) = g(n) = e_2(n) = r_{20}(n) = 0; \quad n < 0$$

For n=0 equations A yield:

$$e_1(0) = -f(0) = -1$$
$$g(0) = -f(0) = -1$$
$$e_2(0) = 0$$
$$r_{20}(0) = 0$$

For n=1 equations A yield:

$$e_1(1) = -1-1 = -2$$
$$g(1) = -1$$
$$e_2(1) = 1$$
$$r_{20}(1) = 0$$

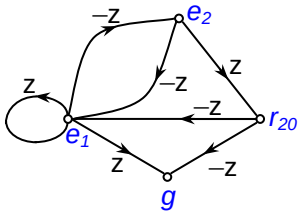
For n>1 all input terms are zero, and taking the values $e_1(1)$, g(l), $e_2(1)$, $r_{20}(1)$ as initial values, we can find all succeeding values by the matrix methods section 6.2.

$$\begin{bmatrix} e_1(n+1) & g(n+1) & e_2(n+1) & r_{20}(n+1) \end{bmatrix} =$$

$$\begin{bmatrix} e_1(1) & g(1) & e_2(1) & r_{20}(1) \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ -1 & -1 & 0 & 0 \end{bmatrix}^n$$

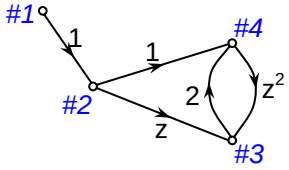
for $n \geq 0$

The graph associated with this matrix description is:



6.4 Problems

6.4.l.a) Find the system of difference equations indicated by the graph:



Write the equations also in the transfer domain.

b. By solving the transfer domain equations, find the node #1 to node #4 transfer function.

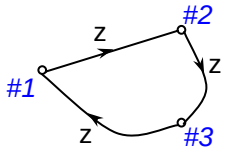
c. Find the node #1 to node #4 transfer function by flow-graph reduction.

6.4.2) Find a flow-graph representation of the following set of difference equations:

$$0 = e_1(n) - e_2(n-1) + e_3(n+1)$$
$$0 = e_1(n) - e_2(n+1)$$
$$0 = e_1(n) - 2e_2(n+1) + f(n)$$

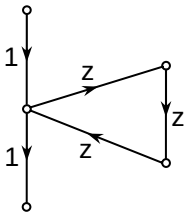
What is the input-to-e_3 node transfer function?

6.4.3.a) Write down the A matrix for the system:



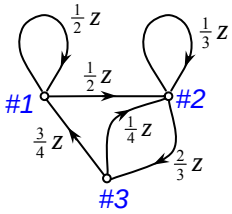
b. Compute the successive powers $\underline{A}^0$, $\underline{A}^1$, $\underline{A}^2$, $\underline{A}^3$, … and deduce the general result, $\underline{A}^n$. Interpret this result in terms of signal flow in the graph.

c. Find the node 1-to-node 1 transfer function by flow-graph reduction, i.e., find the transfer function of the system:



Find the impulse response by dividing out the transfer function. Compare results with those of b.

6.4.4) Given the flow graph:



a. What is the A matrix?

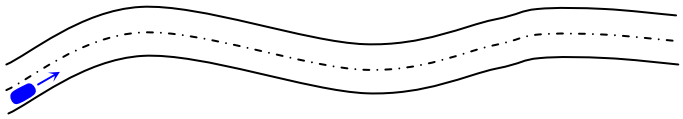
b. Find the node-l-to-nodes 1, 2, 3 transfer functions by matrix methods. As a check,

$$H_{13}(z) = \frac{\frac{1}{3}z^2}{1 - \frac{5}{6}z - \frac{1}{6}z^3}$$

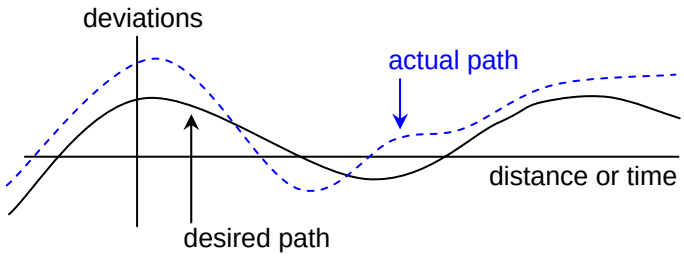
7. System problem I

7.1 One-arm driver

In the next two chapters we investigate two examples of linear system response problems, beginning with physical descriptions, proceeding to system models, and finally analyzing the behavior of the models.



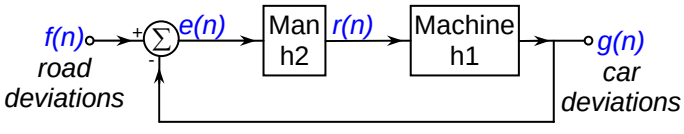
For the first example, consider a car driving down a road that meanders slightly left and right while maintaining generally the same direction. We take this direction as the abscissa and plot the right-left deviations of the road and car as ordinates. The resulting graph shows the desired and actual paths of the car.



Assuming that the car is proceeding at a constant speed, this graph is also a plot of deviations versus time.

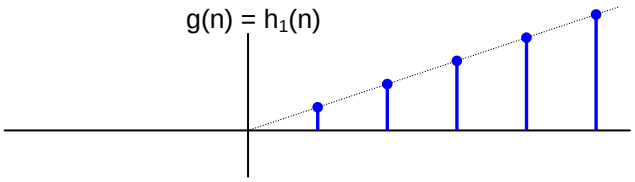
Consider the desired path-time function as the input, f(t), to some system whose output, g(t), is the actual path. The car-driver, man-machine combination is the system that's attempting to reproduce the road deflection, f(t) in the motion of the car, g(t). This kind of a system is called a control system or servomechanism.

Due to circumstances beyond his control the driver manages to get a glimpse of the road only at discrete times. For simplicity we assume that these glimpses are obtained at regular intervals. In effect, then, the driver acts on sampled data. We will only desire to know the position of the car at these same times so that the problem consists of relating the output samples, g(n), to the input samples, f(n). A reasonable model for this system is shown in the following block diagram:



This diagram states that the man observes the difference between the desired positions, f(n), and the actual positions g(n), that he processes these deviations in some way to determine how much to turn the wheel of the car, and that the car then responds to produce the output motion, g(n). We assume that the wheel is moved abruptly at the sampling times so that the car changes direction quickly at these times. Between samples the car proceeds in a straight line. The steering wheel motion is then a momentary right or left twist, the magnitude of which we call the signal r(n). (r(n) is positive for left turns, negative for right turns.)

To describe the machine portion of the system we require the relation between car motion g(n) and wheel motion r(n). Evidently a unit sample input (r(n)=δ(n)) to the quiescent machine system results in some changes in direction to the left which produces a linearly increasing deflection, $h_1(n)$.



By conveniently defining r(n) physically we can insure that r(n)=aδ(n) produces a linear ramp, g(n)=ah_1(n), with a times the rate of increase of the response to δ(n). Moreover, we see that since we are considering only small deflections, the response to r(n)=aδ(n)+bδ(n–1) will nearly be g(n)=ah_1(n)+bh_1(n–1). Generalizing this result, we see that machine superposes the partial responses to separate input samples to obtain the over-all response. In other words the machine system is linear (also time-invariant) and is therefore characterized by the impulse response $h_1(n)$. The transfer function is:

$$H_1(z) = C(z + 2z^2 + 3z^3 + \cdots) = \frac{Cz}{(1-z)^2}$$

where C is some constant.

We will postulate also that the man system is a linear, time-invariant system. Under these restrictions what is the best way for the man to generate turning signals, r(n), from the observed errors, e(n)=f(n)–g(n)? To make the problem simple, we consider only logics of restricted form. We will permit the man to make r(n) proportional to the error e(n), to the change of error e(n)–e(n-1) or

more generally to a linear combination of these. Therefore,

$$r(n) = C_1 e(n) + C_2(e(n) - e(n-1))$$

The transfer function, $H_2(z)$, of the man system is then:

$$H_2(z) = C_1 + C_2(1-z)$$

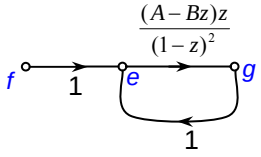
In other words the man makes use of at most the current and immediately preceding errors to compute a turning signal. We should expect C_1 and C_2 to be positive; then an error or changing error would bring about the correct compensating action.

Since H_1 and H_2 are in series they may immediately be combined into an over-all transfer function:

$$H_1(z)H_2(z) = (C_1 + C_2(1-z))\frac{Cz}{(1-z)^2}$$

By writing $A = C(C_1+C_2)$, $B=CC_2$ we simplify the expression to:

$$H_1 H_2 = \frac{(A-Bz)z}{(1-z)^2}$$



By reducing the graph we find the system transfer function:

$$H_g(z) = \frac{G(z)}{F(z)} = \frac{(A-Bz)z}{(1-z)^2 + (A-Bz)z}$$

If we take e(n) as the output the transfer function is:

$$H_e(z) = \frac{E(z)}{F(z)} = \frac{(1-z)^2}{(1-z)^2 + (A-Bz)z}$$

(Verify these reductions.)

All of the subsystems of the above system were physically realizable, i.e., did not respond until excited. Therefore, the transfer functions H_1, H_2, H_g, H_e all have regions of convergence near z=0 in the z-plane where the power series represent the time functions. This means that we can expand H_e or H_g by long division, since only series ascending in powers of z are required. We note also that H_e and H_g have the same denominator polynomial:

$$D(z) = (1-z)^2 + (A-Bz)z = 1 + (A-2)z + (1-B)z^2$$

which means that they have the same poles (two poles).

7.2 Stability

It's essential that the system be stable. Otherwise the system may oscillate wildly because the errors are overcorrected or may compound the error by corrections in the wrong direction.

To test the system for stability we shock it by applying a unit sample at t=0 and then sit back to see whether the resulting error eventually decays to zero or whether it builds up with increasing time. In other words, we inspect $H_e(z)$ to see whether the time signals, $h_e(n)$, tend to zero. $H_e(z)$ can be expanded in a partial fraction expansion of the form:

$$H_e(z) = a_0 + \frac{a_1}{1-b_1 z} + \frac{a_2}{1-b_2 z}$$

$$= a_0 + a_1(1 + b_1 z + b_1^2 z^2 + \ldots) + a_2(1 + b_2 z + b_2^2 z^2 + \ldots)$$

The condition that $h_e(n) \to 0$ as $n \to \infty$ is just that $|b_1| < 1$, and $|b_2| < 1$. Since $(1-b_1 z)\cdot(1-b_2 z) = D(z)$, the denominator polynomial of $H_e(z)$, the condition for stability is that $H_e(z)$ have all poles outside the unit circle in the z-plane. We could also have used $H_g(z)$ to determine stability.

Now we want to determine those values of A and B in our system that result in stable operation. We find those values of A, B for which both zeros of:

$$D(z) = 1 + (A-2)z + (1-B)z^2$$

lie outside the unit circle. It will be more convenient, however, to consider the equivalent problem of finding A, B leading to zeros of the polynomial

$$D_1(w) = w^2 + (A-2)w + (1-B) = (w-b_1)(w-b_2)$$

inside the unit circle in the w-plane.

The roots of $D_1(w)$ are complex if:

$$(A-2)^2 - 4(1-B) = A^2 - 4A + 4B < 0$$

or

$$B < A\left(1-\frac{A}{4}\right)$$

In this case (1–B) is the square of the magnitude of each root and hence (1–B)≥0 and the stability condition is:

$$|1-B|<1$$

or

$$0<B\le 1$$

The roots of $D_1(w)$ are both real if.

$$B \ge A\left(1-\frac{A}{4}\right)$$

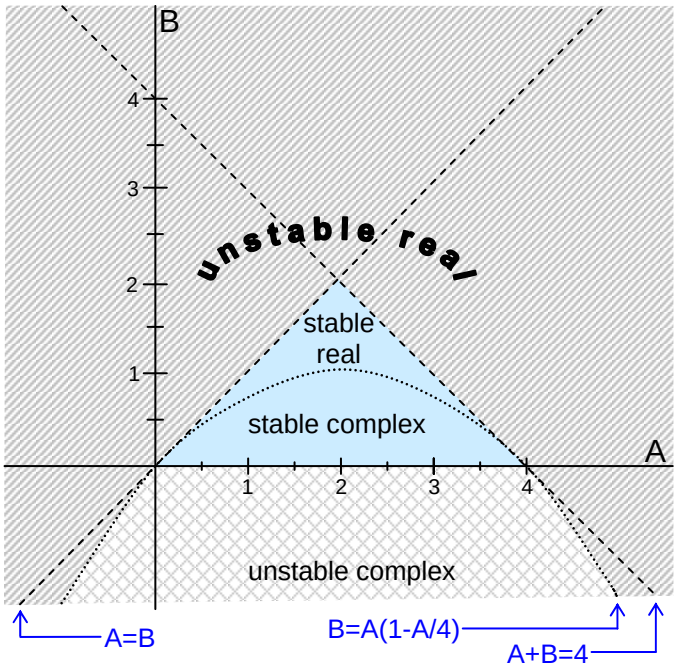
Boundary curves separating stable and unstable values of A, B are gotten by setting w=1 as one root,

$$D_1(1)=0=1+(A-2)+\ (1-B)$$
$$0=A-B$$

and w=–1 as another root.

$$D_1(-1)=0=1-(A-2)+\ (1-B)$$
$$0=4-A-B$$

All relations above may be given in a plot of the AB-plane showing the loci separating values corresponding to complex and real roots and to stability and instability. (We leave further details to the reader.) The diagram shows a triangle of stable values of A, B.



In more complicated problems than the above we can still readily delineate the regions of stability and instability by algebraic expressions involving the coefficients of the transfer function denominator polynomial. This determination is based on the classic test for Hurwitz polynomials (polynomials with all zeros in the left-half plane). The procedure is given in appendix I.

7.3 Interpretations

We are now able to draw some conclusions about the effectiveness of proposed procedures that the man could use to control the car. Suppose first, B=0. The reader will verify that this condition means that the driver generates a turn at t=n which is proportional to the error e(n) observed at this time. The velocity estimate e(n)–e(n-1) is not used. As might be expected, the stability diagram shows that the system is unstable if A is negative. In that case the driver turns in the direction increasing the error and the car continues to diverge from the proper path.

For example, take A=–½ and B=0. Then,

$$H_e(z)=\frac{1-2z+z^2}{1-\frac{5}{2}z+z^2}=1+\frac{\frac{1}{2}z}{(1-2z)(1-\frac{1}{2}z)}$$

$$=1+\frac{-\frac{1}{3}}{1-\frac{1}{2}z}+\frac{\frac{1}{3}}{1-2z}$$

The term (1/3)/(1–2z) provides a geometrically increasing error component.

Now consider the case 0<A<4, and B=0. Here the man acts in the correct sense to compensate for the error, but the degree of correction is held within moderate bounds. After a cursory look at the problem, this procedure might naturally seem to be the best. However, our analysis shows that the system would verge on instability. As the stability diagram shows, the polynomial roots are complex and lie on the unit circle in the z-plane (since the system is on the border between stability with roots outside and instability with roots inside). The result is that under this type of control the car weaves from side-to-side after a disturbance, the magnitude of the oscillations neither increasing nor decreasing with time. For example take A=2, B=0.

$$H_e(z)=\frac{(1-z)^2}{1+z^2}=1+\frac{-2z}{1+z^2}$$

$$=1-2z+2z^3-2z^5+2z^7\cdots$$

Here the oscillations are twice the amplitude of the original disturbance and have a period between maxima of four sampling periods.

As A varies from A=0 to A=4 (B=0), the oscillations change from slow to rapid oscillations. At A=4 the successive maxima occur every two sampling periods. For A>4 the oscillations have the same frequency but increase in amplitude. This behavior results from a pole of the response on the negative real axis inside the unit circle in the z-plane. For example take A= 9/2 and B=0.

$$H_e(z) = \frac{(1-z)^2}{1-\frac{5}{2}z+z^2} = 1 + \frac{-\frac{9}{2}z}{(1-\frac{1}{2}z)(1+2z)}$$

$$= 1 + \frac{-3}{1+\frac{1}{2}z} + \frac{3}{1+2z}$$

Expansion of the last term provides the increasing oscillation component.

A glance at the stability plot shows further that if we choose A, B so that A=B the system will still verge on instability at best. The case A=B results from using only the difference e(n)–e(n-1) to generate turning corrections. That is, only the rate of change of the error is used to provide corrections. For example take A=B=2, then

$$H_e(z) = \frac{(1-z)^2}{1-z^2} = -1 + \frac{2}{1+z}$$

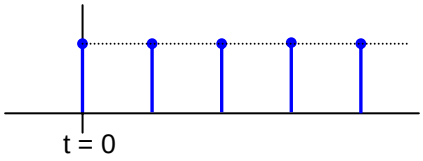
The last term yields a steady amplitude of oscillation.

It's apparent, then, that both error and velocity of error will be needed to properly stabilize the system. Appropriate A,B will be selected from the interior of the stability region.

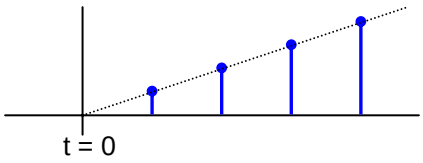
7.4 Test function inputs

To make a more specific choice of A, B we will apply some test inputs to the system and attempt to gauge the quality of the responses. These test inputs are chosen to be simple but also to strain the mathematical mechanism of the system so that weaknesses become apparent. A particularly appropriate set of test inputs consists of the unit sample, unit step, unit linear ramp, unit parabola etc. (We have used the unit sample in our stability discussion.)

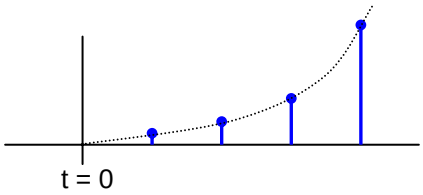
Input	Transform
Unit sample	1
Unit step	$\frac{1}{1-z} = 1 + z + z^2 + \cdots$
Unit ramp	$\frac{z}{(1-z)^2} = z + 2z^2 + 3z^3 + \cdots$
Unit parabola	$\frac{z(1+z)}{(1-z)^3} = z + 4z^2 + 9z^3 + \cdots$



The unit step results from a roadway that is suddenly displaced to the left.



The unit ramp results from a sharp change of direction in the road.



The unit parabola serves as an approximation to a constant radius turn (beginning at t=0).

Applying the unit step to the system, the error is:

$$E(z) = \frac{1}{1-z} \cdot H_e(z) = \frac{1-z}{D(z)}$$

where $D(z)=(1–z)^2 + (A–Bz)z$. The error for a unit ramp is:

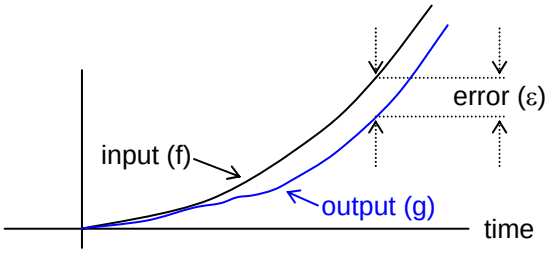
$$E(z) = \frac{z}{(1-z)^2} \cdot H_e(z) = \frac{z}{D(z)}$$

The error for a unit parabola is:

$$E(z) = \frac{z(1+z)}{(1-z)^3} \cdot H_e(z) = \frac{z(1+z)}{(1-z)D(z)}$$

Now if the system is stable the factors of D(z) contribute only decaying transient components to each of these e(n) responses. So for the unit step and unit ramp, the error tends to zero with increasing time and the car eventually gets back on the proper path. However, for the parabola, there is an additional pole in the response at z=1 contributed by the (1–z) factor which appears with D(z). This adds a term of the form ε/(1–z) in the partial

fraction expansion, and the result is a steady-state lag error of ε in following the curve after the initially important transients have decayed.



The magnitude of this lag is a gauge of the tightness of the response.

We may evaluate it directly from the parabola error response,

$$\varepsilon = (1-z)E(z)\Big|_{z=1} = \frac{2}{D(1)} = \frac{2}{A-B}$$

To reduce the lag error, then, we should make A large and B small. The stability diagram shows that the best point to pick is in the corner A=4, B=0. However we know that the response is here subject to undiminished oscillations. Therefore, in adjusting A, B we must strike a balance between loose control leading to big lag errors and tight control leading to pronounced oscillations.

7.5 Sums of squares of error

To estimate the amount of oscillation or other error in the response we can compute the sums of squares of all the error samples e(n). Suppose e(n) is the result of a test input like the above. Then its transform has the expanded form:

$$E(z) = e(0) + e(1)z + e(2)z^2 + e(3)z^3 + \cdots$$

Change z to z^{-1} in E(z),

$$E(z^{-1}) = e(0) + e(1)z^{-1} + e(2)z^{-2} + e(3)z^{-3} + \cdots$$

and multiply E(z) and $E(z^{-1})$ term by term. We write only the term of zero order in z explicitly.

$$E(z)E(z^{-1}) = e(0)e(0) + e(1)e(1) + e(2)e(2) + \cdots$$
$$+[\ ]z + [\ ]z^2 + [\ ]z^3 + \cdots + [\ ]z^{-1} + [\ ]z^{-2} + [\ ]z^{-3} + \cdots$$

The zero order term is evidently the sum of squares we seek. We assume that the system is stable, i.e., that $e(n) \to 0$ geometrically as $n \to \infty$ and, therefore, the function $E(z)E(z^{-1})$ has expansion coefficients which also tend to zero as $n \to \infty$ and $n \to -\infty$. The zero order term can be found by an appropriate partial fraction expansion of $E(z)E(z^{-1})$.

As an example of a sum of squares calculation consider the error response to the unit ramp in our control system

$$E(z) = \frac{z}{(1-z)^2 + (A-Bz)z}$$

To simplify the algebra we restrict ourselves to A=2 (the center line of the stability region) and only allow B to vary. Then,

$$E(z) = \frac{z}{1+(1-B)z^2} = z - (1-B)z^3 + (1-B)^2 z^5 - \cdots$$

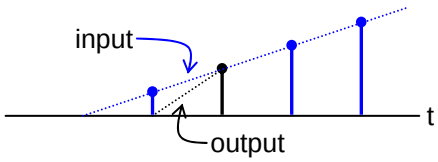
This example does not require us to actually take $E(z)E(z^{-1})$ in closed form and find the central sample by partial fraction expansion since we can sum the squares of the errors directly. Assuming 0<B<2 (stability):

$$\sum_{n=0}^{\infty} e^2(n) = 1 + (1-B)^2 + (1-B)^4 + \cdots = \frac{1}{1-(1-B)^2}$$

The minimum sum of squares of one is obtained with B=1. So A=2, B=1 appears to be a good choice for system constants. For instance, the ramp error is then

$$E(z) = z$$

which means that if the road changes direction abruptly, the car is back on the right track just one sample after the error is observed. A=2, B=1 results in an equal weighting of error and rate of change of error in generating corrective action.



7.6 Problems

In the control system of this example try the case A=1, B=1. What is the error resulting from application of

a) A unit sample at t=0 ?
b) A unit ramp ?
c) A unit parabola ?

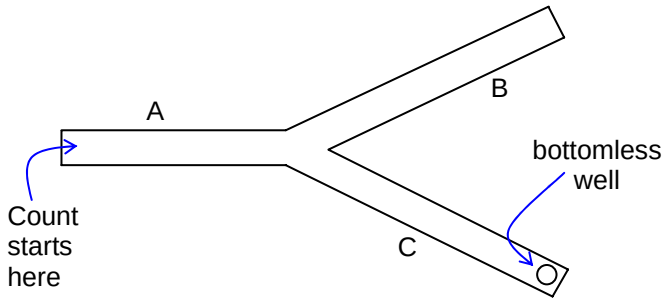
Sketch what would happen (plot road and car positions vs. time) in each case. Why is A=1, B=1 a poor choice of system constants?

8. Systems problem II

8.1 The count of Monte Cristo

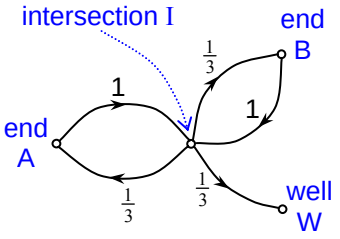
The second example draws a systems problem from the probability theory. The system variables are abstract probabilities rather than physical quantities as in chapter 7. Nevertheless these probabilities are related by a mathematical mechanism that is similar to that which relates physical system quantities.

This is the problem. The count of Monte Cristo has been captured by his enemies and thrown in a dazed condition into a dark dungeon with corridors as shown:



The count starts at the left end of corridor A and proceeds down it until he comes to the intersection of the corridors. Here he stumbles around and chooses to go down corridor A, B, or C with equal probability. At the end of A or B he hits a wall and must return. Each time the intersection is reached he makes a new random choice, as before. Little does he know that at the end of corridor C is a bottomless well set even with the floor. If the count goes down C, he will drop to his death and his enemies, who are betting on the outcome, will celebrate his demise. It takes just one hour to move down the length of each corridor. What is the probability that the Count will leave this world on the nth hour after the beginning of the process? Is the Count sure to go? What is the average time required for such a game to end?

In probability language, the ends of corridors A and B are reflecting barriers and the end of corridor C is an absorption barrier. We will obtain all relevant information by sampling the process at hour intervals after the game begins. At these times the count is either at the end of corridor A or B, at the central intersection, or on his way down the well. Nodes of a flow graph indicate these states of the process. Directed branches indicate the possible movements between these states. No branch returns from the well W.



On each branch we show the transition probabilities for moving in the indicated direction. So all transitions out of the intersection I, have probability 1/3, while the return from ends A, B is certain with probability 1.

This random process is called a discrete Markov process with constant transition probabilities. It's discrete because we only look at the process at hour intervals and we can even consider the count to make abrupt transitions from state to state at these times. The Markov property means that the transitions out of any state are not modified by the previous history of the process.

It should be apparent that if p_A, p_B, p_I, p_W are the probabilities of being in states A, B, I or being absorbed in W, then these quantities at the nth hour are related to their values at the (n-1)st hour by the following set of difference equations:

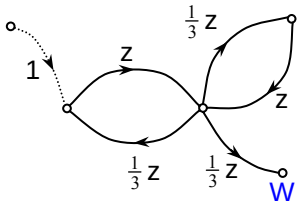
$$\begin{aligned} p_A(n) &= \tfrac{1}{3} p_I(n-1) \qquad n = 1,2,3,\cdots \\ p_B(n) &= \tfrac{1}{3} p_I(n-1) \\ p_I(n) &= p_A(n-1) + p_B(n-1) \\ p_W(n) &= \tfrac{1}{3} p_I(n-1) \end{aligned}$$

In matrix form these equations become:

$$\begin{bmatrix} p_A(n) & p_B(n) & p_I(n) & p_W(n) \end{bmatrix} =$$

$$\begin{bmatrix} p_A(n-1) & p_B(n-1) & p_I(n-1) & p_W(n-1) \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrix of constants is the transition probability matrix. From chapter 6 we know that these difference equations or matrices describe a linear time-invariant system with a graph which in this case is identical in form with the flow graph of the Markov process above. The node probabilities of the process are the time signals of our system. The system flow graph consists of branches with transfer functions of the form $p_{ij}z$, where p_{ij} is the transition probability of the i, j branch and z is the unit delay transfer function.



Since the count starts in state A, $p_A(0)=1$, $p_B(0)=p_I(0)=p_W(0)=0$, and we must externally introduce a unit probability value into node A of the quiescent system at t=0. We accomplish this by adding an extra input branch to the system at A and applying a unit sample at t=0.

8.2 The solution

The transform of the probabilities $p_W(n)$ of being absorbed on the nth hour is just the system transfer function from the input to node W. Using node I as a residual node we reduce the flow graph and find the transfer function

$$P_W(z)=\frac{\frac{1}{3}z^2}{1-\frac{2}{3}z^2}=\frac{1}{3}z^2+\frac{1}{3}\cdot\frac{2}{3}z^4+\frac{1}{3}\left(\frac{2}{3}\right)^2z^6+\cdots$$

So the probability of absorption on the nth hour is:

$$p_W(n)=\begin{cases}0 & n\le 0 \text{ or } n \text{ odd}\\ \frac{1}{3}\left(\frac{2}{3}\right)^{\frac{n-2}{2}} & n \text{ even and positive}\end{cases}$$

The probability of eventual absorption is the sum of these probabilities which we obtain by setting z=1.

$$P_W(1)=\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{3}+\cdots=\left.\frac{\frac{1}{3}z^2}{1-\frac{2}{3}z^2}\right|_{z=1}=1$$

Absorption is certain. The mean time to absorption is by definition,

$$\bar{t}=\sum_{n=0}^{\infty}np_W(n)$$

This average can be generated from $P_W(z)$ (think of the expanded form) by differentiating by z (this brings each n coefficient down) and then by setting z=1 (this adds all the terms). The result is

$$\left[\frac{d}{dz}P_W(z)\right]_{z=1}=\left[\frac{d}{dz}\sum_{n=0}^{\infty}p_W(n)z^n\right]_{z=1}=\sum_{n=0}^{\infty}np_W(n)$$

In particular, for the example,

$$\left.\frac{d}{dz}P_W\right|_{z=1}=\left.\frac{\frac{2}{3}z}{(1-\frac{2}{3}z^2)^2}\right|_{z=1}=6$$

The process will last six hours on the average if repeated again and again. In a similar way, we can compute the variance of absorption times or higher moments of the absorption time distribution.

8.3 Mean absorption time

There is a second method for computing the mean absorption time of a process like that of the example. This alternative also gives us the mean time spent in making transitions to each state of the process.

Since the sum total of all signals entering a given node of the system flow graph at t=n is the probability or average number of times that a transition is made to this state at t=n, the sum of these sums over all time is the mean time spent in making transitions to this state. Then the sum of these quantities over all states is the mean time spent cycling through the entire process. It's often possible to determine the delays associated with each state by a stepwise procedure. Then by adding these delays, the over-all delay is found. In this process we omit counting the unit signal fed to the graph at t=0, for we don't attribute any delay setting the initial conditions.

In the Monte Cristo example we know that a transition into W is certain. Therefore a total signal of 1 enters this node throughout all time and the march of Monte Cristo down the last corridor will contribute an average of 1 hours delay to the entire process. Then since the total signal flow into W comes from I through a transition probability of 1/3 a total of 3 must have started from I. So there will be an average of 3 hours spent in moving into I. The signals at A and B are derived from the signal at I by multiplying by 1/3. Thus the delay entailed in moving into A and B averages one hour each. Adding all of these delays, the over-all mean time to absorption is 1+3+1+1=6 hours, which checks the previous result.

8.4 Problems

8.4.1) Suppose the count has bribed one of his jailers to leave a door open at, the end of corridor B. So if the count wanders down B, he escapes. Assuming that the choice at the corridor intersection, I, is made as before (the count cannot see the door from I and does not know which of the corridors will be his way to freedom)

a.) What is the probability that the count escapes?
b.) That he dies in the well?
c.) That he dies of old age wandering the corridors?
d.) What is the mean time required to escape considering only those instances where escape is the outcome of the game?

Use transform methods.

9. Random Inputs and Correlation Functions

9.1 Random-time functions

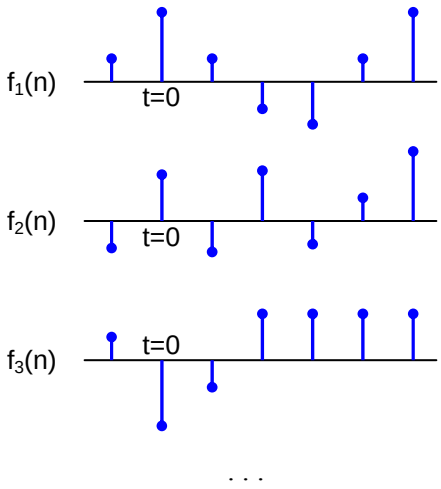
In examples of the last two chapters all system inputs, outputs, and intermediate quantities were prescribed or determined as definite functions of time. In the remaining chapters we will be concerned mainly with system time functions that are described statistically. This means, for example, that if f(n) is the system input, we won't actually know the values of f(n) for n=0, ±1, ±2, ... Instead, only the statistical distributions of the values of f(0),f(l),f(–1),... will be known. Such an f(n) is called a random input.

It's clear that if the input is random, it will be appropriate and generally necessary to discuss the output response and other system quantities likewise in statistical terms. Certain averages of combinations of f(n) will supply sufficient information about f(n) for our purposes.

Suppose f(n) is a random-time function. The values, f(0), f(1), f(–1),… are considered to be random variables. Theoretically, a complete statistical description of this sequence of random variables can be given by prescribing all of the joint probability distributions involving a finite number of the f(n). The only condition to be satisfied is that the set of distributions be so prescribed that if a distribution of lower order (involving fewer f(n)'s) is derived from one of higher order (involving additional f(n)'s), the derived distribution is identical to the distribution originally prescribed for these f(n). Such a set of distributions is said to be consistent. Among the distributions to be so prescribed are the distributions of the individual f(n); n=0,±1,… and the joint distributions of pairs (f(n), f(m)); n,m=0,±1,… .

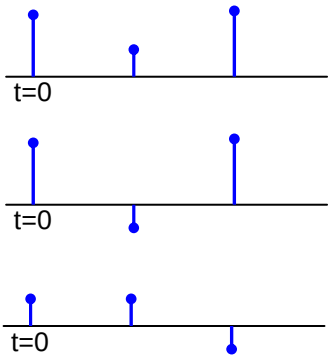
It's not always necessary to define all of the finite order distributions explicitly. For instance, if we postulate that the f(n) values are mutually statistically independent, then a knowledge of just the one-dimensional distributions of f(n); n=0,±1,… is sufficient to determine all higher distributions. A generalization of this procedure of describing random functions by building higher order distributions from lower order ones using supplementary conditions will be discussed later when we show how to generate random f(n) by Markov processes similar to that of chapter 8.

Knowing the defining distributions, we can, in principle, generate a possible sequence of f(n) values by flipping coins, consulting random number tables, or using other randomizing devices in accordance with the frequencies of occurrences specified by the distribution. If we perform this determination over again a new and generally different f(n) is generated. The set of all f(n) which can be so obtained is called an ensemble of random inputs. Although there are generally an uncountable number of such possible functions we will draw a picture of the ensemble by plotting a sequence of the functions with time origins aligned.



. . .

Since we are going to apply f(n) to time-invariant systems we wi1l find it necessary to restrict ourselves to f(n) whose statistics also are invariant in time. This statistical time-invariance, called stationarity, results when the distributions defining the set of f(n) values are the same as those defining the set of f(n+k) values for any fixed k. For example, for stationary f(n), the one-dimensional distributions of etc., are all the same as are the joint distributions of. the pairs (f(0), f(l)), (f(1), f(2)), (f(–1), f(0)), etc. In terms of the ensemble, a shift k units of time to the left or right leads to a new ensemble containing the same functions (in a different order) as the original. Stationarity means statistical homogeneity in time. The following ensemble is apparently not that of a stationary process.



Since alternate sample values are zero, a shift of an odd number of time units to the right or left changes the character of the ensemble. (In particular, the distribution of f(0) is not the same as the distribution of f(1).)

9.2 Time and ensemble averages

Systems analysis based on a response to stationary random inputs must content itself with computing some kind of average system behavior. The exact f(n) applied in some future operation of the system is not known. What is known is that f(n) will be selected from the ensemble according to the probability distributions that describe that collection of inputs. So it's appropriate to average the behavior over all members of the ensemble and use this ensemble average as a gauge of system performance.

To use such averages we must restrict the probability distributions to insure that these averages exist. From now on we will assume that the stationary input values, f(n), have one-dimensional distributions with finite first and second moments (means and mean squares). By stationarity, these moments are independent of n.

The simplest instance of the computation of an average arises if we consider a system with a transfer function of 1 so that the output g(n) is identical to the stationary input f(n), Then suppose we want to find, $\overline{g(n)}$, the average value of the output. $\overline{g(n)}$ is the average (statistical mean) over all members of the input ensemble of the output value at t=n, Since f(n)=g(n), this average is just the average value of f(n) at t=n. This average is determined by the distribution of values f(n) at t=n, Since f(n) is stationary, the average is the same for all n and so $\overline{g(n)}$ is simply a constant equal to the mean of the f(n) distribution.

A second kind of average is the time-average of f(n) (or g(n)). To take a time average, we think of observing the values of just one possible time function, $f_1(n)$, of the ensemble and form the limit as $N\to\infty$ of

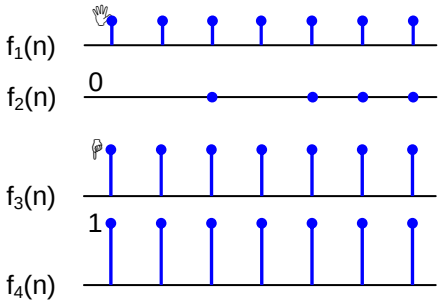
$$\frac{1}{2N+1}\sum_{k=-N}^{N} f_1(k)$$

This limit, if it exists, defines a time average of $f_1(n)$. It can be shown that if f(n) is stationary with a finite mean, then the limit exists for all member functions of the f(n) ensemble except, possibly, for a set of such functions which occurs with only zero probability.

The question arises as to whether ensemble and time averages are the same. (Note $\overline{f(n)}$ is a constant independent of n.)

$$\overline{f(n)} = \overline{f} \stackrel{?}{=} \lim_{n\to\infty}\frac{1}{2N+1}\sum_{k=-N}^{N} f_1(k)$$

That these averages are not necessarily equal is seen by considering the following stationary ensemble.



Here f(0) is uniformly distributed between 0 and +1 and all other f(n) are equal to this value of f(0). Then

$$\overline{f(n)} = \frac{1}{2} \qquad for\ any\ n$$

but the time averages, although they exist for each function, vary from function to function. For f_2 the time average is 0, for f_4 it is 1.

If almost all (i.e. with probability 1) member functions of a stationary ensemble have a time average equal to the ensemble average then it is said to be an ergodic ensemble. This means that (in contrast with the preceding example) nearly all member functions of an ergodic ensemble resemble each other. It can be shown that if the f(n) are independently and identically distributed, then the f(n) ensemble is ergodic.

The detailed verification of ergodic properties is a difficult problem, and we cannot treat it here. Although we would like to have time and ensemble averages the same so that the average behavior of a system over time with a particular input could be computed by averaging at one fixed time over many inputs, we won't demand this equivalence. We will take the ensemble average as the primary means of measuring system performance.

9.3 Correlation functions

The most common, useful, and mathematically tractable measure of system performance with a stationary random input is the average of the squares of some system quantity. In the control system example of chapter 8 the most interesting quantity was e(n), the difference between input and output. Small e(n)'s meant good performance. With a random input, an appropriate measure of the magnitude of the e(n)'s would be the mean square error $\overline{[e(n)]^2}$. Since we expect to find that e(n) is stationary as well as the input, f(n), this mean-square error will be a constant $\overline{e^2}$, independent of n. If we desire an error measure that varies linearly with the size of e(n) then we can use the square root of this mean-square-error, the root-mean-square or r.m.s error, $\sqrt{\overline{e^2}}$. If the e(n) ensemble is ergodic, then e^2 is also the time average of squared errors for almost all member functions.

Besides the mean-square, $\overline{f^2}$, of a stationary random function f(n), we will also be concerned with averages of products such as $\overline{f(n)f(n+1)}$, $\overline{f(n)f(n+2)}$, $\overline{f(n)f(n-1)}$, etc. In general we will seek φ(k), the average of products of f(n) separated by k units of time

$$\varphi(k) = \overline{f(n)f(n+k)}$$

We have made use of the stationarity of f(n) which implies that the averages $\overline{f(n)f(n+k)}$ do not depend on the value of n, but only on the lag k between the two random variables which compose the product. φ(k) is called the autocorrelation function of f(n). φ(k) is symmetric about k=0 since by stationarity,

$$\varphi(-k) = \overline{f(n)f(n-k)} = \overline{f(n+k)f(n)} = \varphi(k)$$

Note that φ(0) is just the mean square value of f(n).

Similarly we can form averages using two functions, f(n) and g(n),

$$\varphi_{fg}(k) = \overline{f(n)g(n+k)}$$

$\varphi_{fg}(k)$ is called the cross correlation function of f(n) and g(n). It's independent of n if we assume f(n) and g(n) are stationary. If the functions are taken in reverse order a generally different cross-correlation function results

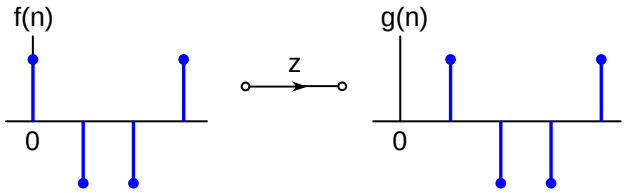
$$\varphi_{gf}(k) = \overline{g(n)f(n+k)}$$

These cross-correlation functions are not necessarily symmetric about k=0 but are related as follows:

$$\varphi_{gf}(k) = \overline{g(n)f(n+k)} = \overline{g(n-k)f(n)} = \varphi_{fg}(-k)$$

Therefore if one cross-correlation function is known the other is found by reflection about k=0.

As correlation function example, suppose that f(n) consists of a sequence of +1 and –1 values. For each n the value is to be selected independently by flipping a balanced coin (probabilities 1/2). This stationary random f(n) is applied to a system with unit delay. The output, g(n), is also obviously stationary.



From the definitions and stationarity property, the autocorrelation functions $\varphi_{ff}(k)$ and $\varphi_{gg}(k)$ of f(n) and g(n), respectively, are identical. We compute values of φ_{ff} by multiplying possible products f(n)f(n+k) by their respective probabilities to obtain the average. Using independence of the component variables for k≠0, we obtain:

$$\varphi_{ff}(0) = \tfrac{1}{2}(1)(1) + \tfrac{1}{2}(-1)(-1) = 1$$

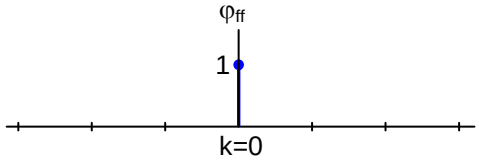
$$\varphi_{ff}(1) = \tfrac{1}{4}(1)(1) + \tfrac{1}{4}(1)(-1) + \tfrac{1}{4}(-1)(1) + \tfrac{1}{4}(-1)(-1) = 0$$

$$\varphi_{ff}(2) = 0$$

…

Therefore,

$$\varphi_{ff}(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

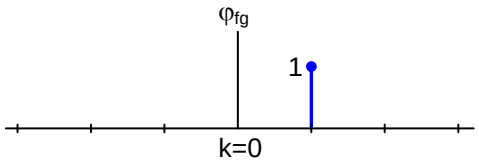


On the other hand, since g(n)=f(n–1) the cross-correlation function is:

$$\varphi_{fg}(k) = \overline{f(n)g(n+k)} = \overline{f(n)f(n+k-1)} = \varphi_{ff}(k-1)$$

so that

$$\varphi_{fg}(k) = \begin{cases} 1 & k = 1 \\ 0 & k \neq 1 \end{cases}$$



As the above example suggests, the autocorrelation function attempts to measure the degree to which the values of the time function at t=n influence their values at t=n+k. A similar comparison between different time functions is made by the cross-correlation function.

The autocorrelation function has a maximum at k=0 and generally tends to decrease as $k \to \infty$, as the outlying samples of the random signal at t=n+k lose their dependence on signal values at t=n. To show that $\varphi_{ff}(0)$ is always a maximum we use the fact that a mean-square is always non-negative.

$$0 \le \overline{[f(n) \pm f(n+k)]^2}$$

$$= \overline{f(n)f(n) \pm 2f(n)f(n+k) + f(n+k)f(n+k)}$$

$$= \varphi_{ff}(0) \pm 2\varphi_{ff}(k) + \varphi_{ff}(0)$$

$$\varphi_{ff}(0) \ge \left|\varphi_{ff}(k)\right|$$

We see that not every function, symmetric about k=0 can be an autocorrelation function. In fact, there are further constraints that $\varphi_{ff}(k)$ must satisfy as we shall see later.

Note that the restriction to f(n) with finite mean-square $\varphi_{ff}(0)$ is sufficient to insure the finiteness of all the autocorrelation values, $\varphi_{ff}(k)$. Therefore, the restrictions noted earlier on moments of the distribution of f(n) ($\overline{f}$, $\overline{f^2}$ exist and are finite) insure that the autocorrelation function of f(n) exists.

9.4 Correlation functions, further properties

We have seen that the autocorrelation function of f(n) provides the mean-square, $\overline{f^2}$ as the value

$$\overline{f^2} = \varphi_{ff}(0)$$

Also if the statistical dependency between f(n) and f(n+k) decreases and becomes negligible with increasing k then we expect to find

$$\varphi_{ff}(k) = \overline{f(n)f(n+k)} \approx \overline{f(n)} \cdot \overline{f(n+k)} = \overline{f}^2$$

(for large k)

That is, the asymptotic value of for φ_{ff} for $k \to \pm\infty$ is the square of the mean value of f.

Next we find the autocorrelation function of s(n)=f(n)–c where c is some constant subtracted from all samples of f(n).

$$\varphi_{ss}(k) = \overline{s(n)s(n+k)} = \overline{(f(n)-c)(f(n+k)-c)}$$

$$= \overline{f(n)f(n+k)} - c\,\overline{f(n)} - c\,\overline{f(n+k)} + c^2$$

$$= \varphi_{ff}(k) - 2c\overline{f} + c^2$$

If $\varphi_{ff}(k) \to \overline{f}^2$ as $k \to \pm\infty$ then by taking $c = \overline{f}$ we find that $\varphi_{ss}(k) \to 0$ as $k \to \pm\infty$. Thus it wi1l sometimes be convenient and often necessary to make the autocorrelation function tend to zero asymptotically by adding or subtracting a constant value from all samples to obtain a new function of zero mean. In most of the future discussion we will take pains to define the input so that it has a zero mean and has an autocorrelation tending to zero as $k \to \pm\infty$.

If the input is composed of the sum of two statistically independent functions with zero mean, f(n)=r(n)+s(n), $\overline{r} = \overline{s} = 0$, then the autocorrelation of f is the sum of those for r and s:

$$\varphi_{ff}(k) = \overline{[r(n)+s(n)][r(n+k)+s(n+k)]}$$

$$= \overline{r(n)r(n+k)} + \overline{r(n)s(n+k)} + \overline{s(n)r(n+k)} + \overline{s(n)s(n+k)}$$

$$= \varphi_{rr}(k) + \varphi_{ss}(k)$$

9.5 Correlation functions and system relations

The reason for describing random signals in terms of correlation functions is that they provide just the required information for computing mean-square values of various signals. For instance suppose we define the system error, e(n), to be the difference between the output, g(n), and the input, f(n+l), one unit of time in the future.

$$e(n) = f(n+1) - g(n)$$

This definition would be appropriate for a system whose job is to predict the future values of the random input. The mean-square error can then be expressed in terms of correlation functions of f(n) and g(n),

$$\overline{e^2} = \overline{[f(n+1) - g(n)]^2}$$

$$= \overline{f(n+1)f(n+1)} - 2\overline{f(n+1)g(n)} + \overline{g(n)g(n)}$$

$$= \varphi_{ff}(0) - 2\varphi_{fg}(-1) + \varphi_{gg}(0)$$

In most problems we will be given the autocorrelation function of the system input and the system impulse response and be required to derive the other correlation functions from this information.

The first question that arises is whether the output, g(n), is defined. Since g(n) is given by the superposition summation,

$$g(n) = \sum_{m=-\infty}^{+\infty} h(m) f(n-m)$$

the question is whether the sum on the right converges for all (or almost all) member functions of the input ensemble. Since f(n) in each such function are identically distributed (stationarity), f(n) doesn't tend to become smaller as n→+∞ or –∞. Therefore, to insure convergence of the sum we must make h(n)→0 as n→±∞. It can be shown that if f(n) is stationary, if f(n) is finite, and if h(n)→0 at a geometrical rate as n→±∞, then g(n) will be defined for almost all (with probability one) input functions. Then since f(n) is stationary, g(n) will also be stationary and φ_{fg} and φ_{gg} can be defined as in section 9.3.

From now on we will assume that the above conditions hold. Furthermore, we will only consider systems whose transfer functions are rational functions of z. The condition that h(n)→0 geometrically as n→±∞ is then equivalent to saying that there are no poles on the unit circle in the z-plane. The transfer function H(z) is related to h(n) by a power series expansion which converges in an annulus containing the unit circle. The following summations and other mathematical operations are justified by these assumed properties of h(n).

To find $\varphi_{gg}(k)$ from $\varphi_{ff}(k)$ we write g(n) in terms of the superposition summation and substitute in the definition of φ_{gg}.

$$\varphi_{gg}(k) = \overline{g(n)g(n+k)}$$

$$= \overline{\sum_{m=-\infty}^{+\infty} h(m) f(n-m) \sum_{j=-\infty}^{+\infty} h(j) f(n+k-j)}$$

Changing the dummy variable in the second summation from j to m+j and summing over the same infinite range of j we obtain,

$$\varphi_{gg}(k) = \overline{\sum_{m=-\infty}^{+\infty} h(m) f(n-m) \sum_{j=-\infty}^{+\infty} h(m+j) f(n-m+k-j)}$$

$$= \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j) \overline{f(n-m) f(n-m+k-j)}$$

$$\varphi_{gg}(k) == \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j)\varphi_{ff}(k-j) \quad (9.5.1)$$

Similarly $\varphi_{fg}(k)$ is obtained as follows:

$$\varphi_{fg}(k) = \overline{f(n)g(n+k)} = \overline{f(n) \sum_{j=-\infty}^{+\infty} h(j) f(n+k-j)}$$

$$= \sum_{j=-\infty}^{+\infty} h(j) \overline{f(n) f(n+k-j)}$$

$$\varphi_{fg}(k) = \sum_{j=-\infty}^{+\infty} h(j)\varphi_{ff}(k-j) \quad (9.5.2)$$

The mean-square error of our unit-time predictor in terms of the system impulse response and input autocorrelation function would be

$$\overline{e^2} = \varphi_{ff}(0) - 2\sum_{j=-\infty}^{+\infty} h(j)\varphi_{ff}(1+j)$$

$$+ \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j)\varphi_{ff}(j)$$

(We have used the symmetry of φ_{ff} is symmetric.) For given $\varphi_{ff}(k)$ we select values for h(n) which minimize $\overline{e^2}$. To make the problem interesting we demand physical realizability and set h(n)=0 for n<0.

Let's try to build a realizable predictor for the penny-pitching signal of section 9.3. We know this will be impossible since this f(n) signal had no time dependence of any kind among samples on which to base a prediction. In $\overline{e^2}$ above we set,

$$h(n) = 0, \quad n < 0 \qquad \varphi_{ff}(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

and obtain

$$\overline{e^2} = 1 + \sum_{m=0}^{+\infty} h^2(m)$$

The minimum error is achieved by taking h(n) identically zero and always predicting, therefore, zero values. The minimum error is the same as the mean-square value of the input itself and our prediction attempt has been futile. In contrast, one can find examples where a near perfect prediction can be made.

We will have occasion to compute the average of g(n) in terms of the average of f(n) and the impulse response. Assuming stationary f(n) and g(n) this relation is:

$$\overline{g} = \overline{\sum_{m=-\infty}^{+\infty} h(m) f(n-m)}$$

$$= \sum_{m=-\infty}^{+\infty} h(m) \overline{f(n-m)} = \overline{f} \cdot \sum_{m=-\infty}^{+\infty} h(m)$$

This shows that if the input has zero mean $(\overline{f} = 0)$ then the output will also have zero mean $(\overline{g} = 0)$.

Note specifically that since $\varphi_{gg}(0)$ is the mean-square value, $\overline{g^2}$, we have from the above formula for $\varphi_{gg}(k)$:

$$\overline{g^2} = \varphi_{gg}(0) = \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j)\varphi_{ff}(-j)$$

$$= \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j)\varphi_{ff}(j)$$

9.6 Generation of random inputs

One way to arrive initially at the random input signals and their autocorrelation functions for use in systems problems is to generate these signals by passing the simple penny-pitching signal through some time-invariant linear system with a given impulse response. The output of this system is then used as the actual f(n) input to the main system under study.

However, let's concentrate on the preconditioning system and apply our simple signal with autocorrelation function $\varphi_{ff}(k)=\delta(k)$.

Its output correlation function φ_{gg} is given by the formula of the preceding section:

$$\varphi_{gg}(k) = \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m)h(m+j)\delta(k-j)$$

$$= \sum_{m=-\infty}^{+\infty} h(m)h(m+k)$$

For example suppose

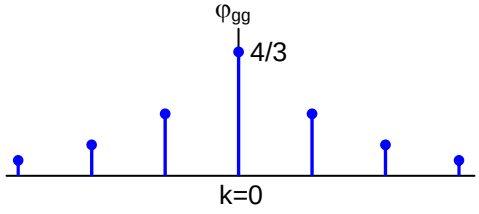
$$h(n) = \begin{cases} 0 & n < 0 \\ (1/2)^n & n \geq 0 \end{cases}$$

then for k≥0

$$\varphi_{gg}(k) = \sum_{m=0}^{+\infty} (\tfrac{1}{2})^m (\tfrac{1}{2})^{m+k} = (\tfrac{1}{2})^k \sum_{m=0}^{+\infty} (\tfrac{1}{4})^m = \tfrac{4}{3}(\tfrac{1}{2})^k$$

For k<0, φ_{gg} is determined by symmetry so,

$$\varphi_{gg}(k) = \tfrac{4}{3}(\tfrac{1}{2})^{|k|}$$



Now g(n) from this process can be used as the input to another system.

Note that since h(n)→0 geometrically as n→±∞ we have,

$$\varphi_{gg}(k) = \sum_{m=-\infty}^{+\infty} h(m)h(m+k)$$

$$\varphi_{gg}(k) \to 0 \quad as \quad k \to \pm\infty$$

Therefore, the output of the preconditioning filter gives the asymptotic property of the correlation function mentioned in section 9.4. In particular, the mean value of the output is zero.

9.7 Review of assumptions

Input, f(n):

1. f(n) is stationary
2. $\overline{f}$ and $\overline{f^2}$ exist and are finite
3. There exists some constant c such that f(n)+c=s(n) has zero mean and $\varphi_{ss}(k)\to 0$ as $k\to\pm\infty$.

System impulse response:

4. The transfer function is a rational function of z.
5. h(n)→0 as n→±∞.

9.8 Problems

9.8.1) A random input function f(n) is generated as follows. Let

$$f(n) = (-1)^{n+\alpha}$$

where α is to be selected either 0 or 1, each with probability 1/2. That is, there are two member functions of the ensemble, one with α=0 and one with α=1. The entire function is selected when α is chosen.

Is the ensemble stationary? Find the autocorrelation function $\varphi_{ff}(k)$. Show that assumption 3 of section 9.7 does not hold. (Therefore, we would not use this ensemble in our system studies.)

Suppose α=0 or 1 with probabilities 1/3 and 2/3, respectively. Is the resulting ensemble stationary?

9.8.2) Show that any stationary f(n) consisting of independent samples with $\overline{f}=0$ and $\overline{f^2}=1$, has the same autocorrelation function as the penny-pitching signal of section 9.3.

$$\varphi_{ff}(k) = \delta(k)$$

Show that any stationary f(n) consisting of uncorrelated samples, (i. e. , $\overline{f(n)f(n+k)} = \overline{f(n)}$ when k≠0) with $\overline{f}=0$, $\overline{f^2}=1$ also leads to $\varphi_{ff}(k)=\delta(k)$. Thus the stronger assumption of independence is not required.

9.8.3) Uncorrelated samples with $\varphi_{ff}(k)=\delta(k)$ are passed through a filter with the impulse response

$$h(n) = \begin{cases} 1 & n = 0,1,2 \\ 0 & otherwise \end{cases}$$

Find the output mean, $\overline{g}$, mean-square, $\overline{g^2}$, and the correlation functions $\varphi_{gg}(k)$ and $\varphi_{fg}(k)$.

9.8.4) The input, f(n), to a system has the correlation function,

$$\varphi_{ff}(n) = (-\tfrac{1}{2})^{|k|}$$

The system impulse response is,

$$h(n) = \begin{cases} 1 & n = 0 \\ \frac{1}{2} & n = 1 \\ 0 & otherwise \end{cases}$$

What is the output autocorrelation, $\varphi_{gg}(k)$?

10. Correlations Transforms

10.1 Definition

Just as we defined the transform, F(z), of a time function, f(n), we can define the transform, Φ(z), of an autocorrelation or cross-correlation function φ(k). We define Φ(z) by the power series:

$$\Phi(z) = \sum_{k=-\infty}^{+\infty} \varphi(k) z^k$$

This series will converge for z with $|z| \approx 1$ provided $\varphi(k) \to 0$ fast enough as $k \to \pm\infty$. We will find that in most practical problems the sum does exist. In fact, φ(k) will ordinarily tend to zero geometrically as $k \to \pm\infty$ (see the example in section 9.5). In that case the series converges for z with $|z| \approx 1$ and the region of convergence is an annulus containing the perimeter of the unit circle in the z-plane. When Φ(z) is continued to the entire z-plane and expressed in closed form, φ(k) can be recovered by expanding Φ(z) in a power series convergent in this annulus.

In the example of section 9.6 we found an autocorrelation function

$$\varphi(k) = \tfrac{4}{3}\left(\tfrac{1}{2}\right)^{|k|}$$

Therefore Φ(z) is

$$\Phi(z) = \tfrac{4}{3}\left(1 + \tfrac{1}{2}z + \tfrac{1}{4}z^2 + \cdots + \tfrac{1}{2}z^{-1} + \tfrac{1}{4}z^{-2} + \cdots\right)$$

$$= \tfrac{4}{3}\frac{1}{1 - \frac{1}{2}z} + \tfrac{4}{3}\frac{\frac{1}{2}z^{-1}}{1 - \frac{1}{2}z^{-1}} = \frac{1}{(1 - \frac{1}{2}z)(1 - \frac{1}{2}z^{-1})}$$

(There is generally no need to explicitly replace z^{-1} by 1/z and then to clear fractions.)

We succeed in defining Φ(z) in the above example because $\varphi(k) \to 0$ geometrically as $k \to \pm\infty$. Also note that Φ(z) is a rational function of z. We will show that these properties are true for the correlation functions of signals generated by passing independent (or uncorrelated) samples of zero mean through linear, time-invariant systems whose transfer functions are a rational function of z and whose impulse responses tend to zero as $n \to \pm\infty$.

10.2 System relations with correlation transforms

Transforms of correlation functions provide a superior way of relating correlation functions of signals at different points of a system. The motive for introducing them is the same as for the ordinary signal transform (where the simple response relation G=FH replaces the more complicated superposition summation.)

We will now derive the simpler transform domain analogs of some of the formulas in sections 9.3 and 9.4. Each derivation results by transforming both sides of the formulas, by multiplying by z^k and by summing over k. We make the assumptions of section 9.7 with regard to the impulse response, h(n).

If f(n)=r(n)+s(n), the sum of two independent (or uncorrelated) signals of zero mean, then

$$\varphi_{ff}(k) = \varphi_{rr}(k) + \varphi_{ss}(k)$$

$$\sum \varphi_{ff}(k) z^k = \sum \varphi_{rr}(k) z^k + \sum \varphi_{ss}(k) z^k$$

or

$$\Phi_{ff}(z) = \Phi_{rr}(z) + \Phi_{ss}(z)$$

If f(n), g(n), and h(n) are the input, output, and impulse response of a system, then using equation 9.5.1 we get:

$$\varphi_{gg}(k) = \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m) h(m+j) \varphi_{ff}(k-j)$$

$$\sum_{k=-\infty}^{+\infty} \varphi_{gg}(k) z^k = \sum_{k=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(m) z^{-m} h(m+j) z^{m+j} \varphi_{ff}(k-j) z^{k-j}$$

$$= \sum_{m=-\infty}^{+\infty} h(m) z^{-m} \cdot \sum_{j=-\infty}^{+\infty} h(j) z^j \cdot \sum_{k=-\infty}^{+\infty} \varphi_{ff}(k) z^k$$

or

$$\Phi_{gg}(z) = H(z^{-1}) H(z)\, \Phi_{ff}(z)$$

Similarly the crosscorrelation transform of f and g is obtained from equation 9.5.2:

$$\varphi_{fg}(k) = \sum_{j=-\infty}^{+\infty} h(j)\varphi_{ff}(k-j)$$

$$\sum_{k=-\infty}^{+\infty} \varphi_{fg}(k)z^k = \sum_{k=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(j)z^j\varphi_{ff}(k-j)z^{k-j}$$

$$= \sum_{j=-\infty}^{+\infty} h(j)z^j \cdot \sum_{k=-\infty}^{+\infty}\varphi_{ff}(k)z^k$$

or

$$\Phi_{fg}(z) = H(z)\,\Phi_{ff}(z)$$

In each of the above transform formulas the left sides exist if the right sides are defined. So if $\varphi_{ff}(k)$ can be transformed to $\Phi_{ff}(z)$ then $\varphi_{gg}(k)$ and $\varphi_{fg}(k)$ can be transformed to $\Phi_{gg}(z)$ and $\Phi_{fg}(z)$. Suppose g(n) is generated by passing an f(n) through a system with transfer function H(z). Then $\varphi_{ff}(k)=\delta(k)$ can be transformed to

$$\Phi_{ff}(z) = 1,$$

$\varphi_{gg}(k)$ can be transformed and

$$\Phi_{gg}(z) = H(z)H(z^{-1})$$

Since H(z), by assumption, is rational, so is $\Phi_{gg}(z)$. Also as $\varphi_{gg}(k)\to 0$ as $k\to\pm\infty$ or else it could not have been transformed; and the rate of decay is geometrical since $\Phi_{gg}(z)$ is rational.

Now that a g(n) has been generated with a $\varphi_{gg}(k)$ having the desired properties, g(n) can be used as the input to other systems, and the outputs of these systems will likewise have the properties of existence and rationality of $\Phi_{ff}(z)$ and $\Phi_{gg}(z)$. We note also that the same arguments apply to $\Phi_{fg}(z)$.

As an example of the above remarks consider the example of 9.6. The given quantities were $\varphi_{ff}(k)=\delta(k)$ and

$$h(n) = \begin{cases} 0 & n < 0 \\ (1/2)^n & n \ge 0 \end{cases}$$

and $\varphi_{gg}(k)$ was required. We now solve the problem by going to the transform domain.

$$\Phi_{ff}(z) = 1$$

$$H(z) = \frac{1}{1-\frac{1}{2}z}$$

Therefore

$$\Phi_{gg}(z) = H(z)H(z^{-1})\,\Phi_{ff}(z)$$

$$= \frac{1}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})} = \frac{\frac{4}{3}}{1-\frac{1}{2}z} + \frac{\frac{2}{3}z^{-1}}{1-\frac{1}{2}z^{-1}}$$

and

$$\varphi_{gg}(k) = \tfrac{4}{3}(\tfrac{1}{2})^{|k|}$$

which checks the previous result.

We can also find $\varphi_{fg}(k)$:

$$\Phi_{fg}(z) = H(z)\,\Phi_{ff}(z) = \frac{1}{1-\frac{1}{2}z}$$

$$\varphi_{fg}(k) = \begin{cases} 0 & k < 0 \\ (1/2)^k & k \ge 0 \end{cases}$$

10.3 Conditions on autocorrelation functions

If a stationary random time function having a given autocorrelation function exists, it can be generated by passing uncorrelated samples of unit variance through an appropriate linear, time-invariant system. (As before, we assume that the time functions have zero mean.) This fact is made plausible by the previous discussion, and we won't prove it rigorously.

Therefore, if $\varphi_{ff}(k)$ is a true autocorrelation function, its transform $\Phi_{ff}(z)$ must have the form

$$\Phi_{ff}(z) = H(z)H(z^{-1})$$

where H(z) is the transfer function of some system. It follows that if we set $z=e^{iw}$, then,

$$\Phi_{ff}(e^{i\omega}) = H(e^{i\omega})H(e^{-i\omega}) = \left|H(e^{i\omega})\right|^2 \geq 0$$

Thus, for all w, $\Phi_{ff}(e^{iw})$ must be real and non negative.

The condition includes the preceding ones mentioned for autocorrelation functions. Use of this condition prevents assuming impossible autocorrelation functions.

For example, consider the correlation function

$$\varphi_{fg}(k) = \begin{cases} 1 & k < 0 \\ A & k = \pm 1 \\ 0 & otherwise \end{cases}$$

Then

$$\Phi_{ff}(z) = 1 + Az + Az^{-1}$$

$$\Phi_{ff}(e^{iw}) = 1 + Ae^{iw} + Ae^{-iw} = 1 + 2A\cos(\omega)$$

Evidently if $\Phi_{ff}(e^{iw})$ is to be non-negative for all w, we must have

$$\left|A\right| \leq \tfrac{1}{2}$$

An autocorrelation with A=1, for example, would be impossible.

10.4 Conversion to uncorrelated samples

In section 9.6 we showed how we could generate complicated random time functions by passing uncorrelated samples through linear systems. Now we reverse this process and convert random time functions to uncorrelated samples through the agency of some other system. This is the key tool of the Bode-Shannon method for deriving optimum filters.

The problem of conversion to uncorrelated samples is particularly simple under the assumptions we have made that $\Phi_{ff}(z)$ must be a rational function of z. By writing numerator and denominator polynomials in factored form, we separate all factors into two groups, one group with poles and zeros outside the unit circle and the other with poles and zeros inside the unit circle.

$$\Phi_{ff}(z) = \frac{P_1(z)}{Q_1(z)} \cdot \frac{P_2(z)}{Q_2(z)}$$

Since $\Phi_{ff}(z)=\Phi_{ff}(z^{-1})$, (symmetry of φ_{ff}), we can make

$$\frac{P_2(z)}{Q_2(z)} = \frac{P_1(z^{-1})}{Q_1(z^{-1})}$$

Now P_1/Q_1 can be considered the transfer function, H(z), of a system that produces f(n) from uncorrelated samples.

$$\Phi_{ff}(z) = H(z)H(z^{-1})$$

$Q_1/P_1 = 1/H(z)$ is also a legitimate transfer function, and if f(n) is passed through a system with this transfer function, the result, g(n) has the correlation

$$\Phi_{gg}(z) = \Phi_{ff}(z) \cdot \frac{1}{H(z)} \cdot \frac{1}{H(z^{-1})} = 1$$

That is, g(n) consists of uncorrelated samples.

Note that we have made the above separation of factors in such a way that 1/H(z) is a physically realizable transfer function. The process we have described of separation of factors is called spectrum factorization.

An exception occurs when $\Phi_{ff}(z)$ has zeros on the unit circle. In this case, no system can strictly be found to convert f(n) to uncorrelated samples. However, such cases may often be successfully treated in practical problems by perturbing the parameters slightly so as to

move the zeros off the circle and then to consider zeros on the circle as a limiting case. The physical significance of this manipulation is usually apparent.

As an example of conversion to uncorrelated samples consider the f(n) previously used with $\varphi_{ff}(k) = \frac{4}{3}(\frac{1}{2})^k$.

$$\Phi_{ff}(z) = \frac{1}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

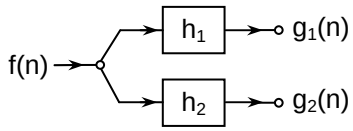
Factoring gives us $P_1=P_2=1$, $Q_1=1-\text{▱}z$, $Q_2=1-\text{▱}z^{-1}$. So the transfer function of the system we seek is:

$$\frac{Q_1}{P_1} = 1-\tfrac{1}{2}z$$

Passing f(n) through this system, in fact, yields g(n) equal to the original uncorrelated samples used to generate f(n) in section 9.6, for the transfer function of the system used there was $1/(1-\text{▱}z)$. In general this does not happen and the g(n), while uncorrelated, may not be identical the initial samples. In fact, spectrum factorization can be performed in more than one way as problem 10.5.3 shows (even if physical realizability of the resulting 1/H(z) is required).

10.5 Problems

10.5.1) An input f(n) with autocorrelation function $\varphi_{ff}(k)$ is applied simultaneously to two separate systems.



The impulse responses of the systems are $h_1(n)$ and $h_2(n)$ and the outputs are $g_1(n)$ and $g_2(n)$. Find $\varphi_{g1g2}(k)$ in terms of $h_1(n)$, $h_2(n)$, and $\varphi_{ff}(k)$ and convert this relation to the transform domain.

10.5.2) Solve problems 3 and 4 of chapter 9 by using transform methods.

10.5.3) If H(z) is the transfer function of a system which converts some f(n) into uncorrelated samples show that a system with transfer function $H_1(z)=z^kH(z)$ has the same property (k is an integer). Show that

$$H_2(z) = \frac{1-\frac{1}{2}z^{-1}}{1-\frac{1}{2}z}$$

H(z) also has the property.

Using the results of problem 1 find the crosscorrelation function between the uncorrelated outputs produced by H(z) and $H_2(z)$ above.

11. Optimum linear filters

11.1 Wiener-Hopf equation

In this chapter we will devise linear, time-invariant systems that operate on stationary random inputs to produce some desired output. However, such a desired output can't always be achieved exactly. It's the task of the system designer to find a linear system whose output is as close as possible to the desired output. To measure this closeness we use the mean-square error - the mean of the square of the difference between the actual and the desired output. As we have seen in chapter 9, this error measure will lead to the use of correlation functions.

Denote by f(n), h(n), and g(n) the input, impulse response, and output of a linear system. Let d(n) be the desired output. We assume that all signals are stationary and random. The system error is g(n)–d(n), the mean-square error is

$$\overline{e^2} = \overline{[g(n)-d(n)]^2}$$

We can express g(n) in terms of f(n) and h(n) by the superposition summation:

$$\overline{e^2} = \overline{\left[\sum_{m=-\infty}^{+\infty} h(m)\, f(n-m) - d(n)\right]^2}$$

Multiplying out the square, interchanging the averaging with the summations, and interpreting the averages as correlation functions we obtain:

$$\overline{e^2} = \overline{d^2(n)} - 2\sum_{m=-\infty}^{+\infty} \overline{h(m)\, f(n-m) d(n)} + \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} \overline{h(m)h(j)\, f(n-m)\, f(n-j)}$$

$$= \varphi_{dd}(0) - 2\sum_{m=-\infty}^{+\infty} h(m)\,\varphi_{fd}(m) + \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(m)\,h(j)\,\varphi_{ff}(m-j)$$

If we assume that the statistical characteristics of f(n) and d(n) are given then $\varphi_{dd}(0)$, $\varphi_{fd}(m)$, and $\varphi_{ff}(m–j)$ are known, and the formula becomes an expression for $\overline{e^2}$ in terms of the values of the impulse response samples h(n). To find the linear system than minimizes the error $\overline{e^2}$, we differentiate the $\overline{e^2}$ expression with respect to h(0), h(l), h(–1), h(2), etc., and so obtain a set of simultaneous equations which determine optimum h(n) values by setting these derivatives to zero.

$$\frac{\partial \overline{e^2}}{\partial h(n)} = -2\varphi_{fd}(n) + 2\sum_{j=-\infty}^{+\infty} h(j)\,\varphi_{ff}(n-j)$$

$$= 0, \text{ for optimum } h(n)$$

Or denoting the optimum h(n) by $h_o(n)$

Equation A

$$\sum_{j=-\infty}^{+\infty} h_o(j)\,\varphi_{ff}(n-j) = \varphi_{fd}(n) \qquad (n = 0, \pm 1, \pm 2, \cdots)$$

Often we want to restrict our optimum systems to be physically realizable. In such cases we set $h_o(j)=0$, $j<0$ and only minimize $\overline{e^2}$ with respect to the remaining $h_o(j)$. We then obtain:

Equation B

$$\begin{cases} \sum_{j=-\infty}^{+\infty} h_o(j)\,\varphi_{ff}(n-j) = \varphi_{fd}(n) & (n = 0,1,2,\cdots) \\ h_o(n) = 0 & (n = -1,-2,\cdots) \end{cases}$$

(Equation B in its integral form for continuous data systems is known as the Wiener-Hopf equation.)

To prove that these equations provide an absolute minimum for $\overline{e^2}$ and not some other type of stationary point we take the expression for $\overline{e^2}$ and substitute h(n) values equal to $h_o(n)$ perturbed by an amount ε(n): $h(n)=h_o(n)+\varepsilon(n)$. (If physical realizability is imposed, we set $h_o(n)=\varepsilon(n)=0$; $n<0$.)

$$\overline{e^2} = \varphi_{dd}(0) - 2\sum_{m=-\infty}^{+\infty} h_o(m)\varphi_{fd}(m) + \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h_o(m)\,h_o(j)\,\varphi_{ff}(m-j)$$
$$-2\sum_{m=-\infty}^{+\infty}\varepsilon(m)\,\varphi_{fd}(m) + 2\sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h_o(m)\,\varepsilon(j)\,\varphi_{ff}(m-j)$$
$$+\sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty}\varepsilon(m)\,\varepsilon(j)\,\varphi_{ff}(m-j)$$

Using equations A (or B if physical realizability is imposed) we obtain:

$$\overline{e^2} = \varphi_{dd}(0) - \sum_{m=-\infty}^{+\infty} h_o(m)\,\varphi_{fd}(m) + \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} \varepsilon(m)\,\varepsilon(j)\,\varphi_{ff}(m-j)$$

The third term on the right may be interpreted as follows:

$$\sum_m\sum_j \varepsilon(m)\,\varepsilon(j)\,\varphi_{ff}(m-j) = \sum_m\sum_j \varepsilon(m)\,\varepsilon(j)\,\overline{f(n+j)\,f(n+m)}$$

$$= \overline{\left[\sum_m \varepsilon(m) f(n+m)\right]\left[\sum_j \varepsilon(j) f(n+j)\right]} = \overline{\left[\sum_m \varepsilon(m) f(n+m)\right]^2}$$

But the average of a square is always non-negative, and so the third term of the $\overline{e^2}$ expression can only increase $\overline{e^2}$ if the ε(n) are not zero. Therefore, equations A or B do provide a minimum, and the minimum error, $\overline{e_o^2}$, is:

Equation C

$$\overline{e_o^2} = \varphi_{dd}(0) - \sum_{m=-\infty}^{+\infty} h_o(m)\varphi_{fd}(m)$$

11.2 A prediction example

Suppose the input f(n) of a system has the correlation function $\varphi_{ff}(k) = \frac{4}{3}(\frac{1}{2})^{|k|}$. We wish to find the optimum, physically realizable, linear system that will predict f(n) p units in advance. The desired output therefore is:

$$d(n) = f(n+p)$$

Other needed quantities are:

$$\varphi_{dd}(0) = \overline{d(n)d(n)} = \overline{f(n+p)\,f(n+p)} = \varphi_{ff}(0) = \frac{4}{3}$$

$$\varphi_{fd}(k) = \overline{f(n)d(n+k)} = \overline{f(n)f(n+k+p)} = \varphi_{ff}(k+p) = \frac{4}{3}\left(\frac{1}{2}\right)^{|k+p|}$$

Applying equation B we obtain:

$$\begin{cases} h_o(j) = 0 & n < 0 \\ \sum_{j=0}^{+\infty} h_o(j)\cdot\frac{4}{3}(\frac{1}{2})^{|j-n|} = \frac{4}{3}(\frac{1}{2})^{|n+p|} & n \ge 0 \end{cases}$$

Writing out the equations for n≥0 (and cancelling the 4/3) we obtain the set of equations:

$$\begin{aligned} h_o(0) + \tfrac{1}{2}h_o(1) + \tfrac{1}{4}h_o(2) + \cdots &= (\tfrac{1}{2})^p \\ \tfrac{1}{2}h_o(0) + h_o(1) + \tfrac{1}{2}h_o(2) + \cdots &= \tfrac{1}{2}(\tfrac{1}{2})^p \\ \tfrac{1}{4}h_o(0) + \tfrac{1}{2}h_o(1) + h_o(2) + \cdots &= \tfrac{1}{4}(\tfrac{1}{2})^p \\ &\cdots \end{aligned}$$

Multiplying the second equation by 1/2 and subtracting from the first; multiplying the third by 1/2 and subtracting from the second and so on we obtain the set

$$\begin{aligned} \tfrac{3}{4}h_o(0) &= \tfrac{3}{4}(\tfrac{1}{2})^p \\ \tfrac{3}{8}h_o(0) + \tfrac{3}{4}h_o(1) &= \tfrac{3}{8}(\tfrac{1}{2})^p \\ \tfrac{3}{16}h_o(0) + \tfrac{3}{8}h_o(1) + \tfrac{3}{4}h_o(2) &= \tfrac{3}{16}(\tfrac{1}{2})^p \\ &\cdots \end{aligned}$$

The solution is seen to be:

$$h_o(0) = (\tfrac{1}{2})^p, \qquad h_o(n) = 0 \ \ otherwise$$

So the optimum predictor is a filter with a simple gain of $(1/2)^p$. If p=0 this result is obviously correct. If p is large the result is also logical, for then the prediction cannot reliably be made, and the system just predicts just zero.

The minimum error of the predictor, from equation C, is:

$$\overline{e_o^2} = \tfrac{4}{3} - (\tfrac{1}{2})^p \cdot \tfrac{4}{3}(\tfrac{1}{2})^p = \tfrac{4}{3}\left[1 - (\tfrac{1}{4})^p\right]$$

When p=0, the error is zero. As p→∞ the error increases to 4/3, the mean-square value of the input itself.

The system we have derived is the optimum linear predictor. It may be possible to improve the prediction by operating non-linearly on the input. Such operations take advantage of statistical characteristics other than correlations to improve predictions. If, however the input samples are governed completely by joint normal distributions, then the optimum linear system is also the absolute optimum system. (We won't prove this fact.) Such input signals are said to be Gaussian.

11.3 Transform domain optimization

If there are no physical realizability restrictions on the optimum system, the optimization problem may be solved easily in the transform domain.

Using equation A, we multiply both sides by z^n and then sum the equations for all n.

$$\sum_{n=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} z^n h_o(j)\,\varphi_{ff}(n-j) = \sum_{n=-\infty}^{+\infty} z^n \varphi_{fd}(n)$$

The right side is just $\Phi_{fd}(z)$, and the left can be rearranged to give:

$$\sum_{j=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} h_o(j)\,z^j \cdot \varphi_{ff}(n-j) z^{n-j} = H_o(z)\Phi_{ff}(z)$$

So, the transfer function of the optimum system is.

Equation D

$$H_o(z) = \frac{\Phi_{fd}(z)}{\Phi_{ff}(z)}$$

As an example of the use of equation D we will find the optimum linear filter that will separate a desired message signal from additive noise. That is, the signal f(n) is composed of the sum of two signals d(n) and r(n).

$$f(n) = d(n) + r(n)$$

r(n) is noise with an autocorrelation function $\varphi_{rr}(k) = \frac{2}{3}\delta(k)$ and d(n) is the desired message signal with $\varphi_{dd}(k) = \frac{10}{27}\left(\frac{1}{2}\right)^{|k|}$. We assume that message and noise are statistically independent (or uncorrelated).

By a result of chapter 10:

$$\Phi_{ff}(z) = \Phi_{dd}(z) + \Phi_{rr}(z)$$

$$= \frac{\frac{5}{18}}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})} + \frac{2}{3} = \frac{\frac{10}{9} - \frac{1}{3}(z+z^{-1})}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

$$= \frac{(1-\frac{1}{3}z)(1-\frac{1}{3}z^{-1})}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

Also we obtain (Note: $\bar{r} = \bar{d} = 0$),

$$\varphi_{fd}(k) = \overline{f(n)d(n+k)}$$

$$= \overline{d(n)d(n+k)} + \overline{r(n)}\cdot\overline{d(n+k)} = \varphi_{dd}(k)$$

or

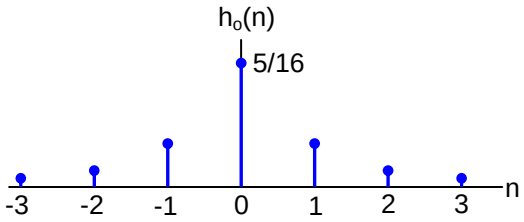
$$\Phi_{fd}(z) = \Phi_{dd}(z) = \frac{\frac{5}{18}}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

Formula D gives us the optimum transfer function:

$$H_o = \frac{\Phi_{fd}(z)}{\Phi_{ff}(z)} = \frac{\frac{5}{18}}{(1-\frac{1}{3}z)(1-\frac{1}{3}z^{-1})}$$

$$= \frac{\frac{5}{16}}{(1-\frac{1}{3}z)} + \frac{1}{3}\frac{\frac{5}{16}z^{-1}}{(1-\frac{1}{3}z^{-1})}$$

The optimum impulse response is $h_o(n) = \frac{5}{16}\left(\frac{1}{3}\right)^{|n|}$



This response is evidently not physically realizable. However, all samples before n=–4, say, are small and neglecting them does not change the behavior of the filter by much. If we delete them and shift the n=0 origin four units to the left we have a new impulse response which has nearly the optimum properties of the original except that the output is delivered after an extra four units of time. Thus the output now is an estimate of the input four units of time ago. This filter is realizable.

So we can use Equation D even if we want physical realizability provided that we are in no hurry to obtain outputs corresponding to current inputs. This equation is sometimes said to represent the infinite lag solution to the optimum filter problem.

In the above filter problem, the error $\overline{e_o^2}$ is given again by equation C.

$$\overline{e_o^{\,2}} = \tfrac{10}{27} - \sum_{m=-\infty}^{+\infty} \tfrac{5}{16}\left(\tfrac{1}{3}\right)^{|m|} \cdot \tfrac{10}{27}\left(\tfrac{1}{2}\right)^{|m|}$$

$$= \tfrac{10}{27} - \sum_{m=-\infty}^{+\infty} \tfrac{25}{216}\left(\tfrac{1}{6}\right)^{|m|} = \tfrac{5}{24}$$

This error is smaller than the 10/27 we would have obtained by refusing to transmit anything through the filter or the 2/3, which is the mean-square value of the original noise errors.

11.4 Problems

11.4.1a) Show that the impulse response of the optimum, physically realizable system which predicts a signal, f(n), one unit of time in advance where

$$\Phi_{ff}(z) = \tfrac{5}{4} - \tfrac{1}{2}z - \tfrac{1}{2}z^{-1}$$

is

$$h_o(n) = \begin{cases} 0 & n < 0 \\ -(1/2)^{n+1} & n \geq 0 \end{cases}$$

b.) Find the minimum mean-square prediction error. How could a signal having the correlation properties of f(n) be simulated using a table of random numbers?

11.4.2) Noise, r(n),with an autocorrelation function

$$\varphi_{rr}(k) = \tfrac{5}{12}\delta(k)$$

is added to the message signal of problem 11.4.1. The noise and message are independent. Find the optimum linear filter (no realizability restrictions) that attempts to reproduce the message and exclude the noise (a smoothing filter).

12. Bode-Shannon optimum realizable filters

12.1 Optimum realizable filters

Our study of optimum filter problems is not yet complete since we have not indicated any practical way to solve equation B of section 11.1, the equations determining $h_o(n)$ under physical realizability conditions. While equations A can be handled in the transform domain as shown in section 11.3, equation B can't be transformed directly. This chapter gives a method of handling equation B in the transform domain and provides a suitable method for their solution when the problem is more complicated than the example of section 11.2.

The first step in the Bode-Shannon procedure for design of optimum realizable filters is to pass the input f(n) through a system which converts it to an output, f'(n), of uncorrelated samples. This system is the one found in section 10.4 by spectrum factorization. Its transfer function, H'(z), is

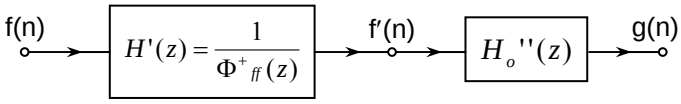
$$H'(z) = \frac{1}{\Phi^{+}_{ff}(z)}$$

where $\Phi^{+}_{ff}(z)$ represents the combination of all numerator and denominator factors of $\Phi_{ff}(z)$ that represent zeros or poles outside the unit circle. The remaining factors will be represented by $\Phi^{-}_{ff}(z)$. Thus

$$\Phi_{ff}(z) = \Phi^{+}_{ff}(z)\Phi^{-}_{ff}(z)$$

H'(z) has the important property that both H'(z) and 1/H'(z) are realizable transfer functions. This means that from the point of view of physical realizability nothing irretrievable has been done to f(n). So, one should be able to approximate the desired d(n) by a realizable filter operating on f'(n) to the same degree as by such a filter operating on f(n) since, if necessary, f(n) can be recovered from f'(n) by a realizable filter.

We have separated the problem of operating on f(n) to produce g(n), an approximation to d(n), into two parts. First, we find a realizable filter to convert f(n) to uncorrelated samples f'(n). Then we follow this filter by another realizable filter that produces an optimum approximation to d(n). Call the transfer function of this second filter $H_o''(z)$.



We can now use the theory of chapter 11 to find $H_o''(z)$ by replacing f(n) with f'(n) as the input to the optimum filter. However, the equations now simplify since

$$h_o''(n)\varphi_{f'f'}(k) = 0 \quad for\ k \neq 0$$

So equation A becomes:

Equation A′

$$h_o''(n)\varphi_{f'f'}(0) = \varphi_{f'd}(n) \qquad (n = 0, \pm 1, \pm 2, \cdots)$$

And equation B becomes:

Equation B′

$$\begin{cases} h_o''(n)\varphi_{f'f'}(0) = \varphi_{f'd}(n) & (n = 0,1,2,\cdots) \\ h_o''(n) = 0 & (n = -1,-2,\cdots) \end{cases}$$

Note that the only difference between the h''(n) values solving A' and B' is that in the realizable case of equation B' the h_o'' values for n<0 are set to zero. Otherwise the optimal solutions are identical, being

$$h_o''(n) = \varphi_{f'd}(n)/\varphi_{f'f'}(0) \qquad for\ n \geq 0$$

Hence, if we know the solution with no realizability restriction, we can immediately find the solution under the realizability restriction by setting $h_o''(n)=0$ for n<0.

Now $H_o''(z)$ with no realizability restriction is easy to obtain since, as we proved in section 11.3, the over-all optimum filter converting f(n) to g(n) should have the transfer function $\Phi_{fd}(z)/\Phi_{ff}(z)$. Since $H'(z)=1/\Phi^{+}_{ff}(z)$, $H_o''(z)$ would be given by:

$$\frac{\Phi_{fd}(z)}{\Phi_{ff}(z)} \div \frac{1}{\Phi^{+}_{ff}(z)} = \frac{\Phi_{fd}(z)}{\Phi^{-}_{ff}(z)}$$

Therefore H_o'' under a realizability restriction, as we have remarked above, may be obtained by converting $\Phi_{fd}(z)/\Phi^{-}_{ff}(z)$ to the time domain, deleting samples for n<0, and returning to the transform domain. The new transform we denote by $\left[\Phi_{fd}(z)/\Phi^{-}_{ff}(z)\right]_{+}$. Thus:

$$H_o''(z) = \left[\frac{\Phi_{fd}(z)}{\Phi^-_{ff}(z)}\right]_+$$

and the transfer function of the over-all optimum realizable filter converting f(n) to an approximation of d(n) is:

Equation E

$$H_o(z) = \frac{1}{\Phi^+_{ff}(z)}\left[\frac{\Phi_{fd}(z)}{\Phi^-_{ff}(z)}\right]_+$$

This is the transform domain solution of equation B.

12.2 Prediction example

We now repeat the prediction example given in section 11.2 using the transform method developed in the previous section. f(n) has the correlation function $\varphi_{ff}(k) = \frac{4}{3}(\frac{1}{2})^{|k|}$ and it was shown in 11.2 that if the desired output d(n) is a prediction of f(n) p units in advance, d(n)=f(n+p), then $\varphi_{fd}(k)=\varphi_{ff}(k+p)$. So we find,

$$\Phi_{ff}(z) = \frac{1}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

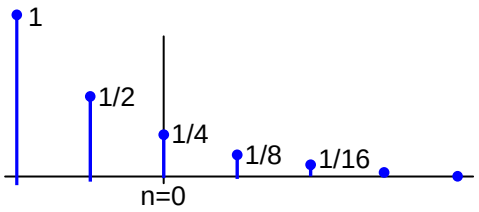
$$\Phi_{fd}(z) = \sum_{k=-\infty}^{+\infty}\varphi_{ff}(k+p)z^k = z^{-p}\Phi_{ff}(z) = \frac{z^{-p}}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

Factoring $\Phi_{ff}(z)$, we obtain

$$\Phi^+_{ff}(z) = \frac{1}{1-\frac{1}{2}z} \qquad \Phi^-_{ff}(z) = \frac{1}{1-\frac{1}{2}z^{-1}}$$

$$\frac{\Phi_{fd}(z)}{\Phi^-_{ff}(z)} = \frac{z^{-p}}{1-\frac{1}{2}z} = z^{-p} + \frac{1}{2}z^{-p+1} + \cdots + \frac{1}{2^p} + \frac{1}{2^{p+1}}z + \cdots$$

In the time domain $\frac{z^{-p}}{1-\frac{1}{2}z}$ represents a sequence of samples which begins at n=–p with an amplitude of 1 and decays by the factor 1/2 for every unit increase of n. For p=2 the picture is



We now throw away all samples before n=0 and return to the transform domain to find

$$\left[\frac{\Phi_{fd}(z)}{\Phi^-_{ff}(z)}\right]_+ = \frac{1}{2^p}\cdot\frac{1}{1-\frac{1}{2}z}$$

So by equation E of the last section the optimum realizable predictor is given by the transfer function

$$H_o'(z) = (1-\tfrac{1}{2}z)\cdot\frac{\frac{1}{2^p}}{1-\frac{1}{2}z} = \frac{1}{2^p}$$

This answer is the same as that previously obtained.

12.3 Noise filtering example

In this section we will repeat the noise filtering example of section 11.3 under the additional condition that the optimum filter be physically realizable, (and that the output be an up-to-date approximate reproduction of the desired message). Under the data given it was found that

$$\Phi_{ff}(z) = \frac{(1-\frac{1}{3}z)(1-\frac{1}{3}z^{-1})}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

$$\Phi_{fd}(z) = \frac{\frac{5}{18}}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}$$

Thus

$$\Phi^{+}{}_{ff}(z) = \frac{1-\frac{1}{3}z}{1-\frac{1}{2}z} \qquad \Phi^{-}{}_{ff}(z) = \frac{1-\frac{1}{3}z^{-1}}{1-\frac{1}{2}z^{-1}}$$

$$\frac{\Phi_{fd}(z)}{\Phi^{-}{}_{ff}(z)} = \frac{\frac{5}{18}}{(1-\frac{1}{2}z)(1-\frac{1}{3}z^{-1})} = \frac{\frac{1}{3}}{1-\frac{1}{2}z} + \frac{\frac{1}{9}z^{-1}}{1-\frac{1}{3}z^{-1}}$$

Hence

$$\left[\frac{\Phi_{fd}(z)}{\Phi^{-}{}_{ff}(z)}\right]_{+} = \frac{\frac{1}{3}}{1-\frac{1}{2}z}$$

and the optimum realizable noise filter has the transfer function (equation E)

$$H_o(z) = \frac{1-\frac{1}{2}z}{1-\frac{1}{3}z}\cdot\frac{\frac{1}{3}}{1-\frac{1}{2}z} = \frac{\frac{1}{3}}{1-\frac{1}{3}z}$$

12.4 A special example

We remarked in chapter 10 that if $\Phi_{ff}(z)$ has zeros on the unit circle, no system can be found which converts f(n) to uncorrelated samples. To clarify this remark and to show how the difficulty carries over to the optimum filter problem, we will work a special example,

Assume that f(n) has been generated by passing a signal consisting of mutually independent +1's and –1's occurring each with probability 1/2 (the penny-pitching signal of chapter 9) through a system with the transfer function (1–z). We want to design an optimum, linear, physically realizable predictor to predict f(n) one unit of time in the future. That is, the desired output is d(n)=f(n+l). Then

$$\Phi_{ff}(z) = (1-z)(1-z^{-1}) = 2-z-z^{-1}$$

and

$$\varphi_{fd}(k) = \overline{f(n)d(n+k)} = \overline{f(n)f(n+1+k)} = \varphi_{ff}(k+1)$$

or

$$\Phi_{fd}(z) = z^{-1}\,\Phi_{ff}(z)$$

Note that $\Phi_{ff}(z)$ has a double zero on the unit circle at z=1.

If we blindly use the results of section 12.1, equation E, we obtain,

$$\Phi^{+}{}_{ff}(z) = 1-z \qquad \Phi^{-}{}_{ff}(z) = 1-z^{-1}$$

$$\left[\frac{\Phi_{fd}(z)}{\Phi^{-}{}_{ff}(z)}\right]_{+} = \left[z^{-1}-1\right]_{+} = -1$$

$$H_o(z) = -\frac{1}{1-z}$$

The factorization here of Φ_{ff} into $\Phi^{-}{}_{ff}$ and $\Phi^{+}{}_{ff}$ is attempted by associating one of the zeros on the unit circle with the interior of circle and one with the exterior. The $H_o(z)$ we obtain is the transform of the impulse response

$$h_o(n) = \begin{cases} 0 & n<0 \\ -1 & n\geq 0 \end{cases}$$

Clearly, this "optimum" impulse response does not tend to zero geometrically as $n\to\pm\infty$ but maintains the value –1 as $n\to\pm\infty$. Therefore, the theory given in section 11.1 leading to equations A and B doesn't hold true. The theory requires the absolute convergence of infinite sums involving the impulse response and wouldn't apply for the $h_o(n)$ we have derived.

Furthermore, we will see that it is obvious that this $h_o(n)$ can't be a predictor of f(n). f(n) is a statistically stationary sequence of samples which neither tend to

decrease or increase in magnitude as n→±∞. The output of the filter, g(n), according to $h_o(n)$, is the negative of the sum of the present and all past samples. Because of stationarity this sum cannot converge (except with probability zero). Our optimum filter has too long a memory for the influence of past samples. The above $h_o(n)$ won't produce a defined output unless we forego stationarity and give precise information on how f(n) started up in the past. Such information is not often available in physical problems.

However, these remarks do not imply that we can't design a good predictor for f(n). It turns out that the minimum mean-square error can be approached as closely as we please but not actually attained. One approach is to replace $h_o(n)$ above with the well-behaved response:

$$h_o(n) = \begin{cases} 0 & n < 0 \\ -(a^n) & n \geq 0 \quad with\ 0 < a < 1 \end{cases}$$

where $h(n) \rightarrow h_o(n)$ as $a \rightarrow 1$.

h(n) has proper convergence as n→±∞ and the equation for $\overline{e_o^2}$, section 11.1 becomes

$$\overline{e_o^2} = \varphi_{dd}(0) - 2\sum_{m=-\infty}^{+\infty} h(m)\,\varphi_{fd}(m) + \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(m)\,h(j)\,\varphi_{ff}(m-j)$$

$$= \varphi_{dd}(0) - 2\sum_{m=0}^{+\infty} -(a^m)\,\varphi_{ff}(m+1) + \sum_{m=0}^{+\infty}\sum_{j=0}^{+\infty} a^{m+j}\varphi_{ff}(m-j)$$

$$= 2 - 2a^0 + 2\sum_{m=-0}^{+\infty} a^{2m} - 2\sum_{m=-0}^{+\infty} a^{2m+1}$$

$$= 2\left[\frac{1}{1-a^2} - \frac{a}{1-a^2}\right] = \frac{2}{1+a}$$

By choosing (a) close to 1 we can nearly achieve the minimum error of $\overline{e_o^2} = 1$.

This technique of modifying the formal optimum solution is generally applicable when Φ_{ff} has zeros on the unit circle.

12.5 Problems

12.5.1) Solve problem 11.4.1a using the transform methods developed in chapter 12.

12.5.2a) Show that the optimum linear predictor that is physically realizable and that predicts p units of time in advance has the transfer function

$$H_o(z) = \frac{1}{\Phi^+_{ff}(z)}\left[z^{-p}\,\Phi^+_{ff}(z)\right]_+$$

(Notation as in equation E).

b. Show that the minimum mean-square prediction error is equal to the sum of the squares of the samples for n<0 thrown away when $z^{-p}\,\Phi^+_{ff}(z)$ is converted into $\left[z^{-p}\,\Phi^+_{ff}(z)\right]_+$. Apply this result to problem 1 and check with problem 1b of chapter 11.

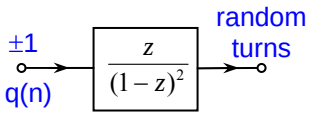
12.5.3) Solve problem 2, chapter 11 under the additional constraint that the optimum filter be physically realizable.

13. Control systems problem

13.1 Problem formulation

In this chapter we renew the analysis of the driver control system of chapter 7. We will be particularly interested in applying random inputs to the system and in optimizing system performance by adjusting A and B, the constants describing the action of the man in the control loop (see section 7.1). A special technique of mean-square error evaluation will be given.

To determine characteristics of the system input f(n), consider the following conditions. The driver attempts to follow the white line in the middle of the road. The road itself makes an abrupt right or left turn of unit magnitude at each sampling point. These right or left turns are chosen independently and with equal probability (1/2) at each point. As before, we assume that the turns are of small magnitude and, therefore the transverse displacement of the roadway at sampling times can be generated as a signal by applying a random, independent sequence of + and – 1's (each with probability 1/2) to a filter with transfer function $z/(1–z)^2$. The ±1 sequence we will denote by q(n).

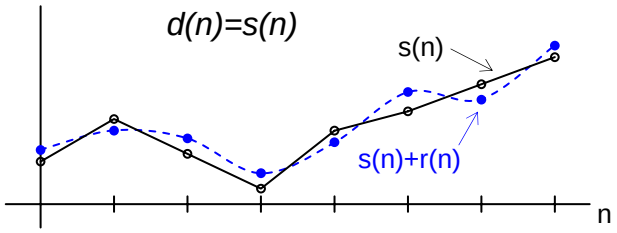


Remember that this transfer function generates a ramp for each input sample. The output is the superposition of these ramps - the required series of random turns.

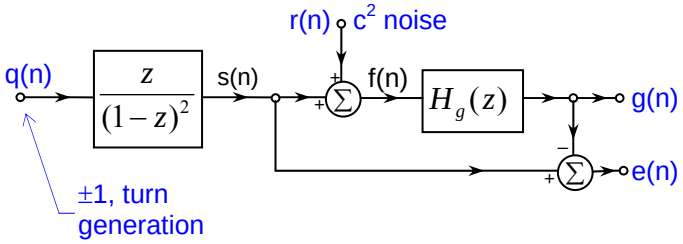
In addition, the man who drew the white line in the road did so erratically, making an independent random transverse position error of zero mean and variance c^2 at each sampling point. We assume that these errors have no statistical connection with the turns in the roadway. Hence, the white line displacement has two statistically independent components, one due to true turns in the road and a second due to the white line error from true road center. The control system input is the sum of these two partial inputs. Call s(n) the turn displacement and r(n) the white line error. The system input f(n) is given by

$$f(n) = s(n) + r(n)$$

s(n) is the turn displacement signal generated above. It's the desired actual motion of the car d(n).



r(n) represents false information, noise, in the system input. The system output error, e(n), is
e(n) = d(n)–g(n) = s(n)–g(n).



13.2 Optimum filter derivation

To determine the best control system for the above road conditions, we will use the theory of the previous chapters to find the optimum transfer function,

$$\left[H_g(z) \right]_{opt} = H_o(z)$$

This means that we won't initially restrict ourselves to a specific form for $H_o(z)$ involving A and B constants.

However, we will impose physical realizability on $H_o(z)$. A second natural restriction is also required. The position of the car g(n) at any time n can't be influenced by the input value f(n) at the same time. One unit of time must pass before the effect of input deviations can be felt in the output. This is physically apparent and leads to the condition that the impulse response value $h_o(0)$ is zero. This together with physical realizability means that we must set $h_o(n)=0$ for $n\leq 0$ and only adjust $h_o(n)$ for $n>0$ to obtain optimum performance.

The second restriction requires a modification of our optimum realizable filter technique. According to chapter 12 equation E we would have (with only the realizability condition,

$$H_o(z)=\frac{1}{\Phi^{+}{}_{ff}(z)}\left[\frac{\Phi_{fd}(z)}{\Phi^{-}{}_{ff}(z)}\right]_{+}$$

An inspection of the proof of equation E shows that the modification required is in the evaluation of the factor $\left[\Phi_{fd}(z)/\Phi^{-}{}_{ff}(z)\right]_{+}$. We threw away any samples occurring before n=0 of the time domain representation of the bracketed quantity because of realizability. Now we modify this procedure to also throw away the sample for n=0 because of our second restriction above. The new transform we will call $\left[\Phi_{fd}(z)/\Phi^{-}{}_{ff}(z)\right]_{++}$. Therefore, in our problem,

$$H_o(z)=\frac{1}{\Phi^{+}{}_{ff}(z)}\left[\frac{\Phi_{fd}(z)}{\Phi^{-}{}_{ff}(z)}\right]_{++}$$

Another difficulty in our particular problem arises because s(n) (and hence f(n)) does not really have a correlation function in the sense we have defined them. The difficulty is that s(n) is not stationary, since in making random turns, the road tends to deviate further and further from the central axis. Mathematically the difficulty lies in the double pole on the unit circle introduced by the $z/(1-z)^2$ transfer function. This carries over into the determination of $\Phi_{ss}(z)$ and invalidates our preceding theory, which requires all correlation transforms to be analytic in some annulus containing the unit circle boundary in its interior.

We circumvent this difficulty as follows. Instead of our original turn generation through $z/(1-z)^2$ we use the transfer function $z/(1-az)^2$, where $0<a<1$, Having solved the problem with this input, which is a legitimate one, we let $a\rightarrow 1$ and use the resulting limiting solution as the final answer. It turns out that an explicit use of this limiting process is not necessary, and that extending the mechanisms to the present problem without change can get the same result. If this procedure makes you uneasy, imagine the convergence factor, a, to be present in the expressions, even though we don't actually include it. No operations will be found which depend critically on whether a is one or a little less than one.

We have, since the ±1 input has the correlation transform, 1,

$$\Phi_{ss}(z)=\frac{z}{(1-z)^2}\cdot\frac{z^{-1}}{(1-z^{-1})^2}\cdot 1=\frac{1}{(1-z)^2(1-z^{-1})^2}$$

Also since r(n) consists of independent samples with a mean-square of c^2.

$$\Phi_{rr}(z)=c^2$$

Since f(n)=s(n)+r(n), and s and r are independent,

$$\Phi_{ff}(z)=\Phi_{ss}(z)+\Phi_{rr}(z)$$

$$=\frac{1}{(1-z)^2(1-z^{-1})^2}+c^2=\frac{1+c^2(1-z)^2(1-z^{-1})^2}{(1-z)^2(1-z^{-1})^2}$$

$$=\frac{(L+Mz+Nz^2)(L+Mz^{-1}+Nz^{-2})}{(1-z)^2(1-z^{-1})^2}$$

The last representation of $\Phi_{ff}(z)$ is in factored form. We have introduced coefficients L, M. N which must be determined by using the identity:

$$1+c^2(1-z)^2(1-z^{-1})^2=(L+Mz+Nz^2)(L+Mz^{-1}+Nz^{-2})$$

Setting z=1 we see immediately that $(L+M+N)^2=1$ and we lose no generality by setting L+M+N=1. Matching the coefficients of z^2 and z (or z^{-1}) on both sides we obtain, respectively, $LN=c^2$ and $M(L+N)=-4c^2$. In summary three equations determining L, M, N in terms of c^2 are:

$$\begin{cases}L+M+N=1\\ LN=c^2\\ M(L+N)=-4c^2\end{cases}$$

Factoring $\Phi_{ff}(z)$ we obtain

$$\Phi^{+}{}_{ff}(z)=\frac{L+Mz+Nz^2}{(1-z)^2};\quad \Phi^{-}{}_{ff}(z)=\frac{L+Mz^{-1}+Nz^{-2}}{(1-z^{-1})^2}$$

Since d(n)=s(n), we find,

$$\varphi_{fd}(k)=\overline{[s(n)+r(n)][s(n+k)]}$$
$$=\overline{s(n)s(n+k)}+\overline{r(n)}\cdot\overline{s(n+k)}=\overline{s(n)s(n+k)}$$
$$=\varphi_{ss}(k)$$

or

$$\Phi_{fd}(z)=\Phi_{ss}(z)=\frac{1}{(1-z)^2(1-z^{-1})^2}$$

Now the computation of the optimum filter follows (eliminating M using L+M+N=1):

$$\frac{\Phi_{fd}(z)}{\Phi^{-}_{ff}(z)} = \frac{1}{(1-z)^2(L+Mz^{-1}+Nz^{-2})}$$

$$= \frac{z}{(1-z)^2} + \frac{(L-N)z}{1-z} + \left\langle \begin{matrix} terms\ whose\ expansions \\ yield\ \ powers,\ z^k, with\ k \le 0 \end{matrix} \right\rangle$$

$$\left[\frac{\Phi_{fd}(z)}{\Phi^{-}_{ff}(z)}\right]_{++} = \frac{z}{(1-z)^2} + \frac{(L-N)z}{1-z}$$

$$= \frac{(1+L-N)z-(L-N)z^2}{(1-z)^2}$$

Therefore, the optimum transfer function is,

$$H_o(z) = \frac{1}{\Phi^{+}_{ff}(z)}\left[\frac{\Phi_{fd}(z)}{\Phi^{-}_{ff}(z)}\right]_{++} = \frac{(1+L-N)z-(L-N)z^2}{L+Mz+Nz^2}$$

13.3 Transfer function comparisons

The $H_o(z)$ transfer function above represents the best linear control system characteristics under the assumed restrictions and inputs. How does $H_o(z)$ compare with the actual $H_g(z)$ that we assumed in chapter 7? We have, (eliminating M),

$$H_o(z) = \frac{(1+L-N)z-(L-N)z^2}{L+(1-L-N)z+Nz^2}$$

and from section 7.1,

$$H_g(z) = \frac{Az-Bz^2}{1+(A-2)z+(1-B)z^2}$$

Identifying,

$$A = \frac{1+L-N}{L}; \qquad B = \frac{L-N}{L}$$

or

$$L = \frac{1}{A-B}; \qquad N = \frac{1-B}{A-B}$$

We see that $H_o(z)$ and $H_g(z)$ are of the same form. So the generality of $H_g(z)$ is sufficient to obtain the actual optimum linear filter by suitably adjusting A and B.

Using the relations between L, N, M, and c^2 we find,

$$M(L+N) = -4c^2 = -4LN$$
$$(1-L-N)(L+N) = -4LN$$
$$L+N-L^2-2LN-N^2 = -4LN$$
$$L+N = (L-N)^2$$

In terms of A and B this yields.

$$(2-B)(A-B) = B^2$$

$$2A-AB-2B = 0$$

$$B = \frac{2A}{A+2}$$

Thus, for our optimum filter B and A must be related by this formula. Developing a further relation,

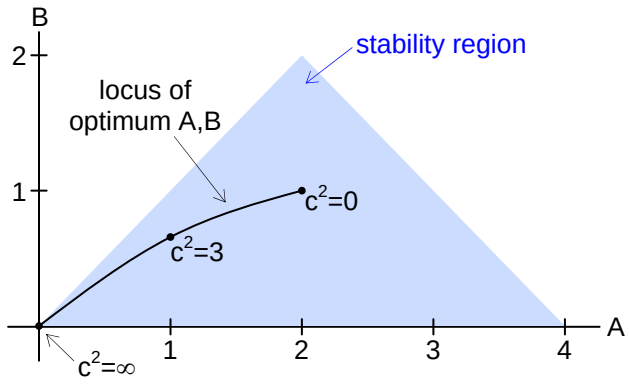
$$LN = \frac{1-B}{(A-B)^2} = c^2$$

The equations,

$$B = \frac{2A}{A+2}; \qquad \frac{1-B}{(A-B)^2} = c^2$$

determine optimum A, B, when c^2 is given. When c>0 we also have: $A^2 = (\sqrt{16c^2+1}-1)/2c^2$.

As c^2 varies, the point (A, B) runs over a locus in the A,B plane diagram of section 7.2. Each A,B pair on the locus is associated with a particular c^2 value. The locus looks like this:



For example, if there is no noise (perfectly painted white line), $c^2=0$ and optimum constants are A=2, B=1. These are just the constants suggested after the discussion of section 7.5. (There was no noise in that problem.) For $c^2=3$, we obtain A=l, B=2/3, and for $c^2=+\infty$, A=0, B=0. The greater the noise, the smaller are the optimum A, B.

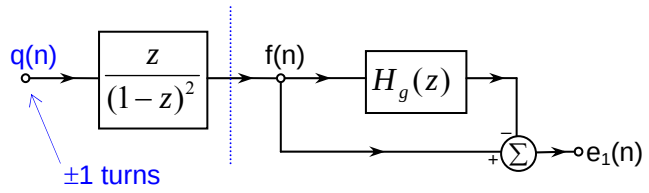
Note that we had fixed the turn magnitudes at unity (positive or negative). We can generalize this as follows. If the turn magnitudes had been d (with a turn generation transfer function of $dz/(1-z)^2$ the above results can be retained if we regard c as normalized relative to the turn magnitudes d. The formulas for $d\neq 1$ would follow by replacing c^2 by c^2/d^2 in the preceding results.

13.4 Mean square error

For any system transfer function, $H_g(z)$, a certain mean-square error is obtained. Since $H_g(z)$ contains A and B as parameters, the mean-square error should also contain A, B as parameters.

The system error, e(n) is due to two causes. First the turning of the roadway is not followed exactly (even in the absence of noise). Secondly, the noise introduced by the white line errors induces an error in the output (even in the absence of turns).

If noise is absent the block diagram of section 13.1 reduces to:



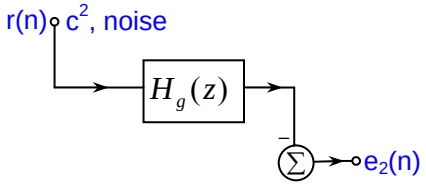
The control system proper is represented to the right of the dotted line. We found its error transfer function in section 7.1.

$$H_e(z)=\frac{(1-z)^2}{(1-z)^2+(A-Bz)z}$$

Multiplying by $z/(1-z)^2$ we obtain the over-all transfer function, $H_1(z)$ which relates the ± turn generation samples, q(n), to the final error they cause, $e_1(n)$. This is the first error component.

$$H_1(z)=\frac{z}{(1-z)^2}H_e(z)=\frac{z}{1+(A-2)z+(1-B)z^2}$$

If the turns are absent and noise is the only input, the block diagram of section 13.1 reduces to



So the transfer function relating the error component due to noise, $e_2(n)$, to the noise samples, r(n), is

$$H_2(z)=-H_g(z)=\frac{-(A-Bz)z}{1+(A-2)z+(1-B)z^2}$$

By the superposition principle for a linear system, the total error, e(n), is the sum of these component errors (call them $e_1(n)$ and $e_2(n)$).

$$e(n)=e_1(n)+e_2(n)$$

$e_1(n)$ and $e_2(n)$ are statistically independent because the signals which generate them are independent, there being no statistical connection between the noise samples and the turn-generating samples. Thus,

$$\overline{e^2(n)}=\overline{e_1^2(n)}+\overline{e_2^2(n)}+\overline{2e_1(n)}\cdot\overline{e_2(n)}$$

$$=\overline{e_1^2(n)}+\overline{e_2^2(n)}$$

(since $\overline{e_1(n)}=\overline{e_2(n)}=0$). Therefore, the total mean-square error is the sum of the two component mean-square errors.

The component mean-square errors are the mean-square values of the outputs of systems with transfer functions $H_1(z)$ and $H_2(z)$ where the inputs are sequences of independent samples. For $H_1(z)$ these samples are ±1 in magnitude. For $H_2(z)$ the samples have a standard deviation of c. So the autocorrelation functions of the inputs g(n), r(n) are:

$$\varphi_{gg}(k)=\delta(k);\qquad \varphi_{rr}(k)=c^2\delta(k)$$

Using a result of section 9.5 we can determine the output mean-square values, $\overline{e_1^2}$ and $\overline{e_2^2}$ in terms of these known correlation functions and the impulse responses, $h_1(n)$, $h_2(n)$ corresponding to $H_1(z)$, $H_2(z)$.

$$\overline{e_1^2} = \varphi_{e_1 e_1}(0) = \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h_1(m)\, h_1(m+j)\, \varphi_{gg}(j)$$

$$= \sum_{m=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h_1(m) h_1(m+j) \delta(j) = \sum_{m=-\infty}^{+\infty} h_1^2(m)$$

Similarly,

$$\overline{e_2^2} = c^2 \cdot \sum_{m=-\infty}^{+\infty} h_2^2(m)$$

These results indicate the mean-square error may be determined by finding the sum of the squares of impulse response values. So, the problem is essentially equivalent to that discussed briefly in section 7.5 with regard to non-random inputs. Here, however, we want to give a more practical method for the sum of squares calculation.

13.5 Sum of squares

The sum of squares of the impulse response values corresponding to $H_l(z)$ or $H_2(z)$ is most easily determined by reformulating the systems problem in matrix form by the methods of chapter 6.

First of all, however, let's note an immediate slight simplification. $H_1(z)$ has a z factor in the numerator. If this factor is removed, the only influence on $h_1(n)$ values is to shift them all one unit of time to the left. Since the sum of squares is unchanged, in place of the original $H_1(z)$ we may use the transfer function

$$H_1'(z) = \frac{1}{1+(A-2)z+(1-B)z^2} = \frac{G_1(z)}{F_1(z)}$$

Similarly the –z factor may be removed in $H_2(z)$ to obtain

$$H_2'(z) = \frac{A-Bz}{1+(A-2)z+(1-B)z^2} = \frac{G_2(z)}{F_2(z)}$$

In each case we have written the transfer function in terms of the ratio of transforms of an input and output signal of the H_1 and H_2. For inputs we want to use $f_1(n)=f_2(n)=\delta(n)$. The outputs $g_1(n)$ and $g_2(n)$ will be system impulse responses. We want to find:

$$\sum_{n=0}^{\infty} g_1^2(m) \qquad \textit{and} \qquad \sum_{n=0}^{\infty} g_2^2(m)$$

We illustrate the method with $H_1(z)$. The expression for G_1/F_1 yields:

$$G_1(z) + (A-2)zG_1(z) + (1-B)z^2 G_1(z) = F_1(z)$$

In the time domain (as in section 6.1) the equivalent difference equation is

$$g_1(n) + (A-2)g_1(n-1) + (1-B)g_1(n-2) = f_1(n)$$

As in section 6.3, we introduce an extra variable, u, by the definition $u(n) = g_1(n-1)$ and rewriting the above equation, we obtain a set of equations which can be written in matrix form.

$$\begin{cases} g_1(n) = (2-A)g_1(n-1) + (B-1)u(n-1) + f_1(n) \\ u(n) = g_1(n-1) \end{cases}$$

Since the system is dead ($g_1(n)=u(n)=0$) until excited by non-zero $f_1(n)$ values and since we are taking $f_1(n)=\delta(n)$, the equations show that

$$g_1(n) = u(n) = 0 \qquad \textit{for } n<0$$
$$g_1(0) = 1; \qquad u(n) = 0$$

For $n \geq 1$, $f_1(n)$ vanishes, and the values for n=0 may be interpreted as initial values, using only the equations,

$$\begin{cases} g_1(n) = (2-A)g_1(n-1) + (B-1)u(n-1) \\ u(n) = g_1(n-1) \qquad\qquad (\textit{for } n \geq 1) \end{cases}$$

to find successive values of $g_1(n)$ and u(n).

We are, however, interested in squares $g_1^2(n)$, not in the $g_1(n)$ values themselves. Therefore, we convert these last equations to equations in the variables $g_1^2(n)$, $g_1(n)u(n)$, $u^2(n)$ by squaring both sides of each equation and by multiplying equations to obtain three equations in the three new unknowns.

$$g_1^2(n) = (2-A)^2 g_1^2(n-1) + 2(2-A)(B-1)g_1(n-1)u(n-1) + (B-1)^2 u^2(n-1)$$
$$g_1(n)u(n) = (2-A)g_1^2(n-1) + (B-1)g_1(n-1)u(n-1)$$
$$u^2(n) = g_1^2(n-1) \qquad (\textit{for } n \geq 1)$$

Summing each equation for all n from 1 to $+\infty$ and adding and subtracting the known values $g_1^2(0)=1$, $g_1(0)u(0)=u^2(0)=0$ from the left sums to extend the range of the sums to n=0, we obtain,

$$\sum_0^\infty g_1^{\,2}(n) - 1 = (2-A)^2 \sum_0^\infty g_1^{\,2}(n-1) +$$

$$2(2-A)(B-1)\sum_0^\infty g_1(n)u(n) + (B-1)^2 \sum_0^\infty u^2(n)$$

$$\sum_0^\infty g_1(n)u(n) = (2-A)\sum_0^\infty g_1^{\,2}(n) + (B-1)\sum_0^\infty g_1(n)u(n)$$

$$\sum_0^\infty u^2(n) = \sum_0^\infty g_1^{\,2}(n-1)$$

This is a set of three equations in the three unknowns, $\sum g_1^{\,2}$, $\sum g_1 u$, $\sum u^2$. We solve for the first of these, the required sum of squares by employing Cramer's rule with determinants. So, if

$$\Delta = \begin{vmatrix} (2-A)^2 - 1 & 2(2-A)(B-1) & (B-1)^2 \\ 2-A & B-2 & 0 \\ 1 & 0 & -1 \end{vmatrix}$$

$$M = \begin{vmatrix} -1 & 2(2-A)(B-1) & (B-1)^2 \\ 0 & B-2 & 0 \\ 0 & 0 & -1 \end{vmatrix}$$

we have

$$\sum_0^\infty g_1^{\,2}(n) = \frac{M}{\Delta}$$

Direct evaluation of Δ and M yields

$$\Delta = B(B-A)(4-A-B)$$
$$M = B-2$$

So, the mean-square error component due to turns is,

$$\overline{e_1^{\,2}} = \sum_0^\infty h_1^{\,2}(n) = \sum_0^\infty g_1^{\,2}(n) = \frac{B-2}{B(B-A)(4-A-B)}$$

The denominator is not surprising. Each factor, when set to zero, yields a relation between A and B that represents one of the three lines bounding the stability region in the A, B plane. It's reasonable that $\overline{e_1^{\,2}}$ should tend to infinity as the boundary of the stability region is approached.

In the special case, A=2, $\overline{e_1^{\,2}}$ becomes,

$$\overline{e_1^{\,2}} = \frac{1}{B(2-B)} = \frac{1}{1-(1-B)^2}$$

which is the same formula we derived in section 7.5. Now, we have determined the general result. In section 7.5, the sum of squares represented the actual sum of squares of errors of the control system under a unit ramp input. In the current problem it represents the mean-square error for random unit turns. Thus, these two types of errors are numerically equal.

An analysis similar to the above leads to a determination of the sum of squares in the second case with $H_2(z)$. The only variation is that the direct influence of the forcing term $f_2(n)$ disappears only after the step with n=2 is reached. So it's convenient to take the sums over the range 1 to ∞ and determine

$$\sum_1^\infty g_2^{\,2}(n)$$

Then the value of $g_2^{\,2}(0)$ is added to the final sum. The determinant, Δ, is the same as before. The answer is:

$$\overline{e_2^{\,2}} = c^2 \sum_0^\infty g_2^{\,2}(n) = c^2 \left(g_2^{\,2}(0) + \sum_1^\infty g_2^{\,2}(n) \right)$$

$$= c^2 \frac{B(B-2) + A(B+2)}{B(4-A-B)}$$

The total mean-square error is,

$$\overline{e^2} = \overline{e_1^{\,2}} + \overline{e_2^{\,2}}$$

$$= \frac{B-2}{B(B-A)(4-A-B)} + c^2 \frac{B(B-2) + A(B+2)}{B(4-A-B)}$$

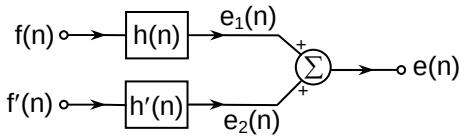
The minimum mean-square error $\overline{e_0^{\,2}}$ can be found by substituting the optimum A, B in this expression. If random turns of magnitude d are made instead of turns of unit magnitude, the $\overline{e_1^{\,2}}$ term above would be multiplied by an extra d^2 factor.

14. Systems with random switching

14.1 Mean-square error components

In this chapter we will discuss the analysis and design of systems in which the signal flow at some point is interrupted by a switch that opens and closes randomly. This switch action may represent the actual operation of a physically existing system or it may be introduced to obtain a mathematical model which accounts for certain missing signal samples or other phenomena.

First we need to generalize this result from section 13.4: If two inputs are simultaneously applied to a system, then the mean-square value of the output $\overline{e^2}$ is the sum of the mean squares, $\overline{e_1^2}$ and $\overline{e_2^2}$ obtained by separate application of each input provided that these inputs are not statistically dependent on one another. (We assume, of course, that mean values of the input and output are zero.) We now show that this relation still holds if the two inputs, say f(n) and f'(n) are not crosscorrelated. That is, if $\varphi_{ff'}(k)=0$ for all k.



Using the superposition summation, we write,

$$\overline{e_1(n)e_2(n)} = \overline{\sum_{m=-\infty}^{+\infty} h(m)\,f(n-m)\cdot \sum_{j=-\infty}^{+\infty} h'(m)\,f'(n-m)}$$

$$= \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(m)h'(j)\,\overline{f(n-m)\,f'(n-j)}$$

$$= \sum_{m=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty} h(m)h'(j)\varphi_{ff'}(m-j)$$

If $\varphi_{ff'}(k)=0$ for all k, we obtain

$$\overline{e_1(n)e_2(n)} = 0$$

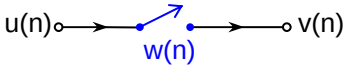
which is sufficient to prove the required result through the following formula.

$$\overline{e^2} = \overline{(e_1+e_2)^2} = \overline{e_1^2} + \overline{e_2^2} + \overline{2e_1e_2} = \overline{e_1^2} + \overline{e_2^2}$$

This result is also extendable to situations with more than two inputs.

14.2 Random switching

This diagram indicates the basic switching operation.



The output sample v(n) is equal to the input sample u(n) if the switch is closed. If the switch is open the output is zero. w(n) represents the switch modulation factor. We set w(n) to zero for every n where the switch is open and to one when the switch is closed. Therefore

$$v(n) = w(n)\,u(n)$$

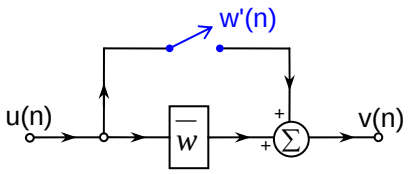
We have deliberately avoided the standard f, g input/output notation here because the switch system shown may actually be imbedded in a larger system. So u(n) may not be one of the over-all system inputs and v(n) may not be the over-all system output. Often, there may be feedback from v(n) to u(n).

We make the following assumptions about the switching function w(n). We assume the switch value is selected at random from a common distribution of values (0 and +1) by an independent trial each sampling time. Thus w(n) is a stationary random signal consisting of a sequence of statistically independent 0's and 1's. Furthermore, we assume w(n) is statistically independent of any of the inputs to the over-all system in which the switch is imbedded. This does not mean that w(n) and u(n) are not statistically related since u(n), as noted above, may not be such an input. In fact, if u(n) is influenced by v(n) through feedback, w(n) and u(n) will usually be related.

We also impose a condition on the type of system in which the switch is imbedded. We assume that either feedback from v(n) to u(n) is absent or it has the character that a unit sample sent out at the v(n) terminal must not evoke any response at the u(n) terminal until at least one unit of time has elapsed. This means that the effect of any switch action cannot be transmitted around the feedback loop without at least one unit of time delay. So we may conclude under this assumption that the samples u(n) and w(n+k) are statistically independent as

long as k>0, for whatever the value of u(n) is, it cannot be statistically dependent on a present or future switch value, w(n+k), which is randomly selected. This system assumption will be satisfied naturally in most physically motivated problems.

The crux of the method for handling random switching is to replace the switch by an equivalent system without the switch. Two steps are needed,



This diagram shows a system equivalent to the original switch.

The original switch has been replaced by a parallel combination of a constant gain, $\overline{w}$, equal to the mean of the switching variable distribution and a second switching function w'(n) related to w(n) by

$$w(n) = \overline{w} + w'(n)$$

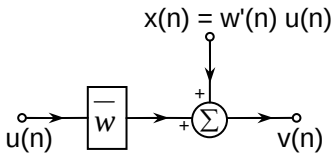
where w'(n) is just the sequence of original 0's and 1's with their mean value, $\overline{w}$, subtracted. The switch in the new diagram is only symbolic. It signifies that a modulation operation, multiplying f(n) by w'(n) is taking place, where w'(n) is $-\overline{w}$ or $(1-\overline{w})$ according to whether w(n) is 0 or +1.

w'(n) now has all the previously assumed properties of w(n) except that its mean, $\overline{w'}$ is zero.

v(n) is now written as the sum of two signals.

$$v(n) = w(n)\,u(n) = \overline{w}\,u(n) + w'(n)\,u(n)$$

The final version of a system equivalent to the original switch is diagrammed as:



This diagram differs from the previous one only in that we do not represent the explicit mechanism by which the signal x(n)=w'(n)u(n) is obtained, but feed it in as if it were a system input. To justify this view we must demonstrate that x(n) has such simple properties that a full analysis of its origins is not necessary.

To this end, we determine the autocorrelation function $\varphi_{xx}(k)$ and the crosscorrelation function $\varphi_{xf}(k)$ where f(n) is any input to the over-all system containing the switch. We find,

$$\varphi_{xx}(k) = \overline{x(n)\,x(n+k)} = \overline{w'(n)\,u(n)\,w'(n+k)\,u(n+k)}$$

For k>0 our assumptions about w(n) and the system properties, insure that the quantities [w'(n)·u(n)·u(n+k)] and [w'(n+k)] are statistically independent. So for k>0

$$\varphi_{xx}(k) = \overline{w'(n)\,u(n)\,u(n+k)} \cdot \overline{w'(n+k)} = 0$$

The same result holds for k<0, by symmetry. For k=0 the quantities [w'(n)·w'(n)] and [u(n)·(n)] are statistically independent and

$$\varphi_{xx}(0) = \overline{w'^2} \cdot \overline{u^2}$$

Therefore, $\varphi_{xx}(k) = \overline{w'^2} \cdot \overline{u^2} \cdot \delta(k)$ and x(n) is revealed to be a sequence of uncorrelated samples with zero mean and a mean-square of $\overline{w'^2} \cdot \overline{u^2}$

The crosscorrelation $\varphi_{xf}(k)$ is similarly evaluated:

$$\varphi_{xf}(k) = \overline{x(n)\,f(n+k)} = \overline{w'(n)\,u(n)\,f'(n+k)}$$

Our assumptions insure that the variables [w'(n)] and [u(n) f(n+k)] are statistically independent for all k. Thus

$$\varphi_{xf}(k) = \overline{w'(n)} \cdot \overline{u(n)\,f(n+k)} = 0$$

Therefore, x(n) is not correlated with any other system input.

These results are physically simple and convincing. They mean that the switch action can be replaced by

1) a direct transmission gain equal to $\overline{w}$, which is just the fraction of the time the switch is closed and

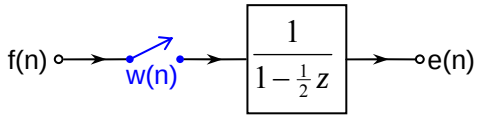
2) an additive, independent source of uncorrelated samples which acts as noise introduced by the switching fluctuation.

Since $\varphi_{xf}(k)=0$ for all k, the considerations of section 14.1 show that x(n) contributes an additive component to the total mean-square error of the over-all system equal to the mean-square error caused by x(n) alone.

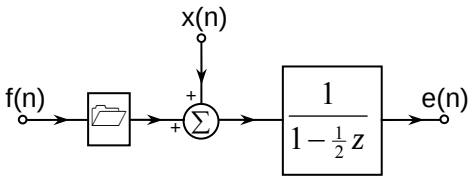
All of the above results can easily be generalized for the case where w(n) takes any number of values instead of just the two values 0 and +1. In fact, all the formulas hold without change. w(n) then no longer represents a simple switch, but a more complex device or phenomenon.

14.3 Elementary example

As a simple example, consider the effect of applying f(n) consisting of uncorrelated samples [$\varphi_{ff}(k)=\delta(k)$], to a switch which opens and shuts at random (each possibility with probability 1/2) and then to a filter with the transfer function 1/(1–0.5z). The switch action is independent of f(n). Call the output e(n).



Then $\overline{w}=\frac{1}{2}$, $\overline{f^2}=1$, and $\overline{w'^2}=\frac{1}{4}$ leads to the equivalent system:

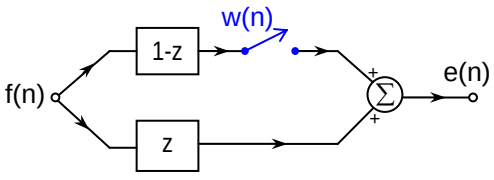


And we have $\overline{x^2}=\overline{w'^2}\cdot\overline{f^2}$ =1/4. The component of mean-square error provided by f(n) is the sum of squares (see section 13.4) of the impulse response corresponding to 0.5/(1–0.5z). The component due to x(n) is 1/4 times the sum of squares corresponding to 1/(1–0.5z).

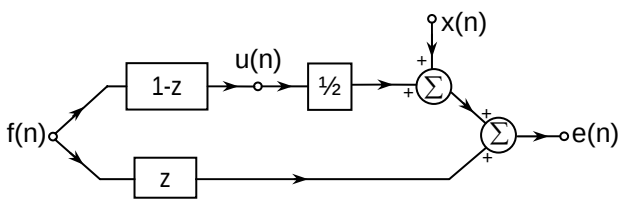
$$\overline{e^2}=\tfrac{1}{4}+\tfrac{1}{16}+\tfrac{1}{64}+\cdots+\tfrac{1}{4}(1+\tfrac{1}{4}+\tfrac{1}{16}+\cdots)=\tfrac{1}{3}+\tfrac{1}{4}(\tfrac{4}{3})=\tfrac{2}{3}$$

If the switch were left closed, $\overline{e^2}$ would be 4/3. If the switch were permanently open, $\overline{e^2}$ would be 0. Evidently $\overline{e^2}$ is an increasing function of the fraction of time the switch is closed.

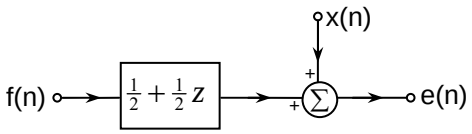
A second example shows how to build a model of a random delay by using a random switch.



When the switch is closed, the –z in the upper branch cancels the z in the lower branch and e(n)=f(n) for this n. If the switch is opened, f(n) reaches the output after a unit delay and e(n)=f(n–1). Assuming the same f(n) and w(n) as in the first example, we get:



We find $\overline{u^2}=2$, $\overline{x^2}=\overline{u^2}\cdot\overline{w'^2}=\frac{1}{2}$. A simplified equivalent system is:



so that

$$\overline{e^2}=(\tfrac{1}{4}+\tfrac{1}{4})\,\overline{f^2}+\overline{x^2}=1$$

14.4 Control system example with switching

The examples of section 14.3 are easy to work out even without the special use of equivalent systems as derived in the preceding discussion. There are examples where the answers are not so easy and where the machinery developed in 14.1 and 14.2 is more clearly essential.

To invent such an example, we return to the driver control system of chapters 7 and 13. Suppose that random turns are applied, and that $c^2=0$ (no white line position error) in the system of chapter 13, but we add a difficulty of a different kind: that the white line painter has failed to paint along some sections. Since we are only concerned with the sampling points we assume that the presence or absence of the white line at each such point is a random event occurring independently at each point. The probability of occurrence of the line is the same constant, β, at each sampling point.

Furthermore, the presence or absence of the line is assumed statistically independent of the random turns of the road, the remaining input of chapter 13.

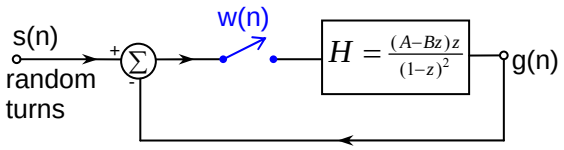
We assume that the driver meets the additional difficulty in the following way. If no line is seen at a sampling time, n, he assumes that he is headed correctly (i.e. that the error, e(n), is zero). If the line is seen, he measures e(n) and proceeds to generate a turning signal as before.

Thus the system operation and the effect of the white line misses can be simulated in a system model by inserting a random switch which opens and disconnects the e(n) signal whenever the white line is missing.

We can now solve two types of problems - the analysis of a particular system performance, or the design of an optimum system. The required design technique is that of chapters 11 and 12 with the use of the equivalent switch system, which introduces an extra noise source in

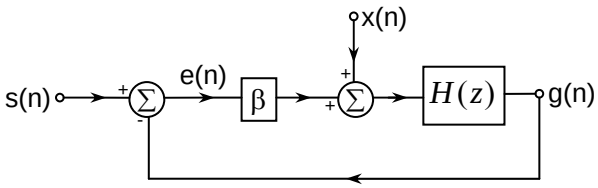
the system. In the following we restrict ourselves to the analysis problem.

The system containing the random switch effect, which we wish to analyze, is as follows.

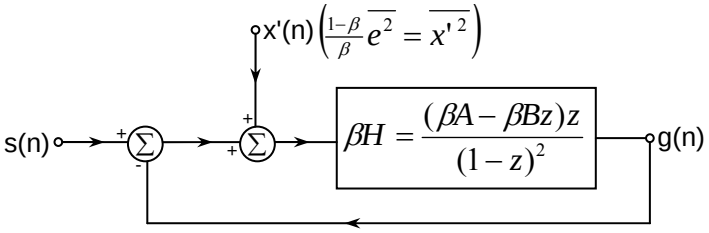


This is the basic system of chapter 7 with random turns as generated in chapter 13, with no white line position error, but with random white line omissions.

β the probability of occurrence of the line is equal to $\overline{w}$ the mean time the switch is closed. Also $\overline{w'^2} = \beta(1-\beta)$. Thus the use of an equivalent switch system gives,



We have u(n)=e(n) and so $\overline{x^2} = \overline{e^2} \cdot \overline{w'^2} = \beta(1-\beta)\overline{e^2}$. Now we move the x(n) input to the left of the β box and divide it by β (or its mean-square by β^2). This new signal input, x'(n), has the same effect on e(n) and g(n) as the original. Then we combine β and H(z) in series to obtain



Note that H(z) has a z factor in the numerator so any signal passing through the switch is delayed at least one unit of time around the feedback loop to the switch input. Therefore, this property together with the preceding assumptions assure that x'(n) consists of uncorrelated samples with a mean-square of $\frac{1-\beta}{\beta}\overline{e^2}$ and is independent of the other system input s(n). Thus, mean-square error from both sources can be found separately and added.

We see that βH has the same form as H except that the constants A, B now become βA and βB. Hence, in reducing the system to determine the transfer functions relating s(n) with e(n) and x'(n) with e(n) we find, as in section 7.1,

$$\frac{G(z)}{S(z)} = \frac{(1-z)^2}{1+(\beta A-2)z+(1-\beta B)z^2}$$

$$\frac{E(z)}{X'(z)} = \frac{-(\beta A-\beta B z)z}{1+(\beta A-2)z+(1-\beta B)z^2}$$

These were just the two transfer functions used in section 13.4 to determine components of the mean-square error except that A, B are now βA, βB. The effect of x'(n) is the same as r(n), the white line error before. The random turn input is the same as before. Hence to find $\overline{e^2}$ in the present case we can use the total mean-square error formula at the end of section 13.5 with $c^2 = \frac{1-\beta}{\beta}\overline{e^2}$ and A, B replaced by βA, βB to obtain:

$$\overline{e^2} = \frac{\beta B-2}{\beta B(\beta B-BA)(4-\beta A-\beta B)}$$

$$+\frac{1-\beta}{\beta}\cdot\overline{e^2}\cdot\frac{B(\beta B-2)+A(\beta B+2)}{B(4-\beta A-\beta B)}$$

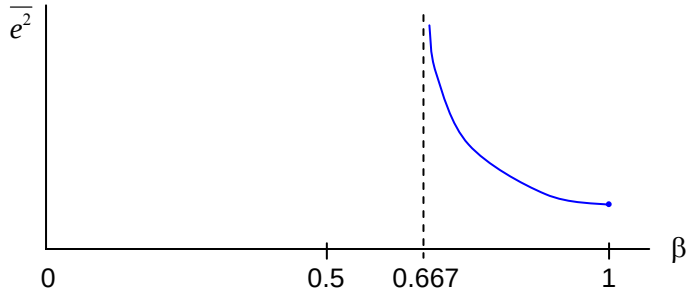
We note that we have determined the error in terms of itself. So we must solve this equation for $\overline{e^2}$ to obtain:

$$\overline{e^2} = \frac{\beta B-2}{\beta(B-A)\left[\beta(A+B)(2-B)+2(B-A)\right]}$$

Trying the case A=2, B=1, which we have found as optimum constants for β=1,

$$\overline{e^2} = \frac{2-\beta}{\beta(3\beta-2)}$$

Thus, as β drops from 1 to 2/3, $\overline{e^2}$ tends to ∞ showing the sensitivity of system performance to β. The formula fails for β<2/3. $\overline{e^2}$ is then still infinite.)



Appendix I. Stability criteria

Given a physically realizable transfer function, H(z), which is a rational function of z (ratio of polynomials), the question often arises whether H(z) represents the response of a stable system; that is, whether the impulse response, h(n), tends to zero as $n\to\infty$. As we saw in chapter 7, an equivalent question is whether the poles of H(z) are outside the unit circle in the z-plane.

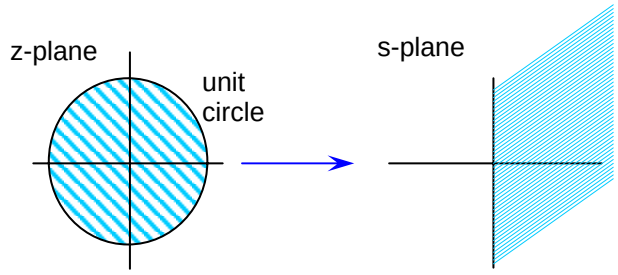
In chapter 7 we made a detailed examination of the location of the poles, the roots of the denominator polynomial of H(z). Generally it would be difficult to continue the offhand logic used there to cases where the denominator is of a degree higher than two. A more systematic procedure is required.

Our object here is to develop an algebraic connection between the coefficients of the denominator polynomial and stability. This is a more general problem than trying to determine stability represented by a given polynomial with numerical coefficients.

Suppose the denominator polynomial of H(z) is

$$D(z) = b_n z^n + \cdots + b_1 z + b_0$$

Substitute z=(1–s)/(1+s) and clear fractions, This transformation takes the interior of the unit circle in the z-plane into the right-hand s-plane.



Zeros of D(z) inside the unit circle in the z-plane map into zeros of D[(1–s)/(1+s)] in the right-hand s-plane. We can simplify considerations by throwing away the $(1+s)^n$ denominator of D[(1–s)/(1+s)] since this does not alter any zero positions. We also change the sign of D, if necessary to $a_0\geq 0$. In this way we arrive at a polynomial in s:

$$\pm(1+s)^n D\left(\tfrac{1-s}{1+s}\right) = P(s) = a_n s^n + \cdots + a_1 s + a_0$$

where the $a_0,\ldots,a_n$ are algebraic functions of the $b_0,\ldots,b_n$ and $a_0\geq 0$.

The determination of right-half plane zeros of P(s) is a well-known problem in continuous data system stability studies. We give the test without proof (see Guillemin, "The Mathematics of Circuit Analysis"). For P(s) to be free of right-half plane zeros, or zeros on the boundary, the necessary and sufficient conditions are that $a_0>0$ and that the following determinants be positive:

$$D_1 = a_1 > 0$$

$$D_2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix} > 0$$

$$D_3 = \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix} > 0$$

. . .

$$D_n = \begin{vmatrix} a_1 & a_0 & 0 & . & . & . & & & 0 \\ a_3 & a_2 & a_1 & a_0 & 0 & . & . & . & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 & a_0 & 0 & . & . & . & 0 \\ . & . & . & . & . & & & & \\ a_{2n-1} & a_{2n-2} & . & . & . & & & & a_n \end{vmatrix} > 0$$

(We set $a_m=0$ if m>n),

These determinants may be written as algebraic expressions in the $b_0,\ldots,b_n$.

Example

Consider a transfer function with the denominator

$$D(z) = 1 + (a + \alpha + \tfrac{A}{2} - 3)z + (3 - 2a - \alpha + \tfrac{A}{2})z^2 + (a - 1)z^3$$

Setting z=(1–s)/(l+s) and eliminating the $(l+s)^3$ factor we obtain, assuming A>0,

$$(1+s)^3 D\left(\tfrac{1-s}{1+s}\right)$$
$$= (8 - 4a - 2\alpha)s^3 + (4a - A)s^2 + 2\alpha s + A$$

Then we obtain the stability conditions,

$$a_0 = A > 0$$

$$D_1 = 2\alpha > 0$$

$$D_2 = 4(2a\alpha - 2A + aA) > 0$$

$$D_3 = 2(4 - 2a - \alpha) \cdot D_2 > 0$$

or

$$A > 0, \quad \alpha > 0, \quad 2a\alpha - 2A + aA > 0, \quad 4 - 2a - \alpha > 0$$

If we attempt to assume A<0 and use

$$-(1+s)^3 D\left(\tfrac{1-s}{1+s}\right)$$
$$= -(8 - 4a - 2\alpha)s^3 - (4a - A)s^2 - 2\alpha s - A$$

we find the resulting conditions incompatible. Hence it is necessary to set A>0.

Appendix II. Flow-graph reduction

In chapter 5 we found the system transfer function by progressively simplifying the flow graph until only a single branch bearing the over-all transfer function remained. While systematic, that method leads to long algebraic manipulations in complex examples.

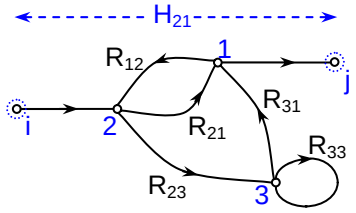
A simpler method for flow-graph reduction exists (Mason, I.R.E., July 1956) where the over-all transfer function can be written down almost by inspection. There are occasionally examples of flow graphs with an infinite number of branches which are better handled with the method of chapter 5. For simple examples like the control system of chapter 7 either method is suitable.

Method

Let $R_{ij}(z)$ be the transfer function of the system branch connecting node i (branch input) with node j (branch output). If more than one branch connects i with j then as a preliminary step we combine these branches and add the parallel transfer functions to determine $R_{ij}(z)$.

Let $H_{mn}(z)$ be the desired over-all node-to-node transfer function where the input is applied to node m and the output is observed at node n.

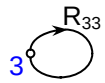
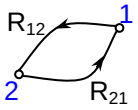
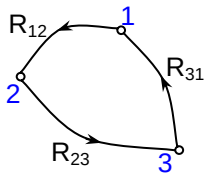
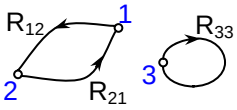
Let's use the following example with input-output connections and three nodes.



The transfer function $H_{mn}(z)$ is now given by the ratio of two expressions:

$$H_{mn}(z) = \frac{N(z)}{D(z)}$$

The denominator, D(z), is a sum of products of the R_{ij}. Each term of the sum is a product of R_{ij} around closed feedback loops of the flow graph. More than one loop may be involved in the product, but then the separate loops lending R_{ij}'s to any product must not touch or coincide at any point.
In the above example the following loops and their associated products are permissible.

3-3	R_{33}	
1-2-1	$R_{12}R_{21}$	
2-3-1-2	$R_{23}R_{31}R_{12}$	
1-2-1, 3-3	$R_{12}R_{21}R_{33}$	

To this list of product terms we add the term, 1, which we may consider as the fictitious transfer function of a loop consisting of no branches.

Then D(z) is the sum of all signed products of the above type. The sign is to be taken positive if the product term is associated with an even number of non-touching loops. If the number of non-touching loops is odd, the minus sign is taken. In the example,

$$D(z) = 1 - R_{33} - R_{12}R_{21} - R_{23}R_{31}R_{12} + R_{12}R_{21}R_{33}$$

The numerator N(z) of $H_{mn}(z)$ is likewise, a sum of products of the R_{ij}. This sum is to be determined as follows. All possible pathways from node m to node n are considered, where a pathway is not to contain any closed loop (i.e. a node once visited is not to be revisited). The product of R_{ij} along this pathway is recorded. The example yields,

2-1	R_{21}
2-3-1	$R_{23}R_{31}$

To form N(z) we multiply each of the pathway products thus obtained by a sum of loop products

(signed) obtained as in the determination of D(z) except that in each case we only use closed loops which do not touch the corresponding pathway from m to n. The fictitious loop contributing the 1 factor is not considered to touch any pathway.

Thus in the example, the pathway 2-3-1 touches every loop, and pathway 2-1 touches every loop except 3-3. Thus:

$$N(z) = R_{21}(1 - R_{33}) + R_{23}R_{31}(1)$$

The end result is

$$H_{21}(z) = \frac{R_{21}(1 - R_{33}) + R_{23}R_{31}}{1 - R_{33} - R_{12}R_{21} - R_{23}R_{31}R_{12} + R_{12}R_{21}R_{33}}$$

It may sometimes be possible to reduce the fraction N/D to lowest terms by removing a common factor.

Proof

The proof of the above method will be based on the properties of determinants. It's more direct than the proof originally given in Mason's paper.

Since we want to find $H_{mn}(z)$, we apply a unit sample to node m at t=0 and observe the output at node n. The transform of this response is the required $H_{mn}(z)$. In addition, let's consider the responses at all other nodes. For the sake of definiteness we assume that there are a total of N nodes and number the nodes so that m=1. Then we consider $H_{1n}(z)$.

The following set of equations expresses the transform domain relationships among the $H_{1n}(z)$ signals. One equation is written for each node.

$$H_{11}(z) = H_{11}R_{11} + H_{12}R_{21} + H_{13}R_{31} + \cdots + H_{1N}R_{N1} + 1$$

$$H_{12}(z) = H_{11}R_{12} + H_{12}R_{22} + H_{13}R_{32} + \cdots + H_{1N}R_{N2}$$

. . .

$$H_{1N}(z) = H_{11}R_{1N} + H_{12}R_{2N} + H_{13}R_{3N} + \cdots + H_{1N}R_{NN}$$

These equations can be rearranged to show that they are a set of N linear algebraic equations in the N variables, $H_{11}, \ldots, H_{1N}$.

$$1 = (1 - R_{11})H_{11} - R_{21}H_{12} - \cdots - R_{N1}H_{1N}$$

$$0 = -R_{12}H_{11} + (1 - R_{21})H_{12} - \cdots - R_{N2}H_{1N}$$

. . .

$$0 = -R_{1N}H_{11} - R_{21}H_{12} - \cdots + (1 - R_{NN})H_{1N}$$

The solution for H_{1n} is given by Cramer's rule as the ratio of determinants. We want to show these determinants, $-\Delta_n$ and Δ are just the numerator and denominator, N(z) and D(z) used in the above method.

We have $H_{1n} = -\Delta_n / \Delta$, where

$$\Delta = \begin{vmatrix} 1 - R_{11} & -R_{12} & \cdots & -R_{1N} \\ -R_{21} & 1 - R_{22} & \cdots & -R_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ -R_{N1} & -R_{N2} & \cdots & 1 - R_{NN} \end{vmatrix}$$

and Δ_n is the cofactor of the element in the first column and nth row of Δ. (Note that we have used the transpose of the matrix of original coefficients. The determinant values, of course, are not affected.)

Now it is evident that Δ is the sum of products of the R_{ij}'s. Each such product is formed by selecting elements of Δ such that one and only one R_{ij} is taken from each row and column except that for an element on the main diagonal we may choose the 1 value instead of the R_{ij}. It's clear that we must form products of the form $R_{ij}\, R_{kl}\, R_{mn}\, R_{op} \ldots$ where the indices i, k, m, o, … are all different and where the indices j, l, n, p, … are all different.

But such a product is the product of R_{ij} around one or more non-touching closed loops as described in the method above. We see this as follows. First of all, no R_{ij} can appear more than once in the product. This follows from the rule of formation of Δ and means that every product corresponds uniquely to a prescribed pathway in the graph. This pathway can't have loose ends, i.e., every node along the pathway must be led into and out of by some pathway branch. Otherwise, there would be an R_{ij}, $i \neq j$ in the product without a corresponding R_{jk} or R_{li}. This situation is impossible for according to the rules for determining terms of Δ, if R_{ij} is selected and $i \neq j$ then we must sti11 select some element in the remaining rows and columns; and since $i \neq j$, we must take an element from

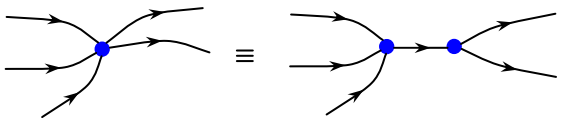
the j'th row and one from the i'th column. Also since i≠j these elements can't be the l's on the diagonal but must be of form R_{jk} or R_{1i}.

We have just shown that each product term of Δ corresponds to R_{ij}'s along closed pathways. To show that these pathways are simple non-touching loops we note that if two loops touch, then there are two pathway branches leading into some node. This means that the product of the R_{ij}'s contains two R_{ij}'s with the same second index. By the method of forming Δ this is clearly impossible. Hence Δ provides the same type of terms as the procedure we have given for D(z).

We must now show that, like D(z), Δ contains all products of R_{ij}'s along non-touching loops. This is seen to be true since any set of non-touching loops according to the above remarks yields a set of i, j indices corresponding to a legitimate selection of R_{ij} values to form a term of Δ. Now we only need to note that for any indices that don't appear we select the corresponding diagonal 1 values and thereby avoid using any R_{ij} with these indices. By using al1 l's down the main diagonal, we obtain the extra 1 term in D(z).

The final detail of proving Δ=D(z) is to show that the signs of the products incorporated into Δ are the same as those in D(z). The signs of terms in Δ are determined by counting permutations from natural order of the indices i, k, m, o, ... and j, 1, n, p, … corresponding to the product $R_{ij}\ R_{kl}\ R_{mn}\ R_{op}$ … . Since the R_{ij} elements of the determinant are prefixed with a minus sign the rule is that the sign of the product is + if the sum of i, k, m, o, ... permutations, the j, 1, n, p,... permutations and the number of terms in the product is an even integer. Otherwise the sign is –. Now since the product is formed from a number of non-touching loop R_{ij}'s let's consider the R_{ij}'s from only one such loop. This partial product has the form $R_{ij}\ R_{jk}\ R_{kl}$ … $R_{mn}\ R_{ni}$. So the sets of indices to consider are i, j, k, l, … m, n and j, k, 1, ... m, n, i. We see that the number of permutations in the first sequence differs from that in the second by the number of terms less one since by shifting the first index, i, in the first set through this number of positions we obtain the second set. Hence such a loop sequence of R_{ij}'s contributes an odd number toward the final sign determination in which all loops entering into the product are counted. Therefore, the final sign will be + if the number of non-touching loops is even and – if odd. This rule is exactly that used in forming D(z). We have now shown that Δ(z)=D(z).

Now we prove that $-\Delta_n(z)=N(z)$. Δ_n can be obtained in the same way as Δ. In fact, the determinant Δ will reduce to the cofactor Δ_n by a suitable replacement of the R_{ij} in the nth row by new values so that value of the element in the first column is 1 and in succeeding columns 0, For convenience we may assume that n≠l, i.e., that the input and output nodes are not the same. For if they are the same we may, as a preliminary step to the entire problem, break apart this node, separating it into two nodes one receiving all of the incoming branches and the other the outgoing branches and use a branch having a unity transfer function to connect the two new nodes.



To accomplish the reduction of Δ to Δ_n, then, we take $R_{nl}=1$, $R_{nn}=1$, $R_{nj}=0$ for j≠l or n. In system terms we have replaced our original graph with a new one in which all original connections from the output node have been severed, where a return branch with $R_{nl}=1$ has been added from output to input, and where a feedback loop with $R_{nn}=1$ has been attached to the output node.

By this artifice the products representing direct pathways from input to output as explained in the determination of N(z) now arise in Δ_n from the loops that have been closed by adding $R_{nl}=1$. The numerical value of a product is not disturbed except that the inclusion of the extra loop changes the sign of all products.

The addition of $R_{nn}=1$ has the following effect. If a set of loops does not touch a loop completed by $R_{nl}=1$ as explained above, then it does not touch the R_{nn} loop either. In forming products of such loop R_{ij} there will be one product without R_{nn} and one product with R_{nn}. But these terms cancel because they differ only in sign (the use of R_{nn} adds another loop.) So, no products survive in the result unless they contain as a factor the R_{ij} product along a loop completed by $R_{nl}=1$. (This behavior also results in the suppression of the 1 term in Δ_n.)

Now if we factor out these products we are left with a form of Δ_n which is identical to N(z) except for sign since the loop countings differ by one. Thus $-\Delta_n=N(z)$ and the method is proved.

Appendix III. An efficient sum of squares method

In section 7.5 we showed that the sum of squares of the response values was a suitable measure for evaluating errors in a control system response to a transient input. Again in section 13.5 we showed that the problem of evaluating mean-square errors with random inputs was equivalent to the calculation of sums of squares of the values of some function.

There were three methods outlined in these sections for the computation of sums of squares of the response, g(n) for n=0, 1, 2,… .

$$\sum_{n=0}^{\infty} g^2(n) = \left\langle coefficient\ of\ \ z^0\ in\ G(z)G(z^{-1}) \right\rangle$$

Direct calculation of $g^2(0)+g^2(1)+\ldots$ and summation of the series in closed form by inspection.

Derivation of difference equations involving $g^2(n)$ instead of g(n) as an output function and evaluation of the sum by manipulating these equations.

Method 2 can only be used in the simplest examples. Method 3 leads to the evaluation of large determinants and is rather clumsy. Method 1, using a partial fraction expansion to obtain the required central sample also leads to involved calculations.

We will now derive a more efficient procedure for evaluating sums of squares based on method 1. We have not given the method before, because its derivation is quite formal and difficult to motivate.

Suppose that G(z) is the rational transform of g(n), a signal which is zero for n<0.

$$G(z) = \sum_{n=0}^{\infty} z^n g(n)$$

Then

$$G(z)G(z^{-1}) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} z^{n-k} g(n)g(k)$$

$$(1) \qquad = \sum_{n=0}^{\infty} g^2(n) + (z+z^{-1})\sum_{n=0}^{\infty} g(n)g(n+1) + \cdots$$

Denoting $G(z)G(z^{-1})$ by Q(z), we note that $Q(z)=Q(z^{-1})$. Then by proper factorization and partial fraction expansion, Q(z) could be written in the forms,

$$(2) \qquad Q(z) = \frac{\sum_{k=-M}^{+M} b_k z^k}{\left(\sum_{k=0}^{N} a_k z^k\right)\left(\sum_{k=0}^{N} a_k z^{-k}\right)}$$

$$(3) \qquad = \sum_{k=0}^{N}\left(\frac{A_k}{1-\lambda_k z} - \tfrac{1}{2}A_k + \frac{A_k}{1-\lambda_k z^{-1}} - \tfrac{1}{2}A_k\right)$$

where a_k, b_k, A_k, λ_k are constants, k, N, M are integers and $b_k=b_{-k}$.

Expanding the terms of (3) by long division, we get

$$Q(z) = \sum_{k=0}^{N}\left(A_k + (z+z^{-1})\lambda_k A_k + \cdots\right)$$

Comparing with (1), we see that the desired sum of squares, which we denote by E_0 is

$$(4) \qquad E_0 = \sum_{n=0}^{\infty} g^2(n) = \sum_{k=0}^{N} A_k$$

Now $\sum_{k=0}^{N} a_k z^k$ contains the factor $(1-\lambda_k z)$; so we can write

$$(5) \qquad \frac{\sum_{k=0}^{N} a_k z^k}{1-\lambda_k z} = \sum_{r=0}^{N-1} B_{rk} z^r$$

In particular we have $B_{0k}=a_0$. A similar expression follows by replacing z by z^{-1}.

We assume that M≤N. If this condition does not hold originally, Q(z) can be modified by subtracting a polynomial of the form $c_0+c_1(z+z^{-1})+\ldots+c_j(z^j+z^{-j})$ such that the condition holds for the remaining Q(z). Then we add c_0 to the result of the sum-of-squares calculation to obtain the final answer.

In the succeeding operations we sum over all integer values of the indices, assuming that all coefficients outside the ranges indicated in (2), (3), and (5) vanish. Thus,

$$
\begin{aligned}
&b_k = 0 && |k| > M \\
&\left.\begin{aligned} a_k &= 0 \\ A_k &= 0 \\ \lambda_k &= 0 \end{aligned}\right\} && k > N \ \ or \ \ k < 0 \\
&B_{rk} = 0 && k > N \ \ or \ \ k < 0 \ \ or \\
& && r > N-1 \ \ or \ r < 0
\end{aligned}
\tag{6}
$$

Then multiplying (2) and (3) by the denominator of (2) and using (5) we obtain

$$
\sum_k b_k z^k = \sum_k A_k \left[\begin{array}{l} \sum_r B_{rk} z^r \sum_j a_j z^{-j} - \frac{1}{2} \sum_r a_r z^r \sum_j a_j z^{-j} \\ + \sum_r B_{rk} z^{-r} \sum_j a_j z^j - \frac{1}{2} \sum_r a_r z^{-r} \sum_j a_j z^j \end{array} \right]
$$

$$
= \sum_k A_k \left[\sum_r \sum_j (B_{rk} - \tfrac{1}{2} a_r) a_j z^{r-j} + \sum_r \sum_j (B_{rk} - \tfrac{1}{2} a_r) a_j z^{-r+j} \right]
$$

$$
= \sum_k A_k \left[\sum_r \sum_s (B_{rk} - \tfrac{1}{2} a_r) a_{r-s} (z^s + z^{-s}) \right]
$$

Comparing coefficients of like powers of z on both sides of this expression we obtain,

$$
b_s = \sum_k A_k \left[\sum_r (B_{rk} - \tfrac{1}{2} a_r)(a_{r+s} + a_{r-s}) \right] \tag{7}
$$

Denote by E_r the expression

$$
E_r = \frac{2}{a_0} \sum_k A_k (B_{rk} - \tfrac{1}{2} a_r) \tag{8}
$$

In particular,

$$
E_0 = \frac{2}{a_0} \sum_k A_k (a_0 - \tfrac{1}{2} a_0) = \sum_k A_k
$$

which checks with our previous definition of E_0.

Now (7) may be written,

$$
b_s = \frac{a_0}{2} \sum_r E_r (a_{r+s} + a_{r-s}) \tag{9}
$$

This is a system of linear equations (one for each value of s) to be solved for the E_r when the a and b coefficients are known. We are principally interested in just E_0. From (8) and (6) we see that E_r=0 for r<0 or r>N. Therefore, there are precisely N+1 nontrivial equations in N+1 unknowns in (9) which can be written in the matrix form.

$$
\frac{2}{a_0} [b_0, b_1, \cdots b_N] = [E_0, E_1, \cdots E_N] (a_{r+s} + a_{r-s})
$$

Using Cramer's rule to solve for for E_0 we obtain

$$
E_0 = \frac{2}{a_0} \cdot \frac{\left| \begin{array}{c} b_0 \ b_1 \cdots b_N \\ \hline a_{r+s} + a_{r-s} \end{array} \right|}{\left| a_{r+s} + a_{r-s} \right|} = \sum_{n=0}^{\infty} g^2(n)
$$