Introduction to
Feedback Control Systems

Lecture by *D Robinson* on *Feedback Control Systems* for *Biomedical Engineering*, Fall 1994
Re-Edited by T Haslwanter

1. Introduction

Your body is jammed full of feedback control systems. A few examples of things so controlled are: Blood pressure; blood volume; body temperature; circulating glucose levels (by insulin); blood partial pressures of carbon dioxide and oxygen (P_{CO2} , P_{O2}); pH; hormone levels; thousands of proteins that keep cells alive and functioning. Without these control systems, life would not be possible. These are just a few examples. Open a physiology textbook at random and you will be reading about some part of some control system. (As a challenge, try to think of a process that is not under feedback control.)

It would be impossible to understand biological systems without using systems analysis and understanding negative feedback.

1.1. Block Diagrams

The first step in analyzing any system is to lay it out in a block diagram. Here is an example:

The Baroreceptor Reflex:

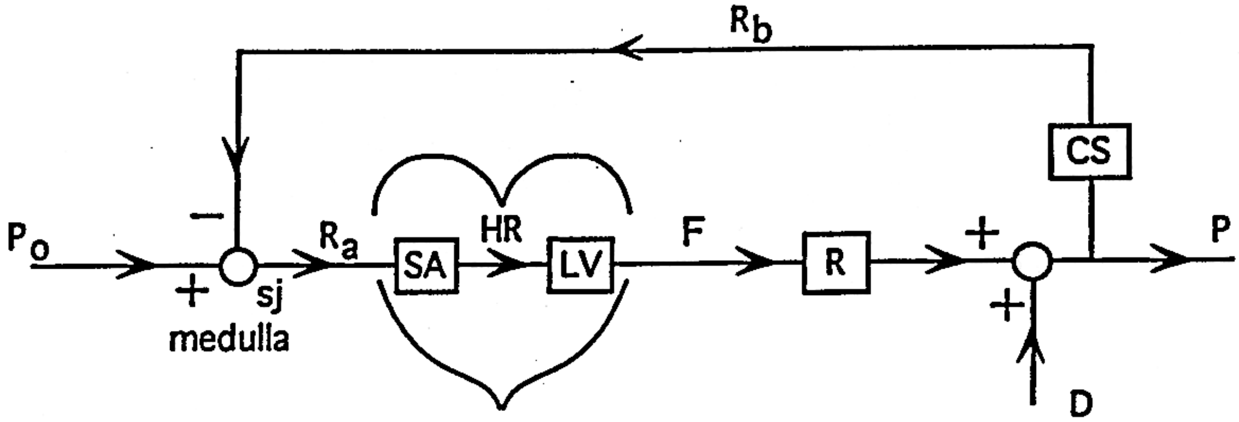


Figure 1

Variables or signals (e.g., *HR*, *F*, *P*) flow along the lines from box to box. The boxes or blocks represent physical processes (a muscle, a gland, a nerve, an electric circuit, a motor, your furnace, etc.) that receives an input signal, does something to it, and produces an output signal.

Conventionally, signals along the main path flow from left to right so at the right is the final output: the variable being controlled is, in this case, arterial blood pressure, *P*. On the left is the input, desired blood pressure, P_o, that drives the whole system.

CS: Blood pressure, *P*, is sensed by the brain by ingenious transducers called *baroreceptors* located in the carotid sinus (*CS*). These are nerve cells which discharge with action potentials all the time. When P goes up, the walls of the *CS* are stretched, depolarizing the branches of these nerve cells so they discharge at a faster rate, R_b. Thus R_b reports to the brain what *P* is.
Obviously if one is to control anything, you first have to measure it.

Medulla: This is the part of the brainstem to which R_b is sent. Its cells have a built-in discharge rate we call P_o since it represents the desired blood pressure. The total system wants to keep *P* equal to P_o (e.g., 100 mm Hg). P_o is called a *setpoint*. These cells are represented by a summing junction (sj), which shows that their discharge rate R_a ("a" for autonomic nervous system) is the difference P_o- R_b. Since R_b represents P, R_a reflects P_o - P. If $P \neq P_o$, there is an error, so R_a is an error signal.

SA: A specialized patch in the heart called the *sino-atrial node* is your pacemaker. It is a clock that initiates each heartbeat. The autonomic nervous system (R_a) can make the heart rate, *HR*, speed up or slow down.

LV: The left ventricle is a pump. With each beat it ejects a volume of blood called the *stroke volume*. The output flow F is a product:

stroke volume [ml/beat] x *HR* [beats/sec] = F [ml/sec].

R: The resistance of the vascular beds (arteries, capillaries, veins) relates pressure to flow. $P=RF$ is the Ohm's law of hydraulics ($E=RI$).

D: If nothing ever went wrong, P would always equal P_o and there would be no need for feedback. But that is not the case. There are many things that will disturb (D) P. Exercise is the obvious example; it causes blood vessels in muscle to open up, which would cause P to drop to dangerous levels. Thus D represents anything that would change P.

In words - if P drops for some reason, R_b decreases, R_a increases as does *HR* and F, thus increasing P back towards its original value. This is an obvious example of how negative feedback works and why it is useful.

Obviously, one must know the anatomy of a system, but an anatomical diagram is not a substitute for a block diagram, because it does not show the variables which are essential for analysis.

Most other biological control systems can be laid out in the form of Fig. 1 for further analysis. What follows is how to do that analysis and what it tells us. A basic concept for each block is the ratio of the output signal to the input, which is loosely called the *gain*. If, for example, *HR* doubles, so will F. Consequently, *F/HR* is more or less a constant, at least in steady state, and is called the *gain of that box*. The gain of an electronic amplifier is *(output voltage)/(input voltage)* and is dimensionless, but that's a special case. In general the gains of boxes can be anything. The gain of *CS* in Fig. 1 is *(spikes /sec)/mmHg*. The gain of *SA* is *(beats/sec)/(spikes/sec)* and so on. We will go into this in more detail later but get used to the general idea: *gain = output/input*.

1.2. What Good is Negative feedback?

Mother Nature discovered negative feedback millions of years ago and obviously found it to be GOOD (we discovered it in the 1930s in the labs of Ma Bell). So what's so good about it?

Let's take Fig. 1 and reduce it to its bare, universal form. Let the gains of *SA*, *LV* and *R* be multiplied and call the net result G. Let the feedback gain R_b/P be 1.0 for this example.

So a reduced version of Fig. 1 is:

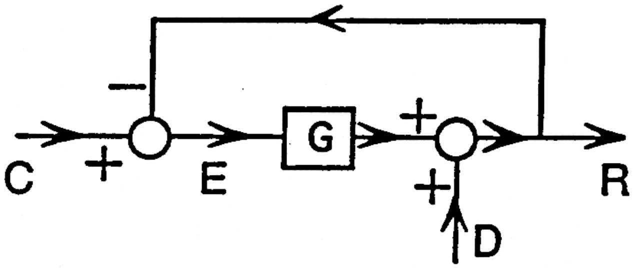


Figure 2

where

C is a command input
R is a response or output
D is a disturbance
E is the error, *C-R*
G is the net forward gain, sometimes called the *open loop gain* because if you removed the feedback path (which, with no block, has an implied gain of 1.0), the open loop gain *R/C* would be *G*, usually much larger than 1.0.

We want to find the output in terms of the inputs C and D:

On the left, E=C-R
Eliminate E;

On the right R=D + GE

$$R = D + G \cdot (C - R)$$

or

$$R \cdot (1 + G) = D + GC$$

or

$$R = \frac{1}{1+G} \cdot D + \frac{G}{1+G} \cdot C \tag{1}$$

This equation says it all (it's recommended you memorize it). Any disturbance *D* is attenuated by *1/(1 + G)*. If *G* is 100, then only 1% of *D* is allowed to affect *R*. Even if *G* is only 10, then only 10% of *D* gets through, which is a lot better than no protection at all. In words, if *D* should cause *R* to decrease, an error *E* is immediately created. That is amplified by *G* to create an output of *G* opposite and nearly equal to *D*, thereby returning *R* close to its original value. (This protection is so important for living systems, Mother Nature adopted it when we were just single cell animals floating in the sea and has used it ever since.)

But external disturbances are not the only source of problems. Suppose the parameters of the system change. Consider a system that does not use negative feedback:

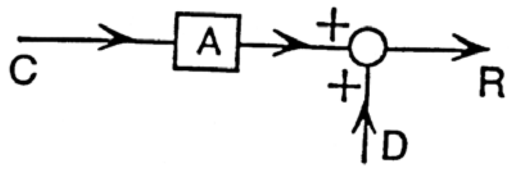


Figure 2a

$$R = D + AC$$

To make this comparable to Fig. 2, let the nominal value of *A* be 1.0.

As we already can see, *R* is affected 100% by *D*. There is no protection against external disturbances. But what if *A* gets sick and its gain dropped by 50%? Then the output drops by 50%. No protection against changes in parameters (in Fig. 2a, the gain *A* is the only parameter, in any real system there would be many more parameters such as the *R's*, *L's* and *C's* of an electric circuit).

What happens with feedback? What happens if G changes in Fig. 2? From eq (1) (let $D = 0$ for now)

$$\frac{R}{C} = \frac{G}{1+G}$$

If G has a nominal value of 100, $R/C = 100/101 = 0.99$, close to the desired closed-loop gain of 1.00. Now let G drop by 50%. Then $R/C = 50/51 = 0.98$. So a 50% drop in G caused only a 1% drop in closed-loop performance, R/C. Even if G dropped by a factor of 10, to a gain of 10, still $R/C = 10/11 = 0.91$, an 8% decrease. In general, so long as $G > 1.0$, then $G/(1+G)$ will be close to 1.0.

To repeat

$$R = \left(\frac{1}{1+G}\right)\cdot D + \left(\frac{G}{1+G}\right)\cdot C$$

Feedback protects system against external disturbances

Feedback protects system against changes in internal parameters

This is why Mother Nature uses NEGATIVE FEEDBACK.

Summary

1. You can't study biological systems without systems analysis.
2. The first thing to do is draw the block diagram showing the negative feedback path explicitly.
3. And now you know why negative feedback is used again and again in biological (or any other) systems.

So far, we have not needed to worry about dynamics (differential equations), just algebra because that's all you need for the main message. But feedback greatly affects the dynamics of a system. For example, feedback can cause a system to oscillate. So from here on, we study system dynamics.

2. Reminder & A Little History

You have probably already learned about Laplace transforms and Bode diagrams. These tools are essential for systems analysis so we review what is relevant here, with some historical comments thrown in to show you where Laplace transforms came from.

You know that, given any time course f(t), its Laplace transform F(s) is,

$$F(s) = \int_0^\infty f(t) \cdot e^{-st} dt \tag{2a}$$

while its inverse is,

$$f(t) = \int_S F(s) \cdot e^{st} ds \tag{2b}$$

where the integration is over the s-plane, *S*. From this you should recall a variety of input transforms and their output inverses. If, for example, the input was a unit impulse function, $f(t) = \delta(t)$, then $F(s) = 1$. If an output function *F(s)* in Laplace transform was $\dfrac{A}{(s+\alpha)}$, then its inverse, back in the time domain, was $f(t) = Ae^{-\alpha t}$.
Using these transforms and their inverses can allow you to calculate the output of a system for any input, but without knowing where eqs 2 came from, it is called the *cook book method*.

I have a hard time telling anyone to use a method without knowing where it came from. So the following is a brief historical tale of what led up to the Laplace transform - how it came to be. You are not responsible for the historical developments that follow but you are expected to know the methods of Laplace transforms.

2.1. Sine Waves

Sine waves almost never occur in biological systems but physiologists still use them experimentally. This is mostly because they wish to emulate systems engineering and use its trappings. But why do systems engineers use sine waves? They got started because the rotating machinery that generated electricity in the late 19th century produced sine waves. This is still what you get from your household outlet. This led, in the early 20th century, to methods of analyzing circuits with impedances due to *resistors* (R), *capacitors* ($1/j\omega C$) and *inductors* ($j\omega L$) - all based on sine waves.[1]

These methods showed one big advantage of thinking in terms of sine waves: it is very easy to describe what a linear system does to a sine wave input. The output is another sine wave of the same frequency, with a different amplitude and a phase shift. The ratio of the output amplitude to the input amplitude is the *gain*. Consequently, the transfer function of a system, at a given frequency, can be described by just two numbers: *gain* and *phase*.

[1] Note that in engineering, $\sqrt{-1} = j$. In other fields it is often denoted by *"i"*.

Example:

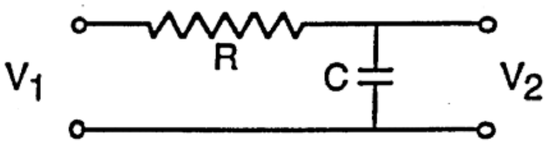


Figure 3

Elementary circuit analysis tells us that the ratio of the output to the input sine wave is

$$\frac{V_2(j\omega)}{V_1(j\omega)} = \frac{\dfrac{1}{j\omega C}}{R + \dfrac{1}{j\omega C}} = \frac{1}{j\omega RC + 1} = G(j\omega) \tag{3}$$

(If this isn't clear, you had better check out an introduction to electronic circuits). Note we are using complex notation and what is sometimes called *phasors*. Through Euler's identity

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t) \tag{4}$$

we can use the more compact form $e^{j\omega t}$ for the input instead of the clumsy forms, cosine and sine. The only purpose is to keep the arithmetic as simple as possible.

Now we want to *generalize* and get away from electric circuits. As we shall see, many other systems, mechanical and chemical as well as electrical, have similar transfer functions $G(j\omega)$. That is why we replace V_2/V_1 with G to generalize to all such other systems. We can even replace RC with the *time constant T* since it is the more intrinsic parameter characterizing this class of systems. So, for this example, we think in terms of:

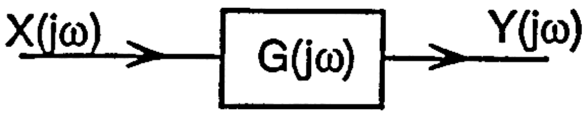


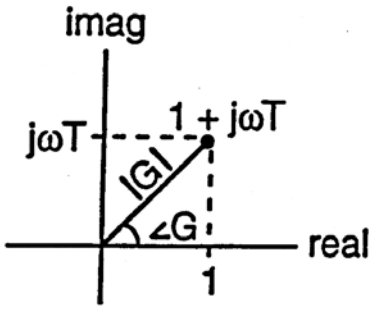
Figure 4

where

$$G(j\omega) = \frac{1}{j\omega T + 1} \tag{5}$$

and X and Y can be any of a wide variety of variables as illustrated in Fig. 1.

This type of system is called a *first-order lag* because the differential equation that describes it is first-order and because it creates a phase lag.



From (5), the gain is

$$|G| = \frac{1}{|j\omega T + 1|} = \frac{1}{\sqrt{(\omega T)^2 + 1}} \tag{6}$$

The phase, or angle, of G, is

$$\angle G = -\tan^{-1}\omega T \tag{7}$$

For any frequency ω, these two equations tell you what the first-order lag does to a sine wave input.

2.2. Bode Diagram

There is an insightful way of displaying the information contained in eqs (6) and (7):

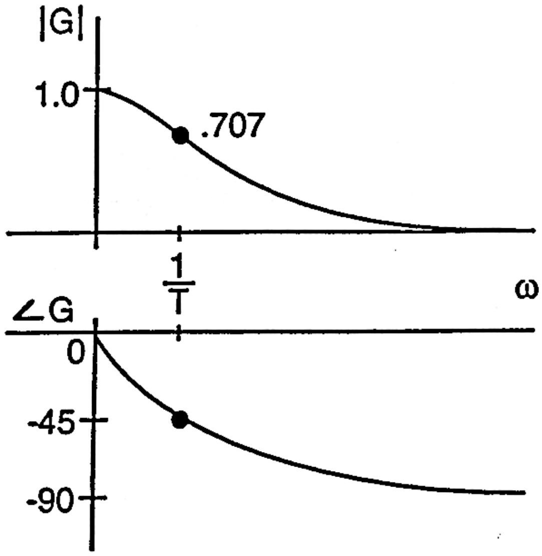


Figure 5

If you plot them on a linear-linear scale they look like this:

You can see that as frequency $\omega = (2\pi f)$ goes up, the gain goes down, approaching zero, and the phase lag increases to -90°.

Bode decided to plot the gain on a *log-log* plot. The reason is that if you have two blocks or transfer functions in cascade, *G(jω)* and *H(jω)*, the Bode plot of their product is just the sum of each individually; that is,

$$\log(G(j\omega) \times H(j\omega)) = \log(G(j\omega)) + \log(H(j\omega)$$

Also by doing this you stretch out the low-frequency part of the ω axis and get a clearer view of the system's frequency behavior. The result will appear:

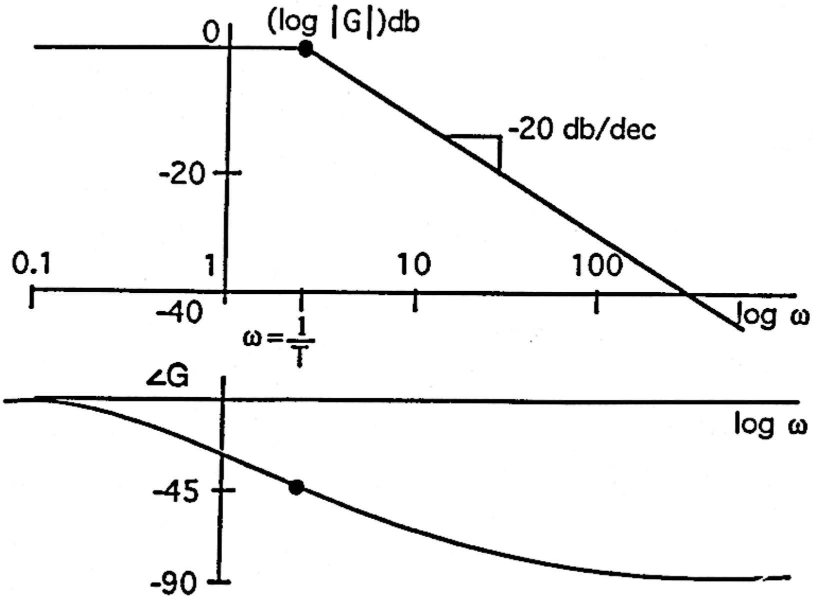


Figure 6

An interesting frequency is $\omega = \frac{1}{T}$, since then, from (6), $|G| = 1/\sqrt{2} = 0.707$ and $\angle G = -\tan^{-1}(1) = -45^\circ$.
Below this frequency, *log(|G|)* can be closely approximated by a horizontal straight line at zero (*log(1)=0*). At high frequencies, above $\omega = 1/T$, $|G|$ falls off in another straight line with a slope of -20db/dec, which means it falls by 10 (20db) if ω increases by 10. The phase is a linear-log plot. The main point in Fig. 6 is a very simple way to portray what this lag element does to sine wave inputs ($e^{j\omega t}$) of any frequency.

|G| is often measured in *decibels*, defined as:

$$decibels = db = 20\ log(|G|)$$

The *20* is *10 × 2*. The *10* is because these are *decibels*, not *bels* (in honor of Alexander Graham), and the *2* is because this was originally defined as a power ratio as in $(V_2 / V_1)^2$.)
The next problem is: suppose the input isn't a sine wave? Like speech signals and radar.

2.3. Fourier Series

Well, a *periodic signal*, like a square wave or a triangular wave, would be a step up to something more interesting than just a sine wave. Fourier series is a way of approximating most *periodic signals as a sum of sine waves.*

The sine waves are all *harmonics* of the *fundamental frequency* ω_0. Specifically, the periodic wave form *f(t)* can be written

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_0 t) \tag{8}$$

The problem is only to find the right constants a_0, a_n, b_n. They can be found from

$$\begin{aligned} a_0 &= \frac{1}{T}\int_0^T f(t)dt \\ a_k &= \frac{2}{T}\int_0^T f(t)\cos(k\omega_0 t)\,dt \\ b_k &= \frac{2}{T}\int_0^T f(t)\sin(k\omega_0 t)\,dt \end{aligned} \tag{9}$$

where *T* is the period of *f(t)* so that the frequency $\omega_0 = \frac{2\pi}{T}$. The square wave is a common example (Fig.7):

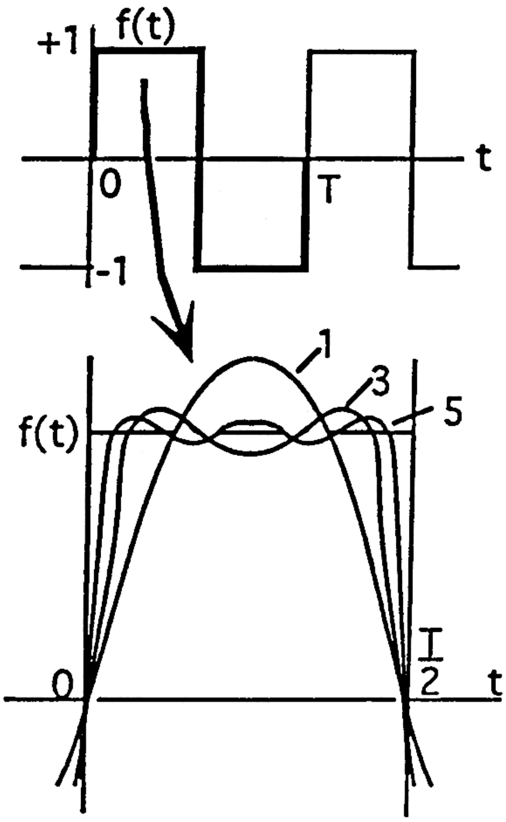


Figure 7

If you plug this into (9) you get

$$f(t) = \frac{4}{\pi}[\sin \omega t + \frac{1}{3}\sin 3\omega t + \frac{1}{5}\sin 5\omega t + \cdots]$$

This figure also illustrates how adding the 3rd and 5th harmonics to the fundamental frequency fills out the sharp corners and flattens the top, thus approaching a square wave.

So how does all this help us to figure out the output of a transfer function if the input is a square wave or some other periodic function? To see this, we must recall *superposition*.

2.4. Superposition

All *linear systems* obey *superposition*. A linear system is one described by linear equations. $y = kx$ is linear (k is a constant). $Y = sin(x)$, $y = x^2$, $y = log(x)$ are obviously not. Even $y = x+k$ is not linear; one consequence of superposition is that if you double the input (x), the output (y) must double and, here, it doesn't. The differential equation

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = y$$

is linear.

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt}x + cx^2 = y$$

is not, for two reasons which I hope are obvious.

The definitive test is that if input $x_1(t)$ produces output $y_1(t)$ and $x_2(t)$ produces $y_2(t)$, then the input $ax_1(t)+bx_2(t)$ must produce the output $ay_1(t)+by_2(t)$. This is *superposition* and is a property of *linear systems* (which is only what we are dealing with).

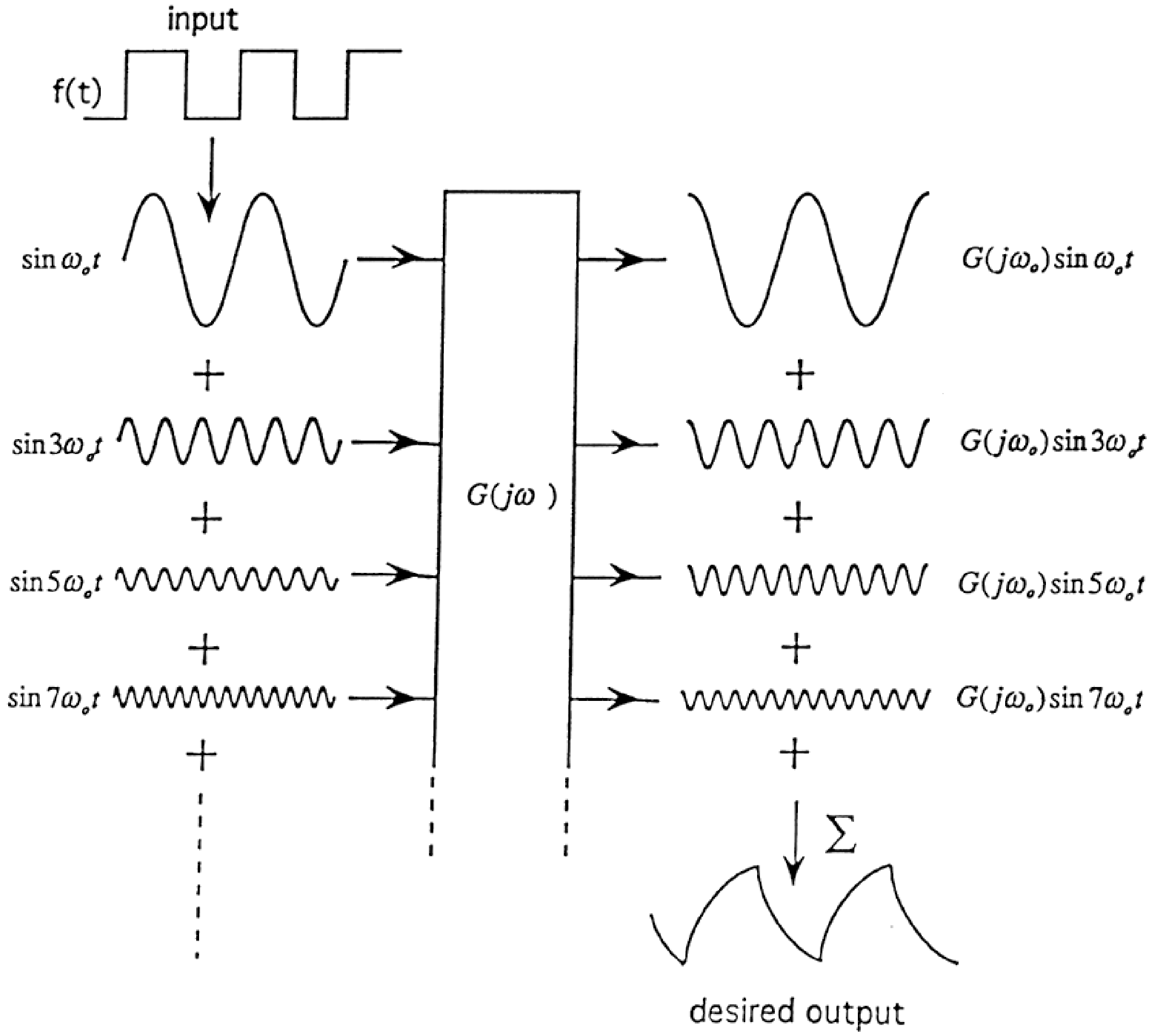


Figure 8

So, if we can *break down f(t)* into a bunch of *sine waves*, the Bode diagram can quickly give us the gain and phase for each harmonic. This will give us all the *output sine waves* and all we have to do is *add them all up* and - voila! -the desired output!

This is illustrated in Fig. 8. The input *f(t)* (the square wave is only for illustration) is *decomposed* into the sum of a lot of *harmonics* on the left using (9) to find their amplitudes. Each is passed through $G(j\omega)$. $G(jk\omega_0)$ has a *gain* and a *phase shift* which, if $G(j\omega)$ is a first-order lag, can be calculated from (6) and (7) or read off the Bode diagram in Fig. 6. The resulting sinusoids $G(jk\omega_0)$ sin $k\omega_0 t$ can then all be added up as on the right to produce the final desired output shown at lower right.

Fig. 8 illustrates the basic method of all transforms including Laplace transforms so it is important to understand the concept (if not the details). In different words, *f(t)* is taken from the *time domain* by the *transform* into the *frequency domain*. There, the *system's transfer function operates* on the frequency components to produce *output components* still in the *frequency domain*. The *inverse transform assembles* those components and *converts* the result back into the *time domain,* which is where you want your answer. Obviously you couldn't do this without *linearity* and *superposition*. For Fourier series, eq (9) is the transform, (8) is the inverse.

One might object that dealing with an infinite sum of sine waves could be tedious. Even with the rule of thumb that the first 10 harmonics is good enough for most purposes, the arithmetic would be daunting. Of course with modern digital computers, the realization would be quite easy. But before computers, the scheme in Fig. 8 was more conceptual than practical.

But it didn't matter because Fourier series led quickly to Fourier transforms. After all, periodic functions are pretty limited in practical applications where *aperiodic signals* are more common. Fourier transforms can deal with them.

2.5. Fourier Transforms

These transforms rely heavily on the use of the exponential form of sine waves: $e^{j\omega t}$. Recall that (eq 4),

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$$

If you write the same equation for $e^{-j\omega t}$ and then add and subtract the two you get the inverses:

$$\begin{aligned} \cos(n\omega_0 t) &= \frac{e^{jn\omega_0 t} + e^{-jn\omega_0 t}}{2} \\ \sin(n\omega_0 t) &= \frac{e^{jn\omega_0 t} - e^{-jn\omega_0 t}}{2} \end{aligned} \tag{10}$$

where we have used harmonics $n\omega_0$ for ω.

It is easiest to derive the Fourier transform from the Fourier series, but first we have to put the Fourier series in its complex form. If you now go back to (8) and (9) and plug in (10) you will eventually get

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \tag{11}$$

and

$$c_n = \frac{1}{T}\int_0^T f(t)e^{-jn\omega_0 t}\,dt \tag{12}$$

(The derivation can be found in textbooks). These are the inverse transform and transform respectively for the *Fourier series* in *complex notation*. The use of (10) introduces negative frequencies ($e^{-j\omega t}$) but they are just a mathematical convenience. It will turn out in (11), that when all the positive and negative frequency terms are combined, you are left with only real functions of positive frequencies. Again, the reason for using (11) and (12) instead of (8) and (9) is, as you can easily see, mathematical compactness.

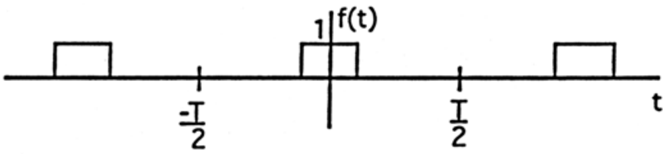


Figure 9

To get to the Fourier transform from here, do the obvious, as shown in Fig. 9. We take a rectangular pulse for *f(t)* (only for purposes of illustration). *f(t)* is a periodic function with a period of *T* (we've chosen *-T/2* to *T/2* instead of 0 to T for simplicity). Now keep the rectangular pulse constant and *let T get larger and larger*. What happens in (11) and (12)? Well, the difference between harmonics, ω_0, is getting smaller. Recall that $\omega_0 = \frac{2\pi}{T}$ (*frequency = inverse of period*), so in the limit $\omega_0 \Rightarrow d\omega$. Thus,

$$\frac{1}{T} = \frac{d\omega}{2\pi}$$

The harmonic frequencies $n\omega_0$ merge into the continuous variable ω,

$n\omega_0 \Rightarrow \omega$ From (12), as $T \Rightarrow \infty$, c_n would => 0 but their product Tc_n does not and it is called *F(ω)* or the *Fourier transform*. Making these substitutions in (11) and (12) gives

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \tag{13}$$

and

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{14}$$

These are the *transform* (14) and its *inverse* (13). See the subsequent pages for examples. *F(ω)* is the spectrum of *f(t)*. It shows, if you plot it out, the frequency ranges in which the energy in *f(t)* lies, so it's very useful in speech analysis, radio engineering and music reproduction.

In terms of Fig. 8, everything is conceptually the same except we now have all possible frequencies - no more harmonics. That sounds even worse, computationally, but if we can express *F(ω)* mathematically, the integration in (13) can be performed to get us back into the time domain. Even if it can't, there are now *FFT* computer programs that have a *fast* method of finding the *Fourier transform*, so using this transform in practice is not difficult.

A minor problem is that if the area under *f(t)* is infinite, as in the unit step, *u(t)*, *F(ω)* can blow up. There are ways around this, but the simplest is to move on to the Laplace transform.

2.6. Laplace Transforms

You already know the formulas (2a, 2b), and how to use them. So here we discuss them in terms of Fig. 8 and superposition.

Again the inverse transform is

$$f(t) = \int_S F(s)e^{st}ds \tag{15}$$

Until now we have dealt only with sine waves, $e^{j\omega t}$. Put another way, we have restricted s to $j\omega$ so that e^{st} was restricted to $e^{j\omega t}$. But this is unnecessary, we can let s enjoy being fully complex or $s=\sigma+j\omega$. This greatly expands the kinds of functions that e^{st} can represent.

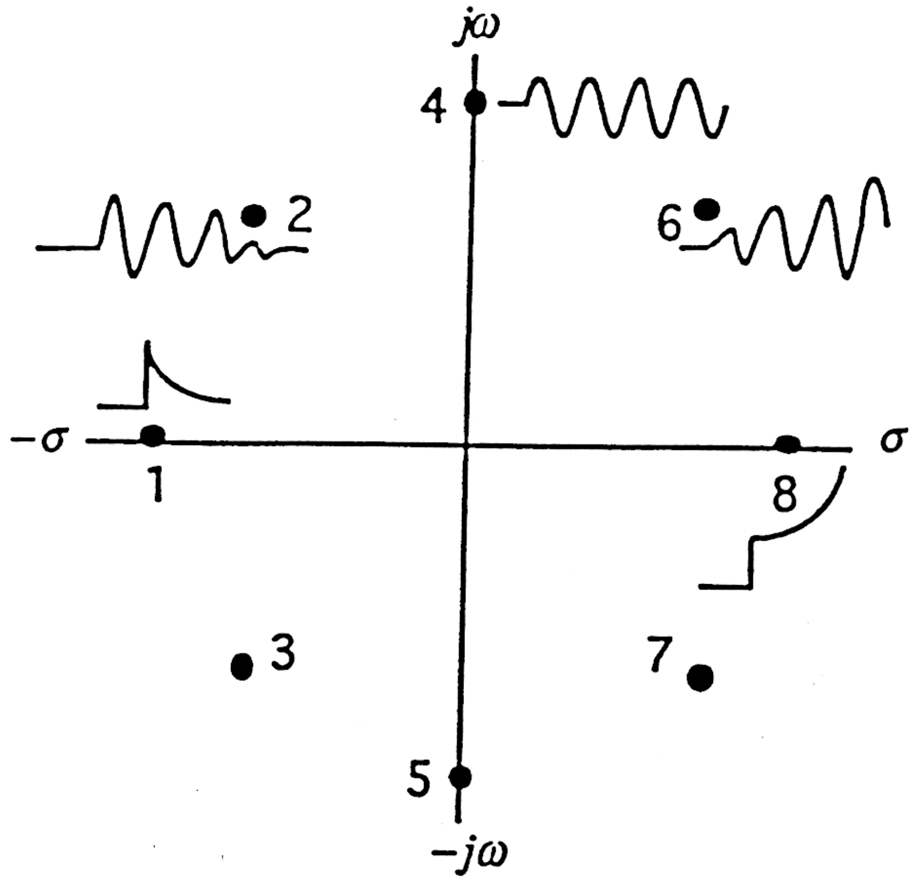


Figure 10

Fig. 10 is a view of the *s-plane* with its real axis (σ) and imaginary axis (jω). At point 1, ω=0 and -σ is negative so $e^{st}=e^{-\sigma t}$, which is a simple decaying exponential as shown. At points 2 and 3 (we must always consider pairs of complex points - recall from (10) that it took an $e^{j\omega t}$ and an $e^{-j\omega t}$ to get a real *sin ωt* or *cos ωt*) we have -σ<0 and $\omega \neq 0$, so $e^{-\sigma t}e^{j\omega t}$ is a damped sine wave as shown. At points 4 and 5, $\sigma=0$ so we are back to simple sine waves. At points 6 and 7, $\sigma > 0$ so the exponential is a rising oscillation. At 8, $\sigma > O$, $\omega=0$ so we have a plain rising exponential. So Fig. 10 shows the variety of waveforms represented by e^{st} .

So (15) says that *f(t)* is made up by summing an infinite number of infinitesimal wavelets of the forms shown in Fig. 10. *F(s)* tells you how much of each wavelet e^{st} is needed at each point on the *s-plane*. That weighting factor is given by the transform

$$F(s) = \int_{0}^{\infty} f(t)e^{-st}dt \tag{16}$$

In terms of Fig. 8, *f(t)* is *decomposed* into an infinite number of wavelets as shown in Fig. 10, each weighted by the complex number *F(s)*. They are then passed through the *transfer function G* which now is no longer *G(jω)* (defined only for sine waves) but *G(s)* defined for $e^{\sigma t}e^{j\omega t}$. The result of *F(s)G(s)* which tells you the *amount of* e^{st} at each point on the s-plane *contained in the output*. Using (15) on *F(s)G(s)* takes you back to the *time domain* and gives you the *output*. If, for example, the output is h(t) then

$$h(t) = \int_{S} F(s)G(s)e^{st}ds \tag{16a}$$

In summary, I have tried to show a logical progression form the *Fourier series* to the *Fourier transform* to the *Laplace transform*, each being able to deal with *more complicated waveforms*. Each method transforms the input time signal $f(t)$ into an infinite sum of infinitesimal wavelets in the frequency domain (defining s as a "frequency"). The transfer function of the system under study is expressed in that domain $G(j\omega)$ or $G(s)$. The frequency signals are passed through G, using superposition, and the outputs are all added up by the inverse transform to get back to $h(t)$ in the time domain.

This is the end of the historical review, and we are now going back to the cook-book method (which you already know – see p6).

2.7. Cook-Book Example

For a first order lag in Laplace, (5) becomes G(s),

$$G(s) = \frac{1}{sT+1} \tag{17}$$

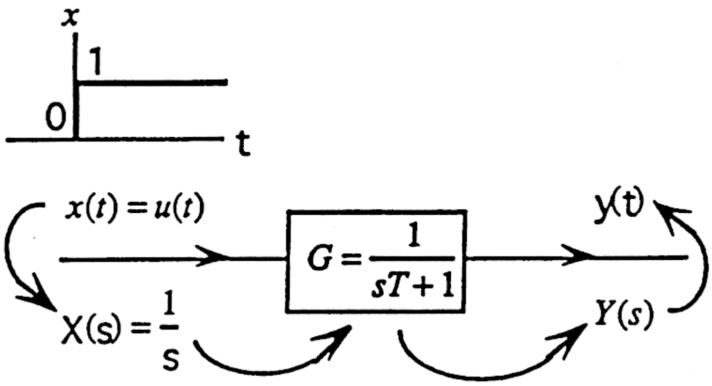


Figure 11

Let's find its step response. A unit step, $u(t)$ has a Laplace transform of $X(s)= 1/s$. So the Laplace transform of the output Y(s) is

$$Y(s) = \frac{1}{s(sT+1)} \tag{18}$$

So far this is easy but how to get from $Y(s)$ (the frequency domain) to $y(t)$ (the time domain)? We already know that:

(19)

f(t)	*F(s)*
$\delta(t)$	1
$U(t)$	$1/s$
t	$1/s^2$
e^{-at}	$1/(s+a)$
$\sin(\omega t)$	$\dfrac{\omega}{s^2+\omega^2}$
etc.	*etc*

So here is the cook-book; if we could rearrange $Y(s)$ so that it could contains forms found in the $F(s)$ column, it would be simple to invert those forms. But that's easy; it's called: *Partial Fraction Expansion*. $Y(s)$ can be rewritten

$$Y(s)=\frac{1}{s(sT+1)}=\frac{1}{s}-\frac{T}{(sT+1)}$$

Well that's close and if we rewrite it so,

$$Y(s)=\frac{1}{s}-\frac{1}{s+\frac{1}{T}} \tag{20}$$

then, from the Table (19)

$$Y(t)=u(t)-e^{-\frac{t}{T}}=1-e^{-\frac{t}{T}} \tag{21}$$

which looks like

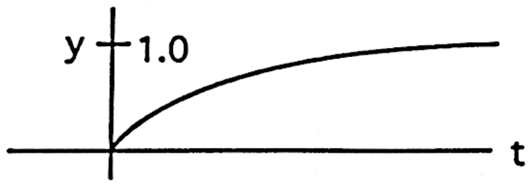


Figure 12

We were not just lucky here. Equation (18) is the *ratio of two polynomials*; 1 in the numerator and s^2T+s in the denominator. This is the usual case. Suppose we consider another system described by the differential equation:

$$a_2\frac{d^2y}{dt^2}+a_1\frac{dy}{dt}+a_0y=b_1\frac{dx}{dt}+b_0x$$

Take the Laplace transform,

$$a_2s^2Y(s)+a_1sY(s)+a_0Y(s)=b_1sX(s)+b_0X(s)$$

(I assume you recall that if *F(s)* is the transform of *f(t)* then *sF(s)* is the transform of $\frac{df}{dt}$).

or

$$\left(a_2s^2+a_1s+a_0\right)Y(s)=\left(b_1s+b_0\right)X(s)$$

or

$$\frac{Y(s)}{X(s)}=G(s)=\frac{b_1s+b_0}{a_2s^2+a_1s+a_0}$$

This is also the ratio of polynomials. Notice that the transforms of common inputs (19) are ratios of polynomials. So *Y(s)=X(s)G(s)* will be another *ratio of polynomials*,

$$Y(s) = \frac{b_n s^n + b_{n-1} s^{n-1} + ... + b_1 s + b_0}{a_m s^m + a_{m-1} s^{m-1} + ... + a_1 s + a_0} \tag{22}$$

Both polynomials can be characterized by their *roots*: those values of s that make them zero:

$$Y(s) = \frac{B(s+z_1)(s+z_2)\cdots(s+z_n)}{A(s+p_1)(s+p_2)\cdots(s+p_m)} \tag{23}$$

The roots of the numerator, when $s = -z_k$ causes $Y(s)$ to be zero, are called the *zeros* of the system. The points in the s plane where $s = -p_k$ are the roots of the denominator, where Y(s) goes to infinity, are the *poles*.

Thus any transfer function, or its output signal can be characterized by *poles* and *zeros*. And the *inverse transform* can be effected by *partial fraction expansion*.

But how did we make the sudden jump from the inverse transform of (16a) to the cook-book recipe of partial fraction expansion?

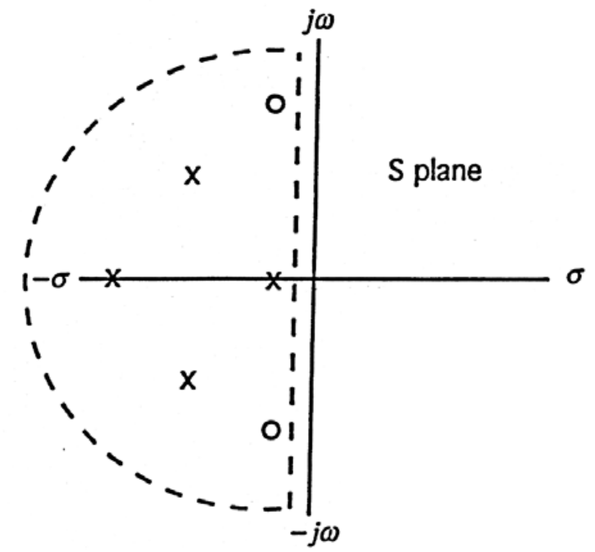


Figure 12a

Equation (16a) says integrate $F(s)G(s)e^{st}\,ds$ over the entire s-plane. It turns out that this integral can be evaluated by integrating over a contour that encloses all the poles and zeros of $F(s)G(s)$. Since their poles lie in the left-hand plane ($e^{+\sigma t}$ blows up) they can be enclosed in a contour such as C, as shown. But even better, the value of this integral is equal to the evaluation of the "residues" evaluated at each pole. This is a consequence of conformal mapping in the complex plane and is, we confess, a branch of mathematics that we just don't have time to go into. The residue at each pole is the coefficient evaluated by partial fraction expansion since it expresses $F(s)G(s)$ as a sum of the poles:

$$\frac{A_1}{(s+p_1)} + \frac{A_2}{(s+p_2)} + \cdots$$

Thus the integration over the s-plane in (16a) turns out to be just the same as evaluating the coefficients of the partial fraction expansions.

Again, it will be assumed that you know the cook-book method of Laplace transforms and Bode diagrams. These tools are an *absolute minimum* if you are to understand systems and *do systems analysis*. And this includes *biological systems*.

Before we get back to feedback, we must first show that the methods of analysis you learned do not apply only to electric circuits.

2.8. Mechanical Systems

In these systems one is concerned with force, displacement and its rate of change, velocity. Consider a simple mechanical element - a *spring*. Symbolically, it appears:

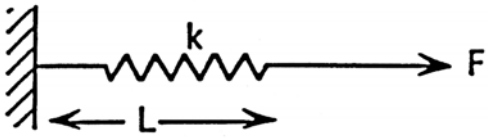


Figure 13

F is the force, L is the length. *Hook's law* states

$$F = kL \tag{24}$$

where k is the spring constant.

Another basic mechanical element is a *viscosity* typical of the shock absorbers in a car's suspension system or of a hypodermic syringe.

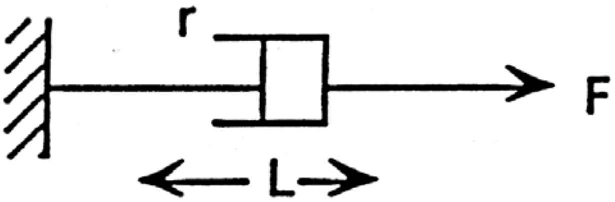


Figure 14

Its symbol is shown in Fig. 14 as *plunger in a cylinder*. The relationship is

$$F = r\frac{dL}{dt} = rsL \tag{25}$$

That is, a *constant force* causes the element to change its length at a constant velocity. r is the viscosity. The element is called a *dashpot*.

Let's put them together,

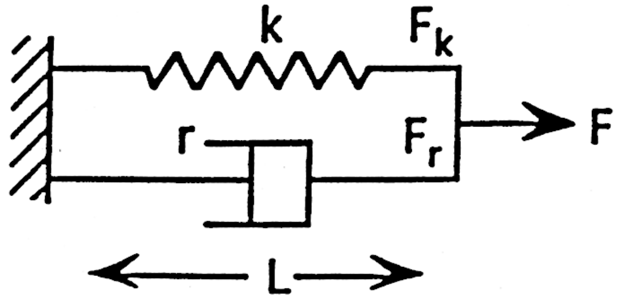


Figure 15

In Fig. 15 the force F, is divided between the two elements, $F = F_k + F_r = kL + rsL$ (in Laplace notation), or

$$F(s) = (rs + k)L(s)$$

If we take F to be the input and L the output

$$\frac{L(s)}{F(s)} = G(s) = \frac{1}{sr + k} = \frac{1/k}{s\frac{r}{k} + 1} = \frac{1/k}{sT + 1}$$

where $T = r/k$ is the system time constant. This is, of course, a *first-order lag*, just like the circuit in Fig. 3 governed by eq (3), with $j\omega$ replaced by s, plus the constant $1/k$. Fig. 15 is a simplified *model of a muscle*. If you think of F as analogous to voltage V, length L as the analog of electric charge Q, and velocity $\frac{dL}{dt}$ as current I $\left(\frac{dQ}{dt}\right)$, then Figures 3 and 15 are interchangeable with the *spring* being the *analog* of a *capacitor* (an energy storage element) and the *dashpot* the *analog* of a *resistor* (energy dissipator). So it's no wonder that you end up with the same mathematics and transfer functions. Only the names change.

2.9. Membrane Transport

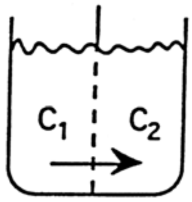


Figure 16

Two liquid compartments containing a *salt solution* at two different concentrations C_1 and C_2 are separated by a permeable membrane. The salt will diffuse through the membrane at a *rate* proportional to the concentration *difference*. That is:

$$\frac{dC_2}{dt} = k(C_1 - C_2)$$

$$or \quad \frac{dC_2}{dt} + kC_2 = kC_1$$

$$or \quad (s+k)C_2(s) = kC_1(s)$$

$$or \quad \frac{C_2(s)}{C_1(s)} = G(s) = \frac{k}{s+k}$$

another *first order lag*.

These examples are just to remind you that what you learned about Laplace transforms *do not apply just to electric circuits*. It applies to any sort of system that can be described by linear differential equations, put into a block in a block diagram and described as a transfer function.

3. Feedback and Dynamics

In deriving equation (1) to explain the major reasons that feedback is used, dynamics were ignored. But *feedback greatly alters a system's dynamics* and it is important to understand and anticipate these effects. In some technological (but not biological) situations, feedback is used solely to achieve a dynamic effect.

3.1. Static Feedback

Let's start with an example: the simple first-order lag.

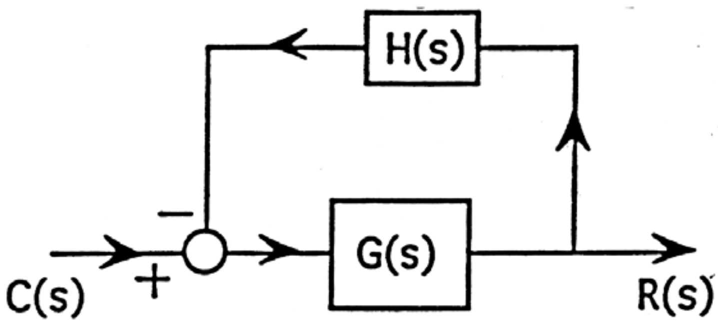


Figure 17

In Fig. 2 we assumed unity feedback (*H(s) =1*) for simplicity but now we must be more general. A little algebra will show that in this case the ratio of the response, *R(s)* to the command *C(s)* is

$$\frac{R}{C}=\frac{G(s)}{1+G(s)H(s)} \tag{26}$$

H(s) is the gain of the feedback path. *G(s)H(s)* is called the *loop gain*. Let's take a specific example

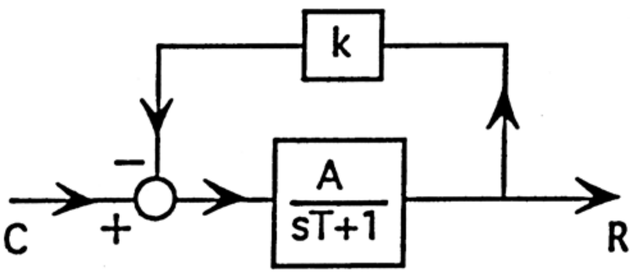


Figure 18

Let *G(s)* be $\frac{A}{sT+1}$ and *H(s)* be a simple gain *k* with no dynamic. From (26),

$$\frac{R}{C}=\frac{\frac{A}{sT+1}}{1+\frac{Ak}{sT+1}}=\frac{A}{sT+1+Ak}=\frac{\frac{A}{1+Ak}}{s\left(\frac{T}{1+Ak}\right)+1} \tag{27}$$

In this form *A /(1 + Ak)* is the gain at *dc* (zero frequency, $s = 0$) and *T/(1 + Ak)* is the new time constant. As *k* increases from zero, the closed loop dc gain decreases. This is a universal property of negative feedback, the more the feedback the lower the closed-loop gain.

The new time constant *T/(1+Ak)* also decreases so the system gets faster. This can clearly be seen in the step response.

If we used partial fraction expansion (see eq 20), it is easy to show that if $C(s)=\frac{1}{s}$ then

$$R(t)=\frac{A}{1+kA}\left(1-e^{-t/(\frac{T}{1+kA})}\right) \tag{28}$$

which will appear, for different values of k, or amount of feedback,

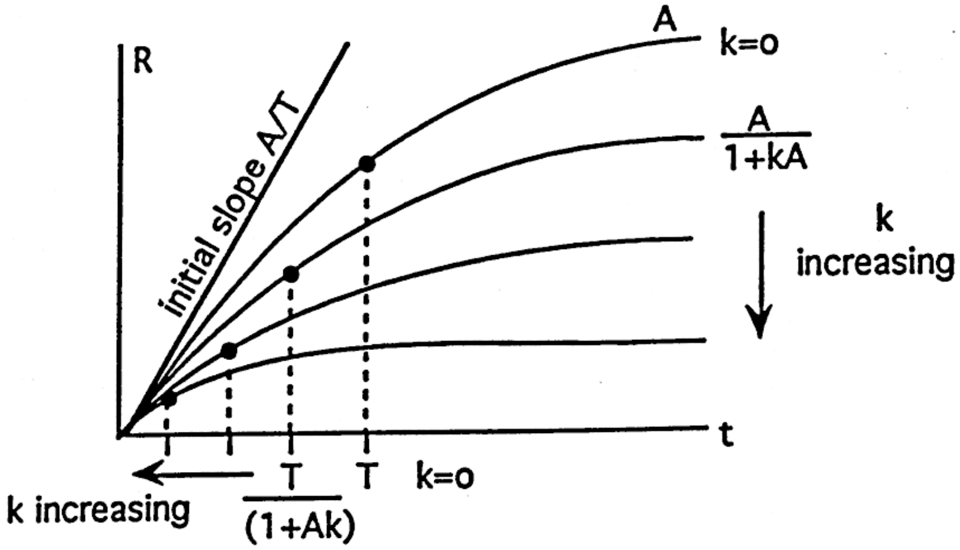


Figure 19

When $k=0$, the steady-state (dc) gain is A, the time constant is T, (when $t=T$, the term in parentheses in (28) is $1- e^{-1} = 0.63$, so $t=T$ is when the response is 63% of its way to steady state.) The initial slope from (28) is A/T and is independent of k. So the response starts off at the same rate but as k increases, the feedback kicks in and lowers the steady state level which is consequently reached sooner.

This phenomenon can also be clearly seen in the Bode diagram (Fig. 20). The high frequency behavior can be found by letting $s = j\omega$ in (27) and then letting ω become very large. Then $\frac{R}{C} \Rightarrow \frac{A}{j\omega T}$ which is independent of k.

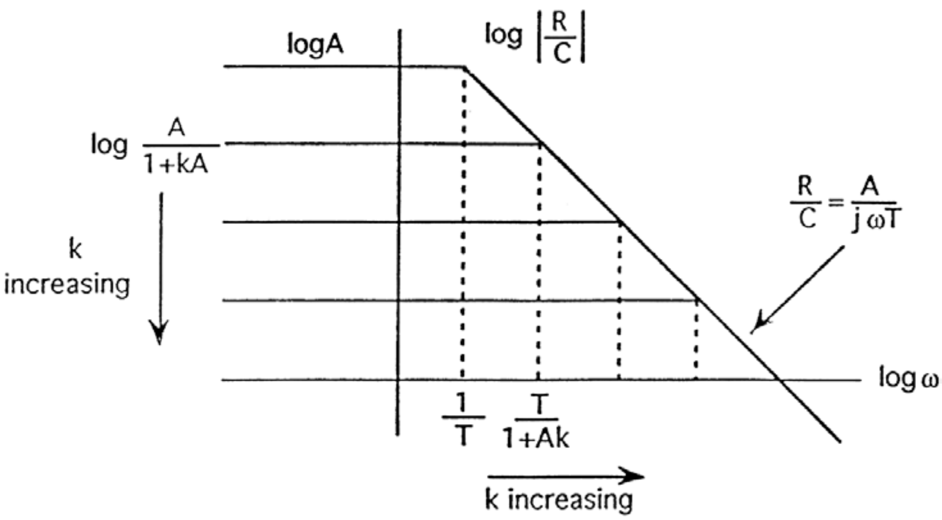


Figure 20

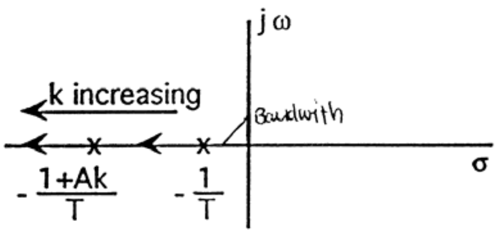


Figure 21

Thus as the low-frequency gain is pressed down as k increases, the intersection with the high frequency line $\frac{A}{j\omega T}$ occurs at higher and higher frequencies (or smaller and smaller time constants). Thus, the *bandwidth* (the break frequency $\frac{(1+Ak)}{T}$) *increases as k increases*.

This can also be visualized on the s-plane. Equ. (27) has one pole at $s=-\frac{(1+Ak)}{T}$. As k increases, the pole moves left on the negative real axis ($-\sigma$) showing in yet another way that the system is getting faster with a smaller time constant and a wider bandwidth.

3.2. Feedback with dynamics

From Fig. 17 and (26) recall that

$$\frac{R}{C}=\frac{G(s)}{1+G(s)H(s)} \tag{29}$$

If *IGHI* becomes very large (loop gain >>1) $\frac{R}{C}\Rightarrow\frac{1}{H(s)}$. This is a general and useful property of feedback.

If you put a differentiator in the feedback loop - you get an *integrator*.
If you put an integrator in the feedback loop - you get a *differentiator*.
If you put a *high gain* in the feedback loop you get a *low gain* for *R/C* as shown in Figs. 19 and 20.
If you put a *lead element* in the feedback loop - you get a lag.
And so on …

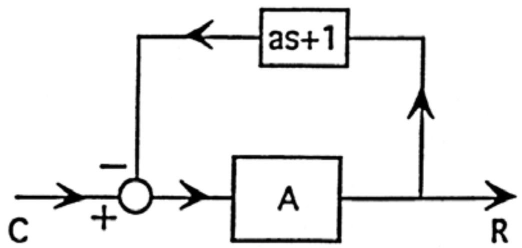


Figure 22

Let's illustrate the last point as an exercise. Here is *H(s) = as+1*, a lead element.

$$|H(j\omega)=\sqrt{(a\omega)^2+1}$$
$$\angle H(j\omega)=\tan^{-1}(a\omega)$$

The Bode diagram appears:

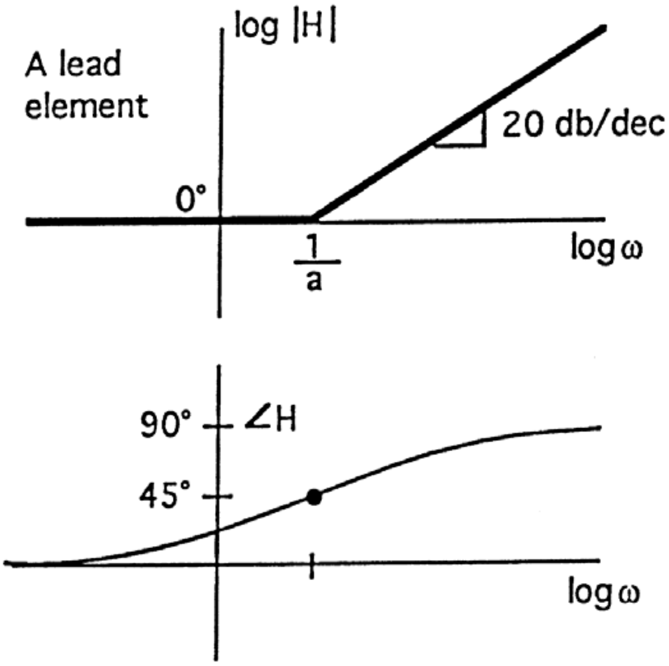


Figure 23

For $\omega < 1/a$, the gain is close to 1.0 (0 db) and the phase is near zero.

For $\omega > 1/a$, the gain rises linearly with ω so it is described by a straight line *rising* at a slope of 20 db/decade. The phase is *leading* and approaches + 90° which, of course, is *why it's called a lead*. Compare with the lag shown in Fig. 6.

From Fig. 22, the closed loop response is:

$$\frac{R}{C} = \left(\frac{G}{1+HG}\right) = \frac{A}{1+A(sa+1)} = \frac{A}{saA+1+A} = \frac{\dfrac{A}{1+A}}{s\left(\dfrac{aA}{1+A}\right)+1} \qquad (30)$$

This is a first-order lag, illustrating the main point that a lead in the feedback path turns the whole system into a lag. The steady state gain (when $s = j\omega = 0$) is, as expected, $\frac{A}{(1+A)}$ which, if A is large, will be close to *1.0* . The new time constant is $\frac{aA}{(1+A)}$. Again, if A is large, this will be close to a. Increasing a increases the lead action by shifting the curves in Fig. 23 to the left so the phase lead occurs over a wider frequency range. It also increases the time constant of the lag in (30) so the lag also covers a wider frequency range. The Bode diagram of (30) will look just like Fig. 23 but with everything upside down, again reflecting that $\frac{R}{C} \Rightarrow \frac{1}{H(s)}$.

3.3. Oscillations

While negative feedback is, in general, good and useful, one must be aware that if incorrectly designed, or in biology, if a change occurs due to, say, a disease, feedback systems can become unstable and oscillate. In engineering, it is always wise in designing a system, to make sure it is stable *before* you build it. In biology, if you see oscillations occurring, it is nice to know what conditions could lead to such behavior.

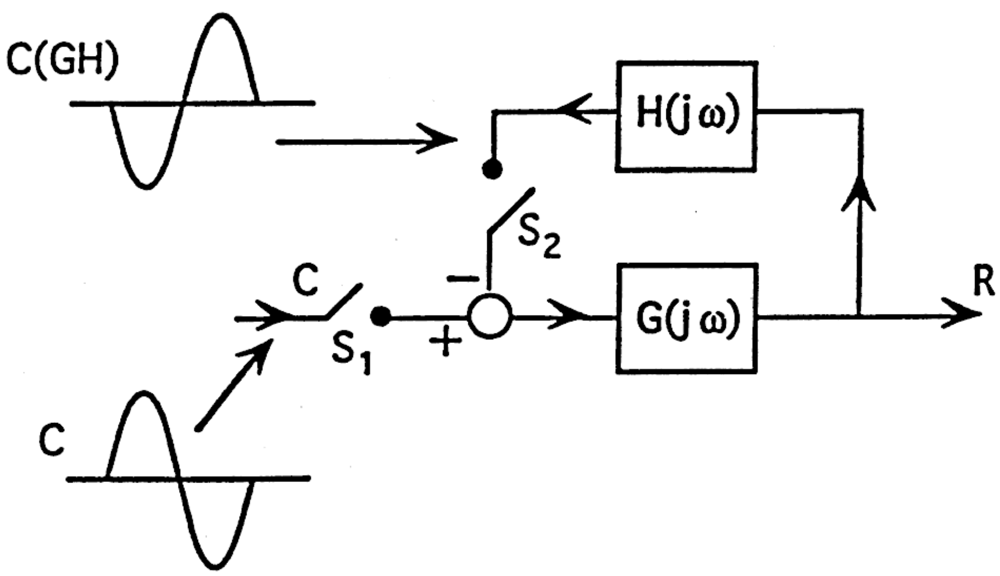


Figure 24

Consider the system in Fig. 24. Let switch S_2 be open and S_1 closed so that a sine wave put in at C flows through $G(j\omega)$ and then back around through $H(j\omega)$. This output is shown as $C(GH)$. Suppose that the gain $|GH|$ at some frequency ω_0 is exactly 1.0 and the phase is exactly -180° as shown. Now instantaneously open S_1 and close S_2. After going through the -1 at the summing junction, the signal $C(GH)$ will look exactly like C and can substitute for it. Thus the output of H can serve as the input to G which then supplies the output of H which then ... etc., etc., the sine wave goes round and round - the system is oscillating.

The key is that there must exist a frequency, ω_0, at which the loop gain $G(j\omega)H(j\omega)$ is -180°. At this frequency, evaluate the gain $|G(j\omega_0)H(j\omega_0)|$. If this gain is >1, oscillations will start spontaneously and grow without limit: the system is unstable.
If the gain is <1, the system is stable. (There are some complicated systems where this is an oversimplification, but it works for most practical purposes).

Thus, the Bode diagram can be used to test for stability. One needs to plot the Bode diagram of the loop gain. As an example, consider the system in Fig. 18 where $G(s)$ is a first-order lag and $H(s)=k$. The loop gain is $\frac{kA}{(sT+1)}$ and its Bode diagram will resemble Fig. 6. This figure shows that the phase lag never exceeds -90°. It never gets to -180°. Therefore it can never become unstable, no matter how much the loop gain kA is increased.

Consider a double lag (Fig. 25).

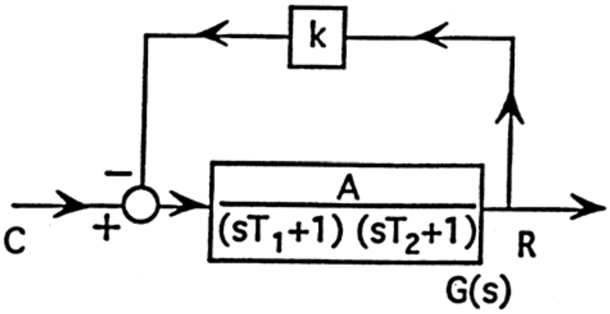


Figure 25

Its loop Bode diagram will look like this:

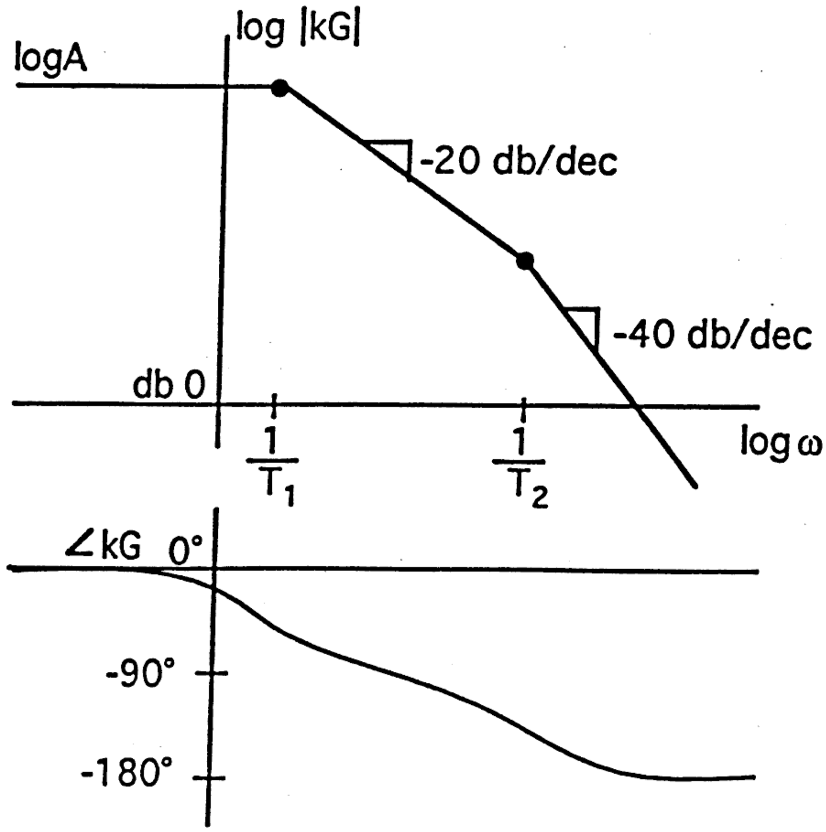


Figure 26

Each pole at $-\frac{1}{T_1}$ and $-\frac{1}{T_2}$ causes the gain to add a decrease of -20db/dec and a phase lag of -90°. Thus the phase approaches -180° but never gets there. So technically, this system too can never be made unstable.
However, when the phase gets near -180° and $|kG|$ is >1, the system "rings"; its step response would look like Fig. 27. Such a system would be pretty useless even though it is technically stable.

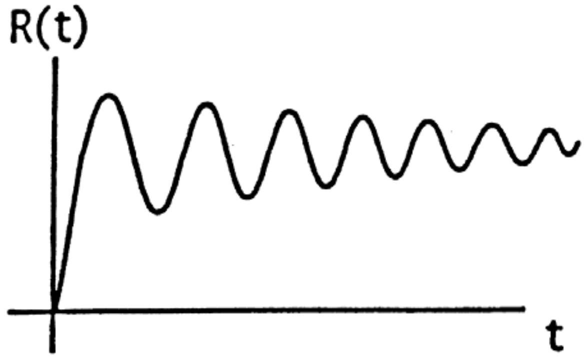


Figure 27

Obviously if we add another pole (another term $\frac{1}{(sT_3+1)}$ in Fig. 25) the phase shift will approach 3x90°, or -270°. Such a system can become unstable if the gain is too high. Most biological control systems, such as those mentioned on page 1, are dominated by a single pole, have modest gains and are very stable. But there is one element that presents a real danger for stability: delays.

3.4. Delays – A Threat to Stability

The output of a pure delay is the same as the input, just translated in time by the delay τ. This operator has the Laplace transfer function $e^{-s\tau}$.

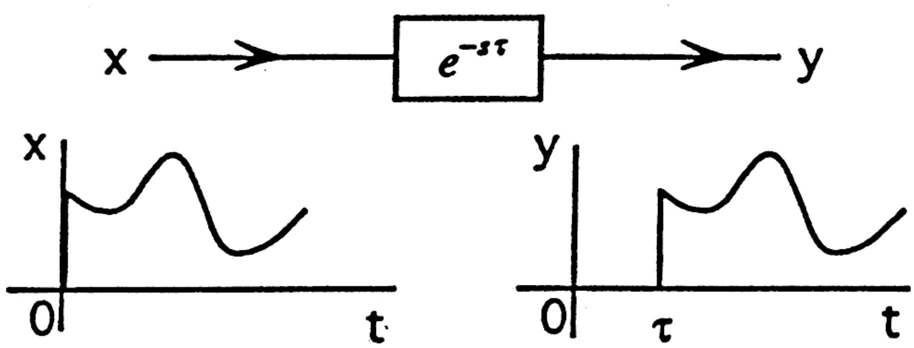


Figure 28

Delays are not uncommon in biology. For example, an endocrine gland dumps a hormone into the blood stream. It is carried in the blood stream to a distant receptor. This could take several seconds. This is called a *transport delay*. In engineering, for another example, if you wanted to incorporate earth into the control loop of a lunar rover (not a good idea) you would have to cope with the few seconds it takes for radio signals to get to the moon and back.

To see why delays are bad, look at its Bode diagram

$$\left|e^{-j\omega\tau}\right| = 1 \qquad \angle e^{-j\omega\tau} = -\omega\tau$$

Its gain is 1 at all frequencies. Its phase is directly proportional to frequency. (Think of τ as a time window - the higher the frequency, the more cycles fit in this window so the greater the phase lag.)

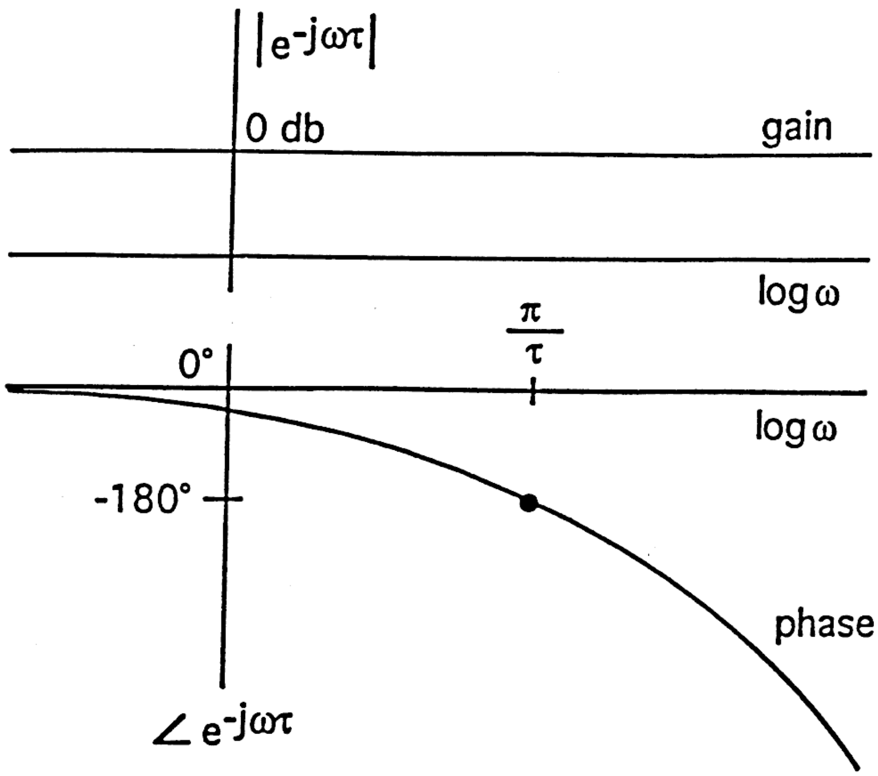


Figure 29

Fig. 29 shows the gain to be independent of ω at 0 db (*gain=1*). The phase lag is a linear function of ω but on a linear-log plot, it appears as an exponential increase in lag. This means that for any system with a delay in the loop, there will *always* exist a frequency ω_0 where the phase lag is 180°. So instability is always a possibility.

Consider a lag with a delay:

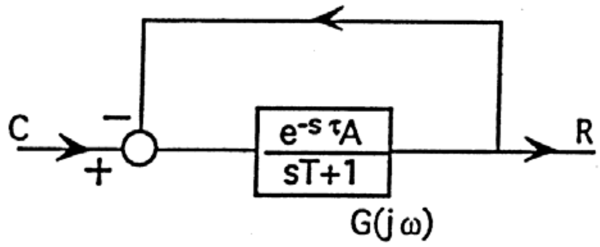


Figure 30

Look at the Bode diagram of the loop gain (Fig. 31). The phase lag is the sum of that due to the lag $\dfrac{A}{(j\omega T+1)}$ and that due to the delay $e^{-j\omega\tau}$ as shown by the two dashed curves. As illustrated, at ω_0 each contributes about -90° so the total is -180°. At ω_0 the gain is about 1 (*0 db*) so this system, if not unstable, is close to it.

To make it stable we could decrease A or increase T which would cause the gain at ω_0 to become <1. This would be the practical thing to do although either move would decease the closed loop bandwidth which is where $|G(j\omega)| = 1$.

Note that if $\tau << T$ the curve $\angle e^{-j\omega\tau}$ moves to the right and the point where the net lag reaches -180° moves to high frequencies where the gain is much less than 1.0. But as $\tau \Rightarrow T$, we rapidly get into trouble.

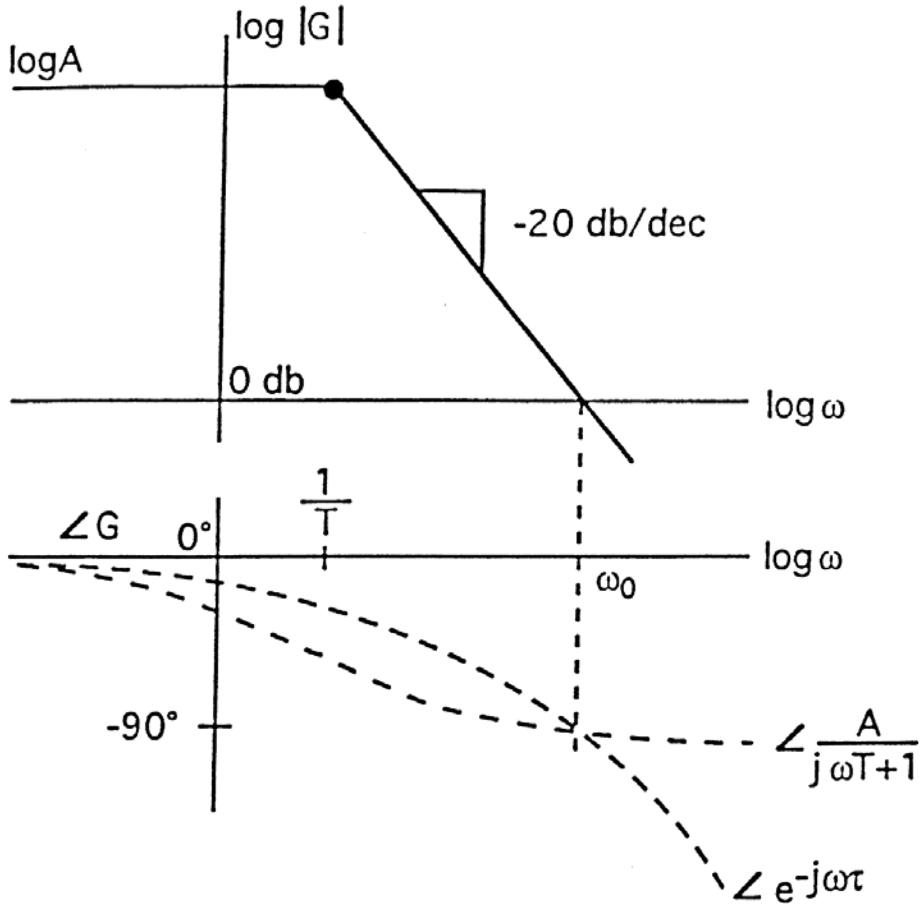


Figure 31

In the regulating systems mentioned on page 2, T is usually much larger than τ, so stability is seldom a problem. In neuromuscular control, however, where τ is due to synaptic delays and axonal conduction delays, T can become small (the time constant of a rapid arm, finger or eye movement) and problems can arise. To deal with this situation, Mother Nature has evolved another type of feedback called *parametric adaptive feedback* - but that is another story.

4. Summary

1. To start to analyze a system you draw its block diagram, which shows the processes and variables of interest, and make explicit the feedback pathway and error signal.
2. The input-output relationship of each block is described by a linear differential equation. Taking the Laplace transform of it gives one its transfer function $G(j\omega)$ or more generally $G(s)$, which tell you what $G(s)$ will do to any signal of the form e^{st}.
3. Collecting all these blocks together to form a net forward gain $G(s)$ and feedback gain $H(s)$, you know that the closed-loop transfer function is $\frac{G}{(1+GH)}$.
4. A transform takes $C(t)$ from the time domain to the frequency domain $C(s)$. This is passed through the system, simply by multiplication, to create the output $R(s)$. The inverse transform brings one back to the time domain and gives the answer $R(t)$. In practice, this is done by partial fraction expansion.
5. Most important, from equation (1), you know why negative feedback is used and you also know how to check for its major potential problem - instability.

Literature

- G. F. Franklin, J. D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Addison-Wesley Publishing Company, 1995.
- B. C. Kuo. *Automatic Control Systems*. John Wiley & Sons, Inc., 1995. ("my favourite")
- JW Nilsson, S Riedel. *Electric Circuits Revised and PSpice Supplement Package*. Prentice Hall; 6th edition, 2001
- Harris CM. The Fourier analysis of biological transients. J Neurosci.Methods 83 (1):15-34, 1998. ("read this if you want to really work with FFTs")