336: Systems and Control

Negative Feeback

(v1.0)

Contents

1 Feedback Control

Feedback is widespread in biochemical networks and physiological systems. Some form of feedback permeates almost every known biological process. On the face of it, feedback is a simple process that involves sending a portion of the output back to the input. If the portion sent back reduces the input then the feedback is called negative feedback otherwise it is called positive feedback.

The concept of feedback control goes back at least as far as the Ancient Greeks. Of some concern to the ancient Greeks was the need for accurate time keeping. In about 270 BC the Greek Ktesibios invented a float regulator for a water clock. The role of the regulator was to keep the water level in a tank at a constant depth. This constant depth yielded a constant flow of water through a tube at the bottom of the tank which filled a second tank at a constant rate. The level of water in the second tank thus depended on time elapsed.

Philon of Byzantium in 250 BC is known to have kept a constant level of oil in a lamp using a float regulator and in the first century AD Heron of Alexandria experimented with float regulators for water clocks. Philon and particularly Heron (13 AD) have left us with an extensive book (Pneumatica) detailing many amusing water devices that employed negative feedback.

It wasn't until the industrial revolution that feedback control, or devices for automatic control, became economically important. Probably the most famous modern device that employed negative feedback was the governor. Thomas Mead in 1787 took out a patent on a device that could regulate the speed of windmill sails. His idea was to measure the speed of the mill by the centrifugal motion of a revolving pendulum and use this to regulate the position of the sail. Very shortly afterwards in early 1788, James Watt is told of this device in a letter from his partner, Matthew Boulton. Watt recognizes the utility of the governor as a device to regulate the new steam engines that were rapidly becoming an important source of new power for the industrial revolution.

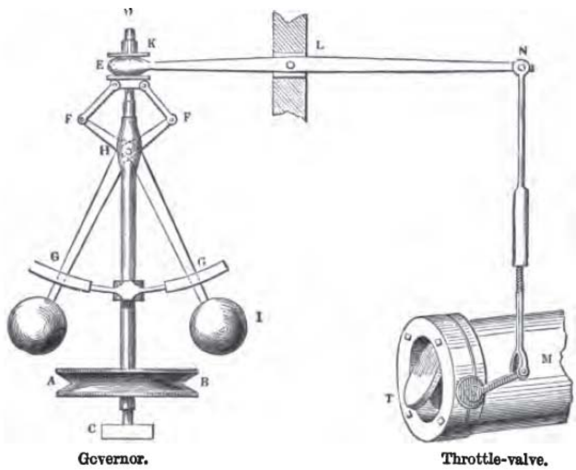

Figure 1: A typical governor from J. Farley, A Treatise on the Steam Engine: Historical, Practical, and Descriptive (London: Longman, Rees, Orme, Brown, and Green, 1827, p436

The device employed two pivoted rotating fly-balls which were flung outward by centrifugal force. As the speed of rotation increased, the flyweights swung further out and up, operating a steam flow throttling valve which slowed the engine down. Thus, a constant speed was achieved automatically. So popular was this innovation that by 1868 it is estimated that 75,000 governors (A History of Control Engineering, 1800-1930 By Stuart Bennett, 1979) were in operation in England. Many similar devices were subsequently invented to control a wide range of processes, including water wheels, telescope drives and temperature and pressure control.

The description of the governor illustrates the operational characteristics of **negative feedback**. The output of the device, in this case the steam engine speed, is Ťfed backŤ to control the rate of steam entering the steam engine and thus influence the engine speed.

During this period devices for automatic control were designed through trial and error and little theory existed to understand the limits and behavior of feedback control systems. One of the difficulties with feedback control is the potential for instability. As the governor became more widespread, improvements were made in manufacturing mechanical devices which reduced friction. As a result engineers began to notice a phenomena they termed hunting. This was where after a change in engine load, the governor would begin to 'hunt' in an oscillatory fashion for the new stream rate that would satisfy the load. This effect caused considerable problems with maintaining a stable engine speed and resulted in James Maxwell and independently Vyshnegradskii, undertaking the first theoretical analysis of a negative feedback system.

Until the 20th century, feedback control was generally used as a means to achieve automatic control, that is to ensure that a variable, such as a temperature or a pressure was maintained at some set value. However, an entirely new application for feedback control was about to emerge with the advent of electronics in the early part of the 20th century.

Feedback Amplifiers Amplification is one of the most fundamental tasks one can demand of an electrical circuit. One of the challenges facing engineers in the 1920's was how to design amplifiers whose performance was robust with respect to the internal parameters of the system and which could overcome inherent nonlinearities in the implementation. This problem was especially critical to the effort to implement long distance telephone lines across the USA.

These difficulties were overcome by the introduction of the feedback amplifier, designed in 1927 by Harold S. Black (Mindell, 2000), who was an engineer for Western Electric (the forerunner of Bell Labs). The basic idea was to introduce a negative feedback loop from the output of the amplifier to its input. At first sight, the addition of negative feedback to an amplifier might seem counterproductive. Indeed Black had to contend with just such opinions when introducing the concept. His director at Western Electric dissuaded him from following up on the idea and his patent applications were at first dismissed. In his own words, “our patent application was treated in the same manner as one for a perpetual motion machine” (Black, 1977).

While Black’s detractors were correct in insisting that the negative feedback would reduce the gain of the amplifier, they failed to appreciate his key insight that the reduction in gain is accompanied by increased robustness of the amplifier and improved fidelity of signal transfer.

Unlike the steam engine governor which is used to stabilize some system variable, negative feedback in amplifiers is used to accurately track an external signal. These two applications highlight the two main ways in which negative feedback can be used, namely as a **regulator** or as a **servomechanism**.

As a regulator, negative feedback is used to maintain a controlled output at some constant desired level, whereas a servomechanism will slavishly track a reference input. We can see both applications at work in the eye. On the one hand there is the need to control the level of light entering the pupil. The diameter of the pupil is controlled by two antagonistic muscles. If the external light intensity increases, the muscles respond by reducing the pupil diameter, whereas the muscles increase the pupil diameter if the light intensity falls. The pupil reflex serves as an example of negative feedback using in a regulator mode. In contrast tracking an object involves maintaining the eyeball fixed on the object. In this mode the eye functions as a servomechanism.

Both regulator and servomechanism are implemented using the same operational mechanism. Figure 2 shows a generic negative feedback circuit. On the left of the figure can be found the input, sometimes terms the desired value or more often the set point. If the circuit is used as a servomechanism then the output tracks the set point. As the set point changes the output follows. If the circuit is used as a regulator or homeostatic device then the set point is held constant and the output is maintained at or near the set point even in the face of disturbances.

The central mechanism in the feedback circuit is the generation of the error signal, that is the difference between the desired output (set point) and the actual output. The error is fed into a controller (often something that simply amplifies the error) which is used to increase or decrease the process. For example, if a disturbance on the process block reduces the output, then the feedback operates by generating a positive error, this in turn increases the process and restores the original drop in the output.

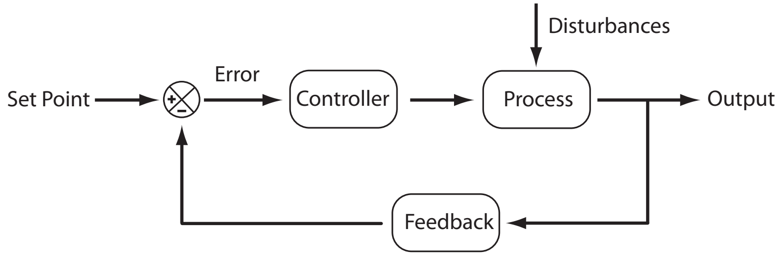

Figure 2: Generic structure of a negative feedback system.

Simple Quantitative Analysis

The figure of the generic negative feedback circuit (Figure 2) is highly stylized which makes it difficult to identify the various component in a real biological system. In addition, biological system are invariably more complex with multiply nested feedback loops and multiple inputs and outputs. If it remarkable that even after 50 or 60 years of research, the role of many of the feedback systems in biochemical networks take is still highly speculative.

In the remainder of this section we will consider some basic properties of negative feedback systems. We will revisit feedback systems again towards the end of the chapter when we consider the more advanced topic of frequency response. The simplest way to think about feedback quantitatively is by reference to Figure 3.

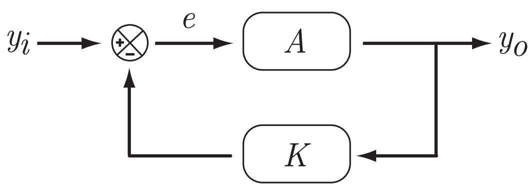

Figure 3: Generic structure of a negative feedback system.

We will assume some very simple rules that govern the flow of information in this feedback system. For example, the output signal, y_o will be given by the process A multiplied by the error, e. The feedback signal will be assumed to be proportional to y_o, that is Ky_o. Finally, the error signal, e will be given by the difference between the set point, y_i and the feedback signal, Ky_o (Figure 4).

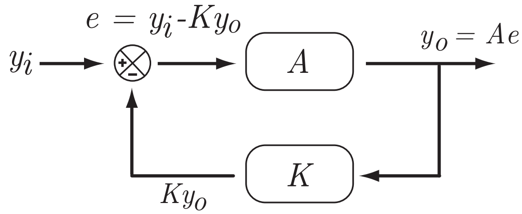

Figure 4: Generic structure of a negative feedback system.

From these simple relations it is straightforward to show that:

$$y_o = \frac{Ay_i}{1 + AK} \qquad \text{or more simply} \quad y_o = Gy_i \tag{1}$$

G is called the gain of the feedback loop, often called the **closed loop gain**. *Gain* is a term that is commonly used in control theory and refers to the scalar change between an input and output. Thus a gain of 2 simply means that a given output will be twice the input. In addition to the close loop gain, engineers also define two other gain factors, the **open loop gain** and the **loop gain**. The open loop gain is simply the gain from process, A, alone. It is the gain one would achieve if the feedback loop were absent. The loop gain is the gain from the feedback and process A combined, AK. The loop gain is a significant quantity when discussing the stability of feedback circuits. Figure 4 illustrates the different types of gain in a feedback circuit.

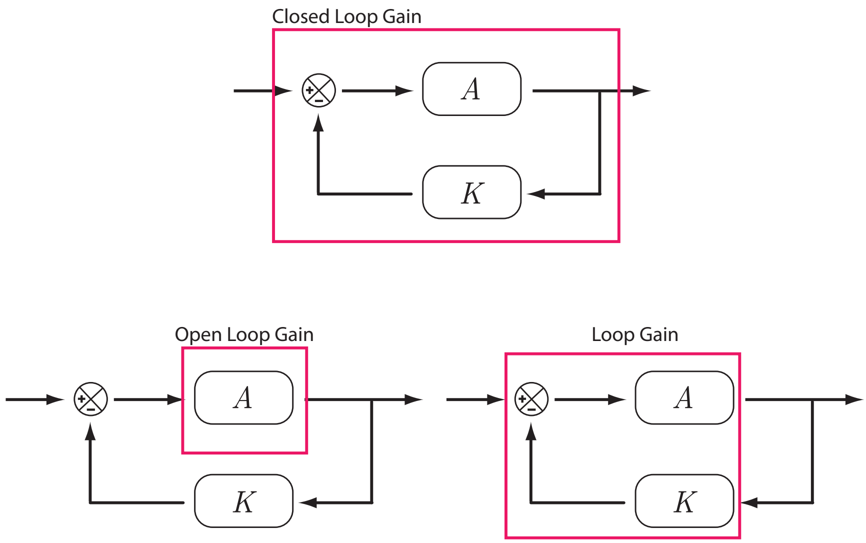

Figure 5: Generic structure of a negative feedback system.

We can use equation 1 to discover some of the basic properties of a negative feedback circuit. The first thing to note is that as the loop gain, AK, increases, the system behavior becomes more dependent on the feedback loop and less dependent on the rest of the system:

$$\text{when} \quad AK \gg 1 \quad \text{then} \quad G \simeq \frac{A}{AK} = \frac{1}{K}$$

This apparently innocent effect has significant repercussions on other aspects of the circuit. To begin with, as the system becomes less important on A, so does variation in the properties of A. Feedback makes the performance of the system independent of any variation in A. Such variation might include noise or variation as a result of the manufacturing process or in the case of biological systems, genetic variation. To be more precise we can compute the sensitivity of the gain G with respect to variation in A.

$$\frac{\partial G}{\partial A} = \frac{\partial}{\partial A} \frac{A}{1 + AK} = \frac{1}{(1 + AK)^2}.$$

If we consider the relative sensitivity we find:

$$\frac{\partial G}{\partial A} \frac{A}{G} = \frac{1}{1 + AK}$$

In addition to resistance to parameter variation, feedback also confers a resistance to disturbances in the output. Suppose that a nonzero disturbance d affects the output. The system behavior is then described by

$$y = Ae - d \qquad e = u - Ky.$$

Eliminating e, we find

$$y = \frac{Au - d}{1 + AK}.$$

The sensitivity of the output to the disturbance is then

$$\frac{\partial y}{\partial d} = -\frac{1}{1 + AK}.$$

The sensitivity decreases as the loop gain AK is increased. In practical terms, this means that the imposition of a load on the output, for example a current drain in an electronic circuit, protein sequestration on a signaling network or increased demand for an amino acid will have less of an effect on the circuit as the feedback strength increases. In electronics this property essentially modularizes the network into functional modules.

Last but not least, feedback also improves the fidelity of the response. That is, for a given change in the input, a system with feedback is more likely to faithfully reproduce the input at the output that a circuit without feedback. An ability to faithfully reproduce signals is critical in electronics communications and in fact it was this need that was the inspiration for the development of negative feedback in the early electronics industry.

Consider now the case where the amplifier A is nonlinear. For example a cascade pathway exhibiting a sigmoid response. Then the behavior of the system G (now also nonlinear) is described by

$$G(y_i) = y_o = A(e) \qquad e = y_i - Ky_o = y_i - KG(y_i).$$

Differentiating we find

$$G'(y_i) = A'(y_i)\frac{de}{dy_i} \qquad \frac{de}{dy_i} = 1 - KG'(y_i).$$

Eliminating $\frac{de}{dy_i}$, we find

$$G'(y_i) = \frac{A'(y_i)}{1 + A'(y_i)K}.$$

We find then, that if $A'(y_i)K$ is large ($A'(y_i)K \gg 1$), then

$$G'(y_i) \approx \frac{1}{K},$$

so, in particular, G is approximately linear. In this case, the feedback compensates for the nonlinearities $A(\cdot)$ and the system response is not distorted. (Another feature of this analysis is that the slope of $G(\cdot)$ is less than that of $A(\cdot)$, i.e. the response is "stretched out". For instance, if $A(\cdot)$ is saturated by inputs above and below a certain "active range", then $G(\cdot)$ will exhibit the same saturation, but with a broader active range.)

A natural objection to the implementation of feedback as described above is that the system sensitivity is not actually reduced, but rather is shifted so that the response is more sensitive to the feedback K and less sensitive to the amplifier A. However, in each of the cases described above, we see that it is the nature of the loop gain AK (and not just the feedback K) which determines the extent to which the feedback affects the nature of the system. This suggests an obvious strategy. By designing a system which has a small "clean" feedback gain and a large "sloppy" amplifier, one ensures that the loop gain is large and the behavior of the system is satisfactory. Engineers employ precisely this strategy in the design of electrical feedback amplifiers, regularly making use of amplifiers with gains several orders of magnitude larger than the feedback gain (and the gain of the resulting system).

1. Amplification of signal.
2. Robustness to internal component variation.
3. High fidelity of signal transfer.
4. Low output impedance so that the load does not affect the performance of the circuit.

These are the main advantages of negative feedback but as we will

Type	Time Domain	s Domain
Multiplier	$y = Kv$	$Y(s) = K V(s)$
Summer	$y = v_1 + v_2$	$Y(s) = V_1(s) + V_2(s)$
Cascade		$Y(s) = V_1(s) V_2(s)$

see in a later section on frequency response, feedback can confer additional useful features.

2 Block Diagrams in the Laplace Domain

In the previous section we saw the use of block diagrams to understand more complex systems. In control theory, block diagram can also be constructed in the Laplace Domain. This is possible because of the simple linearity rules that Laplace transforms obey. The table blow summaries the most important.

These rules can be depicted in diagrammatic form.

3 Frequency Response

Up to this point we have considered how the steady state changes as a result of a step change in some variable, typically an enzyme activity or external species. Such changes were described by the control coefficients.

Many studies on dynamical systems either focus on the steady state or on the time evolution of the system. Such approaches are easy to comprehend and measure and are therefore popular. The study of systems with respect to time is often referred to as the analysis of systems in the **time domain**. There is however an alternative

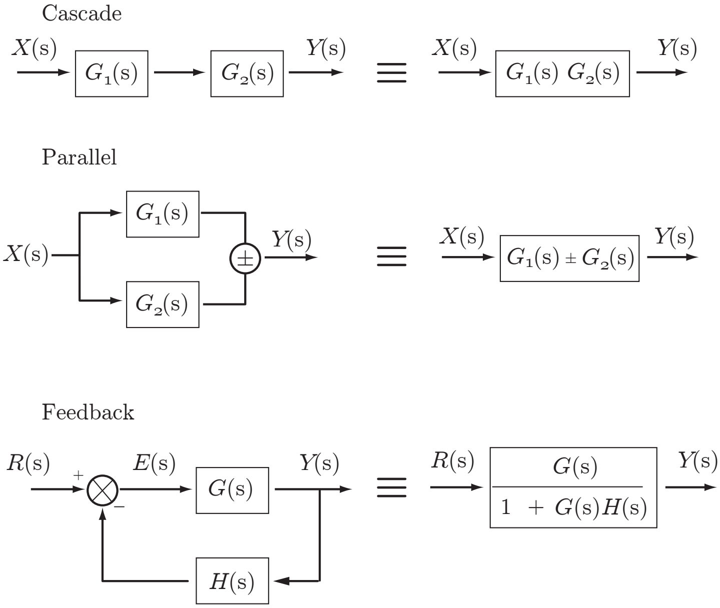

Figure 6: Block diagram in the Laplace Domain.

approach to view a dynamical system and that is its steady state response to inputs that vary in a sinusoidal fashion. More specifically the input to a system is subject to a sinusoidal signal that varies in frequency and the response of one or more systems variables is recorded. Such an analysis is referred to as a **frequency response** or **frequency domain** analysis.

The frequency response is used extensively by engineers to study the performance and detailed characteristics of a given system. A frequency response can give valuable information on the stability of a system, a systems responsiveness or how a system rejects disturbances. For engineers, a frequency response also gives valuable information that can be used to improve the design of a system. There is an elaborate mathematical machinery that accompanies the study of the frequency response that involves an understanding of complex numbers and analysis. Here we will only give a summary of the main results and show how a frequency response can be used to understand certain dynamic behaviors.

3.1 Sinusoidal Signals

To begin the discussion we must first reacquaint ourselves with the sine wave or sinusoidal signal. The sinusoidal signal is the most fundamental periodic signal. Any other periodic signal can be constructed from a summation of sinusoidal signals. We can describe a sinusoidal signal by an equation of the general form:

$$y(t) = A \sin(\omega t + \theta)$$

This equation describes how an output, y, varies in time, t. The equation has three terms, the amplitude (A), the angular frequency (ω) and the phase (θ). We will describe the phase in more detail below but essentially it indicates how delayed or advanced and periodic signal may be. The angular frequency is the rate of the periodic signal and is expressed in radians per second. A sine wave traverses a full cycle (peak to peak) in 2π radians (circumference of a circle) so that the number of complete cycles traversed in one second is then $\omega/2\pi$. This is termed the frequency, f, (cycles sec^{-1}) of a sine wave and has units of Hertz. The angular frequency is then conveniently expressed as $\omega = 2\pi f$.

The amplitude is the extent to which the periodic function changes in the y direction, that is the maximum height of the curve from the origin (Figure 8). The horizontal distance peak to peak is referred to as the period, T and is usually expressed in seconds. The inverse of the period, $1/T$ sec^{-1} is equal to the frequency, f.

Figure 7 shows a typical plot of the sinusoidal function, $y(t) = A \sin(\omega t + \theta)$, where the amplitude is set to one, the frequency to 2 cycles per second and the phase to zero.

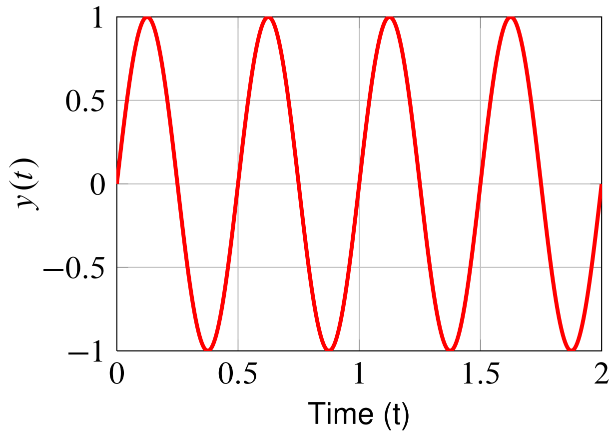

Figure 7: $y(t) = A\sin(\omega t + \theta)$ where $A = 1$, $f = 2$ Hz so that $\omega = 4\pi, \theta = 0$.

The left panel in Figure 8 shows the affect of varying the amplitude and right panel two signals of different frequency.

Sinusoidal signals do not need to start at time zero. The two sine waves shows in Figure 9 have the same frequency and amplitude but one of them is shifted to the right by 90 degrees (or $\pi/2$ radians), that is phase shifted.

One important property of sinusoidal signals is that the sum of two sinusoidal signals of the same frequency but different phase and amplitude will result in another sinusoidal frequency with a different phase and amplitude but identical frequency.

$$A_1\sin(\omega t + \theta_1) + A_2\sin(\omega t + \theta_2) = A_3\sin(\omega t + \theta_3)$$

In fact any linear operation on a sinusoid will only change the amplitude or phase. For example, multiply by a constant only changes the amplitude.

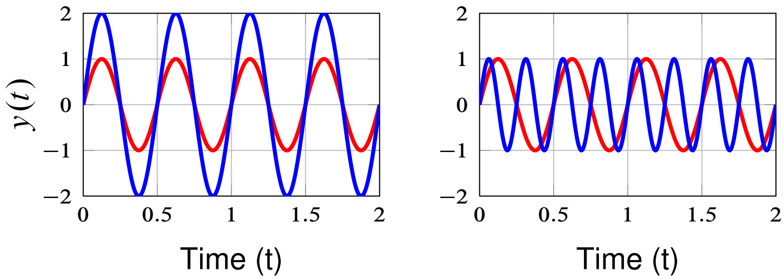

Figure 8: Left Panel: Amplitude change $y(t) = A\sin(\omega t + \theta)$ where $A = 2$, $f = 2$ Hz, $\theta = 0$ Right Panel: Frequency Change $y(t) = A\sin(\omega t + \theta)$ where $A = 1$, $f = 4$ Hz, $\theta = 0$

Linear Systems and Sinusoidals

Given that linear systems are composed of combinations of linear operations such as addition, multiplication by a constant or integration we can be sure that any sinusoidal input to such a system will only experience changes to the phase and amplitude of the signal. The frequency will remain unchanged. Of particular interest is how the steady state responds to a sinusoidal input, termed the **sinusoidal steady state response**.

We can illustrate this by way of an example. Consider the linear first-order differential equation:

$$\frac{dy}{dt} + ay = b\sin(\omega t)$$

where the input to the equation is a sinusoidal equation, $\sin(\omega t)$. This equation is of the standard linear form:

$$\frac{dy}{dt} + P(t)y = Q(t)$$

We can therefore use the integrating factor technique to solve this

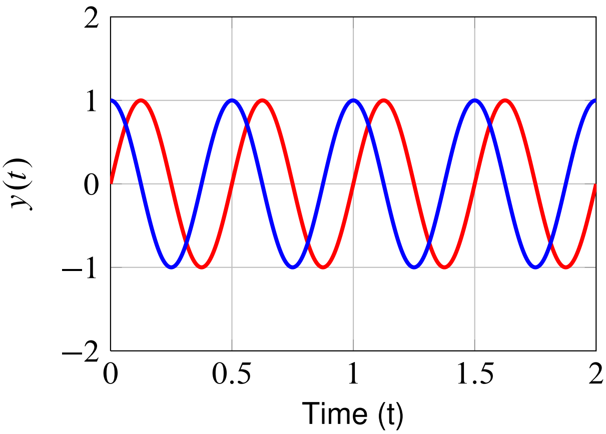

Figure 9: Phase change: $y(t) = A\sin(\omega t + \theta)$ where $A = 1, f = 2$ Hz, $\theta = 90°$. The red curve is shifted 90° to the right relative to the blue curve.

equation. The integrating factor is given by

$$\rho = e^{\int P(t)dt} = e^{\int a dt} = e^{at}$$

Multiplying both sides by ρ and noting that

$$\frac{d}{dt}\left(\int P(t)dt\right) = P(t)$$

we obtain

$$\frac{d}{dt}(ye^{at}) = e^{at} b \sin(\omega t)$$

(Note that $\frac{d}{dt}(ye^{at}) = \frac{dy}{dt}e^{at} + yae^{at}$).
Assuming an initial condition of $y(0) = 0$ and integrating both sides gives:

$$y = b\frac{\omega e^{-at} + a\sin(\omega t) - \omega\cos(\omega t)}{a^2 + \omega^2}$$

At this point we only want to consider the steady state sinusoidal response, hence as $t \to \infty$, then

$$y = \frac{b}{\sqrt{a^2 + \omega^2}}\frac{a\sin(\omega t) - \omega\cos(\omega t)}{\sqrt{a^2 + \omega^2}}$$

To show that the frequency of the input signal is unaffected, we proceed as follows. We start with the well known trigonometric identity:

$$\begin{aligned} A\sin(\beta - \alpha) &= A\cos(\alpha)\sin(\beta) - A\sin(\alpha)cos(\beta) \\ &= a\sin(\beta) + \omega\cos(\beta) \end{aligned}$$

where $a = A\cos(\alpha)$ and $\omega = \sin(\alpha)$. If we sum the squares of a and ω we obtain $a^2 + \omega^2 = A^2(\sin^2(\alpha) + \cos^2(\alpha)) = A^2$. That is:

$$A = \sqrt{a^2 + \omega^2}$$

Similarly, $\omega/a = (A\sin(\alpha))/(A\cos(\alpha)) = \tan(\alpha)$. That it is:

$$\alpha = \tan^{-1}\left(\frac{\omega}{a}\right)$$

Since $a\sin(\beta) + \omega\cos(\beta) = A\sin(\beta - \alpha)$ where $\beta = \omega t$ then

$$y = \frac{b}{\sqrt{a^2 + \omega^2}}\sin(\omega t - \alpha)$$

This final result shows us that the frequency, ω remains unchanged but the amplitude is scaled by $\sqrt{a^2 + \omega^2}$ and the phase shifted by α. In summary the amplitude change is given by:

$$\frac{A_{out}}{A_{in}} = \frac{1}{\sqrt{a^2 + \omega^2}}$$

and the phase shift by:

$$\alpha = -\tan^{-1}\left(\frac{\omega}{a}\right)$$

3.2 Frequency Response

A sinusoidal signal that is applied to some specific input of a system will propagate itself throughout the system to varying degrees. For example we could apply a sinusoidal input to the input concentration, X_o of a simple linear chain. We could then examine the response in each of the species, $S_1, S_2, \ldots$. Upon application of a sinusoidal signal, the system would initially go through a transition as it accommodates the input, eventually reaching some steady periodic behavior. Most likely the pathway species would display some kind of sinusoidal behavior in response to the varying input. Those species nearest the source, X_o, are likely to respond more fully to the input, while species further away are likely to be affected to a lesser degree. The frequency response of a system is the steady state response of a system to a sinusoidal input. In practice a range of frequencies is applied to the system and the response recorded.

Of interest is what can happen to a sinusoidal system as it traverses a pathway. In the most general case, anything. For example the signal could change amplitude, phase or frequency and it could even change it sinusoidal shape. Although we will not give a proof here, linear systems have the special property that when subjected to a sinusoidal input, only the amplitude and phase change, the frequency of the signal remains the same. This makes the analysis much simpler but also forces us to either study pure linear systems or turn nonlinear systems into linear ones through a process of linearization.

In summary, the frequency response of a system concerns itself with how the amplitude and phase of an input signal of a given frequency propagates to all the state variables in a linear system. Often a graph is made over a range of frequencies, such a plot is called a Bode Plot.

> If a system is linear then the response of the system to a sinusoidal input is a sinusoidal output with the same frequency as the input but differing in amplitude and phase.

Figure 10: Effect of a linear system on a sinusoidal signal.

- Provides an indication of how close to instability we might be and how to move closer or further away.
- Provides an approach for measuring the degree of modularity in a circuit.
- Allows us to explain the onset of oscillations in a feedback circuit.
- Allows us to understand the properties of negative feedback in more detail.
- Gives us the machinery to reconstruct the internal structure of a system from the input/output response.
- Relates the DC component of the frequency domain to the existing field of metabolic control analysis.

3.3 Review of Complex Numbers

Imaginary numbers are solutions to equations such as $\sqrt{-x}$ and are usually represented for convenience by the symbol, $\sqrt{x}i$. Thus the imaginary number of $\sqrt{-1} = i$ and for $\sqrt{-9} = 3i$ (We ignore the fact that there are two solution, $+3i$ and $-3i$.

Although i is often used to represent the imaginary unit number, in engineering j is often used instead to avoid confusion with electrical current, i.

Imaginary numbers can also be paired up with real numbers to form **complex numbers**. Such number have the form:

$$a + bj$$

where a represents the **real part** and b the **imaginary part**. This notation is actually a short-hand for the more general statement:

$$(a, 0j) + (0, bj)$$

that is vector addition. For convenience the 0 values are omitted and the notation shortened to $a + bj$.

A conjugate complex pair is given by the pair of complex numbers:

$$a + bj \qquad a - bj$$

Polar form

We can express a complex number on a two dimension plane where the horizontal axis represents the real part and the vertical axis the imaginary part. A complex number can therefore represented as a point on the plane.

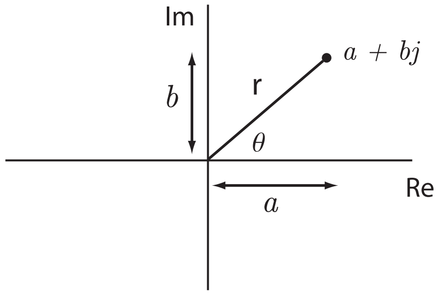

Figure 11: Argand Plane

We can also express a complex number in terms of the distance the point is away from the origin and the angle is has with respect to the horizontal axis. In this way we can express the real and imaginary parts using trigonometric functions:

$$b = r\sin\theta \qquad a = r\cos\theta$$

where r is the length of the line from the origin to the point and θ the angle. The following two representations are therefore equivalent.

$$a + bj = r(\cos\theta + j\sin\theta)$$

When written like this, r is also known as the magnitude or modulus of the complex number, A, that is:

$$|A| = r = \sqrt{a^2 + b^2}$$

The notation $|A|$ is often used to denote the magnitude of a complex number. The angle, θ is known as the argument or phase and is given by:

$$\theta = \tan^{-1}\left(\frac{b}{a}\right)$$

In calculating the angle we must be careful about the sign of b/a. Figure 12 illustrates the four possible situations. If the point is in the second quadrant, 180^o should be added to the $\tan^{-1}$ result. If the point is in the third quadrant, then 180^o should be subtracted from the $\tan^{-1}$ result. The 1st and 4th quadrants need no adjustments.

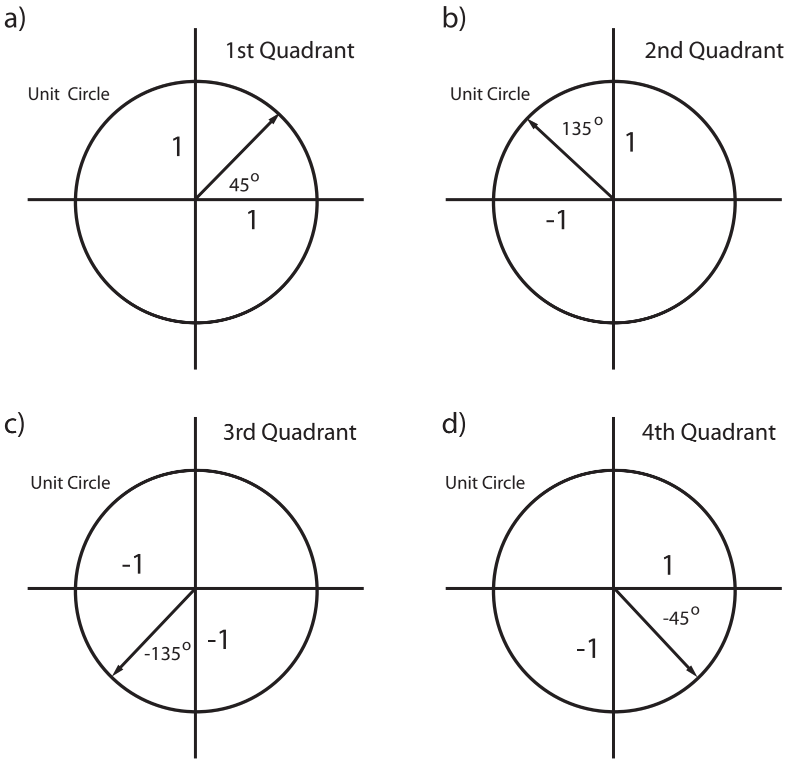

Figure 12: a) Both axes are positive, $\arctan(1/1) = 45^o$; b) Horizontal axis is negative -1, $\arctan(1/-1) = -135^o$; c) Vertical axes is negative, $\arctan(-1/1) = -45^o$; d) Both axes are negative, $\arctan(-1/-1) = -135^o$

The rules for computing the angle are summarized in the list below. The atan2 function often found in software such as Matlab will usually automatically take into consideration the signs.

$$\text{atan2}(y, x) = \begin{cases} \tan^{-1}(\frac{y}{x}) & x > 0 \\ \pi + \tan^{-1}(\frac{y}{x}) & y \geq 0, x < 0 \\ -\pi + \tan^{-1}(\frac{y}{x}) & y < 0, x < 0 \\ \frac{\pi}{2} & y > 0, x = 0 \\ -\frac{\pi}{2} & y < 0, x = 0 \\ \text{undefined} & y = 0, x = 0 \end{cases}$$

Basic Complex Arithmetic

Let $a + bj$ and $c + dj$ be complex numbers. Then:

1. $a + bj = c + dj$ if and only if $a = c$ and $b = d$ (i.e. the real parts are equal and the imaginary parts are equal)
2. $(a + bj) + (c + dj) = (a + c) + (b + d)j$ (i.e. add the real parts together and add the imaginary parts together)
3. $(a + bj) - (c + dj) = (a - c) + (b - d)j$
4. $(a + bj)(c + dj) = (ac - bd) + (ad + bc)j$
5. $(a + bj)(a - bj) = a^2 + b^2$
6. $\dfrac{a + bj}{c + dj} = \dfrac{(ac + bd) + (bc - ad)j}{c^2 + d^2}$

Division is accomplished by multiplying the top and bottom by the conjugate. Note that the product of a complex number and its conjugate gives a real number, this allows us to eliminate the imaginary part from the denominator.

$$
\begin{aligned}
\frac{a+bj}{c+dj} &= \frac{a+bj}{c+dj} \cdot \frac{c-dj}{c-dj} \\
&= \frac{(ac-b(-d)) \; + \; (a(-d)+bc)j}{c^2+d^2} \\
&= \frac{(ac+bd) \; + \; (bc-ad)j}{c^2+d^2} \\
&= \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}
\end{aligned}
$$

Exponential Form

By Euler's formula:

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

we can substitute the sine/cosine terms in the polar representation to give:

$$a+bj = re^{j\theta}$$

This gives us three ways to represent a complex number:

$$a+bj = r(\cos(\theta) + j\sin(\theta) = re^{j\theta}$$

In the previous section we used a relatively laborious approach that determined how a sinusoidal signal was changed by a linear system. For larger systems, this approach becomes too unwieldy. Instead we can get the same information by using the unilateral Fourier Transform. This transform takes a sinusoidal input signal, applies it to a linear system and computes the resulting phase and amplitude change at the given frequency of the sinusoidal input. The transform can be applied at all frequencies so that complete frequency response can be computed indicating how the system alters sinusoidal signals at different frequencies. Analytically, the unilateral Fourier Transform is given by:

$$F(j\omega) = \int_0^\infty x(t)e^{-j\omega t}dt$$

If we compare this to the Laplace transform:

$$X(s) = \int_0^\infty x(t)e^{-st}dt$$

we see that they are very similar. s in the Laplace transform is usually a complex number $\sigma + j\omega$ where the real part

In the case of the unilateral Fourier transform, $s = j\omega$. To compute the Fourier transform we can therefore take the Laplace transform and substitute s with $j\omega$. The reason this works is because the real part represents the transient or exponential decay of the system to steady state, whereas the imaginary part represents the steady state itself. When injecting a sinusoidal signal into a system we are primarily interested in the **sinusoidal steady state**, hence we can set $\sigma = 0$.

In Fourier analysis, harmonic sine and cosines are multiplied into the system function, $f(t)$ and then integrated. The act of integration picks out the strength of the response to a give frequency.

Let us use the Fourier transform to obtain the amplitude and phase change for a general linear first-order differential equation:

$$\frac{dy}{dt} + ay = f(t)$$

We will denote $\mathcal{L}(f(t))$ to means the Laplace transform of $f(t)$. The table below shows a very short list of Laplace transforms.

$f(t)$	$F(s)$
$f(t) + g(t)$	$\mathcal{L}[f(t)] + \mathcal{L}[g(t)]$
$af(t)$	$a\mathcal{L}[f(t)]$
y	$Y(s)$
dy/dt	$sY(s) + y(0)$

Taking Laplace transforms on both sides yields:

$$sY(s) + aY(s) = \mathcal{L}(f(t))$$

so that

$$Y(s) = \frac{\mathcal{L}(f(t))}{s + a}$$

The transfer function of the system, $T(s)$, is however the ratio of $\mathcal{L}$(output/(input) so that

$$T(s) = \frac{Y(s)}{\mathcal{L}(f(t))} = \frac{1}{s + a}$$

To obtain the frequency response we set $s = j\omega$:

$$T(j\omega) = \frac{1}{j\omega + a}$$

From this complex number we can compute both the amplitude and phase shift. First we must get the equation into a standard form:

$$T(j\omega) = \frac{1}{a + j\omega}\frac{(a - j\omega)}{(a - j\omega)} = \frac{a}{a^2 + \omega^2} - \frac{j\omega}{a^2 + \omega^2}$$

From this we can easily compute the amplitude change to be:

$$A = \sqrt{\frac{a^2}{(a^2+\omega^2)^2} + \frac{\omega^2}{(a^2+\omega^2)^2}} = \sqrt{\frac{1}{a^2+\omega^2}} = \frac{1}{\sqrt{a^2+\omega^2}}$$

The phase shift can be computed using $\tan^{-1}(b/a)$ so that

$$\alpha = -\tan^{-1}\left(\frac{\omega}{a}\right)$$

Note that these results are identical to the results obtained in the previous section when the differential equation was integrated directly.

In conclusion we can determine the frequency response of a linear system from the Laplace transform.

3.4 Bode Plots

Plots of amplitude and phase versus frequency are called **Bode plots**. The amplitude plot is plotted using decibels (dB) which is a logarithmic unit for expressing the magnitude of a quantity, in this case the **change** in the amplitude, given by the formula $20\log_{10}(|A|)$. The **bandwidth** of a system is the frequency at which the gain drops below the 3 db peak. This is also the frequency where the signal is $1/\sqrt{2}$ of the maximum signal amplitude (about 70% of the signal strength).

Note that the phase shift starts at zero. That is as the frequency is reduced the phase shift gets smaller. At high frequencies the phase shift tends to -90^o. The way to understand this is to look at the phase expression and realize that the smaller k_2 the more likely the phase shift will be -90^o. If k_2 is very small then we can assume there is very little degradation flux, this means that the change in s is dominated by the input sine wave. The maximum rate of increase in s is when the sine wave is at its maximum peak. As the input sine

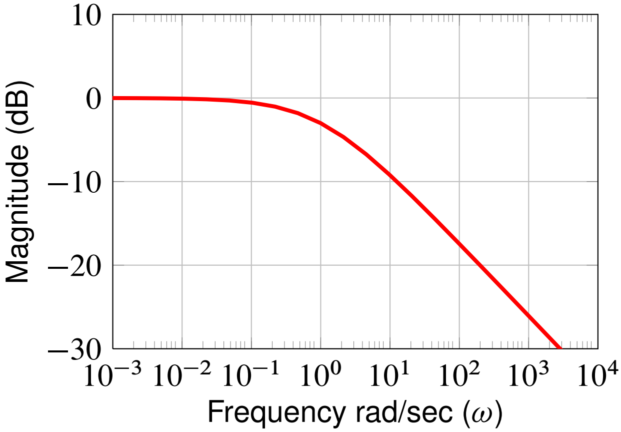

Figure 13: Bode Plot: Magnitude (dB)

wave decreases the rate of increase in s slows until the input sine wave crosses the steady state level of s. Once the input sine wave reaches the steady state level, the level of s also peaks. Thus the input sine wave and the concentration of s will be -90^o out of phase with the concentration of s lagging. The frequency point at which the phase reaches -90^o will depend on the k_2 value. Figure 15 illustrates the phase shift argument. Also note that the amplitude tends to zero as the frequency is increased.

One question that remains is what is the amplitude change at zero frequency?

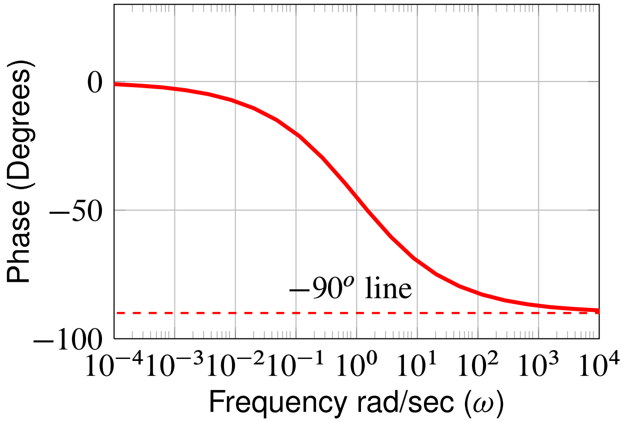

Figure 14: Bode Plot: Phase Shift

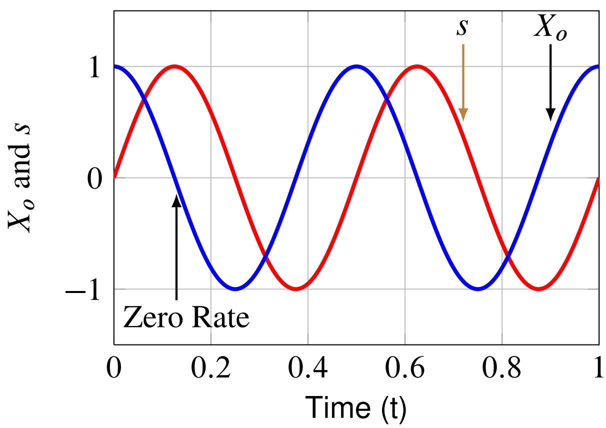

Figure 15: -90^o phase shift at high frequencies or low degradation rates. Note that the concentration of s peaks when the input rate is at zero.