Linear Systems and Control

Systems Overview

Rev 1.1

July 14, 2024

Thomas B Fetter

Table of Contents

1 Introduction

This paper is a short overview of the mathematical tools and methods that are used throughout the papers by this author on the topics of rocket rotational dynamic stability, rocket trajectory analysis, and vertical trajectory control systems. The concepts come from both mathematics and engineering. This paper is intended as a brief introduction for someone new to these topics, or as a quick review for someone who has studied these topics. By showing the context of how these concepts are useful in modeling real systems, the intension is to provide the motivation to study these topics further for someone not familiar with them.

The first section is an overview of linear systems; physical systems that can be described by linear algebraic and differential equations and the tools and methods used to analyze their behavior. There are many math texts that cover differential equations[1], and engineering texts that cover the linear systems concepts that apply specifically to engineering[2], in much greater depth than is presented here that can be used for further study.

The second section is an overview of control system theory. This is used in the design of any form of flight control system such as a canard based vertical trajectory system or a thrust vector system. For now, this section only covers high gain type control loops. An overview of state space controllers, frequently used in aerospace engineering control systems, will be added later. Again, other texts cover these topics in much greater detail for further study[3].

Feedback Control of Dynamic Systems, by Franklin, Powell, and Emami-Naeini[4], of all the cited texts, probably has the most comprehensive coverage of the topics presented in this paper. It covers linear systems, high gain control systems, and state-space control systems very thoroughly. It takes a practical and easy to understand engineering approach rather than a more theoretical mathematical approach and is an excellent starting point for further study.

---

[1] (Boyce & DiPrima, 1969)

[2] (Liu & Liu, 1975)

[3] (Roberge, 1975), (Chen, 1993), (Friedland, 1986), (Franklin, 2019), (Astrom, 2011)

[4] (Franklin, 2019)

2 Linear Systems

2.1 Introduction

Linear systems refer to systems that can be described by linear differential and algebraic equations. A linear system is one where a linear combination of inputs produces a linear combination of outputs (superposition) and a scaling of the input produces the same scaling of the output (homogeneity). Although the series of differential equations that model a rocket's rotational dynamics and flight trajectory are nonlinear, the equations can be linearized about certain operating points and then analyzed using the tools and methods described in this section. Without being linearized, the nonlinear system can only be solved numerically, which does completely describe the state of the system at any point in time, but without the insight that can be learned from a linearized closed-form solution. Also, many of the tools and methods used to design control systems (see Section 3) are based on linear system models.

2.2 Linear Systems and the Time and Frequency Domains

Differential equations are used to model electrical and mechanical systems. To model the flight of a rocket, a system of differential equations is used to both describe the location of the center of mass of the rocket along its flight path, as well as the orientation of the rocket as it rotates about its center of gravity.

A differential equation, rather than a simple algebraic equation, is required to describe any system where the current behavior depends upon something that happened at an earlier point in time. A differential equation has memory of a prior state that it uses to calculate the current state of the system. If the system started at some non-zero initial condition, the complete solution of the differential equation then describes the state of the object over time in response to that initial condition.

As an example, starting with Newton's Law, $F = m \cdot a$, the differential equation that describes the position of a mass on a spring can be constructed. There are two forces that act on the mass, the linear displacement force, which is the displacement of the mass times the spring constant, K, and the damping frictional force, which is the velocity of the mass times the coefficient of friction, CF. The frictional damping is due to resistive losses in the spring, as well as the drag force on the mass object due to the air. Velocity is the derivative of position, $v(t) = dx(t)/dt$, and acceleration is the derivative of velocity, or second derivative of position, $a(t) = dv(t)/dt = d^2x(t)/dt^2$.

$$F_{forcing}(t) + F_{damping}(t) = m \cdot a(t)$$

$$F_{forcing}(t) = K \cdot x(t)$$

$$F_{damping}(t) = CF \cdot v(t)$$

$$\Rightarrow K \cdot x(t) + CF \cdot \frac{dx(t)}{dt} = m \cdot \frac{d^2x(t)}{dt^2} \tag{2.1-1}$$

The solution to Equation (2.1-1) describes the position, $x(t)$, of the mass over time. If the spring and mass is compressed to a distance of 6 units by an external force, so that $x(0) = 6$, the initial condition of the system, and then released, Figure 2-1 shows the result is a sinusoid starting at 6 units, that diminishes in amplitude over time, due the damping force.

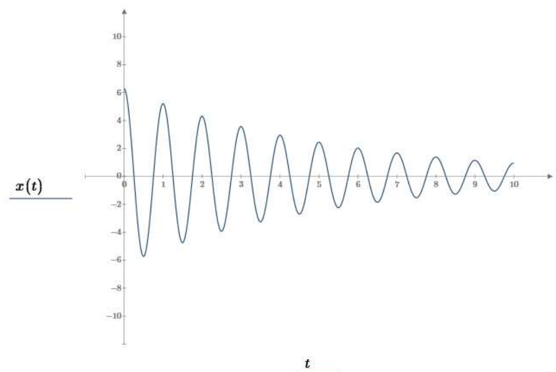


Figure 2-1 Position of a mass on a spring

This is a second order differential equation because the highest order derivative is the second order derivative of the position of the mass object, $x(t)$. The second order differential equation describes many physical systems. The motion of any system containing a mass, a spring, and damping, can be described by this equation. In electronics, an inductor, capacitor, and resistor forming a resonant circuit can be described by a second order differential equation. And several second order differential equations will be used in the rocket flight dynamics model, each one to describe one degree of the rocket's motion.

Because the differential equation describes the state of the system over time, it is called a time-domain equation. Time, *t*, is the independent variable in the time domain.

If the only input to the system described by a differential equation occurs at $t = 0$, like the example above with the mass on a spring that is pulled 6 inches and then released at $t = 0$, then the differential equation can be solved using an initial condition. But a physical system can also have a continuous input, which is a driving term. The differential equation describing this system is written in terms of a driving, or input variable on one side of the equals sign, and the differential equation that describes the response on the other side of the equals sign that includes the desired output variable, where both the input and output variables are parametric functions of time. The general form of the second order differential equation for a system with a continuous driving function is

$$C_3 \cdot \frac{d^2\alpha_{out}(t)}{dt^2} + C_2 \cdot \frac{d\alpha_{out}(t)}{dt} + C_1 \cdot \alpha_{out}(t) = C_0 \cdot \alpha_{in}(t) \tag{2.1-2}$$

where $\alpha_{in}(t)$ is the input, and $\alpha_{out}(t)$ is the resulting output. The input can be any function in time. The solution of the differential equation then describes the system's response to the input. In the solution, $\alpha_{out}(t)$ will try to follow $\alpha_{in}(t)$, but with the dynamics determined by the derivative terms in the second order equation. If $\alpha_{in}(t)$ is a transient signal, like a step function that takes on a constant value, then $\alpha_{out}(t)$ will eventually take on the value $C_0/C_1 \cdot \alpha_{in}$. If the input function is a just step at $t = 0$ that goes from a starting value for the input variable to 0, then the resulting solution will be the same as the system with the input set to zero and an initial starting condition for the output variable.

The input and output do not need to have the same units. For example, the input could be a linear displacement of an object and the output the angular position of a rotating shaft. If they are not the same units, the coefficients will reflect the appropriate units to balance the units of the overall equation.

C_0 and C_1 are called the forcing coefficients because, in a mechanical system, they describe the magnitude of the force that drives the output. C_2 is called the damping coefficient. It describes the magnitude of the first derivative of the output. In a second order equation, this term has the effect of damping the amplitude of the output of the overall system.

Not all differential equations, including most non-linear differential equations, can be solved in closed form. For those systems, numerical methods can be used to solve for the system response. But if all the coefficients are constants, then equation (2.1-2) is a linear differential equation. The easiest way to solve a linear differential equation of this form is to use the Laplace transform[5] The Laplace transform transforms the linear differential equation into a linear algebraic equation that can easily be solved. The inverse Laplace transform is then applied to the result to give a solution that describes the system's state over time. If the system is described by a nonlinear differential equation where the coefficients are not constant but vary with time or some other parameter, the equation can often be linearized for a small region about an operating point, and an acceptable closed form solution can be found using the Laplace transform. Or a numerical method can be used to solve the system of nonlinear differential equations directly. A numerical solution can completely describe the response of a system, but a closed

[5] (Boyce & DiPrima, 1969, pp. 222-258).

form solution, even if it is approximate, can sometimes give more insight into how each of the model's parameters affects the behavior of the system.

The equation that results from applying a Laplace transform to a time domain differential equation has a physical meaning as well. The Laplace transform of a time domain system is a frequency domain system. As the time domain describes the output of a system for a given input versus time, the frequency domain describes the output of a system for a given input versus frequency. For an input that is a single frequency sinusoidal signal, the frequency response will give the magnitude and phase of the output relative to the input. For a linear system, the frequency of the output will always be the same frequency as the input, but its magnitude and relative phase can be different, so two numbers are required to represent the output, a gain and a phase, and those number can be different for each frequency of the sinusoidal input. More complex input signals can be represented as a sum of individual sinusoidal signals at multiple frequencies.

The frequency domain equation has an independent variable that is complex frequency, *s*, which is the frequency multiplied by the imaginary number *i*. The letter *j* is often used in place of *i* in engineering, so $s = j\omega$, where ω is frequency in radians/sec. The quantity *i*, or *j*, is the square root of -1. The set of imaginary numbers are separate from the set of real numbers. A complex number has both a real and imaginary part, so they can be used as a convenient mathematical notation to represent a quantity that has two separate parts. In this case, for the frequency domain, the two parts represent the magnitude and a phase of a sinusoid at each frequency. Because the real and imaginary parts of a complex number are independent, they are said to be orthogonal. When representing a complex number on a graph, the real part is usually plotted along the x-axis, and the imaginary part is plotted along the y-axis, as will be seen in section 2.5.

When taking the Laplace transform of a linear differential equation, each of the terms is replaced by the Laplace transform of that term. The Laplace transform of $\alpha(t)$ is

$$\mathcal{L}\left[\alpha(t)\right] = \alpha(s) \qquad (2.1\text{-}3)$$

The Laplace transform for a first derivative is

$$\mathcal{L}\left[\frac{d\alpha(t)}{dt}\right] = s \cdot \alpha(s) \qquad (2.1\text{-}4)$$

and the transform of a second derivative is

$$\mathcal{L}\left[\frac{d^2\alpha(t)}{dt^2}\right] = s^2 \cdot \alpha(s) \tag{2.1-5}$$

Taking the Laplace transform of the second order equation (2.1-1) by replacing the terms with their transforms results in

$$\mathcal{L}\left[C_3 \cdot \frac{d^2\alpha_{out}(t)}{dt^2} + C_2 \cdot \frac{d\alpha_{out}(t)}{dt} + C_1 \cdot \alpha_{out}(t) = C_0 \cdot \alpha_{in}(t)\right]$$

$$\Rightarrow C_3 \cdot s^2 \cdot \alpha_{out}(s) + C_2 \cdot s \cdot \alpha_{out}(s) + C_1 \cdot x_{out}(s) = C_0 \cdot \alpha_{in}(s) \tag{2.1-6}$$

The frequency domain equation is a linear algebraic equation, so the output variable, $\alpha_{out}(s)$ can be factored out, and the equation can be written in the form

$$G(s) = \frac{\alpha_{out}(s)}{\alpha_{in}(s)} = \frac{C_0}{C_1} \cdot \frac{1}{\frac{C_3}{C_1} \cdot s^2 + \frac{C_2}{C_1} \cdot s + 1} \tag{2.1-7}$$

where the equation is the ratio, or gain, between the input and the output. Because of the complex frequency, the gain function has both a magnitude and a phase.

Gain functions can be used to build up complex systems of differential equations in the frequency domain using a block diagram representation. In the frequency domain, each block simply multiplies the signal coming from the previous block. A block can also be an input signal for certain types of signals. The corresponding operation of multiplying in the frequency domain is convolution in the time domain, a more complicated and less intuitive mathematical operation. Therefore, the design of complex mechanical and electrical systems is typically done in the frequency domain, and then the final time domain result is found by taking the inverse Laplace transform of the output as the last step.

The gain is function of the complex frequency $s = j\omega$, so the gain has both a magnitude and a phase, which is the physical meaning of the mathematical notation of complex frequency. The magnitude of the gain function is the ratio of the amplitude of the sinusoidal output signal to the amplitude of a sinusoidal input signal at each frequency, and the phase is the phase shift between the input and output sinusoid at each frequency.

The magnitude and phase are found from the real and imaginary parts of the complex function by

$$\mathrm{Mag}\left[G(s)\right] = \sqrt{\mathrm{Re}\left[G(s)\right]^2 + \mathrm{Im}\left[G(s)\right]^2} \tag{2.1-8}$$

$$\text{Phase}\left[G(s)\right] = \text{atan}\left(\frac{\text{Im}\left[G(s)\right]}{\text{Re}\left[G(s)\right]}\right) \tag{2.1-9}$$

where the phase is in radians. Figure 2-2 shows a typical frequency domain magnitude and phase plot for a gain function. When the log of the magnitude is plotted against the log of the frequency and linear phase is plotted against the log of the frequency, this is called a Bode plot.

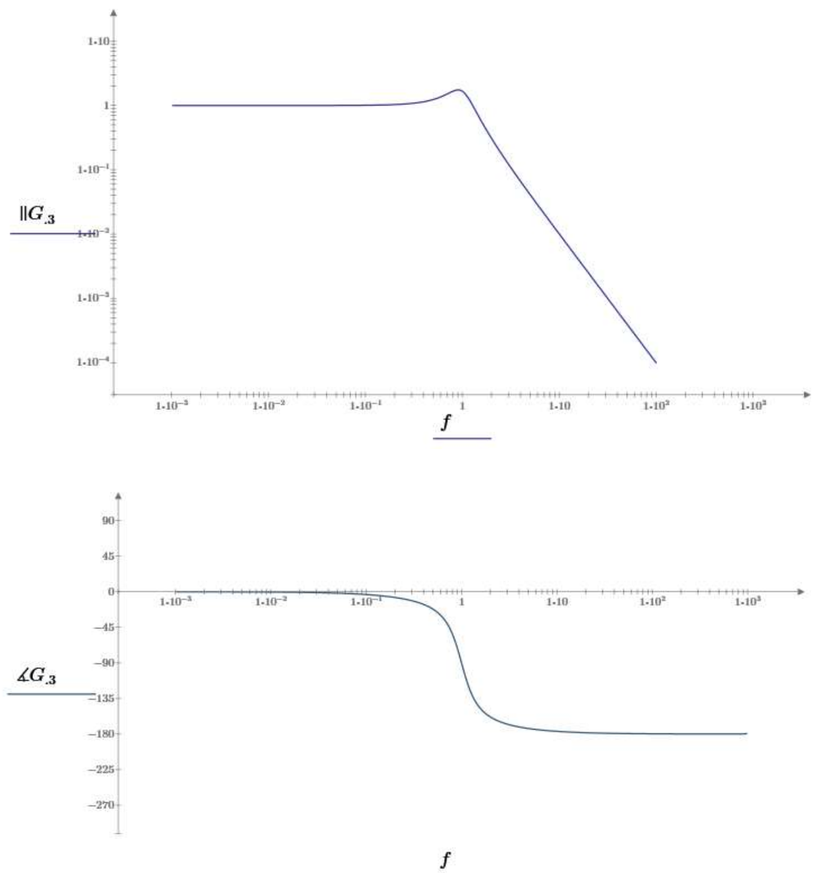


Figure 2-2 Typical Bode plot - frequency domain magnitude and phase plot

Since the gain function is the ratio of the input to the output, it does not, by itself, provide the solution to the response to a specific input waveform, other than for purely sinusoidal signals. For signals for which a Laplace transform exists, the frequency domain transform of that signal can be multiplied by the gain function, and the inverse Laplace transform of that product can be taken to determine the time domain response for that waveform.

An impulse function is one type of signal that can be used as the input to characterize a system. It is the equivalent of striking a bell with a hammer. The resulting ringing is the impulse response of the bell. The frequency and the length of the ringing describes the characteristics of the bell. An impulse is zero for all time except for the time of the impulse, where the amplitude is infinite. The area, amplitude times duration, of a unit impulse is defined as 1. Having a finite area in zero time is a mathematical construct that can only be approximated in the real world.

$$impulse(t) = 0\big|_{t<0}, \infty\big|_{t=0}, 0\big|_{t>0} \tag{2.1-10}$$

The Laplace transform of an impulse is 1, so, in the frequency domain, it has equal amplitude at all frequencies, or a flat frequency response. That makes it good for characterizing a system because it stimulates the system at all frequencies with equal amplitude.

$$Impulse(s) = \mathcal{L}\left(impulse(t)\right) = 1 \tag{2.1-11}$$

The response to an input signal is determined by multiplying the input by the gain of the system in the frequency domain. Since the impulse function is 1 in the frequency domain, the gain of the system is the impulse response. And the inverse Laplace transform of the gain function is the time domain impulse response. And conversely, the impulse response is the gain of the system.

$$\mathcal{L}^{-1}\left(1 \cdot G(s) \cdot\right) = \alpha_{out}(t)\big|_{impulse} \tag{2.1-12}$$

A time domain impulse function can be used to determine the gain function of a system. Figure 2-3 shows a typical time domain impulse response for a second order system. Since the long-term value of the impulse is zero, the long-term value of the impulse response is zero.

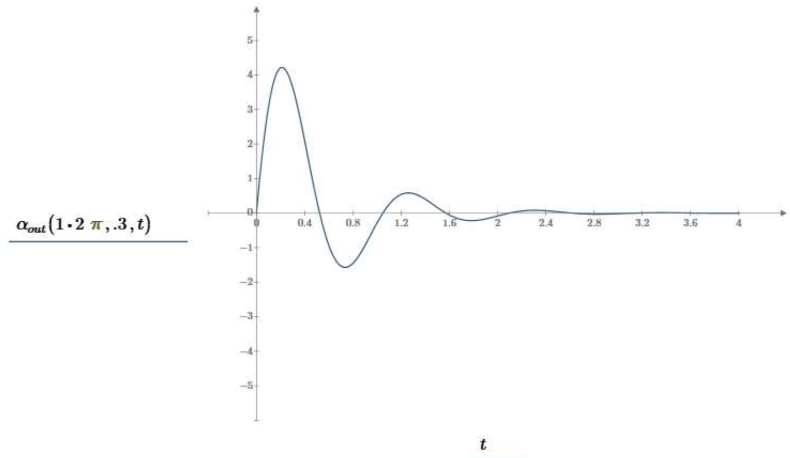


Figure 2-3 Typical impulse response

In the real world, an impulse must be approximated by a pulse function of finite duration and amplitude, which is a more complicated function in the frequency domain than a true impulse. A step function, which is half of a finite duration pulse, is mathematically simpler and is easy to generate. A step function is a commonly used signal to characterize a system in the time domain. A unit step function has a value of 0 for all time < 0, and a value of 1 for all time ≥ 0.

$$step(t) = 0\big|_{t<0}, 1\big|_{t\geq 0} \tag{2.1-13}$$

The Laplace transform of a step function is

$$Step(s) = \mathcal{L}\left(step(t)\right) = \frac{1}{s} \tag{2.1-14}$$

which is a function that rolls off in amplitude with frequency. Most real-world systems are low pass in nature, so the roll off does not generally limit the use of the step function in characterizing a system.

Multiplying the gain function by the frequency domain step, $1/s$, and taking the inverse Laplace transform gives the time domain step response of the system

$$\mathcal{L}^{-1}\left(\frac{1}{s} \cdot G(s) \cdot \right) = \alpha_{out}(t)\big|_{step} \tag{2.1-15}$$

The step response fully characterizes a system as does the impulse response. Figure 2-4 shows the typical time domain step response for a second order system. Since the long-term value of the unit step is 1, the long-term value of the impulse response is 1.

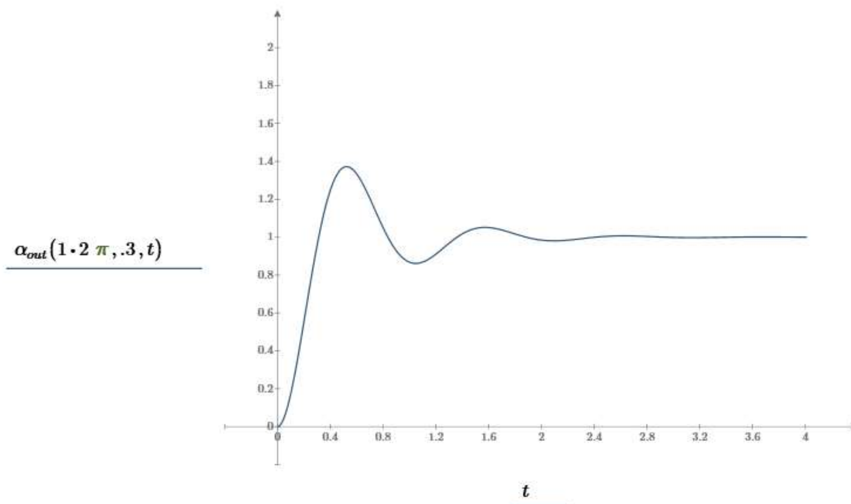


Figure 2-4 Typical time domain step response

The frequency response of a system can also be measured by stimulating the system with a sinusoidal signal and measuring the relative amplitude and phase of the output over all the frequencies of interest.

A rocket sees a step in the lateral wind velocity as it leaves the launch guide, so the step function is the natural choice to use as the input to the rotational system representing the velocity of the wind.

For a system described by linear differential equation, going from the time domain to the frequency domain is straight forward by replacing each n^{th} order derivative with s^n. Many engineering approaches start with a frequency domain design because of the ease of working with linear algebraic equations and gain functions. Getting back to the time domain by taking the inverse Laplace transform is typically more tedious. The result of building up a complex frequency domain system out of many gain blocks means the inverse Laplace transform rapidly becomes very large, containing many terms, as the order of the system solution increases. Traditionally, for first and second order systems, known relationships between the frequency domain parameters and time domain step response were used to relate the time and frequency domain responses. But now, the symbolic processing capabilities of software like Mathcad[6] is powerful enough to solve the inverse Laplace transform for more complex systems symbolically or

[6] (Mathcad Home Page, n.d.)

for very complex systems, numerically. For systems still too complicated for Mathcad to solve, simplifying assumptions can often be made that allow Mathcad to approximate the solution adequately.

2.3 First Order System Response

A first order system has a first order derivative as its highest derivative. Starting with the general form of the first order differential equation

$$C_2 \cdot \frac{d\alpha_{out}(t)}{dt} + C_1 \cdot \alpha_{out}(t) = C_0 \cdot \alpha_{in}(t) \tag{2.2-1}$$

then, taking the Laplace transform

$$C_2 \cdot s \cdot \alpha_{out}(s) + C_1 \cdot \alpha_{out}(s) = C_0 \cdot \alpha_{in}(s) \tag{2.2-2}$$

Rearranging terms, the frequency domain gain equation is

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\left(\frac{C_2}{C_1} \cdot s + 1\right)} \tag{2.2-3}$$

Equation (2.2-3) can be also be written in terms of one constant, the corner frequency ω_c where

$$\frac{C_2}{C_1} = \frac{1}{\omega_c} \tag{2.2-4}$$

equation (2.2-3) becomes

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\left(\frac{s}{\omega_c} + 1\right)} \tag{2.2-5}$$

Figure 2-5 shows the Bode plot of the magnitude and phase of equation (2.2-5) for $\omega_c = 1 \cdot 2\pi \cdot rad/\sec$ (1 Hz in radians/sec). As the frequency approaches the corner frequency, the magnitude starts to roll off. At the corner frequency, the gain is 0.707, which is the half power point, or 3dB point, where the gain in dB = $20 \cdot \log\left(\alpha_{out}/\alpha_{in}\right)$.

The bandwidth of a system is defined as the frequency of the 3dB point, or between the two 3dB points for a bandpass system. The gain rolls off at 20dB (a factor of 10) per decade in frequency above the corner frequency. The phase starts to roll off at 1/10th the corner frequency, reaches 45 deg at the corner frequency, and has a final value of 90 deg at 10 times the corner frequency.

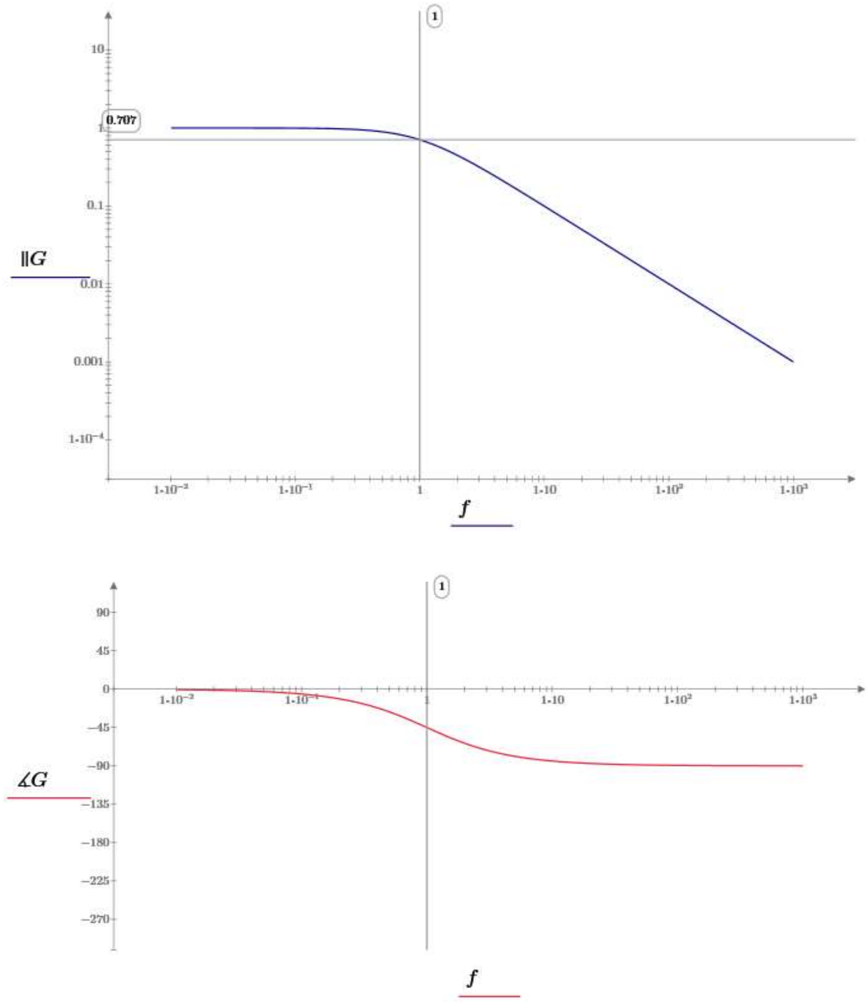


Figure 2-5 Magnitude and phase of the frequency response of a first order system for $C_1 = 1$ and $\omega_c = 1 \cdot 2\pi$

As shown in equation (2.1-15), the time domain step response of a system is found by taking the inverse Laplace transform of the frequency domain step function, $1/s$, multiplied by the frequency domain gain function. The step response for the first order equation (2.2-5) is

$$\mathcal{L}^{-1}\left(\frac{1}{s}\cdot G(s)\right)=\mathcal{L}^{-1}\left(\frac{1}{s}\cdot\frac{C_0}{C_1}\cdot\frac{1}{\left(\dfrac{s}{\omega_c}+1\right)}\right)=\alpha_{out}(t)\Big|_{step}=\frac{C_0}{C_1}\cdot\left[1-e^{-\omega_c\cdot t}\right] \qquad (2.2\text{-}6)$$

which results in an exponential with a time constant of $\tau = 1/\omega_c$.

Figure 2-6 shows the plot of the step response for the first order system. The 10% to 90% rise time of a first order system step[7] is $t_r = 2.2/\omega_c$, so for this system $t_r = 2.2/1\cdot 2\pi = 0.35\sec$.

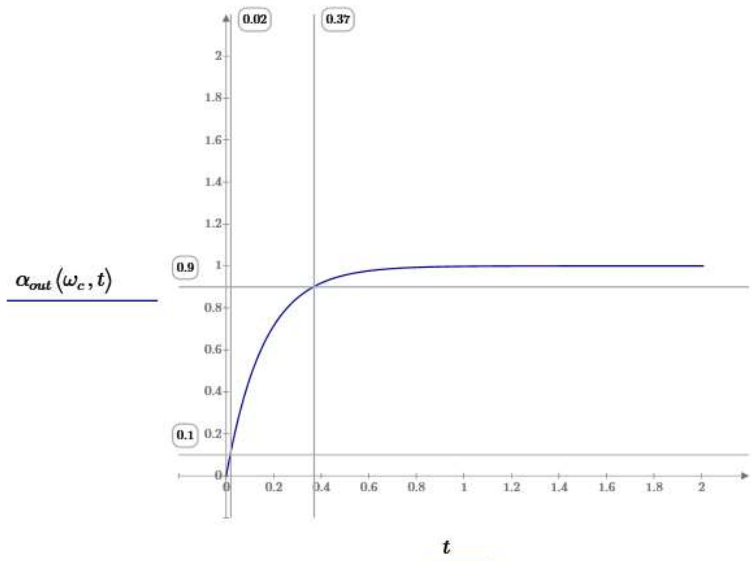


Figure 2-6 Step response of a first order system for $\dfrac{C_0}{C_1} = 1$ and $\omega_c = 1\cdot 2\pi$

2.4 Second Order System Response

A second order system has both a first order and second order derivative. Starting with the general form of the second order differential equation

---

[7] (Roberge, 1975, p. 95)

$$C_3 \cdot \frac{d^2\alpha_{out}(t)}{dt^2} + C_2 \cdot \frac{d\alpha_{out}(t)}{dt} + C_1 \cdot \alpha_{out}(t) = C_0 \cdot \alpha_{in}(t) \tag{2.3-1}$$

then, taking the Laplace transform, the frequency domain gain equation is

$$C_3 \cdot s^2 \cdot \alpha_{out}(s) + C_2 \cdot s \cdot \alpha_{out}(s) + C_1 \cdot \alpha_{out}(t) = C_0 \cdot \alpha_{in}(t) \tag{2.3-2}$$

In this equation C_0 and C_1 are the forcing coefficients, as the 0th order terms of the equation drive the system response, and C_2 is the damping coefficient, as the 1st order term dampens the system response.

Rearranging terms

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\frac{C_3}{C_1} \cdot s^2 + \frac{C_2}{C_1} \cdot s + 1} \tag{2.3-3}$$

Equation (2.3-3) can be also be written in terms of two constants, the natural frequency ω_n and the damping ratio, ζ

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\frac{s^2}{\omega_n^{\ 2}} + \frac{2 \cdot \zeta \cdot s}{\omega_n} + 1} \tag{2.3-4}$$

where:

$$\omega_n = \sqrt{\frac{C_1}{C_3}} \tag{2.3-5}$$

and

$$\zeta = \sqrt{\frac{C_2^{\ 2}}{4 \cdot C_1 \cdot C_3}} \tag{2.3-6}$$

The time domain equation in terms of the natural frequency ω_n and the damping ratio ζ is

$$\frac{1}{\omega_n^{\ 2}} \cdot \frac{d^2\alpha_{out}}{dt^2} + \frac{2 \cdot \zeta}{\omega_n} \cdot \frac{d\alpha_{out}}{dt} + \alpha_{out} = \frac{C_0}{C_1} \cdot \alpha_{in} \tag{2.3-7}$$

Equation (2.3-6) shows that the damping ratio, ζ , is proportional to the damping coefficient, C_2 , and inversely proportional to the square root of the driving coefficient, C_1 , and C_3 . If the damping coefficient increases while C_1 , and C_3 remain constant, the damping ratio will increase by the same factor as the damping coefficient.

The coefficients of the equation are grouped in terms of ω_n and ζ because these two parameters independently describe the shape of both the frequency and time domain responses of the second order system. Figure 2-7 shows plots of the magnitude and phase for two values of ω_n , $1Hz$ and $2Hz$. This is a low-pass filter response where ω_n is the asymptotic corner frequency.

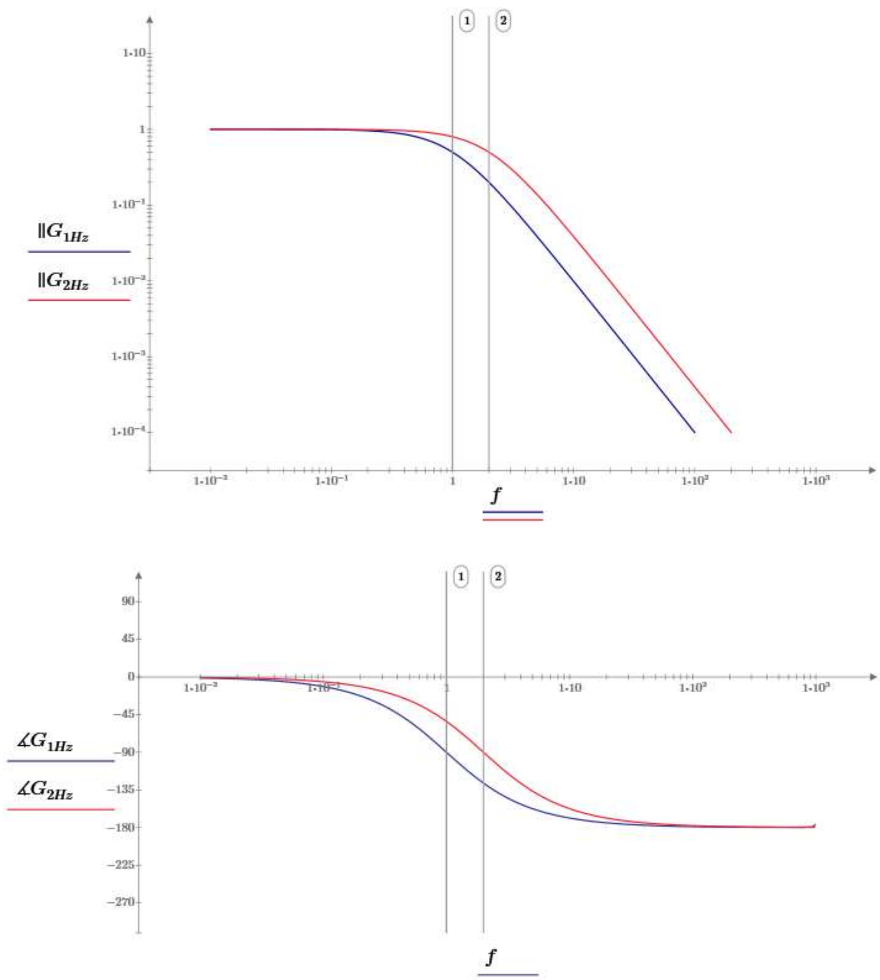


Figure 2-7 Magnitude and phase of the frequency response of a second order system for $\omega_n = 1Hz \,\&\, 2Hz$ and $\zeta = 1$

Figure 2-8 shows the frequency response of the second order equation for ζ equal to 3, 1, 0.3, and 0.03. As the value of ζ is decreased below a value of 1, the gain of the function increases above a gain of 1 at the equation's resonant, or natural frequency ω_n. This is the frequency at which the second order system oscillates in the time domain in response to a perturbation such as an impulse or step function.

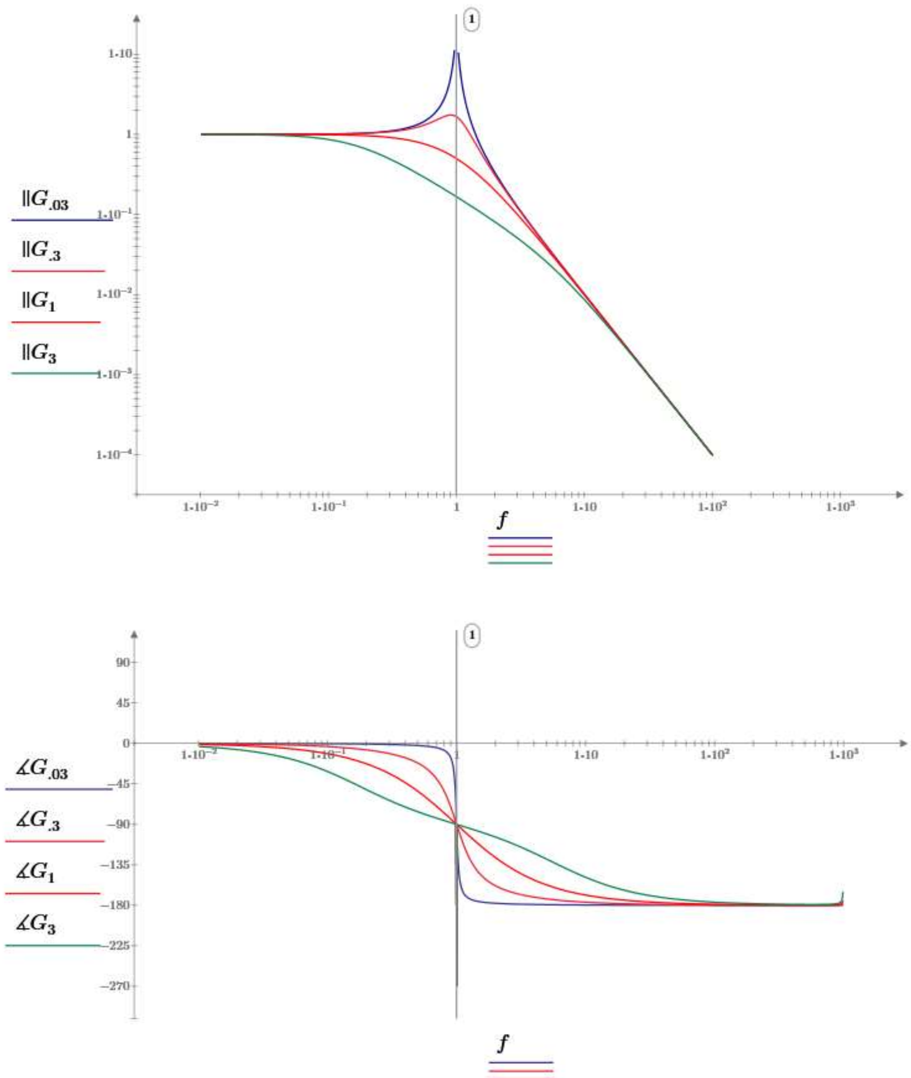


Figure 2-8 Magnitude and phase response of the second order system for $\omega_n = 1Hz$ and $\zeta = .03, .3, 1, 3$

The gain peaking causes the 3dB bandwidth of the system to increase with decreasing ζ , even though the asymptotic corner frequency, ω_n remains constant. The 3dB bandwidth can be calculated from the natural frequency and the damping ratio by[8]

$$\omega_{3dB} = \omega_n \cdot \sqrt{1 - 2 \cdot \zeta^2 + \sqrt{2 - 4 \cdot \zeta^2 + 4 \cdot \zeta^4}} \tag{2.3-8}$$

Figure 2-9 shows the 3dB bandwidth of a second order system, for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$, equals 1.45 Hz

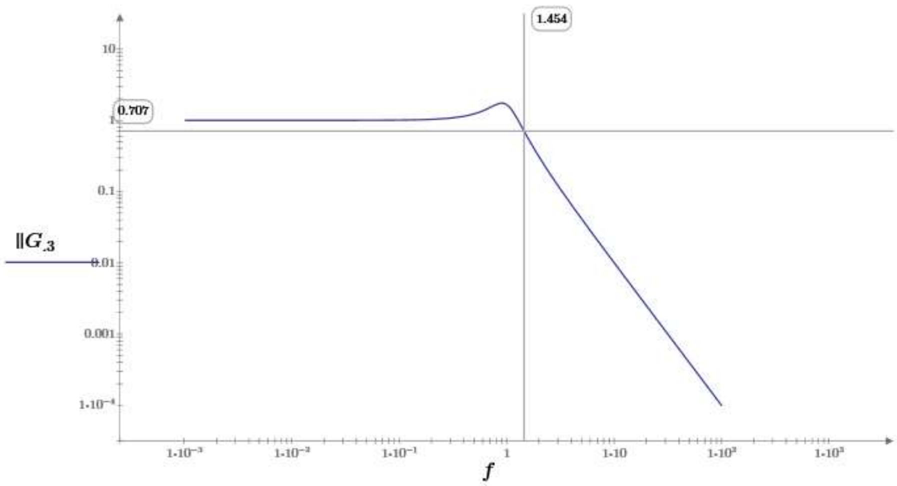


Figure 2-9 3dB bandwidth of a second order system for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$

As shown in equation (2.1-15), the time domain step response of a system is found by taking the inverse Laplace transform of the frequency domain step function, $1/s$, multiplied by the frequency domain gain function. The step response for the second order equation (2.3-4) , assuming the gain $C_0/C_1 = 1$, is

---

[8] (Roberge, 1975, p. 95)

$$\mathcal{L}^{-1}\left(\frac{1}{s}\cdot G(s)\right)=\mathcal{L}^{-1}\left(\frac{1}{s}\cdot\frac{1}{\left(\frac{s^2}{\omega_n{}^2}+\frac{2\cdot\zeta\cdot s}{\omega_n}+1\right)}\right)=$$

$$\alpha_{out}(t)\Big|_{step}=\left(-\frac{\zeta\cdot\omega_n}{\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}}\cdot e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\sin\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)-e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\cos\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)+1\right)$$

(2.3-9)

Figure 2-10 shows the second order system step response for equation (2.3-9) for ζ equals 3, 1 (critically damped), 0.3, and 0.03. The step response with the largest sinusoidal ringing is for a ζ value of .03, while the slowest response with the no ringing is for a ζ value of 3. The amplitude of the decaying oscillations follows an inverse exponential envelope that can be seen in equation (2.3-9). The response of critically damped the system with a ζ value of 1 has the fastest rise time response, for the given value of ω_n, with no overshoot or ringing.

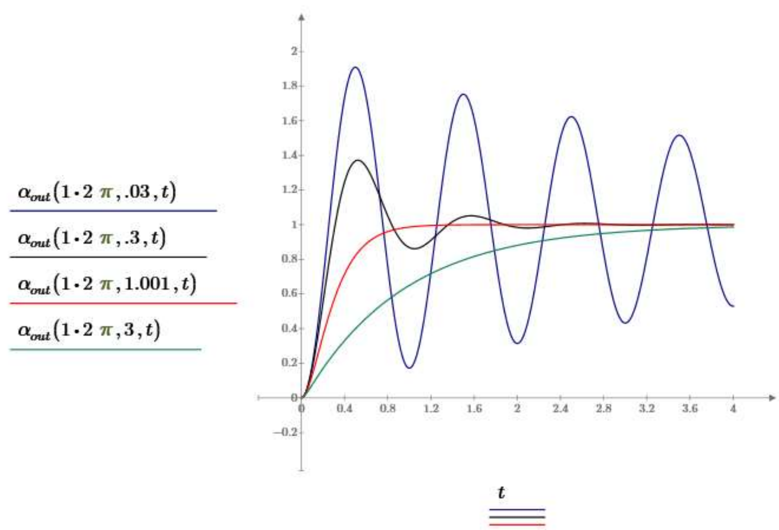


Figure 2-10 Second order system step response for $\omega_n = 1Hz$ and $\zeta = .03, .3, 1, 3$

Increasing the resonant frequency, ω_n, while holding ζ constant, decreases the rise time and increases the frequency of the ringing, but does not impact the overshoot or number of cycles of oscillation in the decay envelop, as shown in Figure 2-11. The slower step response is for $\omega_0 = 2\pi \cdot 1Hz$, while the faster response is for $\omega_n = 2\pi \cdot 2Hz$. In both cases, $\zeta = 0.3$.

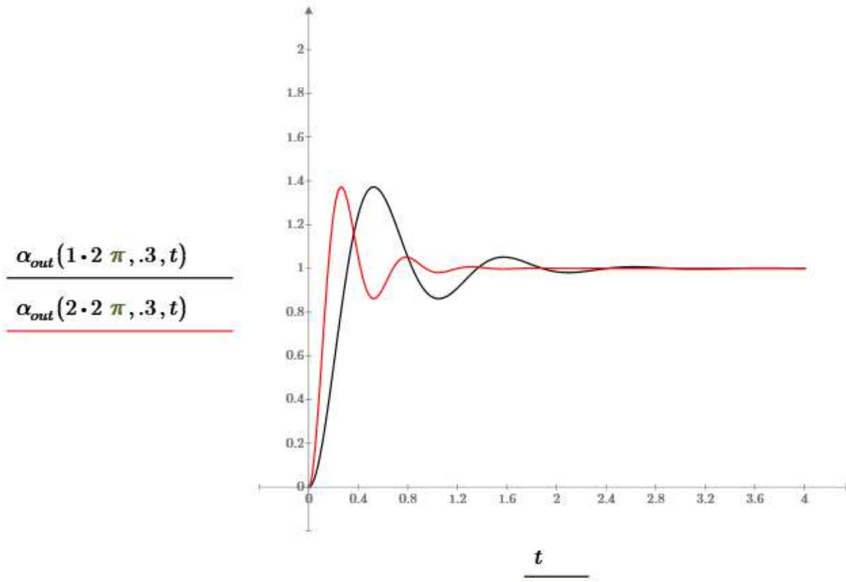


Figure 2-11 Second order system for $\omega_n = 1Hz \,\&\, 2Hz$ and $\zeta = .3$

The rise time of a second order system is defined as the time between the step response reaching 10% and 90% of the final value, and is approximated by[9]

$$t_{rise} \simeq \frac{2.2}{\omega_{3dB}} = \frac{2.2}{\omega_n \cdot \sqrt{1 - 2 \cdot \zeta^2 + \sqrt{2 - 4 \cdot \zeta^2 + 4 \cdot \zeta^4}}} \qquad (2.3\text{-}10)$$

Figure 2-12 shows the 10-90% rise time of a second order system for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$ is approximately 0.24 seconds as calculated by equation (2.3-10).

---

[9] (Roberge, 1975, p. 95)

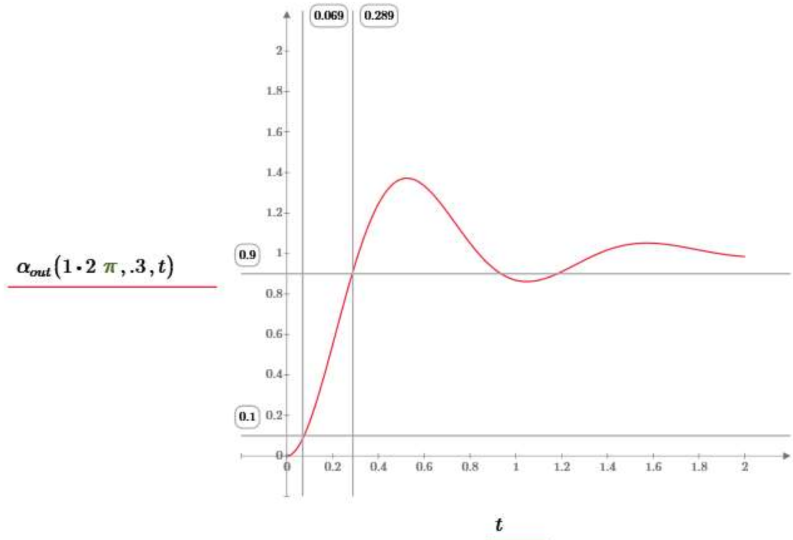
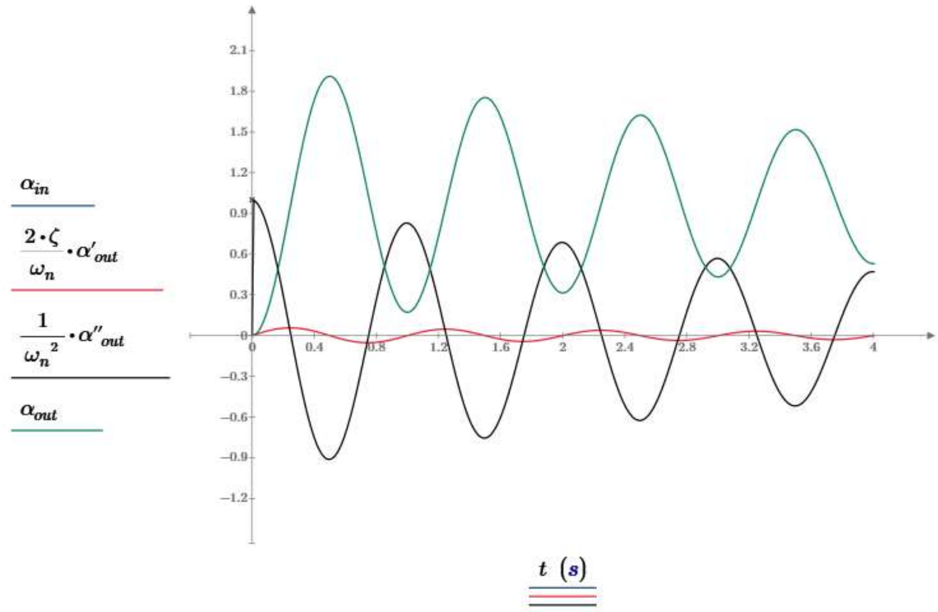
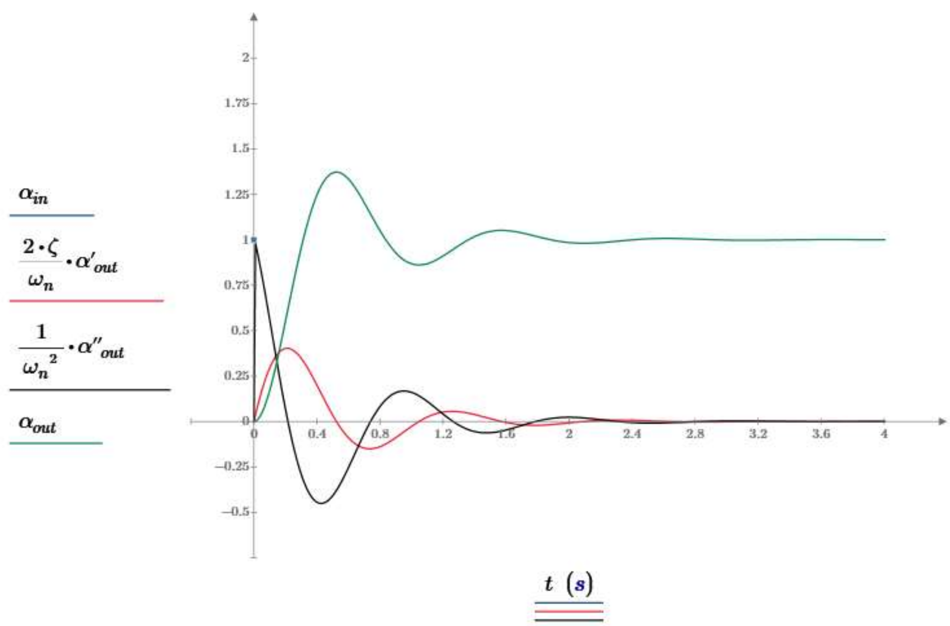


Figure 2-12 10% to 90% rise time of a second order system for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$

Figure 2-13 shows each of the terms of the second order differential equation (2.3-7) for two different values of the damping ratio, $\zeta = 0.03$, and $\zeta = 0.3$, where the input, α_{in}, is a unit step at $t = 0$. The first order damping term leads the forcing term by 90 degrees because α'_{out} is the derivative, or slope of α_{out}. The second order term, in turn, leads the first order term by another 90 degrees, because α''_{out} is the derivative, or slope of the first order term α'_{out}. Comparing the two graphs, the ringing in the step response decays faster when the first order damping term, (red trace), is initially larger compared to the other terms, due to the larger damping ratio. This effect will be seen when analyzing the rotational dynamic stability of a rocket.



$\omega_n = 1 \cdot 2\pi$, $\zeta = 0.03$



$\omega_n = 1 \cdot 2\pi$, $\zeta = 0.3$

Figure 2-13 Plotting the terms of the second-order equation for two different damping ratios

Note: $d\alpha_{out}/dt = \alpha'_{out}$ and $d^2\alpha_{out}/dt^2 = \alpha''_{out}$ in this graph

2.5 Roots of the Frequency Domain Equation

The roots of a differential equation in the frequency domain characterize the behavior of the system. The generalized form of the frequency domain gain equation is a ratio of algebraic polynomials of order m and n

$$G(s) = \frac{K_m \cdot s^m + K_{m-1} \cdot s^{m-1} + \ldots + K_0}{C_n \cdot s^n + C_{n-1} \cdot s^{n-1} + \ldots + C_0} \tag{2.4-1}$$

The roots are the solution of the equation with the input set to zero. This is called the homogeneous form of the equation. An n^{th} order polynomial can be solved numerically by Mathcad or other analysis tools to find the roots, but it is more informative if the solution can be solved symbolically to show the effect each of the parameters has on the solution. Solving the polynomial symbolically involves factoring the polynomial. Many physical systems can be modeled as first or second order systems, so the system equation typically starts out in factored form, made up of first and second order polynomials. If the design process starts out in the frequency domain, the system can be constructed from a series of lower order blocks that are already in factored form. The roots of the polynomial are the roots of each of the polynomial's factors which are found by setting each first or second order factor to zero and solving for *s*.

The roots of the numerator of the system gain equation (2.4-1) are called zeros because the overall gain of the equation goes to zero when the frequency is equal to that root. The roots of the denominator of equation (2.4-1) are called poles. Since the root goes to zero when the frequency is equal to that root, the overall gain of the system goes to infinity at the frequency of a pole. Because the solution to the system equation goes to infinity at its poles, the poles are also called singularities.

For a first order gain equation,

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\left(\frac{C_2}{C_1} \cdot s + 1\right)} = \frac{C_0}{C_1} \cdot \frac{1}{\left(\frac{s}{\omega_c} + 1\right)} \tag{2.4-2}$$

the root of the denominator is found by setting it equal to zero and solving for s

$$\frac{C_2}{C_1} \cdot s + 1 = 0 \tag{2.4-3}$$

the root, or pole is

$$s = -\frac{C_1}{C_2} = -\omega_c \tag{2.4-4}$$

A useful tool in understanding the behavior of a system is a plot of the real and imaginary portions of the roots. This is called an s-plane plot. An s-plane plot is a plot of the real part of the complex number, σ , on the horizontal axis, and the imaginary part, $j\omega$, on the orthogonal vertical axis. The poles are represented by **x**'s, and the zeros by **o**'s.

Figure 2-14 shows an s-plane plot of the pole location for $\omega_c = 1 \cdot 2\pi$. For a physical system, the pole of a first order system is always real so will fall on the real axis of the s-plane. The distance along the real axis from the origin to the pole location is ω_c .

$$\sigma = -\omega_c = -6.283 \tag{2.4-5}$$

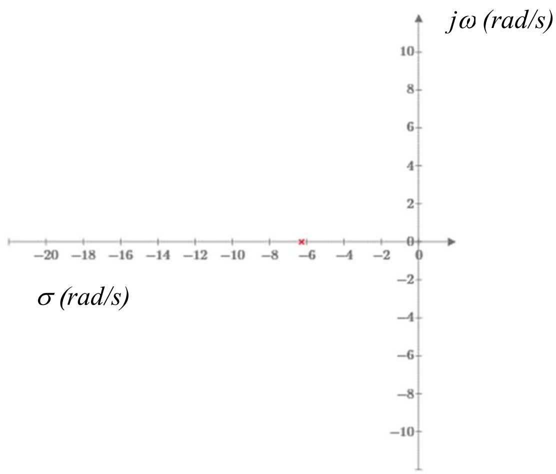


Figure 2-14 s-plane plot of the pole of the first order system for $\omega_c = 1 \cdot 2\pi$

For a second order gain equation,

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G(s) = \frac{C_0}{C_1} \cdot \frac{1}{\frac{C_3}{C_1} \cdot s^2 + \frac{C_2}{C_1} \cdot s + 1} = \frac{C_0}{C_1} \cdot \frac{1}{\frac{s^2}{\omega_n{}^2} + \frac{2 \cdot \zeta \cdot s}{\omega_n} + 1} \tag{2.4-6}$$

the roots of the denominator are found by setting it equal to zero and solving for s

$$C_3 \cdot s^2 + C_2 \cdot s + C_1 = 0 \tag{2.4-7}$$

For a quadratic equation

$$a \cdot x^2 + b \cdot x + c = 0 \tag{2.4-8}$$

the solution for x, the roots are

$$x = \frac{-b \pm \sqrt{b^2 - 4 \cdot a \cdot c}}{2 \cdot a} \tag{2.4-9}$$

So, the roots or poles of the gain equation (2.4-6) are

$$s = \frac{-C_2 \pm \sqrt{-(4 \cdot C_3 \cdot C_1) + C_2^{\ 2}}}{2 \cdot C_3} = -\zeta \cdot \omega_n \pm \sqrt{\left(\zeta^2 - 1\right) \cdot \omega_n^{\ 2}} \tag{2.4-10}$$

The real and imaginary poles in terms of ω_n and ζ are

$$s = \sigma \pm \omega \cdot j = -\zeta \cdot \omega_n \pm \sqrt{\left(\zeta^2 - 1\right) \cdot \omega_n^{\ 2}} \tag{2.4-11}$$

where the real portion, for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$, is

$$\sigma = \mathrm{Re}\left(-\zeta \cdot \omega_n \pm \sqrt{\left(\zeta^2 - 1\right) \cdot \omega_n^{\ 2}}\right) = -\zeta \cdot \omega_n = -1.885 \tag{2.4-12}$$

and the imaginary portion is

$$\begin{aligned} \omega &= \mathrm{Im}\left(-\zeta \cdot \omega_n \pm \sqrt{\left(\zeta^2 - 1\right) \cdot \omega_n^{\ 2}}\right) \\ &= \mathrm{Im}\left(-\zeta \cdot \omega_n \pm \sqrt{-1 \cdot \left(1 - \zeta^2\right) \cdot \omega_n^{\ 2}}\right) \\ &= \mathrm{Im}\left(-\zeta \cdot \omega_n \pm \sqrt{\left(1 - \zeta^2\right) \cdot \omega_n^{\ 2}} \cdot j\right) \\ &= \pm\sqrt{\left(1 - \zeta^2\right) \cdot \omega_n^{\ 2}} = \pm 5.994 \end{aligned} \tag{2.4-13}$$

The roots of equation (2.4-6) are real for $(4 \cdot C_3 \cdot C_1) \le C_2^{\ 2}$ or $\zeta \ge 1$, and are complex conjugates, meaning they both have the same real component, and complimentary imaginary components, for $(4 \cdot C_3 \cdot C_1) > C_2^{\ 2}$ or $\zeta < 1$.

Figure 2-15 shows the s-plane plot of the conjugate pair of roots for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$.

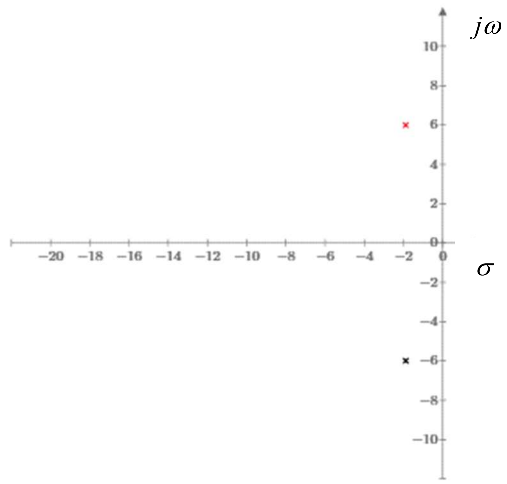


Figure 2-15 s-plane plot of a conjugate pair of roots for $\omega_n = 1 \cdot 2\pi$ and $\zeta = 0.3$

The location of the poles can also be expressed in terms of polar coordinates. Using the functions from equations (2.4-12) and (2.4-11) for $\mathrm{Re}(s)$ and $\mathrm{Im}(s)$, the distance from the origin to the pole location from equation (2.4-14) is ω_n, and the angle between the real axis and the pole from equation (2.4-15) is the arccosine of ζ, as shown in Figure 2-16

$$\sqrt{\left(\mathrm{Re}(s)\right)^2 + \left(\mathrm{Im}(s)\right)^2} = \sqrt{\left(-\zeta \cdot \omega_n\right)^2 + \left(\sqrt{\left(1-\zeta^2\right) \cdot \omega_n^{\ 2}}\right)^2} = \omega_n \tag{2.4-14}$$

and

$$\mathrm{Re}(s) = -\zeta \cdot \omega_n \Rightarrow -\frac{\mathrm{Re}(s)}{\omega_n} = \zeta \Rightarrow \arccos(\zeta) = \phi \tag{2.4-15}$$

or

$$\zeta = \cos(\phi) \tag{2.4-16}$$

So, the natural frequency, ω_n, is the length of the vector from the origin to the pole, and the damping ratio, ζ, is the cosine of the angle between the real axis and the vector to the pole.

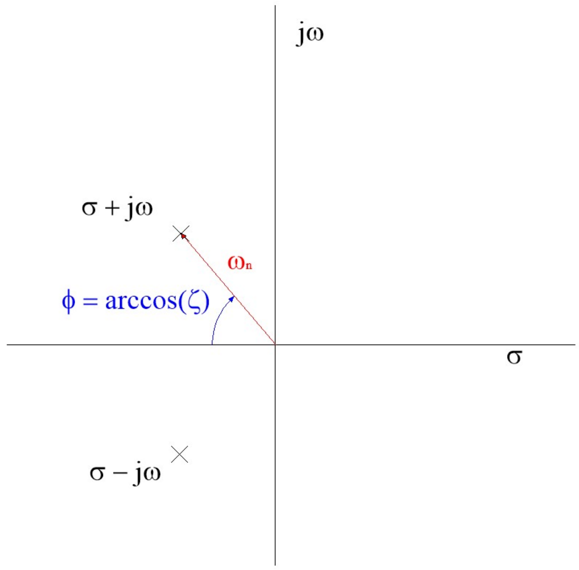


Figure 2-16 s-plane pole locations in terms of ω_n and ζ

The advantage of the polar coordinates is the magnitude is a function of just ω_n , and the angle is a function of just ζ , so it is easy to see the connection between the pole locations and ω_n and ζ . If the poles move radially away from the center of the s-plane, the natural frequency increases, but the damping ratio remains constant. If the poles move along a fixed radius arc, the natural frequency remains constant, but the damping ratio changes with the system becoming less stable as the poles move closer to the imaginary axis. Figure 2-17 shows a plot of the angle of the pole versus the damping ratio. As seen in Figure 2-10, damping ratios above 0.3 have a small amount of overshoot and a fast decay time. Figure 2-17 shows that for systems having damping ratios less than 0.3, the poles are within 17 degrees of the imaginary axis.

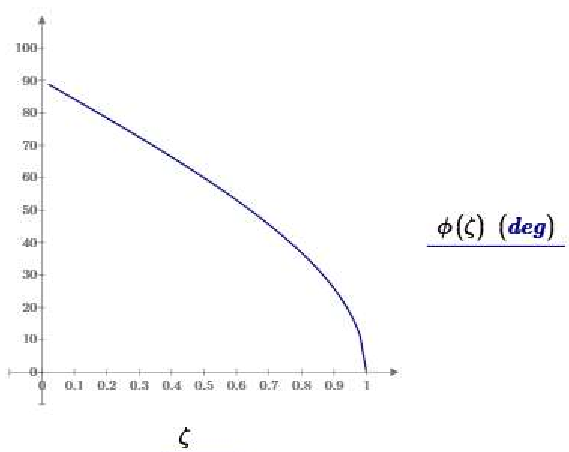


Figure 2-17 Angle of the s-plane pole location as a function of the damping ratio

For physical systems, complex roots will always be complex conjugate pairs. So, in factored form, a higher order system will be comprised of first and second order functions if the functions can be factored. The first order equations cover the real poles and zeros, and the second order equations cover the complex conjugate pair poles and zeros, or pairs of poles on the real axis.

2.6 Right Half vs Left Half Plane Poles

The poles of the system gain equation can fall on the left or right half of the s-plane. Any system with one or more poles that fall in the right half plane will be unstable, meaning the output response to an impulse input will increase exponentially over time, whereas a system with just left half plane poles will be stable, meaning the output response to an impulse input will decrease over time. For a stable system, a system the takes longer to settle to zero is less stable than a system that takes longer to settle to zero.

For the first order system of (2.2-5), when the sign of the coefficient $C_2/C_1 = 1/\omega_c$ is positive, the root is a negative real value

$$\left(\frac{s}{\omega_c}+1\right)=0 \Rightarrow s=-\omega_c \tag{2.5-1}$$

A negative pole falls on the left half of the s-plane. If the coefficient $C_2/C_1 = 1/\omega_c$ is negative, the root is a positive real value

$$\left(\frac{s}{-\omega_c}+1\right)=0 \Rightarrow s=\omega_c \tag{2.5-2}$$

and the pole falls on the right half of the s-plane. For the positive value of ω_c, the exponent of the step response in equation (2.2-6) and (2.5-3) becomes positive, and the step response increases without bound as shown in Figure 2-18.

$$\left.\alpha_{out}(t)\right|_{step}=1-\frac{1}{C_1}\cdot e^{\omega_c\cdot t} \tag{2.5-3}$$

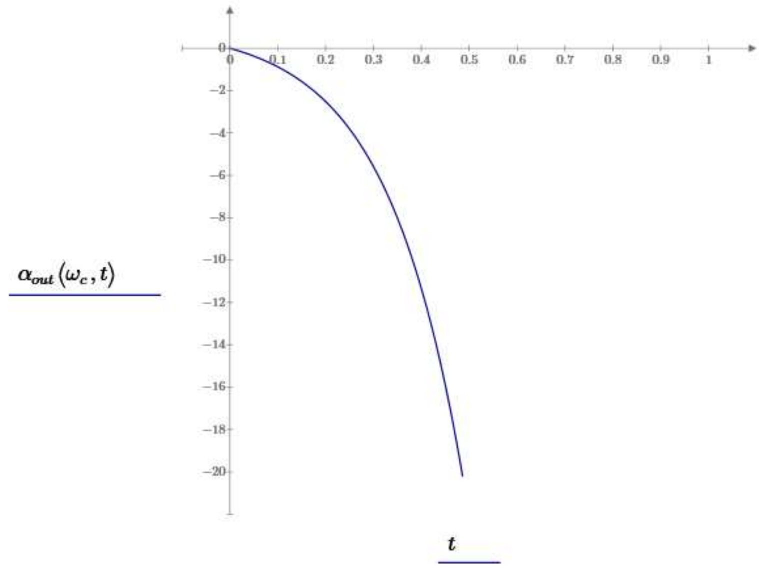


Figure 2-18 Step response of a first order system for $C_1 = 1$ and $\omega_c = -1 \cdot 2\pi$

For a second order system, if ζ goes negative, the poles are in the right half of the s-plane, and the system will produce an exponentially increasing oscillation envelope. Mathematically, this can happen, but in a real-world system, something would eventually limit the amplitude of the oscillations. If ζ is set equal to 0, the poles are on the imaginary axis of the s-plane, and the system will produce a stable oscillation of constant amplitude. For a system with a positive value of ζ, the poles are in the left half of the s-plane, and the system is stable. A system that has a smaller positive value of ζ and more overshoot is said to be less stable than a system with a larger value of ζ, less overshoot, and a faster decay envelope.

For the second order system of equation (2.3-4), the pole locations are

$$\left(\frac{s^2}{\omega_n^{\ 2}} + \frac{2 \cdot \zeta \cdot s}{\omega_n} + 1\right) = 0 \Rightarrow s = -\zeta \cdot \omega_n \pm \sqrt{\left(\zeta^2 - 1\right) \cdot \omega_n^{\ 2}} \qquad (2.5\text{-}4)$$

Table 2-1 shows the location of the poles on the s-plane for all ranges of ζ. If $\zeta < 0$, there will always be at least one right half plane pole.

Case	Criteria	Pole Location
1	$\zeta > 1$	Two negative real axis poles
2	$\zeta = +1$	Two negative real axis poles at the same frequency
3	$0 < \zeta < +1$	Two complex conjugate negative real poles
4	$\zeta = 0$	Two complex conjugate poles on the $j \cdot \omega$ axis
5	$0 > \zeta > -1$	Two complex conjugate positive real poles
6	$\zeta = -1$	Two positive real axis poles at the same frequency
7	$\zeta < -1$	Two positive real axis poles

Table 2-1 Location of the poles of equation (2.5-4) for all ranges of ζ

Figure 2-19 shows the second order step response, equation (2.3-9), for three values of ζ, +0.1, 0.001 (~0), and -0.1, cases 3, 4, and 5, in Table 2-1, producing a decreasing, constant, and increasing envelop, respectively.

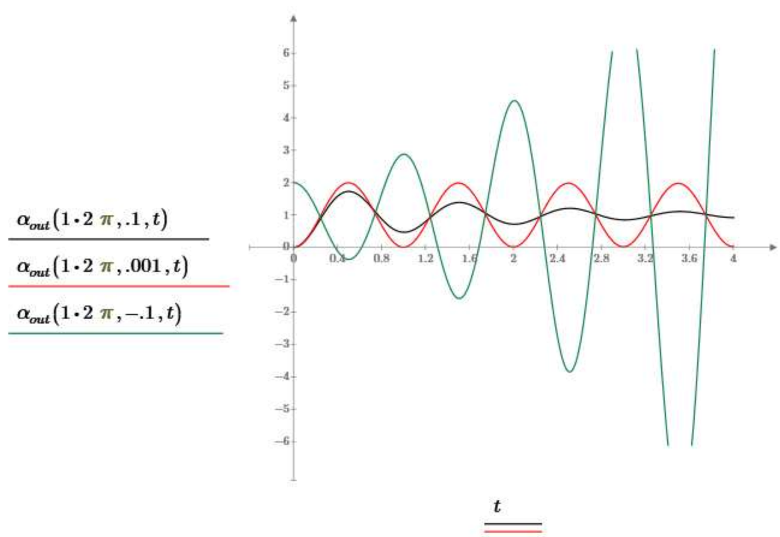


Figure 2-19 Second order step response for of ζ, +0.1, 0.001 (~0), and -0.1, with $\omega_n = 1 \cdot 2\pi$

If the second order system equation is written in the form of equation (2.3-3), then the roots are

$$\left(\frac{C_3}{C_1}\cdot s^2+\frac{C_2}{C_1}\cdot s+1\right)=0\Rightarrow s=\frac{-C_2\pm\sqrt{-4\cdot C_1\cdot C_3+C_2^{\,2}}}{2\cdot C_3} \tag{2.5-5}$$

Using the Mathcad symbolic capability, the step response for the second order system written in this form is

$$\alpha_{out}(t)\Big|_{step}=-\left(\frac{C_2}{2\cdot C_3\cdot\sqrt{\dfrac{4\cdot C_1\cdot C_3-C_2^{\,2}}{4\cdot C_3^{\,2}}}}\right)\cdot e^{\frac{-C_2\cdot t}{2\cdot C_3}}\cdot\sin\left(t\cdot\sqrt{\frac{4\cdot C_1\cdot C_3-C_2^{\,2}}{4\cdot C_3^{\,2}}}\right)-e^{\frac{-C_2\cdot t}{2\cdot C_3}}\cdot\cos\left(t\cdot\sqrt{\frac{4\cdot C_1\cdot C_3-C_2^{\,2}}{4\cdot C_3^{\,2}}}\right)+1$$

(2.5-6)

Table 2-2 shows the location of the poles on the s-plane for various ranges of the coefficients. If $C_2<0$, there will always be at least one right half plane pole. Written in this form as a function of C_1, C_2, and C_3, gives more latitude in the placement of the poles than the equation expressed in terms of just ω_n and ζ. The former allows for a pair of poles on the real axis, one negative and one positive.

Case	**Criteria**	**Pole Location**
1	$4\cdot C_1\cdot C_3<C_2^{\,2}$ & $C_2>0$	Two real axis poles with at least one negative
2	$4\cdot C_1\cdot C_3=C_2^{\,2}$ & $C_2>0$	Two negative real axis poles at the same frequency
3	$4\cdot C_1\cdot C_3>C_2^{\,2}$ & $C_2>0$	Two complex conjugate negative real poles
4	$C_2=0$	Two complex conjugate poles on the $j\cdot\omega$ axis
5	$4\cdot C_1\cdot C_3>C_2^{\,2}$ & $C_2<0$	Two complex conjugate positive real poles
6	$4\cdot C_1\cdot C_3=C_2^{\,2}$ & $C_2<0$	Two positive real axis poles at the same frequency
7	$4\cdot C_1\cdot C_3<C_2^{\,2}$ & $C_2<0$	Two real axis poles with at least one positive

Table 2-2 Location of the poles of equation (2.5-5) for all ranges of C_2

Figure 2-20 shows the step response for equation (2.5-6) for $C_3=2$, $C_2=0.5$, and $C_1=-20$. The pole locations are +3.04 and -3.29 radians/sec. Because the poles are on the real axis, there is no imaginary part that gives rise to a sinusoidal response, there is just an increasing exponential response.

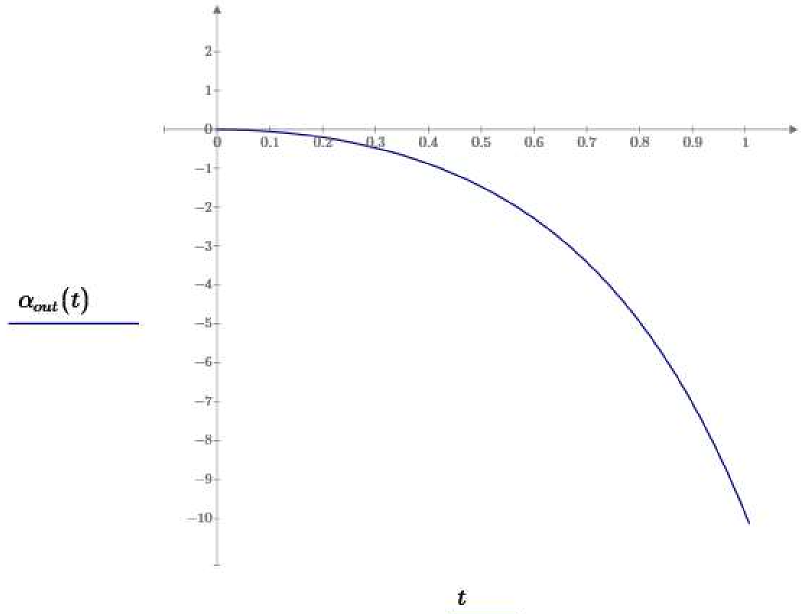


Figure 2-20 Step response of two real poles, one negative and one positive

For the case where there is one positive and one negative pole on the real axis, equation (2.5-4) or (2.5-5) can be simplified by knowing that the solution consists of two first order real poles as shown in equation (2.5-7).

$$G(s) = \frac{1}{\left(\frac{s}{\omega_1}+1\right)\cdot\left(\frac{s}{\omega_2}+1\right)} \tag{2.5-7}$$

The inverse Laplace transform of $G(s)\cdot\frac{1}{s}$ for the step response is

$$\alpha_{out}(t)\Big|_{step} = 1 + \frac{-\omega_2 \cdot e^{-\omega_1 \cdot t} + \omega_1 \cdot e^{-\omega_2 \cdot t}}{\omega_2 - \omega_1} \tag{2.5-8}$$

Even though there is both a positive and negative exponential, the magnitude of the positive exponential quickly exceeds the magnitude of the negative exponential, and the resulting sum grows exponentially as shown in Figure 2-21.

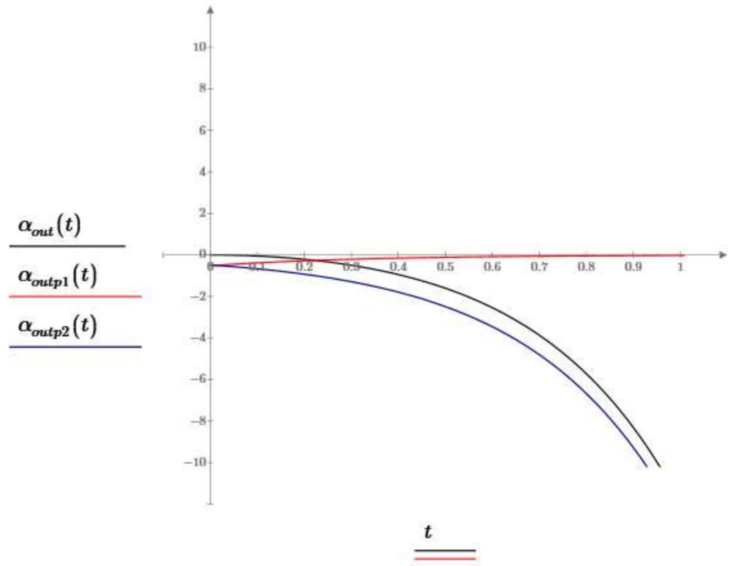


Figure 2-21 Plot of $\alpha_{out}(t)\Big|_{step}$, $\alpha_{outp1} = \frac{\omega_2}{\omega_1 - \omega_2} \cdot e^{-\omega_1 \cdot t}$, and $\alpha_{outp2} = \frac{\omega_1}{\omega_2 - \omega_1} \cdot e^{-\omega_2 \cdot t}$

In summary, any system with all left half s-plane poles will produce a response with an exponentially decaying envelope, which is a stable response. If there is a single pole on the left half real axis, the response is a simple decaying exponential. If there is a pair of left half complex conjugate poles, then the response will a decaying sinusoid. If there is a single pole on the right half real axis, the response is a simple increasing exponential. If there is a pair of right half complex conjugate poles, then the response will be an exponentially increasing sinusoid. Any order system with at least one right half plane pole will produce an exponentially increasing envelope, which is an unstable system. A system with a complex conjugate pole pair on the $j\omega$ axis will produce a stable, constant amplitude sinusoidal response.

2.7 Gain and Phase from Poles & Zeros

The magnitude and phase of the gain of a system can be determined directly from the location of its poles and zeros on the s-plane.

The magnitude of a gain function is the product of the lengths of each of the vectors that extend from all the zeros to a test frequency on the $j\omega$ axis divided by the product of the lengths of each of the vectors that extend from all the poles to that same test frequency on the $j\omega$ axis as shown in Figure 2-22. The phase is the sum the angles of each of the vectors that extend from all the zeros to the test frequency on the $j\omega$ axis minus the sum of the angles of each of the vectors that extend from all the poles to that same test frequency on the $j\omega$ axis. The length of the vector from a pole of zero to the $j\omega$ axis approaches a factor of 10 for each factor of 10 in frequency well past the frequency of the pole or zero. The angles go from -90 deg to + 90 deg for each pole or zero. For frequencies less than the pole and zero frequencies, the ratio of the length of the vector from the zeros and the poles remains relatively constant, so the magnitude of the gain function remains relatively constant.

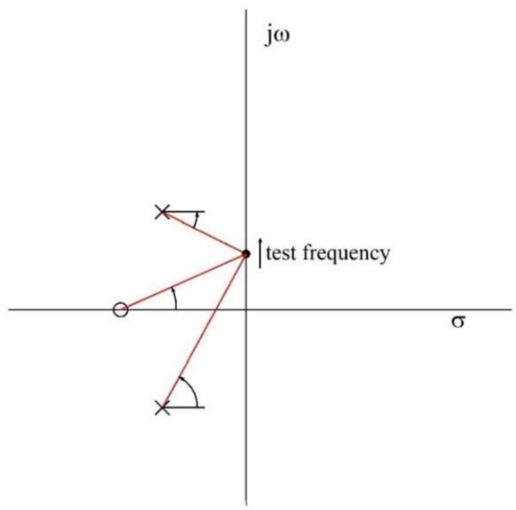


Figure 2-22 Magnitude and phase from s-plane poles and zeros

The Bode plot, then, can be constricted just knowing the locations of the poles and zeros on the s-plane. Each zero contributes +20dB/decade in magnitude above its corner frequency and +90 deg in phase. Each pole contributes -20dB/decade in magnitude and -90 deg in phase. Systems with just poles are low pass in nature as the gain decreases above the frequency of the poles. A system that has only zeros, or zeros at a lower frequency than poles, would be high pass in nature as the gain increases at a stimulus frequency higher than the frequency of the zeros.

Figure 2-23 shows the magnitude and phase of equation (2.6-1), a system consisting of a single pole, for $\omega_c = 1 \cdot 2\pi$. The magnitude decreases at 20dB per decade above ω_c, and the phase increases to -90 deg.

$$G(s) = \frac{1}{\left(\frac{s}{\omega_c} + 1\right)} \tag{2.6-1}$$

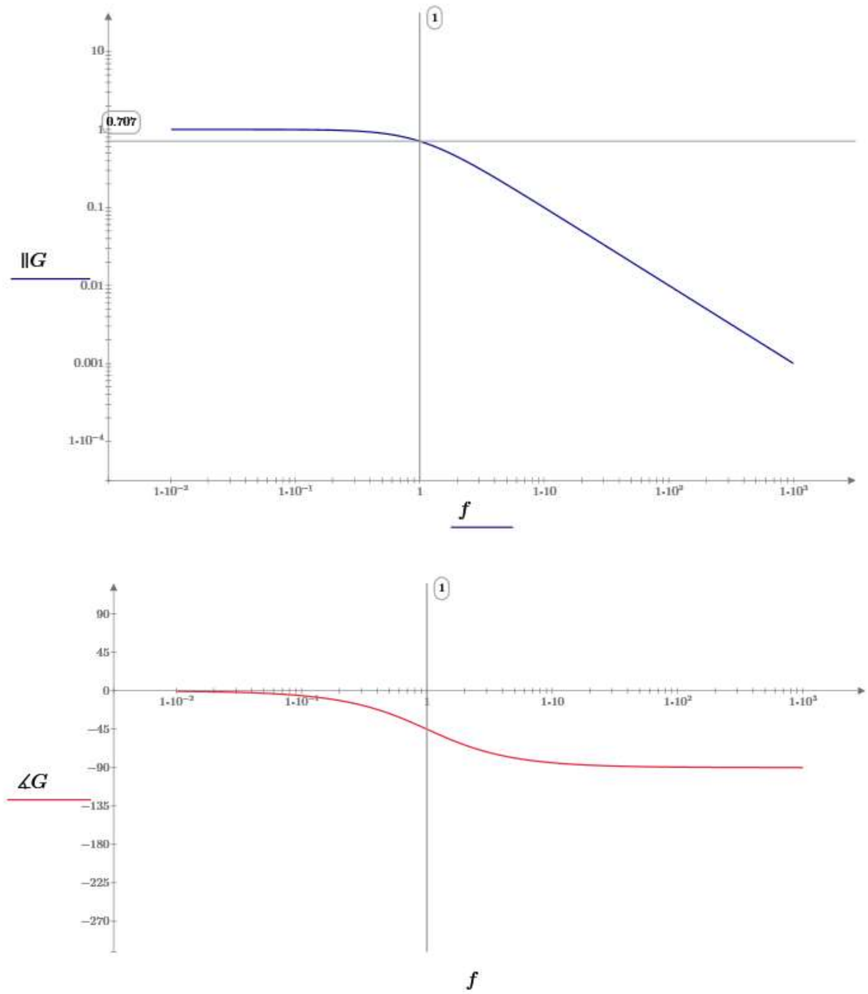


Figure 2-23 Magnitude and phase of the frequency response of a first order system for $C_1 = 1$ and $\omega_c = 1 \cdot 2\pi$

Figure 2-24 shows the magnitude and phase of equation (2.6-2), a system consisting of a single zero, for $\omega_c = 1 \cdot 2\pi$. The magnitude and phase of the gain is the inverse of the single pole function of equation (2.6-1). The magnitude increases at 20dB per decade above ω_c , and the phase increases to +90 deg.

$$G(s) = \left(\frac{s}{\omega_c} + 1\right) \tag{2.6-2}$$

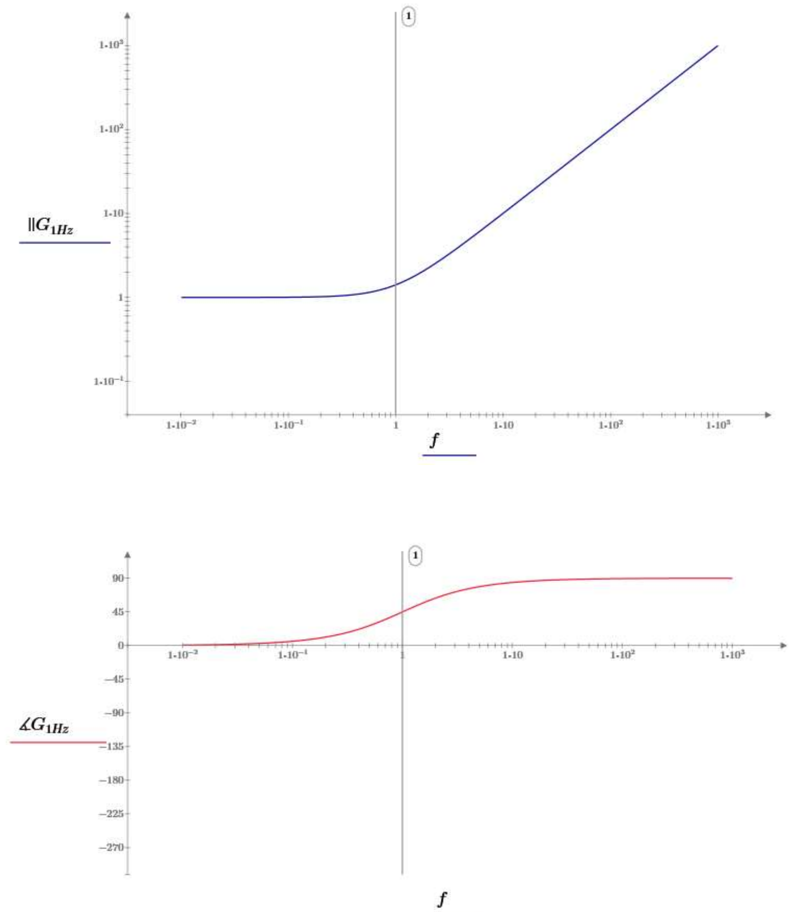


Figure 2-24 Magnitude and phase of equation (2.6-2) for $\omega_c = 1 \cdot 2\pi$

Equation (2.6-3) is a gain function with a real axis zero and a complex pole pair. Figure 2-25 shows a plot of the poles and zero for $\omega_z = 1 \cdot 2\pi$, $\omega_n = 1 \cdot 2\pi$, and $\zeta = 0.5$. The magnitude and phase of that gain function is shown in Figure 2-26. For the function, the net number of poles and zeros is one, that is, there is one more pole than zero, so the function will eventually roll off at 20 dB/decade, and the final phase is 90 deg. There is peaking in the gain as the test frequency passes closest to the complex pole because of the proximity of the pole pair to the $j\omega$ axis when ζ is less than 1. The closer the pole is to the $j\omega$ axis, the greater the peaking as shown Figure 2-8. Since the zero falls at the same frequency as the poles, it has the effect of negating the change in magnitude and phase from one of the poles, but the peaking in magnitude from the poles' proximity to the imaginary axis still occurs.

$$G(s)=\frac{\left(\frac{s}{\omega_z}+1\right)}{\left(\frac{s^2}{\omega_n{}^2}+\frac{2\cdot\zeta}{\omega_n}+1\right)} \tag{2.6-3}$$

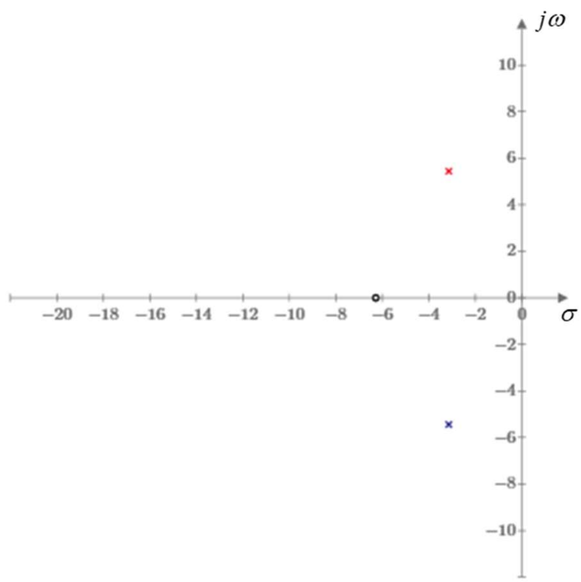
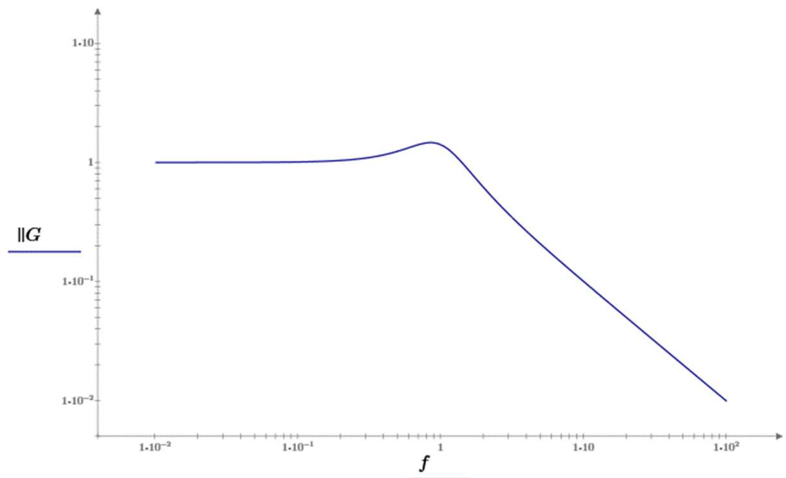


Figure 2-25 s-plane plot of the roots of equation (2.6-3) for $\omega_z = 1\cdot 2\pi$, $\zeta = 0.5$, and $\omega_z = 1\cdot 2\pi$



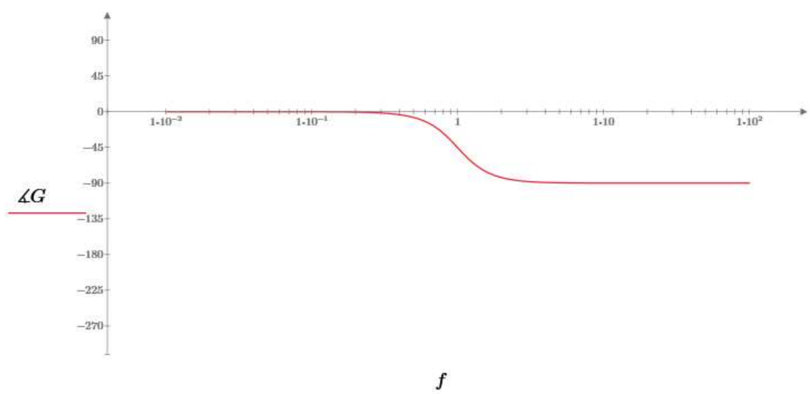


Figure 2-26 Magnitude and phase of the gain of equation (2.6-3) for $\omega_z = 1 \cdot 2\pi$, $\omega_n = 1 \cdot 2\pi$, and $\zeta = 0.5$

2.8 Second Order System with a Zero

One of the models used for the rotational stability system is a second order differential equation that includes a zero in the numerator. This section characterizes that system.

The frequency domain form of the equation is

$$\frac{\alpha_{out}(s)}{\alpha_{in}(s)} = G_2(s) = G_0 \cdot \frac{\dfrac{s}{\omega_z} + 1}{\dfrac{s^2}{\omega_n{}^2} + \dfrac{2 \cdot \zeta \cdot s}{\omega_n} + 1} \tag{2.7-1}$$

Figure 2-27 shows the gain of (2.7-1) for various damping ratios and zero frequencies locations where the frequency of the zero is significantly lower than the frequency of the poles. In these plots, the pole frequency is 1 Hz, the damping ratios are 0.5 and 0.05, the zero frequencies are 0.1 Hz and 0.01 Hz, and the gain $G_0 = 1$. The gain rises at 20 dB/decade starting at the frequency of the zero. The two poles at resonance then cause the gain to turn around and fall at 20 dB/decade. If the frequency of the zero is moved down by a factor of 10, the gain of the function at its peak increases by a factor of 10, regardless of the value of the damping ratio.

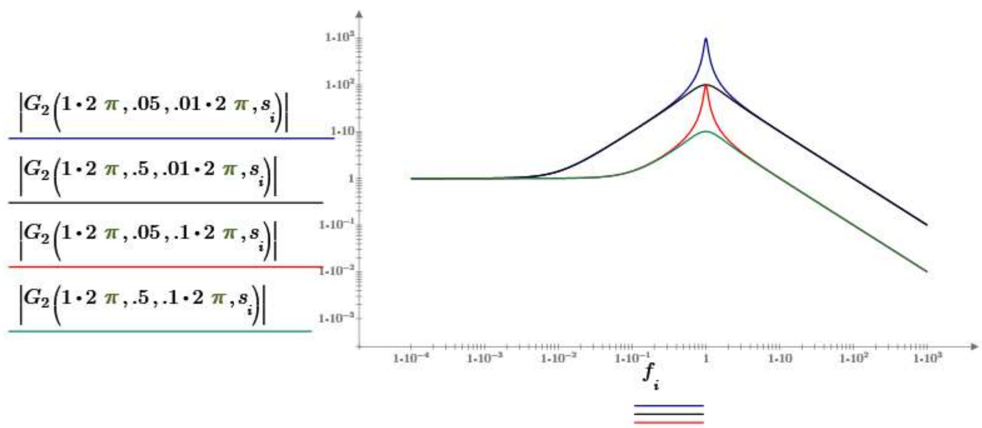


Figure 2-27 Second order equation with varying damping ratio, ζ, and zero location, ω_z

Figure 2-28 compares the second order system response with and without the zero. The effective gain at resonance due to the zero is the ratio of the pole to zero frequencies. If the second order system without the zero is multiplied by that gain, then the peak of its gain at resonance is the same as the second order system with the zero.

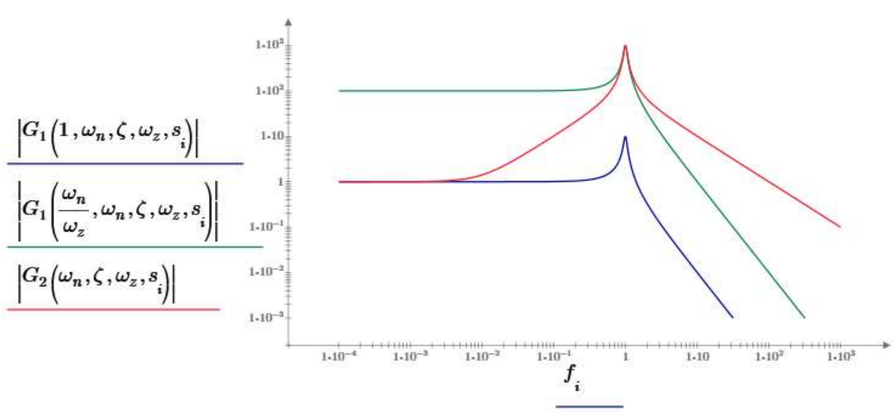


Figure 2-28 Comparing the second order system with and without the zero

The time domain step response of a system is found by taking the inverse Laplace transform of the frequency domain step function, $1/s$, multiplied by the frequency domain gain function. Using the symbolic engine in Mathcad, the step response for the second order equation (2.7-1) , assuming the gain $C_0 = 1$, is

$$\mathcal{L}^{-1}\left(\frac{1}{s}\cdot G_2(s)\right) = \mathcal{L}^{-1}\left(\frac{1}{s}\cdot\frac{\frac{s}{\omega_z}+1}{\frac{s^2}{\omega_n{}^2}+\frac{2\cdot\zeta\cdot s}{\omega_n}+1}\right) =$$

$$\alpha_{out}(t)\Big|_{step} = \frac{\omega_n{}^2-\zeta\cdot\omega_n\cdot\omega_z}{\omega_z\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}}\cdot e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\sin\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)-e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\cos\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)+1$$

(2.7-2)

The step response is plotted in Figure 2-29. The zero being well below the pole frequency results in a final step magnitude that is very small compared to the magnitude of the oscillations. The graph on the right shows the step response for the second order system without the zero for comparison. The frequency of oscillation and the decay envelope for the system with the zero are the same as the second order system without the zero, but the oscillations lead the system without the zero by 90 degrees because of the zero.

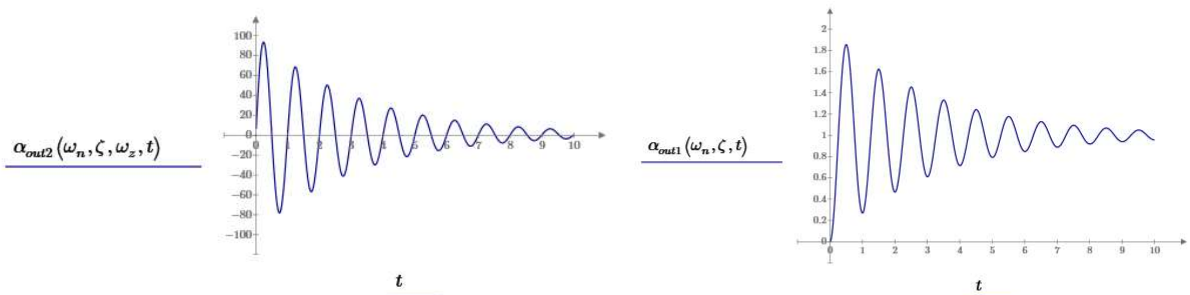


Figure 2-29 Plot of the step response of a second order system with a zero $\omega_z = 0.01Hz$, left, and no zero, right, for $\omega_n = 1Hz$ and $\zeta = 0.05$

Equation (2.7-1) will be used to represent a velocity function in the stability analysis. It will also be helpful to see the linear distance an object with this velocity travels. To find the distance, the velocity function is integrated in the time domain, or multiplied by $1/s$ in the frequency domain

$$\frac{1}{s}\cdot G_2(s) = G_3(s) = K_0\cdot\frac{1}{s}\cdot\frac{\frac{s}{\omega_z}+1}{\frac{s^2}{\omega_n{}^2}+\frac{2\cdot\zeta\cdot s}{\omega_n}+1} \tag{2.7-3}$$

and Figure 2-30 shows a plot of equation (2.7-3)

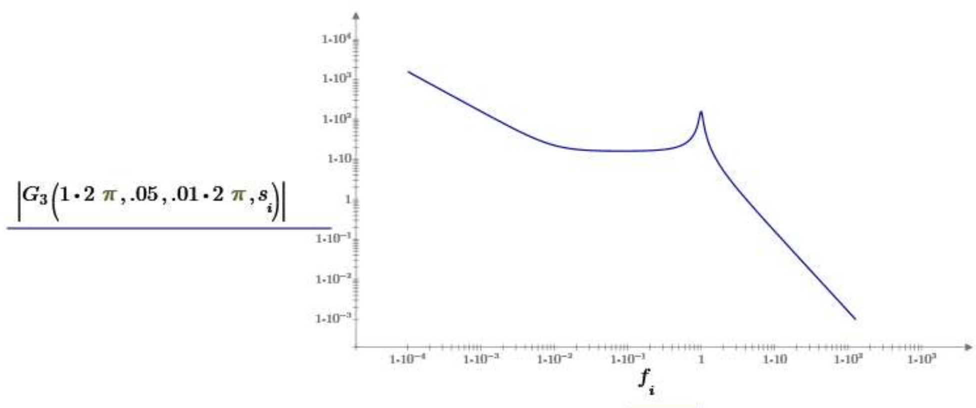


Figure 2-30 Plot of the velocity and distance functions

The step response for equation (2.7-3), as calculated by Mathcad, is given by

$$\mathcal{L}^{-1}\left(\frac{1}{s}\cdot G_3(s)\right) = \mathcal{L}^{-1}\left(\frac{1}{s^2}\cdot\frac{\frac{s}{\omega_z}+1}{\frac{s^2}{\omega_n{}^2}+\frac{2\cdot\zeta\cdot s}{\omega_n}+1}\right) =$$

$$\alpha_{out}(t)\Big|_{step} = \left(\begin{array}{l}\dfrac{\left(2\cdot\zeta^2-1\right)\cdot\omega_n\cdot\omega_z-\zeta\cdot\omega_n{}^2}{\omega_n\cdot\omega_z\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}}\cdot e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\sin\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)+\\ +\dfrac{2\cdot\zeta\cdot\omega_z-\omega_n}{\omega_n\cdot\omega_z}\cdot e^{-(\zeta\cdot\omega_n\cdot t)}\cdot\cos\left(t\cdot\sqrt{(-\zeta^2+1)\cdot\omega_n{}^2}\right)+\dfrac{\omega_n\cdot\omega_z\cdot t+\left(\omega_n-2\cdot\zeta\cdot\omega_z\right)}{\omega_n\cdot\omega_z}\end{array}\right) \tag{2.7-4}$$

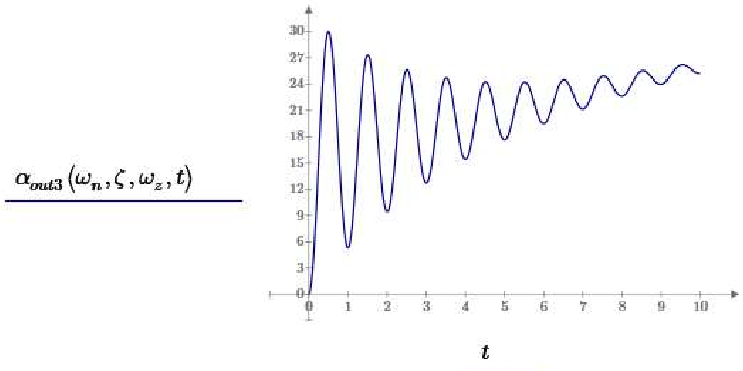


Figure 2-31 Plot of the second order step response for equation (2.7-3) for $\omega_n = 1Hz$ and $\zeta = 0.05$ with the zero $\omega_z = 0.01Hz$

Because there is a small offset in the step response of equation (2.7-1), the step response of the integral of equation (2.7-1) has an initial step and a long term drift due to the pole added by the integration.

3 Control Systems

3.1 Introduction

Feedback control loops are used to correct for the difference between the actual and desired response of a system. Feedback control systems depend upon an accurate measurement of the output of the system. It is often easier to measure the response more accurately than it is to control the system open loop because the system has a non-ideal response, it is difficult to characterize the system response, the system has nonlinearities, or there are external forces that impact the system's response. With feedback, the measured response is compared to the desired response and a correction is generated from the error between the two and applied to the system to force actual response to match the desired response. A feedback control system can be used to steer an inherently stable rocket by using control surfaces on the fins, along a vertical flight path, making it less suitable to weathercocking into the wind. A feedback control loop can also be designed to stabilize an inherently unstable finless rocket by vectoring the thrust of the rocket's motor and direct it along a vertical flight path.

3.2 High Gain Feedback Control Loops

There are two commonly used, fundamentally different approaches to feedback control. The first is the high gain control loop, commonly used in electronics design, and the second is state-space control, commonly seen in process and aerodynamics control systems where the system being controlled has very complex dynamics of its own. This section covers high gain control loop design.

A simple example of a high gain feedback control system is an audio amplifier, where a control loop is used to correct for nonlinearities in the transistors that are used in the high-power output stages of the amplifier.

Figure 3-1 shows the basic form of a feedback control loop[10] where the forward gain is $a(s)$, and the feedback gain is $f(s)$. The closed loop gain for this structure is

$$\frac{\alpha_{out}}{\alpha_{in}} = \frac{a(s)}{1 + a(s) \cdot f(s)} \tag{3.2-1}$$

As the forward gain, $a(s)$, becomes very large, the closed loop gain can be approximated by

[10] See (Roberge, 1975) for a more complete explanation of control systems and stability theory

$$\frac{\alpha_{out}}{\alpha_{in}} \simeq \frac{1}{f(s)} \tag{3.2-2}$$

and the error term, α_{err} , which is $\frac{\alpha_{out}}{a(s)}$, must tend to *0*. With a very large forward path gain, the overall gain of a feedback control loop is almost entirely dependent upon the feedback path gain. If the feedback function is linear and stable, such as a passive resistor network in the case of an audio power amplifier, then the feedback control loop can be used to accurately set the gain and linearize an element that is embedded in the forward path of the loop, such as a power transistor.

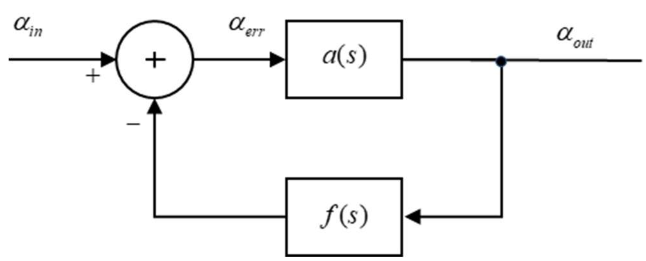


Figure 3-1 Basic control loop structure

The loop transmission is defined as the gain around the loop including the negative sign of the summing junction

$$LT(s) = -a(s) \cdot f(s) \tag{3.2-3}$$

and is used in calculating the stability of the closed loop system. A feedback control loop can become unstable for certain values of $a(s)$ or $f(s)$. From equation (3.2-1), the closed loop gain becomes infinite for a value of $a(s) \cdot f(s) = -1$, or the loop transmission equals +1 . This is called "positive" feedback. When the values of $a(s)$ and $f(s)$ are transfer functions that include poles and zeros in the frequency domain and change in magnitude and phase with frequency, there can be frequencies at which the loop is stable, and frequencies at which the loop becomes unstable. When the loop is unstable, the poles of the closed loop system are either on the $j \cdot \omega$ axis or the right half of the s-plane. In the case where $a(s) \cdot f(s) = -1$, the closed loop poles are on the $j \cdot \omega$ axis of the s-plane. A value of -1 is the equivalent of a gain of 1 with a phase of 180 deg, so a phase shift of 180 degrees in the loop transmission when the loop transmission has a magnitude of 1 produces an unstable loop.

When the magnitude and phase of the loop transmission are plotted on a Bode plot, then the stability of the closed loop system can be determined[11]. If the phase of the loop transmission is less than 180 degrees when the magnitude crosses unity, the closed loop is stable. If it is equal to or greater than 180 deg, then the closed loop is unstable.

The following is a simple example of using the Bode plot to determine the stability of control loop. Equations (3.2-4) through (3.2-7) describe a loop with two poles at 0 Hz (a double integrator), a zero before crossover, and a pole after crossover, where the zero frequency is $\omega_z = 3 \cdot 2\pi$, the pole frequency is $\omega_p = 100 \cdot 2\pi$, and the loop crossover frequency is $\omega_c = 10 \cdot 2\pi$.

$$a(s) = \frac{K_1}{s^2} \cdot \frac{\left(\frac{s}{\omega_z} + 1\right)}{\left(\frac{s}{\omega_p} + 1\right)} \tag{3.2-4}$$

$$f(s) = 1 \tag{3.2-5}$$

$$LT(s) = -a(s) \cdot f(s) = -\frac{K_1}{s^2} \cdot \frac{\left(\frac{s}{\omega_z} + 1\right)}{\left(\frac{s}{\omega_p} + 1\right)} \tag{3.2-6}$$

$$G_\Gamma = \frac{a(s)}{1 + a(s) \cdot f(s)} = \frac{a(s)}{1 - LT(s)} \tag{3.2-7}$$

K_1 will set the crossover frequency, Solving equation (3.2-4) for $s = j \cdot \omega_c$ and $a(s) = 1$, assuming that $\omega_z \ll \omega_c \ll \omega_p$

$$1 = \left| \frac{K_1}{-\omega_c^2} \cdot \frac{\left(\frac{\omega_c \cdot j}{\omega_z} + 1\right)}{\left(\frac{\omega_c \cdot j}{\omega_p} + 1\right)} \right| \simeq \frac{K_1}{\omega_c} \cdot \frac{1}{\omega_z} \Rightarrow K_1 = \omega_c \cdot \omega_z \tag{3.2-8}$$

[11] Using a Bode plot to determine stability assumes the system is a minimum phase system - see (Roberge, 1975) for more detail

Figure 3-2 shows the magnitude and phase of the loop transmission. The standard convention for the phase plot is to plot the phase of $-LT(s)$, so that a magnitude of 1 and a phase of -180 degrees corresponds to a loop transmission of +1, the condition for instability. Since the loop has two poles at 0 Hz, the loop gain rolls off at 40 dB/decade, and then breaks to a single pole, or 20 dB/decade at the zero frequency, 3 Hz. It crosses unity gain as a single pole, and then goes back to 40 dB/decade roll-off at the post crossover pole frequency, 100 Hz. The phase starts at -180 deg due to the two poles at the origin, starts to move toward -90 deg about a decade before the zero, and then moves back toward -180 after crossover due to the post-crossover pole. At unity gain, the phase is -113 deg. The difference between the phase at unit gain and -180 deg is called the phase margin and is an indication of the stability of the closed loop response. In this example, the phase margin is +67 deg. If the loop crosses over with -180 of phase, the loop has 0 degrees of phase margin, the poles are on the $j \cdot \omega$ axis, and the response is a fixed level oscillation. If the loop crosses over at more than -180 deg, the poles are in the right half plane, and the loop is unstable with an increasing exponential envelope response. If the loop crosses unity gain with less than -180 of phase, the loop has a positive phase margin, and the response is a decaying exponential and is stable. The larger the phase margin, the less peaking in the frequency response, and the less overshoot and faster the decay or the ringing on the step response. With 90 deg of phase margin or greater, there is no peaking in the frequency response, and no overshoot in the step response.

Figure 3-3 shows the magnitude of the closed loop gain. With 67 deg of phase margin, there is minimal peaking in the frequency response. Figure 3-4 shows the closed loop step response with minimal overshoot.

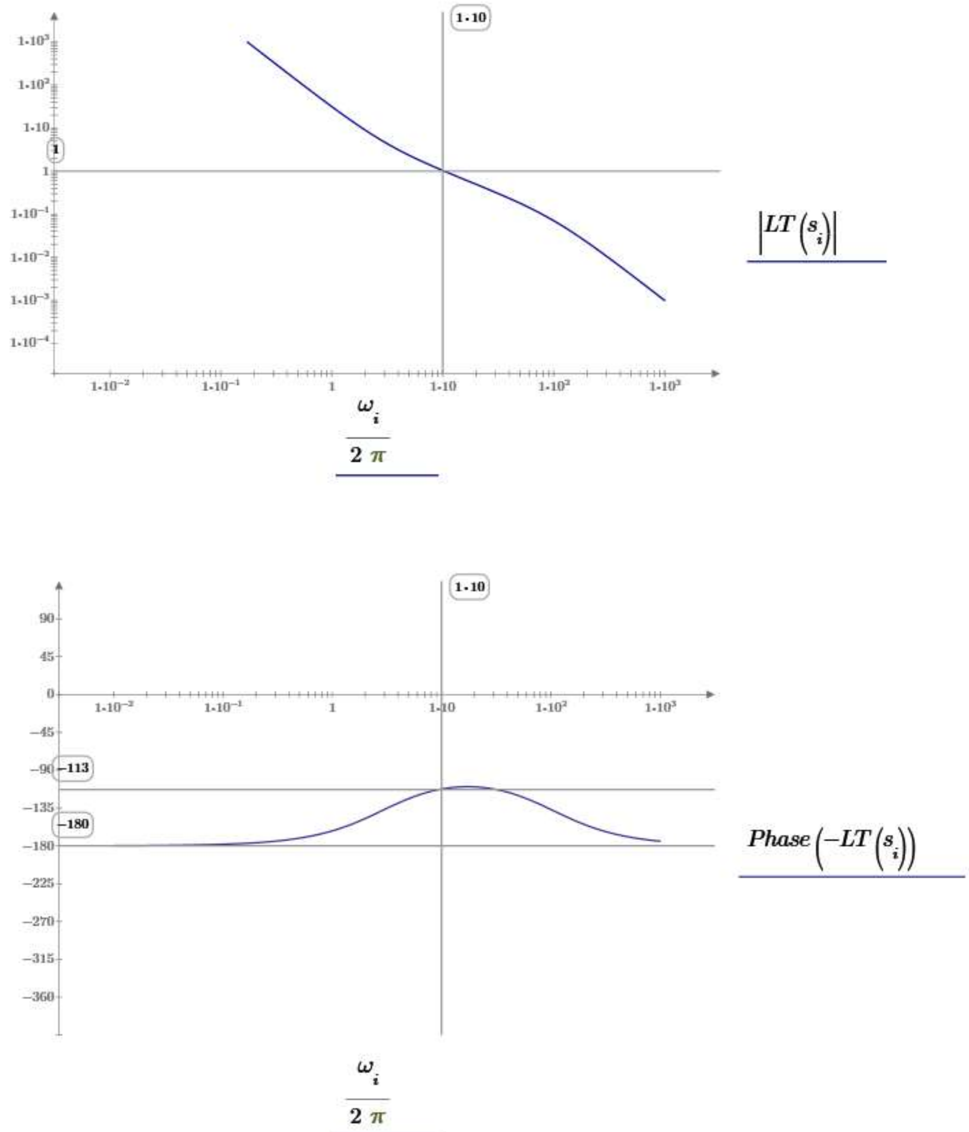


Figure 3-2 Magnitude and phase of the negative of the loop transmission

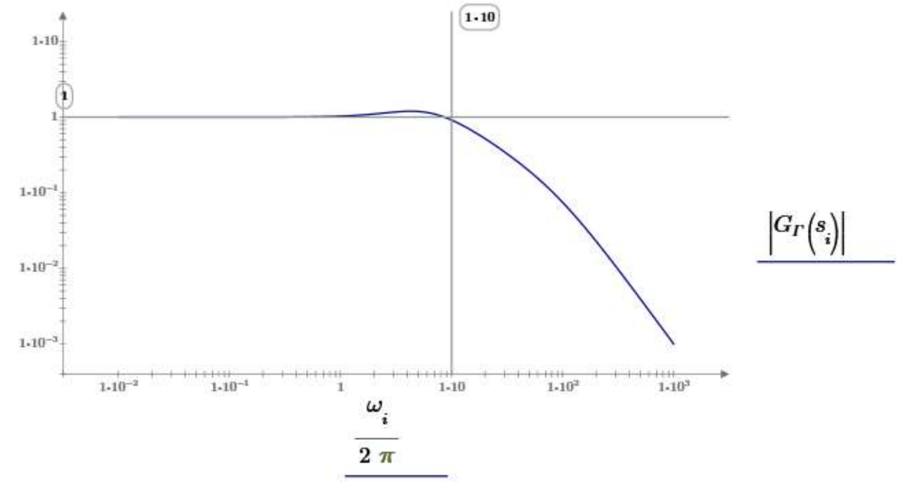


Figure 3-3 Magnitude of the closed loop gain

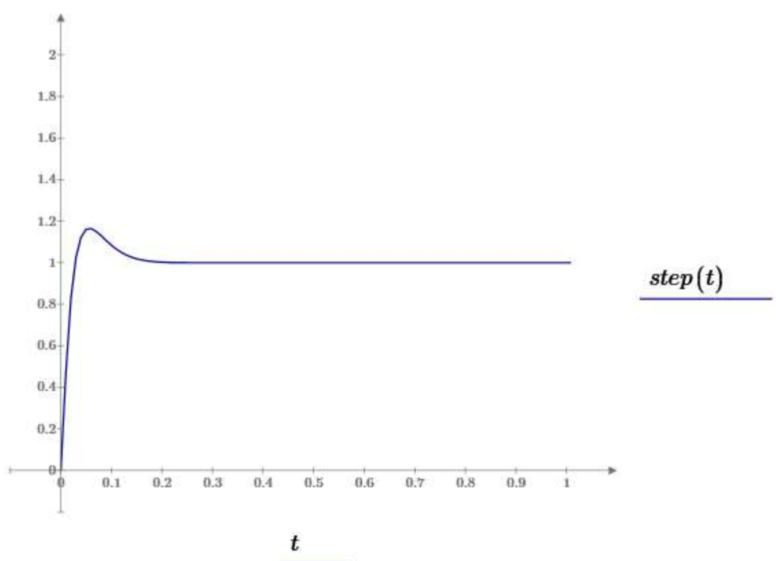


Figure 3-4 Closed loop step response

3.3 Root Locus

Not only can control loops correct for internal nonlinearities or automatically correct for external perturbations, they can also be used to shape the dynamics of the response of the system, where the dynamics are determined by the closed loop poles and zeros, the roots of the differential equation that describes the system. Control loop design, therefore, whether high gain or state space, involves closed loop pole/zero placement. The control loop takes the open loop poles of the system, and, with additional forward path and feedback path functions, moves those poles to a new closed loop location that determines the new dynamic response of the system. The goal is to design the control loop functions to move the poles to a location where the dynamics of the closed loop has the desired system dynamics response.

To help visualize the locations of poles and zeros of the system, and the overall dynamics of the system, an s-plane plot is used (see Section 2.5). Plotting both the open loop and closed loop poles and zeros shows how the control system modifies their location and the impact those changes have on the dynamics of the system.

An s-plane plot of the roots of a differential equation as a function of a parameter that changes the location of the roots is called the root locus. The root locus can be used to determine how that parameter impacts the stability of the system. Root locus plots are commonly used in designing control systems, where the design parameter being varied is the loop gain of the control loop. But root locus plots are also useful for understanding the stability of any system with complex poles where the location of the poles can be expressed as a function of any design parameter that impacts the pole locations. For the control gain function, there are a series of heuristic rules that govern the trajectory of the poles as a function of the loop gain as described below. But Mathcad can be used to solve for the root locus numerically for most any gain function.

The characteristics of a second order pair of complex poles are quantified by the natural frequency and damping ratio as described in Section 2.5. From the location of the poles on the s-plane, the natural frequency, ω_n, is the length of the vector from the origin to the pole, as shown in Figure 2-16, and the damping ratio, ζ, is the cosine of the angle between the real axis and the vector to the pole. The greater the distance the poles are from the origin, the faster the response of the system response, and the greater the angle between the real axis and the vector to the pole, the smaller the damping ratio, as shown in Figure 2-17, and greater the ringing and the less stable the system.

Figure 3-5 shows the trajectory of the poles for a constant ζ while varying ω_n. The poles move along a straight line at a distance of ω_n from the origin. These are lines of constant ζ. If the trajectory of the poles moves off a straight-line trajectory and move toward the imaginary axis, the system becomes less stable as the parameter is being varied. If the trajectory of the poles moves off the straight-line trajectory away from the imaginary axis, then the system becomes more stable as that parameter is varied. If the trajectory remains on the straight line, the bandwidth of the system increases or decreases while the damping ratio and stability remain constant.

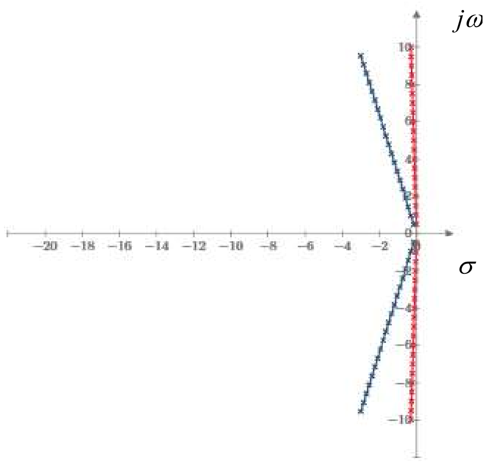


Figure 3-5 Locus of poles varying $0.5 \le \omega_n \le 10$ for $\zeta = 0.3$ (blue) and $\zeta = 0.03$ (red)

Figure 3-6 shows the trajectory of the polls for constant ω_n while varying ζ. For values of $\zeta \ge 1$ the poles are on the real axis. For values of $0 < \zeta < 1$ the poles split into a complex conjugate pair and move toward the imaginary axis at a radius of ω_n. In this plot, ζ is varied in equal increments, with the locations marked by the x's. This shows that the poles move much faster along the real axis as ζ approaches 1, and then more slowly again with decreasing ζ as they move away from the real axis as a complex conjugate pair and approach the imaginary axis. Once the poles leave the real axis and move closer to the imaginary axis, the damping ratio and stability of the system decreases as the controlling parameter is varied, while the natural frequency remains constant.

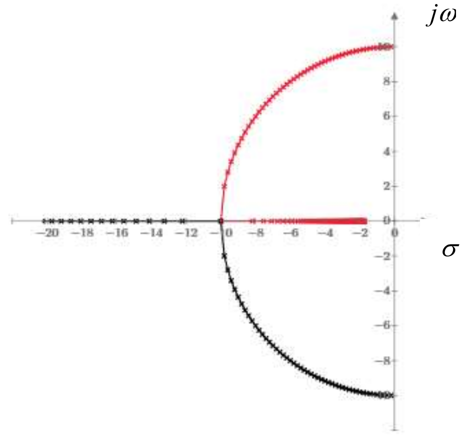


Figure 3-6 Locus of poles varying $3 \geq \zeta \geq 0.2$ for $\omega_n = 10$

The closed loop poles of a high gain feedback control system can be modified by the selection of poles, zeros, and gain used in the forward path of the control loop. The poles, or roots of the denominator of the closed loop gain equation move as a function of the gain of the loop transmission (value of $a(s) \cdot f(s)$ for $s = 0$). A plot of the location of the closed loop poles as a function of the loop gain, $a(0) \cdot f(0)$ is the traditional definition of the control system root-locus. Changing the loop gain impacts the closed loop dynamics and system stability. Looking at the trajectory of the closed loop pole locations as a function of the loop gain is a useful tool in choosing the appropriate loop gain to produce a stable system with the desired dynamics.

Solving for the poles of the closed loop gain can be difficult as the closed loop equation can be complex to factor. The functions for $a(s)$ and $f(s)$ generally start in factored form as they are constructed from factored poles and zeros, so the poles and zeros of the open loop transmission $-a(s) \cdot f(s)$ are known. But the denominator of the closed loop gain function $a(s)/(1 + a(s) \cdot f(s))$ is needed to determine the closed loop poles. The closed loop poles are the sum of the open loop zeros polynomial and the open loop poles polynomial, which, when added together, can be difficult to factor. To make the job of finding the root locus easier, there are a few simple heuristic rules that the locus of the roots follow as the value of $a(0) \cdot f(0)$ is increased. The rules allow constructing the root-locus without actually factoring the closed loop denominator. The rules were developed before tools like Mathcad were widely available that make it easy to solve for the roots of complex polynomial equations numerically. Although Mathcad is used to generate the root locus plots presented here, knowing the key rules gives insight into how to change the loop parameters to affect the trajectory of the roots and move them toward a desired location.

The key rules are[12]:

1. For small values of loop gain, $a(0) \cdot f(0)$, the closed loop poles start at the poles of $a(s) \cdot f(s)$
2. For large values of loop gain, the closed loop poles approach the zeros of $a(s) \cdot f(s)$ or infinity if the number of poles exceed the number of zeros.
3. The root locus always moves to the left of an odd number of singularities (poles and zeros) of $a(s) \cdot f(s)$.
4. The root locus will move in a way to attempt to preserve the center of mass of the poles (that's an oversimplification, but the effect will be seen in the root locus plots as poles moving in opposite directions toward or away from each other in pairs as the loop gain increases)

3.4 Sampled Data Systems

A control system can be implemented using analog circuitry or using a digital microcontroller. The analog circuitry is a continuous time system, whereas the digital microcontroller is a discrete time system. The advantage of using a microcontroller is all the sensor interfaces and loop filter equations can be implemented in code which can be easily modified, and the hardware is much simpler. Limiting and other non-linear control functions are much easier to implement in code than they are in analog hardware. Analog hardware can have advantages in applications requiring high frequency signals and wide loop bandwidths, but the speed of the Arduino microcontrollers, such as the SAMD21, are fast enough to run a control loop at a high enough sample rate, at or above 1 KHz, which is well above the bandwidth required for many mechanical servo systems, which is typically in the range of 1-10 Hz. For a canard based vertical trajectory system or a thrust vector control system, the microcontroller-based system is the better choice.

For a vertical trajectory system, the control loop microcontroller reads the attitude sensors, runs the control code, and controls the rotational angle of the canards. Being a discrete time control system, the control code samples the sensor data, calculates the control parameters, and adjusts the servo location once each pass through the loop. This makes the control system a sampled data system. The controller runs at a fixed loop rate to create a fixed sample time or sample rate. The control loop samples the sensors at the beginning of the control loop and uses those values for calculations done within that pass of the loop. At the end of the loop, the controller sets the new value for the servo angle based on the calculations done during that cycle. A sampler that holds a value for a sample period is called a sample and hold. Figure 3-7 shows the samples of a continuous time function using a sample-and-hold.

[12] See (Roberge, 1975) for a more complete explanation of root locus theory and techniques

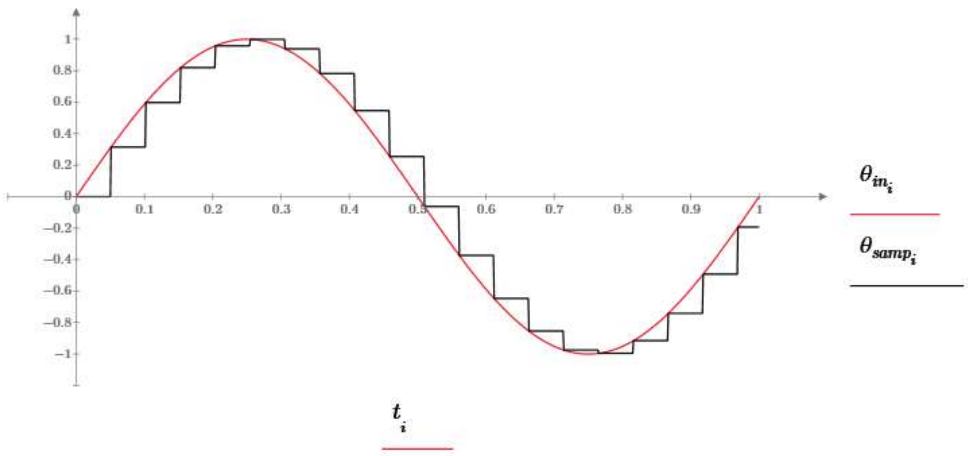


Figure 3-7 Sample and hold function for a 0.05 sec sample period

Unlike a continuous time system, a sampled data system folds frequencies that are above half the sampling frequency down into the frequency range below the sampling frequency. This frequency, $f_{nyquist} = \frac{1}{2} \cdot \frac{1}{t_{samp}}$ is called the Nyquist rate, where t_s is the sample period. To prevent higher frequencies signals from folding down into the operating frequency range of the control system, an anti-aliasing low-pass filter is placed ahead of the sampler. For sensors such as gyroscopes and accelerometers that are used to measure the rockets position, the actual signal sampling and digitization is done in the sensor IC. The result is passed to the microcontroller via an I2C, SPI, or serial data bus. The anti-aliasing filter is typically done on the sensor IC and is controlled as a part of the sensor IC setup. The maximum sample rate will be determined by how fast the control code can run on the selected microcontroller or by the maximum sample rate of the sensor IC. The maximum bandwidth of the control loop will need to be less than half the sampling frequency, or Nyquist rate.

A sample-and-hold has dynamics of its own. Equation (3.4-1) gives the frequency response of a sample and hold. In this function, $\frac{1}{s}$ is the transform of a step function, and $\frac{1}{s} \cdot e^{-s \cdot t_s}$ is the transform of a step function delayed by t_s. Subtracting these two step functions give a single sample. Multiplying by $\frac{1}{t_s}$ normalized the function to unity gain. Figure 3-8 and Figure 3-9 show the magnitude and phase of the sampler for a 1 ms sample period. The sample-and-hold phase starts to accumulate a decade below the sample frequency and reaches 90 degrees at half the sampler frequency, the Nyquist rate. That excess phase will impact the stability of the control loop, as will be seen in the next section. The maximum bandwidth of the control loop will need to be well below the Nyquist rate due to the excess phase of the sampler. Practically, the loop bandwidth will need to be less than a tenth the sampler frequency.

$$S(s) = \frac{1}{s \cdot t_s} \cdot (1 - e^{-s \cdot t_s}) \tag{3.4-1}$$

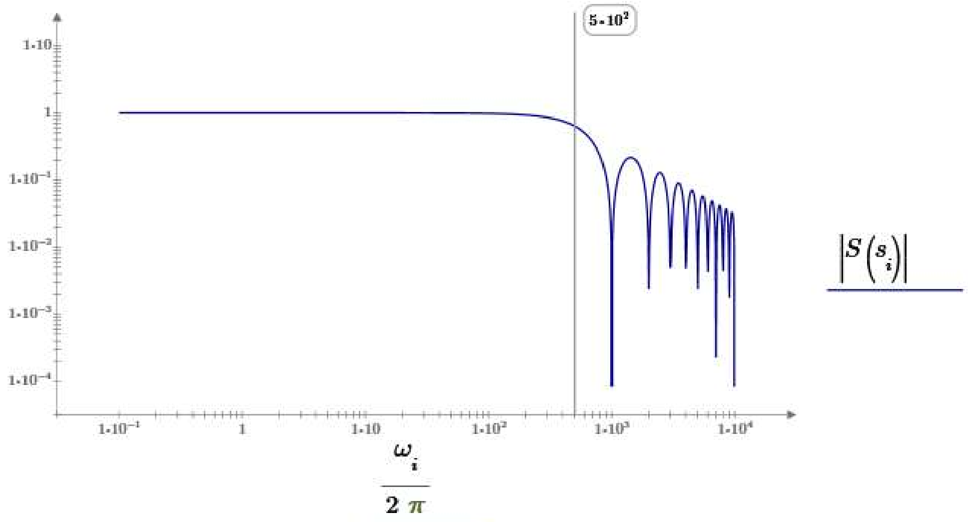


Figure 3-8 Magnitude of the sample-and-hold frequency response for a 1 ms sample time

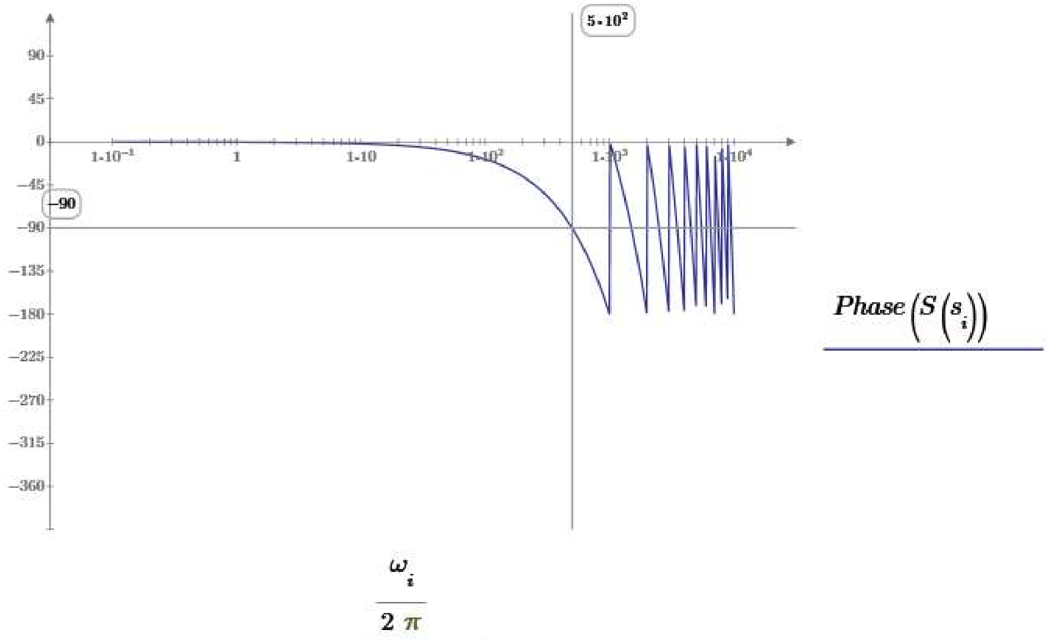


Figure 3-9 Sample-and-hold phase frequency response for a 1 ms sample time

3.5 Time-domain Control Loop Implementation Techniques

The easiest way to implement the control loop hardware is to use a microcontroller like an Arduino or a Raspberry Pi. An Arduino can easily interface to the sensors needed to determine the orientation of the rocket, such as a gyroscope, as well as other sensors to determine the rockets position, altitude, velocity, and acceleration. If an Arduino is used, the controller will be written in C. If a Raspberry Pi is used, the loop filter will most likely be written in Python. For modeling, the equations can be written using the Mathcad programming feature. The code can then be implemented in the real-time control system by easily translating that Mathcad code into C or Python.

As discussed in the section on the Sampled Data Systems, the speed of the inexpensive Arduino microcontrollers, such as the SAMD21, are fast enough to run a control loop at a high enough sample rate, at or above 1 KHz, which is well above the required thrust vector control loop bandwidth, which will be in the range of 1-10 Hz. The impact of the gain and phase of the sampler, as shown in Figure 3-8 and Figure 3-9 is minimal when the sampler frequency is ~100 times loop crossover frequency.

The design process starts with a frequency domain block diagram of the control loop. Then the control loop is implemented by translating the frequency domain blocks into their time domain equivalent. For a continuous-time system, the components would be integrators and differentiators, but because this is a discrete time implementation, the components are difference and summation equations that approximate their continuous time equivalents. There are multiple techniques for implementing difference equations that have different impacts on their behavior as the frequency of operation approaches the sample rate. But the various methods have little difference when the operating frequency is well below the sample rate[13], and the most straight forward method is used here.

3.5.1 Methodology – Basic Building Blocks

The time domain building blocks are integrators, differentiators, and gain blocks. The poles and zeros of the frequency domain block diagram are built out of these three basic building blocks. Because this is a real-time discrete time system, the differentiator and integrator and are implemented as difference and summation equations.

For the differentiator, the differentiation becomes a difference, expressed by

$$y = \frac{dx}{dt} \rightarrow \frac{\Delta x}{\Delta t} \tag{3.5-1}$$

and is implemented in code by

$$y_i = \frac{x_i - x_{i-1}}{t_i - t_{i-1}} = \frac{x_i - x_{i-1}}{t_{samp}} \tag{3.5-2}$$

where i indicates the current sample, $i-1$ indicates the prior time sample value , and t_{samp} is the sample time, or the time that it takes to make one pass through the control code loop. In a real-time system, a timer is used as a part of the control code to ensure the execution time for a pass through the loop is predictable and consistent as the sample time is a part of the calculations.

The integrator becomes a summer

$$y = \int x dt \rightarrow \sum x \cdot \Delta t \tag{3.5-3}$$

and is implemented in code by

$$y_i = y_{i-1} + x_i \cdot t_{samp} \tag{3.5-4}$$

---

[13] See (Astrom, 2011) Chapter 8 for a more complete explanation of time-domain approximations

It is easiest to build up the system by starting with the frequency domain block diagram, where the individual blocks are either integrators, differentiators, or gain blocks. Each block is then translated into the lines of time domain code necessary to implement that block, with the blocks connected together to implement the complete block diagram. Figure 3-10 to Figure 3-12 show the basic building blocks and the code needed to implement those blocks

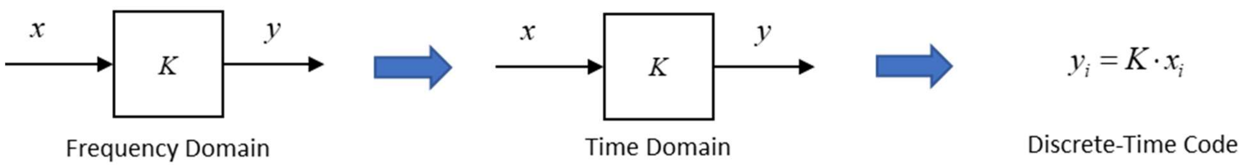


Figure 3-10 Gain block implementation

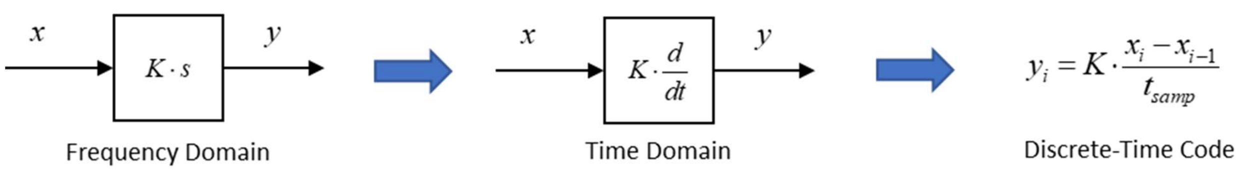


Figure 3-11 Differentiator implementation

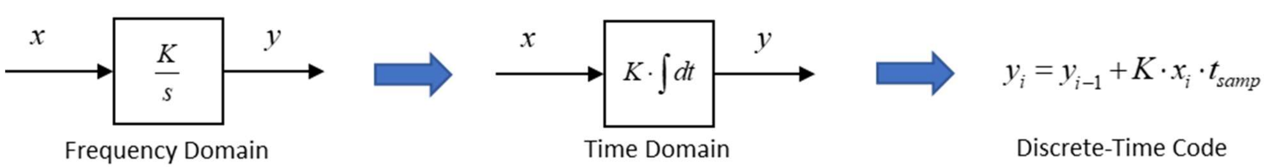


Figure 3-12 Integrator implementation

3.5.2 Intermediate Blocks

Figure 3-13 to Figure 3-16 shows a selection of intermediate building blocks that may be used to implement a system. Each figure shows the frequency domain block diagram, frequency domain transfer function, and the discrete-time time-domain coded needed to implement that block diagram.

Figure 3-15 shows the PID function which provides an integrator with two zeros. The PID configuration is just one of many configurations of blocks that used to implement a high gain loop filter. The zeros can be either real axis zeros at different frequencies or a complex zero pair. The transfer function equations show the coefficients needed to do either.

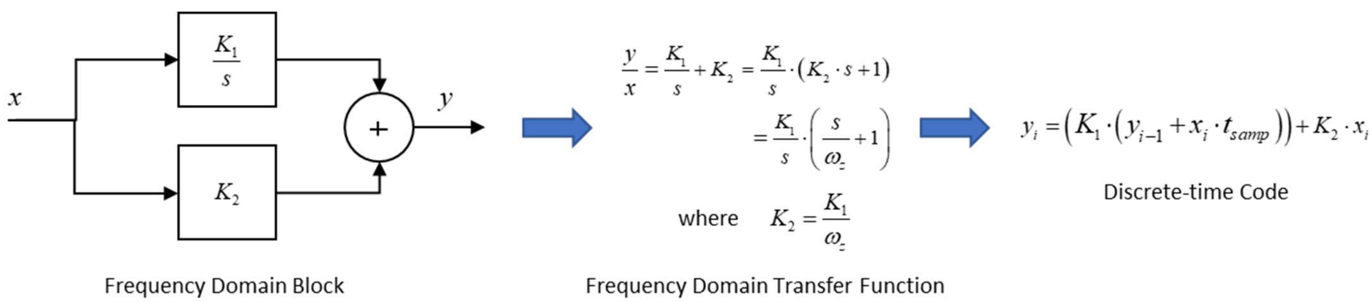


Figure 3-13 Integrator with gain and zero (PI function)

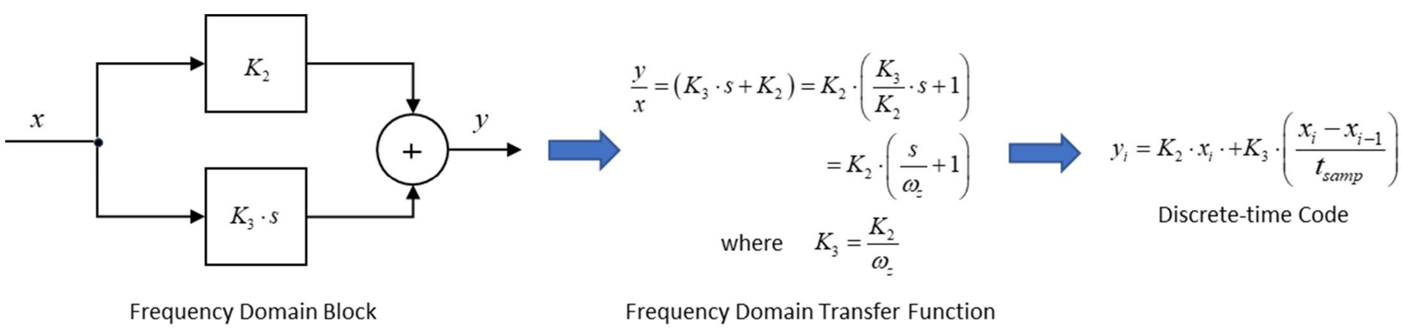
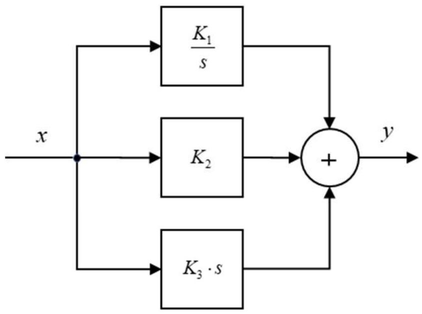


Figure 3-14 Gain with zero



Frequency Domain Block

$$\frac{y}{x} = \frac{K_1}{s} + K_2 + K_3 \cdot s = \frac{1}{s} \cdot \left(K_3 \cdot s^2 + K_2 \cdot s + K_1 \right)$$
$$= \frac{K_1}{s} \cdot \left(\frac{K_3}{K_1} \cdot s^2 + \frac{K_2}{K_1} \cdot s + 1 \right)$$
$$= \frac{K_1}{s} \cdot \left(\frac{s^2}{{\omega_{zn}}^2} + \frac{2 \cdot \zeta}{\omega_{zn}} + 1 \right)$$
$$= \frac{K_1}{s} \cdot \left(\frac{s}{\omega_{z1}} + 1 \right) \cdot \left(\frac{s}{\omega_{z2}} + 1 \right)$$
$$= \frac{K_1}{s} \cdot \left(\frac{1}{\omega_{z1} \cdot \omega_{z2}} \cdot s^2 + \frac{\omega_{z1} + \omega_{z2}}{\omega_{z1} \cdot \omega_{z2}} \cdot s + 1 \right)$$

where $K_3 = \frac{K_1}{{\omega_{zn}}^2}$ $\quad K_2 = \frac{2 \cdot \zeta \cdot K_1}{\omega_{zn}}$

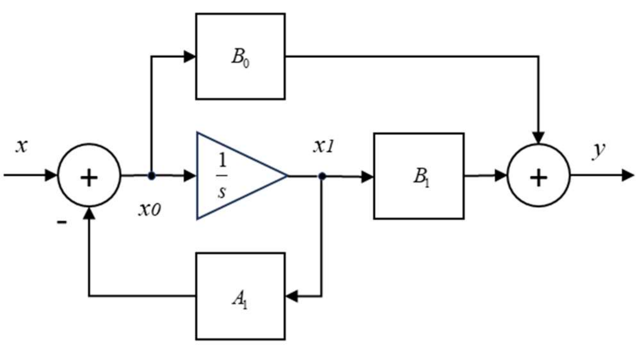
or $K_3 = \frac{K_1}{\omega_{z1} \cdot \omega_{z2}}$ $\quad K_2 = K_1 \cdot \frac{\omega_{z1} + \omega_{z2}}{\omega_{z1} \cdot \omega_{z2}}$

Frequency Domain Transfer Function

$$y_i = \left(K_1 \cdot \left(y_{i-1} + x_i \cdot t_{samp} \right) \right) + K_2 \cdot x_i + K_3 \left(\frac{x_i - x_{i-1}}{t_{samp}} \right)$$

Discrete-time Code

Figure 3-15 Integrator with gain and 2 zeros (PID function)



Frequency Domain Block

$$\frac{y}{x} = \frac{B_0 \cdot s + B_1}{s + A_1} = \frac{B_1}{A_1} \cdot \frac{\frac{B_0}{B_1} \cdot s + 1}{\frac{1}{A_1} \cdot s + 1}$$
$$= K \cdot \frac{\frac{s}{\omega_z} + 1}{\frac{s}{\omega_p} + 1}$$

For unity gain

$A_1 = B_1 = \omega_p$ $\quad b_0 = \frac{\omega_p}{\omega_z}$

Frequency Domain Transfer Function

$$x0_i = x_i - x1_{i-1} \cdot A_1$$
$$x1_i = x1_{i-1} + x0_i \cdot t_{samp}$$
$$y_i = x0_i \cdot B_0 + x1_i \cdot B_1$$

Discrete-time Code

Figure 3-16 Unity gain pole & zero (lead or lag network)

4 References

Astrom, K. J. (2011). *Computer-Controlled Systems, Theory and Design.* Mineola, New York: Dover.

Boyce, W. E., & DiPrima, R. C. (1969). *Elementary Differential Equations.* New York: John Wiley & Sons.

Chen, C.-T. (1993). *Analog and Digital Conrol System Design, Transfer-Function, State Space, and Algebraic Methods.* New York: Saunders College Publishing.

Franklin, G. F.-N. (2019). *Feedback Control of Dynamic Systems.* New York: Pearson.

Friedland, B. (1986). *Control System Design, An Introduction to State-Space Methods.* Mineola, New York: Dover Publications, Inc.

Liu, C., & Liu, J. W. (1975). *Linear Systems Analysis.* New York: McGraw-Hill.

Mathcad Home Page. (n.d.). Retrieved from https://www.mathcad.com/en/

Roberge, J. (1975). *Operational Amplifiers Theory & Practice.* New York: John Wiley & Sons.