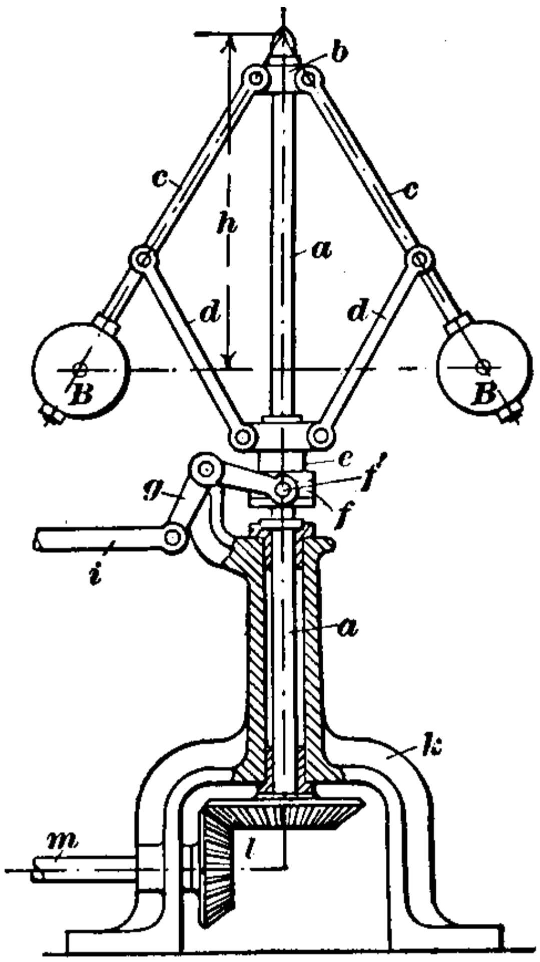
Control Theory Fundamentals

Seminar Manual



Author: Richard Poley

Mailing address:
Texas Instruments Inc.
MS.102
13905 University Blvd.
Sugar Land
TX 77479
USA

e-mail: poleyshark@gmail.com

Module Q3/1

Version 1

Revision 1

March 2015

Seminar materials may be downloaded at…
www.controltheoryseminars.com

Contents

Scope & Objectives

This seminar concerns control using feedback of linear time invariant systems. We will consider both continuous time (analogue) and discrete time (digital) systems.

The principal objectives of this seminar are to...

- Understand the concepts and key ideas in control
- Review the basic mathematical theory
- Know the language and the jargon

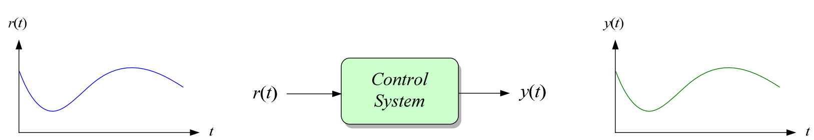
What is not covered in this seminar:

- System modelling
- Non-linear or time varying systems
- Advanced topics: e.g. adaptive, stochastic, optimal control, ...

What is Control?

"A control system is considered to be any system which exists for the purpose of regulating or controlling the flow of energy, information, money, or other quantities in some desired fashion."

William L. Brogan, *Modern Control Theory*, 1991



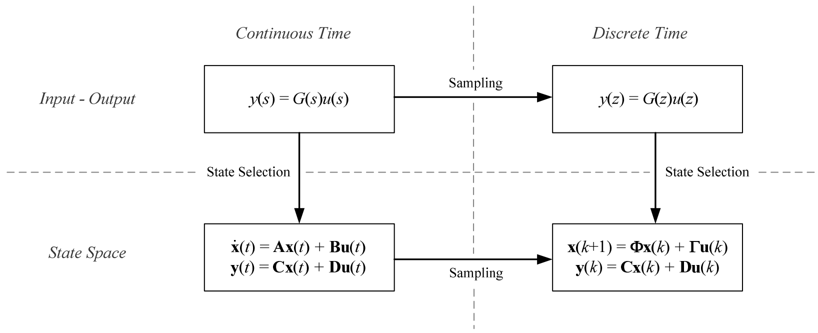
- The principal objective of control is to force the system output to accurately follow an input (**reference tracking**).

- Perfect reference tracking must be maintained even in the presence of disturbances (e.g. noise, load change, parameter variation, etc.) entering the system (**disturbance rejection**).

Limitations in physical systems makes perfect control impossible: compromises must be made!

Modelling Paradigms

In this seminar we will consider two different modelling paradigms: input-output and state space. Each may be used to model continuous time or discrete time systems.



- The process of **sampling** converts a continuous time to a discrete time system representation

- **State selection** is required to arrive at an equivalent state space representation.

Notation

❖ Important points are marked with a blue quad-bullet. Keywords are **highlighted** in this colour.

Signals are always represented by lower case symbols and transfer functions by upper case symbols.

$$y(s) = G(s)\, u(s)$$

$$g(t) = \mathscr{L}^{-1}\{\, G(s)\, \}$$

The independent variable may be omitted where the meaning is obvious from the context.

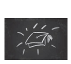
$$G = \frac{y}{u}$$

Matrices and vectors are represented by non-italic bold case. Matrices are upper case.

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t)$$

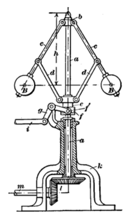
Differentiation will be denoted using prime or dot notation as appropriate:

$$x'(t) \qquad \dot{\mathbf{x}}(t)$$


0.0 Slides with associated tutorials are marked in the lower left corner.

Control Theory Seminar

1. Fundamental Concepts



- Linear Systems
- Transfer Functions
- Dynamic Response
- Classification of Systems

"Few physical elements display truly linear characteristics.... however, by assuming an ideal, linear physical element, the analytical simplification is so enormous that we make linear assumptions wherever we can possibly do so in good conscience."

Robert H. Cannon, Dynamics of Physical Systems, 1967

Linear Systems

All physical systems are inherently non-linear. Some examples of non-linearity are:

- Viscous drag coefficients depend on flow velocity
- Amplifier outputs saturate at supply voltage
- Coulomb friction present in mechanical moving objects
- Temperature induced parameter changes

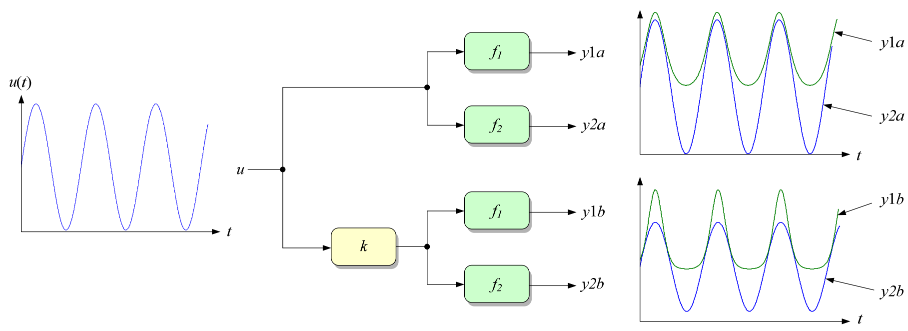
We approximate physical systems with **linear** models because convenient mathematical methods are available to handle them.

Linearisation of a non-linear model about an operating point can help to understand local behaviour, however global non-linear phenomena cannot be predicted by linear models.

- Multiple equilibriums
- Domains of attraction
- Chaotic response
- Limit cycles

Linearity

If a scaling factor is applied to the input of a **linear** system, the output is scaled by the same amount.



This is the **homogeneous** property of a linear system $\boxed{f(k\,u) = k\,f(u)}$

The **additive** property of a linear system is $\boxed{f(u_1 + u_2) = f(u_1) + f(u_2)}$

Terminology of Linear Systems

- Homogeneous and additive properties combine to form the principle of **superposition**, which all linear systems obey

$$f(k_1 u_1 \pm k_2 u_2) = k_1 f(u_1) \pm k_2 f(u_2)$$

The dynamics of a linear system may be captured in the form of an ordinary differential equation...

$$a_n \frac{d^n y}{dt^n} + \ldots + a_1 \frac{dy}{dt} + a_0 y = b_m \frac{d^m u}{dt^m} + \ldots + b_1 \frac{du}{dt} + b_0 u$$

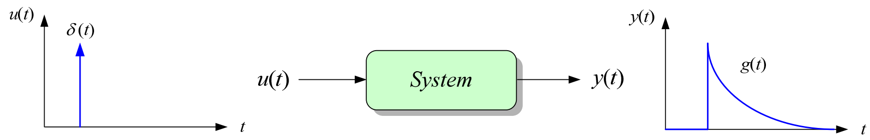
...or, using a more compact notation...

$$a_n y^{(n)} + \ldots + a_1 y' + a_0 y = b_m u^{(m)} + \ldots + b_1 u' + b_0 u$$

- If all the coefficients a_0, a_1, ... a_n and b_0, b_1, ... b_m are (real) constants, this equation is termed a **constant coefficient differential equation**, and the system is said to be **linear, time invariant** (LTI).

Convolution

The **impulse response** of a system is it's response when subjected to an impulse function, $\delta(t)$.



- If the impulse response $g(t)$ of a system is known, it's output $y(t)$ arising from any input $u(t)$ can be computed using a **convolution integral**

$$y(t) = g(t) * u(t) = \int_{-\infty}^{t} g(t-\tau) u(\tau) \, d\tau$$

This integral has a distinctive form, involving time reversal, multiplication, and integration over an infinite interval. It is cumbersome to apply for every $u(t)$.

The Laplace Transform

If $f(t)$ is a real function of time defined for all $t > 0$, the **Laplace transform** $f(s)$ is...

$$f(s) = \mathscr{L}\{f(t)\} = \int_{0^+}^{\infty} f(t) e^{-st} dt$$

...where s is an arbitrary complex variable.

Convolution	$\mathscr{L}\left\{ \int_0^t f_1(t-\tau) f_2(\tau) \, d\tau \right\} = f_1(s) f_2(s)$
Linearity	$\mathscr{L}\{ k_1 f_1(t) \pm k_2 f_2(t) \} = k_1 f_1(s) \pm k_2 f_2(s)$
Final value theorem	$\lim_{t \to \infty} f(t) = \lim_{s \to 0} s\, f(s)$
Shifting theorem	$\mathscr{L}\{ f(t-T) \} = e^{-sT} f(s)$

The Laplace transform converts time functions to frequency dependent functions of a complex variable, s.

Poles & Zeros

$$a_n y^{(n)}(t) + \ldots + a_1 y'(t) + a_0 y(t) = b_m u^{(m)}(t) + \ldots + b_1 u'(t) + b_0 u(t)$$

For zero initial conditions, the differential equation can be written in Laplace form as...

$$a_n s^n y(s) + \ldots + a_1 s\, y(s) + a_0 y(s) = b_m s^m u(s) + \ldots + b_1 s\, u(s) + b_0 u(s)$$

$$(\, a_n s^n + \ldots + a_1 s + a_0 \,)\, y(s) = (\, b_m s^m + \ldots + b_1 s + b_0 \,)\, u(s)$$

$$\alpha(s)\, y(s) = \beta(s)\, u(s)$$

The dynamic behaviour of the system is characterised by the roots of the two polynomials:

$$\beta(s) = b_m s^m + \ldots + b_1 s + b_0 = 0$$

$$\alpha(s) = a_n s^n + \ldots + a_1 s + a_0 = 0$$

- The m roots of $\beta(s)$ are called the **zeros** of the system
- The n roots of $\alpha(s)$ are called the **poles** of the system

The Transfer Function

$$u(s) \longrightarrow \boxed{G(s)} \longrightarrow y(s)$$

The ratio $\dfrac{\beta(s)}{\alpha(s)}$ is called the **transfer function** of the system.

$$G(s) = \frac{y(s)}{u(s)} = \frac{\beta(s)}{\alpha(s)} = \frac{b_m s^m + \ldots + b_1 s + b_0}{a_n s^n + \ldots + a_1 s + a_0}$$

- The transfer function of a system is the Laplace transform of its impulse response

$$y(t) = g(t) * u(t) = \mathscr{L}^{-1}\{\, G(s)\, u(s) \,\}$$

The quantity $n - m$ is called the **relative degree** of the system. Systems are classified according to their relative degree, as follows...

- **strictly proper** if $m < n$
- **proper** if $m \leq n$
- **improper** if $m > n$

Transient Response

Numerator & denominator can be factorised to express the transfer function in terms of poles & zeros.

$$y(s) = k\,\frac{(s-z_1)(s-z_2)\ldots(s-z_m)}{(s-p_1)(s-p_2)\ldots(s-p_n)}\,u(s)$$

This rational function yields q terms through partial fraction expansion

$$y(s) = \frac{\varepsilon_1}{s-r_1} + \frac{\varepsilon_2}{s-r_2} + \ldots + \frac{\varepsilon_q}{s-r_q}$$

❖ Since all a_i, b_i are real, $r_1 \ldots r_n$ are always either real or complex conjugate pairs

The time response is a sum of exponential terms, where each index is a denominator root.

$$y(t) = \underbrace{\varepsilon_1 e^{r_1 t} + \varepsilon_2 e^{r_2 t} + \ldots + \varepsilon_n e^{r_n t}}_{\substack{y_c(t) \\ \text{Transient response}}} + \underbrace{\varepsilon_{n+1} e^{r_{n+1} t} + \ldots + \varepsilon_q e^{r_q t}}_{\substack{y_p(t) \\ \text{Steady state response}}}$$

The n terms in $y(t)$ with roots originating from $G(s)$ comprise the **transient response**, while the $q-n$ terms originating from $u(s)$ comprise the **steady state response**.

Stability

For stability we require that the transient part of the response decays to zero, *i.e.* $y_c(t) \to 0$ as $t \to \infty$.

The transient response is defined by the first n exponential terms in $y(t)$

$$y_c(t) = \varepsilon_1 e^{r_1 t} + \varepsilon_2 e^{r_2 t} + \ldots + \varepsilon_n e^{r_n t}$$

...where complex roots have the form $r_i = \sigma_i \pm j\omega_i$

For real systems complex roots always arise in conjugate pairs, so terms involving complex exponential pairs arise in the time response.

$$y_c(t) = \varepsilon_1 e^{r_1 t} + \varepsilon_1^* e^{r_1^* t} + \ldots$$

Therefore the transient part of the response will include oscillatory terms, the amplitude of each being constrained by an exponential.

$$y_c(t) = A_1 e^{\sigma_1 t} \sin(\omega_1 t + \phi_1) + \ldots$$

❖ For stability we require that the real part (σ_i) of every r_i in $G(s)$ be negative.

First Order Systems

The dynamics of a classical first order system are defined by the differential equation

$$\tau\, y'(t) + y(t) = u(t)$$

...where the parameter τ represents the **time constant** of the system.

Taking Laplace transforms and re-arranging to find the transfer function...

$$\tau\, s\, y(s) + y(s) = u(s)$$

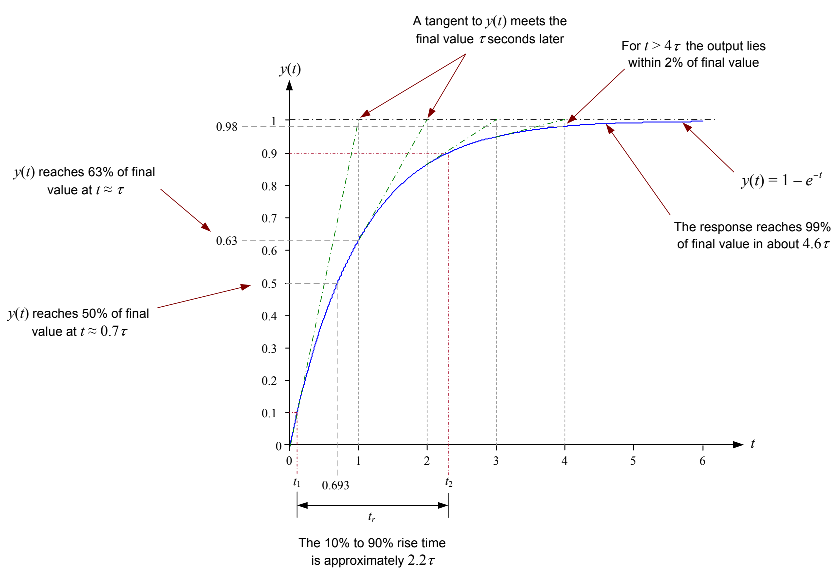
$$\frac{y(s)}{u(s)} = \frac{1}{s\tau + 1}$$

The output $y(t)$ for any input $u(t)$ can be found using the method of Laplace transforms.

$$y(t) = \mathscr{L}^{-1}\left\{u(s)\frac{1}{s\tau+1}\right\}$$

The response following a unit step input is: $y(t) = 1 - e^{-\frac{t}{\tau}}$

First Order Step Response



Unit step response for first order system with $\tau = 1$.

Second Order Systems

Linear constant coefficient second-order differential equations of the form

$$y''(t) + 2\zeta\omega_n y'(t) + \omega_n^2 y(t) = \omega_n^2 u(t)$$

are important because they often arise in physical modelling.

Dynamic behaviour is defined by two parameters:

ζ is called the **damping ratio**

ω_n is called the **un-damped natural frequency**

The transfer function of the second order system is $\dfrac{y(s)}{u(s)} = \dfrac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$

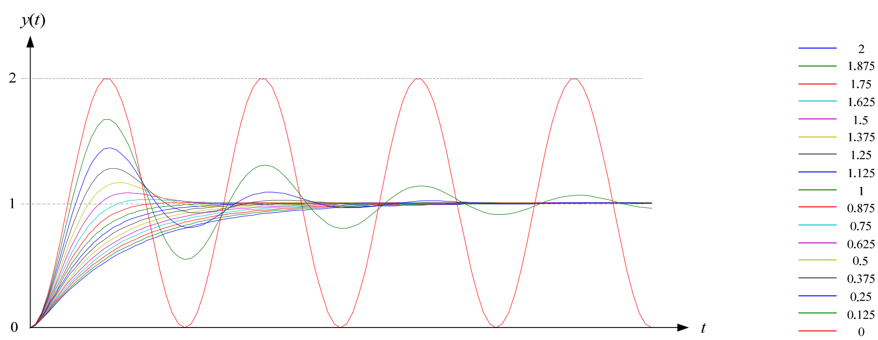
...from which we get the **characteristic equation** $s^2 + 2\zeta\omega_n s + \omega_n^2 = 0$

The poles of the second-order linear system are at $s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$

Classification of Second Order Systems

Dynamic response of the second order system is classified according to damping ratio.

Damping ratio	Roots	Classification
$\zeta > 1$	$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$	**over-damped**
$\zeta = 1$	$s = -\omega_n$	**critically damped**
$0 < \zeta < 1$	$s = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$	**under-damped**
$\zeta = 0$	$s = \pm j\omega_n$	**un-damped**



The Under-Damped Response

In the under-damped case ($0 < \zeta < 1$) we have a pair of complex conjugate roots at

$$s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$$

Real and imaginary parts are denoted $s = -\sigma \pm j\omega_d$

σ is termed the **damping coefficient**. It is related to a virtual "time constant" by $\tau = \frac{1}{\sigma}$

ω_d is the **damped natural frequency** of the system: $\omega_d = \omega_n\sqrt{1-\zeta^2}$

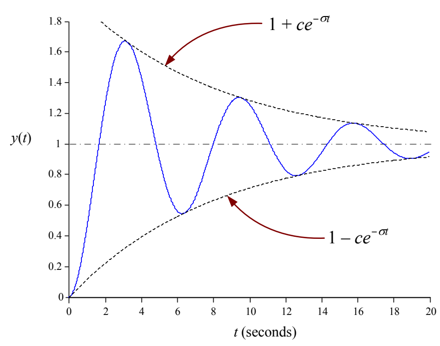
❖ The unit step response of the under-damped second order system is

$$y(t) = 1 - \frac{\omega_n}{\omega_d} e^{-\sigma t} \sin(\omega_d t + \phi)$$ ...where $\phi = \cos^{-1}\zeta$

Transient Decay Envelope

$$y(t) = 1 - \frac{\omega_n}{\omega_d} e^{-\sigma t} \sin(\omega_d t + \phi)$$

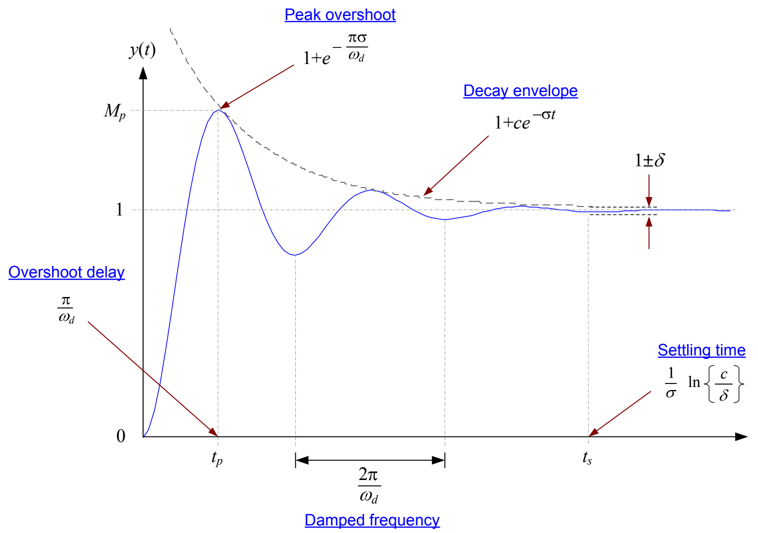
The under-damped unit step response comprises an oscillation of frequency ω_d and phase ϕ, constrained within a decaying exponential envelope determined by σ and ζ.



$c = \frac{\omega_n}{\omega_d}$

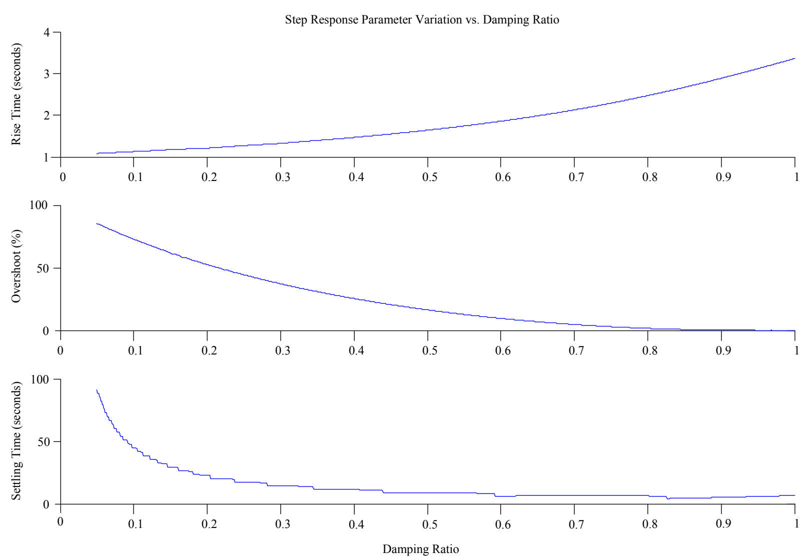
Unit step response with $\omega_n = 1$ & $\zeta = 0.125$

Second Order Step Response



Characteristics of the unit step response of the under-damped ($\zeta = 0.25$) second order system $\frac{1}{s^2+0.5s+1}$

Step Response Specifications



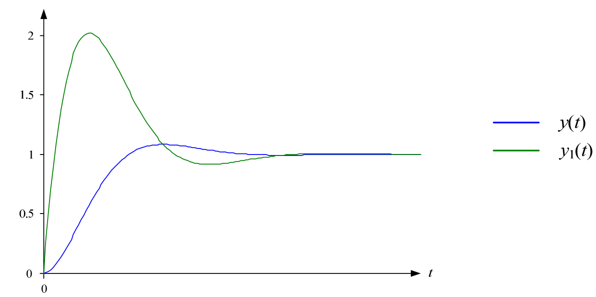
Plots show variation in rise time, over-shoot, and settling time for a second order system with $\omega_n = 1$

Effects of LHP Zero

Adding a LHP zero at $s = -z$ to the original transfer function: $G_1(s) = \left(1 + \frac{s}{z}\right) G(s)$

For a unit step input $y_1(s) = \frac{1}{s} G_1(s) = y(s) + \frac{1}{z} s\, y(s)$

Taking inverse Laplace transforms $y_1(t) = y(t) + \frac{1}{z} y'(t)$

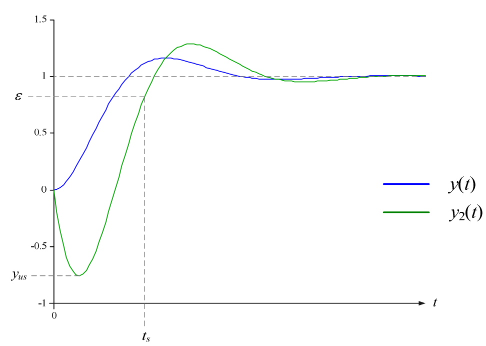


The effect of adding a LHP zero is to add a derivative term to the step response of the original system

❖ In general, rise time is decreased and overshoot increased by a LHP zero

Effects of RHP Zero

Similarly, adding a RHP zero at $s = z$ changes the response according to $y_2(t) = y(t) - \frac{1}{z} y'(t)$



Peak under-shoot is bounded by:

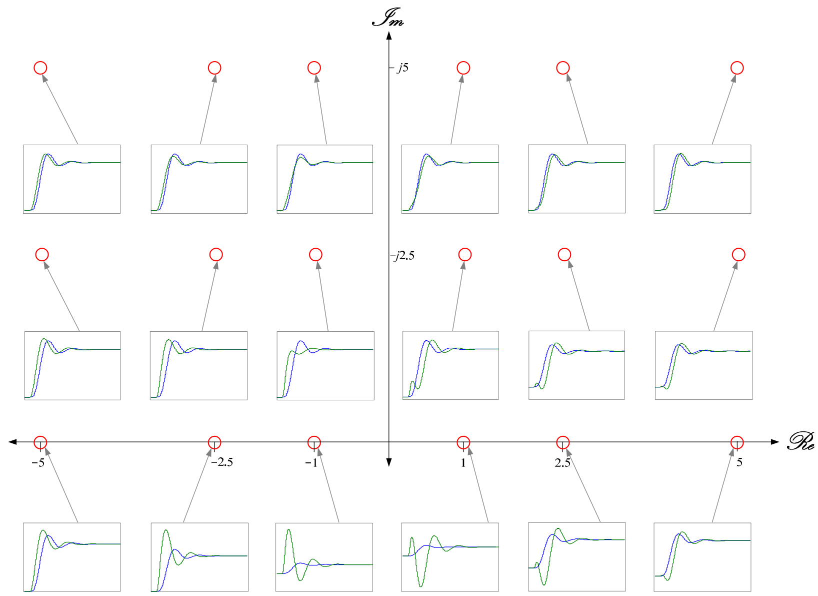
$$|y_{us}| \geq |y_f| \, \frac{1-\varepsilon}{e^{\alpha t_s} - 1}$$

y_f = final value
t_s = settling time
ε = error bound
α = Re(z)

The step response of a stable plant with n real RHP zeros will cross it's starting value at least n times.

❖ The effect of adding an RHP zero is to increase rise time (make the response slower) and induce undershoot

Effect of Zero Location on Transient Response



Effect on step response of adding a zero pair to a stable system

Time Delay Approximation

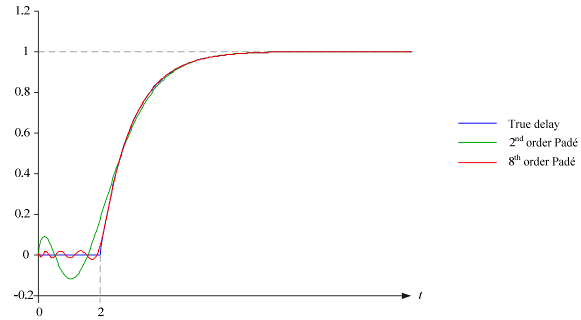
Time delay can be approximated by a rational transfer function with n real RHP zeros...

$$e^{-\theta s} \approx \frac{\left(1 - \frac{\theta}{2n} s\right)^n}{\left(1 + \frac{\theta}{2n} s\right)^n}$$

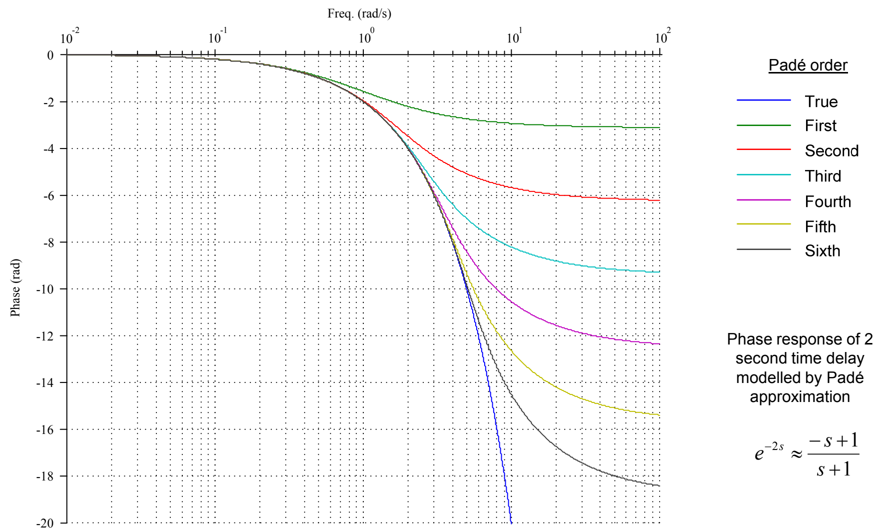
The **Padé approximation** is only valid at low frequencies, so it is important to compare the true and modelled responses to choose the right approximation order and check its' validity.

Plot shows 2^{nd} & 8^{th} order Padé approximations to the transfer function

$$G(s) = \frac{e^{-2s}}{s+1}$$

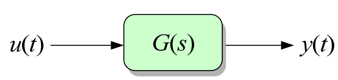


Phase Delay



❖ The phase of the Padé approximation departs from that of true time delay at frequencies beyond that of the pole-zero pairs.

Frequency Response



If the steady state sinusoid $u(t) = u_0 \sin(\omega t + \alpha)$ is applied to a linear system $G(s)$, the output is

$$y(t) = y_0 \sin(\omega t + \beta)$$

- The amplitude is modified by $\dfrac{y_0}{u_0}$
- The phase is shifted by $\phi = \beta - \alpha$

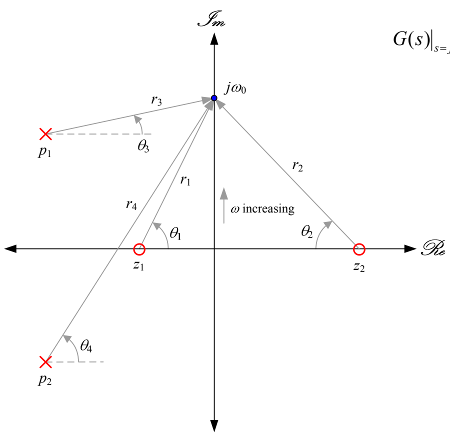
❖ Amplitude and phase change from input to output are determined by $G(j\omega)$:

$$\boxed{\frac{y_0}{u_0} = |G(j\omega)|} \qquad \boxed{\phi = \angle G(j\omega)}$$

Frequency Response

Response of $G(s)$ at each frequency can be determined directly from the pole-zero map

For example, at frequency ω_0 the transfer function $G(s)=\dfrac{(s-z_1)(s-z_2)}{(s-p_1)(s-p_2)}$ has the response



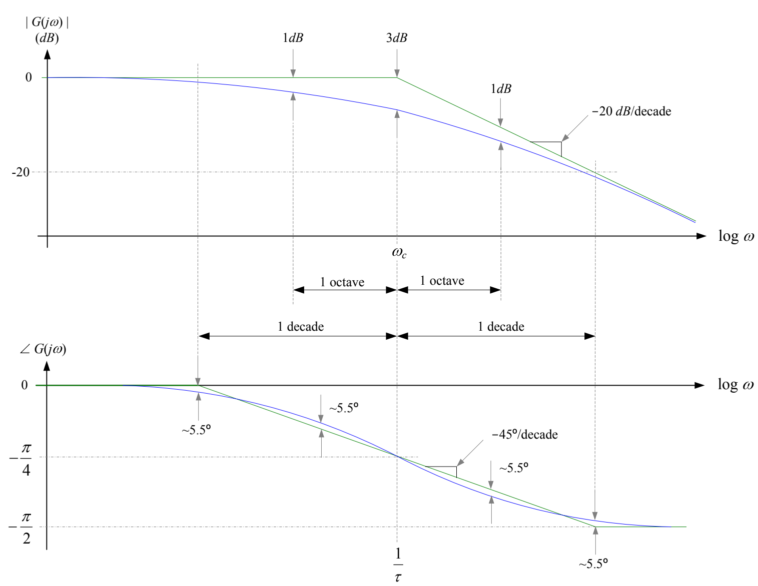
$$G(s)\big|_{s=j\omega_0}=\frac{(j\omega_0-z_1)(j\omega_0-z_2)}{(j\omega_0-p_1)(j\omega_0-p_2)}=|G(j\omega_0)|\angle G(j\omega_0)$$

Modulus and argument are found from:

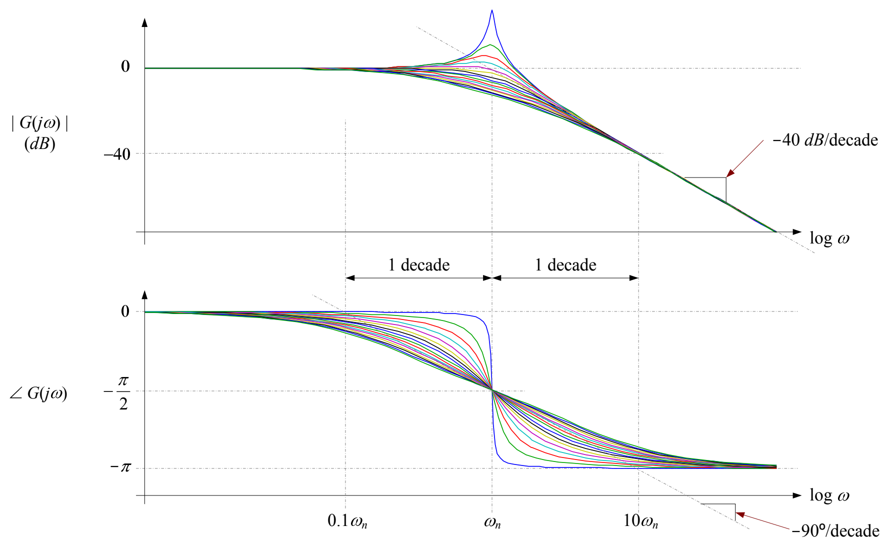
$$|G(j\omega_0)|=\frac{r_1\,r_2}{r_3\,r_4}$$

$$\angle G(j\omega_0)=\theta_1+\theta_2-\theta_3-\theta_4$$

First Order Bode Asymptotes



Second Order Bode Asymptotes

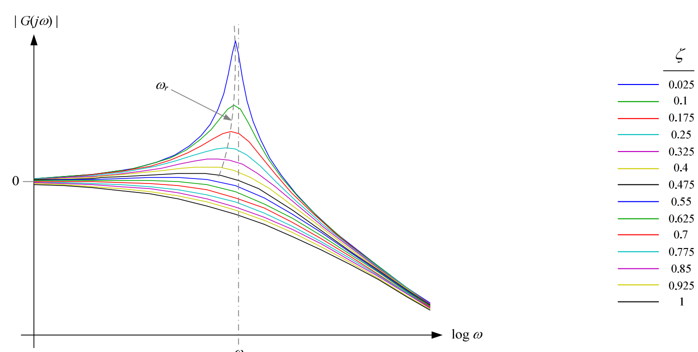


Plots shown for damping ratios: $0.025 \le \zeta \le 2$

Resonant Peak

The resonant peak occurs at frequency $\omega_r = \omega_n\sqrt{1-2\zeta^2} \quad : 0 < \zeta < \frac{1}{\sqrt{2}}$

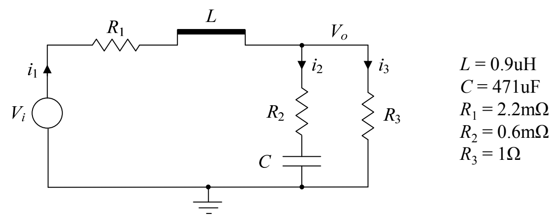
Resonant peak magnitude is given by $M_r = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$



The resonant peak ω_r approaches ω_n as damping ratio approaches zero: $\omega_r \to \omega_n$ as $\zeta \to 0$

Electrical Network Example

The network shown below represents the output filter of a switching power converter operating in continuous conduction mode.



The transfer function contains a single real LHP zero and a pair of complex conjugate poles lying relatively far to the left in the LHP.

$$\frac{V_o(s)}{V_i(s)} = \frac{sCR_2R_3 + R_3}{s^2LC(R_2+R_3) + s(CR_1(R_2+R_3) + CR_2R_3 + L) + (R_1+R_3)}$$

This implies an oscillatory transient response and peaking in the Bode plot.

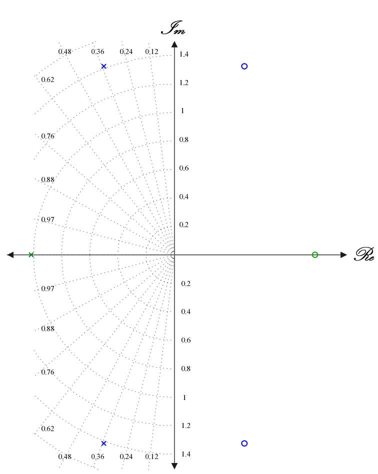

1.1

All-Pass Transfer Functions

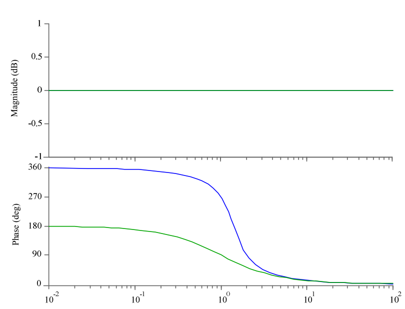
An **all-pass** transfer function G_{ap} passes all frequencies with the same attenuation.

Such a transfer function has pole-zero symmetry about the imaginary axis.

i.e. if s_0 is a zero, then $-s_0*$ is a pole.



Examples are: $\dfrac{s-1}{s+1}$ $\dfrac{s^2-s+2}{s^2+s+2}$



Minimum Phase Systems

A **minimum phase** transfer function G_{mp} meets the following criteria:

- No time delay
- No RHP zeros
- No poles on the imaginary axis (except the origin)
- No unstable poles

Examples are: $1 \quad \frac{1}{s} \quad \frac{s}{s+1} \quad \frac{s^2+2}{s^2+s+2}$

For a minimum phase system, total phase variation is $(n-m)\frac{\pi}{2}$ over $0 < \omega < \infty$.

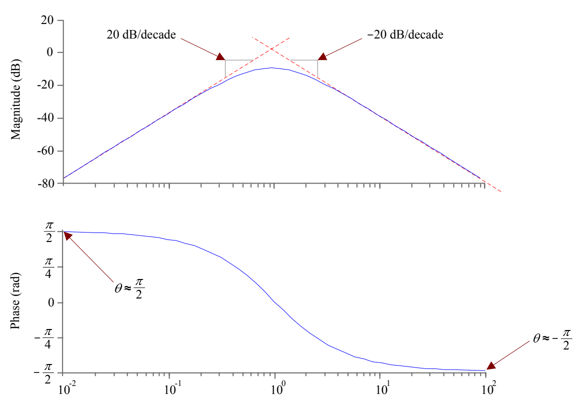
A **non-minimum phase** transfer function exhibits more negative phase.

❖ Any stable, proper, real-rational transfer function G can always be written in terms of minimum-phase and all-pass transfer functions: $G = G_{ap}G_{mp}$

Phase Area Formula

For minimum phase systems, gain and phase curves on the Bode plot are approximately related through a derivative:

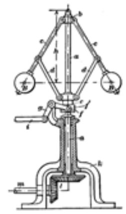
$$\angle G(j\omega) \approx \frac{\pi}{2}\frac{d\log|G(j\omega)|}{d\log(\omega)}$$



For a constant gain slope p, the phase curve has the asymptotic value $\theta = p\frac{\pi}{2}$

Control Theory Seminar

2. Feedback Control

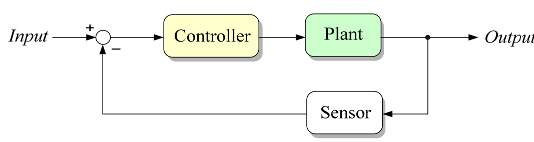


- Effects of Feedback
- The Nyquist Plot
- Phase Compensation
- Sensitivity & Tracking
- Internal Model Control

"...by building an amplifier whose gain is deliberately made, say 40 decibels higher than necessary, and then feeding the output back to the input in such a way as to throw away that excess gain, it has been found possible to effect extraordinary improvement in constancy of amplification and freedom from non-linearity."

Harold S. Black, *Stabilized Feedback Amplifiers*, 1934

Effects of Feedback

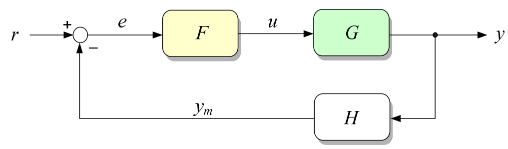


Feedback (also called "closed loop control") is a simple but tremendously powerful idea which has revolutionised many engineering applications.

When properly applied, feedback can...

- Reduce or eliminate steady state error
- Reduce the sensitivity of the system to parameter changes
- Change the gain or phase of the system over some desired frequency range
- Cause an unstable system to become stable
- Reduce the effects of load disturbance and noise on system performance
- Linearise a non-linear component

Notation



Signals

r = reference input

e = error signal

u = control effort

y = output

y_m = feedback

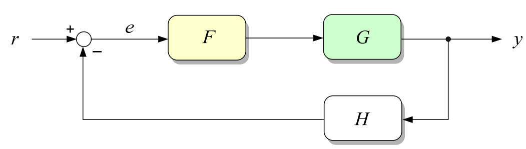
Transfer Functions

F = ***controller***

G = ***plant***

H = ***sensor***

Negative Feedback



Error equation is $e = r - Hy$

Output equation is $y = FGe$

Combining error and output equations gives $y = FG\,(r - Hy)$

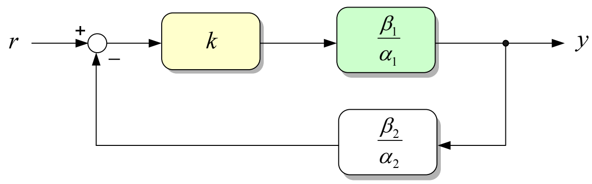
$$y\,(1 + FGH) = FGr$$

- The **closed loop** transfer function is $$\frac{y}{r} = \frac{FG}{1 + FGH}$$

- The **open loop** transfer function is $$L = FGH$$

Closed Loop Transfer Function

Define the transfer functions of the forward and feedback elements as $F = k$, $G = \dfrac{\beta_1}{\alpha_1}$ and $H = \dfrac{\beta_2}{\alpha_2}$



$$\frac{y}{r} = \frac{k\dfrac{\beta_1}{\alpha_1}}{1 + k\dfrac{\beta_1}{\alpha_1}\dfrac{\beta_2}{\alpha_2}}$$

The closed loop transfer function is

$$\boxed{\frac{y}{r} = \frac{k\,\beta_1\,\alpha_2}{\alpha_1\,\alpha_2 + k\,\beta_1\,\beta_2}}$$

Closed Loop Stability

❖ The criterion for **external stability** is that the closed loop transfer function has no RHP poles

There are two equivalent ways to pose this question:

1. Writing the closed loop transfer function as...

$$\frac{y}{r} = \frac{FG}{1 + L}$$

...we require that there are no RHP zeros in $1 + L$

2. Writing the closed loop transfer function as...

$$\frac{y}{r} = \frac{k\,\beta_1\,\alpha_2}{\alpha_1\,\alpha_2 + k\,\beta_1\,\beta_2}$$

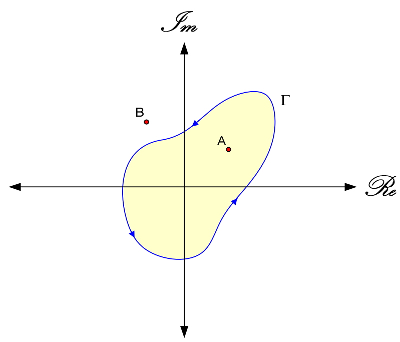
...we require that there are no positive roots of $\alpha_1\alpha_2 + k\,\beta_1\,\beta_2 = 0$

In this section we address the first question from the perspective of the frequency domain. A method which answers the second question in terms of the time domain is outlined in section 3.

Encirclement & Enclosure

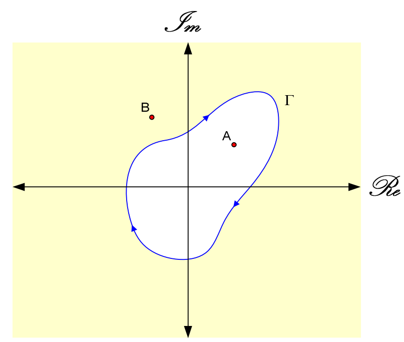
A complex point or region is **encircled** if it is found inside a closed path.

A complex point or region is **enclosed** if it is found to the left of the path when the path is traversed in the CCW direction.



A encircled & enclosed

B not encircled or enclosed

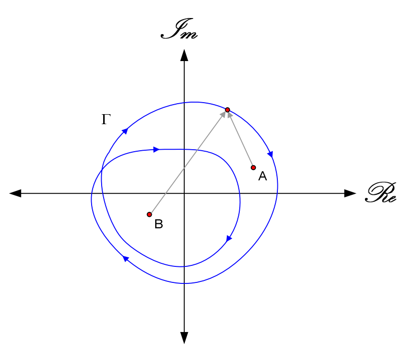


A encircled but not enclosed

B enclosed but not encircled

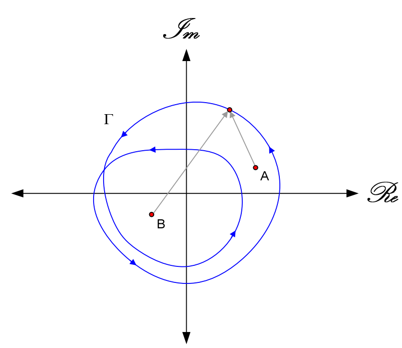
Multiple Encirclements

A point in the complex plane can be encircled multiple times.



A encircled once
B encircled twice

A not enclosed
B not enclosed

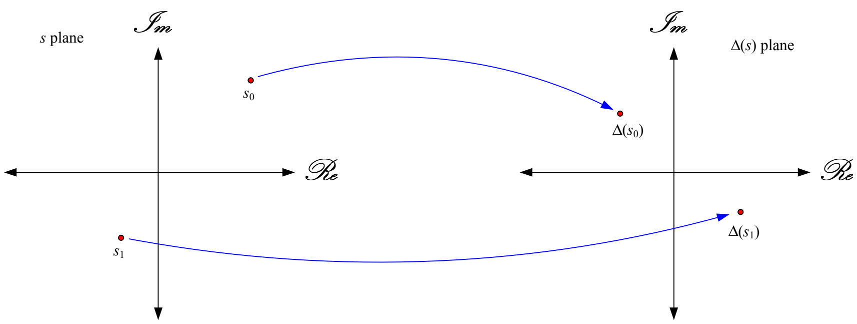


A encircled once
B encircled twice

A enclosed
B enclosed

Mapping

A complex function of a complex variable cannot be plotted on a single set of axes. We need two separate complex planes: the s plane and the function plane. The correspondence between points in the two planes is called **mapping**.



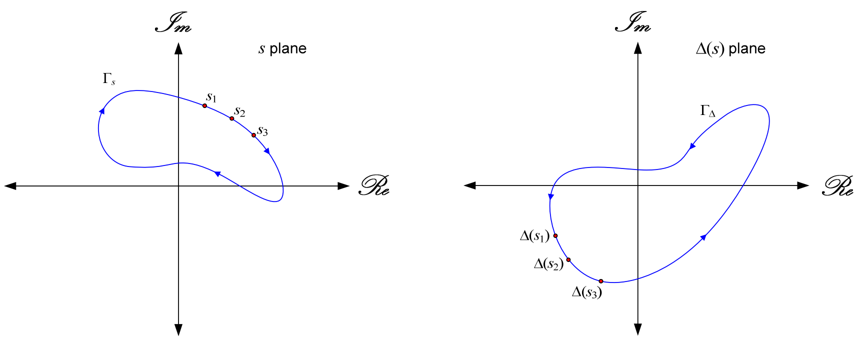
If each point in the s plane maps to one (and only one) point in the function space, the function is called **single valued**. A transfer function is an example of a single valued complex function.

The transfer function $\Delta(s)$ uniquely maps points in the s plane to points in the $\Delta(s)$ plane.

Contour Mapping

Let $\Delta(s)$ be a single valued function, and Γ_s represent an arbitrary closed contour in the s plane.

If Γ_s does not pass through any poles of $\Delta(s)$, then its image Γ_Δ is also closed.



Depending on $\Delta(s)$, the direction of Γ_Δ can be the same as, or opposite to that of Γ_s.

Principle of the Argument

Assuming that Γ_s encircles Z zeros and P poles of $\Delta(s)$, define the integer N :

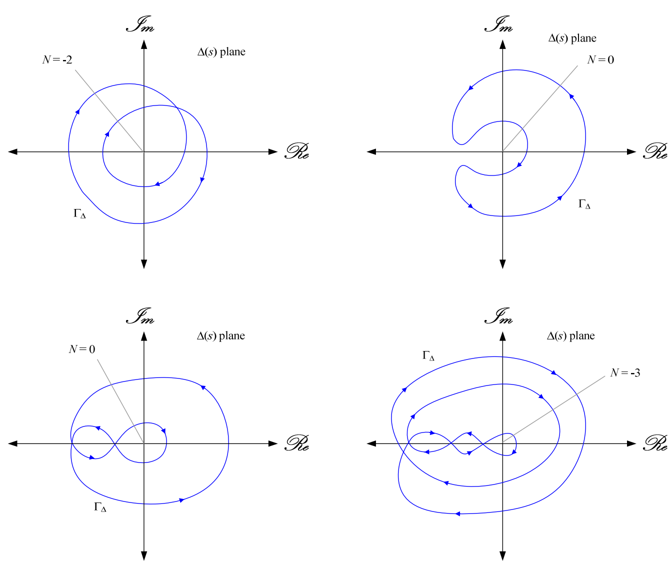
$$N = Z - P$$

The **principle of the argument** states that Γ_Δ will encircle the origin of the $\Delta(s)$ space exactly N times

The direction of encirclement is as follows:

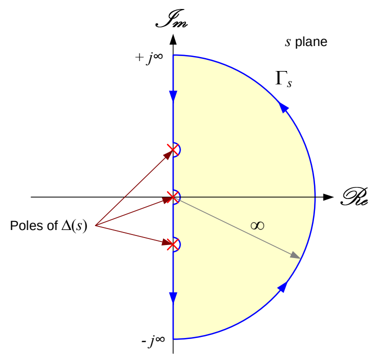
- $N > 0 \quad (Z > P)$: Γ_Δ encircles the origin N times in the same direction as Γ_s
- $N = 0 \quad (Z = P)$: Γ_Δ does not encircle the origin of the $\Delta(s)$ space
- $N < 0 \quad (Z < P)$: Γ_Δ encircles the origin N times in the opposite direction to Γ_s

Determination of N



❖ By convention, counter-clockwise encirclement is regarded as positive.

The Nyquist Path

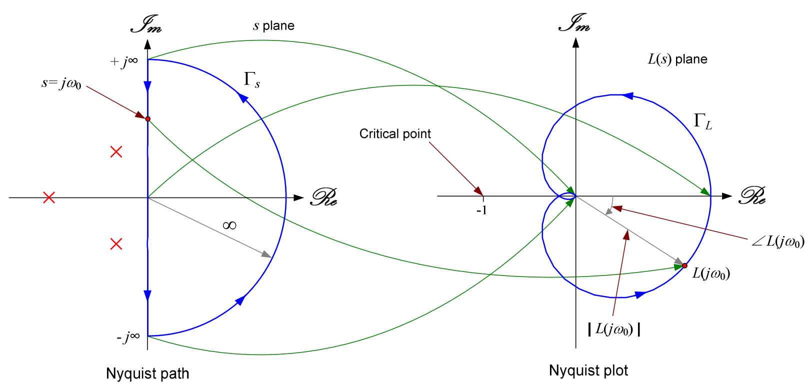


Indentations on the imaginary axis are necessary to ensure Γ_s does note pass through any poles of $\Delta(s)$.

Any RHP pole or zero of $\Delta(s)$ is enclosed by the Nyquist path.

The Nyquist Plot

The **Nyquist plot** is the image of the loop transfer function $L(s)$ as s traverses the Γ_s contour.



Since we are interested in roots of $1 + L(s)$ we examine enclosure relative to the point $[-1,0]$.

Nyquist Stability Criteria

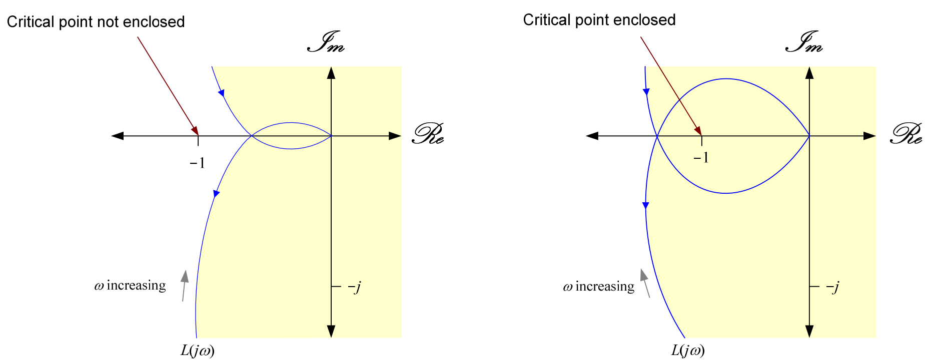
Recall, for closed loop (external) stability we require no RHP zeros in $1 + L(s)$. *i.e.* $N = -P$

- For closed loop stability, the Nyquist plot must encircle the critical point once for each RHP pole in $L(s)$, and any encirclement must be made in the opposite direction to Γ_s.

For minimum phase systems: $N = 0$

- The simplified Nyquist stability criterion for minimum phase systems states that the feedback system is stable if the Nyquist plot does not enclose the critical point.

Enclosure of the Critical Point



Closed loop stable

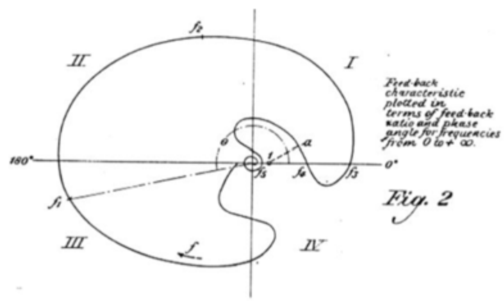
Closed loop unstable

Note: The convention of CCW traversal of the Nyquist path means the direction of Γ_L follows decreasing positive frequency.

Nyquist's Paper

Nyquist, H. 1932. Regeneration Theory. Bell System Technical Journal, 11, pp. 126-147

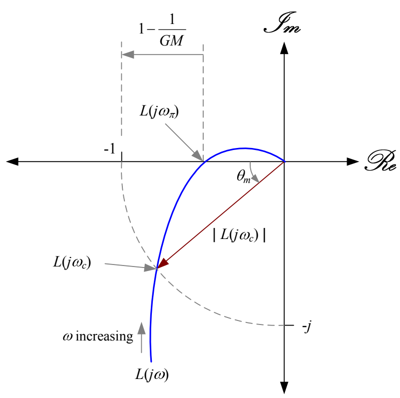
Nyquist's paper changed the process of feedback control from trial-and-error to systematic design.



Nyquist had the critical point at +1. Bode changed it to -1.

Relative Stability

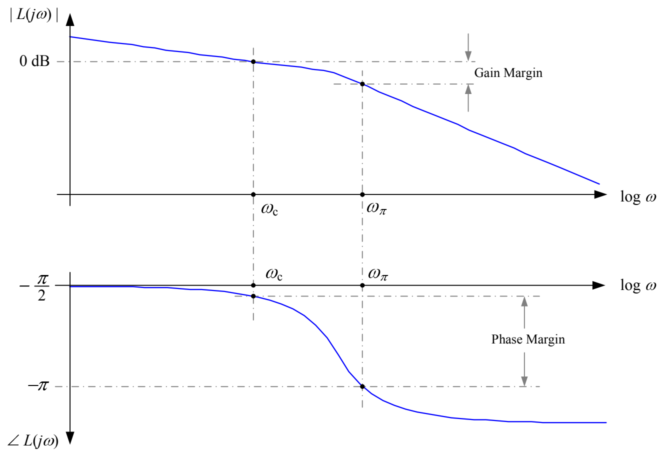
The proximity of the $L(s)$ curve to the critical point is a measure of **relative stability**, which is often used as a performance specification on the feedback system



- Gain Margin (GM) is defined as: $GM = \dfrac{1}{|L(j\omega_\pi)|}$ ... where ω_π is the **phase crossover frequency**
- Phase Margin (PM) is defined as: $PM = \angle L(j\omega_c) + \pi$ ... where ω_c is the **gain crossover frequency**

Stability Margins

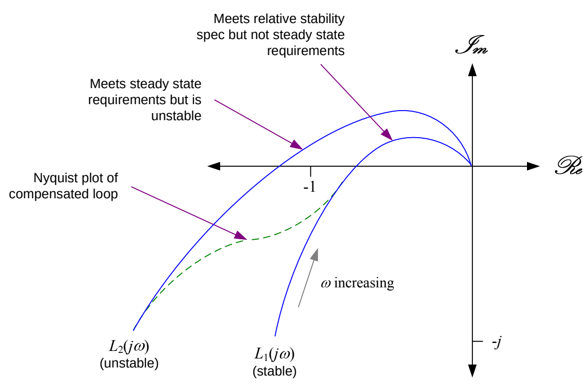
Gain & phase margins can be read directly from the Bode plot.



A rule-of-thumb for minimum phase systems is that the closed loop will be stable if the slope of $|L(j\omega)|$ is -2 or less at the cross-over frequency (ω_c). This follows from the phase area formula.

Phase Compensation

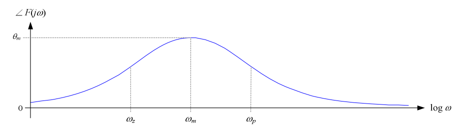
When relative stability specifications cannot be met by gain adjustment alone, **phase compensation** techniques may be applied to change the Nyquist curve in some frequency range.



The terms “controller” and “compensator” are used interchangeably.

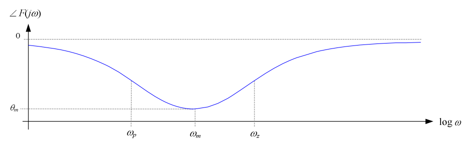
Phase Compensation Types

- Start with gain k_1 and introduce **phase lead** at high frequencies to achieve specified PM, GM, M_p, ...etc.



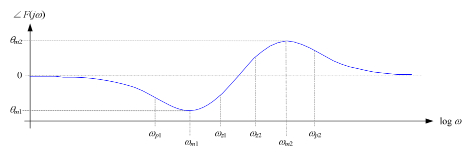
$$F(s) = \frac{s+\omega_z}{s+\omega_p}$$ ...where ($\omega_z < \omega_p$)

- Start with gain k_2 and introduce **phase lag** at low frequencies to meet steady-state requirements



$$F(s) = \frac{s+\omega_z}{s+\omega_p}$$ ...where ($\omega_z > \omega_p$)

- Start with gain between k_1 and k_2 and introduce phase lag at low frequencies and lead at high frequencies (**lag-lead** compensation)



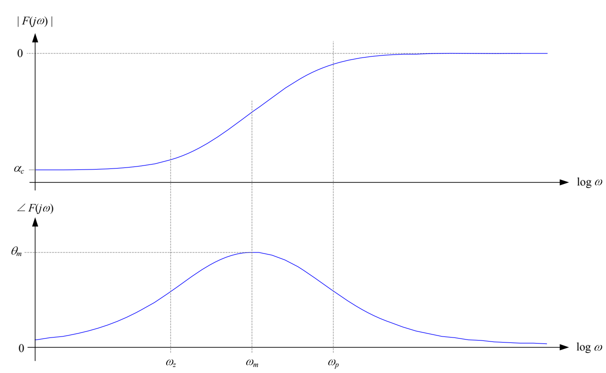
$$F(s) = \frac{(s+\omega_{z1})(s+\omega_{z2})}{(s+\omega_{p1})(s+\omega_{p2})}$$

For unity gain: $\omega_{p1}\,\omega_{p2} = \omega_{z1}\,\omega_{z2}$

Phase Lead Compensation

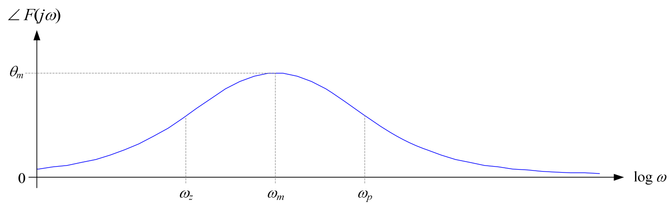
The first order phase lead compensator has one pole and one zero, with the zero frequency lower than that of the pole.

The simple lead compensator transfer function is: $F(s) = \frac{s+\omega_z}{s+\omega_p}$ ...where ($\omega_z < \omega_p$)



Lead Compensator Design

The passive phase lead compensator is given by $F(s) = \frac{1}{\alpha}\frac{1+\alpha cs}{1+cs}$ ...where $\alpha > 1$



Maximum phase lead of $\sin\theta_m = \frac{\alpha - 1}{\alpha + 1}$ occurs at frequency $\omega_m = \frac{1}{c\sqrt{\alpha}}$

❖ Fix α using $\alpha = \frac{1+\sin\theta_m}{1-\sin\theta_m}$, then calculate c using $c = \frac{1}{\omega_m\sqrt{\alpha}}$

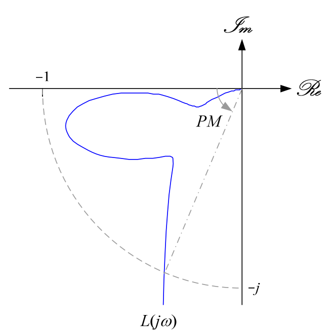
Note that cross-over frequency will typically fall so the process will need to iterate to find an acceptable design.

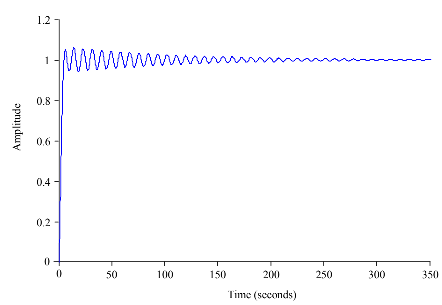

2.1, 2.2

A Problem with Stability Margins

Care should be taken when relying solely on gain and phase margins to determine stability and performance. These evaluate the proximity of $L(j\omega)$ to the critical point at (at most) two frequencies, whereas the closest point may occur at any frequency and be considerably less than that at either GM or PM, as the example below illustrates.

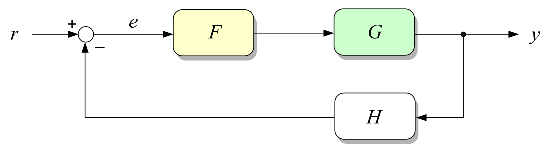
$$L(s) = \frac{0.38(s^2 + 0.1s + 0.55)}{s(s+1)(s^2 + 0.06s + 0.5)}$$





In this example, although gain and phase margins are adequate (GM infinite, $PM \approx$ 70 deg.), simultaneous change of both gain and phase over a narrow range of frequency leads to poor relative stability. The step response exhibits a fast rise time but with considerable oscillation.

Error Ratio



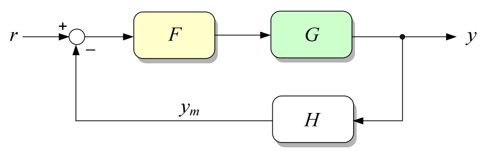
The **error ratio** plays a fundamental role in feedback control.

$$\frac{e}{r} = \frac{1}{1+FGH} = \frac{1}{1+L}$$

- The error ratio is also called the **sensitivity function** as it determines loop sensitivity to disturbance

$$\boxed{S = \frac{1}{1+L}}$$

Feedback Ratio



The **feedback ratio** or **complementary sensitivity function** is

$$\frac{y_m}{r} = \frac{FGH}{1+FGH} = \frac{L}{1+L}$$

- The feedback ratio determines the reference tracking accuracy of the loop

$$\boxed{T = \frac{L}{1+L}}$$

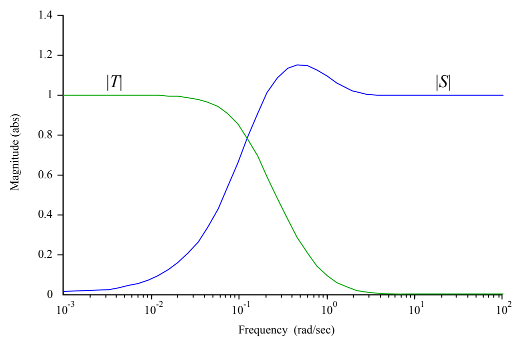
- The closed loop transfer function is related to T by: $\dfrac{y}{r} = \dfrac{T}{H}$

S + T = 1

Sensitivity function is: $S = \frac{1}{1+L}$ Complementary sensitivity function is: $T = \frac{L}{1+L}$

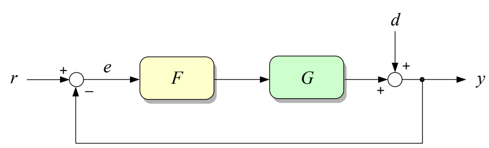
$$\boxed{S + T = \frac{1+L}{1+L} = 1}$$

- The shape of $L(j\omega)$ means we cannot maintain a desired S or T over the entire frequency range



Control with Output Disturbance

Consider the case of a unity feedback loop with disturbance acting at the output...



Superposition allows reference and disturbance effects to be included in y...

$$y = d + FG\,(r - y)$$

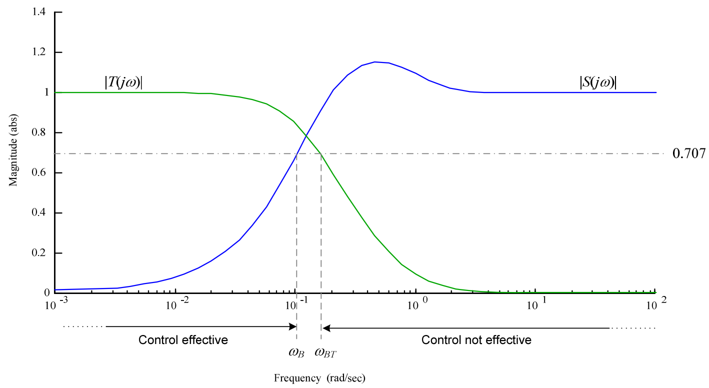
Substituting S and T gives

$$\boxed{y = S\,d + T\,r}$$

- S determines the ability of the loop to reject disturbance acting at the output
- T determines the ability of the loop to track a reference input

Bandwidth

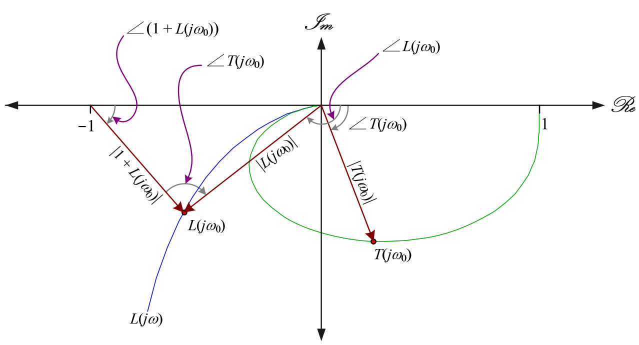
Bandwidth (ω_B & ω_{BT}) can be defined in terms of the frequencies at which $|S|$ & $|T|$ first cross $\frac{1}{\sqrt{2}}$



- Below ω_B performance is improved by control
- Between ω_B and ω_{BT} control affects response but does not improve performance
- Above ω_{BT} control has no significant effect

Closed Loop Properties from the Nyquist Plot

The vectors $|L(j\omega_0)|$ and $|1+L(j\omega_0)|$ can be obtained directly from the Nyquist plot for any frequency ω_0



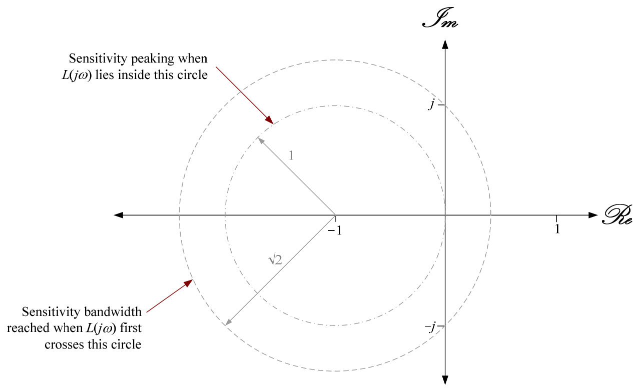
For the unity feedback system...

Closed loop magnitude is given by: $$|T(j\omega)| = \frac{|L(j\omega)|}{|1+L(j\omega)|}$$

Closed loop phase is given by: $$\angle T(j\omega) = \angle L(j\omega) - \angle(1+L(j\omega))$$

Nyquist Plot: Sensitivity Function

Peaking & bandwidth properties of the sensitivity function can be inferred from the Nyquist diagram.

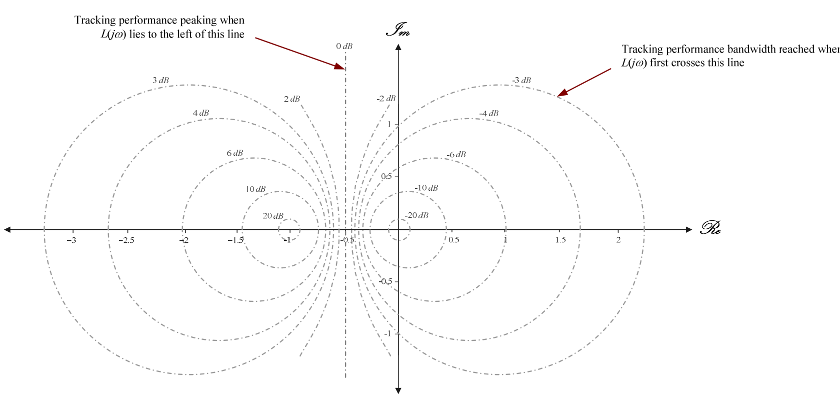


Sensitivity bandwidth reached when $|S| = \dfrac{1}{|1+L|}$ first crosses $\dfrac{1}{\sqrt{2}}$ from below: $|S| > \dfrac{1}{\sqrt{2}} \Rightarrow |1+L| < \sqrt{2}$

Sensitivity peaking occurs when $|1+L| < 1$

Nyquist Plot: Tracking Performance

Peaking & bandwidth of the tracking response can also be determined from the Nyquist diagram.



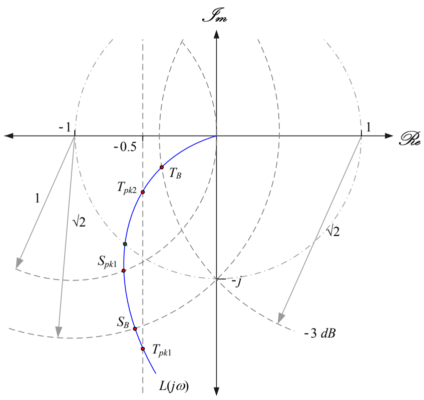
Tracking bandwidth reached when $|T| = \dfrac{|L|}{|1+L|}$ first crosses $\dfrac{1}{\sqrt{2}}$ from above: $|T| < \dfrac{1}{\sqrt{2}} \Rightarrow |1+L| > \sqrt{2}|L|$

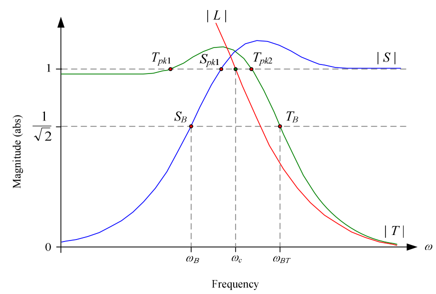
Tracking peaking occurs when $Re\{\, L(j\omega)\,\} < -0.5$

Note: lines of constant T are known as "M-circles"

Nyquist Plot: S & T Curves

Key features of the S & T curves such as peaking and bandwidth are available from the Nyquist plot.





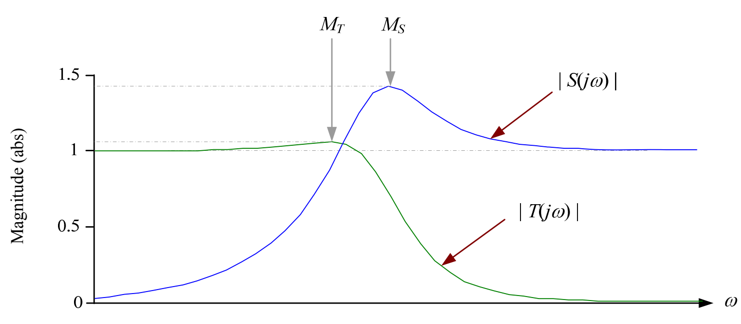
The trajectory of $L(j\omega)$ can be determined from the S & T curves, since $L(j\omega) = \dfrac{T(j\omega)}{S(j\omega)}$

For example, at cross-over: $|L(j\omega_c)| = 1 \iff |T(j\omega_c)| = |S(j\omega_c)|$

Maximum Peak Criteria

The maximum peaks of sensitivity and complementary sensitivity are:

$$M_S = \sup_{\omega} |S(j\omega)| = \|S\|_\infty \qquad M_T = \sup_{\omega} |T(j\omega)| = \|T\|_\infty$$

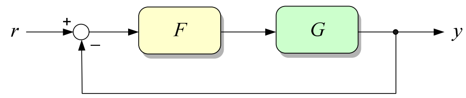


Phase margin and gain margin are loosely related to the $|S|$ & $|T|$ peaks

- Typical design requirements are: $M_S < 2$ $(6dB)$ and $M_T < 1.25$ $(2dB)$

2.1a

Plant Model Sensitivity



For the unity feedback system, tracking performance is given by $T = \dfrac{FG}{1+FG}$

If we differentiate T with respect to the plant G, we find ...

$$\frac{\partial T}{\partial G} = \frac{F(1+FG) - FGF}{(1+FG)^2} = \frac{ST}{G}$$

$$\boxed{S = \frac{\partial T}{\partial G}\frac{G}{T}}$$

❖ The sensitivity function S represents the relative sensitivity of the closed loop to relative plant model error

Sensitivity & Model Error

Let the model error in G be represented by the multiplicative output term ε.

$$\widetilde{G} = G\,(1+\varepsilon)$$

$$1+\widetilde{L} = 1+FG(1+\varepsilon)$$

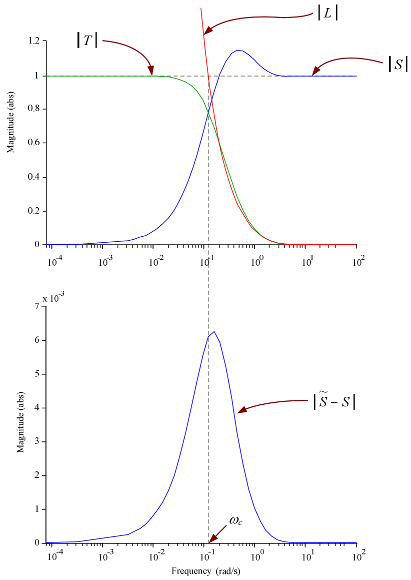
Therefore loop sensitivity including model error is:

$$\widetilde{S} = \frac{1}{1+\widetilde{L}} = \frac{1}{1+FG+FG\varepsilon}$$

$$\boxed{\widetilde{S} = \frac{S}{1+T\varepsilon}}$$

❖ The major effect of model error is in the cross-over region, where $S \approx T$

Effect of Plant Model Error



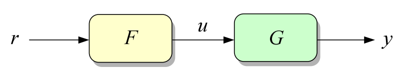
$$\tilde{S} = \frac{S}{1 + T\varepsilon}$$

The effect of plant model error is most severe around cross-over - exactly where the stability and performance properties of the loop are determined.

Evaluating and accounting for model uncertainty is therefore an important step in design.

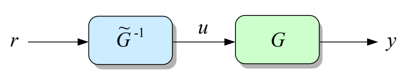
The process of modelling plant uncertainty and designing the control system to be tolerant of it is known as **robust control**.

Internal Model Principle



In open loop control: $y = FGr$

The basis of the Internal Model Principle is to determine the plant model $\tilde{G}$ and set $F = \tilde{G}^{-1}$



$$y = \tilde{G}^{-1}Gr = r$$

i.e. perfect control is achieved without feedback!

The practical value of this approach is limited because...

- Information about the plant may be inaccurate or incomplete
- The plant model may not be invertible or realisable
- Control is not robust, since any change in the process results in output error

Internal Model Control

An alternative to shaping the open loop is to directly synthesize the closed loop transfer function. The approach is to specify a desirable closed loop shape Q, then solve to find the corresponding controller.

$$Q = \frac{FG}{1+FGH}$$

$$\Rightarrow \; F = G^{-1}\frac{Q}{1-QH}$$

This method is known as **Internal Model Control** (IMC), or **Q-parameterisation**.

In principle, any closed-loop response can be achieved providing the plant model is accurate and invertible, however the plant might be difficult to invert because...

- RHP zeros give rise to RHP poles – $i.e.$ the controller will be *unstable*
- Time delay becomes time advance – $i.e.$ the controller will be *non-causal*
- If the plant is strictly proper, the inverse controller will be *improper*

 **2.3**

Non-Minimum Phase Plant Inversion

- Step 1: factorise G into invertible and non-invertible ($i.e.$ non-minimum phase) parts: $G = G_m G_n$

  ...where the non-invertible part is given by $G_n = e^{-\theta s}\prod_{i=1}^{q}\frac{s-z_i}{s+z_i}$

  This is an all-pass filter with delay. Any new LHP poles in G_n can be cancelled by LHP zeros in G_m

- Step 2: write the desired closed loop transfer function to include G_n: $Q = f\,G_n$

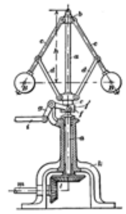
- Step 3: substitute into the controller equation: $F = G_m^{-1}\,G_n^{-1}\,\frac{f\,G_n}{1-f\,G_n\,H}$

  Non-minimum phase terms cancel to leave an equation which does not require inversion of G_n

$$\boxed{F = G_m^{-1}\,\frac{f}{1-f\,G_n\,H}}$$

Control Theory Seminar

3. Transient Response

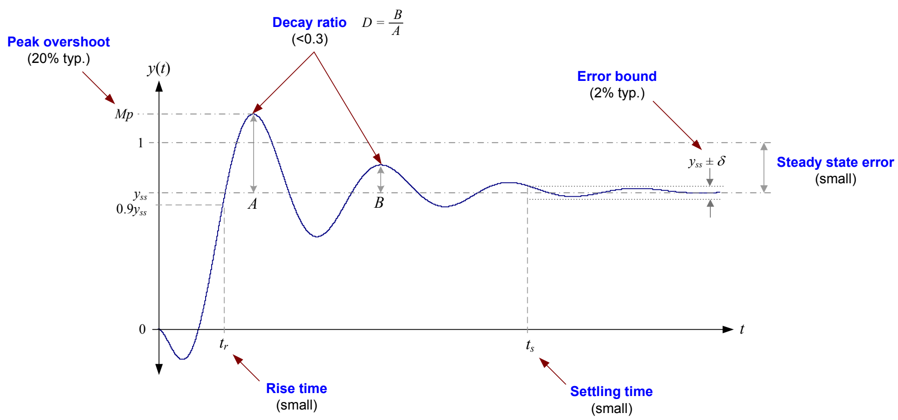


- Transient Specifications
- Steady State Error
- PID Control
- Root Locus Analysis

"It don't mean a thing if it ain't got that swing."

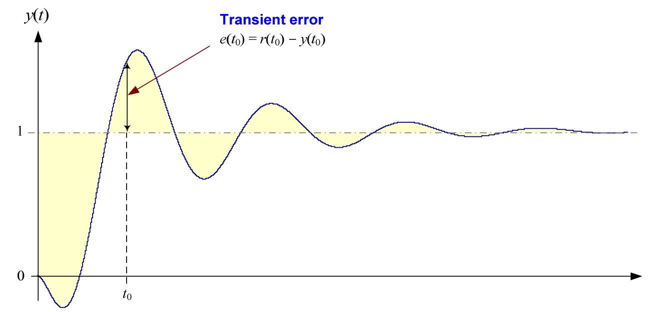
Duke Ellington (1899 – 1974)

Transient Response Specifications



- Transient response tuning is typically a compromise between competing objectives
- Results are highly subjective: different users may select very different controller settings
- Optimality only possible when some form of performance index is specified

Transient Performance Index



A **performance index** can be defined based on the integral of the closed loop error:

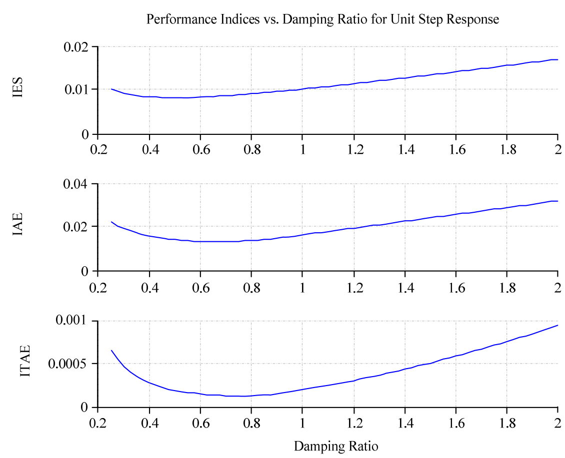
IES = Integral of the Error Squared $\int_0^\infty e(t)^2 dt$

IAE = Integral of the Absolute Error $\int_0^\infty |e(t)| dt$

ITAE = Integral of Time x Absolute Error $\int_0^\infty t\,|e(t)| dt$

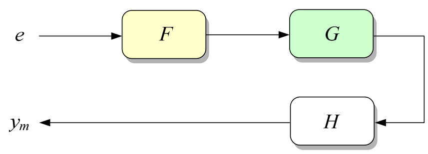
Quality of Response

A plot of performance index against variation of a key parameter typically yields a convex curve with a well defined minimum.



The parameter setting which yields minimum performance index represents an optimal controller choice.

Classification by Type



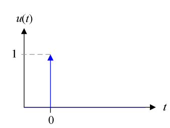
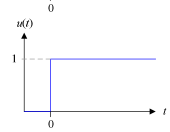
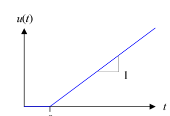
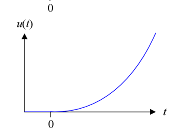
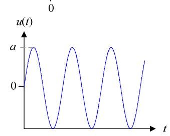
A canonical feedback system with open-loop transfer function

$$\frac{y_m(s)}{e(s)} = L(s) = \frac{k\,\beta(s)}{s^n\,\alpha(s)}$$

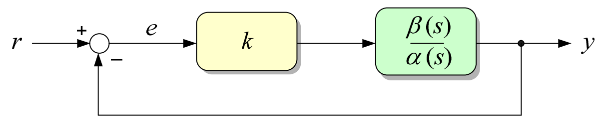
...where $n \geq 0$ is called a “type n” system.

- The **type number** denotes the number of integrators in the open-loop transfer function, $L(s)$
- Closed loop steady state error will be zero, finite or infinite, depending on the type number, n

Input Stimuli

Impulse		$u(t) = \delta(t)$	$u(s) = 1$
Unit step		$u(t) = 1(t)$	$u(s) = \frac{1}{s}$
Unit ramp		$u(t) = t$	$u(s) = \frac{1}{s^2}$
Parabola		$u(t) = t^2$	$u(s) = \frac{1}{s^3}$
Sine		$u(t) = a\sin(\omega t)$	$u(s) = a\frac{\omega}{s^2+\omega^2}$

Type 0 Systems



Error ratio is given by:

$$\frac{e(s)}{r(s)} = \frac{1}{1 + k\dfrac{\beta(s)}{\alpha(s)}} = \frac{\alpha(s)}{\alpha(s) + k\beta(s)}$$

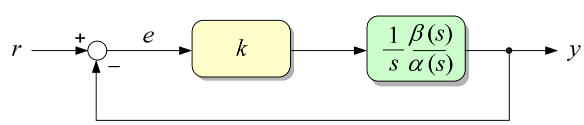
Steady state error following a step input is found by applying the final value theorem to $e(s)$

$$e_{ss} = \lim_{s \to 0} s \left\{ \frac{1}{s} \frac{\alpha(s)}{\alpha(s) + k\beta(s)} \right\}$$

$$e_{ss} = \frac{\alpha(0)}{\alpha(0) + k\beta(0)}$$

- For a type 0 system there is always a steady state error following a step input which is inversely related to loop gain, k

Type 1 Systems



Error ratio is given by:

$$\frac{e(s)}{r(s)} = \frac{1}{1 + k\dfrac{1}{s}\dfrac{\beta(s)}{\alpha(s)}} = \frac{s\alpha(s)}{s\alpha(s) + k\beta(s)}$$

Again, steady state error following a step input is found from the final value theorem:

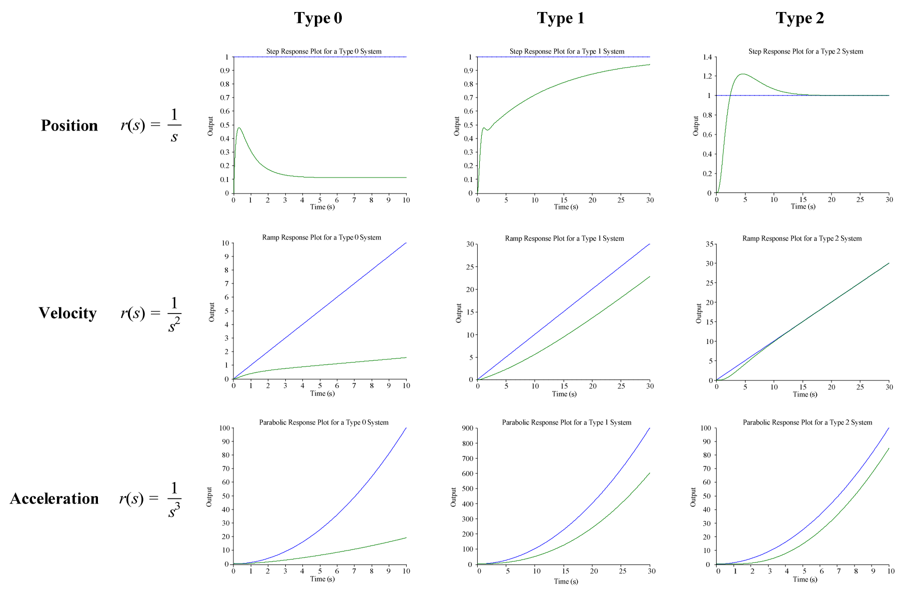
$$e_{ss} = \lim_{s \to 0} s \left\{ \frac{1}{s} \frac{s\alpha(s)}{s\alpha(s) + k\beta(s)} \right\}$$

$$e_{ss} = 0$$

- The presence of an integrator in the loop eliminates steady state error following a step input

- To avoid steady state error $L(s)$ must contain at least as many integrators as $r(s)$

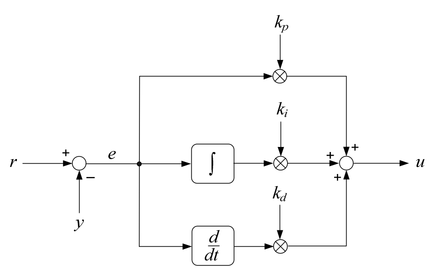
Response Type Summary



PID Controllers

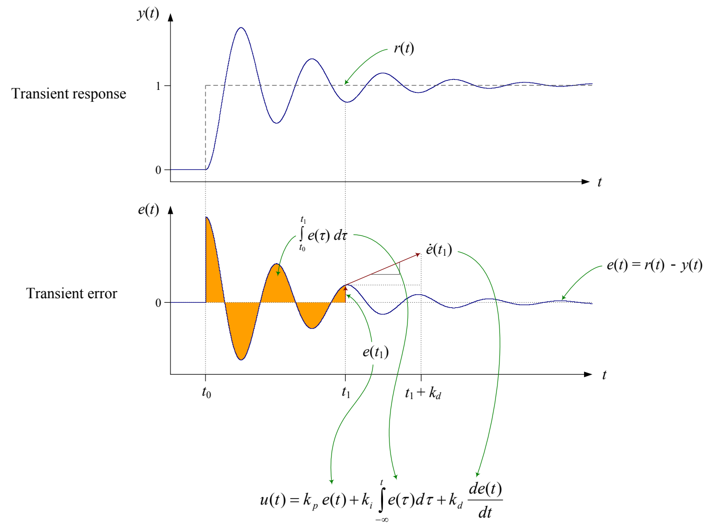
PID (Proportional + Integral + Derivative) controllers allow intuitive tuning of the transient response.

The **parallel** PID form is: $u(t) = k_p\, e(t) + k_i \int_{-\infty}^{t} e(\tau)\, d\tau + k_d \frac{de(t)}{dt}$



- The proportional term k_p directly affects loop gain
- Integral action increases low frequency gain and reduces/eliminates steady state errors, however this can have a de-stabilizing effect due to increased phase lag
- Derivative action introduces a predictive type of control which tends to damp oscillation & overshoot but can lead to large control effort

PID Control Action

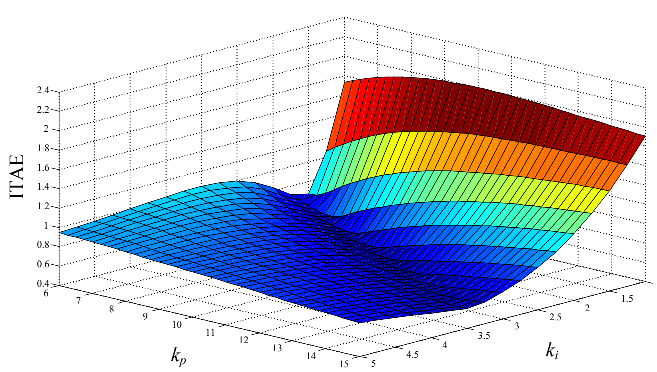


Many guidelines exist (Ziegler-Nichols, Cohen-Coon, etc.) but PID tuning is typically an iterative process.

 **3.1**

Optimal PID Tuning

Optimal controller settings can be sought based on a transient response cost function, such as ITAE.

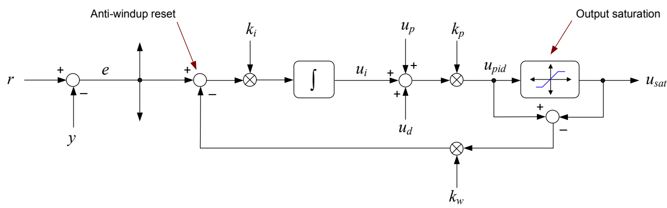


A simple minimum search algorithm reveals the controller terms which yield the smallest cost function.

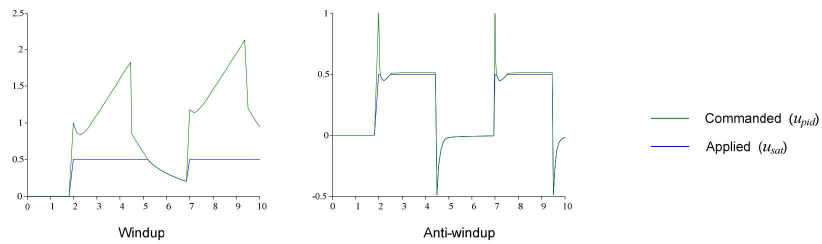
Pairs of tuning parameters, such as proportional and integral gain terms, can be found in this way.
For larger numbers of tuning parameters, iteration using multiple plots is required.

Integrator Windup

If a component in the loop saturates control will be lost. The integrator continues to accumulate error, increasing corrective effort even though the plant output does not change. This effect is called **windup**.

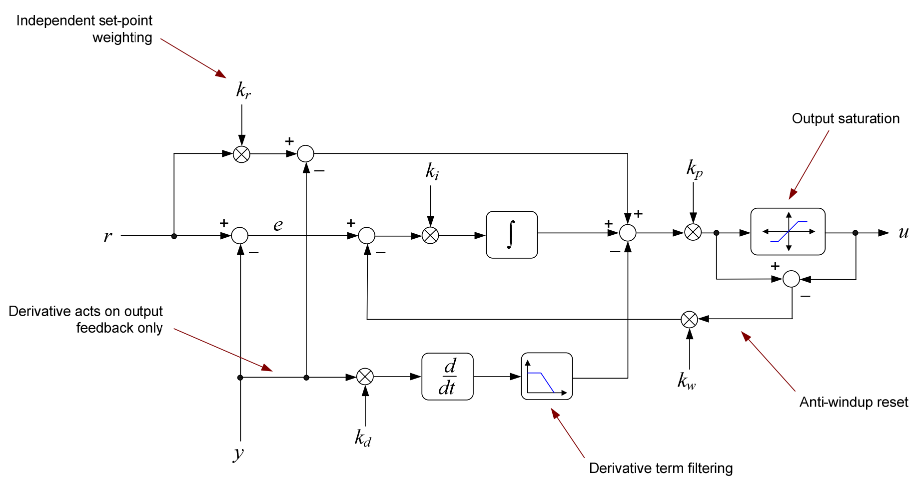


Modern industrial PID controllers incorporate an anti-windup feature which clamps the integrator input when saturation occurs.



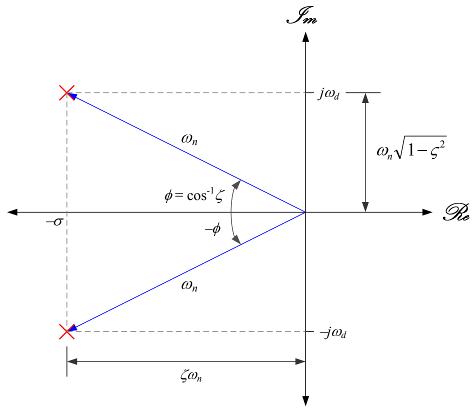
PID Controller Refinements

Practical PID controllers incorporate various refinements to improve performance and avoid specific difficulties. Some of these are shown below.



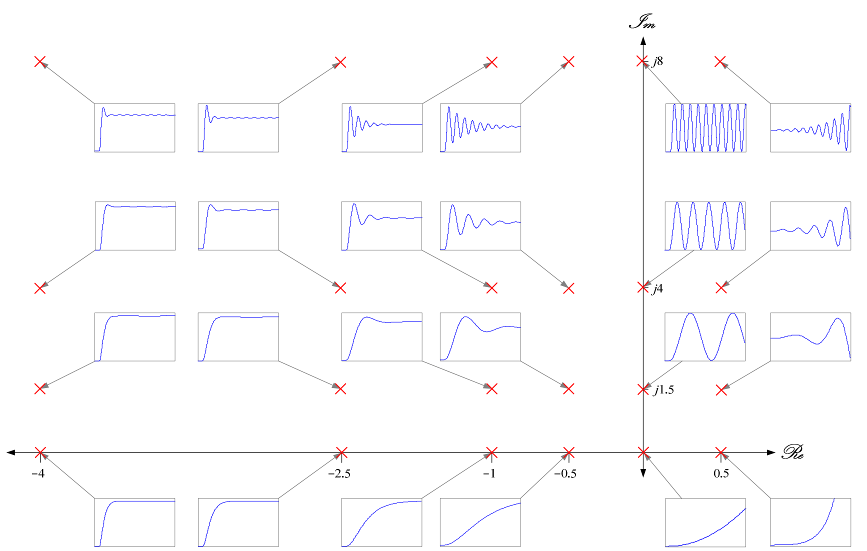
Complex Pole Interpretation

Recall, for the under-damped second order case poles are located at $s = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$



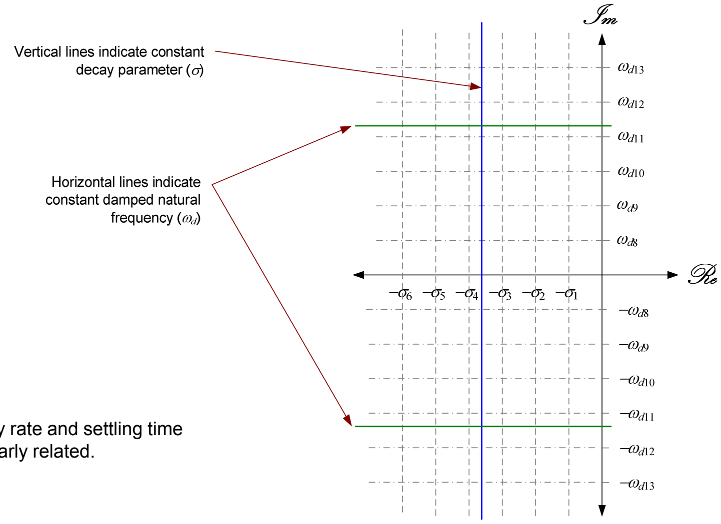
- The damping coefficient and damped natural frequency are the real and imaginary parts of the poles
- Un-damped natural frequency and transient phase represent the modulus and argument of the poles

Influence of Pole Location on Transient Response



Plot shows unit step response of second order system with varying pole location. Stable poles positioned further to the left exhibit faster decay, while those with larger imaginary part have a higher frequency of oscillation.

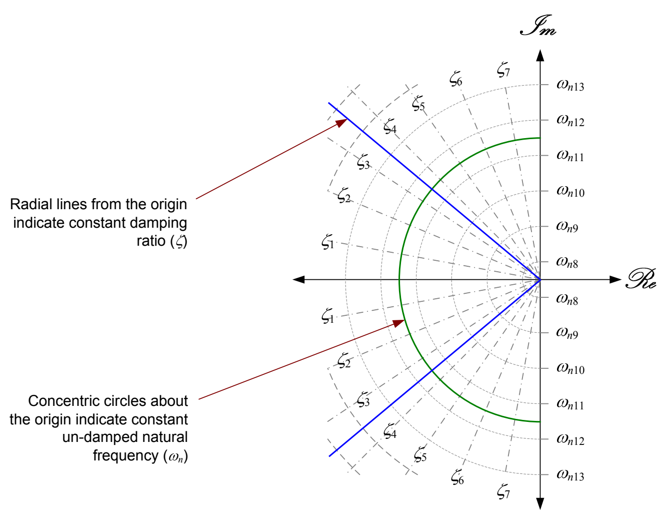
Constant Parameter Loci



Note: decay rate and settling time are not linearly related.

- Poles located further to the left have faster decay rate
- Poles with larger imaginary component are more oscillatory

Constant Parameter Loci



This is the usual grid drawn on a pole-zero map to aid in transient response estimation.

Root Locus Design

We have seen how key properties of the transient response can be inferred from the location of under-damped second order poles in the complex plane.

- In a root locus plot, the closed loop pole paths are plotted in the complex plane as some free parameter (often loop gain, k) is varied

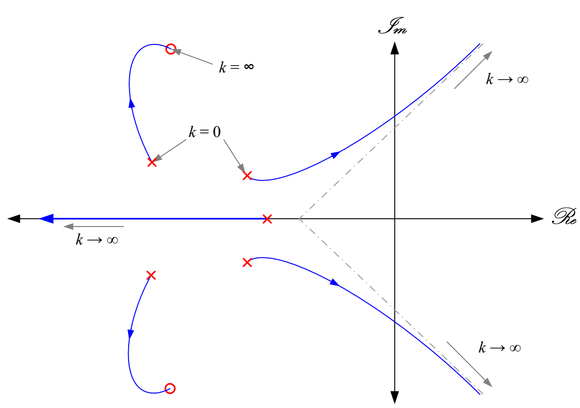
The **root locus method** is a graphical procedure for controller design based on the locations of the dominant closed loop poles.

Recall, closed loop poles are the roots of $\alpha_1\alpha_2 + k\,\beta_1\beta_2 = 0$

- When $k = 0$ the roots are $\alpha_1\alpha_2 = 0$ $\quad$ *i.e.* at open loop poles
- As $k \to \infty$ the roots tend towards $\beta_1\beta_2 = 0$ $\quad$ *i.e.* at open loop zeros
- For $0 < k < \infty$ the roots follow well defined paths called "loci"

Root Locus Plots

- Every root locus begins at an open loop pole when $k = 0$, and either ends at an open loop zero or follows a high gain asymptote to infinity



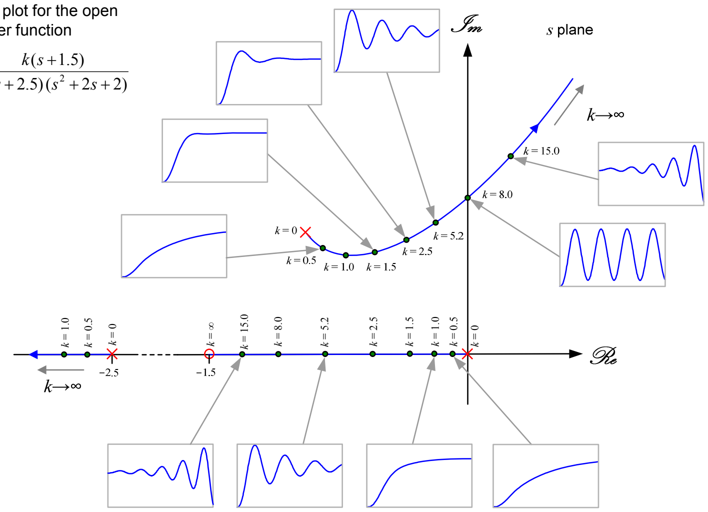
Example root locus plot for system with two closed loop zeros and five poles (*i.e.* relative degree three)

At each value of k, features of the closed loop transient response can be inferred from location of the dominant poles.

Root Locus Example

Root locus plot for the open loop transfer function

$$L(s) = \frac{k(s+1.5)}{s(s+2.5)(s^2+2s+2)}$$

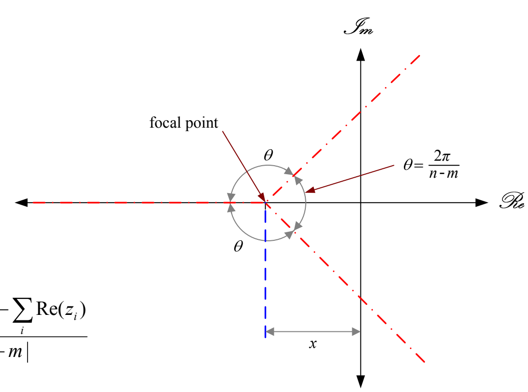


Association of step response with closed loop root location for varying controller gain.

High Gain Asymptotes

The number of high gain asymptotes is equal to the relative degree of $L(s)$, $n - m$.

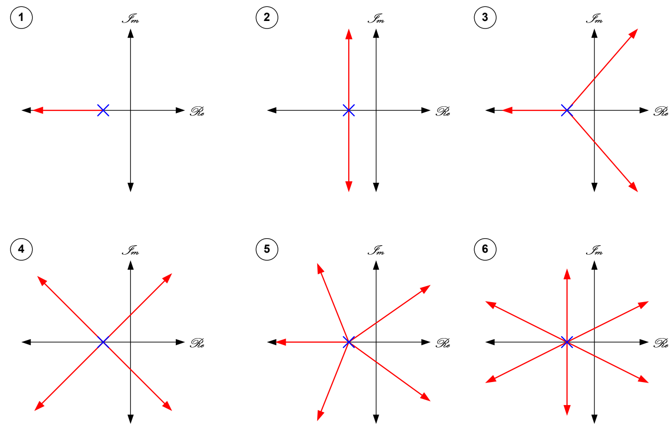
Asymptotes are distributed symmetrically around a focal point on the real axis. The angle of separation of the asymptotes and their point of intersection on the real axis depend on the relative degree of the closed loop transfer function.



$$x = \frac{\sum_i \mathrm{Re}(p_i) - \sum_i \mathrm{Re}(z_i)}{|n-m|}$$

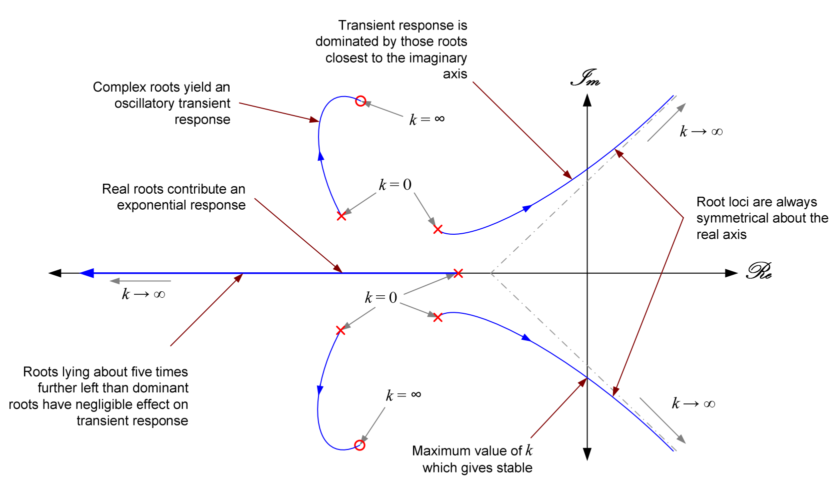
High Gain Asymptotes

High gain root locus asymptotes shown by closed loop relative degree



Note that for relative degree of 3 or greater loci move into the RHP, causing instability at high gain

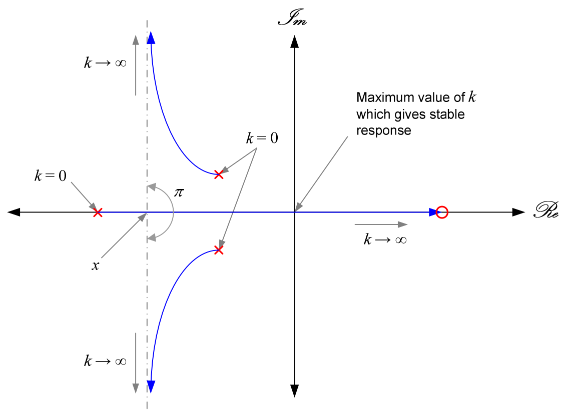
Properties of the Root Loci



❖ The number of root loci in the s plane is the same as the order of $L(s)$

Limitation of RHP Zero

As open loop gain increases, each root locus tend towards either an infinite asymptote or an open loop zero. *i.e.* for proper systems, each zero accommodates a closed loop pole at infinite gain.



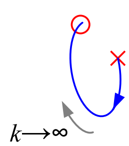
❖ For each RHP open loop zero one locus crosses into the RHP, so at sufficiently high gain the closed loop will become unstable

Pole-Zero Cancellation

When a pole and zero lie on top of one another their combined effect on closed loop response is zero.

$$G(s) = \frac{s+q}{s+q} = 1\angle 0$$

When a poles and a zero lie close to one another they generate a short locus which has little overall effect on the closed loop response.

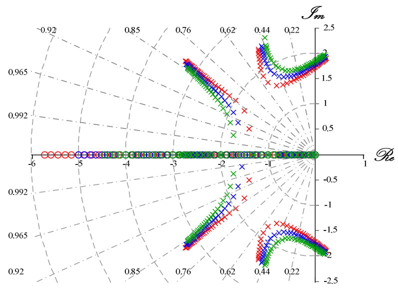


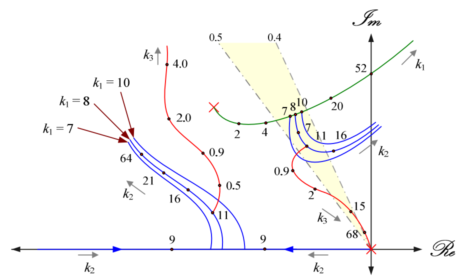
Pole-zero cancellation means placing controller poles and zeros to cancel out undesirable poles and zeros in the plant. Additional controller poles & zeros can then be placed in more desirable locations in the complex plane.

3.2, 3.3

Tuning Multiple Parameters

Interpretation of the root loci may be difficult if more than one parameter is varied. Simulation packages contain no native tools to display root loci for multiple free parameters, or **root contours**.



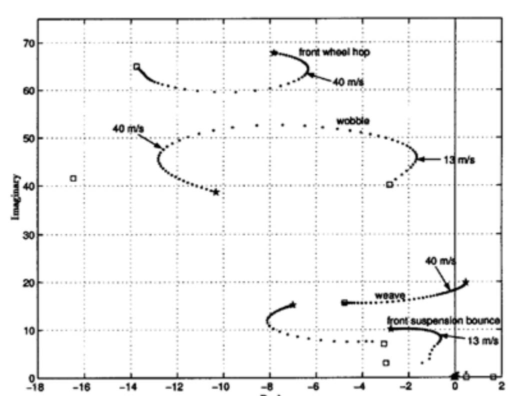


Standard PID controller parameter tuning for the plant $G(s) = \dfrac{1}{(s^2+4s+8)(s+1)}$

❖ The presence of closed loop zeros means tuning choices should be supported by other data.

Root Locus Example

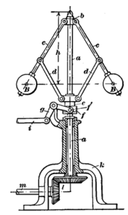
"The Stability of Motorcycles Under Acceleration,"
by D J N Limebeer, R S Sharp and S Evangelou,
Journal of Applied Mechanics, Vol. 69, 2002
Original publisher: ASME



Root-locus for a fixed roll angle of 30°. The speed is increased from 6 m/s (□) to 60 m/s (*).

Control Theory Seminar

4. Discrete Time Systems

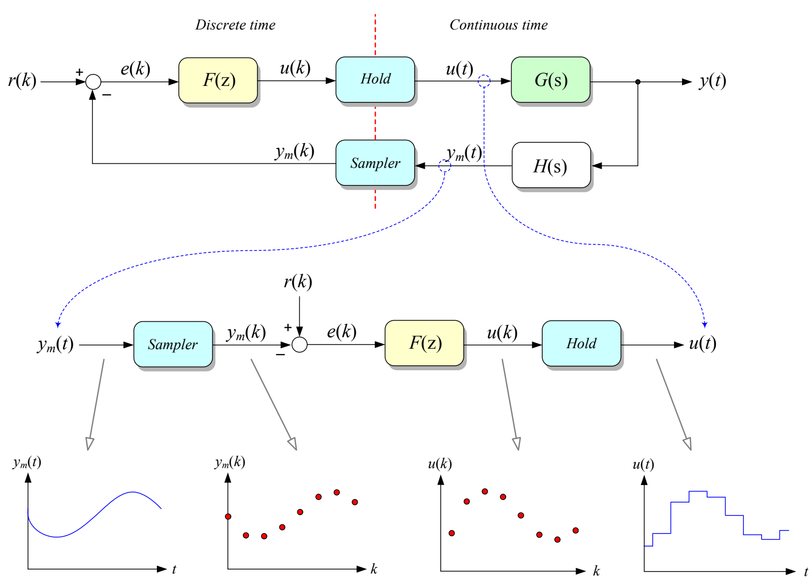


- Sampled Systems
- The z Transform
- The Frequency Domain
- Delay & Reconstruction
- Discrete Time Controller Design

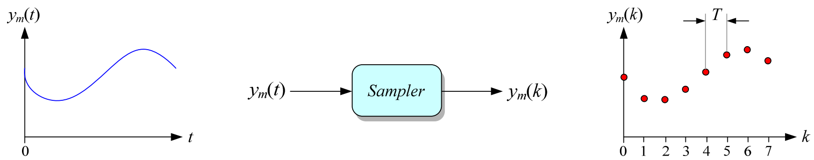
"...in recent times, almost all analogue controllers have been replaced by some form of computer control. This is a very natural move since control can be conceived as the process of making computations based on past observations of a systems behaviour. The most natural way to make these computations is via some form of computer."

Goodwin, Graebe & Salgado, *Control System Design*, 2000

The Digital Control System



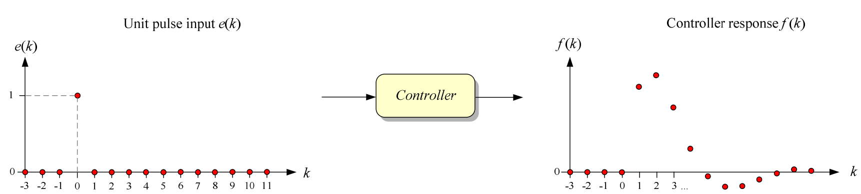
The Sampler



- The sampler converts a continuous function of time $y_m(t)$ into a discrete time function $y_m(kT)$
- Almost all samplers operate at a fixed rate $f_s = \frac{1}{T}$
- The T is implicit in notation, so for example $y_m(k)$ is equivalent to $y_m(kT)$
- The dynamic properties of the signal are changed as it passes through the sampler

The Unit Pulse Response

The **unit pulse response** $f(nT)$ is the response of the controller output following an input which has unit value at time $kT = 0$, and zero at all other times.

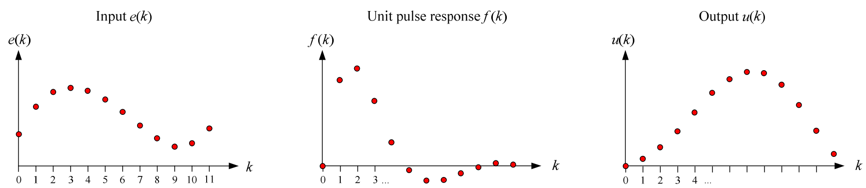


Providing the controller is stable, its unit pulse response will converge as k becomes large.

The transient properties of the controller are captured in the sequence $f(k)$.

Digital Controller Operation

The input sequence $e(k)$ is convolved with the unit pulse response $f(k)$ to form the controller output $u(k)$



The sequence of events which take place inside the digital controller is tabulated below.

$k = 0$: $\quad u(0) = f(0)e(0)$

$k = 1$: $\quad u(1) = f(1)e(0) + f(0)e(1)$

$k = 2$: $\quad u(2) = f(2)e(0) + f(1)e(1) + f(0)e(2)$

$k = 3$: $\quad u(3) = f(3)e(0) + f(2)e(1) + f(1)e(2) + f(0)e(3)$

$k = n$: $\quad u(n) = f(n)e(0) + f(n-1)e(1) + \ldots\ldots\ldots\ldots + f(0)e(n)$

Discrete Convolution

The digital controller computes this n-term **sum-of-products** for each input sample.

$k = n$: $\quad u(n) = f(n)e(0) + f(n-1)e(1) + \ldots\ldots\ldots\ldots + f(0)e(n)$

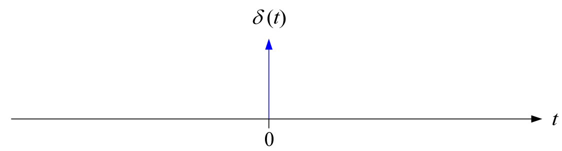
In practice, the number of terms (n) in the sequence is limited by the available computation time and memory.

Once the unit pulse response $f(nT)$ is known, the controller output $u(nT)$ arising from any arbitrary input $e(nT)$ can be found using a convolution summation.

$$u(nT) = \sum_{k=0}^{n} e(kT)\, f([n-k]T)$$

The design task is to find the $f(nT)$ coefficients which deliver a desired output $u(nT)$ for some $e(nT)$.

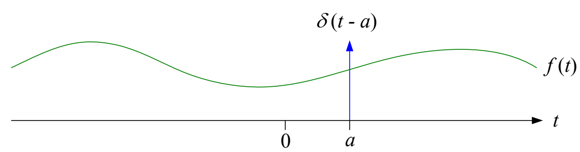
The Delta Function



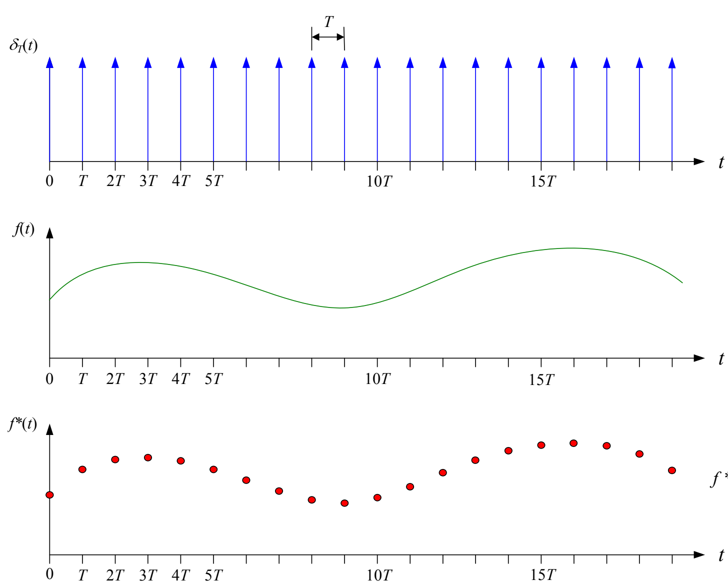
The delta function, denoted $\delta(t)$, represents an impulse of infinite amplitude, zero width, and unit area.

If a delta impulse is combined with a continuous signal the result is given by the **screening property**.

$$\int_{-\infty}^{\infty} \delta(t-a) f(t) dt = f(a)$$



Impulse Modulation



$$\delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

$$f^*(t) = f(t) \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

The z Transform

Applying the screening property of the delta function at each sample instant, we find

$$f^*(t) = \sum_{n=-\infty}^{\infty} f(nT)\,\delta(t-nT)$$

The shifting theorem allows us to take the Laplace transform of this series term-by-term...

$$f^*(s) = \mathscr{L}\{\ldots + f(-2T)\,\delta(t+2T) + f(-T)\,\delta(t+T) + f(0)\,\delta(t) + f(T)\,\delta(t-T) + f(2T)\,\delta(t-2T) + \ldots\}$$

$$f^*(s) = \sum_{n=-\infty}^{\infty} f(nT)e^{-snT}$$

The z transform of $f(t)$ is found from the above series after making the substitution $z = e^{sT}$

$$f(z) = \sum_{n=-\infty}^{\infty} f(nT)z^{-n}$$

Properties of the z Transform

$$\boxed{f(z) = \mathcal{Z}\{\, f(nT)\,\} \triangleq \sum_{n=-\infty}^{\infty} f(nT)z^{-n}}$$

Convolution: $$\mathcal{Z}\left\{ \sum_{k=0}^{n} f_1(kT)\, f_2([n-k]T) \right\} = f_1(z)\, f_2(z)$$

Linearity: $$\mathcal{Z}[a_1 f_1(nT) \pm a_2 f_2(nT)] = a_1 f_1(z) \pm a_2 f_2(z)$$

Final value theorem: $$\lim_{n\to\infty} f(nT) = \lim_{z\to 1}(z-1)f(z)$$

Time shift: $$\mathcal{Z}\{f(n+k)\} = z^k f(z)$$

Note: Compare the above properties with those of the Laplace transform.

Transfer Functions

A linear continuous time system may be represented in transfer function form as

$$G(s) = A_s \frac{(s - z_{c1})(s - z_{c2}) \ldots (s - z_{cm})}{(s - p_{c1})(s - p_{c2}) \ldots (s - p_{cn})}$$

An equivalent sampled data system can be found using a **discrete transformation**, which yields a transfer function in the complex variable z.

$$G(z) = A_z \frac{(z - z_{d1})(z - z_{d2}) \ldots (z - z_{dm})}{(z - p_{d1})(z - p_{d2}) \ldots (z - p_{dn})}$$

Comparing the continuous time and discrete time representations of the same system:

- Poles & zeros are in different positions in the complex plane
- The relative degree may not be the same
- Dynamic performance is different

The Difference Equation

The 2-pole 2-zero transfer function is written $\dfrac{u(z)}{e(z)} = \dfrac{\beta_0 z^2 + \beta_1 z + \beta_2}{\alpha_0 z^2 + \alpha_1 z + \alpha_2}$

Normalizing for the term involving the highest denominator power (α_0) gives

$$\frac{u(z)}{e(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

Re-arranging to find an expression for $u(z)$...

$$u(z)\ \{\ 1 + a_1 z^{-1} + a_2 z^{-2}\ \} = e(z)\ \{\ b_0 + b_1 z^{-1} + b_2 z^{-2}\ \}$$

$$u(z) = e(z)\ \{\ b_0 + b_1 z^{-1} + b_2 z^{-2}\ \} - u(z)\ \{\ a_1 z^{-1} + a_2 z^{-2}\ \}$$

$$u(z) = b_0 e(z) + b_1 z^{-1} e(z) + b_2 z^{-2} e(z) - a_1 z^{-1} u(z) - a_2 z^{-2} u(z)$$

Applying the shifting property of the z-transform term-by-term yields the difference equation

$$\boxed{u(k) = b_0 e(k) + b_1 e(k-1) + b_2 e(k-2) - a_1 u(k-1) - a_2 u(k-2)}$$

Common Controller Types

The digital control law is a difference equation involving input & output data weighted by a set of coefficients derived from those of the transfer function.

Controller	Transfer Function	Control Law
PID	$k_p + \frac{k_i}{s} + k_d s$	$u(k) = u(k-1) + b_0 e(k) + b_1 e(k-1) + b_2 e(k-2)$
Type II	$k_p \frac{(s+z_1)}{s(s+p_1)}$	$u(k) = a_1 u(k-1) + a_2 u(k-2) + b_0 e(k) + b_1 e(k-1) + b_2 k(k-2)$
2-Pole 2-Zero	$k_p \frac{(s+z_1)(s+z_2)}{s(s+p_1)}$	$u(k) = a_1 u(k-1) + a_2 u(k-2) + b_0 e(k) + b_1 e(k-1) + b_2 k(k-2)$
Type III	$k_p \frac{(s+z_1)(s+z_2)}{s(s+p_1)(s+p_2)}$	$u(k) = a_1 u(k-1) + a_2 u(k-2) + a_3 u(k-3)$ $+ b_0 e(k) + b_1 e(k-1) + b_2 e(k-2) + b_3 e(k-3)$
3-Pole 3-Zero	$k_p \frac{(s+z_1)(s+z_2)(s+z_3)}{s(s+p_1)(s+p_2)}$	$u(k) = a_1 u(k-1) + a_2 u(k-2) + a_3 u(k-3)$ $+ b_0 e(k) + b_1 e(k-1) + b_2 e(k-2) + b_3 e(k-3)$

Notice the similarities between the first three and the last two control laws: the structure of the difference equation is the same - only the coefficients are different.

Discrete Time Stability

Consider the first order transfer function $\frac{u(z)}{e(z)} = \frac{b}{z-a}$

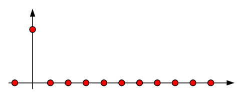
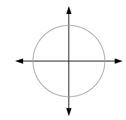
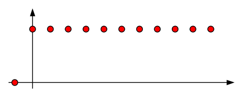
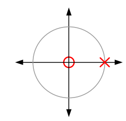
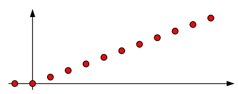
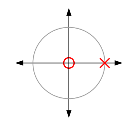
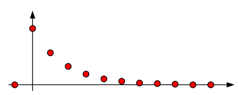
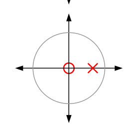
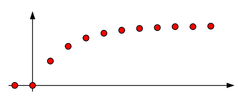
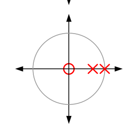
The corresponding difference equation is: $u(k) = be(k-1) + au(k-1)$

The evolution of the time sequence is:

k	$u(k)$
1	$be(0)$
2	$be(1) + abe(0)$
3	$be(2) + abe(1) + a^2be(0)$
4	$be(3) + abe(2) + a^2be(1) + a^3be(0)$
$\vdots$	$\vdots$
n	$b\sum_{\kappa=1}^{n-1} a^\kappa e(n-\kappa)$

- The presence of the a^κ term means that the output $u(k)$ will remain bounded (stable) as $k \to \infty$ providing $|a| \leq 1$. This is the stability constraint for discrete time systems.

Common z Transforms

Data	$f(nT)$	$F(z)$	z-plane
	$\delta[T]$	1	
	1	$\dfrac{z}{z-1}$	
	nT	$\dfrac{z}{(z-1)^2}$	
	a^n	$\dfrac{z}{z-a}$	
	$1-a^n$	$\dfrac{z(1-a)}{(z-a)(z-1)}$	

Complex Poles

As for continuous time systems, discrete time complex poles always arise in conjugate pairs.

$$G(z) = \frac{z^2}{(z-ae^{j\omega})(z-ae^{-j\omega})}$$

The transient part of the response is given by

$$y(z) = \frac{\varepsilon_1}{(z-ae^{j\omega})} + \frac{\varepsilon_1^*}{(z-ae^{-j\omega})} + \ldots$$

$$y(k) = \varepsilon_1\left(ae^{j\omega}\right)^k + \varepsilon_1^*\left(ae^{-j\omega}\right)^k + \ldots$$

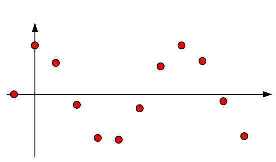
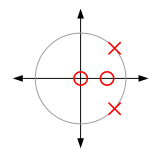
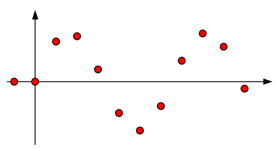
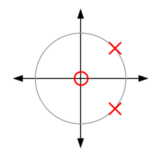
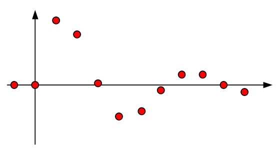
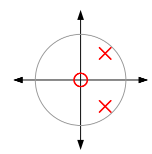
...where the residual ε_1 has the form $Ae^{j\theta}$

The time sequence is always oscillatory and of the form

$$y(k) = Ba^k \sin(k\omega + \theta) + \ldots$$

- In order that $y(k)$ remain bounded, every pole in $G(z)$ must be constrained by $|a| \le 1$

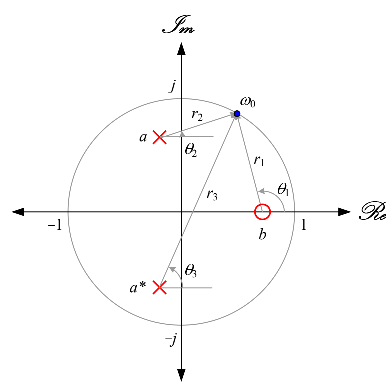
Common z Transforms

Data	$f(nT)$	$F(z)$	z-plane
	$\cos anT$	$\dfrac{z(z-\cos aT)}{z^2-2z\cos aT+1}$	
	$\sin anT$	$\dfrac{z\sin aT}{z^2-2z\cos aT+1}$	
	$a^n \sin bnT$	$\dfrac{az\sin bT}{z^2-2az\cos bT+a^2}$	

Frequency Response

❖ The response of the discrete time system $G(z)$ at frequency $\omega=\omega_0$ is evaluated by $G(z)\big|_{z=e^{j\omega_0 T}}$

Consider the system $G(z)=\dfrac{z-b}{(z-a)(z-a^*)}$



Magnitude is found from...

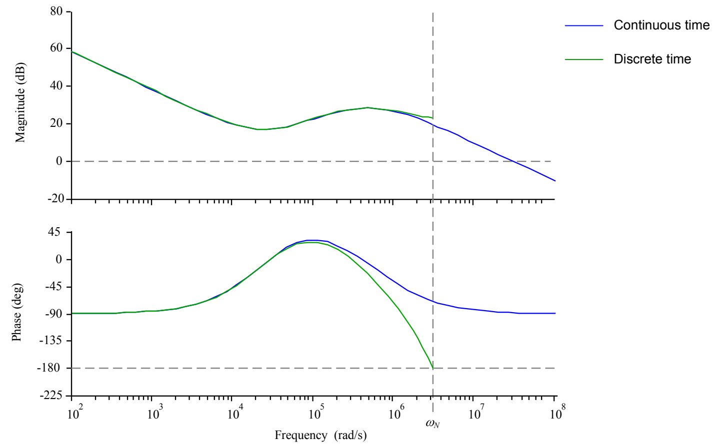
$$\left|G(e^{j\omega_0 T})\right|=\frac{\left|e^{j\omega_0 T}-b\right|}{\left|e^{j\omega_0 T}-a\right|\left|e^{j\omega_0 T}-a^*\right|}=\frac{r_1}{r_2 r_3}$$

Phase is found from...

$$\angle G(e^{j\omega_0 T})=\angle\left(e^{j\omega_0 T}-b\right)-\angle\left(e^{j\omega_0 T}-a\right)-\angle\left(e^{j\omega_0 T}-a^*\right)=\theta_1-\theta_2-\theta_3$$

The Bode Plot

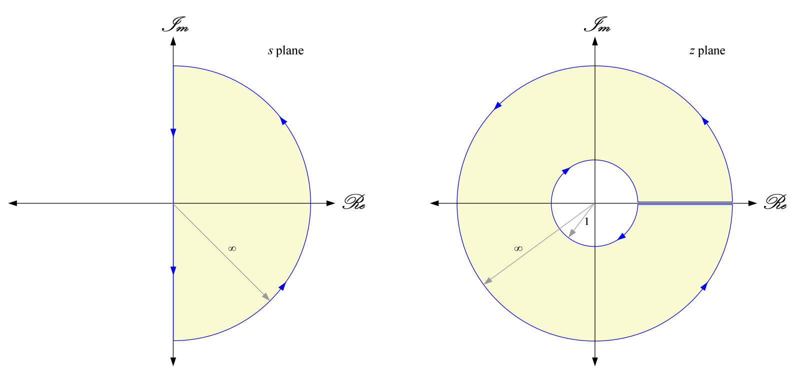
The frequency response of a discrete time system may be represented in Bode plot form, however the maximum unique frequency is limited by the sampling theorem. Typically only those frequencies below the Nyquist limit (ω_N) are shown.



Notice that the response of the discrete time system typically exhibits greater phase lag. This is due to sample-to-output delay and the effects of reconstruction.

The Nyquist Path

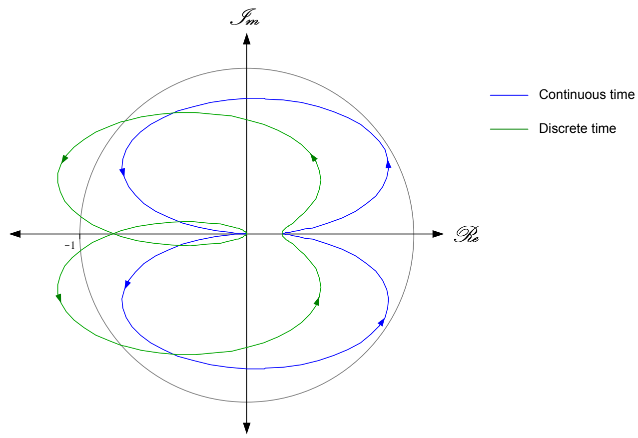
Nyquist analysis can be used with discrete time systems in a similar way to continuous systems. The region of unstable roots of $L(z)$ is shown shaded in the diagram below.



Recall, if the open loop is stable we look for enclosure of the critical point by the above contour after mapping by $L(z)$. If the open loop is unstable, we determine closed loop stability by counting encirclements of the critical point relative to the number of unstable poles of $1 + L(z)$.

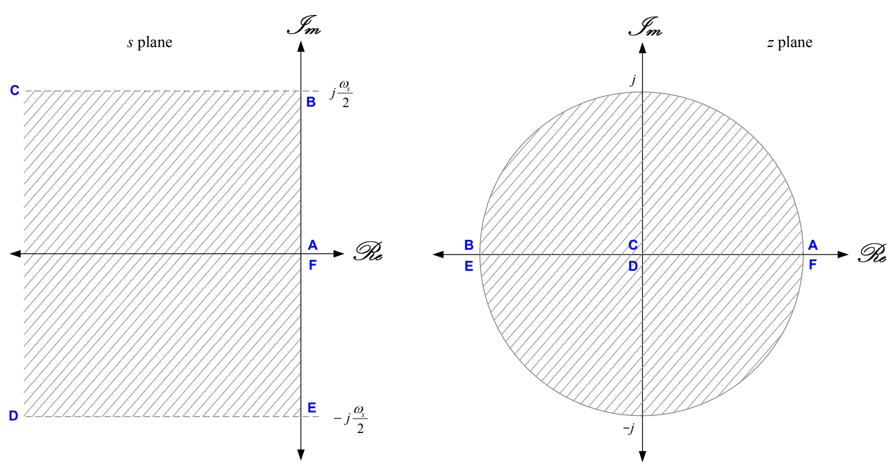
The Nyquist Plot

The frequency response of discrete time systems may be represented using the Nyquist plot in the same way as continuous time systems.



Plot shows the Nyquist curve for the system $\frac{1}{s^2 + 0.3s + 5}$ together with its discrete time equivalent after transformation by the matched pole-zero method for a sample rate of 2Hz.

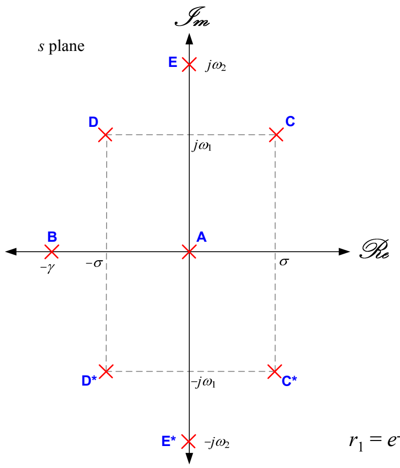
Complex Plane Mapping

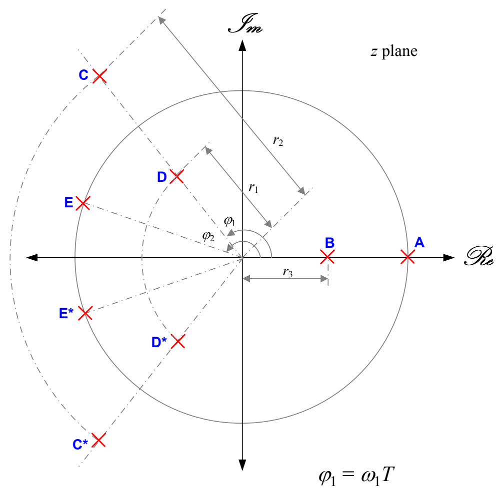


Equivalent regions shown cross-hatched

Complex Plane Mapping

❖ Points in the s-plane are mapped according to: $z = e^{sT} = e^{(a+jb)T} = e^{aT}e^{jbT} = re^{j\varphi}$

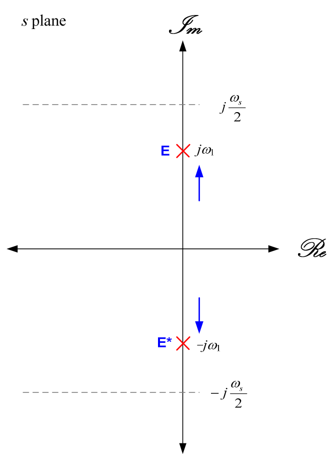


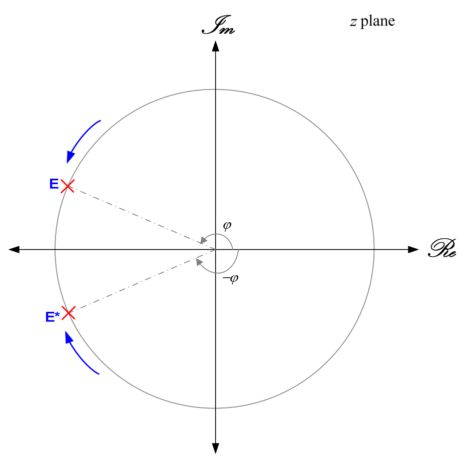


$$r_1 = e^{-\sigma T}$$
$$r_2 = e^{\sigma T}$$
$$r_3 = e^{-\gamma T}$$

$$\varphi_1 = \omega_1 T$$
$$\varphi_2 = \omega_2 T$$

The Nyquist Frequency

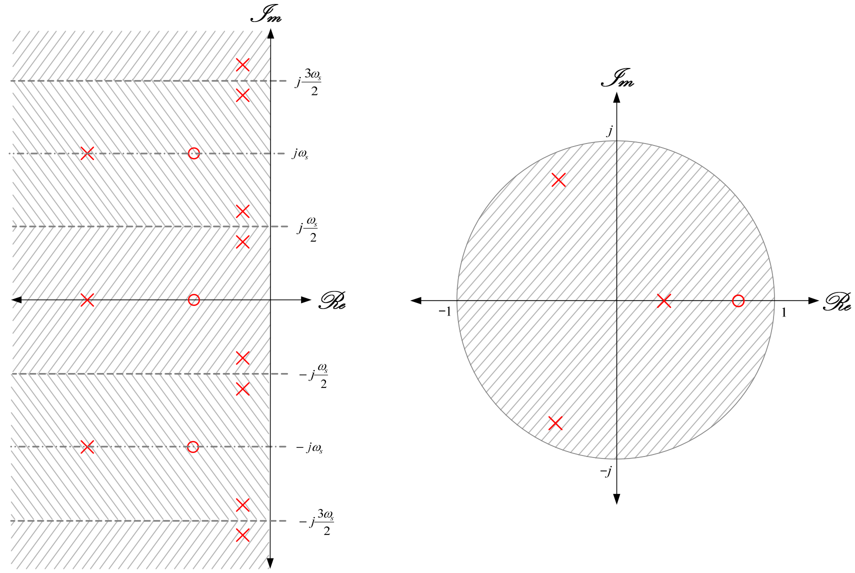




❖ The **Nyquist frequency** represents the highest unique frequency in the discrete time system

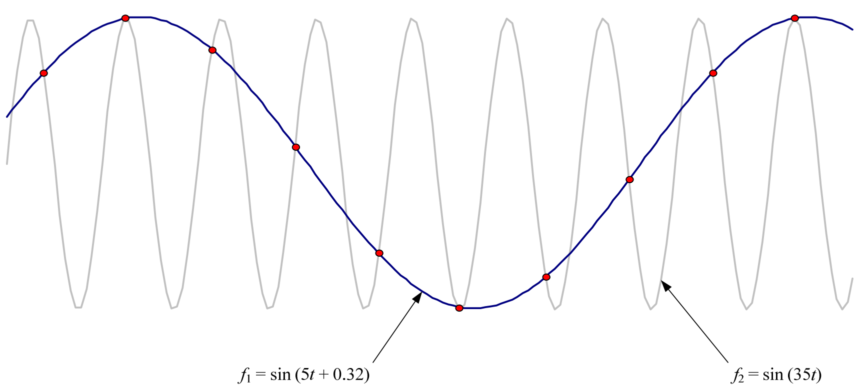
❖ Uniqueness is lost for higher continuous time frequencies after sampling

Aliasing



Loss of uniqueness means an infinite number of congruent strips are mapped into the unit circle.

Frequency Ambiguity



Both f_1 & f_2 give rise to exactly the same set of samples. After sampling it is impossible to determine which frequency was sampled. In fact, any of an infinite number of possible sine waves could have produced these samples. This effect is known as **aliasing**.

Sampled Frequency Response

The sampled signal is given by $y^*(t) = y(t)\sum_{k=-\infty}^{\infty}\delta(t-kT)$

The sampler is periodic so can be represented by the Fourier series $\sum_{k=-\infty}^{\infty}\delta(t-kT) = \sum_{n=-\infty}^{\infty} C_n\, e^{jn\omega_s t}$

...where the Fourier coefficients are given by $C_n = \frac{1}{T}\int_{-T/2}^{T/2}\sum_{k=-\infty}^{\infty}\delta(t-kT)\ e^{-jn\omega_s t}dt$

Only one term is within range of the integration, so $C_n = \frac{1}{T}\int_{-T/2}^{T/2}\delta(t)\, e^{-jn\omega_s t}dt$

We can integrate this easily using the screening property of the delta function

$$C_n = \frac{1}{T}\left[e^0\right]_{-T/2}^{T/2} = \frac{1}{T}$$

So, the Fourier series representing the sampler is given by $\sum_{k=-\infty}^{\infty}\delta(t-kT) = \frac{1}{T}\sum_{n=-\infty}^{\infty} e^{jn\omega_s t}$

Sampled Frequency Response

$$\sum_{k=-\infty}^{\infty}\delta(t-kT) = \frac{1}{T}\sum_{n=-\infty}^{\infty} e^{jn\omega_s t}$$

We can now find the Laplace transform of the sampled system $\mathscr{L}\{f(t)\} = \int_{0+}^{\infty} f(t)\, e^{-st}dt = f(s)$

$$y^*(s) = \mathscr{L}\{\, y^*(t)\,\} = \int_0^{\infty} y(t)\left\{\frac{1}{T}\sum_{n=-\infty}^{\infty} e^{jn\omega_s t}\right\}e^{-st}dt$$

$$y^*(s) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\int_0^{\infty} y(t)\, e^{-(s-jn\omega_s)t}dt$$

The integral term is the same as the Laplace transform of $y(t)$, but with a change of complex variable

$$y^*(s) = \frac{1}{T}\sum_{n=-\infty}^{\infty} y(s-jn\omega_s)$$

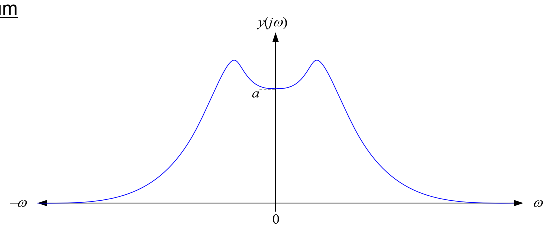
The frequency response of the samples signal is:

$$y^*(j\omega) = \frac{1}{T}\sum_{n=-\infty}^{\infty} y(\, j[\omega - n\omega_s])$$

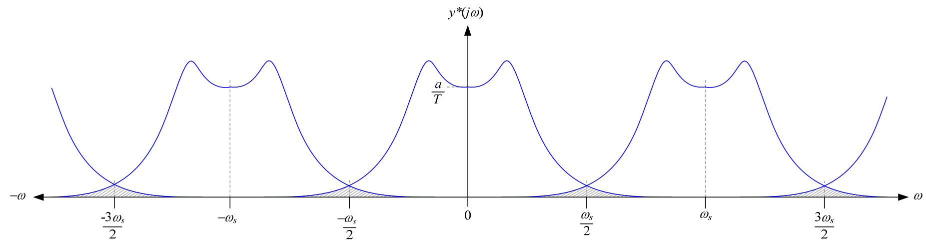
Each term in the infinite summation corresponds to the response of the continuous system, shifted along the frequency axis by $\pm n\omega_s$

Sampled Frequency Response

Continuous Spectrum

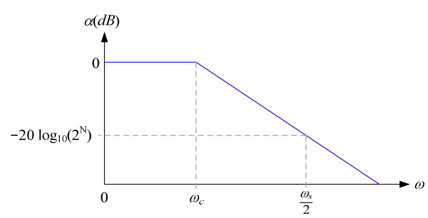


Sampled Spectrum

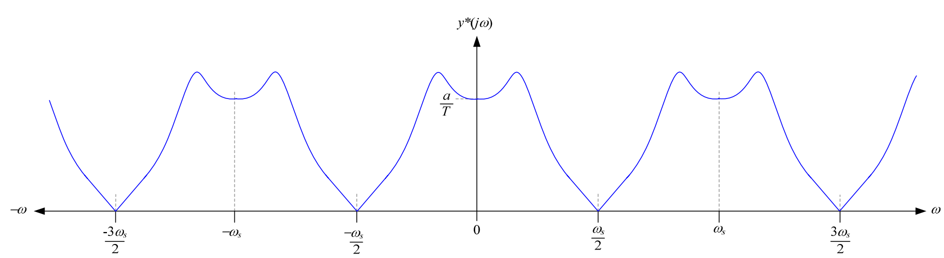


Anti-Aliasing Filter

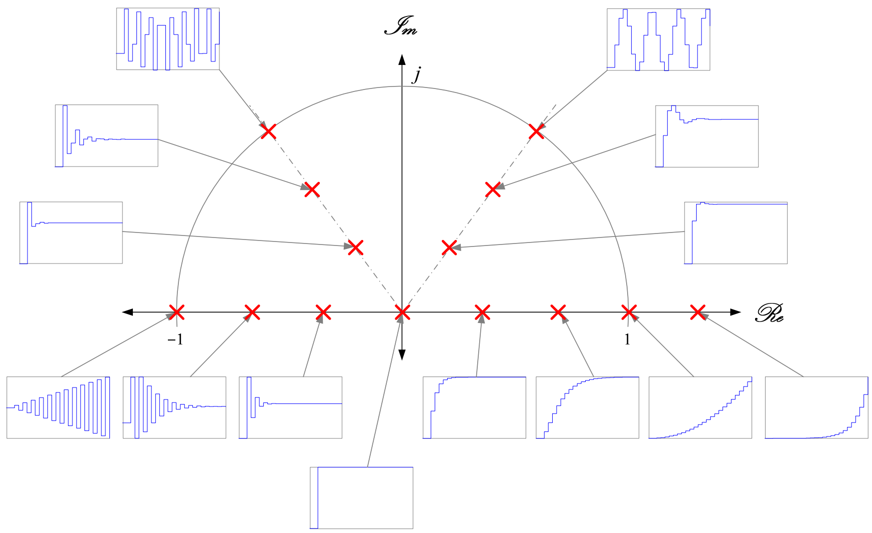
- To prevent aliasing, we need to attenuate the input signal to less than 1 converter bit at $\frac{\omega_s}{2}$ before sampling.



Filter constraints can be relaxed if a faster sample rate is selected.



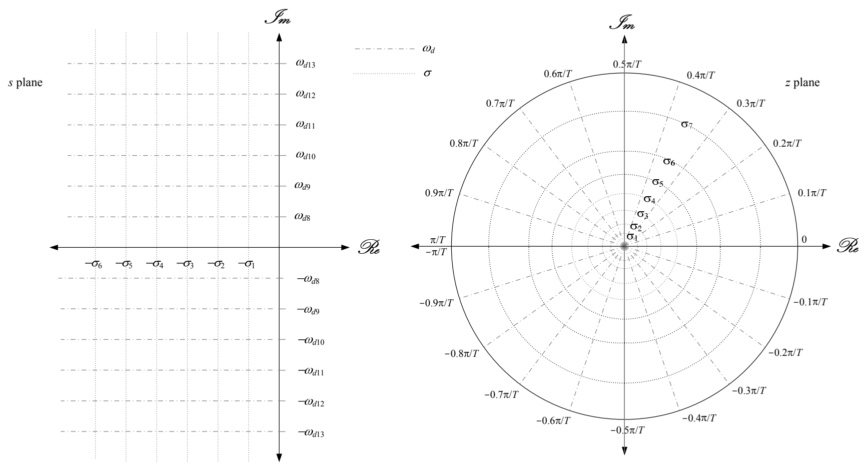
Pole Location vs. Step Response



Unit step response as a function of pole location for a second order system.

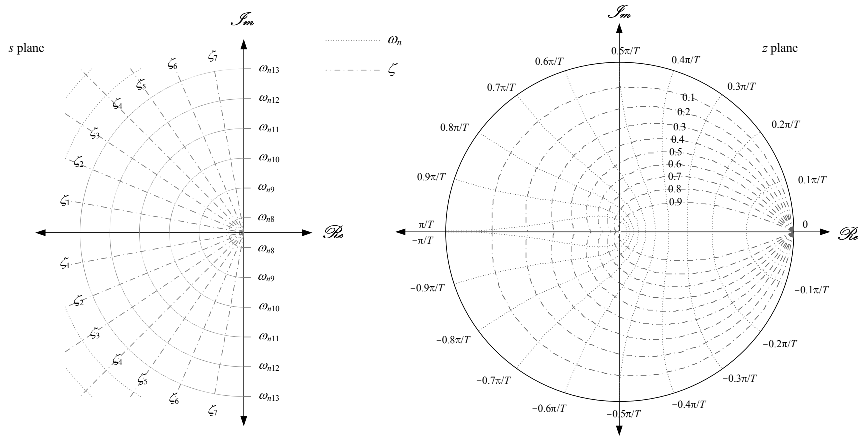
Complex Plane Grid

Lines of constant decay parameter (σ) and damped natural frequency (ω_d)

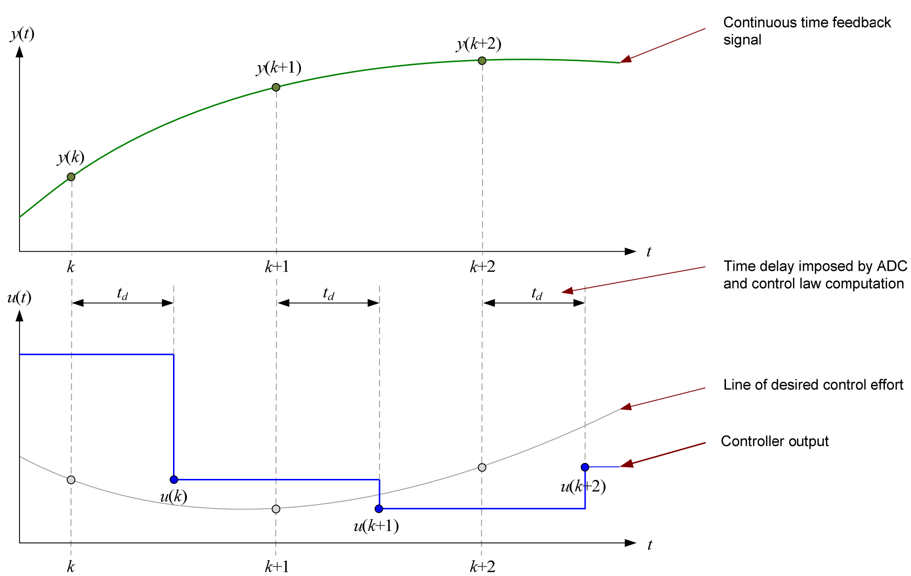


Complex Plane Grid

Lines of constant damping ratio (ζ) and un-damped natural frequency (ω_n)



Delay & Reconstruction



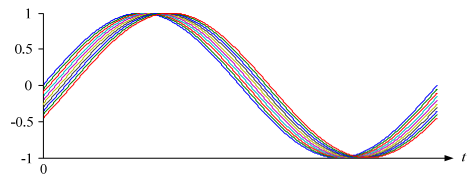
t_d = sample to output delay

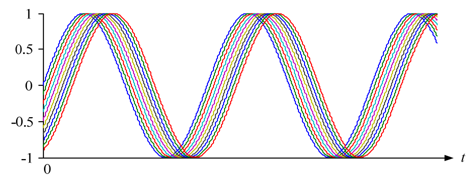
Time Delay

Consider a continuous signal $y(t)$ to which a fixed delay ϕ seconds is applied.

From the shifting property of the Laplace transform we know that $\mathscr{L}\{y(t-\phi)\} = e^{-s\phi} y(s)$

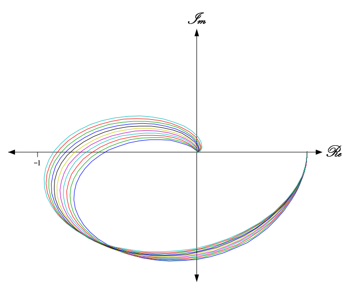
The influence of time delay is to change the phase of the signal by $-\omega\phi$, while the amplitude is unaffected.





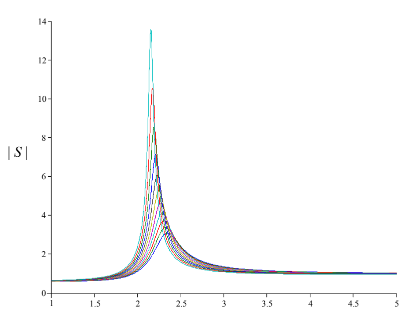
- Phase lag is indistinguishable from delay in the time domain: Delay in the time domain translates into frequency dependent phase lag in the frequency domain.

Effect of Time Delay



Plots show the effect of adding a progressively longer time delay to a stable third order system

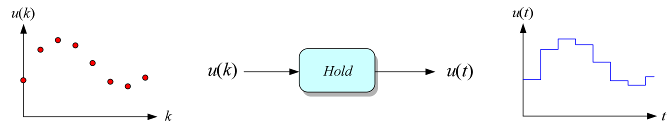
$$L(s) = e^{-\theta s} \frac{6}{(s+2)(s^2+s+4.25)}$$



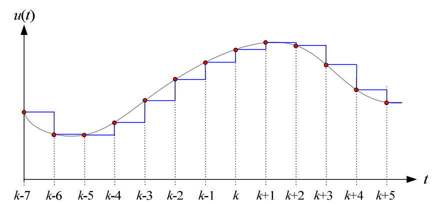
θ (sec)	PM (deg)	M_S
0	42.1174	3.05143
0.025	39.1419	3.34803
0.05	36.1664	3.6989
0.075	33.1909	4.1195
0.1	30.2154	4.63152
0.125	27.2398	5.26635
0.15	24.2643	6.07107
0.175	21.2888	7.11953
0.2	18.3133	8.53442
0.225	15.3378	10.5353
0.25	12.3623	13.5574

Reconstruction

Hold functions attempt to reconstruct a smooth continuous time signal from a discrete time sequence.



The only practical hold function considered is the Zero Order Hold (ZOH) which delivers a piece-wise constant output over the unknown interval $kT \le t \le (k+1)T$



The frequency response of the Zero Order Hold is modelled by that of a unit pulse over the sampling interval T.

Zero Order Hold

The frequency response of the Zero Order Hold can be modelled by that of a unit pulse over the sampling interval T.

$$F_{ZOH}(j\omega) = \frac{1 - e^{-j\omega T}}{j\omega}$$

This can be simplified using the exponential form of the sine function

$$F_{ZOH}(j\omega) = \frac{2j}{j\omega}\left(\frac{e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}}}{2j}\right)e^{-j\omega\frac{T}{2}} = \frac{2}{\omega}\sin\left(\frac{\omega T}{2}\right)e^{-j\omega\frac{T}{2}}$$

This is a complex number expressed in polar form, where the phase angle is given by

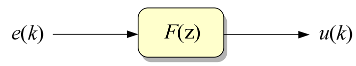
$$\angle F_{ZOH}(j\omega) \approx -\omega\frac{T}{2}$$

The approximation is accurate up to about half the Nyquist limit.

- The Zero Order Hold contributes a frequency dependent phase lag to the open loop response

Discrete Time Controller Design

The result of discrete time controller design is a difference equation involving current and previous terms in $e(k)$ and $u(k)$.



There are two approaches to the discrete time design:

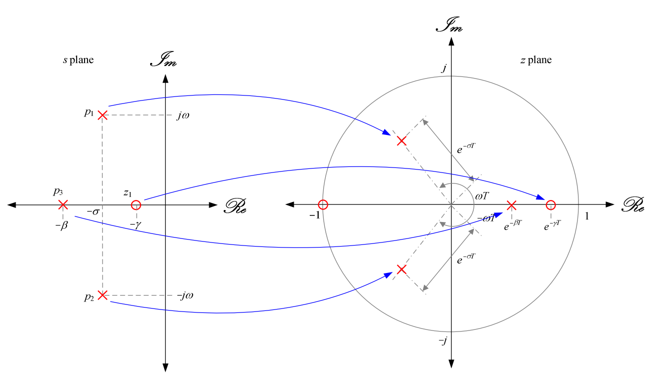
- In **design by emulation**, we transform an existing controller design into the z domain, then find a corresponding difference equation. The following methods are common:
  - Pole-zero matching
  - Numerical approximation
  - Hold Equivalent

- In **direct digital design**, we carry out the entire controller design in the z domain using one of the methods previously described (Nyquist, root locus, ...etc.).

In general, direct design methods yield superior performance for the same sample rate, however access to computer design tools is very desirable.

Pole-Zero Matching

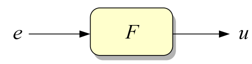
1. Transform the poles & zeros of the transfer function using $z = e^{sT}$
2. Map any infinite zeros to $z = -1$ (but maintain a relative degree of 1)
3. Match the gain of the transformed system at $z = 1$ to that of the original at $s = 0$



$$F(s) = A_s \frac{(s - z_1)}{(s - p_1)(s - p_2)(s - p_3)} \quad \Longrightarrow \quad F(z) = A_z \frac{(z+1)(z - e^{-z_1 T})}{(z - e^{-p_1 T})(z - e^{-p_2 T})(z - e^{-p_3 T})}$$

4.1, 4.2

Numerical Approximation

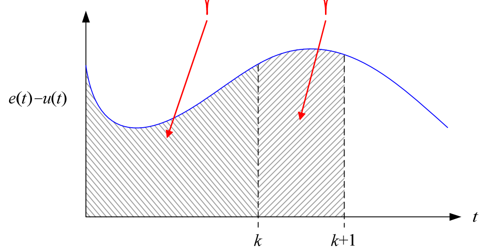


Starting with the simple controller $F(s) = \dfrac{a}{s+a}$ we get the differential equation $u'(t) + au(t) = ae(t)$

The solution to the continuous equation is $u(t) = a\int_0^t (e(\tau) - u(\tau))\,d\tau$

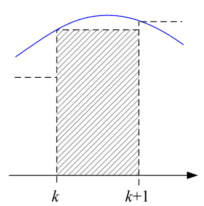
An equivalent discrete time controller performs this integration in discrete time:

$$u(k+1) = \underbrace{u(k)} + \underbrace{a\int_k^{k+1} \big(e(\tau) - u(\tau)\big)\,d\tau}$$



Forward Approximation

$$u(kT+T) = u(kT) + a\int_{kT}^{kT+T} (e(\tau) - u(\tau))\,d\tau$$



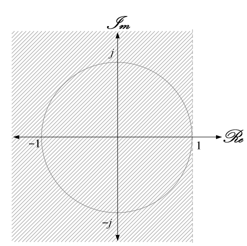
The integral portion can be approximated by a rectangle area:

$$u(k+1) = u(k) + aT\big[e(k) - u(k)\big]$$

Using the shifting property of the *z*-transform: $\mathcal{Z}\{f(k+n)\} = z^n F(z)$

$$zu(z) = u(z) + aTe(z) - aTu(z)$$

$$F(z) = \frac{u(z)}{e(z)} = \frac{a}{\left(\frac{z-1}{T}\right) + a}$$



The forward approximation method implies we can find the z-transform directly from the Laplace transform by making the substitution:

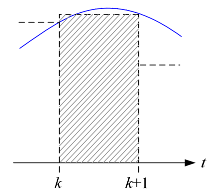
$$s \leftarrow \frac{z-1}{T}$$

The forward approximation rule maps the ROC of the s plane into the region shown. The unit circle is a subset of the mapped region, so stability is not necessarily preserved under this mapping.

Backward Approximation

Approximating the unknown area using a rectangle of height $a\{e(k+1) - u(k+1)\}$...

$$u(k+1) = u(k) + aT[e(k+1) - u(k+1)]$$

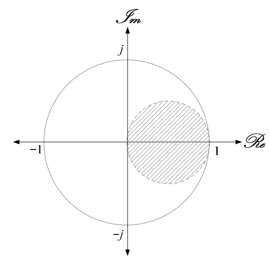


Application of the shifting theorem and simple algebra leads to...

$$F(z) = \frac{a}{\left(\frac{z-1}{Tz}\right) + a}$$

The backward approximation method implies we can find the z-transform directly from the Laplace transform by making the substitution:

$$s \leftarrow \frac{z-1}{Tz}$$

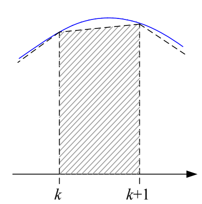


The backward approximation rule maps the ROC of the s plane into a circle of radius 0.5 within the z plane unit circle. Pole-zero locations are very distorted under this mapping.

Trapezoidal Approximation

Approximating the unknown area using a trapezoid...

$$u(k+1) = u(k) + \frac{aT}{2}[e(k) - u(k) + e(k+1) - u(k+1)]$$

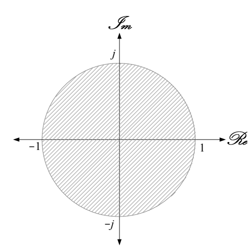


Application of the shifting theorem and simple algebra leads to...

$$F(z) = \frac{a}{\left(\frac{2}{T}\frac{z-1}{z+1}\right) + a}$$

The trapezoidal approximation method implies we can find the z-transform directly from the Laplace transform by making the substitution:

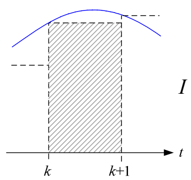
$$s \leftarrow \frac{2}{T}\frac{z-1}{z+1}$$



Trapezoidal approximation maps the ROC of the s plane exactly into the unit circle.

This method is also known as **Tustin's method** or the **bi-linear transform**.

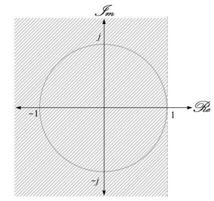
Numerical Approximation Methods

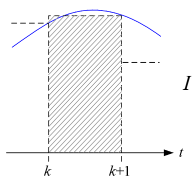


Forward approximation

$$I = aT\left[e(k) - u(k)\right]$$

$$s \leftarrow \frac{z-1}{T}$$

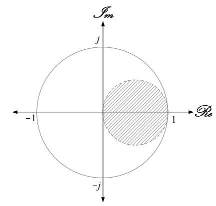


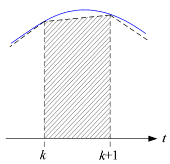


Backward approximation

$$I = aT\left[e(k+1) - u(k+1)\right]$$

$$s \leftarrow \frac{z-1}{Tz}$$

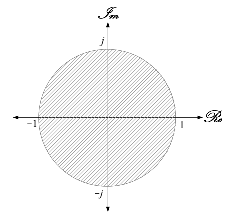




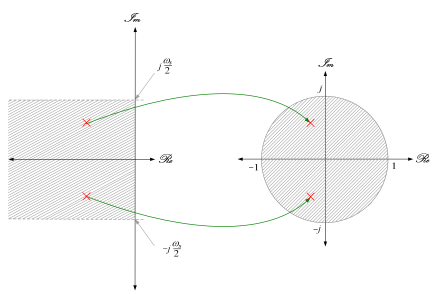
Trapezoidal approximation

$$I = a\frac{T}{2}\left[e(k+1) - u(k+1) - e(k) + u(k)\right]$$

$$s \leftarrow \frac{2}{T}\frac{z-1}{z+1}$$

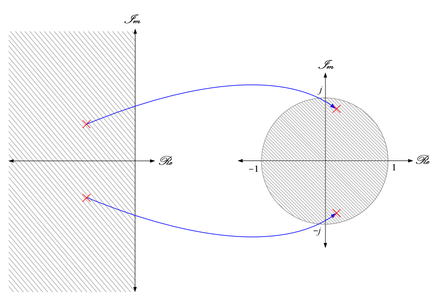


Frequency Warping



$$z = e^{sT}$$

The correct transformation maps only the primary strip inside the unit circle.



$$z = \frac{2 + sT}{2 - sT}$$

The Tustin transformation maps the entire LHP inside the unit circle. Pole & zero frequencies are said to be **warped** by the transformation.

Frequency Warping

The Tustin transformation is: $s \leftarrow \frac{2}{T}\frac{z-1}{z+1}$

The frequency response of the continuous time prototype $F(s) = s$ is evaluated as

$$F(s)\big|_{s=j\omega} = j\omega$$

Evaluating the frequency response of the equivalent discrete time system...

$$F(z)\big|_{z=e^{j\omega T}} = \frac{2}{T}\frac{e^{j\omega T}-1}{e^{j\omega T}+1} = j\frac{2}{T}\tan\omega\frac{T}{2}$$

Compared with the continuous time system, we see that the frequency response of the discrete time system has been "warped" by the above formula.

This effect can be compensated by **pre-warping** the pole-zero frequencies of the original system prior to transformation by the Tustin method.

Pre-Warping

The technique of **pre-warping** changes the s-plane location of each pole such that it is mapped by the Tustin transformation to the correct place in the z-plane.

1. Re-write the desired characteristic in the form $G\left(\frac{s}{\omega_1}\right)$
2. Replace ω_1 by a, such that $a = \frac{2}{T}\tan\frac{\omega_1 T}{2}$
3. Transform using the Tustin method $s \leftarrow \frac{2}{T}\frac{z-1}{z+1}$
4. Match the gain of the original system at $s = 0$ with that of the transformed system at $z = 1$

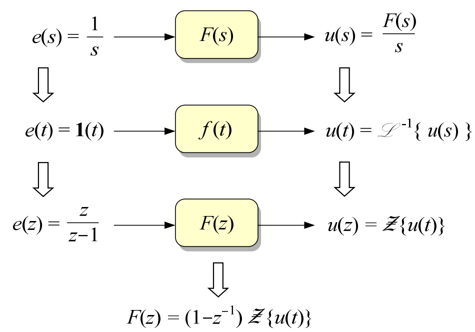
For systems with multiple critical frequencies which must be preserved, each frequency must be warped using the formula in step 2 prior to design in the continuous domain.

4.3

Step Invariant Method

Invariant methods emulate the response of the continuous system to a specific input.

1. Determine the output of the output of the continuous time system for the selected hold input
2. Find the corresponding ***z***-transform of the response
3. Divide by the ***z***-transform of the selected input



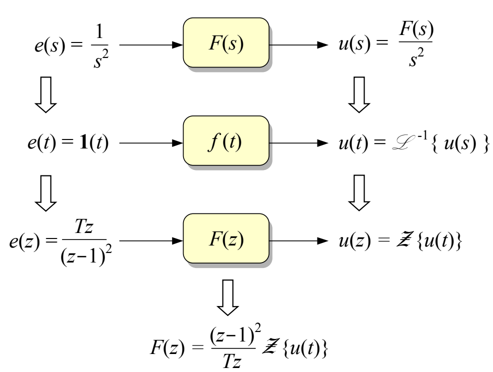
The step invariant method is also known as the **ZOH equivalent** method.

Invariant methods capture the gain & phase characteristics of the respective hold unit.

Ramp Invariant Method

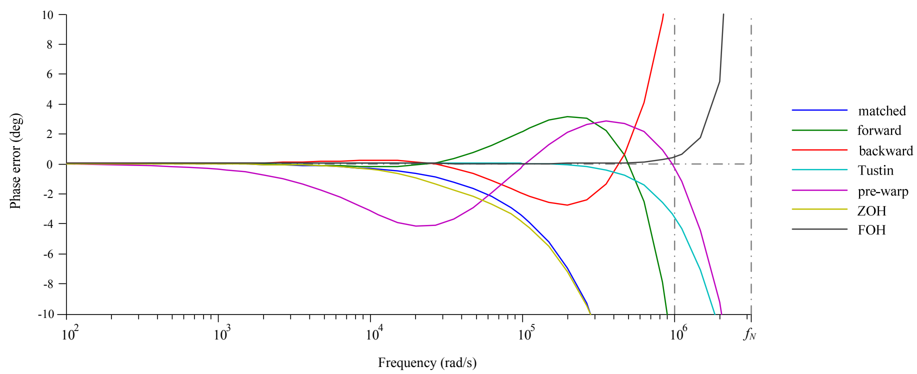
The ramp invariant method emulates the response of the continuous system to a ramp input.

Except for the input reference the method is identical to the step invariant method.



The ramp invariant method is also known as the **FOH equivalent** method.

Phase Error Comparison



Comparison of phase performance for various discrete transformation methods.

Summary of Emulation Methods

Matched pole-zero	Relatively easy hand calculation with good performance, but computation delay imposes significant phase lag.
Forward approximation	Easiest method for hand calculation, but performance is very dependent on sample rate. Can potentially convert a stable design into an unstable one.
Backward approximation	Produces significant phase error at low frequencies due to warping of the stability region during mapping.
Tustin's method	Best compromise between ease-of-calculation and performance. Pre-warping enables phase to be preserved at specific frequencies.
Step invariant	Most accurate, since it accounts for phase shift induced by the ZOH unit. Used for direct digital design methods.
Ramp invariant	Best overall performance, but need access to design tools for computation.

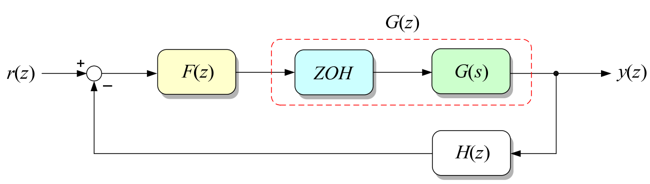
Recommendations

- If a zero order hold element is present, use the step invariant (ZOH equivalent) method <u>once</u> in the design. This will capture phase lag effects introduced by the ZOH.

- If multiple elements must be transformed and the ZOH effect has already been accounted for, use the ramp invariant (FOH equivalent) method for the remaining elements.

- If computer design tools are not available, Tustin's method represents a good compromise between performance and ease of calculation. Remember to account for ZOH phase effects separately.

Direct Digital Design



- In direct design we begin by transforming the plant model into discrete form using the step invariant method. This captures the action of the ZOH element which precedes the plant.

- Any of the standard design techniques (e.g. root locus, Bode, etc.) can then be used to synthesize the controller.

- The design cycle iterates as many times as necessary until a satisfactory controller is found.

- For the same sample rate, control performance with the direct method can be significantly better than with emulation methods.

4.4, 4.5

Discrete Time Control

- Sampling
  - The sampling process changes the frequency characteristics of the feedback signal. Understanding of the relationship between s- and z-planes is key to good digital design.
  - Careful selection of sample rate is the first and most critical step in design.
- Controller design
  - Emulation techniques allow legacy analogue controller designs to be re-used. Trade-offs exist between computational complexity and performance of each method.
  - Design in the digital domain yields superior performance for the same sample rate. Classical design techniques (Bode, Nyquist, root locus, ...) can be used, with modifications to account for the discrete time nature of the signals and sub-systems.
  - State space design methods for continuous and discrete time systems are similar.
- Time delay
  - Conversion and computational delays are unavoidable in practice. These contribute a net phase lag to the open loop response which is proportional to frequency. Phase margin is eroded!
  - Reconstruction using zero order hold contributes additional phase lag.

Agenda Review

Day 1: *Control Theory Fundamentals*

1. **Fundamental Concepts**
   Linear systems, the Laplace transform, dynamic response, classification of systems
2. **Feedback Control**
   Effects of feedback, the Nyquist plot, phase compensation, sensitivity & tracking, robustness
3. **Transient Response**
   Transient specifications, steady state error, PID controllers, root locus analysis
4. **Discrete Time Systems**
   Sampled systems, the z-transform, complex plane mapping, aliasing, discrete transformations

Day 2: *State Space Control*

5. **State Space Systems**
   Co-ordinate transformations, eigenvalues & eigenvectors, lumped parameter systems, realisations
6. **Properties of Linear Systems**
   Stability, modal decomposition, controllability, observability, canonical forms
7. **State Feedback Control**
   State feedback, pole placement, eigenstructure assignment, input matrix design, integral control
8. **Linear State Estimators**
   State reconstruction, state estimator design, reduced order observers, separation principle

Recommended Reading

Control Theory Fundamentals

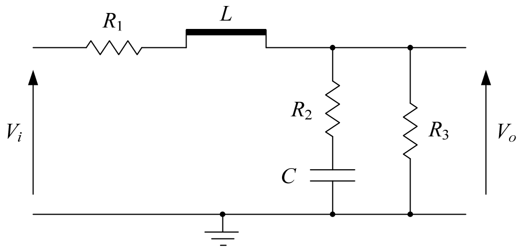
- J.J.DiStefano, A.R.Stubberud & I.J.Williams, ***Feedback & Control Systems***, Schaum, 2011
- J.Doyle, B.Francis & A.Tannenbaum, ***Feedback Control Theory***, Macmillan, 1990
- J.Schwarzenbach & K.F.Gill, ***System Modelling & Control***, Edward Arnold, 1992

State Space Control

- W.L.Brogan, ***Modern Control Theory***, Prentice-Hall, 1991
- G.F.Franklin, J.D.Powell & M.L.Workman, ***Digital Control of Dynamic Systems***, Addison-Wesley, 1998
- K.J.Åstrom & R.M.Murray, ***Feedback Systems***, Princeton, 2010

Day 1 Tutorials

1.1 The electrical network shown below forms the output filter of a switching power converter. Use Matlab to investigate its dynamic properties.



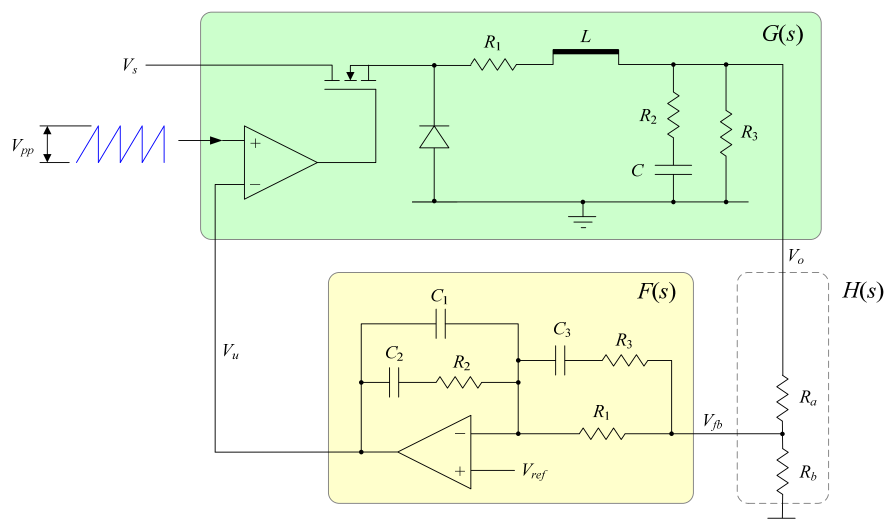
Passive component values are:
$R_1 = 2.2\text{m}\Omega$
$R_2 = 0.6\text{m}\Omega$
$R_3 = 1\Omega$
$L = 0.9\mu\text{H}$
$C = 471\mu\text{F}$

1.2 Investigate the transient response characteristics of the second order system with damping ratio 0.2 and natural frequency 1 radian/second.

2.1 Design a phase compensator for the following plant with unity feedback which achieves a phase margin of at least 45º at a cross-over frequency of 3 rad/s, and steady state error following a unit step of no more than 0.01.

$$G(s) = \frac{1}{(10s+1)(s+1)}$$

2.2 The diagram below shows a type 3 compensator applied to a switching buck regulator.



The filter stage components are as shown in tutorial 1.1. The feedback divider has an attenuation of 0.5. Design a phase lead compensator to achieve a phase margin of at least 45º and a gain margin of greater than 10dB. Ensure the cross-over frequency is above 10kHz (6.28 x 10^4 rad/s).

2.3 Use the Q-parameterisation technique to design a closed loop controller for the switching power supply described in tutorial 2.2. Select a first order target closed loop response with unity gain and roll-off above 2,000 rad/s. Verify the step response of the completed design.

3.1 Find the parallel form PID controller settings which optimise the unit step response of the following plant. Repeat the PID tuning exercise for the switching converter described in tutorial 2.2.

$$G(s) = \frac{s+1}{s^3 + 3s^2 + 5s + 1}$$

3.2 Design a unity feedback controller for following plant which achieves a rise time following a unit step input of less than 0.5 seconds, with overshoot of less than 20% and steady state error less than 2%.

$$G(s) = \frac{1}{s^2 + 4s + 8}$$

3.3 Investigate the dynamic properties of the following plant under feedback control and design a compensator to correct any undesirable transient response characteristics.

$$G(s) = \frac{s^2 + 2s + 4}{s(s+4)(s+6)(s^2 + 1.4s + 1)}$$

4.1 Use the matched pole-zero method to transform the following continuous time controller with sample frequency 5Hz.

$$F(s) = \frac{10s+1}{s+2}$$

4.2 Draw the Bode plot of the continuous time controller designed in tutorial 2.2 and investigate the performance of equivalent discrete time controllers created using the emulation design techniques described in the seminar.

4.3 Use pre-warping to transform the following simple lag filter to preserve the frequency of the pole at 4 rad/s. Use a sample frequency of 10Hz.

$$G(s) = \frac{1}{s+4}$$

4.4 Design a 1Hz digital controller for the following plant which achieves a peak overshoot following a step input of less than 16%. Also, ensure the settling time to within 1% of steady state is less than 10 seconds.

$$G(s) = \frac{1}{s(10s+1)}$$

4.5 Design a discrete time controller for the buck converter of tutorial 2.2 with a sampling frequency of 1MHz. The closed loop system should have zero steady state error, and a phase margin of at least 40 degrees. Gain cross-over should be at least 10kHz.

This page intentionally left blank

This page intentionally left blank