Control Systems

IMAD M. JAIMOUKHA

EE2/ISE2

Room 1111C

Extension 46279

E-mail: i.jaimouka@imperial.ac.uk

Organisation

- 18 lectures and 9 tutorials/study groups.
- Notes to be handed out.
- The notes are only notes, use books for background.
- Supplementary notes (*NOT EXAMINABLE*) occasionally handed out.
- Problem sheets (together with answers) will be handed out.
- Past exam papers (from 3 and 4 years ago) will be handed out.
- The problems will be of type you would expect to answer in the exam, with some exceptions:
  - There are book-work type questions which make sense in an exam but which do not on a problem sheet.
  - Exam questions are generally easier.
  - Computationally intensive questions are not suitable for examination.

Useful Books

(Use latest editions)

- *Modern Control System Theory and Design*, S.M. Shinners, Wiley. Good for the state-space approach.
- *Modern Control Systems*, R.C. Dorf and R.H. Bishop, Addison-Wesley. Good for the transfer function approach.
- *Feedback Control of Dynamic Systems*, G.F. Franklin, J.D. Powell and A. Emami-Naeini, Addison-Wesley. Good for practical design issues.
- *Feedback Control Systems*, C.L. Phillips and R.D Harbor, Prentice-Hall. Follows the lecture course closely, and is useful for third year course too.

Course Outline

Page

1. Introduction 6
   - Aims of control
   - Examples of Control Systems
   - The feedback design dilemma
   - Control design tasks
   - Feedback for static systems
2. Mathematical Models of Dynamic Systems 13
   - Laplace transform
   - Transfer functions
   - Electric circuit models
   - Mechanical systems
   - Control System Modelling
3. Block Diagram Algebra and Signal Flow Graphs (**Not examinable**) 25
   - Prototype Feedback System
   - Block Diagram Algebra
   - Signal Flow Graphs and Mason's Rule
4. State-Variable Models (**Not examinable**) 41
   - General state-variable form
   - Simulation diagrams
   - Solution of state-variable equations
   - Relation to transfer functions
   - Stability and the matrix exponential function

5. System Responses 57
- Test signals
- First order system
- Steady-state response
- Second order system
- Time response specifications
- Frequency response of systems
- Steady-state Accuracy

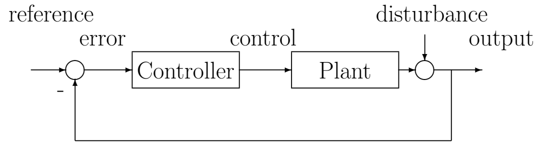
6. Stability Analysis 77
- The problem of stability in control
- Routh-Hurwitz stability criterion

7. Root-locus Analysis and Design Techniques 97
- Stability of feedback systems
- The root-locus
- Rules for plotting the root-locus
- Root-locus design tools

8. Frequency Response Methods 141
- Conformal mapping
- The Nyquist stability criterion
- Phase lead - phase lag compensation
- PID control

Aims of Control

Use of feedback to force a given plant to exhibit desired characteristics



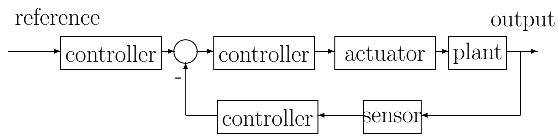
- Examples of control systems include:
  - Domestic hot water heater - temperature feedback used to control the water temperature.
  - Control of the concentration of a fluid - opacity used to control concentration.
  - Lateral and longitudinal control of a Boeing 747.
  - Active control of a segmented optical system.

The Control Problem

- Typical aims of feedback:
  - disturbance rejection
  - tracking improvement
  - transient response shaping
  - sensitivity reduction
  - reduction in effect of nonlinearity
  - closed-loop stability
- The feedback design dilemma:
  Almost all the benefits of feedback can be achieved, provided that the loop gain is sufficiently high (over a certain frequency range). *Unfortunately, for most plants, high loop gain tends to drive the system into instability.*

Control Design Tasks

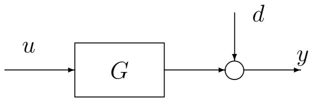
Most control systems contain some of the following:



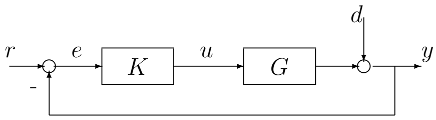
- Output and sensor selection
- Input and actuator selection
- Model building
- Controller design
- Simulation testing
- Hardware testing

Feedback for Static Systems

– Consider the open loop static system :



– Introducing feedback compensation:



– The tracking error signal is given by

$$e = r - (d + GKe)$$
$$\Rightarrow \quad e = \frac{1}{1+GK}\,(r-d)$$

– Hence, the output signal is

$$y = d + GKe$$
$$\Rightarrow \quad y = \frac{GK}{1+GK}\,r + \frac{1}{1+GK}\,d$$

– Finally, the control signal is

$$u = Ke = \frac{K}{1+GK}\,(r-d)$$

Advantages of Feedback

– Tracking and disturbance rejection:

$$y = \frac{GK}{1+GK}\,r + \frac{1}{1+GK}\,d$$

Good tracking and disturbance rejection both require high loop gain

– Sensitivity reduction (neglecting the disturbance):

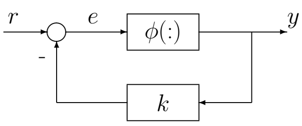
$$y = Cr, \qquad C = \frac{GK}{1+GK}$$

The sensitivity of the closed loop to changes in G is then given by

$$\begin{aligned} S &= \frac{\partial C/C}{\partial G/G} = \frac{\partial C}{\partial G}\frac{G}{C} \\ &= \frac{K(1+GK) - KGK}{(1+GK)^2}\,\frac{G}{GK}\,(I+GK) \\ &= \frac{1}{1+GK} \end{aligned}$$

Sensitivity reduction requires high loop gain

– Reducing the effects of nonlinearity:



Here,

$$y = \phi(e), \qquad e = r - k\phi(e)$$

– Hence

$$\frac{y}{r} = \frac{\phi(e)}{e + k\phi(e)} = \frac{\phi(e)/e}{1 + k\phi(e)/e}$$

– It follows that

$$k\,\frac{\phi(e)}{e} \gg 1 \quad \Rightarrow \quad \frac{y}{r} \approx \frac{1}{k}$$

Reducing nonlinear effects requires high loop gain

– Other advantages
- ∗ Improvement of transient response
- ∗ Stabilisation of unstable systems

The Cost of Feedback

– Increase in system complexity:
- ∗ Sensors
- ∗ Actuators

– Loss of gain: In the previous figure:
- ∗ Open loop gain : $\phi(e)/e$
- ∗ Closed loop gain : $\frac{\phi(e)/e}{1+k\phi(e)/e} \to \frac{1}{k}$ as $\frac{k\phi(e)}{e} \to \infty$
- ∗ Normally: $\phi(e)/e >> k^{-1}$

– Control signal energy (absent in open loop):

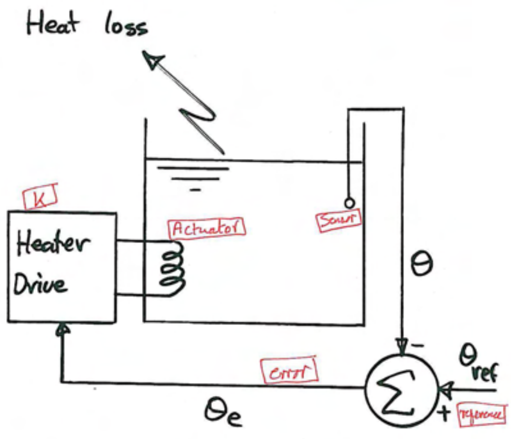
$$u = \frac{K}{1+GK}(r-d)$$

– Possible introduction of instability: what happens when $1 + GK = 0$?

– Other cost include:
- ∗ Sensor noise: requires low loop gain
- ∗ Design exercise

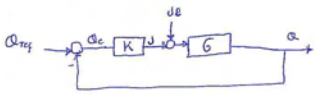
IN MOST CASES THE ADVANTAGES OF FEEDBACK FAR OUTWEIGH THE COST AND FEEDBACK IS UTILISED

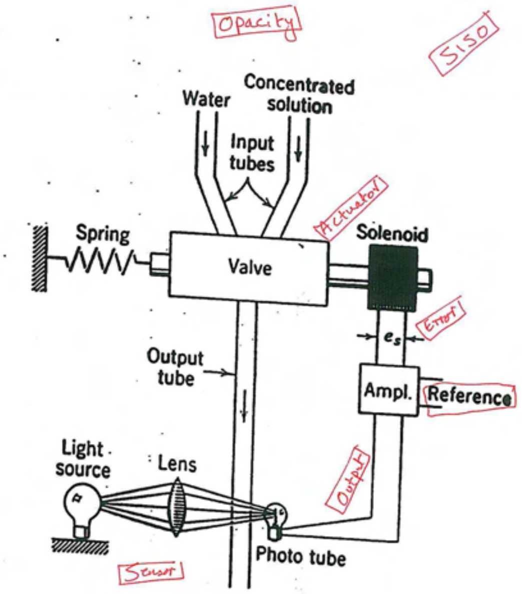
Simple on-off | bang-bang



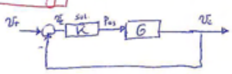
Water heating controller

Domestic



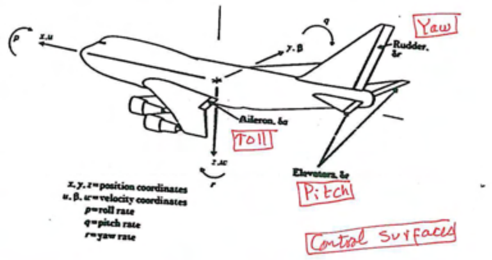


Process Industry



More Complex

FIGURE 7.29
Definition of aircraft coordinates.



Rudder → Yaw → Roll → Multivariable, Interaction

Aerospace Industry

Alternative to High Quality Mirror

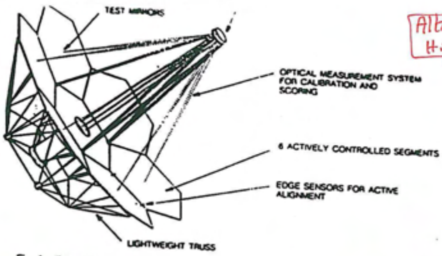


Fig. 1. Schematic view of ASCIE test bed illustrating important features for a segmented optical system.

$3 \times 4 \times 2 = 24$ EDGE SENSORS
$6 \times 3 = 18$ ACTUATORS
$6 + 1 = 7$ OPTICAL SENSORS

Alignment
Disturbances

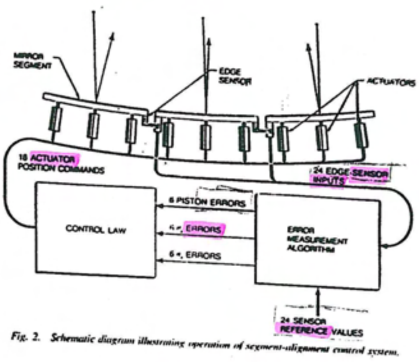
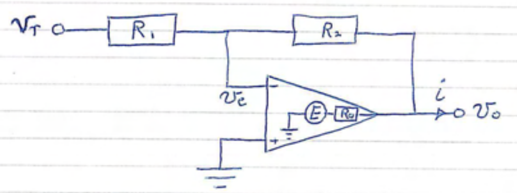


Fig. 2. Schematic diagram illustrating operation of segment-alignment control system.

Astronomy

+ Transport, Power, Electronics, Economics

Voltage Regulator



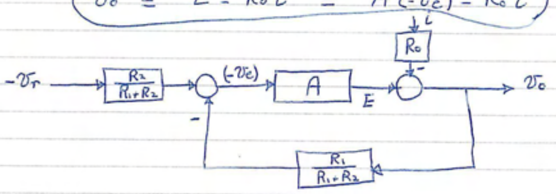
$$\boxed{E = -A v_e}$$

$$\frac{v_e - v_r}{v_o - v_r} = \frac{R_1}{R_1+R_2} \cancel{v_o}$$

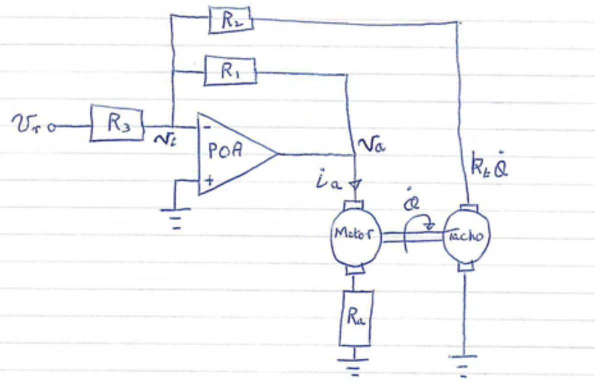
$$\therefore \quad v_e = v_r + \frac{R_1}{R_1+R_2} v_o - \frac{R_1}{R_1+R_2} v_r = \frac{R_2}{R_1+R_2} v_r + \frac{R_1}{R_1+R_2} v_o$$

$$\therefore \quad -v_e = \frac{R_2}{R_1+R_2}(-v_r) - \frac{R_1}{R_1+R_2} v_o$$

$$v_o = E - R_o i = A(-v_e) - R_o i$$



Voltage Feedback for DC Motor

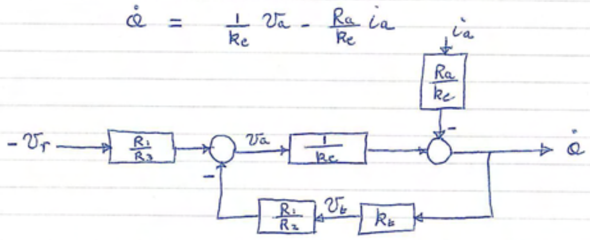


Motor Equation : $\quad v_a = k_e \dot{\theta} + R_a i_a$

Opamp Equation : $\quad \frac{v_a}{R_1} + \frac{v_r}{R_3} + \frac{k_t \dot{\theta}}{R_2} = 0$

$$\therefore \quad v_a = \frac{R_1}{R_3}(-v_r) - k_t \frac{R_1}{R_2} \dot{\theta}$$

$$\dot{\theta} = \frac{1}{k_e} v_a - \frac{R_a}{k_e} i_a$$



Loop gain ~~large~~ : $\frac{1/k_e}{1 + \frac{R_1}{R_2} \frac{k_t}{k_e}} \simeq \frac{R_2}{R_1 k_t} \rightarrow \dot{\theta} \simeq -\frac{R_1}{R_3} \frac{R_2}{R_1 k_t} \cdot v_r$

$$\simeq -\frac{R_2}{R_3} \frac{1}{k_t} v_r$$

Mathematical Models of Dynamic Systems

- Introduction
- Laplace Transform
- Inverse Laplace Transform
- Applications to Differential Equations
- Models for Electric Circuits
- Transfer Functions
- Models for Mechanical Systems

Introduction

There are many ways of describing linear dynamical systems, but the following three occur most frequently:

- **A system of ordinary differential equations** often the 'initial' description of a system, in terms of its physics and interconnections. The variables are just the inputs and outputs.

- **A transfer function model**. A single input singe output process is described by its transfer function: the ratio of the Laplace transform of output and input. Particularly convenient for the design of controllers for dynamical systems.

- **A state variable model**: a system of first order ordinary (linear) differential equations. Descriptions of the first form can be reduced to this form by the introduction of the intermediate variables called 'states'. This form is convenient for simulation, computation and certain forms of analysis.

Laplace Transforms

- Used to solve differential equations via algebraic equations.

- Given $f(t), 0 \leq t \leq \infty$, with $f(t) = 0$ for $t \leq 0$, we define the **Laplace transform** of $f(t)$ via:
$$F(s) = \mathcal{L}\{f(t)\} = \int_0^\infty e^{-st} f(t) dt.$$

- **Example:** $f(t) = 1$ for $t \geq 0$:
$$F(s) = \int_0^\infty e^{-st} dt = -\frac{1}{s} e^{-st}|_0^\infty = \frac{1}{s} \cdot$$

- **Example:** $f(t) = e^{at}$, for $t \geq 0$:
$$F(s) = \int_0^\infty e^{-st} e^{at} dt = \int_0^\infty e^{-(s-a)t} dt = \frac{1}{s-a} \cdot$$

- The Laplace transform is a **linear operator**:
$$\mathcal{L}\{k_1 f_1(t) + k_2 f_2(t)\} = k_1 \mathcal{L}\{f_1(t)\} + k_2 \mathcal{L}\{f_2(t)\}$$

- **Example:** $f(t) = \cos \omega t = 0.5(e^{j\omega t} + e^{-j\omega t})$:
$$F(s) = \frac{0.5}{s - j\omega} + \frac{0.5}{s + j\omega} = \frac{s}{s^2 + \omega^2}.$$

- There is a table of standard transforms and manipulative rules-properties.

- **Example:** Suppose $\mathcal{L}\{f(t)\} = F(s)$. Then
$$\mathcal{L}\{e^{-\alpha t} f(t)\} = F(s + \alpha).$$
Proof: $\mathcal{L}\{e^{-\alpha t} f(t)\} = \int_0^\infty e^{-st} e^{-\alpha t} f(t) dt$
$$= \int_0^\infty e^{-t(s+\alpha)} f(t) dt = F(s + \alpha).$$

- **Example:** $\mathcal{L}\{e^{-\alpha t} \cos \omega t\}$:
$$\mathcal{L}\{\cos \omega t\} = \frac{s}{s^2 + \omega^2}$$
$$\Rightarrow \quad \mathcal{L}\{e^{-\alpha t} \cos \omega t\} = \frac{s + \alpha}{(s + \alpha)^2 + \omega^2}$$

- **Example:** $\mathcal{L}\{\dot{x}(t)\} = s\mathcal{L}\{x(t)\} - x(0)$.
$$\begin{aligned} Proof: \mathcal{L}\{\dot{x}(t)\} &= \int_0^\infty e^{-st} \dot{x}(t) dt \\ &= e^{-st} x(t)|_0^\infty + s \int_0^\infty e^{-st} x(t) dt \\ &= s\mathcal{L}\{x(t)\} - x(0) \end{aligned}$$

- **Example:** $\mathcal{L}\{\ddot{x}(t)\} = s^2 \mathcal{L}\{x(t)\} - sx(0) - \dot{x}(0)$.

Inverse Laplace Transform

- Let $F(s) = \mathcal{L}\{f(t)\}$. Then

$$f(t) = \frac{1}{2\pi j} \int_{\sigma-j\omega}^{\sigma+j\omega} F(s)e^{st}ds, \quad \omega \to \infty.$$

In general, this expression is difficult to evaluate. However, we can usually use partial fraction expansion and tables of Laplace transforms.

- **Example:** Using the "covering rule":

$$F(s) = \frac{s+1}{s(s+2)} = \frac{0.5}{s} + \frac{0.5}{s+2} \Rightarrow f(t) = \frac{1+e^{-2t}}{2}$$

- **Example:**

$$\begin{aligned} F(s) &= \frac{s+3/5}{(s+1)^2+2^2} = \frac{(s+1)-2/5}{(s+1)^2+2^2} \\ &= \frac{s+1}{(s+1)^2+2^2} - \frac{1}{5}\frac{2}{(s+1)^2+2^2} \\ \Rightarrow \quad f(t) &= e^{-t}\cos 2t - 0.2e^{-t}\sin 2t \end{aligned}$$

- **Example:**

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

$$\Rightarrow \; f(t) = e^{-2t} + e^{-t} - te^{-t}$$

- *The inverse Laplace transform of any strictly proper rational function of s is a combination of exponentials, sinusoids and polynomials in t.*

Applications to Differential Equations

- Consider the differential equation:

$$\ddot{y}(t) + 4\dot{y}(t) + 5y(t) = u(t), \; y(0) = y_0, \; \dot{y}(0) = 0$$

- Taking Laplace transforms:

$$s^2Y(s) - sy_0 + 4(sY(s) - y_0) + 5Y(s) = U(s)$$

$$\Rightarrow \quad Y(s) = \frac{U(s)}{s^2+4s+5} + \frac{(s+4)y_0}{(s^2+4s+5)}$$

- Suppose that $U(s) = 1/s$:

$$\begin{aligned} Y(s) &= \frac{1}{s(s^2+4s+5)} + \frac{(s+4)y_0}{(s+2)^2+1} \\ &= \frac{1}{5}\{\frac{1}{s} - \frac{s+4}{(s+2)^2+1}\} + \frac{(s+4)y_0}{(s+2)^2+1} \\ &= \frac{1}{5}\{\frac{1}{s} - \frac{s+2}{(s+2)^2+1} - \frac{2}{(s+2)^2+1}\} \\ &\quad + \frac{(s+2)y_0}{(s+2)^2+1} + \frac{2y_0}{(s+2)^2+1} \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad y(t) &= \frac{1}{5}\{1 - e^{-2t}\cos t - 2e^{-2t}\sin t\} \\ &\quad + y_0e^{-2t}(\cos t + 2\sin t) \end{aligned}$$

Models for Electric Circuits

- We know that:

$$v(t) = Ri(t) \; \Rightarrow \; V(s) = RI(s)$$

is a model for a resistor.

- For a capacitor (assuming zero initial conditions):

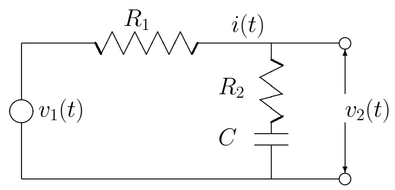
$$i(t) = C\dot{v}(t) \; \Rightarrow \; V(s) = \frac{1}{sC}\,I(s)$$

- For an inductor (assuming zero initial conditions):

$$v(t) = L\frac{di(t)}{dt} \; \Rightarrow \; V(s) = sL\,I(s)$$

- The **impedance** $Z(s)$ of an electrical component is the ratio of the Laplace transform of the voltage (across the component) to the Laplace transform of the current (through the component), with all initial conditions assumed to be zero:
  - For a resistor: $Z(s) = R$
  - For an inductor: $Z(s) = sL$
  - For a capacitor: $Z(s) = \frac{1}{sC}$

- A more complex example is provided by the following RC circuit:

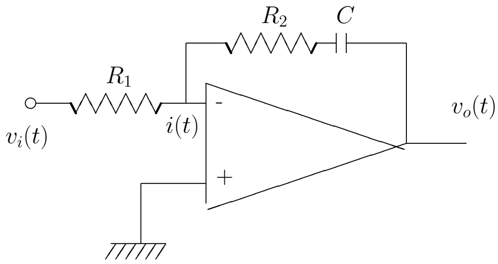


$$V_1(s) = I(s)(R_1 + R_2 + \frac{1}{sC})$$

$$V_2(s) = I(s)(R_2 + \frac{1}{sC})$$

$$\Rightarrow \quad \frac{V_2(s)}{V_1(s)} = \frac{1 + sCR_2}{1 + sC(R_1+R_2)}$$

- Another example is a circuit containing an op-amp:



- Making the 'virtual earth' assumption:

$$\begin{aligned} V_i(s) &= I(s)\ R_1 \\ V_o(s) &= -\ I(s)(R_2 + \frac{1}{sC}) \\ \Rightarrow \quad \frac{V_o(s)}{V_i(s)} &= -\ \frac{R_2 + \frac{1}{sC}}{R_1} \\ &= -\ \frac{1 + sCR_2}{sCR_1} \end{aligned}$$

Transfer Functions

- The **transfer function** $G(s)$ of a linear system is the ratio of the Laplace transform of the output to the Laplace transform of the input, with all initial conditions assumed to be zero:

- For the first example:

$$G(s) = \frac{V(s)}{I(s)} = R$$

- For the second:

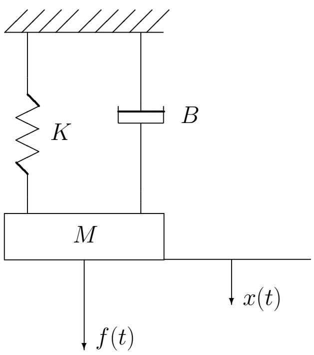
$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{1 + sCR_2}{1 + sC(R_1 + R_2)}$$

- For the third:

$$G(s) = \frac{V_o(s)}{V_i(s)} = -\ \frac{1 + sCR_2}{sCR_1}$$

Models for Mechanical Systems

Example: Assume $x(0) = \dot{x}(0) = 0$.



Balancing forces on M:

$$f(t) = Kx(t) + B\dot{x}(t) + M\ddot{x}(t)$$

Taking Laplace transform:

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{Ms^2 + Bs + K}$$

Example: Balancing forces on M_2 and M_1:

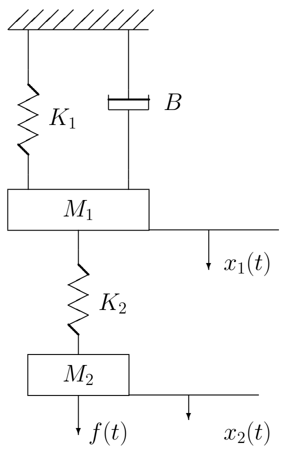
$$\begin{aligned} f(t) &= M_2\ddot{x}_2 + K_2(x_2 - x_1) \\ 0 &= M_1\ddot{x}_1 + K_2(x_1 - x_2) + K_1x_1 + B\dot{x}_1 \end{aligned}$$

Taking Laplace transforms (assume $x_i(0) = \dot{x}_i(0) = 0$)

$$\begin{aligned} (s^2M_2 + K_2)X_2(s) &= F(s) + K_2X_1(s) \\ K_2X_2(s) &= X_1(s)(s^2M_1 + Bs + K_1 + K_2) \end{aligned}$$

$$\Rightarrow \frac{X_1(s)}{K_2} = \frac{F(s) + K_2X_1(s)}{(s^2M_2 + K_2)(s^2M_1 + Bs + K_1 + K_2)}$$

$$\Rightarrow \frac{X_1(s)}{F(s)} = \frac{K_2}{(s^2M_2 + K_2)(s^2M_1 + Bs + K_1 + K_2) - K_2^2}$$



APPENDIX A

Laplace Transform Pairs

TABLE A.1

$F(s)$	$f(t),\ t \ge 0$
1. 1	$\delta(t_0)$, unit impulse at $t = t_0$
2. $1/s$	1, unit step
3. $\dfrac{n!}{s^{n+1}}$	t^n
4. $\dfrac{1}{(s+a)}$	e^{-at}
5. $\dfrac{1}{(s+a)^n}$	$\dfrac{1}{(n-1)!}t^{n-1}e^{-at}$
6. $\dfrac{a}{s(s+a)}$	$1 - e^{-at}$
7. $\dfrac{1}{(s+a)(s+b)}$	$\dfrac{1}{(b-a)}(e^{-at} - e^{-bt})$
8. $\dfrac{s+\alpha}{(s+a)(s+b)}$	$\dfrac{1}{(b-a)}[(\alpha - a)e^{-at} - (\alpha - b)e^{-bt}]$
9. $\dfrac{ab}{s(s+a)(s+b)}$	$1 - \dfrac{b}{(b-a)}e^{-at} + \dfrac{a}{(b-a)}e^{-bt}$
10. $\dfrac{1}{(s+a)(s+b)(s+c)}$	$\dfrac{e^{-at}}{(b-a)(c-a)} + \dfrac{e^{-bt}}{(c-a)(a-b)} + \dfrac{e^{-ct}}{(a-c)(b-c)}$
11. $\dfrac{s+\alpha}{(s+a)(s+b)(s+c)}$	$\dfrac{(\alpha-a)e^{-at}}{(b-a)(c-a)} + \dfrac{(\alpha-b)e^{-bt}}{(c-b)(a-b)} + \dfrac{(\alpha-c)e^{-ct}}{(a-c)(b-c)}$
12. $\dfrac{ab(s+\alpha)}{s(s+a)(s+b)}$	$\alpha - \dfrac{b(\alpha-a)}{(b-a)}e^{-at} + \dfrac{a(\alpha-b)}{(b-a)}e^{-bt}$
13. $\dfrac{\omega}{s^2+\omega^2}$	$\sin \omega t$
14. $\dfrac{s}{s^2+\omega^2}$	$\cos \omega t$
15. $\dfrac{s+\alpha}{s^2+\omega^2}$	$\dfrac{\sqrt{\alpha^2+\omega^2}}{\omega}\sin(\omega t + \phi),\ \phi = \tan^{-1}\omega/\alpha$

(continued)

TABLE A.1 *Continued*

$F(s)$	$f(t),\ t \ge 0$
16. $\dfrac{\omega}{(s+a)^2+\omega^2}$	$e^{-at}\sin \omega t$
17. $\dfrac{(s+a)}{(s+a)^2+\omega^2}$	$e^{-at}\cos \omega t$
18. $\dfrac{s+\alpha}{(s+a)^2+\omega^2}$	$\dfrac{1}{\omega}[(\alpha-a)^2+\omega^2]^{1/2}e^{-at}\sin(\omega t+\phi)$, $\phi = \tan^{-1}\dfrac{\omega}{\alpha - a}$
19. $\dfrac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\dfrac{\omega_n}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin \omega_n\sqrt{1-\zeta^2}\,t,\quad \zeta < 1$
20. $\dfrac{1}{s[(s+a)^2+\omega^2]}$	$\dfrac{1}{a^2+\omega^2} + \dfrac{1}{\omega\sqrt{a^2+\omega^2}}e^{-at}\sin(\omega t - \phi)$, $\phi = \tan^{-1}\omega/-a$
21. $\dfrac{\omega_n^2}{s(s^2+2\zeta\omega_n s+\omega_n^2)}$	$1 - \dfrac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}\,t+\phi)$, $\phi = \cos^{-1}\zeta,\ \zeta < 1$
22. $\dfrac{(s+\alpha)}{s[(s+a)^2+\omega^2]}$	$\dfrac{\alpha}{a^2+\omega^2} + \dfrac{1}{\omega}\left[\dfrac{(\alpha-a)^2+\omega^2}{a^2+\omega^2}\right]^{1/2}e^{-at}\sin(\omega t+\phi)$, $\phi = \tan^{-1}\dfrac{\omega}{\alpha-a} - \tan^{-1}\dfrac{\omega}{-a}$
23. $\dfrac{1}{(s+c)[(s+a)^2+\omega^2]}$	$\dfrac{e^{-ct}}{(c-a)^2+\omega^2} + \dfrac{e^{-at}\sin(\omega t+\phi)}{\omega[(c-a)^2+\omega^2]^{1/2}},\ \phi = \tan^{-1}\dfrac{\omega}{c-a}$

SUPPLEMENTARY NOTES

(*NOT EXAMINABLE*)

Block Diagram Algebra and Signal Flow Graphs

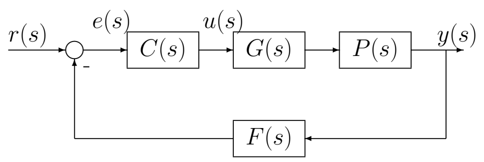
- Introduction
- Prototype Feedback System
- Block Diagram Algebra
- Block Diagram Transformations
- Examples
- Signal Flow Graphs
- Mason's Rule
- Examples

Introduction

- Many practical control systems consist of a complicated interconnection of smaller subsystems.
- Before tackling a control system design for such systems, it is usually helpful to simplify the complex interconnection of subsystems.
- Essentially, we seek a systematic way of eliminating variables (signals) we do not want to control or measure.
- We discuss two methods of performing these simplifications:
  - Block diagram algebra.
  - Signal flow graphs.

Prototype Feedback System

- Usually, we will want to reduce the interconnection to a diagram of the form:



- In this diagram:

$$\begin{aligned} G(s) &= \text{system or plant} \\ P(s) &= \text{post} - \text{compensator} \\ C(s) &= \text{pre} - \text{compensator} \\ F(s) &= \text{feedback compensator} \\ r(s) &= \text{reference signal} \\ e(s) &= \text{tracking error signal} \\ u(s) &= \text{input (or actuator) signal} \\ y(s) &= \text{output signal} \end{aligned}$$

- We will often make use of the following definitions which refer to this diagram:
  - **Loop gain function**: $C(s)G(s)P(s)F(s)$
  - **Closed loop transfer function**: $y(s)/r(s)$
  - Tracking error ratio (or **sensitivity function**):

$$e(s)/r(s)$$

- From the diagram

$$e(s) = r(s) - F(s)P(s)G(s)C(s)e(s)$$

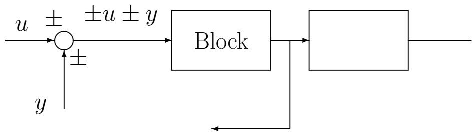
So, the sensitivity function is given by

$$\frac{e(s)}{r(s)} = \frac{1}{1 + F(s)P(s)G(s)C(s)}$$

- It follows that the closed loop transfer function is

$$\frac{y(s)}{r(s)} = \frac{C(s)G(s)P(s)}{1 + C(s)G(s)P(s)F(s)}$$

Block Diagram Algebra



- Block diagrams consist of:

  1. **Blocks**: these give a description of subsystem dynamics.

  2. **Summers**: add or subtract two or more signals.

  3. **Arrows**: these give the direction of signal propagation.

  4. **Take off points**.

Block Diagram Transformations

- The reduction of complex block diagrams is facilitated by a series of easily derivable transformations which are summarised in diagrams 1 and 2 below.

- The following steps may be used to simplify complicated block diagrams:

  – Combine cascaded blocks using 1.

  – Combine parallel blocks using 2.

  – Eliminate minor loops using 4.

  – Shift summers left using 7 and shift take off points right using 10 and 12.

  – Go around the block again if necessary.

- General methods best learned from examples.

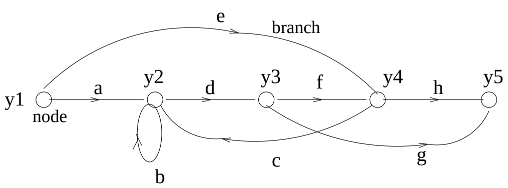
Signal Flow Graphs

- A **signal flow graph** is a pictorial representation of a set of simultaneous equations describing a system.

- Equations are represented with the aid of **branches** and **nodes**:
  – nodes represent variables.
  – branches relate variables.

- **Example:** The signal flow graph below is a pictorial representation of:

$$\begin{aligned} y_2 &= ay_1 + by_2 + cy_4 \\ y_3 &= dy_2 \\ y_4 &= ey_1 + fy_3 \\ y_5 &= hy_4 + gy_3 \end{aligned}$$



- Before proceeding further, we will define some terms which we will need later:

  – A **source** is a node which has outgoing branches only (y_1).

  – A **sink** only has incoming branches (y_5).

  – A **path** is a set of branches having the same sense of direction ($adfh$, eh and $adfc$).

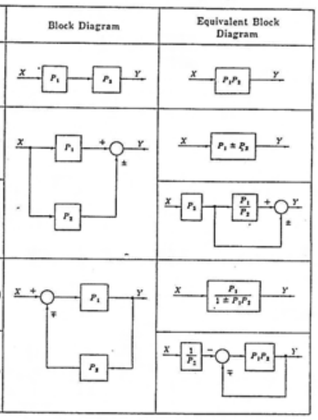
  – A **forward path** originates from a source and terminates in a sink. No node may be encountered more than once (eh, $ecdg$ and $adfh$).

  – The **path gain** is the product of the coefficients associated with the branches of the path.

  – A **feedback loop** is a path that begins and ends at the same node; additionally no node may be encountered more than once (b and dfc).

  – The **loop gain** is the path gain of a feedback loop.

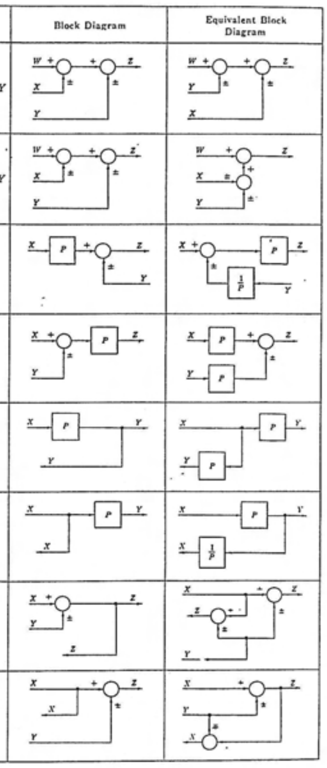
(7)



	Transformation	Equation
1	Combining Blocks in Cascade	$Y = (P_1P_2)X$
2	Combining Blocks in Parallel; or Eliminating a Forward Loop	$Y = P_1X \pm P_2X$
3	Removing a Block from a Forward Path	$Y = P_1X \pm P_2X$
4	Eliminating a Feedback Loop	$Y = P_1(X \mp P_2Y)$
5	Removing a Block from a Feedback Loop	$Y = P_1(X \mp P_2Y)$

Diagram 1

(8)

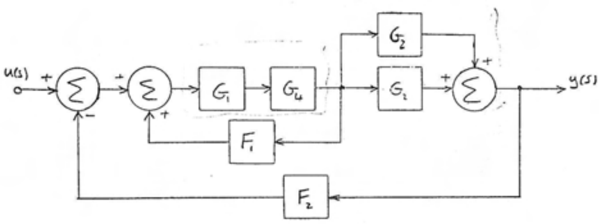


	Transformation	Equation
6a	Rearranging Summing Points	$Z = W \pm X \pm Y$
6b	Rearranging Summing Points	$Z = W \pm X \pm Y$
7	Moving a Summing Point Ahead of a Block	$Z = PX \pm Y$
8	Moving a Summing Point Beyond a Block	$Z = P[X \pm Y]$
9	Moving a Takeoff Point Ahead of a Block	$Y = PX$
10	Moving a Takeoff Point Beyond a Block	$Y = PX$
11	Moving a Takeoff Point Ahead of a Summing Point	$Z = X \pm Y$
12	Moving a Takeoff Point Beyond a Summing Point	$Z = X \pm Y$

Diagram 2

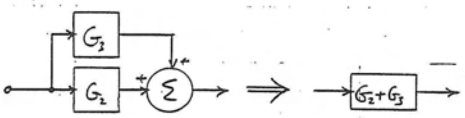
Example (9)

Reduce the following block diagram to simple feedback form.



Step 1:

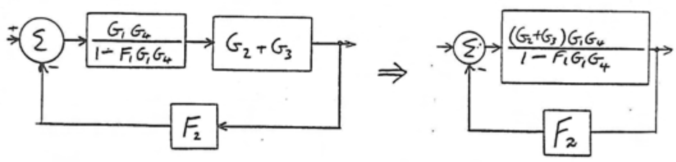
Step 2:



Step 3: (10)

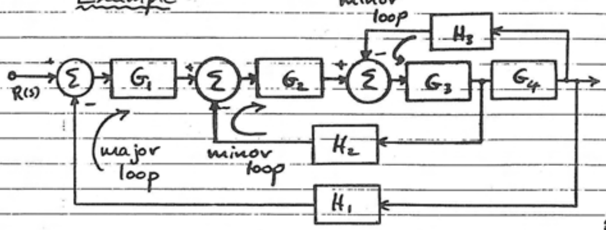
Step 4: Does not apply

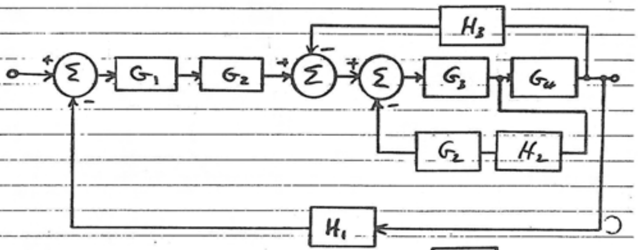
Step 5:

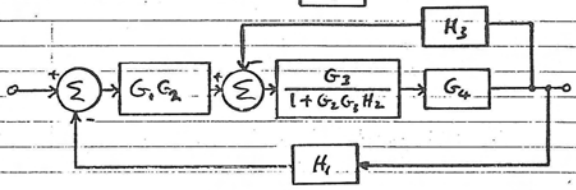


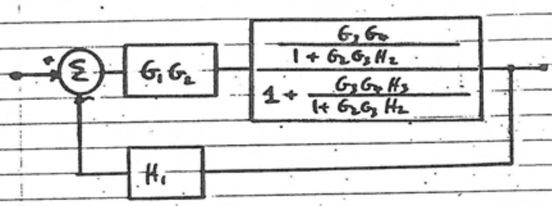
Signal flow graph

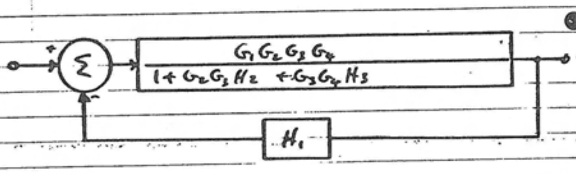
Example

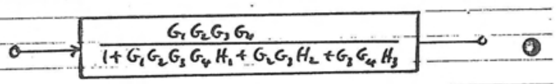

minor loop
H_3
$R(s)$
Σ
G_1
G_2
G_3
G_4
H_2
major loop
minor loop
H_1


H_3
G_1
G_2
G_3
G_4
G_2
H_2
H_1


H_3
G_1G_2
$\frac{G_3}{1+G_2G_3H_2}$
G_4
H_1


G_1G_2
$\frac{\frac{G_3G_4}{1+G_2G_3H_2}}{1+\frac{G_3G_4H_3}{1+G_2G_3H_2}}$
H_1


$\frac{G_1G_2G_3G_4}{1+G_2G_3H_2+G_3G_4H_3}$
H_1


$\frac{G_1G_2G_3G_4}{1+G_1G_2G_3G_4H_1+G_2G_3H_2+G_3G_4H_3}$

Mason's Rule

The general formula for the path gain between the single source and the single sink is:

$$G = \frac{\Sigma_k P_k \Delta_k}{\Delta}$$

where:

$$\Delta = 1 - \Sigma L_1 + \Sigma L_2 - \Sigma L_3 + ... + (-1)^m \Sigma L_m$$

and

$L_1 =$ gain of each closed loop in the graph

$L_2 =$ product of loop gains of any two nontouching loops. (loops are called nontouching if they have no node on common)

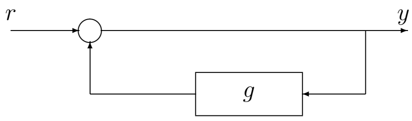
$\vdots$

$L_m =$ product of loop gains of any m nontouching loops

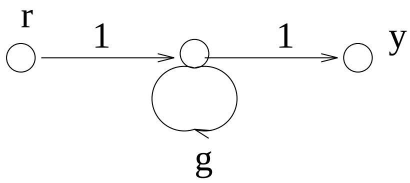
$P_k =$ gain of the k^{th} forward path

$\Delta_k =$ the value of Δ remaining with the loops touching the path P_k are removed

Example

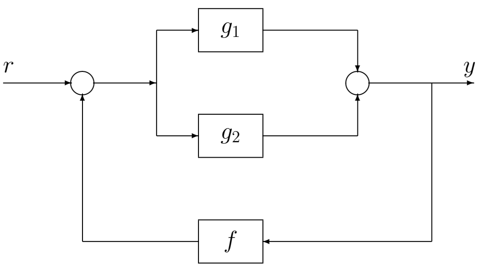


$$y = r + gy$$
$$\frac{y}{r} = \frac{1}{1-g}$$

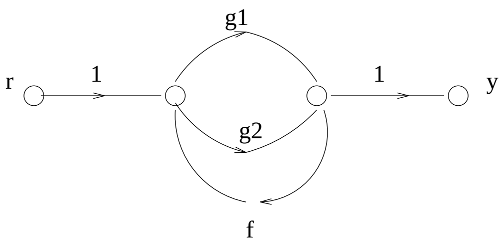


$$\Delta = 1 - g$$
$$P_1 = 1$$
$$\Delta_1 = 1$$
$$\frac{y}{r} = \frac{1}{1-g}$$

Example



$$\frac{y}{r} = \frac{g_1 + g_2}{1 - f(g_1 + g_2)}$$

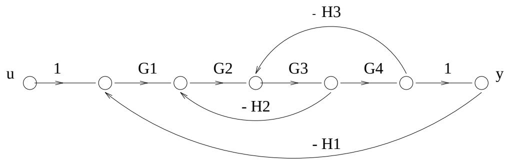


$$P_1 = g_1, \quad \Delta_1 = 1$$

$$P_2 = g_2, \quad \Delta_2 = 1$$

$$\Delta = 1 - g_1 f - g_2 f \quad \Rightarrow \frac{y}{r} = \frac{g_1 + g_2}{1 - f(g_1 + g_2)}$$

Example



- Forward paths:

$$P_1 = G_1 G_2 G_3 G_4.$$

- Closed loops:

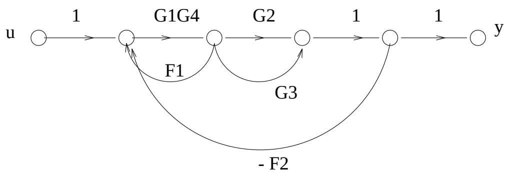
$$-G_2 G_3 H_2, \quad -G_3 G_4 H_3, \quad -G_1 G_2 G_3 G_4 H_1$$

- $L_1 = -(G_2 G_3 H_2 + G_3 G_4 H_3 + G_1 G_2 G_3 G_4 H_1)$
- Non-touch loops: None

- $\Delta = 1 + G_2 G_3 H_2 + G_3 G_4 H_3 + G_1 G_2 G_3 G_4 H_1$
- $\Delta_1 = 1$

$$\Rightarrow \; G = \frac{y}{u} = \frac{G_1 G_2 G_3 G_4}{1 + G_2 G_3 H_2 + G_3 G_4 H_3 + G_1 G_2 G_3 G_4 H_1}$$

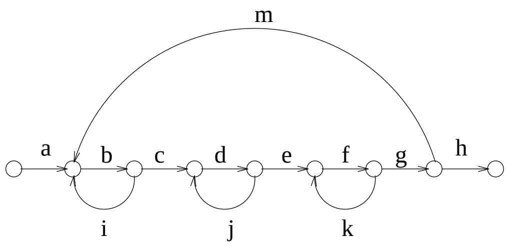
Example



- Forward paths: $P_1 = G_1 G_2 G_4$, $P_2 = G_1 G_3 G_4$
- Closed loops: $G_1 G_4 F_1, -G_1 G_2 G_4 F_2, -G_1 G_3 G_4 F_2$
- $L_1 = G_1 G_4 F_1 - G_1 G_4 F_2 (G_2 + G_3)$
- $\Delta = 1 - G_1 G_4 F_1 + G_1 G_4 F_2 (G_2 + G_3)$
- $\Delta_1 = \Delta_2 = 1$

$$\Rightarrow \; G = \frac{y}{u} = \frac{G_1 G_4 (G_2 + G_3)}{1 - G_1 G_4 F_1 + G_1 G_4 F_2 (G_2 + G_3)}$$

Example



$$\begin{aligned} \Delta = \; & 1 - (bi + dj + fk + bcdefgm) + \\ & (bidj + bifk + djfk) - (bidjfk) \end{aligned}$$

$$P_1 = abcdefgh$$

$$\Delta_1 = 1$$

$$G = \frac{abcdefgh}{1 - bi - dj - fk - bcdefgm + bidj + bifk + djfk - bidjfk}$$

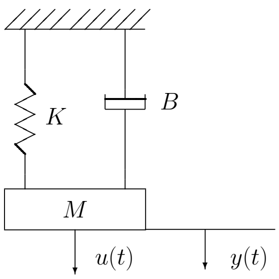
State-variable Models

- Introduction
- Standard State-variable Format
- Advantages of State-variable Models
- Simulation Diagrams
- State-variable Models from Transfer Functions
- Transfer Functions from State-variable Models
- The Matrix Exponential Function
- Solving the State-variable Equations
- Stability and the Matrix Exponential
- Transfer Functions from State-variable Models
- Computing the Matrix Exponential Function

Introduction

- We have already come across:
  - **linear differential equation models**
  - **transfer function models**
- We now consider the **state-variable model**, also called the **state-space model**.
- This is a differential equation model, but the equation is written in a specific format.
- The idea is that an nth order differential equation is decomposed into a set of n 1st order equations written in matrix-vector form.
- This decomposition is achieved by defining internal variables, called **states**, whose time evolution completely characterises the behaviour of the system.

Example



$$M\ddot{y}(t) + B\dot{y}(t) + Ky(t) = u(t)$$

- Define $x_1(t) = y(t)$, $\quad x_2(t) = \dot{y}(t)$.

$$\Rightarrow \quad \dot{x}_1(t) \;=\; x_2(t)$$

$$\Rightarrow \quad \dot{x}_2(t) \;=\; -\frac{K}{M}x_1(t) - \frac{B}{M}x_2(t) + \frac{u(t)}{M}$$

or in matrix form:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{K}{M} & -\frac{B}{M} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{M} \end{bmatrix} u(t)$$

$$y(t) \;=\; \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

- This is an example of a **single input/single output** state-variable model.

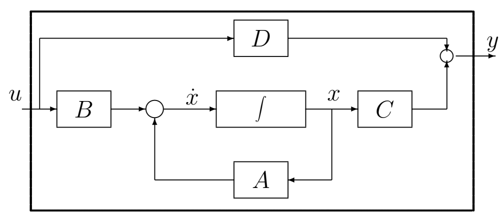
Standard State-variable Format

- **The state of a system at time t_1 is the amount of information at t_1 that, together with all inputs for $t \geq t_1$, uniquely determines the system behaviour for all $t \geq t_1$.**
- State-variable models have the form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) && (\textbf{state equation}) \\ y(t) &= Cx(t) + Du(t) && (\textbf{output equation}) \\ x(0) &= x_0 && (\textbf{initial condition}) \end{aligned}$$

where

$$\begin{aligned} A &: n \times n && (\textbf{system matrix}) \\ B &: n \times r && (\textbf{input matrix}) \\ C &: p \times n && (\textbf{output matrix}) \\ D &: p \times r && (\textbf{direct feedthrough matrix}) \\ x(t) &: n \times 1 && (\textbf{state vector}) \\ u(t) &: r \times 1 && (\textbf{input vector}) \\ y(t) &: p \times 1 && (\textbf{output vector}) \end{aligned}$$



Example

- Consider the coupled differential equations:

$$\begin{aligned}\ddot{y}_1 + k_1\dot{y}_1 + k_2 y_1 &= u_1 + k_3 u_2\\ \dot{y}_2 + k_4 y_2 + k_5\dot{y}_1 &= k_6 u_1\end{aligned}$$

- Define: $x_1 = y_1,\ x_2 = \dot{y}_1,\ x_3 = y_2$

- This gives the state equations:

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= -k_2 x_1 - k_1 x_2 + u_1 + k_3 u_2\\ \dot{x}_3 &= -k_5 x_2 - k_4 x_3 + k_6 u_1\end{aligned}$$

  and the output equations:

$$y_1 = x_1, \qquad y_2 = x_3$$

- Collecting these into matrix form gives the 2-input 2-output state-variable model

$$\overbrace{\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}}^{\dot{x}} = \overbrace{\begin{bmatrix}0 & 1 & 0\\ -k_2 & -k_1 & 0\\ 0 & -k_5 & -k_4\end{bmatrix}}^{A}\overbrace{\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}}^{x} + \overbrace{\begin{bmatrix}0 & 0\\ 1 & k_3\\ k_6 & 0\end{bmatrix}}^{B}\overbrace{\begin{bmatrix}u_1\\ u_2\end{bmatrix}}^{u}$$

$$\underbrace{\begin{bmatrix}y_1\\ y_2\end{bmatrix}}_{y} = \underbrace{\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}}_{C}\underbrace{\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}}_{x}$$

Advantages of State-variable Models

- State-variable models give knowledge about the **internal structure** as well as the input-output characteristics of the system.

- There are **computational** advantages:
  - **time-domain** matrix methods lend themselves naturally to computer solution, especially for high order models.
  - matrix methods enable us to easily determine the **transient response** and evaluate the system performance.

- A good unified framework for several advanced control theories, such as **optimal control** design methods.

- A natural framework for system **simulation**.

- Extensions to:
  - **multivariable** systems
  - **nonlinear** systems
  - **time-varying** systems

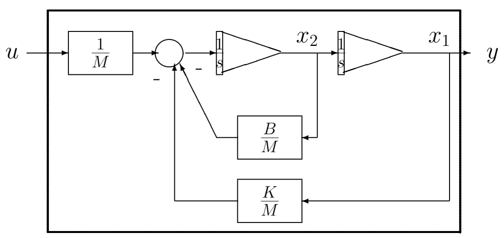
  are (relatively) straightforward.

Simulation Diagrams

- These are block diagrams (or signal flow graphs) that are constructed to have a given transfer function or to model a set of differential equations.
- There are three elements in a simulation diagram:
  - an **integrator** (whose transfer function is $\frac{1}{s}$)
  - a pure **gain**
  - a **summer**

  All these components can be easily constructed using simple electronic devices.
- They are useful in constructing computer simulations (digital or analogue) of a given system.
- For the mass-spring system we have:

$$\begin{aligned}\dot{x}_1(t) &= x_2(t)\\ \dot{x}_2(t) &= -\frac{K}{M}x_1(t) - \frac{B}{M}x_2(t) + \frac{u(t)}{M}\end{aligned}$$



State-variable Models from Transfer Functions

- There are simulation diagrams which can be derived from general transfer functions of the form:

$$\frac{y(s)}{u(s)} = g(s) = \frac{b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0}{s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0}$$

- Divide numerator and denominator by s^n:

$$y(s) = \frac{b_{n-1}s^{-1} + b_{n-2}s^{-2} + \cdots + b_0 s^{-n}}{1 + a_{n-1}s^{-1} + a_{n-2}s^{-2} + \cdots + a_0 s^{-n}}\ u(s)$$

- Set

$$e(s) := \frac{u(s)}{1 + a_{n-1}s^{-1} + a_{n-2}s^{-2} + \cdots + a_0 s^{-n}}$$

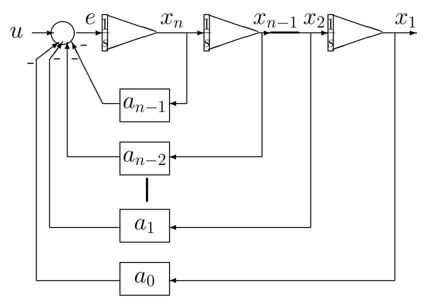
  so that

$$y(s) = \left[b_{n-1}s^{-1} + b_{n-2}s^{-2} + \cdots + b_0 s^{-n}\right] e(s)$$

- This gives

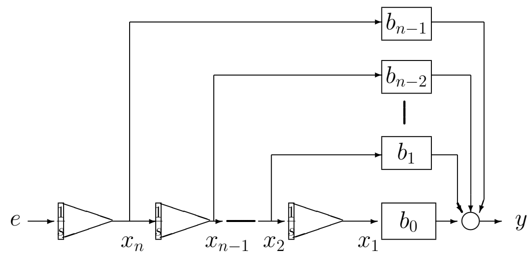
$$e(s) = u(s) - \left[a_{n-1}s^{-1} + a_{n-2}s^{-2} + \cdots + a_0 s^{-n}\right] e(s)$$

- This has the following simulation diagram:



- Complete the diagram by setting

$$y(s) = [b_{n-1}s^{-1} + b_{n-2}s^{-2} + \cdots + b_0 s^{-n}]\, e(s)$$



- The state variables satisfy:

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{n-1} &= x_n \\
\dot{x}_n &= -a_0x_1 - a_1x_2 - \cdots - a_{n-2}x_{n-1} \\
&\quad -a_{n-1}x_n + u
\end{aligned}$$

while the output is

$$y = b_0x_1 + b_1x_2 + \cdots + b_{n-2}x_{n-1} + b_{n-1}x_n$$

- In matrix form this yields the state-variable model:

$$\begin{aligned}
\dot{x} &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u \\
y &= \begin{bmatrix} b_0 & b_1 & \cdots & b_{n-2} & b_{n-1} \end{bmatrix} x
\end{aligned}$$

- Note the direct connection with the coefficients of the transfer function:

$$g(s) = \frac{b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_0}{s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0}$$

The Matrix Exponential Function

- Suppose that A is a square matrix and let t denote the time variable and I denote the **identity matrix**.

- Define the **matrix exponential function** e^{At} via the power series

$$e^{At} = I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \cdots + \frac{A^kt^k}{k!} + \cdots$$

- It can be shown this series converges for any square matrix A and any scalar t.

- Some, but not all, of the properties of the scalar exponential function are inherited by the matrix exponential function.

- A most useful property concerns the derivative:

$$\begin{aligned}
\frac{d}{dt}e^{At} &= A + A^2t + \frac{A^3t^2}{2!} + \cdots \\
&= A(I + At + \frac{A^2t^2}{2!} + \cdots) \\
&= Ae^{At} = e^{At}A
\end{aligned}$$

- The most useful properties are summarised below:

  - P1: $e^0 = I$
  - P2: $e^{T\Lambda T^{-1}} = Te^{\Lambda}T^{-1}$ for any nonsingular T
  - P3: $e^{(\alpha+\beta)A} = e^{\alpha A}e^{\beta A}$
  - P4: $e^{-A} = (e^A)^{-1}$
  - P5: $\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A$
  - P6: $\mathcal{L}(e^{At}) = (sI - A)^{-1}$

Solving the State-variable Equations

- Consider the state-variable model:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \qquad x(0) = x_0 \\ y(t) &= Cx(t) + Du(t)\end{aligned}$$

- Suppose first that $u(t) = 0$ for all t:

$$\dot{x}(t) = Ax(t), \qquad x(0) = x_0$$

Trial solution: $x(t) = e^{At}x_0$. Check:

$$\Rightarrow \quad \dot{x}(t) = A \underbrace{e^{At}x_0}_{x(t)} = Ax(t), \quad x(0) = \underbrace{e^0}_{I} x_0 = x_0$$

- When $u(t) \neq 0$ we proceed as follows:

$$\begin{aligned} e^{-At}\dot{x}(t) &= e^{-At}Ax(t) + e^{-At}Bu(t) \\ \text{(P5)} \Rightarrow \quad \frac{d}{dt}[e^{-At}x(t)] &= e^{-At}Bu(t) \\ \Rightarrow \quad \int_0^t \frac{d}{dt}[e^{-At}x(t)]dt &= \int_0^t e^{-A\tau}Bu(\tau)d\tau \\ \text{(P1)} \Rightarrow \quad e^{-At}x(t) - x_0 &= \int_0^t e^{-A\tau}Bu(\tau)d\tau \\ \text{(P4, P3)} \Rightarrow \quad x(t) &= e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \end{aligned}$$

- Finally

$$y(t) = \underbrace{Ce^{At}x_0}_{\substack{\text{initial cond.}\\ \text{response}}} + \overbrace{\underbrace{\int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau}_{\text{convolution integral}} + Du(t)}^{\text{input response}}$$

Stability and the Matrix Exponential

- Consider the eigenvalue decomposition of A:

$$\begin{aligned} A &= T\Lambda T^{-1} \\ \Lambda &= \text{diag}\underbrace{(\lambda_1, \lambda_2, \cdots, \lambda_n)}_{\text{eigenvalues of } A} \end{aligned}$$

- Using property P2:

$$\begin{aligned} e^{At} &= e^{T\Lambda T^{-1}t} \\ &= I + T\Lambda T^{-1}t + \frac{T\Lambda T^{-1}T\Lambda T^{-1}t^2}{2!} + \cdots \\ &= T[I + \Lambda t + \frac{\Lambda^2 t^2}{2!} + \cdots]T^{-1} = Te^{\Lambda t}T^{-1} \\ &= T \text{ diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \cdots, e^{\lambda_n t})\, T^{-1} \end{aligned}$$

since Λ is a diagonal matrix.

- Consider the response of the **unforced system**

$$\begin{aligned} \dot{x}(t) &= Ax(t), \qquad x(0) = x_0 \\ \Rightarrow \quad x(t) &= e^{At}x_0 = Te^{\Lambda t}T^{-1}x_0 \end{aligned}$$

- It follows that

$$\lim_{t\to\infty} x(t) = 0, \forall\, x_0 \quad \Longleftrightarrow \quad \text{Re}[\lambda_i(A)] < 0, \forall\, i$$

The unforced system is stable if and only if all the eigenvalues of A lie in the left half of the complex plane

Transfer Functions from State-variable Models

- Consider a state-variable model initially at rest:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \qquad x(0) = 0 \\ y(t) &= Cx(t) + Du(t)\end{aligned}$$

- Taking Laplace transforms:

$$\begin{aligned} & sx(s) = Ax(s) + Bu(s) \\ \Rightarrow \; & (sI - A)x(s) = Bu(s) \\ \Rightarrow \; & x(s) = (sI - A)^{-1}Bu(s) \end{aligned}$$

and so

$$\begin{aligned} & y(s) = Cx(s) + Du(s) \\ \Rightarrow \; & y(s) = [D + C(sI - A)^{-1}B]u(s) \end{aligned}$$

- The transfer function from $u(s)$ to $y(s)$ is then

$$g(s) = [D + C(sI - A)^{-1}B]$$

- Note that $(sI - A)^{-1}$ will exist for all values of s which are *not* eigenvalues of A. We call $\lambda_i(A)$ the **poles** of $g(s)$

Computing the Matrix Exponential Function

1. **Truncation** of power series expansion:

$$e^{At} \simeq I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \cdots + \frac{A^kt^k}{k!}$$

   - Usually, only a few terms are needed

2. Eigenvalue-eigenvector decomposition:
   - Let

$$\begin{aligned} A &= T\Lambda T^{-1} = T\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)T^{-1} \\ \Rightarrow e^{At} &= Te^{\Lambda t}T^{-1} = T\text{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, \cdots, e^{\lambda_n t})T^{-1} \end{aligned}$$

   - The eigenvalue matrix Λ is not always diagonal
   - Method not always accurate

3. Inverse Laplace transform:
   - Since $\mathcal{L}\{e^{At}\} = (sI - A)^{-1}$

$$\Rightarrow \quad e^{At} = \mathcal{L}^{-1}\{(sI - A)^{-1}\}$$

   - Feasible only for matrices of small dimension
   - Almost never used in practice

4. Matlab: Just type $E = expm(A * t)$

System Responses

- Introduction
- Test Signals
- Response of First Order Systems
- Steady-state Response
- Step Response of Second Order Systems
- Time Response Specifications
- Interpretation of Pole Locations
- Step Response for Higher Order Systems
- Introduction to Frequency Response
- Steady-state Accuracy of Feedback Systems

Introduction

- For control system design, we need a basis of **comparison of performance** of different designs.
- One way of setting up this basis is to specify particular **test signals** and compare the response of different designs to these signals.
- The use of test signals is justified since there is a close correlation between the capability of systems to respond to actual input signals and the response characteristics to test signals.
- We analyse the **transient response** as well as the **steady-state response** of simple systems.
- Although practical control systems are generally of **high order**, we will be mainly concerned with **first & second order** systems, since these capture the main features of practical control systems.

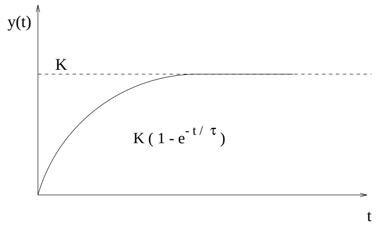
Test Signals

- All the following test signals are defined for $t \geq 0$.
- The **unit impulse function** is based on a rectangular function $\delta_\epsilon(t)$ such that
$$\delta_\epsilon(t) = \epsilon^{-1}, \quad 0 \leq t \leq \epsilon, \quad \epsilon > 0$$
As $\epsilon \to 0$, $\delta_\epsilon(t)$ approaches the unit impulse $\delta(t)$:
$$\int_0^\infty \delta(t)dt = 1, \quad \int_0^\infty \delta(t-a)g(t)dt = g(a), \quad \mathcal{L}\{\delta(t)\} = 1$$
It is useful for modeling 'shock' inputs.
- The **unit step function:**
$$r(t) = 1, \ \ t \geq 0 \quad \Rightarrow \quad \mathcal{L}\{r(t)\} \ = \ \frac{1}{s}$$
is useful for modeling sudden disturbances.
- The **unit ramp function:**
$$r(t) = t, \ \ t \geq 0 \quad \Rightarrow \quad \mathcal{L}\{r(t)\} \ = \ \frac{1}{s^2}$$
is useful for modeling gradually changing inputs.
- The **sinusoidal functions**
$$\cos \omega t \ = \ \text{Re } e^{j\omega t} \quad \Rightarrow \quad \mathcal{L}\{\cos \omega t\} = \frac{s}{s^2 + \omega^2}$$
$$\sin \omega t \ = \ \text{Im } e^{j\omega t} \quad \Rightarrow \quad \mathcal{L}\{\sin \omega t\} = \frac{\omega}{s^2 + \omega^2}$$
are important in frequency response techniques.

Response of First Order Systems

- Consider the first order transfer function
$$g(s) = \frac{y(s)}{r(s)} = \frac{K}{s\tau + 1}$$
- The differential equation representation is
$$\tau \dot{y}(t) + y(t) = Kr(t)$$
- When $r(t)$ is a unit step
$$y(s) \ = \ \frac{K}{s(s\tau + 1)} = K\left(\frac{1}{s} - \frac{1}{s + 1/\tau}\right)$$
$$\Rightarrow \quad y(t) \ = \ K(1 - e^{-t/\tau}) = K - Ke^{-t/\tau}$$
- If $\tau > 0$, the second term tends to 0 as $t \to \infty$.
- The first term is called the **steady-state response** and K is called the **steady-state value**.
- The second term is called the **transient response** and τ is called the **time constant** of the system.

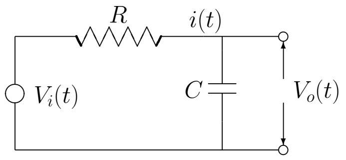
- The following figure shows the step response.



- The exponential function decays to less than 2% of its initial value within 4 time constants.

- The output $y(t)$ reaches about 63% of its final value when $t = \tau$.

- An example of a first order system is provided by the following RC circuit.



$$\frac{V_0(s)}{V_i(s)} = \frac{1}{1+sCR} \qquad (K = 1, \quad \tau = RC)$$

Steady-state Response

- The concept of the steady-state response can be generalized to transfer functions of any order.

- Suppose that

$$y(s) = g(s)r(s)$$

where $g(s)$ is a given transfer function.

- Suppose that the limit

$$\lim_{t \to \infty} y(t)$$

exists.

- Using the **final value theorem**:

$$\lim_{t \to \infty} y(t) = \lim_{s \to 0} sy(s) \;=\; \lim_{s \to 0} sg(s)r(s)$$

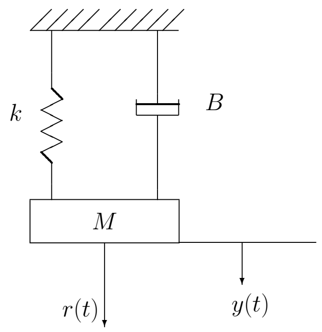
- If $r(t)$ is a unit step, this becomes:

$$\begin{aligned}\lim_{t\to\infty} y(t) &= \lim_{s\to 0} sg(s)\frac{1}{s} \;=\; \lim_{s\to 0} g(s) \\ &= g(0)\end{aligned}$$

- $g(0)$ is often called the **d.c. gain** of the system.

Response of Second Order Systems

- Consider the damped mass spring system again



- The transfer function can be written as

$$\begin{aligned} g(s) = \frac{y(s)}{r(s)} &= \frac{1}{Ms^2 + Bs + k} \\ &= \frac{\frac{1}{M}}{s^2 + \frac{B}{M}s + \frac{k}{M}} \\ &= \frac{1}{k}\,\frac{\frac{k}{M}}{s^2 + \frac{B}{M}s + \frac{k}{M}} \\ &= K\,\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \end{aligned}$$

where

$$K = \frac{1}{k}, \quad \omega_n = \sqrt{\frac{k}{M}}, \quad \zeta = \frac{B}{2\sqrt{kM}}$$

- The **standard form** for second order systems is:

$$g(s) = K\,\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

where

ζ : **damping ratio**
ω_n : **undamped natural frequency**
K : **d.c. gain**

- Since K only determines the d.c. magnitude of the response, we will set

$$K = 1$$

and so

$$\begin{aligned} g(s) &= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{\omega_n^2}{(s-p_1)(s-p_2)} \end{aligned}$$

where

$$p_1, p_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

are the **poles** of $g(s)$.

Pole Locations

- The system characteristics depend on ζ and ω_n (equivalently p_1, p_2), since they are the only parameters that appear in the transfer function.

- There are three cases of practical interest:

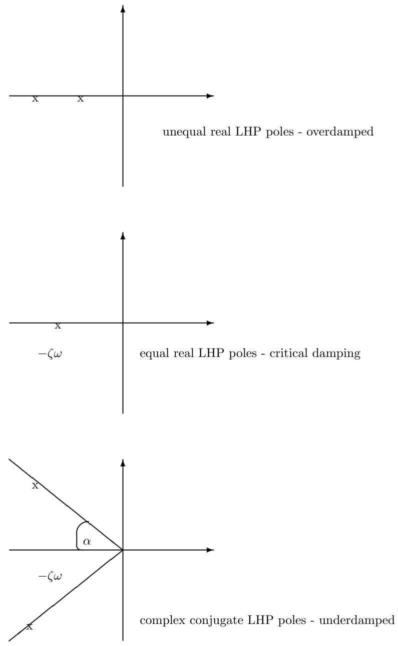
  1. $\zeta > 1$ **(overdamped system)**:
  $$p_1, p_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$
  are real, negative and unequal.

  2. $\zeta = 1$ **(critically damped system)**:
  $$p_1, p_2 = -\zeta\omega_n$$
  are real, negative and equal.

  3. $0 \leq \zeta < 1$ **(underdamped system)**:
  $$p_1, p_2 = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$$
  are complex conjugate and have negative real parts (zero real part if $\zeta = 0$).

- Another case is $\zeta < 0$ **(unstable system)**:
$$p_1, p_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$
have positive real parts will be covered later.

- The three cases are illustrated in the figure below.



$$\alpha = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}$$

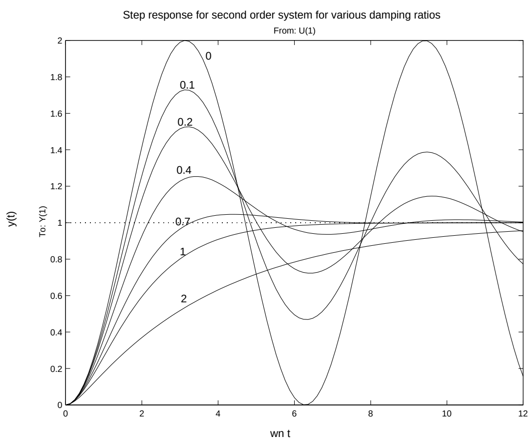
Step Response of Second Order Systems

- If $r(t)$ is a unit step
$$y(s) = \frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$$

- Expanding in partial fractions and completing the square
$$\begin{aligned} y(s) &= \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{1}{s} - \frac{s + \zeta\omega_n}{(s+\zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s+\zeta\omega_n)^2 + \omega_d^2} \end{aligned}$$
where
$$\omega_d = \sqrt{1-\zeta^2}\omega_n$$
is called the **damped natural frequency**.

- Taking inverse Laplace transform
$$\begin{aligned} y(t) &= 1 - e^{-\zeta\omega_n t}(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t) \\ &= 1 - \frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin(\sqrt{1-\zeta^2}\omega_n t + \alpha) \end{aligned}$$
where
$$\alpha = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}$$

- The step responses are shown for several values of ζ as a function of $\omega_n t$.

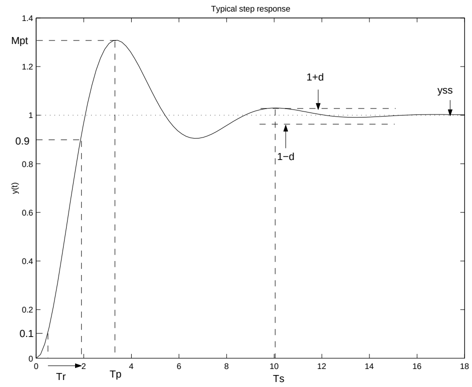


- If $\zeta = 0$, the frequency of the sinusoid is ω_n - hence *undamped natural frequency.*
- If $1 > \zeta > 0$, $\omega_d = \sqrt{1-\zeta^2}\omega_n$ is the frequency of the sinusoid - hence *damped natural frequency.*
- The time constant of the **exponential envelope** is $\tau = \frac{1}{\zeta\omega_n}$
- For $0 < \zeta < 1$ the response is a damped sinusoid - hence *damped (or underdamped) system.*
- As ζ increases from 0 to 1, the response becomes less oscillatory - hence *damping ratio.*
- For $\zeta \geq 1$, the oscillations have ceased - hence *overdamped system* ($\zeta > 1$) and *critically damped system* ($\zeta = 1$).
- When $\zeta < 0$, the response grows without limit - hence *unstable system.*

Time Response Specifications

- Some system design specifications can be described in terms of the step response.
- For underdamped second order systems, we define the following specifications:

$$\begin{aligned}
T_r &: \textbf{rise time } (10\% - 90\%)\\
M_p &: \textbf{peak value}\\
T_p &: \textbf{time to first peak}\\
y_{ss} &: \textbf{steady} - \textbf{state value}\\
\%\ \textbf{overshoot} &: \frac{M_p - y_{ss}}{y_{ss}} \times 100\\
T_s &: \textbf{settling time}
\end{aligned}$$



- T_s is the time required to enter a $y_{ss} \pm d$ tube. Common values are $2\%, 4\%$ and 5%.
- The time constant is $\tau = 1/\zeta\omega_n$ so $T_s \simeq 4\tau = 4/\zeta\omega_n$.
- For overdamped (or critically damped) systems, T_r, T_s and y_{ss} (but not M_p or T_p) are well-defined.
- With $\alpha = \tan^{-1}\sqrt{1-\zeta^2}/\zeta$, the step response for an underdamped 2nd order system is given by:

$$y(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin(\sqrt{1-\zeta^2}\omega_n t + \alpha)$$

- Differentiating $y(t)$ and equating to 0 gives

$$T_p = \frac{\pi}{\sqrt{1-\zeta^2}\omega_n}\,, \qquad M_p = 1 + e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}}$$

- Since $y_{ss} = 1$,

$$\%\ \text{overshoot} = \frac{M_p - y_{ss}}{y_{ss}} \times 100 = 100 e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}}$$

- As ζ increases from 0 to 1, M_p, T_s and the percentage overshoot decrease, while T_p and T_r increase.
- A compromise must be found between the **speed of response** (measured by T_p & T_r) and **tracking properties** (measured by T_s & % overshoot).

Interpretation of Pole Locations

- Suppose that

$$g(s) = \frac{y(s)}{r(s)} = \frac{n(s)}{d(s)} = \frac{(s-z_1)(s-z_2)\cdots(s-z_l)}{(s-p_1)(s-p_2)\cdots(s-p_m)}$$

where $n(s)$ is the **zero polynomial** and $d(s)$ is the **pole polynomial**.

- The roots of $n(s)$ $(z_1, \ldots, z_l)$ are the **zeros** of $g(s)$ while the roots of $d(s)$ $(p_1, \ldots, p_m)$ are the **poles** of $g(s)$.
- Assuming no repeated poles and expanding in partial fractions the response to $r(s)$ has the form

$$y(s) = \frac{r_1}{s-p_1} + \frac{r_2}{s-p_2} + \cdots + \frac{r_m}{s-p_m} + a(s)$$

where $a(s)$ contains terms due to $r(s)$.

$$\Rightarrow \quad y(t) = r_1 e^{p_1 t} + r_2 e^{p_2 t} + \cdots + r_m e^{p_m t} + a(t)$$

where $a(t)$ is the forced response.

- Since $n(s)$ and $d(s)$ are real, poles and zeros are either real or occur in complex conjugate pairs.
- Each real pole contributes a pure exponential, while a pair of complex conjugate poles contribute an exponentially decaying (or expanding) sinusoid.

Therefore, since $c_{ss} = 1$, the percent overshoot is, from (4-26),

$$\text{percent overshoot} = e^{-\zeta\pi/\sqrt{1-\zeta^2}} \times 100 \tag{4-32}$$

The percent overshoot is thus a function only of ζ and is plotted versus ζ in Figure 4.8.

We can express (4-29) as

$$\omega_n T_p = \frac{\pi}{\sqrt{1-\zeta^2}} \tag{4-33}$$

Thus the product $\omega_n T_p$ is also a function of only ζ, and this is also plotted in Figure 4.8. Since T_p is an approximate indication of the rise time, Figure 4.8 also roughly indicates rise time. An example is now given to indicate the usefulness of these curves.

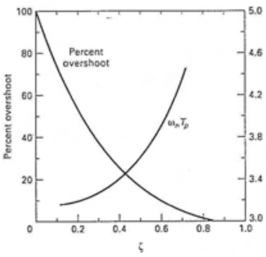


Figure 4.8 Relationship of percent overshoot, ζ, ω_n, and T_p for a second-order system.

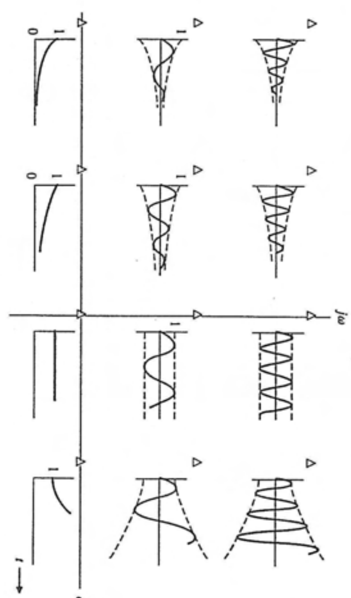


FIGURE 5.17 Impulse response for various root locations in the s-plane. (The conjugate root is not shown.)

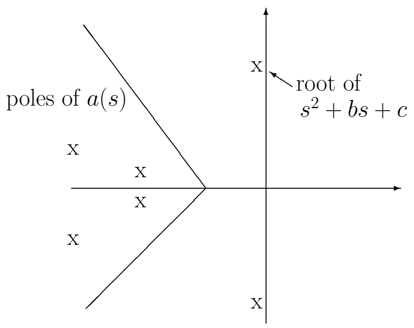
Step Response of Higher Order Systems

- Good test signal since it is easy to interpret.
- Useful if system is predominantly second order.
- A system $g(s)$ is predominantly second order if

$$g(s) = a(s) + \frac{\alpha s + \beta}{s^2 + bs + c}$$

where the poles of $a(s)$ are located far into the left half plane compared with the roots of $s^2 + bs + c$, called the **dominant poles**.

- The response associated with the poles of $a(s)$ will decay fast, and the response will be largely determined by the dominant poles.



Introduction to Frequency Response

- Let $g(s)$ be stable and suppose that

$$u(t) = e^{j\omega t} \Longrightarrow u(s) = \frac{1}{s - j\omega}$$

- Expanding in partial fractions

$$y(s) = g(s)\,\frac{1}{s - j\omega} = \frac{g(j\omega)}{s - j\omega} + a(s)$$

where $a(s)$ includes all the poles of $g(s)$.

- Taking inverse Laplace transform

$$y(t) = g(j\omega)e^{j\omega t} + a(t).$$

Since $a(s)$ is stable the steady-state response is

$$y_{ss}(t) = g(j\omega)e^{j\omega t} = |g(j\omega)|e^{j[\omega t + \phi(\omega)]}$$

where

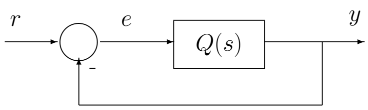
$$g(j\omega) \;=\; \overbrace{|g(j\omega)|}^{\text{gain}}\; e^{j\,\overbrace{\phi(\omega)}^{\text{phase}}}$$

- Since $e^{j\omega t} = \cos\omega t + j\sin\omega t$

$$\begin{aligned} \text{Im}\,\{u(t)\} &= \sin\omega t \\ \text{Im}\,\{y_{ss}(t)\} &= |g(j\omega)|\sin[\omega t + \phi(\omega)] \end{aligned}$$

Steady-state Accuracy of Feedback Systems

- Consider the feedback system shown below. Here, $Q(s)$ represents the loop gain. *Assume the feedback loop is stable.*



- The Laplace transform of the error signal is

$$\begin{aligned} e(s) &= \frac{r(s)}{1 + Q(s)} \\ \Rightarrow\quad e_{ss} &= \lim_{t\to\infty} e(t) = \lim_{s\to 0} \frac{sr(s)}{1 + Q(s)} \end{aligned}$$

- The **system type** N is defined to be the number of **free integrators** in the loop.
- For a unit step

$$e_{ss} = \frac{1}{1 + K_p}, \qquad K_p = \lim_{s\to 0} Q(s)$$

For a type 0 system ($N = 0$), $e_{ss} = 1/[1 + Q(0)]$, while $e_{ss} = 0$ for type 1 and type 2 systems.

- The following table gives e_{ss} for type 0, 1, and 2 systems for step, ramp and parabolic inputs.

Type	$r(s) = \frac{1}{s}$	$r(s) = \frac{1}{s^2}$	$r(s) = \frac{1}{s^3}$	Error constants
0	$\frac{1}{1+K_p}$	∞	∞	$K_p = \lim_{s\to 0} Q(s)$
1	0	$\frac{1}{K_v}$	∞	$K_v = \lim_{s\to 0} sQ(s)$
2	0	0	$\frac{1}{K_a}$	$K_a = \lim_{s\to 0} s^2Q(s)$

Steady-state error constants

Stability Analysis - The Routh-Hurwitz Stability Criterion

- Introduction
- Bounded-input Bounded-output Stability
- General Properties of Polynomials
- The Routh-Hurwitz Stability Criterion
- The Routh Array
- Stability for First & Second Order Systems
- Applications in Feedback Design

Introduction

- One of the fundamental design objectives in control is the maintenance of system **stability**.
- Even if a system is **open-loop stable**, the introduction of feedback may destabilise it.
- Furthermore, most of the benefits of feedback are obtained by increasing the loop gain. Unfortunately, for most feedback systems, this tends to destabilise the closed-loop.
- It is therefore important to investigate the stability properties of control systems.
- For a general (possibly **nonlinear** or **time-varying**) system, the problem of stability is usually one of the most difficult to analyse.
- Fortunately, for **linear, time-invariant** (LTI) systems, the question of stability is much easier to determine.

Bounded-input Bounded-output Stability

- A system is **bounded-input bounded-output stable**, if, for every bounded input, the output remains bounded for all time.
- Let $g(s)$ be a transfer function representation of an LTI system. Then, $g(s)$ can be written as

$$g(s) = \frac{\beta_m s^m + \beta_{m-1}s^{m-1} + \cdots + \beta_1 s + \beta_0}{\alpha_n s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0}$$

  where the denominator polynomial is called the **characteristic polynomial** of $g(s)$.

- The equation

$$\alpha_n s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0 = 0$$

  is called the **characteristic equation** for $g(s)$.

- The roots of the characteristic equation are called the **poles** of $g(s)$.
- The **order** of the system $g(s)$ is defined to be the degree of the characteristic polynomial.

- We already know that
  - **Fact:** *An LTI system is bounded-input bounded-output stable provided all its poles lie in the left half-plane.*
  - **Fact:** *An LTI system is* **marginally stable** *if all its poles lie in the left half-plane, and at least one pole on the imaginary axis.*
  - **Fact:** *An LTI system is* **unstable** *if any of its poles lie in the right half-plane.*
- One way of determining stability is to calculate the roots of the characteristic equation.
- The disadvantage of this is that system parameters must be assigned numerical values, which makes it difficult to find the range of values of a parameter that ensures stability.
- Furthermore, in a feedback control system, even if we know the open-loop system poles, we still need to determine the closed-loop poles.
- An alternative procedure which determines the *location* of the poles is presented next.

General Properties of Polynomials

- A left half-plane root of a characteristic polynomial (pole) is called a **stable root (pole)**.
- A right half-plane root is called an **unstable root**.
- An imaginary axis root is called a **marginally stable root**.
- Assume all polynomial coefficients are real.
- Consider the second order polynomial

$$\begin{aligned} P(s) &= s^2+\alpha_1 s+\alpha_0=(s-p_1)(s-p_2) \\ &= s^2-(p_1+p_2)s+p_1p_2 \\ \Rightarrow \alpha_1 &= -(p_1+p_2), \quad \alpha_0=p_1p_2 \end{aligned}$$

- p_1 and p_2 are either real, or else complex conjugate. If p_1 & p_2 are stable, we have $\alpha_1 > 0$ & $\alpha_0 > 0$.

- Consider next the third order polynomial

$$\begin{aligned} P(s) &= s^3+\alpha_2 s^2+\alpha_1 s+\alpha_0=(s-p_1)(s-p_2)(s-p_3) \\ &= s^3-(p_1+p_2+p_3)s^2+(p_1p_2+p_2p_3+p_1p_3)s \\ &\quad -p_1p_2p_3 \end{aligned}$$

$$\begin{aligned} \Rightarrow\ & \alpha_2=-(p_1+p_2+p_3) \\ & \alpha_1=(p_1p_2+p_2p_3+p_1p_3), \\ & \alpha_0=-p_1p_2p_3 \end{aligned}$$

- Again, if all the roots are stable, all the coefficients will have the same sign.
- **NOT THE OTHER WAY**

- Consider a general nth order polynomial:

$$\begin{aligned} P(s) &= s^n+\alpha_{n-1}s^{n-1}+\ldots+\alpha_1 s+\alpha_0 \\ &= (s-p_1)(s-p_2)\cdots(s-p_n) \end{aligned}$$

- By analogy with the previous examples, we have:

$\alpha_{n-1} = -\sum_{i=1}^{n} p_i$

$\alpha_{n-2} = \Sigma$ product of roots taken 2 at a time

$\alpha_{n-3} = -\Sigma$ product of roots taken 3 at a time

$\vdots$

$\alpha_0 = (-1)^n \prod_{i=1}^{n} p_i$

- We conclude:

1. If all the roots are stable, all the polynomial coefficient will be positive.
2. If any coefficient is negative, at least one root is unstable.
3. If any coefficient is zero, not all roots are stable.

The Routh-Hurwitz Stability Criterion

- The stability of an LTI system requires that all the poles (roots of the characteristic equation) must lie in the left half-plane.
- The **Routh-Hurwitz** stability criterion is an analytical procedure for determining if all roots of a polynomial lie in the left half-plane, *without evaluating these roots.*
- Consider the nth order polynomial

$$P(s) = \alpha_n s^n+\alpha_{n-1}s^{n-1}+\ldots+\alpha_1 s+\alpha_0$$

in which we can always assume $\alpha_0 \neq 0$.

- If $\alpha_0 = 0$, we can write :

$$P(s) = s\,\underbrace{(\alpha_n s^{n-1}+\alpha_{n-1}s^{n-2}+\cdots+\alpha_1)}_{\hat{P}(s)}$$

and work with $\hat{P}(s)$ instead.

- Note that in the case that $\alpha_0 = 0$, $P(s)$ will have at least one root at the origin and we conclude that the LTI system is marginally stable or unstable.

The Routh Array

- The first 2 rows of the **Routh array** are formed from the characteristic polynomial coefficients

$$P(s) = \alpha_n s^n + \alpha_{n-1} s^{n-1} + \alpha_{n-2} s^{n-2} + \alpha_{n-3} s^{n-3} + \alpha_{n-4} s^{n-4} + \cdots + \alpha_1 s + \alpha_0$$

$$\begin{array}{c|cccccc}
s^n & \alpha_n & \alpha_{n-2} & \alpha_{n-4} & \alpha_{n-6} & \dots \\
s^{n-1} & \alpha_{n-1} & \alpha_{n-3} & \alpha_{n-5} & \alpha_{n-7} & \dots \\
s^{n-2} & b_1 & b_2 & b_3 & b_4 & \dots \\
s^{n-3} & c_1 & c_2 & c_3 & c_4 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
s^2 & k_1 & k_2 & & & \\
s^1 & l_1 & & & & \\
s^0 & m_1 & & & &
\end{array}$$

- The remaining rows are formed as follows:

$$b_1 = -\frac{1}{\alpha_{n-1}} \begin{vmatrix} \alpha_n & \alpha_{n-2} \\ \alpha_{n-1} & \alpha_{n-3} \end{vmatrix}, b_2 = -\frac{1}{\alpha_{n-1}} \begin{vmatrix} \alpha_n & \alpha_{n-4} \\ \alpha_{n-1} & \alpha_{n-5} \end{vmatrix}$$

$$c_1 = -\frac{1}{b_1} \begin{vmatrix} \alpha_{n-1} & \alpha_{n-3} \\ b_1 & b_2 \end{vmatrix}, \; c_2 = -\frac{1}{b_1} \begin{vmatrix} \alpha_{n-1} & \alpha_{n-5} \\ b_1 & b_3 \end{vmatrix}$$

$$\vdots$$

$$m_1 = -\frac{1}{l_1} \begin{vmatrix} k_1 & k_2 \\ l_1 & 0 \end{vmatrix}$$

- Once the Routh array is completed, we have the

Routh-Hurwitz Criterion

The number of unstable roots is equal to the number of sign changes in the first column of the array

- **Example**: Consider the third order polynomial:

$$P(s) = s^3 + s^2 + 2s + 8$$
$$\Rightarrow \text{roots} : -2.0, \; \mathbf{0.5 \pm j1.9365}$$

- The Routh array is:

$$\begin{array}{c|cc}
s^3 & 1 & 2 \\
s^2 & 1 & 8 \\
s & -6 & \\
1 & 8 &
\end{array}$$

- The two sign changes (from +1 to −6 and from −6 to +8) indicate two unstable roots.

- **Not all arrays can be completed in this manner**.

- **Case 2 Problems** *If the first element of a row is zero, with at least one nonzero element in the same row, the procedure is modified by replacing that first element with a small number ϵ such that $|\epsilon| \ll 1$ and proceeding as before.*

- **Example:** Consider the polynomial

$$P(s) = s^5 + 2s^4 + 2s^3 + 4s^2 + 11s + 10$$
$$\Rightarrow \text{roots} : -1.31, \; -1.24 \pm j1.04, \; \mathbf{0.89 \pm j1.46}$$

- The modified Routh array is

$$\begin{array}{c|ccc}
s^5 & 1 & 2 & 11 \\
s^4 & 2 & 4 & 10 \\
s^3 & 0 \rightarrow \epsilon & 6 & 0 \\
s^2 & -12/\epsilon & 10 & \\
s & 6 & & \\
1 & 10 & &
\end{array}$$

- There are two sign changes (irrespective of the sign of ϵ) indicating two unstable roots.

- The 1st element of the 4th row is calculated as

$$-\frac{1}{\epsilon} \begin{vmatrix} 2 & 4 \\ \epsilon & 6 \end{vmatrix} = -\frac{1}{\epsilon}(12 - 4\epsilon) = 4 - \frac{12}{\epsilon} \simeq -\frac{12}{\epsilon}$$

- A similar procedure is used to calculate the remaining rows.

- **Case 3 Problems** *If every entry in a row is zero, the last modification will not give useful information and another modification is needed.*

- An **even polynomial** is one in which the powers of s are even integers or zero only: $s^0,\ s^2,\ s^4, \ldots$

- The roots of an even polynomial are symmetric with respect to both the real and imaginary axes.

- **Example:** Consider the even polynomial

$$p(s) \;=\; s^2+1$$
$$\Rightarrow \text{ roots } \;:\; \pm\mathbf{j}$$

$$\begin{array}{r|ll} s^2 & 1 & 1 \\ s & 0 & 0 \\ 1 & ? & \end{array}$$

- Using the previous modification would give

$$\begin{array}{r|ll} s^2 & 1 & 1 \\ s & \epsilon & 0 \\ 1 & 1 & \end{array}$$

- If ϵ is chosen positive - there are no sign changes, while a negative ϵ would give two sign changes.

- The previous test is inconclusive.

- The modification is illustrated with an example.

- **Example:** $P(s) = (s+1)(s^2+2) = s^3+s^2+2s+2$

$$\Rightarrow \text{roots}: \; -1,\ \pm\mathbf{j}\sqrt{\mathbf{2}}$$

$$\begin{array}{r|lll} s^3 & 1 & 2 & \textbf{auxiliary polynomial} \\ s^2 & 1 & 2 & \Rightarrow \;\overbrace{P_a(s) = s^2+2} \\ s & 0 \rightarrow 2 & 0 & \Leftarrow \;\underbrace{2s}_{\frac{dP_a(s)}{ds}} \\ 1 & 2 & & \end{array}$$

- $\dfrac{dP_a(s)}{ds} = 2s$ allows the array to be completed.

- The absence of a sign change indicates no unstable roots, but there are marginally stable roots (due to the auxiliary polynomial).

- Applying the modified criterion to $P_a(s) = s^2+2$:

$$\begin{array}{r|lll} s^2 & 1 & 2 & \Rightarrow P_a(s) = s^2+2 \\ s & 0 \rightarrow 2 & & \Leftarrow \dfrac{dP_a(s)}{ds} = 2s \\ 1 & 2 & & \end{array}$$

- The absence of a sign change indicates no unstable roots, so all the roots are on the imaginary axis. We conclude the system is marginally stable.

- **Example:** This example has a combination of case 2 and 3 problems:

$$P(s) \;=\; s^4+4$$
$$\Rightarrow \text{roots} \;\;:\; -1 \pm j,\ \mathbf{1 \pm j}$$

- The modified Routh array is:

$$\begin{array}{r|lll} s^4 & 1 & 0\ \ 4 & \Rightarrow\ P_a(s) = s^4+4 \\ s^3 & 0 \rightarrow 4 & 0\ \ 0 & \Leftarrow\ \dfrac{dP_a(s)}{ds} = 4s^3 \\ s^2 & 0 \rightarrow \epsilon & 4 & \\ s & -16/\epsilon & & \\ 1 & 4 & & \end{array}$$

- Summary of procedure:
  - All elements of row 2 are zero (case 3 problem)
  - Row 1 defines the auxiliary polynomial $P_a(s)$
  - Row 2 is replaced by coefficients of $\dfrac{dP_a(s)}{ds}$
  - Only element 1 of row 3 is zero (case 2 problem)
  - The zero element is replaced by ϵ with $|\epsilon| \ll 1$

- The two sign changes indicate two unstable roots.

Stability for First & Second Order Systems

- For a 1st order system with characteristic polynomial

$$P(s) = \alpha_1 s + \alpha_0$$

it is clear that stability requires both coefficients to have the same sign.

- A 2nd order system with characteristic polynomial

$$P(s) = \alpha_2 s^2 + \alpha_1 s + \alpha_0$$

has a Routh array

$$\begin{array}{r|ll} s^2 & \alpha_2 & \alpha_0 \\ s & \alpha_1 & \\ 1 & \alpha_0 & \end{array}$$

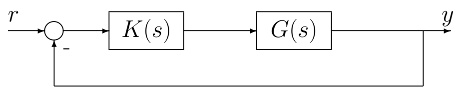
A *necessary and sufficient condition* for stability is that all coefficients have the same sign.

- Thus, for 1st and 2nd order systems, stability properties can be determined by inspection.

- For higher order systems we can only conclude that a *necessary condition* for stability is that all the coefficients have the same sign.

Applications in Feedback Design

- Consider the feedback system involving a plant $G(s)$ and compensator $K(s)$.



- The closed-loop transfer function $H(s)$ is

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

- The closed-loop poles are given by the roots of the **characteristic equation**

$$1 + G(s)K(s) = 0$$

- Suppose we write

$$G(s) = \frac{N_g(s)}{D_g(s)}, \qquad K(s) = \frac{N_k(s)}{D_k(s)}$$

where $N_g(s)$, $D_g(s)$, $N_k(s)$ and $D_k(s)$ are polynomials.

- Then, the closed-loop transfer function is given by

$$\begin{aligned} H(s) &= \frac{\dfrac{N_g(s)N_k(s)}{D_g(s)D_k(s)}}{1 + \dfrac{N_g(s)N_k(s)}{D_g(s)D_k(s)}} \\ &= \frac{N_g(s)N_k(s)}{N_g(s)N_k(s) + D_g(s)D_k(s)} \end{aligned}$$

- The poles of the closed-loop system are also given by the roots of the characteristic polynomial

$$N_g(s)N_k(s) + D_g(s)D_k(s) = 0$$

- The poles of the open-loop system $G(s)$ are the roots of its characteristic polynomial

$$D_g(s) = 0$$

which are generally different from closed-loop poles.

- A fundamental design objective in control is the stabilisation of unstable systems.

- The Routh-Hurwitz stability criterion can be used as an aid in feedback design.

- **Example:** Let $G(s)$ be given by

$$G(s) = \frac{1}{s^3+5s^2+2s-8} = \frac{1}{(s-1)(s+2)(s+4)}$$

$\Rightarrow$ poles : $+\mathbf{1}, -2, -4 \Rightarrow$ unstable

- Let $K(s) = K$ be a constant controller. Find the range of values of K for which the closed-loop is stable.

- The closed-loop poles are the roots of the characteristic equation

$$\begin{aligned} 1 + KG(s) &= 1 + \frac{K}{s^3+5s^2+2s-8} = 0 \\ \Rightarrow &\quad s^3+5s^2+2s+(K-8) = 0 \end{aligned}$$

- The Routh array is:

$$\begin{array}{r|cc} s^3 & 1 & 2 \\ s^2 & 5 & K-8 \\ s & 0.2(18-K) & \\ 1 & K-8 & \end{array}$$

- For stability we need

$$8 < K < 18$$

- **Example:** Let $G(s)$ be given by

$$G(s) = \frac{1}{s^3 - s^2 - 10s - 8}$$

- Since the coefficients of the characteristic polynomial do not have the same sign, we conclude that $G(s)$ is unstable.

- Let $K(s) = K$ be a constant controller. Find the range of values of K for which the closed-loop is stable.

- The characteristic equation is given by

$$\begin{aligned} 1 + KG(s) &= 1 + \frac{K}{s^3 - s^2 - 10s - 8} = 0 \\ \Rightarrow &\quad s^3 - s^2 - 10s + (K-8) = 0 \end{aligned}$$

- It follows that no choice of K can ensure all coefficients have the same sign.

- We conclude that $G(s)$ cannot be stabilised by a constant controller.

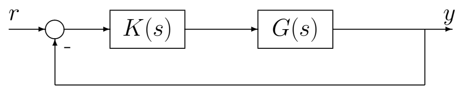
- **We need dynamic compensation**.

Root-Locus Analysis

- Introduction
- The Root-Locus
- Criteria for Plotting the Root-Locus
- Rules for Plotting the Root-Locus
- Derivation of Root-Locus Plotting Rules
- The Magnitude Criterion
- Closed-Loop Stability Range
- Other Feedback Configurations
- Root-Locus Plotting Rules (Negative K)

Introduction

- Consider the following feedback system:



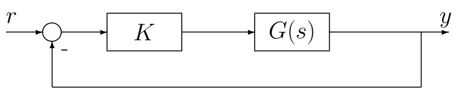
- The closed-loop transfer function is

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

- The closed-loop poles are given by the roots of the characteristic equation
$$1 + G(s)K(s) = 0$$
- We have seen that the response of an LTI system is largely determined by the location of its poles.
- In **control system analysis**, we are interested in how the location of the closed-loop poles is determined by $K(s)$.
- In **control system design**, we are concerned with finding a compensator $K(s)$ that places the closed-loop poles in some desired region of the complex plane.

The Root-Locus

- Suppose first that the feedback compensator is a constant, $K(s) = K$.



- The system characteristic equation becomes
$$1 + KG(s) = 0$$
- **The root-locus (RL) of a system is a plot of the roots of the system characteristic equation (closed-loop poles) as K varies from 0 to ∞.**
- For an nth order system, the root-locus is a family of n curves traced out by the n closed-loop poles as K varies over its range.
- An inspection of the root-locus allows us to find the possible location of the closed-loop poles as a function of K.
- This enables us to determine the nature of the response of the closed-loop system for all possible values of K.

Example

- Suppose that

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

so that the characteristic equation becomes

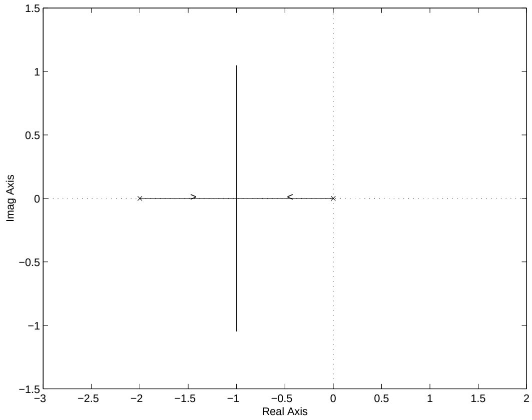
$$1 + \frac{K}{s(s+2)} = 0 \implies s^2 + 2s + K = 0$$

- The closed-loop poles are then given by

$$s = -1 \pm \sqrt{1-K}$$

1. For $0 \le K < 1$, the poles are real and unequal
2. For $K = 1$, the poles are real and equal
3. For $K > 1$, the poles are complex conjugate and lie on the $s = -1$ line in the complex plane.

- The RL is shown on the following page.



Root-Locus Plot for $G(s) = \frac{1}{s(s+2)}$

Criteria for Plotting the Root-Locus

- A point s lies on the RL if

$$1 + KG(s) = 0 \iff K = -\frac{1}{G(s)}$$

$$\iff G(s) = -\frac{1}{K}$$

- Since K is positive, this is equivalent to:

  1. The **magnitude criterion**:

  $$K = \frac{1}{|G(s)|}$$

  2. The **angle criterion**:

  $$\arg G(s) = \pm r\pi, \quad r = 1, 3, 5, \ldots$$

- Since the angle criterion is independent of K, we can use it to:
  1. determine whether a point s lies on the RL,
  2. plot the RL.

- The magnitude criterion can be used to find K corresponding to a point on the RL.

- The figure illustrates the angle criterion for

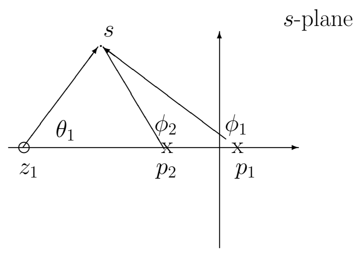
$$G(s) = \frac{s - z_1}{(s - p_1)(s - p_2)}$$

- A point s lies on the RL if

$$\arg \frac{s - z_1}{(s - p_1)(s - p_2)} = \pm\pi$$

- That is, if

$$\underbrace{\arg(s - z_1)}_{\theta_1} - [\underbrace{\arg(s - p_1)}_{\phi_1} + \underbrace{\arg(s - p_2)}_{\phi_2}] = \pm\pi$$



- In general, a point s lies on the RL if

$$\sum_i(\text{angles from } z_i) - \sum_i(\text{angles from } p_i) = \pm(2l+1)\pi$$

Rules for Plotting the Root-Locus

1. The RL is **symmetric** w.r.t. the real axis.

2. The RL **originates** on the poles of $G(s)$ ($K=0$) and **terminates** on the zeros of $G(s)$ ($K \to \infty$).

3. If $G(s)$ has α **zeros at infinity**, the RL will approach α **asymptotes** as K approaches infinity. The angles of the asymptotes are

$$\phi_A = \pm \frac{(2l + 1)\pi}{\alpha}, \quad l = 0, 1, \ldots$$

   These asymptotes intersect the real axis at

$$\sigma_A = \frac{\Sigma \text{ poles } - \Sigma \text{ zeros}}{\text{number of poles } - \text{ number of zeros}}$$

4. The RL includes all points on the real axis to the left of an **odd** number of poles and zeros.

5. The **breakaway points** σ_b are roots of

$$\frac{dG(s)}{ds} = 0$$

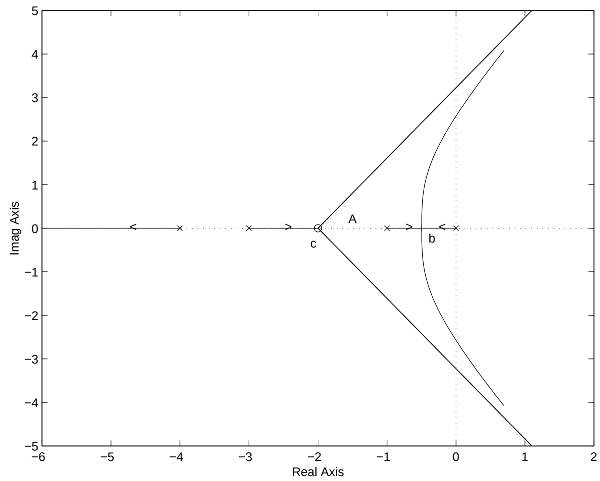
6. The RL **departs** from a pole p_j at an angle

$$\theta_d = \sum_i \theta_{zi} - \sum_{i \neq j} \theta_{pi} + (2l+1)\pi$$

and **arrives** at a zero z_j at an angle

$$\theta_a = \sum_i \theta_{pi} - \sum_{i \neq j} \theta_{zi} + (2l+1)\pi$$

where θ_{pi} is the angle from p_i to p_j and θ_{zi} is the angle from z_i to z_j.



Root-Locus Plot for $G(s) = \frac{s+2}{s(s+1)(s+3)(s+4)}$, $\quad \alpha = 3$

x: pole of $G(s)$

o: zero of $G(s)$

b: breakaway point

c: centre of asymptotes

A: angles of asymptotes

Derivation of Root-Locus Plotting Rules

1. **The RL is symmetric w.r.t. the real axis.**
   - This follows from the assumption that $G(s)$ is a ratio of two polynomials with real coefficients. So, the characteristic polynomial roots are either real or occur in complex conjugate pairs.

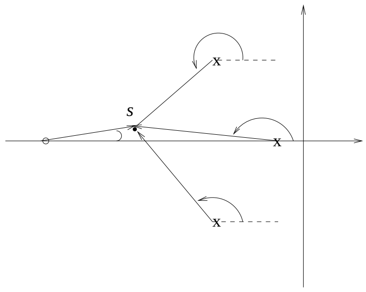
2. **The RL originates on the poles (for $K=0$) and terminates on the zeros (as $K \to \infty$), of $G(s)$.**
   - Suppose that
   $$G(s) = \frac{b_m(s-z_1)(s-z_2)\cdots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)}$$
   - **Assumption:** $b_m > 0$. (If $b_m < 0$, replace with $-b_m > 0$ and apply rules for negative K.)
   - Then the characteristic equation becomes
   $$1 + KG(s) = 0$$
   or
   $$(s-p_1)\cdots(s-p_n)+Kb_m(s-z_1)\cdots(s-z_m)=0$$
   - If $K = 0$ then $p_1, \ldots, p_n$ must lie on the RL.
   - As K approaches infinity but s remains finite, the root-loci approach $z_1, \ldots, z_m$.

3. **If $G(s)$ has α zeros at infinity, the RL will approach α asymptotes as $K \to \infty$. The angles of the asymptotes are**
$$\phi_A = \pm\frac{(2l+1)\pi}{\alpha}, \quad l = 0, 1, \ldots$$
and they intersect the real axis at
$$\sigma_A = \frac{\Sigma \text{ poles } - \Sigma \text{ zeros}}{\text{number of poles } - \text{ number of zeros}}$$

   - Write the open-loop transfer function $G(s)$ as
   $$G(s)=\frac{b_m s^m + b_{m-1}s^{m-1}+\cdots}{s^n + a_{n-1}s^{n-1}+\cdots} = \frac{b_m s^m + \cdots}{s^{m+\alpha}+\cdots}$$
   where
   $$\alpha = n - m > 0 \qquad \text{(no asymptotes if } \alpha = 0)$$
   so that $G(s)$ has α zeros at infinity.
   - Then the RL for large s satisfies
   $$\lim_{s\to\infty}[1+KG(s)] = 1 + \frac{Kb_m}{s^\alpha} = 0 \Rightarrow s^\alpha + Kb_m = 0$$
   which has roots $s^\alpha = -Kb_m$ and angles ϕ_A.
   - In the previous example, $\alpha = 3$ and so
   $$\phi_A = \pm\frac{\pi}{3},\ \pi$$
   $$\sigma_A = \frac{(-1-3-4)-(-2)}{3} = -2$$

4. **The RL includes all points on the real axis to the left of an odd number of poles and zeros**

This follows from the following observations:



- A point s lies on the RL if

  $\sum_i(\text{angles from } z_i) - \sum_i(\text{angles from } p_i) = \pm(2l+1)\pi$

- The point s is shown off the real axis for clarity.

- The angle contribution from a pair of complex conjugate poles or zeros cancels out.

- The angle contribution from a pole or zero to the left of s is $\theta = 0$.

- The angle contribution from a pole or zero to the right of s is $\theta = \pm\pi$

5. **The breakaway points σ_b are among the real roots of**

$$\frac{dG(s)}{ds} = 0$$

This follows from the following observations:

- The breakaway point σ_b must be real.

- The characteristic equation has at least two roots at the breakaway point.

- If a rational function $R(s)$ has at least two roots at σ_b, then $\frac{dR(s)}{ds}$ has at least one root at σ_b:

$$\begin{aligned} R(s) &= (s-\sigma_b)^2 C(s) \\ \frac{dR(s)}{ds} &= 2(s-\sigma_b)C(s) + (s-\sigma_b)^2 \frac{dC(s)}{ds} \\ &= (s-\sigma_b)[2C(s) + (s-\sigma_b)\frac{dC(s)}{ds}] \end{aligned}$$

- Since

$$R(s) = 1 + KG(s)$$

  has at least two roots at σ_b,

$$\frac{dR(s)}{ds} = K\frac{dG(s)}{ds}$$

  has at least one root at σ_b.

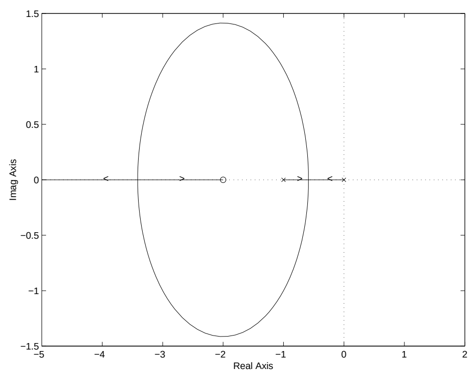
- **Example:** Let

$$\begin{aligned} G(s) &= \frac{s+2}{s(s+1)} = \frac{s+2}{s^2+s} \\ \frac{dG(s)}{ds} &= \frac{s^2+s-(2s+1)(s+2)}{s^2(s+1)^2} \\ &= -\frac{s^2+4s+2}{s^2(s+1)^2} \end{aligned}$$

- The derivative of $G(s)$ vanishes at

$$\sigma_b = -2 \pm \sqrt{2} = -3.4142, -0.5858$$

- The root-locus is shown below.



6. **The RL departs from a complex pole p_j (arrives at a complex zero z_j) at an angle θ_d (θ_a) given by**

$$\begin{aligned} \theta_d &= \sum_i \theta_{zi} - \sum_{i\neq j} \theta_{pi} \pm (2l+1)\pi \\ \theta_a &= \sum_i \theta_{pi} - \sum_{i\neq j} \theta_{zi} \pm (2l+1)\pi \end{aligned}$$

where θ_{pi} (θ_{zi}) is the angle from p_i (z_i) to p_j (z_j)

This follows from the following observations:

- A point s lies on the RL if

  $\sum_i(\text{angles from } z_i) - \sum_i(\text{angles from } p_i) = \pm(2l+1)\pi$

- For departure from p_j (arrival to z_j), s can be taken to be very close to p_j (z_j).

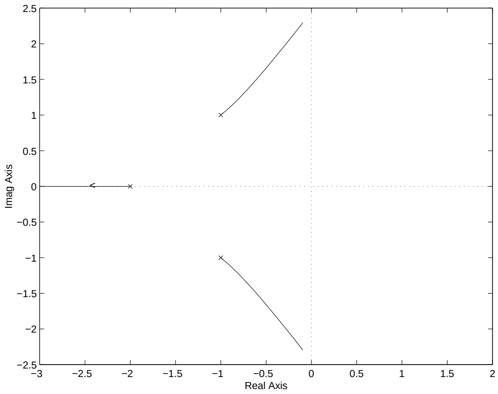
- The angles to s from the other poles $p_i,\ i \neq j$ (other zeros $z_i,\ i \neq j$) can be approximated by the angles from these poles (zeros) to p_j (z_j).

- The rule is illustrated in the following example.

- **Example:** Let

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

- The root-locus is shown below.



- The angle of departure from the pole

$$p_1 = -1 + j$$

can be evaluated as

$$\theta_d = 180^\circ - 90^\circ - 45^\circ = 45^\circ$$

The Magnitude Criterion

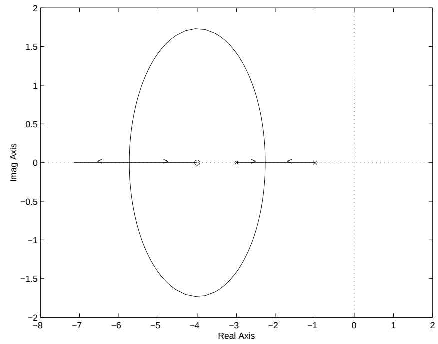
- Suppose that s_1 is a point on the RL. It follows that there must exist a gain K such that one of the closed-loop poles is equal to s_1.

- The required gain can be found from the magnitude criterion:

$$K = 1/\,|G(s_1)|$$

- **Example:** Let $G(s) = \frac{s+4}{(s+1)(s+3)}$.
- The figure shows that $s_1 = -2$ lies on the RL. The K that gives s_1 as one of the closed-loop poles is

$$K = \frac{1}{|G(-2)|} = |\frac{-2+4}{(-2+1)(-2+3)}|^{-1} = 0.5$$



Closed-Loop Stability Range

- The range of K for closed-loop stability is best determined via the Routh-Hurwitz stability criterion.

- **Example:** Let

$$G(s) = \frac{1}{(s-1)(s+2)(s+3)} = \frac{1}{s^3+4s^2+s-6}$$

$$\Rightarrow 1+KG(s)=0 \Rightarrow s^3+4s^2+s+(K-6)=0$$

- The Routh array is given by

$$\begin{array}{c|cc} s^3 & 1 & 1 \\ s^2 & 4 & K-6 \\ s & 0.25(10-K) & \\ 1 & (K-6) & \end{array}$$

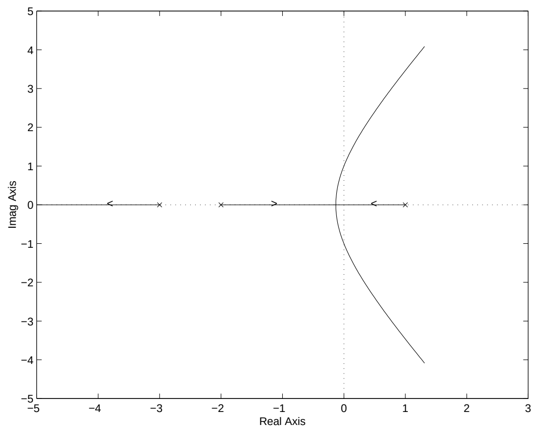
- The closed-loop is stable for $6 < K < 10$. The Routh arrays when $K = 6$ and $K = 10$ are

$$\begin{array}{c|ccl} s^3 & 1 & 1 & \\ s^2 & 4 & 0 & \\ s & 1 & & \Rightarrow \overbrace{s}^{P_a(s)} \\ 1 & 0 \rightarrow 1 & & \Leftarrow \underbrace{1}_{dP_a(s)/ds} \end{array} \qquad \begin{array}{c|ccl} s^3 & 1 & 1 & \\ s^2 & 4 & 4 & \Rightarrow \overbrace{4(s^2+1)}^{P_a(s)} \\ s & 0 \rightarrow 8 & & \Leftarrow \underbrace{8s}_{dP_a(s)/ds} \\ 1 & 4 & & \end{array}$$

- When $K=6$ the RL crosses the $j\omega$-axis at the root of the auxiliary polynomial $P_a(s)=s$, $s=0$.

- When $K = 10$, the RL crosses the $j\omega$-axis at the roots of the auxiliary polynomial $P_a(s) = 4(s^2+1)$, that is, at $s = \pm j$.

- The following figure shows the RL plot for $G(s)$.



Root-Locus Plot for $G(s) = \frac{1}{(s-1)(s+2)(s+3)}$

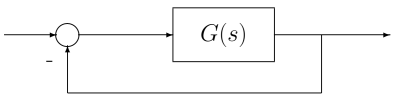
Other Feedback Configurations

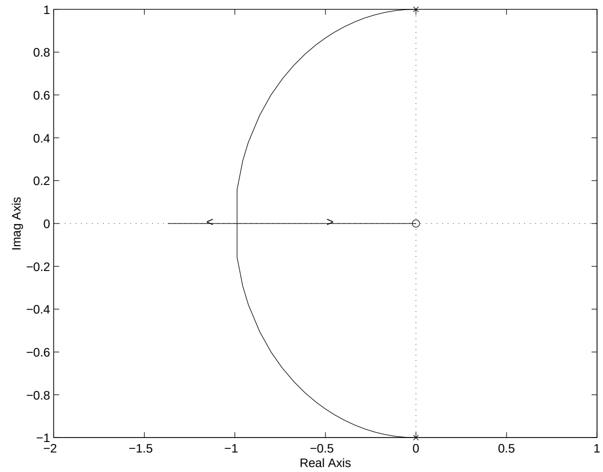
- The RL technique can be used to analyse more general feedback configurations.

- **Example:** Consider the feedback loop in the figure with
$$G(s) = \frac{1}{s(s+k)}$$
and it is required to plot the roots of the characteristic equation (closed-loop poles)
$$1 + G(s) = 1 + \frac{1}{s(s+k)} = 0$$
as k is varied.

- The characteristic equation can be rearranged as
$$s^2 + 1 + ks = 0 \Rightarrow 1 + k\frac{s}{s^2+1} = 0$$

- We can now plot the RL of
$$\hat{G}(s) = \frac{s}{s^2+1}$$

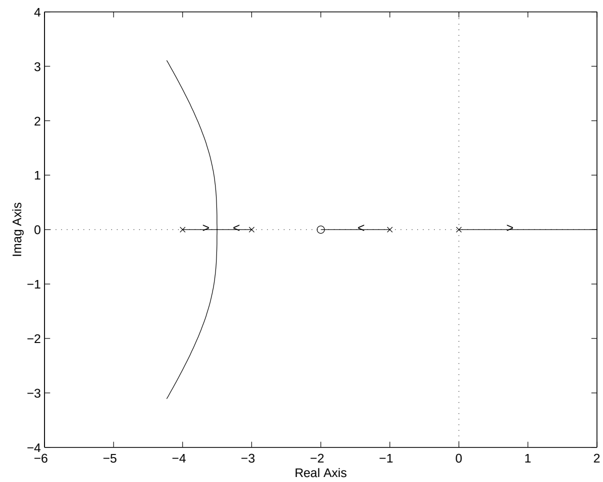




Root-Locus Plot for $\hat{G}(s) = \frac{s}{s^2+1}$

Root-Locus Plotting Rules (Negative K)

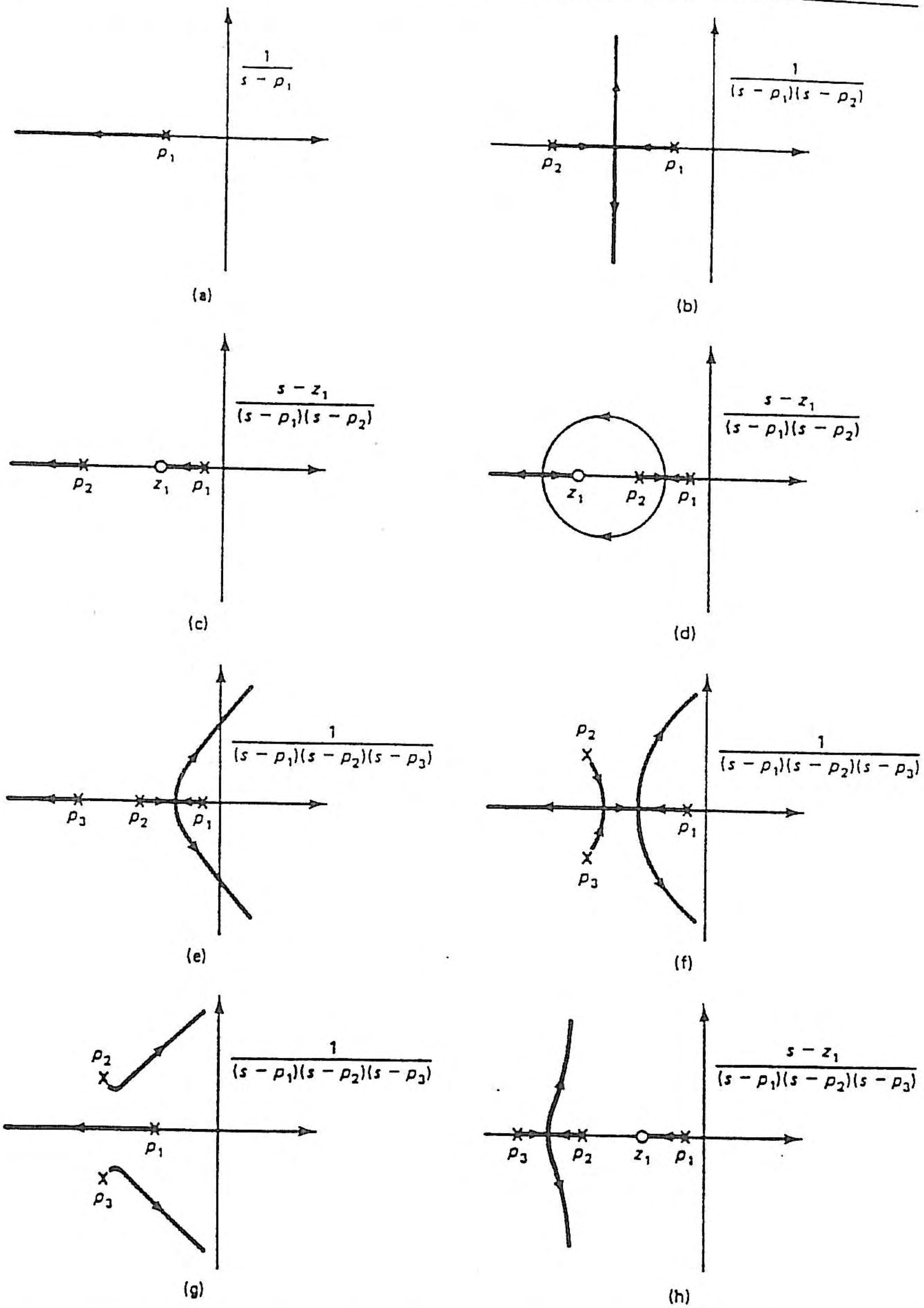
1. The RL is **symmetric** w.r.t. the real axis.
2. The RL **originates** on the poles (for $K = 0$) and **terminates** on the zeros (as $K \to -\infty$), of $G(s)$.
3. If $G(s)$ has α **zeros at infinity**, the RL will approach α **asymptotes** as K approaches infinity. The angles of the asymptotes are
$$\phi_A = \pm\frac{2l\pi}{\alpha}, \quad l = 0, 1, \ldots$$
and these asymptotes intersect the real axis at
$$\sigma_A = \frac{\Sigma \text{ poles } - \Sigma \text{ zeros}}{\text{number of poles } - \text{ number of zeros}}$$
4. The RL includes all points on the real axis to the left of an **even** number of poles and zeros.
5. The **breakaway points** σ_b are roots of
$$\frac{dG(s)}{ds} = 0$$
6. The RL **departs** from a pole p_j at an angle
$$\theta_d = \sum_i \theta_{zi} - \sum_{i \neq j} \theta_{pi} + 2l\pi$$
and **arrives** at a zero z_j at an angle
$$\theta_a = \sum_i \theta_{pi} - \sum_{i \neq j} \theta_{zi} + 2l\pi$$
where θ_{pi} is the angle from p_i to p_j and θ_{zi} is the angle from z_i to z_j.



Root-Locus for $G(s) = \frac{s+2}{s(s+1)(s+3)(s+4)}$ (Negative K)

10c

TABLE 7.6



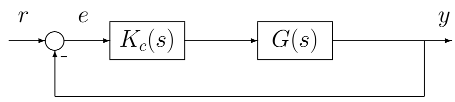
1/(s − p1)
p1
(a)
1/((s − p1)(s − p2))
p2
p1
(b)
(s − z1)/((s − p1)(s − p2))
p2
z1
p1
(c)
(s − z1)/((s − p1)(s − p2))
z1
p2
p1
(d)
1/((s − p1)(s − p2)(s − p3))
p3
p2
p1
(e)
p2
1/((s − p1)(s − p2)(s − p3))
p1
p3
(f)
p2
1/((s − p1)(s − p2)(s − p3))
p1
p3
(g)
(s − z1)/((s − p1)(s − p2)(s − p3))
p3
p2
z1
p1
(h)

Root-Locus Design

- Introduction
- Proportional (P) Controllers
- Phase-Lead Compensation
- Rate Feedback
- Proportional-Plus-Derivative (PD) Controllers
- Phase-Lag Compensation
- Proportional-Plus-Integral (PI) Controllers
- Proportional-Plus-Integral-Plus-Derivative (PID) Controllers
- Compensator Realisation

Introduction

- Consider the feedback control system involving a **plant** $G(s)$ and **compensator** $K_c(s)$.



- The closed-loop transfer function is

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

and the characteristic equation is given by

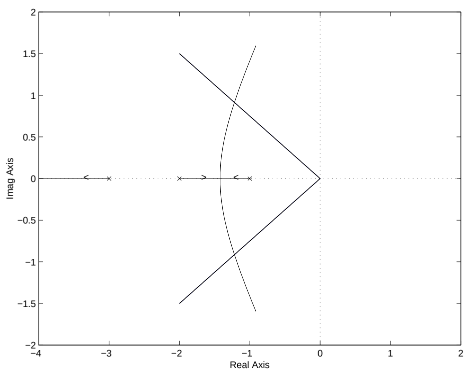
$$1 + G(s)K_c(s) = 0$$

- Suppose that the open-loop response is deemed unsatisfactory and it is required to use **feedback compensation** to improve the response.
- **Root-locus** principles can be used for:

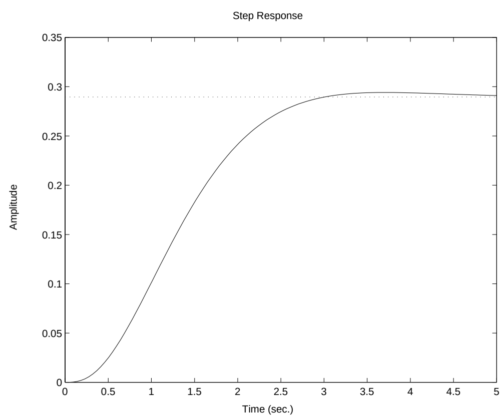
1. closed-loop **pole placement**
2. improving **stability** and **transient response**
3. improving **steady-state response**.

Proportional Controllers

- It may be possible to satisfy the design specifications using a constant compensator $K_c(s) = K_p$, called a **proportional** (P) controller.
- **Example:** Let $G(s) = \dfrac{1}{(s+1)(s+2)(s+3)}$. The design specifications on the closed-loop poles are:
  1. the dominant poles have damping ratio $\zeta = .707$
  2. the third pole has a decay rate faster than e^{-3t}.
- The figure shows the RL with the $\zeta = 0.707$ damping ratio line. It is clear that the specifications can be satisfied using a proportional controller.



- Using the *rlocfind* command in MATLAB gives $K_p = 2.4450$ with the resulting closed-loop poles at $-3.5923, -1.2038 \pm j0.9495$.
- The following figure shows the step response of the resulting closed-loop system.



- Note the small overshoot and the nonzero steady-state error (since the system is type 0).

Phase-Lead Compensation

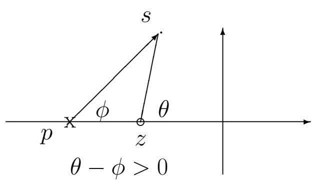
- A **phase-lead** compensator has transfer function

$$K_c(s) = K\,\frac{s-z}{s-p}$$

where z and p are real and negative and

$$|z| < |p|.$$

- A phase-lead compensator has a **positive contribution** to the angle criterion and can be used to shift the RL towards the left in the complex plane.



- Hence a phase-lead compensator will improve the transient response by improving closed-loop stability.

- **Example:** Let

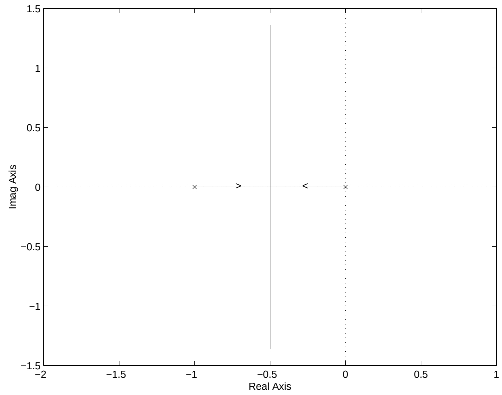
$$G(s) = \frac{1}{s(s+1)}$$

(type 1 system). The design specifications are:

1. The two dominant poles have damping ratio $\zeta = .707$ and settling time of $2s$.

2. The third pole decays at least as fast as e^{-2t}.

- The figure shows the uncompensated system RL.



- We use a phase-lead compensator

$$K_c(s) = K\,\frac{s-z}{s-p}$$

- The first specification is satisfied by placing two closed-loop poles at $s_1, \bar{s}_1$ where

$$\begin{aligned} s_1 = -2 + j2 \;&\Rightarrow\; \zeta = \sqrt{0.5}, \omega_n = 2\sqrt{2} \\ &\Rightarrow\; T_s = \frac{4}{\zeta\omega_n} = 2s \end{aligned}$$

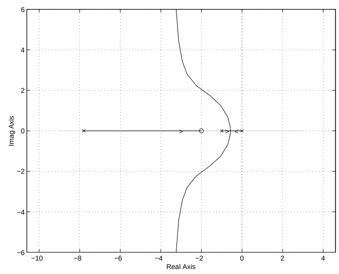
- The second specification (which ensures $s_1, \bar{s}_1$ dominant) is satisfied by setting $z = -2$.

- The compensator pole position can now be determined by the angle criterion:

$$\theta = 90^\circ - (116^\circ + 135^\circ) \pm 180^\circ \sim 19^\circ$$

which is satisfied by $p \sim -7.8$.

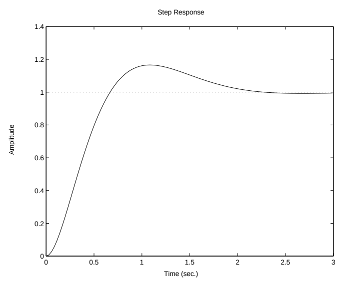
- The following shows the compensated system RL.



- Finally, K is found by the gain criterion:

$$K = -\,\frac{s_1(s_1+1)(s_1-p)}{s_1 - z} \sim 19$$

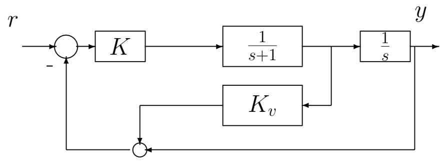
- The following shows the closed-loop step response.



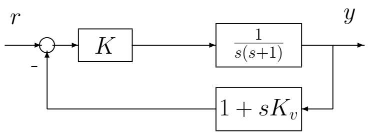
Rate Feedback

- In the last example, the positive angle contribution comes from the compensator zero. It may be possible to obtain a zero, without introducing a pole.

- Let $G(s) = \dfrac{1}{s(s+1)}$ as before and consider the following control system utilising **rate feedback**:



- This reduces to the **non-unity feedback loop**:



- The closed-loop transfer function is given by

$$H(s) = \frac{KG(s)}{1 + KG(s)K_v(s + 1/K_v)}$$

and the characteristic equation is given by

$$1 + KK_v\,\frac{s + 1/K_v}{s(s+1)} = 0$$

- **Example:** Design K and K_v to satisfy the same specifications as in the previous example.

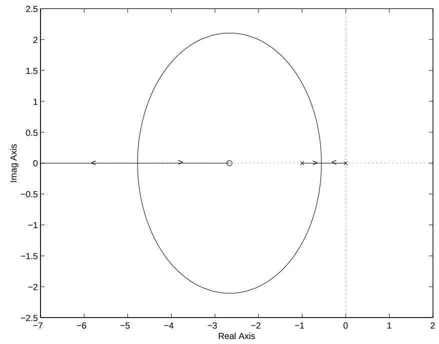
- The location of the zero $z = -1/K_v$ can be determined from the angle criterion:

$$\theta = 116^\circ + 135^\circ - 180^\circ \sim 71.6^\circ$$

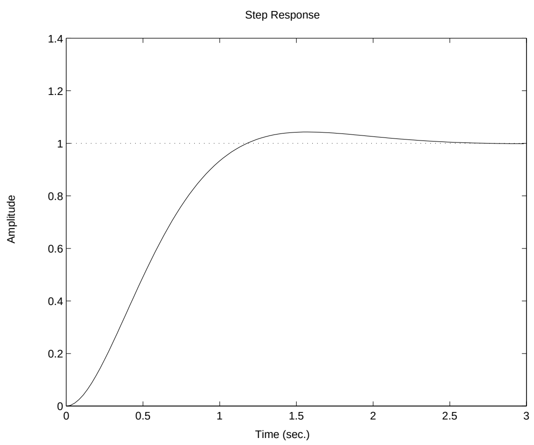
which is satisfied by $z = -8/3$. So, $K_v = 3/8$.

- Finally, K is obtained from the gain criterion:

$$KK_v = -s_1(s_1+1)/(s_1+8/3) = 3 \quad \Rightarrow \quad K = 8$$



- The following figure shows the step response of the closed-loop systems.



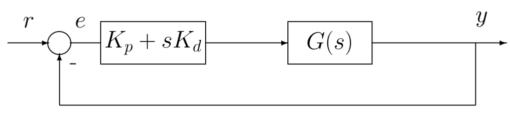
- Note that the overshoot is lower than that of the previous design, but the response is slower.

PD Controllers

- The same positive angle contribution (phase-lead), with the resulting improvement in the transient response, can be obtained by utilising the following **proportional-plus-derivative** (PD) controller

$$K_c(s) = K_p + K_d s = K_d(s + K_p/K_d)$$

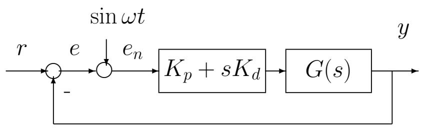
- This is shown in the feedback loop below.



- Note, however, that derivative control may amplify **high frequency noise** in the error signal e:

$$
\begin{aligned}
e_n(t) &= e(t) + \sin \omega t \\
\Longrightarrow \dot{e}_n(t) &= \dot{e}(t) + w \cos \omega t
\end{aligned}
$$

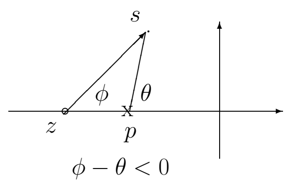
- This can deteriorate the performance of the feedback system if w is large.



Phase-Lag Compensation

- A **phase-lag** compensator has transfer function
$$K_c(s) = K\,\frac{s-z}{s-p}$$
where $z\,\&\,p$ are real and negative and $|p| < |z|$.

- A phase-lag compensator has a *negative* contribution to the angle criterion and will tend to shift the RL towards the right in the complex plane.



$\phi - \theta < 0$

- The negative angle contribution must be small to avoid **destabilising** the feedback loop.

- This angle contribution can be reduced by placing the pole and zeros close to each other.

- Phase-lag compensation can be used to improve the steady-state response of feedback systems.

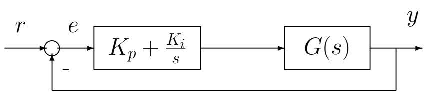
- Refer to [Phillips & Harbor, page 249] for a design procedure and fully worked out example.

Proportional-Plus-Integral Controllers

- The **proportional-plus-integral** (PI) controller

$$K_c(s) = K_p + \frac{K_i}{s} = K_p\,\frac{s + K_i/K_p}{s}$$

is a special type of phase-lag compensator ($p\!=\!0$).



- A PI controller increases the system type by 1, and can be used to improve the steady-state response.

- **Example:** Let
$$G(s) = \frac{1}{s+0.1}, \qquad \tau = 10s,\ \text{type } 0.$$

Design a PI controller such that the closed-loop

1. has zero steady-state error for a constant input

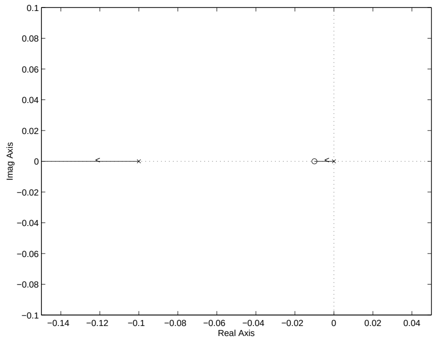
2. has a time constant $\tau = 2.5s$.

- The PI controller increases system type by 1 and ensures the first specification is satisfied.

- Choose the zero near the origin (to minimise the negative angle contribution of the PI controller). One possible choice is

$$K_i = 0.01 K_p$$

- The RL is shown below.



- For a time constant $\tau = 2.5s$, we place one of the closed-loop poles at $s_1 = -0.4$.

- Finally, the proportional gain K_p can be found using the gain criterion:

$$K_p = -\frac{(-0.4)(-0.4+0.1)}{(-0.4+0.01)} \sim 0.3077 \Rightarrow K_i \sim 0.003$$

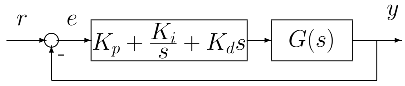
- The closed-loop system has a pole somewhere between 0 and -0.01, which has a large τ. However, its influence is cancelled by the zero at -0.01.

Proportional-Plus-Integral-Plus-Derivative Controllers

- The **proportional-plus-integral-plus-derivative** (PID) controller

$$K_c(s) = K_p + \frac{K_i}{s} + K_d s$$

is employed for systems that require improvements in both the transient and steady-state responses.

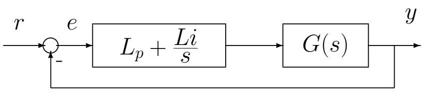


- It is one of the most widely used controllers in feedback control system design.

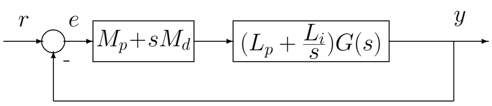
- Intuitively:
  - The proportional term K_p gives a component proportional to the **current value** of $e(t)$.
  - The integrator term K_i/s gives a component related to the accumulated **past values** of $e(t)$ and is used to improve steady-state error.
  - The derivative term $K_d s$ gives a component related to the predicted **future value** of $e(t)$ and is used to improve the speed of response (provided $e(t)$ is not too noisy).

- One PID design method is as follows:

1. Design a PI controller to improve the steady-state properties of the feedback loop:



2. Absorb the PI controller into $G(s)$ and design a PD controller to improve the transient response:



- The overall controller is then given by

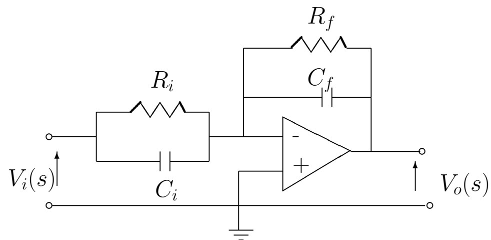
$$(L_p+\frac{L_i}{s})(M_p+M_d s) = \overbrace{(L_pM_p+L_iM_d)}^{K_p} + \frac{\overbrace{L_iM_p}^{K_i}}{s} + \overbrace{L_pM_d}^{K_d}\, s$$

- An additional pole is generally introduced to the PID controller to limit high frequency gain.

- Refer to [Phillips & Harbor, page 256] for an alternative design procedure and fully worked out example.

Controller Realisation

- The controllers considered previously can be **realised** using electronic components:



- The circuit shown has the transfer function

$$\frac{V_o(s)}{V_i(s)} = -\frac{Z_f(s)}{Z_i(s)} = -\frac{C_i(s+1/R_iC_i)}{C_f(s+1/R_fC_f)} = -\frac{C_i s+1/R_i}{C_f s+1/R_f}$$

and will realise the controllers using a suitable choice of components:

1. **Pure Gain**: Remove C_f and C_i ($C_f = C_i = 0$).
2. **Phase-lead**: Choose $R_iC_i > R_fC_f$.
3. **Phase-lag**: Choose $R_iC_i < R_fC_f$.
4. **PI**: Remove R_f ($R_f \to \infty$).
5. **PD**: Remove C_f ($C_f \to 0$).

- Refer to [Phillips & Harbor, page 263] for a circuit realisation of a PID controller.

Frequency Response Methods

- Introduction
- Nyquist Analysis
- Principle of the Argument
- Nyquist Stability Criterion
- Gain Compensation - Modified Nyquist Criterion
- Relative Stability - Gain and Phase Margins

Introduction

- Frequency-response methods are among the most widely used techniques for control system analysis and design.
- There is no one systematic design procedure for all control problems, rather, the different techniques complement each other.
- Root-locus techniques give powerful indicators for closed-loop transient response.
- Unfortunately, we need accurate, hence expensive, plant models.
- One of the advantages of frequency-response techniques is that the response of the system can be inferred experimentally without the need for accurate models.
- In fact, it is possible to design a control system without the need for a transfer function model.

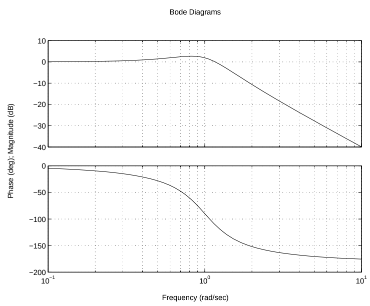
Frequency Response

- Suppose that a plant has a *stable* transfer function $G(s)$. If the input to the plant is a sinusoid, $u(t) = Ae^{j\omega t}$, of frequency ω and amplitude A, then the steady-state output

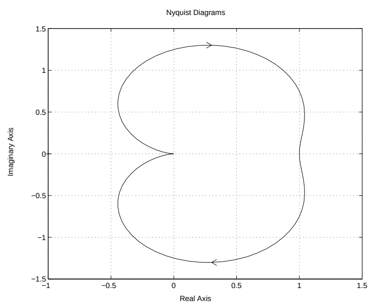
$$y_{ss}(t) = AG(j\omega)e^{j\omega t} = A|G(j\omega)|e^{j[\omega t + \angle G(j\omega)]}$$

  is also a sinusoid of the same frequency ω, amplitude $A|G(j\omega)|$ and phase $\angle G(j\omega)$.

- We call $G(j\omega),\ 0 \leq \omega \leq \infty$ the **frequency-response function**.
- We will see that the frequency-response also gives information about the stability, relative stability and the response for closed-loop systems.
- Since, for a given ω, $G(j\omega)$ is a complex number, two numbers are required to specify $G(j\omega)$:
  1. $|G(j\omega)|$ and $\angle G(j\omega)$ - **Bode diagrams**.
  2. $\mathrm{Re}G(j\omega)$ and $\mathrm{Im}G(j\omega)$ - **Nyquist diagram**.
- The first characterisation is covered elsewhere, and will be covered in more detail next year.
- We will be mainly concerned with the second characterisation.



Bode plots for $G(s) = \dfrac{1}{s^2 + 0.8s + 1}$



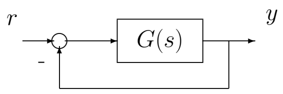
Nyquist diagram for $G(s) = \dfrac{1}{s^2 + 0.8s + 1}$

Nyquist Analysis

- Consider the feedback loop in the figure and let

$$G(s) = \frac{N(s)}{D(s)}$$

where N & D are the zero & pole polynomials.



- The closed-loop transfer function is given by

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

and the characteristic equation is given by

$$1 + G(s) = 0$$

- We derive a graphical method, using the frequency-response $G(j\omega)$, to determine closed-loop stability.

- Note that the frequency-response of stable $G(s)$ can be determined *experimentally.* This implies that we may not need a model for $G(s)$ to analyse closed-loop stability.

- We can write the closed-loop transfer function as

$$\begin{aligned} H(s) &= \frac{G(s)}{1+G(s)} = \frac{N(s)/D(s)}{1+N(s)/D(s)} \\ &= \frac{N(s)}{D(s)+N(s)} \end{aligned}$$

- We can rewrite the characteristic equation as

$$F(s) = 1 + G(s) = 1 + \frac{N(s)}{D(s)} = 0$$

$$\Rightarrow\ F(s) = \frac{D(s) + N(s)}{D(s)} = 0$$

- That is,
  1. **The poles of $F(s)$ are the open-loop poles** (poles of $G(s)$).
  2. **The zeros of $F(s)$ are the closed-loop poles** (poles of $H(s)$).

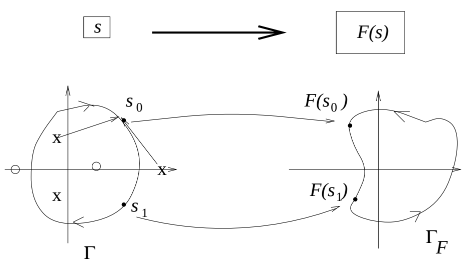
- To determine closed-loop stability, we need to know the number of right half-plane *zeros* of $F(s)$.

- First, we give an analysis of the problem in a more general setting.

Principle of the Argument

1. Let $F(s) = F_0\,\dfrac{(s-z_1)\cdots(s-z_m)}{(s-p_1)\cdots(s-p_n)}$.
2. Let Γ be a closed curve in the s-plane.
3. Let $F(s)$ have Z zeros and P poles inside Γ.
4. Assume $F(s)$ has no poles or zeros *on* Γ.
5. Let Γ_F be the closed curve defined by the mapping $s \to F(s)$ as s traverses Γ clockwise.
6. Let N be the number of net clockwise encirclement of the origin by Γ_F.

Then $N = Z - P$.



- **Proof:** Choose any s_0 on Γ. Then
  $\angle F(s_0) = \sum_i \angle(s_0 - z_i) - \sum_i \angle(s_0 - p_i) + \angle F_0$

- As s_0 traverses Γ clockwise, the total change of phase in $F(s)$ is:
  $\Delta\angle F(s) = \sum_i \Delta\angle(s - z_i) - \sum_i \Delta\angle(s - p_i)$

- Consider the phase contribution of each pole and zero of $F(s)$:
  - *Case 1*: z_i lies inside Γ $\Rightarrow$ $\Delta\angle(s - z_i) = 2\pi$
  - *Case 2*: p_i lies inside Γ $\Rightarrow$ $\Delta\angle(s - p_i) = 2\pi$
  - *Case 3*: z_i lies outside Γ $\Rightarrow$ $\Delta\angle(s - z_i) = 0$
  - *Case 4*: p_i lies outside Γ $\Rightarrow$ $\Delta\angle(s - p_i) = 0$

- That is,
  - **each zero inside Γ contributes one clockwise encirclement**
  - **each pole inside Γ contributes one anticlockwise encirclement**
  - **poles and zeros outside Γ do not have a net contribution to the encirclements**.

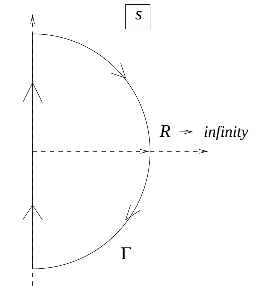
- It follows that

$$N = Z - P$$

- Let us return to the problem of determination of closed-loop stability.

- To exploit the principle of the argument, set
$$F(s) = 1 + G(s)$$
- Recall that:
  – The poles of $F(s)$ are the open-loop poles - known.
  – The zeros of $F(s)$ are the closed-loop poles - unknown.

- Identify the curve Γ with the **Nyquist contour** shown below.

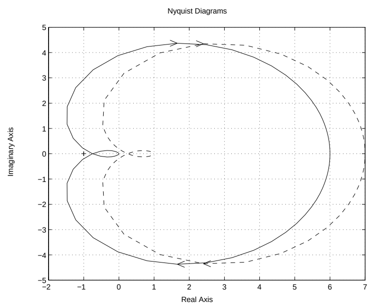


- In the limit, as $R \to \infty$, Γ covers the whole of the right half-plane - the region of instability.

- The principle of the argument requires plotting $F(s)$ as s traverses Γ and counting the number of clockwise encirclement of the origin.

- Suppose that we plot $G(s)$ instead. The resulting plot has the same shape as the plot of $F(s)$, but is shifted 1 unit to the left.

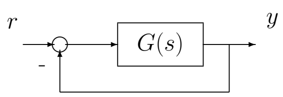
- Hence, rather than plotting $F(s)$ and counting the number of encirclements of the origin, we plot $G(s)$ instead and count the number of encirclements of the $-1 + j0$ point.



Nyquist plot of $G(s)$ (solid line), and
$F(s) = 1 + G(s)$ (dashed line)

Nyquist Stability Criterion

- Consider the feedback loop shown in the figure.
  1. Assume that $G(s)$ has no imaginary axis poles.
  2. Let P be the number of unstable poles of $G(s)$.
  3. Let Z denote the number of unstable poles of the closed-loop system $H(s) = \dfrac{G(s)}{1 + G(s)}$
  4. Let Γ denote the Nyquist contour.
  5. Let Γ_G, called the **Nyquist diagram** of $G(s)$, denote the closed curve defined by the mapping $s \to G(s)$ as s traverses Γ clockwise.
  6. Let N be the number of clockwise encirclement of the $-1 + j0$ point by Γ_G.
- Then
$$N = Z - P$$
- In particular, $H(s)$ is stable if and only if
$$-N = P$$
- **The closed-loop system is stable if and only if the number of *anticlockwise* encirclement of the $-1+j0$ point by Γ_G is equal to the number of unstable poles of $G(s)$.**



- **Example 1:** Let
$$G(s) = \frac{5}{(s+1)^2}$$

  The Nyquist contour is divided into 4 regions:

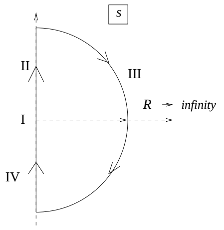
  – **I:** This is the point $s = 0$. $G(s)$ evaluated at this point is $G(0) = 5$ (**d.c. gain**).

  – **II:** This is the $j\omega$-axis. $G(s)$ evaluated along this axis is simply the **frequency-response**
  $$G(j\omega) = 5/(j\omega + 1)^2$$
    1. Each factor in the denominator has a magnitude increasing with ω from 1 to ∞. Therefore $|G(j\omega)|$ decreases from 5 to 0.
    2. The angle of each factor in the denominator increases with ω from $0°$ to $90°$. Therefore $\angle G(j\omega)$ decreases from $0°$ to $-180°$.

  – **III:** This is the infinite arc
  $$s = \{Re^{j\theta} :\ R \to \infty,\ \theta : 90° \to -90°\}.$$
  For physical systems, $\lim_{s\to\infty} G(s) = 0$, and this arc will generally map into the origin.

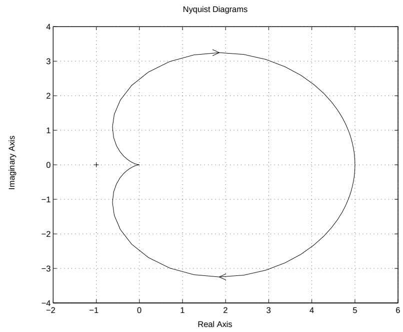
  – **IV:** This is the lower half of the $j\omega$-axis. $G(s)$ evaluated along this path is the complex conjugate of $G(s)$ evaluated along path II.

- The Nyquist contour is shown below.



Γ = Nyquist Contour

- The Nyquist diagram is shown below.



Nyquist Diagram for $G(s) = \frac{5}{(s+1)^2}$

- To apply the Nyquist criterion for $G(s) = \dfrac{5}{(s+1)^2}$:

1. P = (no. of unstable poles of $G(s)$) = 0

2. N = (no. of clockwise encirclements of -1) = 0

- Therefore, the number of unstable closed-loop poles is given by

$$Z = N + P = 0$$

- That is, the closed-loop system is stable.

- **Example 2:** Let

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

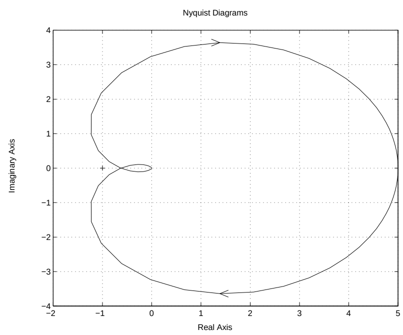
1. **Region I:** This is the d.c. gain, $G(0) = 5$.
2. **Region II:** This is the frequency-response

$$G(j\omega) = 5/(j\omega + 1)^3$$

(a) Each factor in the denominator has a magnitude increasing with ω. Therefore $|G(j\omega)|$ decreases from 5 to 0.

(b) The angle of each factor in the denominator increases with ω from 0° to 90°. Therefore $\angle G(j\omega)$ decreases from 0° to -270°.

3. **Region III:** $\lim_{s\to\infty} G(s) = 0$.



- To evaluate N we need to calculate the intersection with the real axis, the point $G(j\omega_1)$.

- This can be calculated using the Routh-Hurwitz stability criterion as follows:

1. Introduce a gain K into $G(s)$ and evaluate the characteristic equation

$$1 + KG(s) = 1 + \frac{5K}{(s+1)^3} = 0$$
$$\Longrightarrow s^3 + 3s^2 + 3s + 1 + 5K = 0$$

2. Form the Routh array

$$\begin{array}{c|ccl} s^3 & 1 & 3 & \\ s^2 & 3 & 1+5K & \\ s & (8-5K)/3 & 0 & \Longrightarrow K < 1.6 \\ 1 & 1+5K & 0 & \Longrightarrow K > -0.2 \end{array}$$

3. When $K = 1.6$, the system is marginally stable (the Nyquist diagram intersects the -1 point). Thus $1.6 \times G(j\omega_1) = -1$ and

$$G(j\omega_1) = -1/1.6 = -0.625.$$

4. To find ω_1, we form the auxiliary equation

$$3s^2 + 1 + 5 \times 1.6 = 3s^2 + 9 = 3(s^2 + 3)$$

The roots are $s = \pm j\sqrt{3} = \pm j\omega_1$. Check:

$$G(j\sqrt{3}) = 5/(1 + j\sqrt{3})^3 = -0.625$$

- To apply the Nyquist criterion for

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

1. P = (no. of unstable poles of $G(s)$) = 0
2. N = (no. of clockwise encirclements of -1) = 0

- Therefore, the number of unstable closed-loop poles is given by

$$Z = N + P = 0$$

- That is, the closed-loop system is stable.
- Note that when $K = -0.2$ the system is also marginally stable. Thus

$$-0.2G(j\omega_2) = -1 \Rightarrow G(j\omega_2) = 5$$

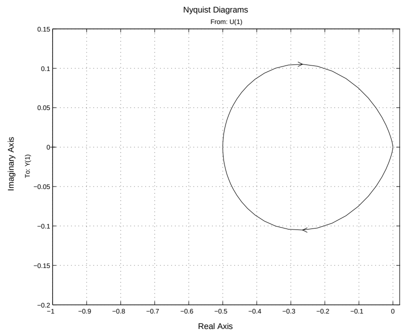
This gives the other intersection with the real axis.

- **Example:** Let $G(s) = \frac{1}{(s+1)(s+2)(s-1)}$.

1. **Region I:** This is the d.c. gain, $G(0) = -0.5$.
2. **Region II:** This is the frequency-response

$$G(j\omega) = 1/(j\omega+1)(j\omega+2)(j\omega-1)$$

   (a) Each factor in the denominator has a magnitude increasing with ω. Therefore $|G(j\omega)|$ decreases from 0.5 to 0.
   (b) The angle of each of the first two factors in the denominator increases with ω from 0° to 90° while the angle of the third factor decreases from 180° to 90°. Therefore $\angle G(j\omega)$ decreases from $-180°$ to $-270°$.
3. **Region III:** $\lim_{s\to\infty} G(s) = 0$.



- To apply the Nyquist criterion for

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

1. P = (no. of unstable poles of $G(s)$) = 1
2. N = (no. of clockwise encirclements of -1) = 0

- Therefore, the number of unstable closed-loop poles is given by

$$Z = N + P = 1$$

- That is, the closed-loop system is unstable.

Poles on the Imaginary Axis

- When $G(s)$ has poles on the imaginary axis, e.g.,

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

it is not possible to plot $G(s)$ on the Nyquist contour.

- It is still possible to determine closed-loop stability, but the Nyquist contour must be modified.
- See [Phillips and Harbor, page 313] for details and examples.

Gain Compensation - Modified Nyquist Criterion

- Consider the feedback loop shown in the figure. Clearly, the Nyquist diagram of $KG(s)$ is the same as the Nyquist diagram of $G(s)$, but scaled by K.

- It follows that the Nyquist diagram of $KG(s)$ encircles the $-1+j0$ point if and only if the Nyquist diagram of $G(s)$ encircles the $-\frac{1}{K}+j0$ point, and we have the following modified Nyquist criterion.

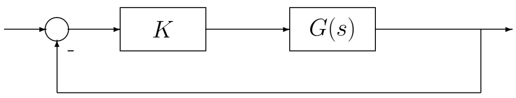
- Consider the feedback loop in the figure. Let
  1. P be the number of unstable poles of $G(s)$.
  2. Z be the number of unstable closed-loop poles.
  3. N be the number of clockwise encirclement of the $-\frac{1}{K}+j0$ point by the Nyquist diagram of $G(s)$.

- Then

$$N = Z - P$$

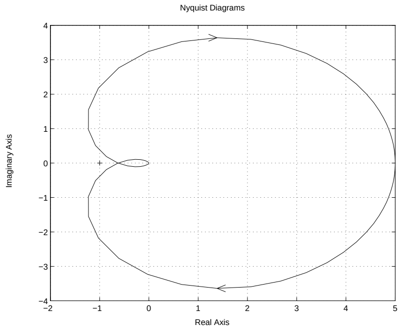
- That is, the closed-loop is stable if and only if

$$-N = P$$



- **Example:** Find the values of (positive and negative) K for which the closed-loop is stable with

$$G(s) = \frac{5}{(s+1)^3} \qquad (\Rightarrow \quad P = 0)$$



1. For $-\frac{1}{K} < -0.625$ ($\Rightarrow K < 1.6$), $N=0$ so $Z=0$.
2. For $-0.625 < -\frac{1}{K} < 0$ ($\Rightarrow 1.6 < K < \infty$), $N=2$ so $Z=2$.
3. For $0 < -\frac{1}{K} < 5$ ($\Rightarrow -\infty < K < -0.2$), $N=1$ so $Z=1$.
4. For $-\frac{1}{K} > 5$ ($\Rightarrow K > -0.2$), $N=0$ so $Z=0$.

- The closed-loop is stable when $-0.2 < K < 1.6$. This is confirmed by the Routh-Hurwitz criterion in the last example.

Relative Stability - Gain and Phase Margins

- Suppose that in Example 2, due to **modeling uncertainty**, the d.c. gain of $G(s)$ is increased by 1.6 ($= 1/0.625$). Then the Nyquist diagram of $1.6 \times G(s)$ will just touch the $-1+j0$ point, and the closed-loop system will be marginally stable.

- We define the **gain margin** as the factor by which the loop gain of a *stable* closed-loop system must be changed to make the system marginally stable.

- The gain margin is a measure of how much extra gain the loop can tolerate without losing stability - the larger the gain margin, the more extra gain the loop can tolerate and the more **robust** the design.

- If the Nyquist diagram of $G(s)$ intersects the negative real axis at $-\alpha$, the gain margin is $1/\alpha$.

- If the Nyquist diagram intersects the negative real axis at more than one point, the gain margin is determined by that point which results in the smallest gain margin.

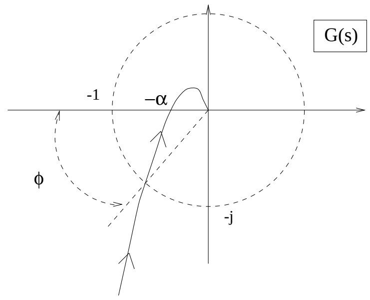
- As a rule of thumb, for many control systems, a gain margin of 2 ($\sim 6dB$) has been observed to give moderately damped response.

- The **phase margin** of a stable closed-loop system is the minimum angle by which the Nyquist diagram must be rotated to intersect the -1 point.

- The phase margin is a measure of how much additional phase-*lag* the loop can tolerate (without changing the gain) before the onset of instability.

- The phase margin is indicated by the angle ϕ in the figure. Note that the magnitude of the Nyquist diagram, $|G(j\omega)|$, is unity at the frequency that the phase margin occurs.

- For many control systems, a phase margin of about $45°$ is considered adequate.



Gain Margin $= 1/\alpha$, Phase Margin $= \phi$

- **Example:** Let

$$G(s) = 4/(s+1)^3.$$

Calculate the gain and phase margins.

- The real axis intercepts can be found using the Routh-Hurwitz criterion as before.

- Alternatively, they can be found by setting the imaginary part of $G(j\omega)$ to zero:

$$\begin{aligned} G(j\omega) &= \frac{4}{(1+j\omega)^3} \\ &= \frac{4}{(1+\omega^2)^3}[1-3\omega^2+j\omega(\omega^2-3)] \end{aligned}$$

This gives intercepts at

$$\omega_i = 0, \pm\sqrt{3}, \infty$$

and so

$$G(j\omega_i) = 4, -0.5, -0.5, 0.$$

- Since the intercept with the negative real axis is at -0.5, the gain margin is 2.

- For the phase margin, we need the intercept with the unit circle centred on the origin. We solve

$$\begin{aligned} |G(j\omega)| &= \frac{4}{(\sqrt{1+\omega^2})^3} \\ &= 1 \end{aligned}$$

This gives

$$\omega_1 = \sqrt{4^{\frac{2}{3}} - 1}$$

and

$$\angle G(j\omega_1) \sim -153^\circ.$$

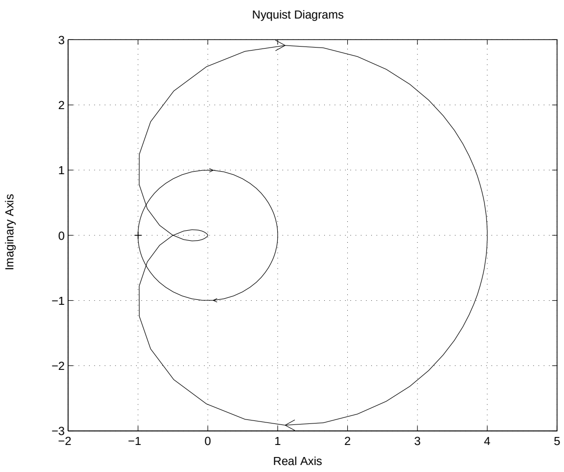
The phase margin is then $\phi \sim 27^\circ$.

- The gain margin is adequate, but the phase margin is too low - further compensation is required.

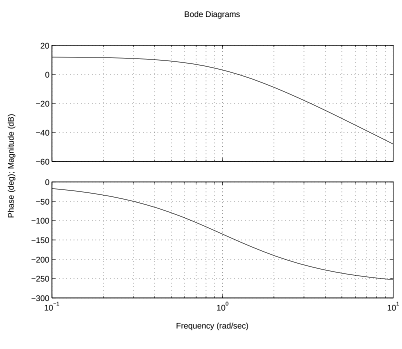
- The following figure shows the Nyquist diagram of

$$G(s) = \frac{4}{(s+1)^3},$$

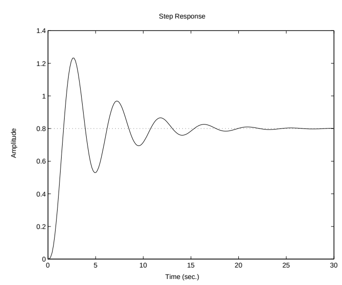
indicating the gain and phase margins.



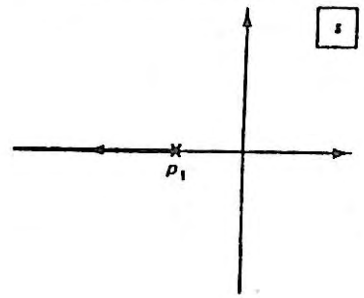
- The next figure shows the bode diagrams for $G(s)$ indicating the gain and phase margins.



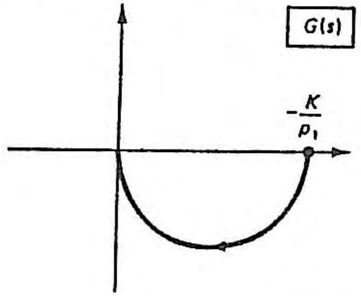
- The next figure shows the closed-loop step response, indicating the need for further compensation.



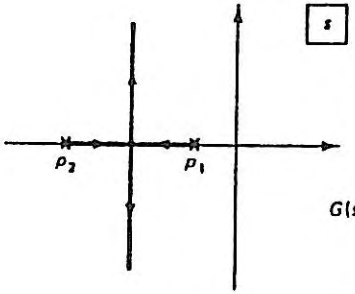
BLE 8.4 COMPARISON OF ROOT LOCI AND NYQUIST DIAGRAMS



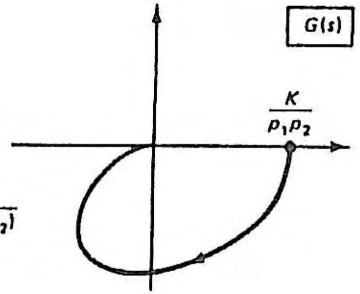
$G(s) = \frac{K}{s - p_1}$



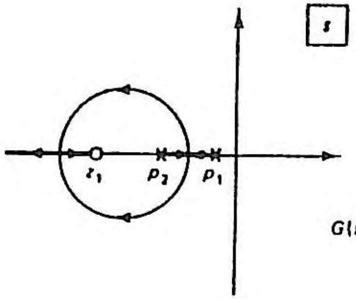
(a)



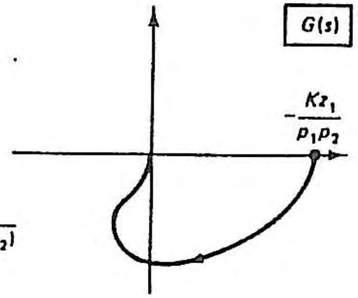
$G(s) = \frac{K}{(s - p_1)(s - p_2)}$



(b)

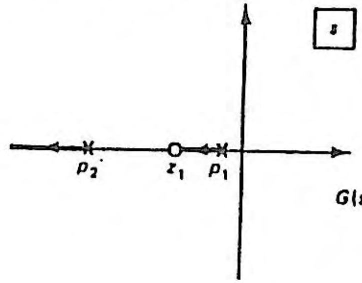


$G(s) = \frac{K(s - z_1)}{(s - p_1)(s - p_2)}$

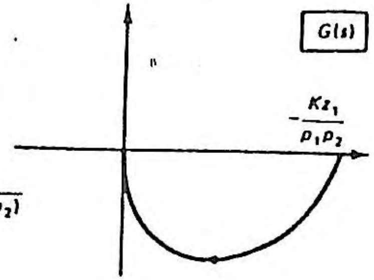


(c)

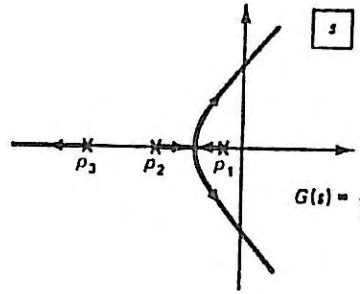
TABLE 8.4 (Continued)



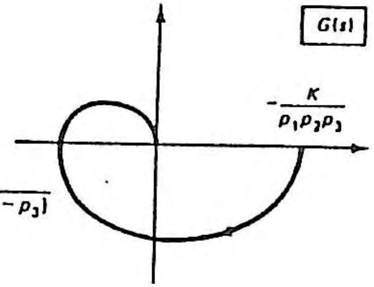
$G(s) = \frac{K(s - z_1)}{(s - p_1)(s - p_2)}$



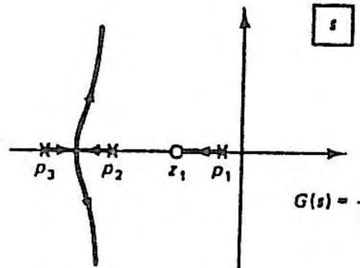
(d)



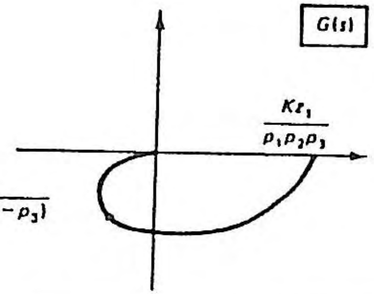
$G(s) = \frac{K}{(s - p_1)(s - p_2)(s - p_3)}$



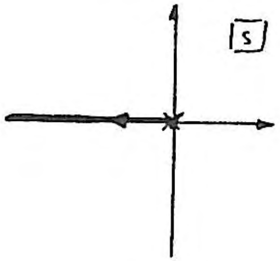
(e)



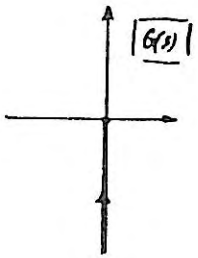
$G(s) = \frac{K(s - z_1)}{(s - p_1)(s - p_2)(s - p_3)}$

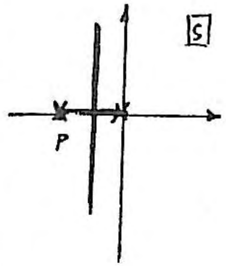


(f)

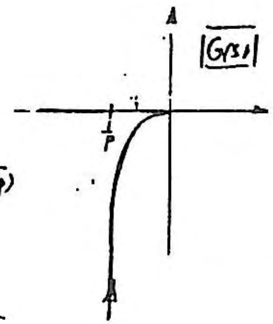


$G(s) = \frac{K}{s}$





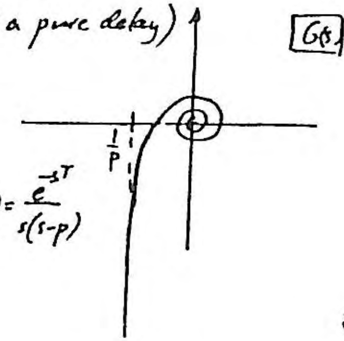
$G(s) = \frac{1}{s(s-p)}$



(An "exotic", with a pure delay)

The root locus is terrifying!

$G(s) = \frac{e^{-sT}}{s(s-p)}$



13 b

Introduction to Frequency Response Design Methods

- A Prototype Feedback System
- Relation between Open-loop and Closed-loop Gain
- Steady-state Accuracy
- Output Disturbance Rejection
- Sensor Noise Attenuation
- Design Trade-offs
- Transient Response & Stability Margins
- Gain Compensation
- Phase-lag Compensation
- Phase-lead Compensation
- Other Types of Compensation
- The Ziegler-Nichols Tuning Rules

A Prototype Feedback System

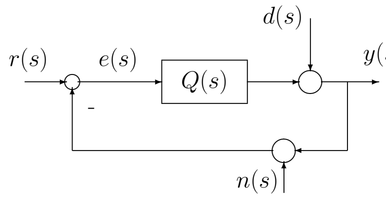
- Consider the feedback system in the figure. Here:

$$\begin{aligned} Q(s) = G(s)K_c(s) &: \text{ open} - \text{loop gain} \\ r(s) &: \text{ reference signal} \\ y(s) &: \text{ output signal} \\ e(s) &: \text{ error signal} \\ d(s) &: \text{ output disturbance} \\ n(s) &: \text{ sensor noise} \end{aligned}$$

- The error and output signals are given by

$$\begin{aligned} e(s) &= \frac{1}{1+Q(s)}\,[r(s)-d(s)-n(s)] \\ y(s) &= \frac{1}{1+Q(s)}\,d(s)+\frac{Q(s)}{1+Q(s)}\,[r(s)-n(s)] \end{aligned}$$

- We consider the following design objectives:
  1. **Steady-state accuracy**.
  2. **Output disturbance rejection**.
  3. **Sensor noise attenuation**.
  4. **Transient Response** & **stability margins**



Relation between Open-loop and Closed-loop Gain

- Define the closed-loop transfer function as

$$H(s)=\frac{Q(s)}{1+Q(s)} \quad \Longrightarrow \quad H(j\omega)=\frac{Q(j\omega)}{1+Q(j\omega)}$$

and the **sensitivity** function as

$$S(s)=\frac{1}{1+Q(s)} \quad \Longrightarrow \quad S(j\omega)=\frac{1}{1+Q(j\omega)}$$

- When the open-loop gain is very large,

$$|Q(j\omega)| \gg 1 \quad \Longrightarrow \quad \begin{cases} |H(j\omega)| \approx 1 \\ |S(j\omega)| \ll 1 \end{cases}$$

- When the open-loop gain is very small,

$$|Q(j\omega)| \ll 1 \quad \Longrightarrow \quad \begin{cases} |H(j\omega)| \approx |Q(j\omega)| \ll 1 \\ |S(j\omega)| \approx 1 \end{cases}$$

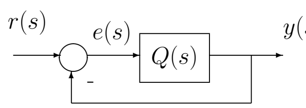
- Note that

$$H(j\omega)+S(j\omega)=1$$

and so we cannot have both $|S(j\omega)|$ and $|H(j\omega)|$ small (or large) at the same frequency.

Steady-state Accuracy

- Consider the simplified loop in the figure. We want y to track r in the steady-state, that is, we want e_{ss} small.



- Now,

$$e(s)=\frac{1}{1+Q(s)}\,r(s)=S(s)r(s)$$

- Suppose that we inject:

$$r(t)=e^{j\omega_0 t} \quad \Longrightarrow \quad r(s)=\frac{1}{s-j\omega_0}$$

Then, *provided the closed-loop is stable*

$$e(s)=\frac{S(s)}{s-j\omega_0}=\frac{S(j\omega_0)}{s-j\omega_0}+A(s)$$

where $A(s)$ is 'stable'.

- The steady-state tracking error is:
$e_{ss}(t)=S(j\omega_0)e^{j\omega_0 t}=|S(j\omega_0)|e^{j[\omega_0 t+\angle S(j\omega_0)]}$

- Thus, $|S(j\omega_0)|$ gives the (steady-state) gain from reference signal r to error signal e at frequency ω_0.
- For good steady-state accuracy, we need

$$|S(j\omega_0)| \ll 1$$

or equivalently,

$$|Q(j\omega_0)| \gg 1$$

- To track sinusoids up to ω_c we need

$$|Q(j\omega)| \gg 1, \quad \omega \leq \omega_c$$

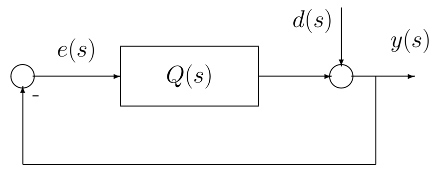
or equivalently,

$$|H(j\omega)| \approx 1, \quad \omega \leq \omega_c$$

- **Good steady-state accuracy requires large open-loop gain over a wide range of frequencies**
- **Equivalently, good steady-state accuracy requires large closed-loop bandwidth**

Output Disturbance Rejection

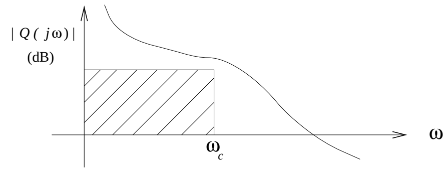
- Consider the simplified loop in the figure.



- Now,

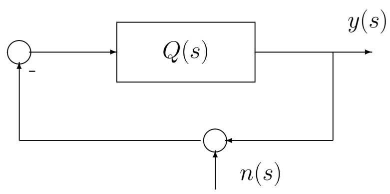
$$y(s) = \frac{1}{1+Q(s)}\, d(s) = S(s)d(s)$$

- It follows that good disturbance rejection is compatible with good steady-state accuracy, both require large open-loop gain, or equivalently, large closed-loop bandwidth.
- The following diagram summarises the open-loop gain requirements for good steady-state tracking and output disturbance rejection.



Sensor Noise Attenuation

- Consider the simplified loop in the figure.



- Here,

$$y(s) = -\frac{Q(s)}{1+Q(s)}\, n(s) = -H(s)n(s)$$

- Let

$$n(t) = e^{j\omega_0 t} \implies n(s) = \frac{1}{s - j\omega_0}$$

- Then, *provided the closed-loop is stable*

$$y_{ss}(t) = H(j\omega_0)e^{j\omega_0 t} = |H(j\omega_0)|e^{j[\omega_0 t + \angle H(j\omega_0)]}$$

- For good sensor noise attenuation we need

$$|H(j\omega_0)| \ll 1$$

or equivalently

$$|Q(j\omega_0)| \ll 1$$

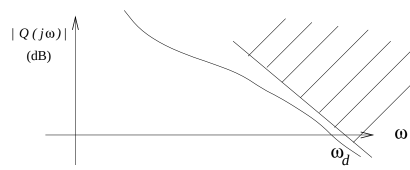
- To attenuate sinusoids beyond ω_d we need

$$|Q(j\omega)| \ll 1, \quad \omega \geq \omega_d$$

or equivalently

$$|H(j\omega)| \ll 1, \quad \omega \geq \omega_d$$

- **Good sensor noise attenuation requires small open-loop gain at high frequencies**
- **Equivalently, good sensor noise attenuation requires small closed-loop bandwidth**
- Note that this is in conflict with the previous two requirements.
- The next diagram summarises the open-loop gain requirements for good sensor noise attenuation.



Design Trade-offs

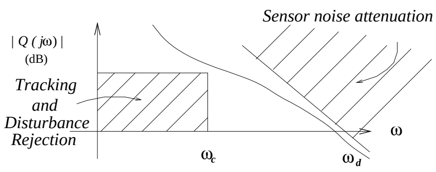
- Good Steady-state accuracy & disturbance rejection are compatible objectives. Both require large open-loop gain (large closed-loop bandwidth).

- Noise attenuation conflicts with these, requiring small open-loop gain (small closed-loop bandwidth).

- For most control systems:
  - References and output disturbances are slow, and therefore low frequency signals.
  - Sensor noise is a fast, high frequency signal.

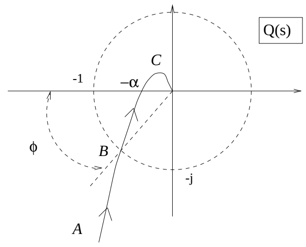
- The conflict can be resolved by requiring
  - Large gain at low frequencies.
  - Small gain at high frequency.

- The following diagram illustrates this **frequency separation** requirement.



Transient Response & Stability Margins

- Our analysis has assumed the closed-loop is stable. The figure shows a typical Nyquist diagram.



- The diagram is divided into three regions:
  - A is the low frequency region ($|Q(j\omega)| \gg 1$).
  - B is the **crossover** region ($|Q(j\omega)| \approx 1$).
  - C is the high frequency region ($|Q(j\omega)| \ll 1$).

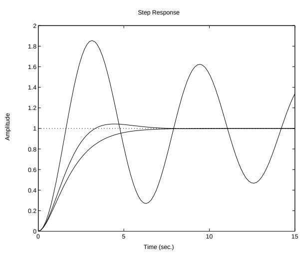
- Large gain at high frequency tends to destabilise the closed-loop: the transient response is more oscillatory, faster and the system is less robust to model uncertainty (small gain margin).

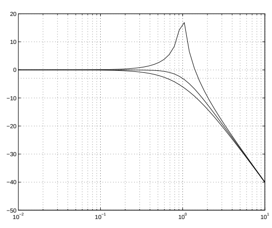
- Moreover, large phase-lag at high frequency has the same effect since the phase margin is reduced.

- Stability sets a limit on the closed-loop bandwidth.

- The figures show step responses and bode plots for the 2nd order system

$$H(s) = \frac{1}{s^2 + 2\zeta s + 1}, \qquad \zeta = 1,\ 0.707,\ 0.05$$



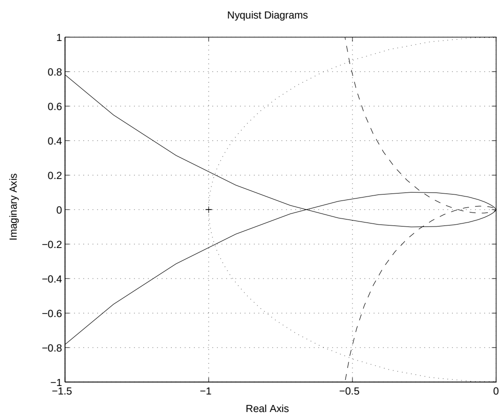


- Note that as ζ tends to zero (towards instability):
  - The response becomes faster (smaller T_r) and more oscillatory (higher %-overshoot).
  - The system bandwidth increases.

Gain Compensation

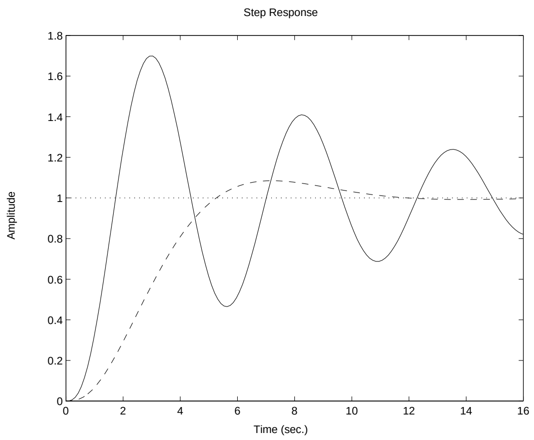
- **Example:** Let $G(s) = 4/s(s+2)(s+4)$. Introducing a constant gain $K_c(s) = K$ increases (or decreases) the gain uniformly for all frequencies.
- Reducing the gain ($K < 1$) will:
  1. increase the stability margins and the rise time
  2. reduce the overshoot and the bandwidth.

- The figure shows the Nyquist diagrams for $K = 1$ (solid) and $K = 0.2$ (dashed). It illustrates the improvement in the stability margins when the gain is reduced.



- The figure shows the step responses for $K = 1$ (solid) and $K = 0.2$ (dashed).

- It illustrates the increase in the rise time (slower response) and the decrease in the overshoot (improved steady state accuracy) as the gain is reduced.



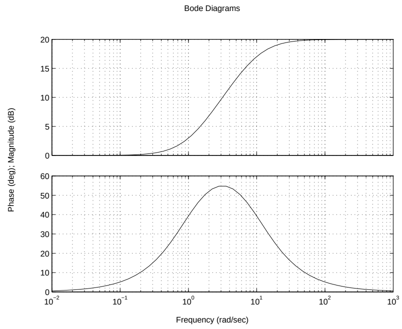
- Improving the response further requires **dynamic compensation** - we need to introduce large gain at low frequency, and small gain at high frequency.

Phase-lead Compensation

- A **phase-lead controller** has transfer function

$$K_c(s) = K\,\frac{1 + s/\omega_0}{1 + s/\omega_p}, \quad \omega_0 < \omega_p$$

- The bode plots (for $K = 1$) are shown below.



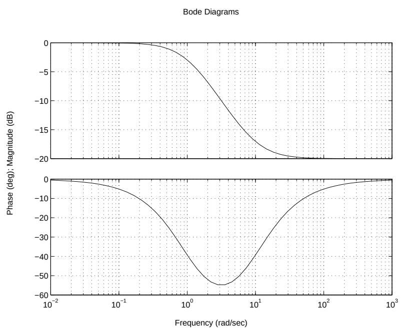
- Phase-lead compensation
  - increases gain at frequencies above ω_p (tending to degrade stability margins)
  - introduces phase-lead between ω_0 and ω_p.
- Since the phase-lead is stabilising, we choose $\omega_0,\ \omega_p$ in the crossover frequency range (region B).
- It is generally difficult to balance the destabilising increase in gain and the stabilising phase-lead.
- See [Phillips and Harbor, page 375] for an example.

Phase-lag Compensation

- A **phase-lag controller** has a transfer function

$$K_c(s) = K\,\frac{1 + s/\omega_0}{1 + s/\omega_p}, \quad \omega_p < \omega_0$$

- The bode plots (for $K = 1$) are shown below.



- Phase-lag compensation
  - reduces gain at frequencies above ω_0 (tending to improve stability margins)
  - introduces phase-lag between ω_p and ω_0.
- Since the phase-lag is destabilising, we choose $\omega_p,\ \omega_0$ in the middle frequency range (region A).
- K is chosen to improve steady-state accuracy by increasing the gain at low frequencies.
- See [Phillips and Harbor, page 367] for an example.

Other Types of Compensation

- Other controllers include:

  - A **lag-lead controller**

    $$K_c(s) = K\frac{1+s/\omega_{0_1}}{1+s/\omega_{p_1}}\,\frac{1+s/\omega_{0_2}}{1+s/\omega_{p_2}},\ \omega_{p_1} < \omega_{0_1}, \omega_{0_2} < \omega_{p_2}$$

    combines phase-lag and phase-lead features.

  - A **PI controller**

    $$K_c(s) = K_p + \frac{K_i}{s} = K_i\,\frac{1 + \dfrac{s}{K_i/K_p}}{s}$$

    is a special form of the phase-lag controller.

  - A **PD controller**

    $$K_c(s) = K_p + K_d s = K_p\left(1 + \frac{s}{K_p/K_d}\right)$$

    is a special form of the phase-lead controller.

  - A **PID controller**

    $$K_c(s) = K_p + \frac{K_i}{s} + K_d s$$

    is a special form of the lag-lead controller.

- See [Phillips and Harbor, page 388] for explanation and design examples.

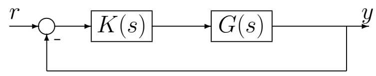
Practical Controller Design
The Ziegler-Nichols Tuning Rules

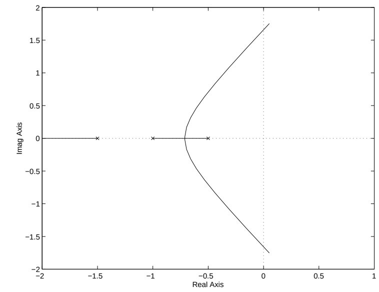
- In the process control industry (steel and paper making; chemical, pharmaceutical and refining industries, etc.), many plants are:
  1. Stable.
  2. Type 0 (no free integrators).
  3. Overdamped (real poles and zeros).
- For such systems, there is a simple practical design technique for tuning a compensator (P, PI or PID) that does not require a model for the plant.
- The Ziegler-Nichols tuning rule is summarised as follows:
  1. Apply a proportional compensator and adjust the gain until the closed loop becomes marginally stable, i.e., when the closed-loop system just starts to oscillate. Let K_{po} be the value of this gain and T_o be the period of oscillations.
  2. The compensator is defined by either:
     – P: $K(s) = 0.5K_{po}$

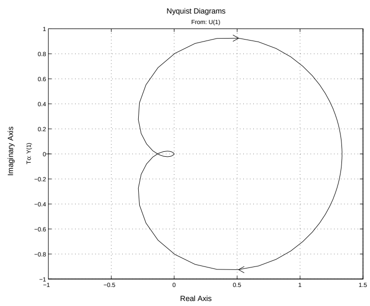
     – PI: $K(s) = 0.45K_{po} + \dfrac{0.54K_{po}/T_o}{s}$

     – PID: $K(s) = 0.6K_{po} + \dfrac{1.2K_{po}/T_o}{s} + 0.075K_{po}T_o s$
- Since the plant is assumed to be type 0, we need integral action for zero steady-state error against a step reference. This rules out a PD compensator.

- **Example:** Consider the feedback loop in the figure and let $G(s) = \frac{1}{(s+0.5)(s+1)(s+1.5)}$. The root-locus and Nyquist diagrams are shown below:





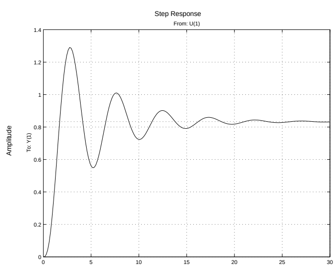


- In practice, the design is carried out experimentally, but since we have a model of the plant, we can do the design analytically.
- To find the critical gain K_{po} we form the Routh-Hurwitz array for the characteristic equation:

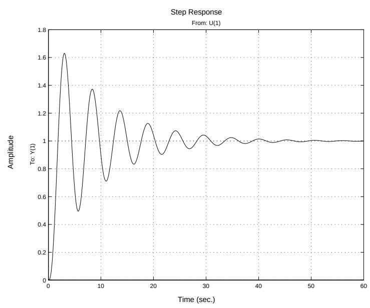
$$1 + K_{po}G(s) = 0 \Rightarrow s^3 + 3s^2 + 2.75s + 0.75 + K_{po} = 0.$$

$$\begin{array}{r|cc} s^3 & 1 & 2.75 \\ s^2 & 3 & 0.75 + K_{po} \\ s & 2.75 - (0.75 + K_{po})/3 & \\ s^0 & 0.75 + K_{po} & \end{array}$$

- Setting the third row to zero: $K_{po} = 7.5$. To find the critical frequency, we set the auxiliary polynomial to zero (second row): $3s^2 + (0.75 + 7.5) = 0$. Thus $s = j\sqrt{8.25/3}$ and so $T_o = 2\pi/\sqrt{8.25/3} = 3.7889$.
- A proportional compensator is then given by $K(s) = 0.5K_{po} = 3.75$. The step response is given below.



- A PI controller is given by $K(s) = 0.45K_{po} + 0.54K_{po}/sT_o = 3.375 + 1.0689/s$. The step response is shown below.



- A PID controller is given by $K(s) = 0.6K_{po} + 1.2K_{po}/sT_o + 0.075K_{po}T_o s = 4.5 + 2.3754/s + 2.1313s$. The step response is given below.