Lecture 1, Jan 8, 2024

Taxonomy of Control Systems

- A control system maintains important process/plant characteristics at desired targets despite external noise, perturbations and uncertainties
- Control systems can be natural or artificial (human-made), manual or automatic; we focus on automatic systems
- *Regulatory control* maintains the plant at a fixed setpoint despite external noise, disturbances, and system uncertainties
- *Tracking (servo) control* tracks the plant's output to a desired trajectory
- *Open-loop* control systems rely only on existing knowledge of the system and not the system's output; issues:
  - System variation: if the system changes, the dynamics will be different, so the controller fails
  - Unsatisfactory dynamics: cannot alter the system dynamics
  - External disturbances and noise: cannot adapt to external disturbances
- *Closed-loop (feedback)* control systems determine the control actions based on the measurements of the plat's output
  - Benefits:
    * *Robustness*: ability to work despite unknown or inaccurate plant models, external disturbances, and noise
      - A robust system is both not very *sensitive* (i.e. doesn't change a lot if system parameters change) and good at *disturbance regulation/rejection* (i.e. eliminating disturbances/noise)
      - Note we usually distinguish between disturbance and noise based on frequency (high frequency disturbances are referred to as noise)
    * Ability to enhance system dynamics and improve *performance* (regulation or tracking)
  - However, they are more complex, and *stability* can be an issue – if the controller is not designed properly, it can work against the goal
- In *classic* feedback control:
  - System parameters are either invariant or varies insignificantly
  - Control actions rely on immediate and not future values of the plant output
  - No guarantees can be made about optimality
  - Relies on a (roughly) linear relationship between input or output, in the operational range or close vicinity of the nominal operating point
    * This implicitly assumes we operate near the nominal output
- In summary, we want our controllers to be *robust*, *stable*, and have the desired transient and steady-state *performance*
- *Devices* are modular components of a process (e.g. sensors); the *process* is the dynamic system we want to control, without the actuator; the *plant* is the process with actuators, but not the controller; the *system* is everything, including plant and controller
  - Note these definitions are not universal

Lecture 2, Jan 11, 2024

Classic Feedback Control Example: Cruise Control

- This is an example of a *regulation* problem since we wish to maintain a fixed speed
- Simplified linear model: at a nominal speed of 60 mph, a 1 degree change in throttle angle (control command u) causes a 10 mph change in speed (output y)
  - Disturbance model: road grade change of 1% causes speed decrease of 5 mph (i.e. half the effect of a 1 degree throttle change)
  - Sensor model: speedometer is accurate and without noise, i.e. we get perfect measurements
- Using open-loop control, we can simply scale down the setpoint by a factor of 10 since that is the plant gain

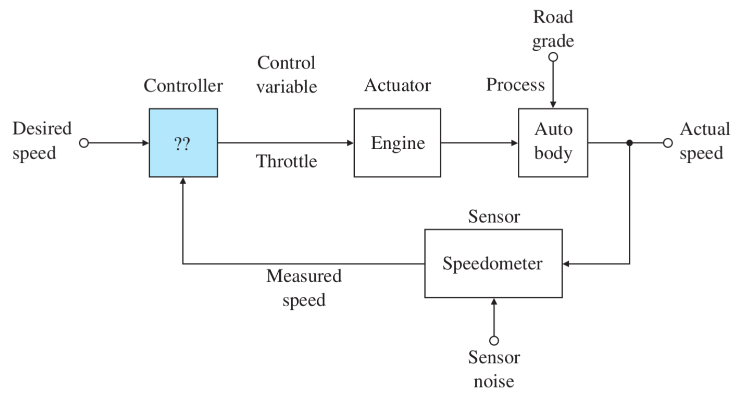


Figure 1: General system block diagram.

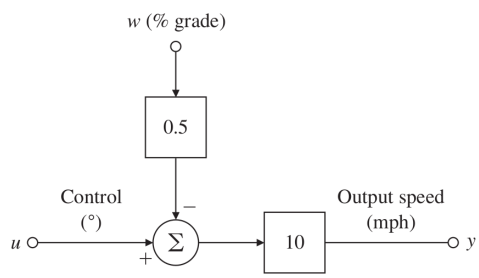


Figure 2: Simplest linearized model of the plant.

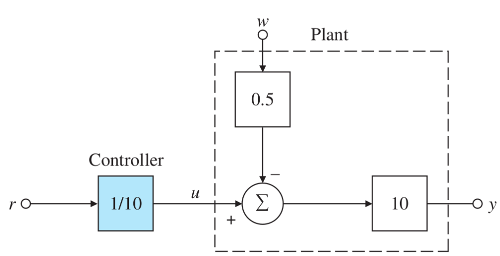


Figure 3: Open-loop controller.

- $y_{ol} = 10(u - 0.5w)$ so if we substitute $y_{ol} = \frac{r}{10}$, $y = r - 5w$
  - If there is no disturbance, the output is perfect, but if there is a road grade change, then we will always be off
  - Define the error as $e_{ol} = r - y_{ol} = 5w$
  - The percentage error is $\frac{e}{r} \times 100\%$
- For open-loop control, if $w \neq 0$, or if we don't know the plant gain exactly or the plant gain changes, then we will have error

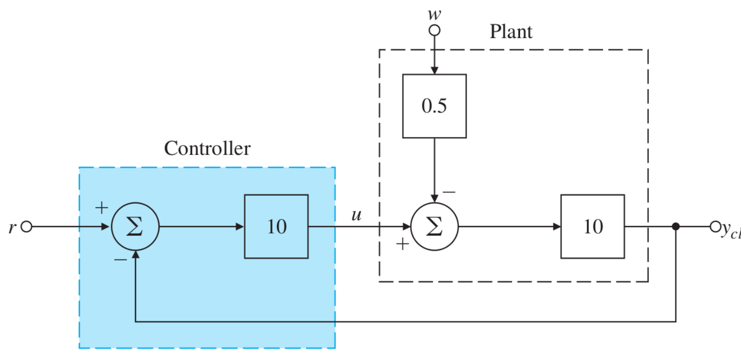


Figure 4: Closed-loop controller.

- Using closed-loop control:
  - $y_{cl} = 10(u - 0.5w) = 10(10(r - y_{cl}) - 0.5w) \implies y_{cl} = \frac{100}{101}r - \frac{5}{101}w$
  - Notice that the factor in front of w is decreased significantly, so we are a lot less susceptible to disturbances
    - If we increase the gain, the effect of noise will be even smaller
  - However, with $w = 0$, we don't have $y_{cl} = r$, but it gets close
  - The error is $e = \frac{1}{100}r$, which reduces if we increase the gain
  - Generally, for a feedback system, we want to increase the gain which would generally decrease the error; however, for many systems increasing the gain makes the transient behaviour worse or even make the system unstable, which holds us back
- The *open-loop gain* of a system is the overall gain of the system (disregarding feedback), often equal to the product of the controller and plant gains
- We want to increase this gain as much as possible, but this involves a tradeoff between decreasing error and getting desirable transient behaviour/stability

System Modelling

- *System*: A collection of components of interest, demarcated by a boundary, interacting through certain physical principles
  - *System parameters* C are properties that define the components of the system
    - e.g. the resistance of a resistor, the mass of an object
  - *State variables* X is the minimal set of variables that completely identify the "state" of the system at each moment
    - The minimality of this set is important!
    - Given system parameters, by knowing the input and state variables, the output can always be identified
    - For many systems, the output are the same as the state variables
    - e.g. positions and velocities of rigid bodies; voltages of nodes and currents through elements
- *Static systems* are where the output $Y(t)$ only depends on the input $U(t)$ at any time t, i.e. the state variables do not change; otherwise the systems are *dynamic*
  - Formulated as $Y(t) = H(U(t), C) \iff y_i(t) = h_i(u_1(t), \dots, u_m(t); c_1, \dots, c_k)$

- When we given a input to a static system at time t_1, we get the output immediately also at time t_1
- Dynamics systems' output also depend on the history of the input
- Example: circuit with R_1 in series with R_2, R_3 in parallel; take output to be the voltage across R_2, R_3
  * The parameters are the resistance values
  * $y(t) = \frac{R_2 R_3}{R_1(R_2 + R_3) + R_2 R_3} u(t)$
  * Since the relationship does not depend on any state variables, the system is static
- Static systems do not have any energy storage or dissipation elements; they don't have "memory"

- *Dynamic systems* are where the current value of the output depends on the past history as well as the present input; we can think of this as the state variables changing through time
  - $Y(t) = H(U(\tau), C), 0 \leq \tau \leq t \iff Y(t) = H(U(t), X(t), t, C)$
    * *Acausal* dynamic systems have outputs that can also depend on future inputs; however all physical systems are causal so we will not worry about this
  - Example: RC circuit
    * System parameters are C and R
    * $y(t) = \frac{1}{C} \int i(t)\, \mathrm{d}t \implies i(t) = C \frac{\mathrm{d}y}{\mathrm{d}t}$
    * We can form an ODE $y(t) + RC \frac{\mathrm{d}y}{\mathrm{d}t} u(t)$
    * Assuming capacitor starts uncharged, $y(t) = u(t) \left(1 - e^{-\frac{t}{RC}}\right)$
- A *system model* is a simplified representation (or abstraction) of a physical system
  - A complete/universal model is often impossible and unnecessary, so in system modelling we abstract away certain details
  - An effective model is the simplest model that does the job
  - Modelling is useful for conceptual analysis (of the problem), controller design, and detailed simulation for verification (whether the model is correct) and validation (whether the design works in the real world) etc

Lecture 3, Jan 15, 2024

Taxonomy of System Models

- Dynamic system models can be physical (i.e. a prototype), mathematical (e.g. simulations) or hybrid (HITL simulation)
- Mathematical models can be one of 3 types:
  - Observation-based (experimental): black-box approach
    * Model is developed only based on empirically observed input-output relations
    * Used when the system's internal dynamics are unknown or too complex to model, e.g. biological systems
  - Knowledge-based: white-box approach
    * Using known dynamics of the system to model it
  - Hybrid: grey-box approach
    * Model is based on both empirical input-output relations and knowledge of the system
    * DEs are used for the model and parameters are identified by experiments
- Within knowledge-based systems, models can be qualitative (expert systems, a set of if-then rules using fuzzy sets/logic) or quantitative (analytical/using DEs, numerical algorithms/simulation, graphical diagrams)
- This course will focus on dynamic, mathematical, knowledge-based, quantitative, analytical models in the time, frequency and Laplace domains
- Systems can be lumped- or distributed-parameter
  - Most physical systems have parameters that are continuously distributed, e.g. a real spring has mass, stiffness, and damping distributed in all 3 axes
  - For such distributed-parameter systems, not only do we need to model the dynamics in time, but

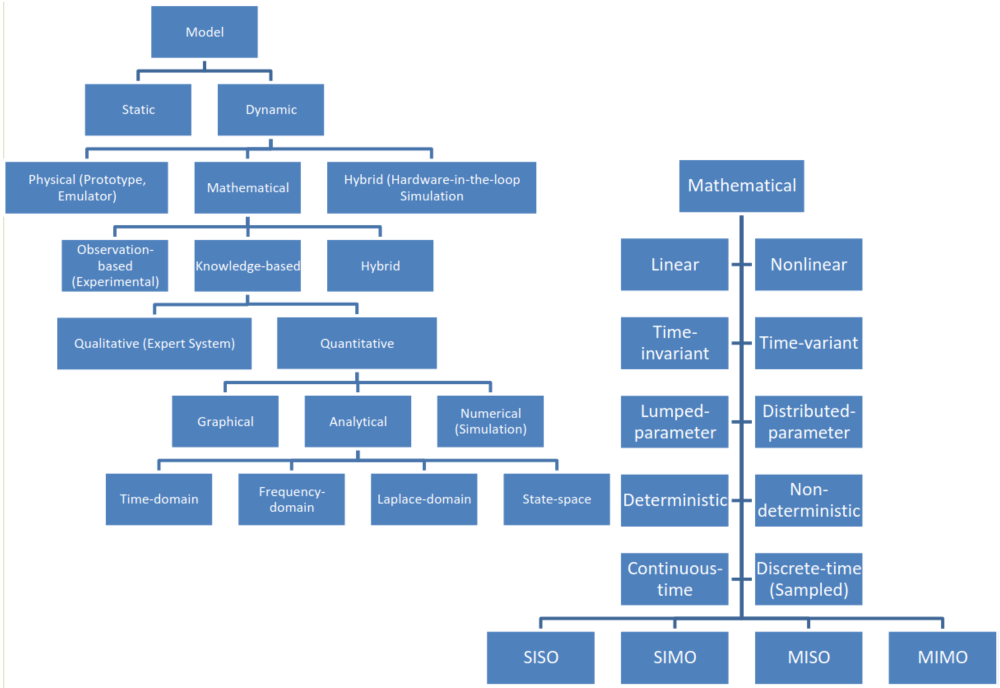


Figure 5: Taxonomy of system models.

also the distribution of properties in space
    * This means using PDEs instead of ODEs and makes things much more complicated
  - Lumped-parameter systems approximate the real physical system with a discrete number of parameters, e.g. reducing a real spring to a point mass, ideal spring, and dashpot
  - When reducing systems and simplifying them, make sure to state all assumptions
- Systems can be linear or nonlinear
  - Physical systems are generally nonlinear, but the relation between input and output can be locally linear within a narrow range (*smooth nonlinearity*)
  - A linear model can approximate the system well in this case if the operation stays within a small range
  - Given a general nonlinear model $Y = \dot{X} = F(X, U)$, we wish to obtain an approximate linear model $\dot{X} = AX + BU$ where A, B are constants, about an operating point X_0, U_0; this can be achieved using a Taylor series
  - Linear systems are desirable due to the principle of superposition
  - If the coefficient on u is 1 (or identity), we can also reorder systems connected in series (principle of interchangeability)
- In time-invariant systems, system parameters stay the same regardless of time, so the input-output relationship does not change in time
  - For a constant time delay T, if $u(t)$ gives output $y(t)$, then $u(t-T)$ gives output $y(t-T)$
  - Practically to check for time invariance, we will give the system some input then delay the output by some time, and compare this against the system's output for a delayed input
- Deterministic models assume that nominal values of inputs, initial conditions, states and parameters (and thus the output) can all be identified without random deviations
  - In probabilistic models, inputs, initial conditions, and/or parameters may vary according to a PDF (noisy), but state variables are still deterministic
    * Note static models can be probabilistic
  - In stochastic models, the system state can vary as well

Lecture 4, Jan 18, 2024

Dynamic System Modelling

- Electrical, mechanical, fluid and thermal systems can be represented by analogous models, regardless of the underlying system, by taking an energy perspective
- We divide basic elements into two groups: energy storage and energy dissipation; within energy storage, elements can be either capacitive or inductive
- Each element is defined by either a *through variable* (aka *t-type*, a property that appears to flow through the element unaltered), or *across variable* (aka *a-type*, a property that is measured as a difference at the two ends of the element)
  - Capacitive elements are represented by t-type variables; inductive elements are represented by a-type variables
  - All energy dissipation elements are represented by t-type variables
- Sometimes we might want to use the integrated version of the t-type and a-type variables

System	Through Variable	Integrated Through Variable	Across Variable	Integrated Across Variable
Electrical	Current, i	Charge, q	Voltage difference, v_{21}	Flux linkage, λ_{21}
Mechanical translational	Force, F	Translational momentum, P	Velocity difference, v_{21}	Displacement difference, y_{21}
Mechanical rotational	Torque, T	Angular momentum, h	Angular velocity difference, ω_{21}	Angular displacement difference, θ_{21}
Fluid	Fluid volumetric rate of flow, Q	Volume, V	Pressure difference, P_{21}	Pressure momentum, γ_{21}
Thermal	Heat flow rate, q	Heat energy, H	Temperature difference, $\mathcal{T}_{21}$	

Figure 6: A-type and t-type state variables for the four types of systems.

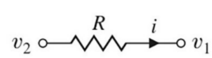
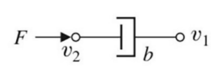
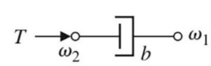
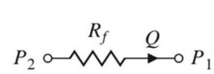
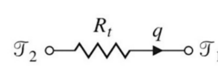
Physical Element	Governing Equation	Energy E or Power $\mathcal{P}$	Symbol
Electrical resistance	$i = \frac{1}{R}v_{21}$	$\mathcal{P} = \frac{1}{R}{v_{21}}^2$	
Translational damper	$F = bv_{21}$	$\mathcal{P} = b{v_{21}}^2$	
Rotational damper	$T = b\omega_{21}$	$\mathcal{P} = b{\omega_{21}}^2$	
Fluid resistance	$Q = \frac{1}{R_f}P_{21}$	$\mathcal{P} = \frac{1}{R_f}{P_{21}}^2$	
Thermal resistance	$q = \frac{1}{R_t}\mathcal{T}_{21}$	$\mathcal{P} = \frac{1}{R_t}\mathcal{T}_{21}$	

Figure 7: Energy dissipation elements.

- Note this is referred to as a force-current analogy; alternatively we can have a force-voltage analogy instead
- Example: cruse control model

Physical Element	Governing Equation	Energy E or Power $\mathcal{P}$	Symbol
Electrical capacitance	$i = C\dfrac{dv_{21}}{dt}$	$E = \dfrac{1}{2}Cv_{21}{}^2$	v_2, i, C, v_1
Translational mass	$F = M\dfrac{dv_2}{dt}$	$E = \dfrac{1}{2}Mv_2{}^2$	F, v_2, M, $v_1 =$ constant
Rotational mass	$T = J\dfrac{d\omega_2}{dt}$	$E = \dfrac{1}{2}J\omega_2{}^2$	T, ω_2, J, $\omega_1 =$ constant
Fluid capacitance	$Q = C_f\dfrac{dP_{21}}{dt}$	$E = \dfrac{1}{2}C_fP_{21}{}^2$	Q, P_2, C_f, P_1
Thermal capacitance	$q = C_t\dfrac{d\mathcal{T}_2}{dt}$	$E = C_t\mathcal{T}_2$	q, $\mathcal{T}_2$, C_t, $\mathcal{T}_1 =$ constant

Figure 8: Capacitive energy storage elements.

Physical Element	Governing Equation	Energy E or Power $\mathcal{P}$	Symbol
Electrical inductance	$v_{21} = L\dfrac{di}{dt}$	$E = \dfrac{1}{2}Li^2$	v_2, L, i, v_1
Translational spring	$v_{21} = \dfrac{1}{k}\dfrac{dF}{dt}$	$E = \dfrac{1}{2}\dfrac{F^2}{k}$	v_2, k, v_1, F
Rotational spring	$\omega_{21} = \dfrac{1}{k}\dfrac{dT}{dt}$	$E = \dfrac{1}{2}\dfrac{T^2}{k}$	ω_2, k, ω_1, T
Fluid inertia	$P_{21} = I\dfrac{dQ}{dt}$	$E = \dfrac{1}{2}IQ^2$	P_2, I, Q, P_1

Figure 9: Inductive energy storage elements.

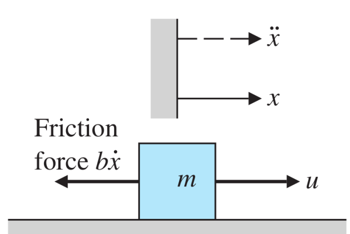


Figure 10: Free body diagram for the cruise control example.

- We apply a force u to the car of mass m, which has a resistive force proportional to the speed
- We want to know how the speed of the car varies in time
- Assumptions:
  * Car is a rigid body
  * Rotational inertia of the wheels is negligible
  * Friction/drag is proportional to speed with a factor of b
- $F = ma \implies u - b\dot{x} = m\ddot{x} \implies \ddot{x} + \frac{b}{m}\dot{x} = \frac{u}{m}$
- Change variable to v: $\dot{v} + \frac{b}{m}v = \frac{u}{m}$
- Typically, we rearrange the system to put all the outputs on the left and all the inputs on the right
- We get a first order linear ODE

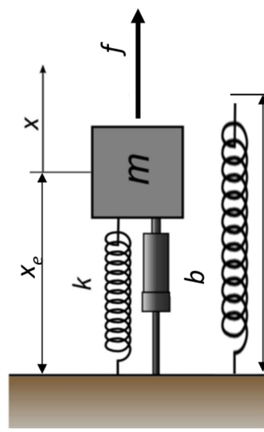


Figure 11: Mass-spring-damper example.

- Example: mass-spring-damper system
  - The input force f is applied at time 0; we want to know how x (measured from equilibrium) varies in time as a result of this force
  - x_e is the equilibrium position of the mass with no force applied; x_0 is the uncompressed length of the spring
  - In equilibrium, $k(x_0 - x_e) = mg$
  - The full FBD would have the external force m upwards, the spring force $k(x_0 - (x_e + x))$ upwards, the gravitational force mg downwards, the damping $b\dot{x}$ downwards
  - $k(x_0 - x_e - x) - b\dot{x} - mg + f = m\ddot{x}$
    * Notice that the equilibrium condition means the $k(x_0 - x_e)$ cancels with mg, so we have no g term in the final expression
  - Final ODE: $m\ddot{x} + b\dot{x} + kx = f$ (second order linear ODE)
  - In general, in mechanical systems moving around their equilibrium state, the holding (static) forces and moments required for maintaining the equilibrium do not contribute to the motion state
    * In this example, the spring force and gravity at equilibrium are the holding forces
    * Therefore we don't have x_0, x_e or g in the model
- Example: automobile suspension system
  - Each wheel of the car is equipped with a suspension system
    * The tire itself acts like a spring
    * The suspensions system consists of a spring and a dashpot
  - Consider the car moving on a road with some profile; we wish to model the vertical movement of the car body
  - This is a single input, two output system because we also need to model the movement of the wheel itself to get the movement of the car body

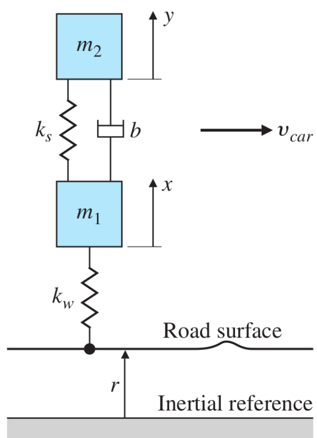


Figure 12: Automotive suspension system example.

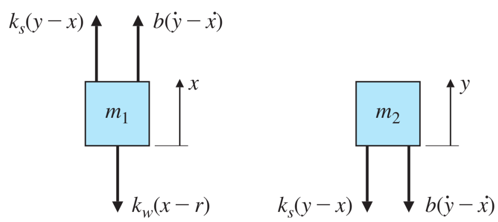


Figure 13: Free body diagram for the example.

- Drawing free body diagrams around the equilibrium allows us to ignore gravity and consider only the forces by the springs and dashpots
- Dynamic equations: $\begin{cases} b(\dot{y}-\dot{x}) + k_s(y-x) - k_w(x-r) = m_1\ddot{x} \\ -k_s(y-x) - b(\dot{y}-\dot{x}) = m_2\ddot{y} \end{cases}$
- Rearrange: $\begin{cases} \ddot{x} + \dfrac{b}{m_1}(\dot{x}-\dot{y}) + \dfrac{k_s}{m_1}(x-y) + \dfrac{k_w}{m_1}x = \dfrac{k_w}{m_1}r \\ \ddot{y} + \dfrac{b}{m_2}(\dot{y}-\dot{x}) + \dfrac{k_s}{m_2}(y-x) = 0 \end{cases}$

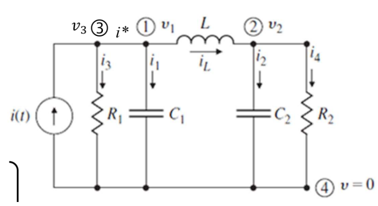


Figure 14: KCL electrical system example.

- Example: electrical system with KCL
  - Node 1: $i^* = i_1 + i_L$
  - Node 2: $i_L = i_2 + i_4$
  - Node 3: $i(t) = i_1 + i_3 + i_L$
  - $i(t) = \dfrac{v_1}{R_1} + C_1\dfrac{\mathrm{d}v_1}{\mathrm{d}t} + i_L$
  - $i_L = C_2\dfrac{\mathrm{d}v_2}{\mathrm{d}t} + \dfrac{v_2}{R_2}$

- $v_1 - v_2 = L\frac{\mathrm{d}i_L}{\mathrm{d}t}$
- $LC_1C_2\frac{\mathrm{d}^3v_2}{\mathrm{d}t^3} + \left(\frac{LC_1}{R_2} + \frac{LC_2}{R_1}\right)\frac{\mathrm{d}^2v_2}{\mathrm{d}t^2} + \left(\frac{L}{R_1R_2} + C_1 + C_2\right)\frac{\mathrm{d}v_2}{\mathrm{d}t} + \left(\frac{1}{R_1} + \frac{1}{R_2}\right) + i(t)$
- This ends up being a third order linear ODE

Lecture 5, Jan 22, 2024

More Dynamic System Examples

- Example: pendulum with point mass m under gravity, input torque T_c around the pivot
  - $T_c - mgl\sin\theta = I\ddot{\theta} \implies \ddot{\theta} + \frac{g}{l}\sin\theta = \frac{T_c}{ml^2}$
    - * We can do this since all the moments act on the same axis in planar motion
  - Linearize using $\sin\theta \to \theta$ for small angles
    - * These assumptions need to be checked later
  - In free oscillation this would oscillate around 0 with $\omega_n = \sqrt{\frac{g}{l}}$
  - If T_c is constant, then this gives a constant bias angle to the oscillation

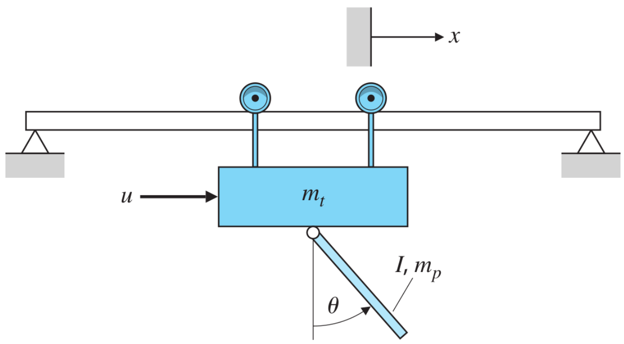


Figure 15: Example: Crane with hanging load.

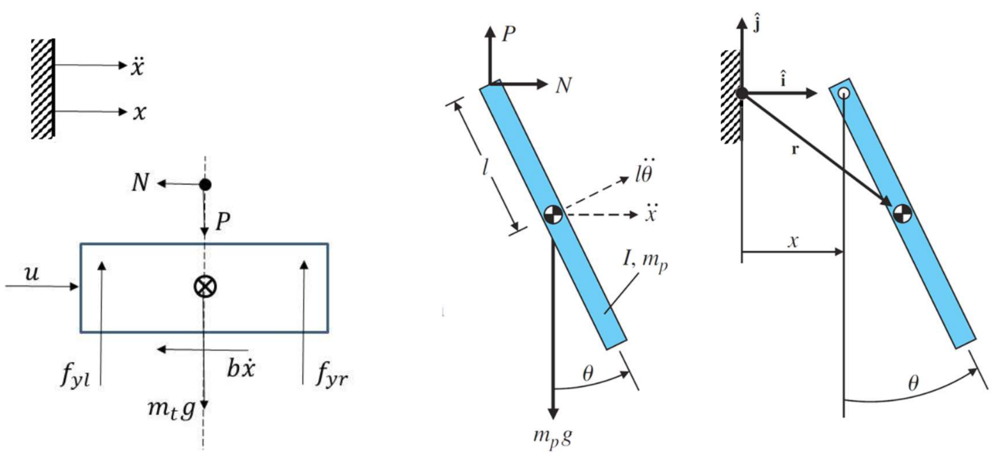


Figure 16: Free body diagram for the example.

- Example: consider a pendulum on a cart as shown above
  - The cart is subject to an applied force and viscous friction (also gravity and normal forces which cancel)
    - * The pendulum also applies unknown forces to the cart at the hinge position

        * $u - N - b\dot{x} = m_t\ddot{x}$
    - For the pendulum we write the equations about the centre of mass for both forces and moments, since we don't have a fixed point on the body anymore
        * To find the accelerations of the centre of mass we differentiate its position
            • $\boldsymbol{r} = x\hat{\imath} + l(\hat{\imath}\sin\theta - \hat{\jmath}\cos\theta)$
            • $\ddot{\boldsymbol{r}} = \ddot{x}\hat{\imath} + l\ddot{\theta}(\hat{\imath}\cos\theta + \hat{\jmath}\sin\theta) - l\dot{\theta}^2(\hat{\imath}\sin\theta - \hat{\jmath}\cos\theta)$
        * $N = m_p\ddot{x} + m_pl\ddot{\theta}\cos\theta - m_pl\dot{\theta}^2\sin\theta$
        * $P - m_pg = m_pl\sin\ddot{\theta} + m_pl\dot{\theta}^2\cos\theta$
        * $-Pl\sin\theta - Nl\cos\theta = I\ddot{\theta}$
    - Clean up to get $\begin{cases}(m_t + m_p)\ddot{x} + b\dot{x} + m_pl\ddot{\theta}\cos\theta - m_pl\dot{\theta}^2\sin\theta = u \\ (I + m_pl^2)\ddot{\theta} + m_pgl\sin\theta + m_pl\ddot{x}\cos\theta = 0\end{cases}$
    - We can linearize this around $\theta = 0$ for a crane system, or around $\theta = \pi$ for a segway (inverted pendulum) system, assuming $\theta, \dot{\theta}$ are small

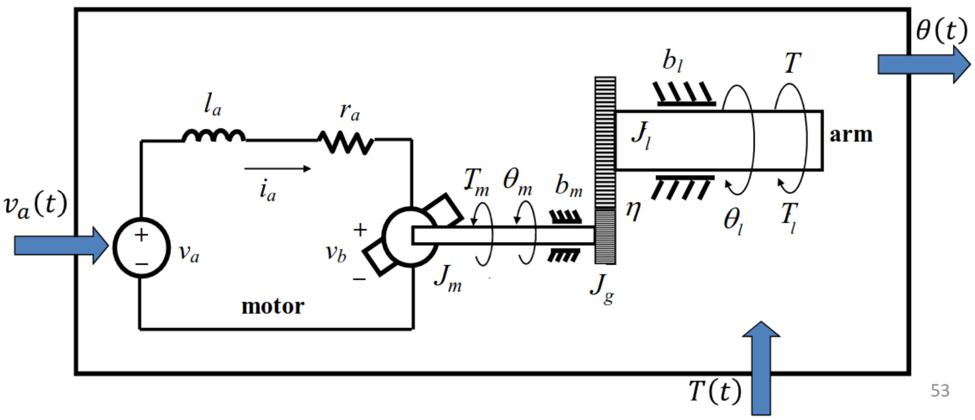


Figure 17: Mechatronics diagram of the DC brushed motor model.

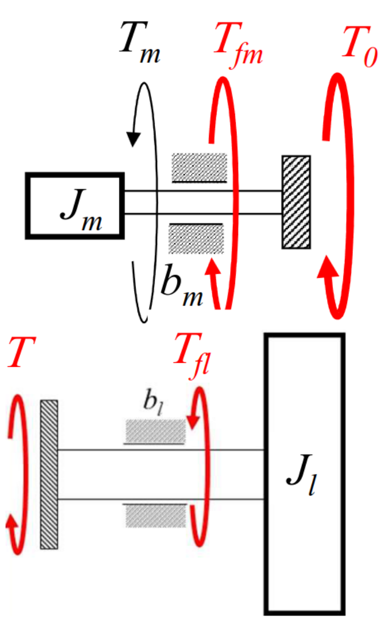


Figure 18: Free body diagrams for the two shafts.

• Example: DC brushed motor
    - The brushes power the rotor and provide the polarity switching needed to maintain a constant

rotation in one direction

- The electric part can be obtained through KVL: $l_a \frac{\mathrm{d}i_a}{\mathrm{d}t} + r_a i_a = v_a - K_e\dot{\theta}_m$
  * Lorentz law: $T_m = K_1\varphi i_a = K_l i_a$ where φ is determined by the field strength, i_a is the current and K_1 is determined by the loop geometry
  * Faraday law: $v_b = K_2\varphi\dot{\theta}_m = K_e\dot{\theta}_m$ where v_b is the back-EMF, a voltage
- The mechanical part has 2 bodies: the shaft directly connected to the rotor and the output arm, connected to the load
  * The motor generates a torque $T_m(t)$, which has to overcome a resistive torque load $T(t)$ to give the final output $\theta_m(t)$
  * For the smaller shaft: $T_m - b_m\dot{\theta}_m - T_0 = J_m\ddot{\theta}_m$
    - T_0 is some torque applied by the larger output shaft
    - $b_m\dot{\theta}_m$ is a viscous friction torque
    - J_m is the moment of inertia of the shaft
  * For the larger shaft: $T - b_l\dot{\theta}_l = J_l\ddot{\theta}_l$
  * θ_l and θ_m are related as $\theta_l = \eta\theta_m$ and $T_0 = \eta T$ since $r_m\theta_m = r_l\theta_l$
    - η is the gear ratio, $\eta = \frac{r_m}{r_i}$
    - Note we always use the smaller radius in the numerator so the gear ratio is always 1 or less
  * This gives $J_m\ddot{\theta}_m + b_m\dot{\theta}_m = T_m - \eta T = K_l i_a - \eta T = K_l i_a - \eta J_l\ddot{\theta}_l - \eta b_l\dot{\theta}_l$
  * Expand and rearrange: $(J_m + \eta^2 J_l)\ddot{\theta}_m + (b_m + \eta^2 b_l)\dot{\theta}_m = K_t i_a$
- Now we can solve for i_a from the mechanical system and its derivative and substitute into the first equation
- This gives us a third order linear ODE, however we can reduce this by noting that the time constant of the electric part is much smaller than the time constant of the mechanical part (i.e. after a voltage change, the current stabilizes much faster than the motor speed), which allows us to ignore the inductance of the coil and reduce the electric part to a static system, which reduces the overall differential equation to 2nd order

Lecture 6, Jan 25, 2024

Linear Time-Invariant Systems

- *Zero state response*: the response of a system to some input when the system is initially "at rest", i.e. all inputs, outputs, states and their derivatives are initially zero
  - When we talk about linear systems, we are usually assuming zero-state
- The most important property of linear systems is homogeneity and superposition – we can scale and add inputs and the outputs will scale and add accordingly
- In a time-invariant system the parameters C are constant in time, so delaying the input will delay the output by the same amount and leave it otherwise unchanged
  - This also works in reverse – if the system output remains the same but delayed when the input is delayed, then the system is time-invariant (we can show that this implies that C is constant)
- These properties let us determine the response of a system to any general input by only knowing its impulse response
  - Any general input $u(t)$ can be approximated by a series of pulses $p_\Delta(t) = \begin{cases} \frac{1}{\Delta} & 0 \leq t \leq \Delta \\ 0 & \text{otherwise} \end{cases}$
    * The input at $t = k\Delta$ has a value $u(t) = u(k\Delta)$, so we can approximate this as $u(k\Delta) \cdot \Delta \cdot p_\Delta(t - k\Delta)$
      - Note we multiply by Δ so the integral remains the same
    * If the system has a response $h_\Delta(t)$ to $p_\Delta(t)$, then due to homogeneity and time-invariance the response to the above input is $y(t) = y(n\Delta) = u(k\Delta) \cdot \Delta \cdot h_\Delta(n\Delta - k\Delta)$
    * Then the total response to all the pulses is $y(t) = \sum_{k=0}^{\infty} u(k\Delta) \cdot \delta \cdot h_\Delta(t - k\Delta)$
  - In the limit, $p_\Delta(t)$ becomes the Dirac delta function $\delta(t)$ (or *unit impulse function*); $h_\Delta(t)$ becomes

the *impulse response* $h(t)$

- Therefore the output is a convolution: $y(t) = \int_0^\infty u(\tau)h(t-\tau)\,\mathrm{d}\tau = u(t) * h(t)$
    * Formally the convolution integral should be from $-\infty$, however we consider the zero-state response so we don't need to consider $t < 0$
    * Furthermore, if $t - \tau < 0$, we would be considering negative time for $h(t)$, which makes no sense for a causal system (in other words $y(t)$ would depend on values of the input in the future); therefore our upper bound is t instead of ∞

- Note this only applies to LTI systems, or upon linearization assuming a small input region

Summary

The response of an LTI system to any arbitrary input $u(t)$ is given by

$$y(t) = \int_0^t u(\tau)h(t-\tau)\,\mathrm{d}\tau = u(t) * h(t)$$

where $h(t)$ is the response of the system to the unit impulse $\delta(t)$.

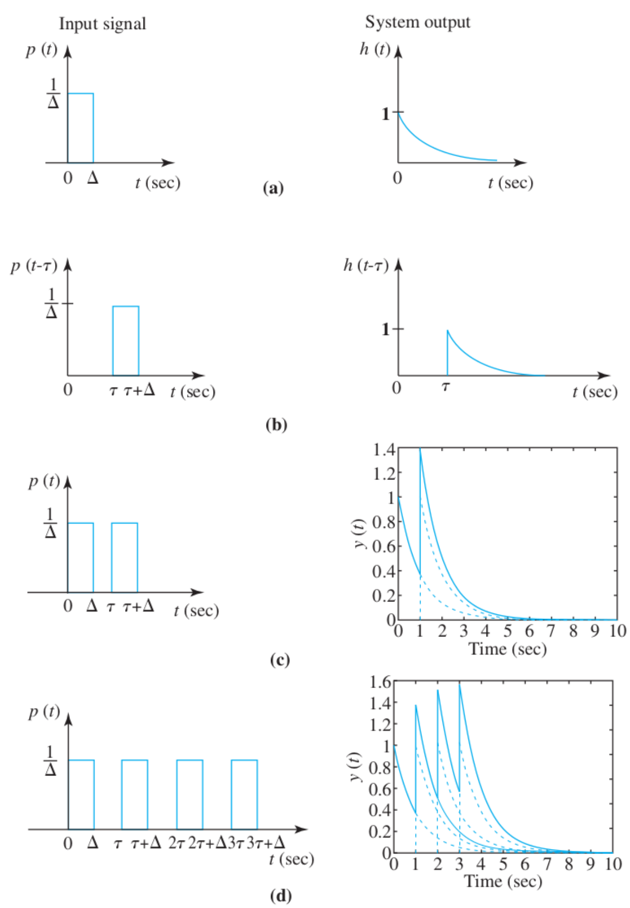


Figure 19: Response of the system to a series of pulses.

- Note convolution has the following properties:
    - Commutativity: $x_1(t) * x_2(t) = x_2(t) * x_1(t)$
        * Obtained by a simple change of variables
    - Associativity: $x_1(t) * [x_2(t) * x_3(t)] = [x_2(t) * x_2(t)] * x_3(t)$

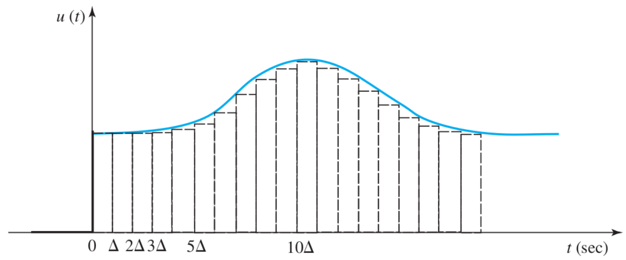


Figure 20: Approximation of any input function as a series of impulses.

  - Distributivity: $x_1(t) * [x_2(t) + x_3(t)] = x_1(t) * x_2(t) + x_2(t) * x_3(t)$
  - Shift: $x_1(t) * x_2(t-T) = x_1(t-T) * x_2(t)$
    * $x_1(t) * x_2(t) = y(t) \implies x_1(t-T_1) * x_2(t-T_2) = y(t-T_1-T_2)$
  - Impulse: $x(t) * \delta(t) = \displaystyle\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau) = x(t)$
  - Width: the convolution of a function covering a length of time T_1 and another function covering T_2 covers a time of $T_1 + T_2$
- Example: find the impulse response of the following system, with $y(0^-) = 0$: $\dot{y} + ky = u(t)$
  - $\displaystyle\int_{0^-}^{0^+} \dot{y}\,\mathrm{d}t + k\int_{0^-}^{0^+} y\,\mathrm{d}t = \int_{0^-}^{0^+} \delta(t)\,\mathrm{d}t$
    * The second term goes to zero since y is a continuous function
    * The right hand side is by definition 1
  - $\displaystyle\int_{0^-}^{0^+} \dot{y}\,\mathrm{d}t = 1 \implies y(0^+) - y(0^-) = 1 \implies y(0^+) = 1$
  - Now we use the model of the system to find other times, which gives $y = Ae^{\alpha t}$
    * $A\alpha e^{\alpha t} + kAe^{\alpha t} = 0 \implies \alpha = -k$
    * $y(0^+) = 1 \implies A = 1$
  - This gives $y(t) = h(t) = e^{-kt}1(t)$ where $1(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$ is the Heaviside step function (sometimes denoted $u(t)$)
    * We need the $1(t)$ because for $t < 0$ we assumed zero-state
  - For a general input $u(t)$, $y(t) = \displaystyle\int_0^{\infty} e^{-k\tau}u(t-\tau)\,\mathrm{d}\tau$ or $\displaystyle\int_0^{t} e^{-k\tau}u(t-\tau)\,\mathrm{d}\tau$ for a causal system
    * The Heaviside function is gone because our bound starts at 0, so it is 1 for the entire integration range

Lecture 7, Jan 29, 2024

Laplace Transform

Definition

The *Laplace transform* for a generic function $f(t)$ is defined as

$$F(s) \equiv \int_{-\infty}^{\infty} f(t)e^{-s}\,\mathrm{d}t$$

The *unilateral* (one-sided) Laplace transform is defined as

$$F(s) = \mathcal{L}\left\{f(t)\right\} \equiv \int_{0^-}^{\infty} f(t)e^{-st}\,\mathrm{d}t$$

where $s = \sigma + j\omega$ is a complex frequency variable with units of inverse time.

- The Laplace transform transforms linear ODEs into algebraic equations
- For our purposes since we only consider $t \geq 0$, we consider all functions to be 0 for $t < 0$ and so the unilateral transform suffices
- $F(s)$ exists (i.e. the integral converges) if for all $\mathrm{Re}(s) > \alpha$ we have $|f(t)| < Me^{\alpha t}$ for all $s \in \mathbb{C}, M \in \mathbb{R}$, i.e. $f(t)$ grows slower than exponential
  - When multiplying transforms, the output is only valid for values of s in the intersection of the regions of convergence
- Some examples:
  - Unit step: $\mathcal{L}\left\{1(t)\right\} = \int_0^{\infty} 1(t)e^{-st}\,\mathrm{d}t = -\frac{1}{s}\left[e^{-st}\right]_0^{\infty} = \frac{1}{s}$
  - Unit impulse: $\mathcal{L}\left\{\delta(t)\right\} = \int_{0^-}^{\infty} \delta(t)e^{-st}\,\mathrm{d}t = e^{-st}\big|_{t=0} = 1$
    - Note that we had to start at 0^- to include 0 in the integration region
  - Exponential: $\mathcal{L}\left\{e^{\alpha t}\right\} = \int_0^{\infty} e^{\alpha t}e^{-st}\,\mathrm{d}t = \int_0^{\infty} e^{-(s-a)t}\,\mathrm{d}t = -\frac{1}{s-a}\left[e^{-(s-a)t}\right]_0^{\infty} = \frac{1}{s-a}$
    - Note we need to assume $\mathrm{Re}(s) > \mathrm{Re}(a)$ so that the exponent has a negative real part
  - Sinusoid:
$$\begin{aligned}\mathcal{L}\left\{\cos(\omega t)\right\} &= \int_0^{\infty} \cos(\omega t)e^{-st}\,\mathrm{d}t \\ &= \int_0^{\infty} \frac{e^{j\omega t} + e^{-j\omega t}}{2}e^{-st}\,\mathrm{d}t \\ &= -\frac{1}{2(s-j\omega)}\left[e^{-(s-j\omega)t}\right]_0^{\infty} - \frac{1}{2(s+j\omega)}\left[e^{-(s+j\omega)t}\right]_0^{\infty} \\ &= \frac{1}{2(s-j\omega)} + \frac{1}{2(s+j\omega)} \\ &= \frac{s}{s^2+\omega^2}\end{aligned}$$
    - Similarly we can show $\mathcal{L}\left\{\sin(\omega t)\right\} = \frac{\omega}{s^2+\omega^2}$
  - Power of t:
$$\begin{aligned}\int_0^{\infty} t^n e^{-st}\,\mathrm{d}t &= \left[-\frac{t^n}{s}e^{-st}\right]_0^{\infty} + \int_0^{\infty} nt^{n-1}\frac{e^{-st}}{s}\,\mathrm{d}t \\ &= \frac{n}{s}\int_0^{\infty} t^{n-1}e^{-st}\,\mathrm{d}t \\ &= \frac{n}{s}\mathcal{L}\left\{t^{n-1}\right\} \\ &= \frac{n!}{s^{n+1}}\end{aligned}$$

    * Therefore the unit ramp function has $\mathcal{L}\{t\} = \dfrac{1}{s^2}$
- Important properties:
  - Linearity/superposition: $\mathcal{L}\{\alpha_1 f_1(t) + \alpha_2 f_2(t)\} = \alpha_1 F_1(s) + \alpha_2 F_2(s)$
  - Time delay: $$\begin{aligned}\mathcal{L}\{f(t-\tau)1(t-\tau)\} &= \int_0^\infty f(t-\tau)1(t-\tau)e^{-st}\,\mathrm{d}t \\ &= \int_\tau^\infty f(t-\tau)e^{-st}\,\mathrm{d}t \\ &= \int_0^\infty f(\lambda)e^{-s(\tau+\lambda)}\,\mathrm{d}\lambda \\ &= e^{-\tau s}\int_0^\infty f(\lambda)e^{-s\lambda}\,\mathrm{d}\lambda \\ &= e^{-\tau s}F(s)\end{aligned}$$
    * A delay in time domain is a multiplication by an exponential in Laplace domain
  - Differentiation: $$\begin{aligned}\mathcal{L}\left\{\frac{\mathrm{d}}{\mathrm{d}t}f(t)\right\} &= \int_0^\infty e^{-st}\dot{f}(t)\,\mathrm{d}t \\ &= \left[f(t)e^{-st}\right]_0^\infty + s\int_0^\infty f(t)e^{-st}\,\mathrm{d}t \\ &= sF(s) - f(0)\end{aligned}$$
    * Note the $f(0)$ term vanishes for a zero-state response
    * For higher derivatives: $\mathcal{L}\left\{\dfrac{\mathrm{d}^2}{\mathrm{d}t^2}f(t)\right\} = s(sF(s) - f(0)) - \dot{f}(0) = s^2F(s) - sf(0) - \dot{f}(0)$
    * Going backwards: $\mathcal{L}\{tf(t)\} = -\dfrac{\mathrm{d}}{\mathrm{d}s}F(s)$
  - Integration: $$\begin{aligned}\mathcal{L}\left\{\int_0^t f(\tau)\,\mathrm{d}\tau\right\} &= \int_0^\infty \int_0^t f(\tau)\,\mathrm{d}\tau e^{-st}\,\mathrm{d}t \\ &= -\frac{1}{s}\left[\int_0^t f(\tau)\,\mathrm{d}\tau e^{-st}\right]_0^\infty + \frac{1}{s}\int_0^\infty f(t)e^{-st}\,\mathrm{d}t \\ &= \frac{1}{s}F(s)\end{aligned}$$
  - Convolution: $$\begin{aligned}\mathcal{L}\{f(t) * h(t)\} &= \int_0^\infty \int_0^t f(t-\tau)h(\tau)\,\mathrm{d}\tau e^{-st}\,\mathrm{d}t \\ &= \int_0^\infty \int_0^t f(t-\tau)h(t)e^{-st}\,\mathrm{d}\tau\,\mathrm{d}t \\ &= \int_0^\infty \int_\tau^\infty f(t-\tau)h(\tau)e^{-st}\,\mathrm{d}t\,\mathrm{d}\tau \\ &= \int_0^\infty \int_0^\infty f(\lambda)h(\tau)e^{-s(\lambda+\tau)}\,\mathrm{d}\lambda\,\mathrm{d}\tau \\ &= \int_0^\infty f(\lambda)e^{-s\lambda}\,\mathrm{d}\lambda \int_0^\infty h(\tau)e^{-s\tau}\,\mathrm{d}\tau \\ &= F(s)H(s)\end{aligned}$$
    * This means we can multiply the Laplace transform of the input by the Laplace transform of the impulse response to get the Laplace transform of the output
    * Note $\mathcal{L}\{f(t)h(t)\} = \dfrac{1}{2\pi j}(F(s) * H(s))$
  - Final Value Theorem: $\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$
    * Recall that $\mathcal{L}\left\{\dfrac{\mathrm{d}}{\mathrm{d}t}f\right\} = sF(s) - f(0)$
    * $\lim_{s\to\infty}(sF(s) - f(0)) = \lim_{s\to 0} sF(s) - f(0) = \lim_{s\to 0}\int_0^\infty e^{-st}\dfrac{\mathrm{d}f}{\mathrm{d}t}\,\mathrm{d}t = \int_0^\infty \dfrac{\mathrm{d}f}{\mathrm{d}t}\,\mathrm{d}t = \lim_{t\to\infty} f(t) - f(0)$

    * Note this requires that $f(t)$ and $\frac{\mathrm{d}f}{\mathrm{d}t}$ have Laplace transforms, and $\lim_{t\to\infty} f(t)$ exists, i.e. it is *stable*
  - Initial Value Theorem: $\lim_{t\to 0^+} f(t) = \lim_{s\to\infty} sF(s)$

Transfer Functions

Definition

Transfer Function: The ratio of the Laplace transforms of the output to the input of a system, assuming that the system was initially at equilibrium (zero state/initial conditions).

- All transfer functions assume zero-state; if we want to look at initial conditions we shouldn't use transfer functions
- Given any input $u(t)$ to the system, the output of the system in time domain is $y(t) = h(t) * u(t)$ where $h(t)$ is the impulse response
- In Laplace domain, the output is $Y(s) = H(s)U(s)$ where $H(s) = \frac{Y(s)}{U(s)}$, the Laplace transform of the impulse response, is the transfer function
- For all LTI systems, the transfer function of the system fully characterizes the system dynamics
- Most transfer functions are rational functions $H(s) = K_H \frac{n_H(s)}{d_H(s)} = K_h \frac{s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n}$
  - *Poles* are the roots of $d_H(s)$
    * These are more important than the zeros
  - *Zeros* are the roots of $n_H(s)$
  - Poles are denoted with an X while zeros are denoted by O on the complex plane when plotting
  - K_H is the transfer function *gain*
  - $d_H(s)$ is the *characteristic equation* of the transfer function/system
    * The system's *order* is the degree of $d_H(s)$
- For all causal systems, the *relative degree* $n - m$ of the transfer function is always greater than or equal to zero
  - Consider $H(s) = s$; then for an input $U(s)$, we get output $Y(s) = sU(s)$, which means $y(t) = \frac{\mathrm{d}}{\mathrm{d}t}u(t)$
    * Such a system cannot be causal, because in order to determine the derivative of the input, the system needs to somehow anticipate the input's behaviour in the future
    * e.g. if we put in a sinusoid, it will be shifted to the left, which is non-causal
    * Generally, zeros tend to push the system towards non-causality by moving the response earlier in time, while poles push the system towards causality by delaying the response
  - The transfer function is a *proper ratio* (if $m < n$, then it is *strictly proper*)
  - Most systems we will study have strictly proper transfer functions
- $H(s) = K_H \frac{\prod_{i=1}^{m}(s - z_i)}{\prod_{i=1}^{n}(s - p_i)} = \left(\frac{K_H \prod_{i=1}^{m} z_i}{\prod_{i=1}^{n} p_i}\right) \frac{\prod_{i=1}^{m}\left(\frac{s}{z_i} - 1\right)}{\prod_{i=1}^{n}\left(\frac{s}{p_i} - 1\right)}$ where z_i are the zeros, p_i the are poles
- For any LTI system, the poles of a system determines its behaviour
  - Note complex poles always come in conjugate pairs
  - Any poles on the right hand plane are unstable, i.e. the output will keep growing
    * Larger real values lead to faster growth
  - Any poles on the left hand plane are convergent, i.e. output eventually settles to 0
    * More negative real values lead to faster decay
  - Poles with zero real part neither grow nor shrink in magnitude
  - Any imaginary component in the pole causes the output to oscillate
    * Larger imaginary component lead to higher oscillation frequency
- When there are multiple poles and zeros, they will interact with each other and lead to more interesting behaviour

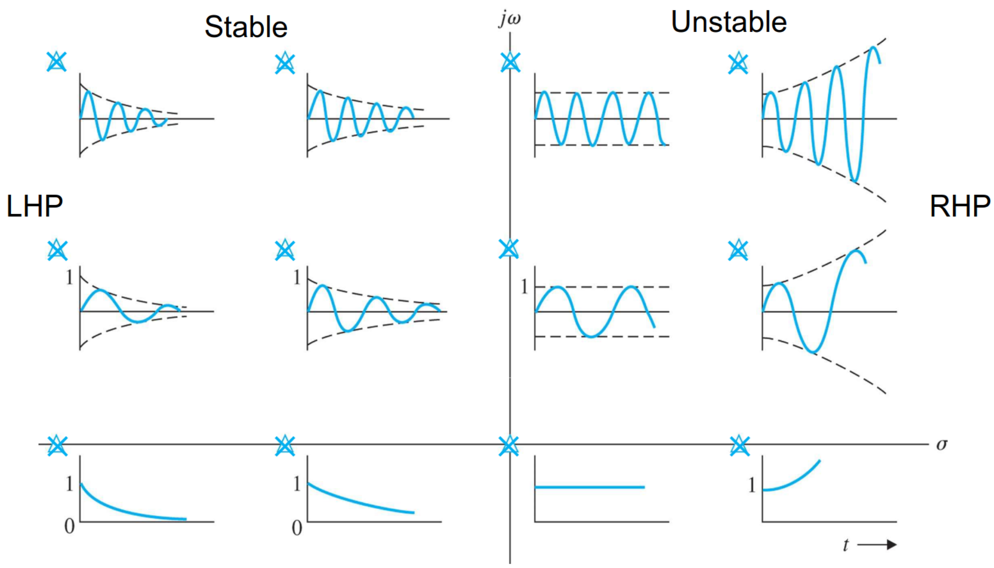


Figure 21: Behaviour of a system according to its poles.

- The *DC gain* (or *static gain*) is the steady-state response of the system to the unit step input
  - This will give an output $Y_s(s) = H(s)U(s) = \frac{1}{s}H(s)$
  - Using FVT, $\lim_{t\to\infty} y_s(t) = \lim_{s\to 0} s\left(\frac{1}{s}H(s)\right) = \lim_{s\to 0} H(s)$
  - This makes the DC gain very easy to find

Lecture 8, Feb 1, 2024

Block Diagrams

- We use block diagrams to depict cause-and-effect relationships within a system
  - Each block shows a function acting on an input to generate output
    - The block is depicted with a transfer function
  - Arrows are used to represent the direction of signals (i.e. information flow)
  - Circles are used for algebraic sum and differences of signals
  - Nodes (aka pick-off points) are used for branching out signals
- Note when we have a feedback system, we usually depict the plant' transfer function by $G(s)$ and the feedback's transfer function by $H(s)$

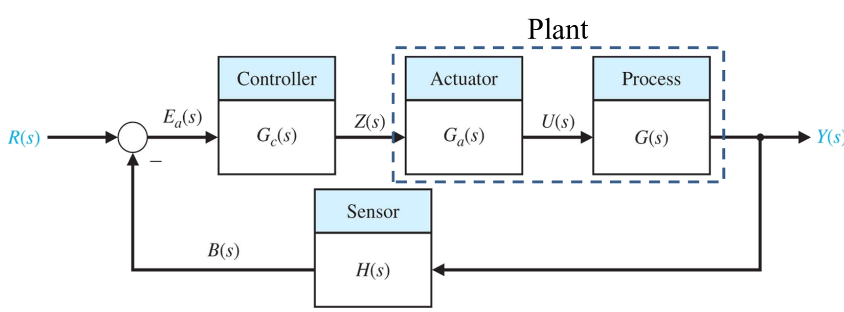


Figure 22: General feedback system.

- Transfer function definitions for a general feedback system:

- Closed-loop TF: $\mathcal{T} = \dfrac{Y(s)}{R(s)} = \dfrac{G_c(s)G_a(s)G(s)}{1 + G_c(s)G_a(s)G(s)H(s)}$
- Open-loop TF: $L(s) = \dfrac{B(s)}{E_a(s)} = G_c(s)G_a(s)G(s)H(s)$
  * Note this is not the output to input without feedback (which would be feedforward TF)
  * This is the ratio of the feedback signal to $E_a(s)$
- Error TF: $\dfrac{E(s)}{R(s)} = \dfrac{R(s) - Y(s)}{R(s)} = \dfrac{1 + G_c(s)G_a(s)G(s)(H(s) - 1)}{1 + G_c(s)G_a(s)G(s)H(s)}$
  * Note the $E(s)$ here is not the same as $E_a(s)$
- Feedforward TF: $\dfrac{Y(s)}{E_a(s)} = G_c(s)G_a(s)G(s)$
  * Note here we use $E_a(s)$ not $E(s)$
- Feedback TF: $\dfrac{B(s)}{R(s)} = \dfrac{G_c(s)G_a(s)G(s)H(s)}{1 + G_c(s)G_a(s)G(s)H(s)}$
  * This is the ratio of feedback signal to input signal
  * We can find this easily by taking $\dfrac{Y(s)}{R(s)}H(s)$
- Sensitivity TF: $\mathcal{S}(s) = \dfrac{1}{1 + G_c(s)G_a(s)G(s)H(s)}$
  * This is important to the robustness of the controller as we will later see
  * This is the inverse of the characteristic equation
- Characteristic equation: $1 + G_c(s)G_a(s)G(s)H(s)$
  * This is the denominator of the closed-loop TF

- Block diagrams can be simplified to find the overall transfer function of the system
  - There are a number of simplification rules

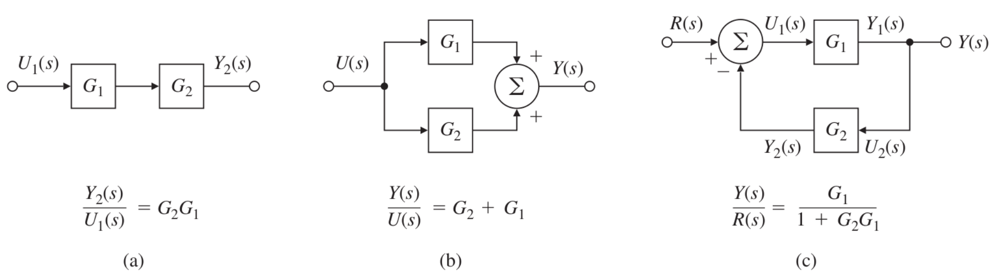


Figure 23: Block diagram reduction rules.

Lecture 9, Feb 5, 2024

First-Order System Response

- Consider a pure integrator: $y(t) = \displaystyle\int_0^t u(t)\,\mathrm{d}t + y(0)$ which has transfer function $H(s) = \dfrac{1}{s}$ if $y(0) = 0$
  - The ODE is $\dot{y}(t) = u(t)$
  - The impulse response is $y_i(t) = \mathcal{L}^{-1}\left\{H(s)\right\} = \mathcal{L}^{-1}\left\{\dfrac{1}{s}\right\} = 1$
  - The step response is $y_s(t) = \mathcal{L}^{-1}\left\{H(s)\dfrac{1}{s}\right\} = \mathcal{L}^{-1}\left\{\dfrac{1}{s^2}\right\} = t$
  - What if the initial condition is not zero?
    * Laplace transform the ODE to get $sY(s) - y(0) = U(s) \implies Y(s) = \dfrac{1}{s}U(s) + \dfrac{1}{s}y(0)$
    * For a step response, $y_s(t) = \mathcal{L}^{-1}\left\{\dfrac{1}{s^2}\right\} + \mathcal{L}^{-1}\left\{\dfrac{1}{s}y(0)\right\} = t + y(0)1(t) = t + y(0)$
- Consider an RC circuit with input voltage $Ku(t)$

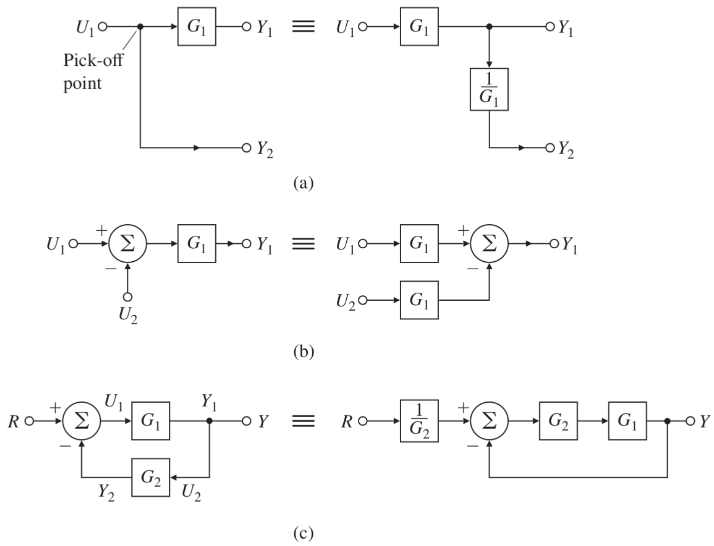


Figure 24: Block diagram reduction rules.

- Form the ODE: $T\dot{y}(t) + y(t) = Ku(t)$ where $T = RC$
- Laplace transform: $TsY(s) + Y(s) = KU(s) \implies H(s) = \frac{Y(s)}{U(s)} = \frac{K}{Ts+1}$
- Impulse response: $y_i(t) = \mathcal{L}^{-1}\{H(s)\} = \frac{K}{T}e^{-\frac{t}{T}}$
- Step response: $y_s(t) = \mathcal{L}^{-1}\left\{\frac{K}{s(Ts+1)}\right\} = K\mathcal{L}^{-1}\left\{\frac{1}{s} - \frac{T}{Ts+1}\right\} = K\left(1 - e^{-\frac{t}{T}}\right)$
- We can see that T is the time constant of the system; the smaller it is, the faster the system evolves
- DC gain: $y_ss = \lim_{t\to\infty} y_s(t) = \lim_{s\to 0} s\frac{K}{s(Ts+1)} = K$

- In general a first-order system has transfer function $H(s) = \frac{b}{s+a}$ and impulse response $h(t) = be^{-at}1(t)$
  - For positive a, this is stable and the system decays to 0; for negative a, this is unstable; for $a = 0$ the system maintains a constant output
    - Positive a gives poles in the LHP and negative a gives poles in the RHP
  - The step response is given by $y_s = \frac{b}{a}(1 - e^{-at})1(t)$
    - For positive a, this converges to the DC gain $\frac{b}{a}$
    - For negative a this diverges exponentially
    - For zero a this gives a linear response (note we can derive this by nothing $H(s) = \frac{b}{s}$ in this case)
- The time constant is given by $T = \frac{1}{a}$
  - The rise time is given by $t_r \approx 2.2T$, which is the time taken for the output to go from 10% to 90% of the DC gain
  - The settling time is given by $t_s \approx \frac{4.6}{a}$, which is the time taken for the output to reach 99% of the DC gain
- In a first-order system, there is never any overshoot or oscillation; the output never passes the steady state value

Second Order System Response

- Consider a spring-mass-dashpot system: $m\ddot{y}(t) + b\dot{y}(t) + ky(t) = kf(t)$

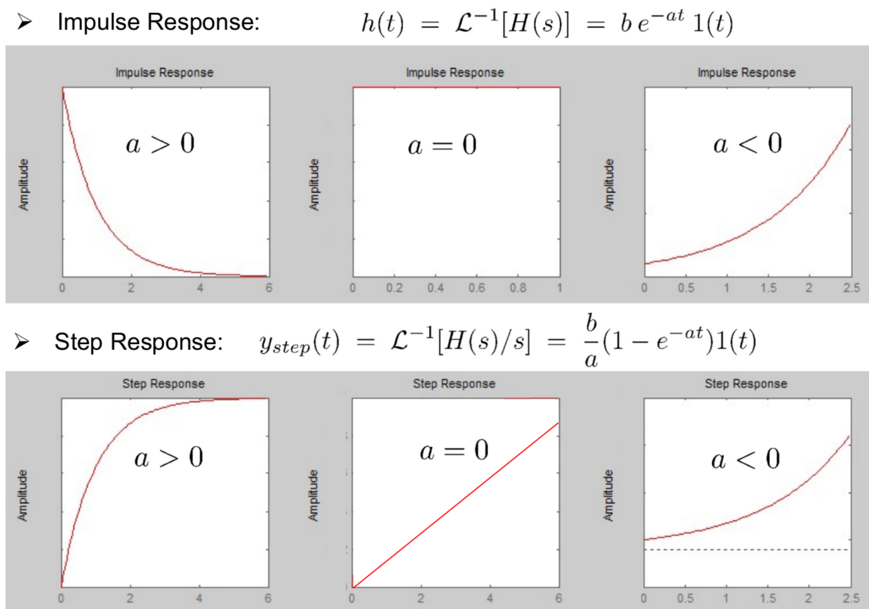


Figure 25: Behaviour of the impulse and step responses for a general (strictly proper) first-order system.

- Laplace transform: $m(s^2Y(s) - sy(0^-) - \dot{y}(0^-)) + b(sY(s) - y(0^-)) + kY(s) = kF(s)$
- $Y(s) = \dfrac{\frac{k}{m}}{s^2 + \frac{b}{m}s + \frac{k}{m}}F(s) + \dfrac{s + \frac{b}{m}}{s^2 + \frac{b}{m}s + \frac{k}{m}}y(0^-) + \dfrac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}}\dot{y}(0^-)$
- Assuming zero state, $H(s) = \dfrac{Y(s)}{F(s)} = \dfrac{\frac{k}{m}}{s^2 + \frac{b}{m}s + \frac{k}{m}} = \dfrac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
  * $\omega_n = \sqrt{\dfrac{k}{m}}$ is the natural frequency
  * $\zeta = \dfrac{b}{2\sqrt{km}}$ is the damping ratio
- The poles are at $-\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$
  * Depending on ζ we can get real or imaginary poles
- If $\zeta > 1$ (i.e. $b > 2\sqrt{km}$) we have two distinct real poles; the system is *overdamped*
  * Let $-\sigma_1 = -\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}$, $-\sigma_2 = -\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}$
  * Then $\omega_n = \sqrt{\sigma_1\sigma_2}, \zeta = \dfrac{\sigma_1 + \sigma_2}{2\sqrt{\sigma_1\sigma_2}}$
- If $\zeta = 1$ (i.e. $b = 2\sqrt{km}$) we have two overlapping real poles; the system is *critically damped*
  * $H(s) = \dfrac{\sigma^2}{(s + \sigma)^2}$ where $\sigma = \omega_n$
- If $0 \leq \zeta < 1$ (i.e. $b < 2\sqrt{km}$) we have two complex conjugate poles; the system is *underdamped*
  * The poles are $s_1, s_2 = -\sigma \pm j\omega_d$ where $\sigma = \zeta\omega_n, \omega_d = \omega_n\sqrt{1 - \zeta^2}$
  * $H(s) = \dfrac{\omega_n^2}{(s - (-\sigma + j\omega_d))(s - (-\sigma - j\omega_d))} = \dfrac{\sigma^2 + \omega_d^2}{(s + \sigma)^2 + \omega_d^2}$
  * In this case the system oscillates
  * ω_d is the oscillation frequency and σ is the decay rate

- Consider the impulse response of the underdamped case
  - $y_i(t) = \mathcal{L}^{-1}\left\{\dfrac{(\sigma^2 + \omega_d^2)}{(s + \sigma)^2 + \omega_d^2}\right\} = \mathcal{L}^{-1}\left\{\dfrac{(\sigma^2 + \omega_d^2)}{\omega_d}\dfrac{\omega_d}{(s + \sigma)^2 + \omega_d^2}\right\} = \dfrac{\sigma^2 + \omega_d^2}{\omega_d}e^{-\sigma t}\sin(\omega_d t)$
  - Alternatively $y_i(t) = \dfrac{\omega_n}{\sqrt{1 - \zeta^2}}e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1 - \zeta^2}t\right)$

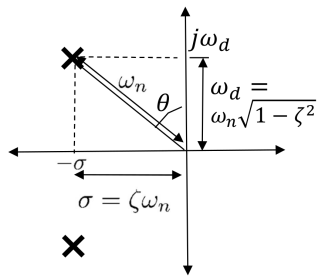


Figure 26: Illustration of the system variables in polar form.

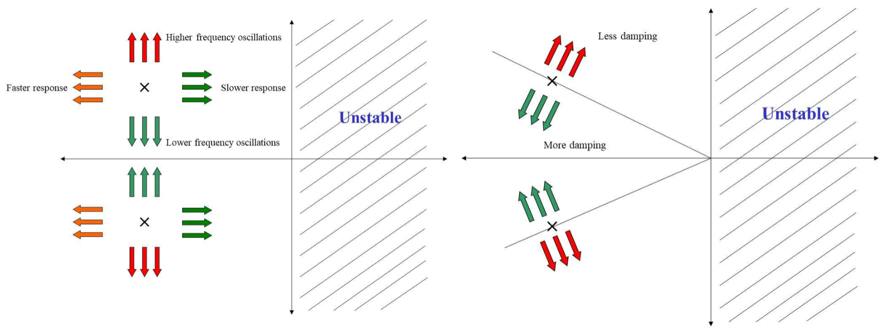


Figure 27: Response of an underdamped second-order system based on pole location.

– The response is a decaying exponential

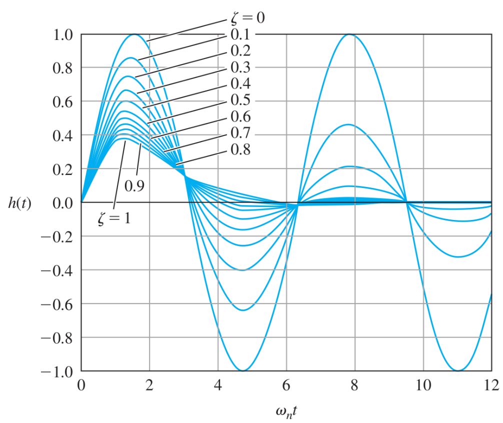


Figure 28: Impulse response of an underdamped second-order system.

Lecture 10, Feb 8, 2024

Second Order System Response (Continued)

- Consider the step response

- $y_s(t) = \mathcal{L}^{-1}\left\{\frac{H(s)}{s}\right\}$
$= \mathcal{L}^{-1}\left\{\frac{\sigma^2+\omega_d^2}{((s+\sigma)^2+\omega_d^2)s}\right\}$
$= \mathcal{L}^{-1}\left\{\frac{1}{s} - \frac{s+2\sigma}{(s+\sigma)^2+\omega_d^2}\right\}$
$= \mathcal{L}^{-1}\left\{\frac{1}{s} - \frac{s+\sigma}{(s+\sigma)^2+\omega_d^2} - \frac{\sigma}{\omega_d}\frac{\omega_d}{(s+\sigma)^2+\omega_d^2}\right\}$
$= 1 - e^{-\sigma t}\left(\cos(\omega_d t) + \frac{\sigma}{\omega_d}\sin(\omega_d t)\right)$
$= 1 - e^{-\sigma t}\frac{\omega_n}{\omega_d}\cos(\omega_d t - \theta)$
  * Where $\theta = \tan^{-1}\left(\frac{\omega_d}{\sigma}\right) = \tan^{-1}\left(\frac{\omega_d}{\zeta\omega_n}\right)$
- For an overdamped system, the two separate poles lie on the real axis, and with decreasing ζ the poles move together until they overlap, and then move radially into the imaginary axis
- The system starts with no oscillation but a slow response to faster responses but oscillations begin; when $\zeta = 0$ the poles are purely imaginary, at which point the response is purely oscillatory and no decay occurs
  - When $\zeta = 1$, the poles overlap, and we get critical damping, which is the fastest possible system response without oscillation

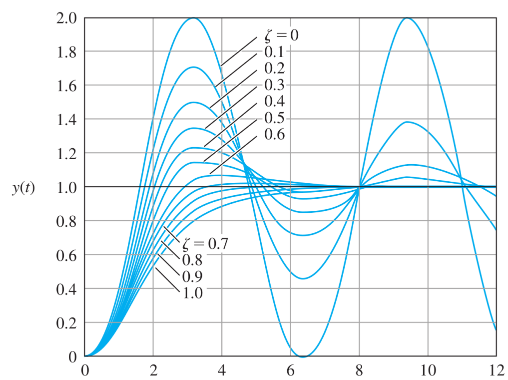


Figure 29: Step response of an underdamped second-order system.

- For the second order system, we can characterize it using the following (for a unit step input):
  - DC gain y_{ss}: the steady-state value of the system output
  - Peak time t_p: time to reach the maximum overshoot/undershoot point
  - Overshoot M_p: the max amount the output overshoots $y_s s$, divided by the steady state value (usually as a percentage)
  - Rise time t_r: the time the system takes to rise from 10% to 90% of $y_s s$
  - Settling time t_s: the time the system takes to reach, and stay within, 1% of $y_s s$ (2% in some texts)

- DC gain: $y_{ss} = \lim_{s\to 0} sY_s(s) = \lim_{s\to 0} s\frac{1}{s}H(s) = \frac{\omega_n^2}{\omega_n^2} = 1$
  - The DC gain here is 1 because when we derived the system, we multiplied u by k
  - Without this scaling the DC gain would be k instead
- Peak time:
  - Take derivative: $\dot{y}_s(t) = \mathcal{L}^{-1}\{sY_s(s)\} = \mathcal{L}^{-1}\left\{s\frac{1}{s}H(s)\right\} = y_i(t)$
    * Note the derivative of the step response is just the impulse response
  - Therefore $y_i(t) = \frac{\omega_n}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2 t}\right)$

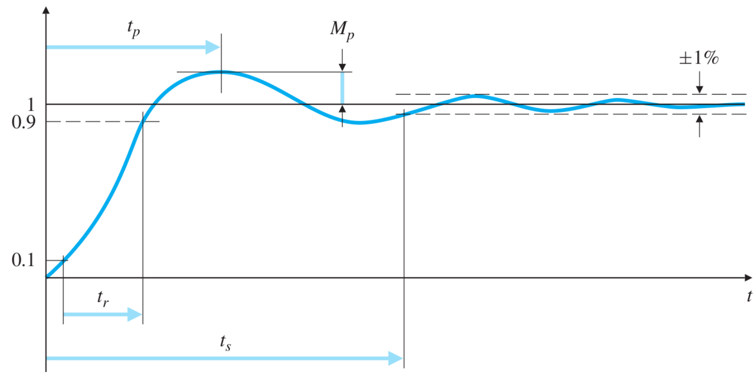


Figure 30: Illustration of the characteristics of second-order system response.

  - $\dot{y}_s(t) = y_i(t) = 0 \implies \omega_n\sqrt{1-\zeta^2}t = n\pi \implies t = \dfrac{n\pi}{\omega_n\sqrt{1-\zeta^2}}$
  - The first peak occurs at $t_p = \dfrac{\pi}{\omega_n\sqrt{1-\zeta^2}} = \dfrac{\pi}{\omega_d}$
  - As we reduce the damping,
- Overshoot:
  - Substitute t_p into the step response to get the peak of the response
  - $y_s(t_p) = 1 - \dfrac{e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}}\cos(\pi-\theta) = 1 + \dfrac{e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}}\cos\theta$
  - We know $\cos\theta = \sqrt{1-\zeta^2}$ so this simplifies to $1 + e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}$
  - The overshoot is therefore $M_p = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}$ (or times 100 for percentage)
  - Notice that this depends only on ζ
    * Usually we're interested in two values: $\zeta = \dfrac{1}{2}$ which gives 16% overshoot, and $\zeta = 0.7$ which gives 5% overshoot
  - The percent overshoot decreases with ζ, but $\omega_n t_p$ increases with ζ

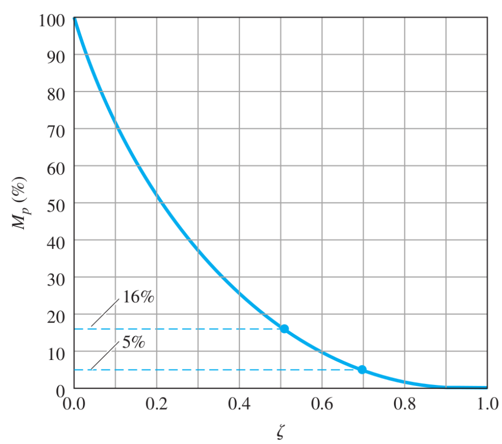


Figure 31: Overshoot as a function of damping ratio.

- Rule of thumb: the response of the second-order underdamped systems (with no finite zeroes) with different damping ratios rises roughly with the same pace
  - Typically we related t_r to only ω_n instead of also ζ, as an approximation
  - For $\zeta = 0.5$, we can approximate $t_r \approx \dfrac{1.8}{\omega_n}$

- We typically choose ζ between 0.5 and 0.7 for a balance between overshoot and rise time
- For settling time we can approximate the deviation of the response by the exponential only
  - Therefore $e^{-\zeta\omega_n t_s} \approx 0.01 \implies t_s \approx \frac{4.6}{\zeta\omega_n} = \frac{4.6}{\sigma}$

Lecture 11, Feb 12, 2024

Second Order System Response (Continued)

- We can now generalize our system to $H(s) = \frac{K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
  - In this case K is the DC gain
- Without any zeroes, we have 3 parameters K, ω_n, ζ to fully specify the system's behaviour
  - In practice, we look for regions in the s plane where we can put the poles
  - e.g. if we want to specify a maximum rise time t_{rd}, settling time t_{sd} and overshoot (corresponding to some damping ζ_d), then we have:
    * $\omega_n \geq \frac{1.8}{t_{rd}}$
    * $\zeta \geq \zeta_d$
    * $\sigma \geq \frac{4.6}{t_{sd}}$
    * Combining these 3 requirements, we see that the allowed region for the pole is indicated in the figure below

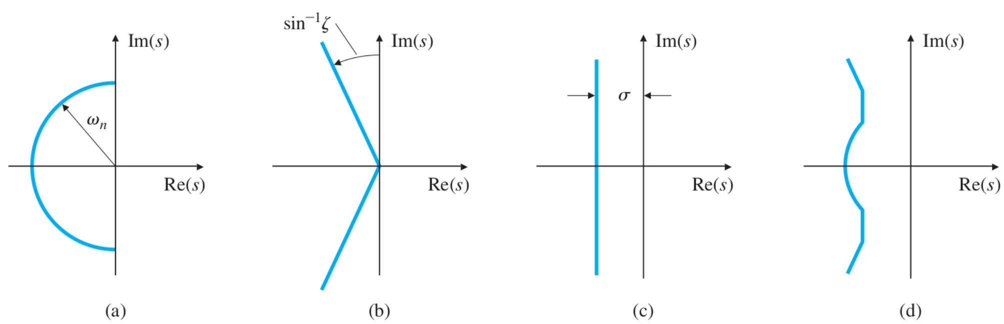


Figure 32: Allowed regions of the s-plane based on the system design requirements.

Effect of Zeroes

- Consider the system $m\ddot{y}(t) + b\dot{y}(t) + ky(t) = kf(t)$ with initial conditions given by $y(0^-) = \frac{k}{b}y_0, \dot{y}(0^-) = 0, f(t) = 0$; consider y_0 as the system input
  - Laplace transform the system: $m(s^2Y(s) - sy(0^-) - \dot{y}(0^-)) + b(sY(s) - y(0^-)) + kY(s) = kF(s)$
  - $Y(s) = \frac{s + \frac{b}{m}}{s^2 + \frac{b}{m}s + \frac{k}{m}} y(0^-)$
  - Again let $\omega_n = \sqrt{k}m, \zeta = \frac{b}{2\sqrt{km}}$
  - Notice that the system now has a zero
  - $\frac{Y(s)}{y_0} = \frac{\frac{\omega_n}{2\zeta}(s + 2\zeta\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
  - In the underdamped case $0 \leq \zeta < 1$, the poles are $-\zeta\omega_n \pm j\omega_n\sqrt{\zeta^2 - 1} = -\sigma \pm j\omega_d$ with a zero at $z_1 = -2\zeta\omega_n = -2\sigma$
  - Normalize by ω_n: $\frac{Y(s)}{y_0} = \frac{\frac{1}{2\zeta}\frac{s}{\omega_n} + 1}{\left(\frac{s}{\omega_n}\right)^2 + 2\zeta\frac{s}{\omega_n} + 1}$

        * By doing this we can ignore ω_n

- More generally, $H(s) = \dfrac{\frac{1}{\alpha\zeta}s + 1}{s^2 + 2\zeta s + 1}$
    - The DC gain is still 1
    - We have generalized the 2 to α and replaced $\dfrac{s}{\omega_n}$ by s (equivalent to $t \leftarrow \omega_n t$)
    - For this system the zero is at $z = -\alpha\sigma$
    - We can write this as $H(s) = H_p(s) + \dfrac{1}{\alpha\zeta}sH_p(s)$ where $H_p = \dfrac{1}{s^2 + 2\zeta s + 1}$
        * $H_p(s)$ is a second-order transfer function with no zeroes
        * We see that the effect of a zero is equivalent to adding s times the transfer function
        * In time domain, this is equivalent to adding the derivative of the response to itself (since multiplication by s is differentiation)
    - The DC gain of the system is $y_{ss} = \lim_{s\to 0} H(s) = \lim_{s \to 0} H_p(s)$
        * The DC gain of the original transfer function is not changed by adding a zero
        * The steady-state response is unaffected by adding zeroes

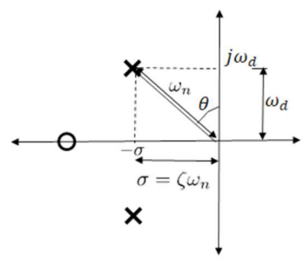


Figure 33: Poles and zeros of the example system.

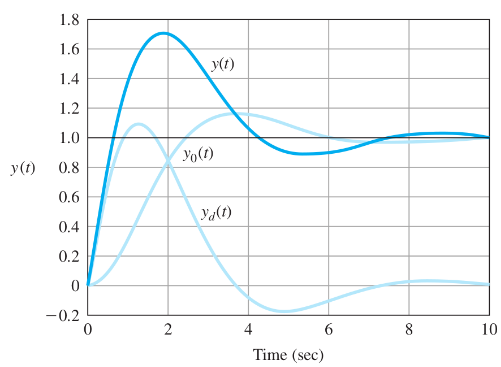


Figure 34: Plot of the system's transient step response. y_0 is the response without the zero, and y_d is its derivative.

- The effect of adding a zero is to add the derivative of the response to itself, resulting in a shorter rise/peak time and larger overshoot
    - With increasing α, the system gets closer to the response without a zero
    - Increasing alpha means the zero moves further into the negative
        * In general, the further the zero gets from the poles, the less its effect will be
    - For ζ values of 0.5 or above, any value of α larger than 4 will have a negligible effect
    - Note adding a zero may inadvertently affect the initial conditions of the system
        * By the initial value theorem we can find $y(0)$ by taking the limit as $s \to \infty$
        * Adding zeros can make $y(0)$ nonzero

- What if α is negative, so the zero is in the right hand plane?

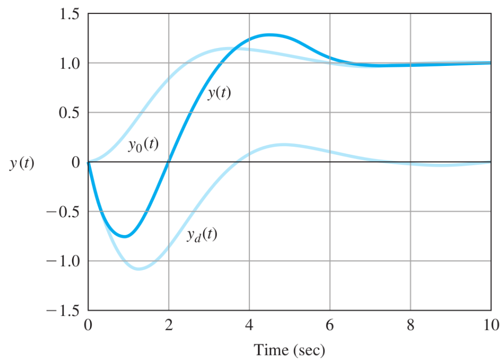


Figure 35: System step response for a nonminimum-phase zero.

- This doesn't make the system unstable (since only the pole locations determine system stability)
- The effect is now subtractive, so the system slows down and the rise/peak time is increased
  * The overshoot is far less than the case where the zero is in the LHP (however it is still more than the case of having no zeroes)
- The system may start in the "wrong direction" – moving in the opposite direction as the equilibrium initially
  * This is often undesirable
- These systems are called *nonminimum-phase zeroes*

- If a zero is close to a pole, it can "neutralize" the effect of the pole
  - We can deliberately place zeros to neutralize poles to change the system behaviour
  - Consider $H_1(s) = \dfrac{2}{(s+1)(s+2)} = \dfrac{2}{s+1} - \dfrac{2}{s+2}$ and $H_2(s) = \dfrac{\frac{2}{1.1}(s+1.1)}{(s+1)(s+2)} = \dfrac{0.18}{s+1} + \dfrac{1.64}{s+2}$
    * Same characteristic equation and DC gain, but the second has a zero very close to a pole
    * Notice in H_2, the part corresponding to the second pole at $s = -2$ stayed roughly the same, while the first pole at $s = -1$ diminished significantly
  - In the figure below, the response of H_2 is much closer to the first-order system H_{12} than H_1
  - Mathematically we can use a zero on the RHP to neutralize an unstable pole, but this should never be done in practice because we never know where exactly the pole is, so the zero may not overlap perfectly
    * This also applies for LHP poles that are close to being unstable

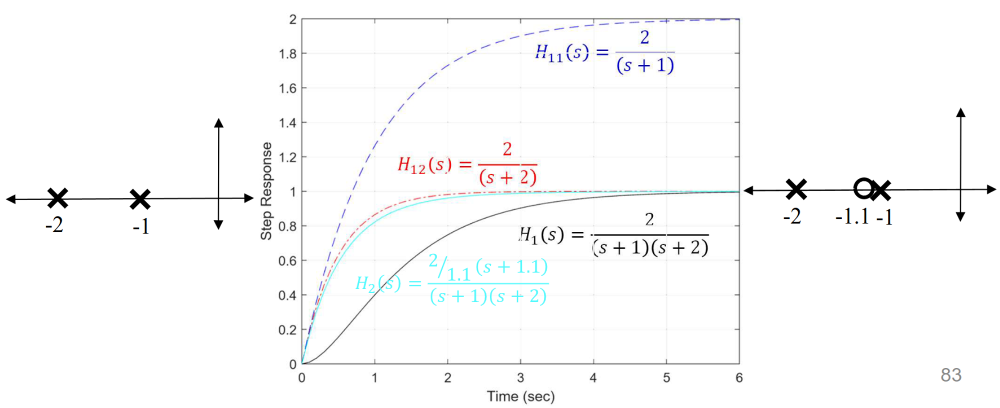


Figure 36: Effect of zeroes close to a pole.

- Now consider the effect of complex poles on the system
  - Example: $H_1(s) = \dfrac{1.01}{\alpha^2+\beta^2} \cdot \dfrac{(s+\alpha)^2+\beta^2}{(s+1)[(s+0.1)^2+1]}$

  * The term in the front normalizes the DC gain to 1
  * The zeroes are at $z_1, z_2 = -\alpha \pm j\beta$
  * The poles are at $p_1 = -1, p_2, p_3 = -0.1 \pm j1$
  * The closer the poles get to the zeroes, the less their effect becomes

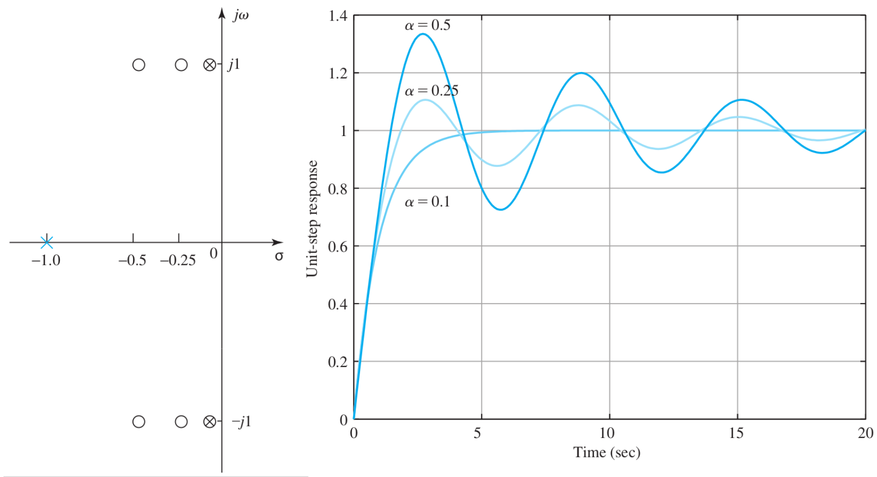


Figure 37: Effect of complex poles; $\beta = 1$ is used in all cases.

Higher Order Systems

- Generally, the higher the system order, the more complex it is and the more lag we will see in the system
- The rise and peak times will generally increase and overshoot deceases as we add more poles
- The transient response is slowed down but they have little effect on the settling time
- Additional poles are more effective the closer they are to the existing second-order poles
  - Generally if they are 4 or more times further, their effect can be ignored
- The overall system response is the sum of terms due to each pole/pair of poles
  - Poles having a real part closer to zero will have a much more pronounced effect on the system
- The effect of the poles is determined by:
  - The real part of the pole, σ, determines both the stability and the system time constant (rate of decay)
  - The imaginary part of pole, ω_d, determines the damped frequency
  - The magnitude of the pole determines the natural frequency of the system
  - The argument/angle of the pole determines the damping ratio
- Based on these, we can approximate the system and reduce its order to make it easier to analyze

Summary

For a second-order system with no finite zeroes, the transient response can be characterized approximately by 3 characteristics:

- Rise time: $t_r \approx \dfrac{1.8}{\omega_n}$ (if the rise time is too long, increase the natural frequency)
- Overshoot: $M_p = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}$ (if there is too much overshoot, increase the damping ratio)
- Settling time: $t_s \approx \dfrac{4.6}{\sigma}$ (if the system takes too long to settle, move the pole to the left)

Real zeroes in the LHP will significantly increase the overshoot but decrease the rise time (if it is within a factor of 4 of the real part of the complex poles); real zeroes in the RHP (nonminimum-phase zeroes) will reduce the overshoot, but may cause the system to start in the wrong direction. Zeroes close to poles may cancel out their effects on the system.
Additional real poles in the LHP will significantly increase the rise time but decrease the overshoot (again, if it is within a factor of 4 of the real part of existing poles).

Lecture 12, Feb 15, 2024

Stability of LTI Systems

- For a transfer function $H(s) = K_H \dfrac{\prod_{i=1}^{m}(s - z_i)}{\prod_{j=1}^{n}(s - p_j)}$, the impulse response will look like a sum of exponentials, $y(t) = \sum_{j=1}^{n} C_j e^{p_j t}$ (assuming poles are distinct)
  - The coefficients C_j depend on the initial conditions and locations of zeroes
  - If a pole is repeated k times, we will have terms $C_{j_0}e^{p_j} + C_{j_1}te^{p_j} + \cdots + C_{j_{k-1}}t^{k-1}e^{p_j}$
  - The system response is bounded if and only if all $\mathrm{Re}(p_j) \leq 0$ (regardless of repeating poles); hence any poles in the RHP are unstable

Definition

We define three types of stability for systems:

- *Bounded-Input-Bounded-Output* (BIBO) Stability: Any bounded input generates a bounded output (with no requirement on convergence).
- *Asymptotic* Stability: Any inital condition generates an output which approaches for zero as time approaches infinity.
- *Marginal* (or *Neutral*) Stability: Any initial condition generates an output which is bounded (for a zero input).

- Asymptotic stability is a generally stronger form of stability than BIBO
  - All asymptotically stable systems are also BIBO stable
- For all LTI systems, all BIBO systems are also asymptotically stable
- If any poles are exactly on the imaginary axis, then if they are non-repeating, the system is marginally/neurally stable, but if they are repeating, then the system is unstable
  - This will result in either a constant output or a free oscillator
- The *Routh-Hurwitz* stability criterion can be used to identify the stability of a system without explicitly factoring the characteristic equation
  - Consider the characteristic equation $s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-1}s + a_n$
  - If all poles are in the LHP, then all coefficients a_i are positive and real
    * Therefore if any $a_i \leq 0$, then the system is always unstable (or marginally stable)
  - If all $a_i > 0$, we need to form the *Routh array* to check for stability
    * The array consists of $n + 1$ rows, with row i corresponding to s^i
    * Row n contains $1, a_2, a_4, \ldots$

    * Row $n-1$ contains $a_1, a_3, a_5, \dots$
    * Row $n-2$ contains $b_1, b_2, b_3, \dots$ where:
      - $b_1 = -\frac{1}{a_1} \det \begin{bmatrix} 1 & a_2 \\ a_1 & a_3 \end{bmatrix}$
      - $b_2 = -\frac{1}{a_1} \det \begin{bmatrix} 1 & a_4 \\ a_1 & a_5 \end{bmatrix}$
      - $b_3 = -\frac{1}{a_1} \det \begin{bmatrix} 1 & a_6 \\ a_1 & a_7 \end{bmatrix}$
    * Row $n-3$ contains $c_1, c_2, c_3, \dots$ where:
      - $c_1 = -\frac{1}{b_1} \det \begin{bmatrix} a_1 & a_3 \\ b_1 & b_2 \end{bmatrix}$
      - $c_2 = -\frac{1}{b_1} \det \begin{bmatrix} a_1 & a_5 \\ b_1 & b_3 \end{bmatrix}$
      - $c_2 = -\frac{1}{b_1} \det \begin{bmatrix} a_1 & a_7 \\ b_1 & b_4 \end{bmatrix}$
    * For x_i, consider the 2×2 matrix formed by taking column 1 and column $i+1$ of the two previous rows, take the negative of its determinant and divide by the bottom left entry
      - We treat any missing entries as zeroes
      - This means for each row starting from $n-2$, we will get one fewer element (zero entry) per row
      - By row 1 we are left with only one element
      - For row 0 we still have one element, obtained by treating the missing entry as a zero when calculating the determinant
      - Past row 0 all entries will be zero, so we stop
    * Note: all elements of each row can be divided by a common factor to simplify computation
  - A system is stable if and only if all elements in the first column of the Routh array are positive
  - The number of roots in the RHP is equal to the number of sign changes in the first column of the Routh array
  - Note the two special cases:
    * One of the elements in the first column is zero
      - Replace this element by some small positive value $\operatorname{var}\epsilon$ and construct the rest of the array
      - Take the limit $\operatorname{var}\epsilon \to 0^+$ and check the sign of the first column
    * An entire row is zero
      - Take the contents of the row above this, and create an auxiliary polynomial with only even powers, using the row as coefficients
      - e.g. if the row above the zero row has 3 and 12, then the auxiliary polynomial is $p(s) = 3s^2 + 12$
      - Differentiate this polynomial and use the coefficients in the derivative as the new contents for the zero row
- Example: unity feedback system, with a PI controller $K + \frac{K_I}{s}$, and a plant $\frac{1}{(s+1)(s+2)}$
  - For what values of K and K_I is the closed-loop system stable?
  - $H(s) = \frac{\left(K + \frac{K_I}{s}\right)\left(\frac{1}{(s+1)(s+2)}\right)}{1 + \left(K + \frac{K_I}{s}\right)\left(\frac{1}{(s+1)(s+2)}\right)} = \frac{Ks + K_I}{s^3 + 3s^2 + (2+K)s + K_I}$
  - From this we can see that a necessary condition is $K_I > 0$ and $K > -2$, but this is not a sufficient condition
  - Form the Routh array:
    *

Row 3	1	$(2+K)$
Row 2	3	K_I
Row 1	$(6+3K-K_I)/3$	
Row 0	K_I	

  - To have all terms in the first row be positive, we require $K_I > 0$ and $K > \frac{1}{3}K_I - 2$

- Note that for this system, even if we took $K_I = 0$, because we'd have a pole at 0 and zero at 0, they cancel out and the overall system is still stable

Lecture 13, Feb 27, 2024

Control System Performance

- In open-loop control, we control the plant without using feedback from its output
- For open-loop control, $Y_{ol} = GD_{ol}R + GW$
  - $E_{ol} = R - Y_{ol} = (1 - GD_{ol})R - GW$
  - Assuming no disturbance so $W(s) = 0$, we can define the open-loop transfer function
  - $T_{ol}(s) = \dfrac{Y(s)}{R(s)} = G(s)D_{ol}(s)$

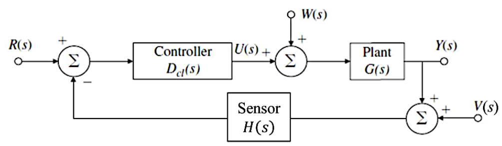


Figure 38: Closed-loop control.

- For closed-loop control, $Y_{cl} = \mathcal{T}R + G\mathcal{S}W - H\mathcal{T}V$
  - $E_{cl} = R - Y_{cl} = (1 - \mathcal{T})R - G\mathcal{S}W + H\mathcal{T}V$
    - * Notice the negative sign in front of the W term, since the output is subtracted from reference
  - In this case, $V(s)$ is noise in our sensor measurements, which we separate from $W(s)$
    - * $W(s)$ is usually low-frequency
    - * $V(s)$ is usually high-frequency
  - Since we have an LTI system, we can consider the input and sources of noise separately
    - * Assuming $W(s) = V(s) = 0$ we can define the closed-loop transfer function
      - $\mathcal{T}(s) = \dfrac{Y(s)}{R(s)} = \dfrac{G(s)D_{cl}(s)}{1 + H(s)G(s)D_{cl}(s)} = T_{cl}(s)$
    - * Assuming $R(s) = V(s) = 0$ we can define the transfer function for process noise
      - $\dfrac{Y(s)}{W(s)} = G(s) \cdot \dfrac{1}{1 + H(s)G(s)D_{cl}(s)} = G(s)\mathcal{S}(s)$
      - Recall that $\mathcal{S}(s)$ is the *sensitivity* transfer function
    - * Assuming $R(s) = W(s) = 0$ we can define the transfer function for measurement noise
      - $\dfrac{Y(s)}{V(s)} = -H(s) \cdot \dfrac{D_{cl}(s)G(s)}{1 + H(s)D_{cl}(s)G(s)} = -H(s)\mathcal{T}(s)$

Stability

- Consider an unstable plant $G(s) = \dfrac{b(s)}{a(s)}$, i.e. $a(s)$ has roots in the RHP; how can we design our controller to make the system stable?
  - Let the controller be $\dfrac{c(s)}{d(s)}$
  - For the open loop control we have $T_{ol} = \dfrac{b(s)}{a(s)}\dfrac{c(s)}{d(s)}$
    - * Theoretically we can design $c(s)$ to cancel the unstable roots of $a(s)$, but as previously mentioned, this is impractical
    - * We can make the same argument for cancelling bad zeroes (zeroes with small real part causing large overshoot)
      - For a nonminimum-phase zero, we can't do this at all because we'd need an unstable pole in the controller

    * Therefore in practice we can't use open-loop control to stabilize a plant
  - For closed-loop controllers, assume $H(s) = 1$ for the sensor, then $T_{cl} = \dfrac{b(s)c(s)}{a(s)d(s) + b(s)c(s)}$
    * Now we have a lot more options for eliminating the unwanted poles
- Example: Inverted pendulum (segway)
  - $\begin{cases} (m_t + m_p)\ddot{x} + b\dot{x} - m_p l\ddot{\theta}_0 = u \\ (I + m_p l^2)\ddot{\theta} - m_p g l\theta - m_p l\ddot{x} = 0 \end{cases}$
    * m_t, m_p are the masses of the cart and pendulum, I is the moment of inertia of the pendulum, l is the length of the pendulum, x is the cart's displacement and θ is the angle of the pendulum from normal
  - $G(s) = \dfrac{\Theta(s)}{U(s)} = \dfrac{m_p l s}{((m_t + m_p)(I + m_p l^2) - m_p^2 l^2)s^3 + b(I + m_p l^2)s^2 - m_p g l(m_t + m_p)s - m_p g b l}$
  - If we assume $b = 0$ then we get a send order system with $((m_t + m_p)(I + m_p l^2) - m_p^2 l^2)s^2 - m_p g l(m_t + m_p)$ in the denominator
    * We can immediately tell that this is unstable by the RH criterion since since the s^1 term is missing
  - Assume $m_p = 1\,\mathrm{kg}, I = 1\,\mathrm{kgm}^2, l = 1\,\mathrm{m}, m_t = 0$ then we get $G(s) = \dfrac{1}{s^2 - 10} = \dfrac{1}{(s + 3.16)(s - 3.16)}$
  - Consider a controller $D_{cl}(s) = \dfrac{K(s + \gamma)}{s + \delta}$ and $H(s) = 1$
    * Choosing $\gamma = -3.16$ cancels the RHP pole, but this is impractical
    * Choose $\gamma = +3.16$ cancels the stable pole, leaving $\dfrac{K}{(s - 3.16)(s + \delta) + K}$
    * Now we can choose δ and K to move both poles of this second-order system to the LHP

Tracking

- We want to make the output follow the reference input as closely as possible, in effect having a unity transfer function from reference to output
- For open-loop control, we again have $T_{ol} = \dfrac{b(s)}{a(s)}\dfrac{c(s)}{d(s)}$
  - Designing the controller to cancel the plant's transfer function is only possible under the constraints:
    * The plant needs to be stable (and stable poles cannot be too close to the imaginary axis)
      - Trying to cancel out stable poles close to the imaginary axis may make the system too sensitive and cause unstable transients
    * The plant should have no zeroes in the RHP (since we'd need an RHP pole to cancel that)
    * The controller transfer function must be proper so it can be physically realized (it must be causal)
      - If the plant is strictly proper, this can be an issue since the controller would have to be improper
      - Digital controllers may be an exception
    * The controller cannot go beyond the plant's actuation limit (the response can't be too fast, or excite plant's resonance modes)
      - This will cause the system to be no longer linear
- For a closed-loop control system, most of the same restrictions apply, but we have more freedom to tune the response

Regulation

- Regulation is the ability of the control system to keep the error small when the input is constant, with added disturbances/noise
- In the open-loop case, the controller has no influence whatsoever on the effect of $W(s)$ on the output
- For the closed-loop controller: $E_{cl} = (1 - \mathcal{T})R - G\mathcal{S}W + H\mathcal{T}V$
  - $E_{cl} = \dfrac{1 + G(s)D_{cl}(s)(H(s) - 1)}{1 + H(s)G(s)D_{cl}(s)}R - \dfrac{G(s)}{1 + H(s)G(s)D_{cl}(s)}W + \dfrac{H(s)G(s)D_{cl}(s)}{1 + H(s)G(s)D_{cl}(s)}V$

- Notice that if D is large, the second term is small so effect of W is small, but the third term gets closer to 1, so the effect of V is not reduced
- Conversely if D is small we have less effect of V but more of W
- To address this, we can design $D_{cl}(s)$ to have large values at low frequencies and small values at high frequencies, since W is often low frequency and V is often high

Sensitivity

- The robustness of the system against variations in the plant behaviour
- Assume that the plant transfer function can change from $G(s)$ to $G(s) + \delta G(s)$
- The *sensitivity* of the (overall) system transfer function T to plant G is defined as $\mathcal{S}_G^T = \frac{\frac{\delta T}{T}}{\frac{\delta G}{G}} = \frac{G}{T} \cdot \frac{\delta T}{\delta G}$
  - This is the ratio of the normalized change to the overall transfer function to the normalized change to the plant transfer function
- For open-loop control:
  - $T_{ol} + \delta T_{ol} = D_{ol}(G + \delta G) = D_{ol}G + D_{ol}\delta G = T_{ol} + D_{ol}\delta G = T_{ol} + \frac{T_{ol}}{G}\delta G \implies \delta T_{ol} = \frac{T_{ol}}{G}\delta G$
    - $\frac{\delta T_{ol}}{T_{ol}} = \frac{\delta G}{G}$ so the sensitivity is 1
  - i.e. whatever change happens in the plant, it will be immediately reflected in the entire system
- For closed-loop control, $\mathcal{S}_G^{T_{cl}} = \frac{G}{T_{cl}} \frac{\delta T_{cl}}{\delta G} = \frac{G}{T_{cl}} \cdot \frac{\mathrm{d}T_{cl}}{\mathrm{d}G}$
  - We can show that $\mathcal{S}_G^{T_{cl}} = \frac{1}{1 + HGD_{cl}}$
    - This is why we define the sensitivity transfer function as $\mathcal{S} = \frac{1}{1 + HGD_{cl}}$
    - The sensitivity is not 1 but is mitigated by the additional term in the denominator
    - The larger the controller D_{cl}, the more robust it is to changes in the plant
  - The complementary sensitivity transfer function is $\mathcal{T} = \frac{GD_{cl}}{1 + HGD_{cl}}$
    - Notice that this is just the closed-loop transfer function
    - This is named so because for the case of a perfect sensor $H(s) = 1$, $\mathcal{S} + \mathcal{T} = 1$

Lecture 14, Feb 29, 2024

Control System Type

- The reference input $R(s)$ can often be approximated by a time domain polynomial $r(t) = Ct^k 1(t)$
  - e.g. for position $k = 0$, for velocity $k = 1$ and for acceleration $k = 2$
- The *type* of a closed-loop controller is the maximum order of the polynomial that the controller can follow with a constant error
  - e.g. if the system can follow a ramp with constant error, then it is a type 1 system
  - Any inputs of a higher order will lead to increasing error
  - Any inputs of a lower order will lead to zero error
- For unity feedback (i.e. $H(s) = 1$ or perfect sensors) and no disturbance ($W = V = 0$), the type of a system depends on the number of poles that its open loop transfer function, $HGD_{cl} = GD_{cl}$, has at the origin
  - $E_{cl}(s) = R(s) - Y(s) = \frac{1}{1 + GD_{cl}} R = \mathcal{S}(s)R(s)$
  - Let the reference input $r(t) = \frac{1}{k!}t^k 1(s) \implies R(s) = \frac{1}{s^{k+1}}$
  - $e_{ss} = \lim_{t \to \infty} e(t) = \lim_{s \to 0} sE_{cl}(s) = \lim_{s \to 0} s \frac{1}{1 + GD_{cl}} \frac{1}{s^{k+1}}$
  - First consider if GD_{cl} has no pole at the origin
    - With $k = 0$, $e_{ss} = \lim_{s \to 0} s \frac{1}{1 + GD_{cl}} \frac{1}{s} = \frac{1}{1 + GD_{cl}(0)} = \frac{1}{1 + K_0}$

        - Therefore for a step input we get a constant steady state error
    * For $k > 0$, $e_{ss} = \lim_{s \to 0} s \frac{1}{1 + GD_{cl}} \frac{1}{s^{k+1}} = \frac{1}{1 + GD_{cl}(0)} \lim_{s \to 0} \frac{1}{s^k} = \infty$
        - For any higher degree input, the error goes to infinity
  - Now consider $GD_{cl}(s) = \frac{\overline{GD}_c(s)}{s^n}$
    * $\overline{GD}_c(s)$ contains all terms of $GD_{cl}(s)$ except for poles at the origin, so $K_n = \overline{GD}_c(0)$ is a finite value
    * For $n = k = 0$ (type 0) we've shown above that $e_{ss} \to 0$
    * For $n = k \neq 0$, $e_{ss} = \lim_{s \to 0} s \frac{1}{1 + \frac{\overline{GD_{cl}}(s)}{s^n}} \frac{1}{s^{k+1}} = \lim_{s \to 0} \frac{s^n}{s^k(s^n + \overline{GD_{cl}}(s))} = \frac{1}{\overline{GD_{cl}}(0)} = \frac{1}{K_n}$
    * For $n > k$, $e_{ss} = \lim_{s \to 0} \frac{s^n}{s^k(s^n + \overline{GD_{cl}}(s))} = \frac{1}{\overline{GD_{cl}}(0)} \lim_{s \to 0} s^{n-k} = 0$
    * For $n < k$, $e_{ss} = \lim_{s \to 0} \frac{s^n}{s^k(s^n + \overline{GD_{cl}}(s))} = \frac{1}{\overline{GD_{cl}}(0)} \lim_{s \to 0} \frac{1}{s^{k-n}} = \infty$
- The type of a system is a *robust property*, i.e. it is independent of the parameters of the system
- For a type 0 system, we can define a *position constant*, $K_p = K_0 = \lim_{s \to 0} GD_{cl}(s)$, so $e_{ss} = \frac{1}{1 + K_0}$ (known as the *position error constant*)
  - Note that this is the only one where the error constant is not a simple reciprocal
- For a type 1 system, we can define a *velocity constant*, $K_v = K_1 = \lim_{s \to 0} sGD_{cl}(s)$, so $e_{ss} = \frac{1}{K_1}$
- For a type 2 system, we can define an *acceleration constant*, $K_a = K_2 = \lim_{s \to 0} s^2 GD_{cl}(s)$, so $e_{ss} = \frac{1}{K_2}$
- Example: plant $G(s) = \frac{A}{\tau s + 1}$ with controller $D_{cl}(s) = k_P + \frac{k_I}{s}$
  - $GD_{cl}(s) = \frac{A(k_P s + k_I)}{s(\tau s + 1)}$ so this is a type 1 system
  - The velocity constant is $K_v = \lim_{s \to 0} sGD_{cl}(s) = Ak_I$ so the steady-state error is $\frac{1}{Ak_I}$
- For non-unity feedback, $E_{cl}(s) = R(s) - Y_{cl}(s) = \frac{1 + (H-1)GD_{cl}}{1 + HGD_{cl}} R = (1 - \mathcal{T}(s))R(s)$
  - $e_{ss} = \lim_{s \to 0} s(1 - \mathcal{T}(s))R(s) = \lim_{s \to 0} \frac{1 - \mathcal{T}(s)}{s^k}$
  - We have to explicitly check the type by finding the largest value of k that keeps e_{ss} finite
  - However, the relationship between the position/velocity/acceleration constants and the steady state error still holds
- Typing a system can also be done with respect to regulation, i.e. setting $R = V = 0$ and finding the highest order of disturbance W that leads to a finite steady state error; in this case the type is determined by the number of zeroes in the error transfer function
  - $E_{cl}(s) = R(s) - Y(s) = -\frac{G(s)}{1 + H(s)G(s)D_{cl}(s)} W \implies \frac{E_{cl}(s)}{W(s)} = -\frac{G(s)}{1 + H(s)G(s)D_{cl}(s)} = -T_w(s)$
    * Note the negative sign in the definition, so that $Y(s) = T_w(s)W(s)$
  - The type is the number of zeroes of $T_w(s)$ at the origin (instead of poles!)
  - Let $W(s) = \frac{1}{s^{k+1}}$ and $T_w(s) = s^m \tilde{T}_w(s)$ where $\tilde{T}_w(0) = \frac{1}{K_{m,w}}$
  - $-e_{ss} = y_s s = \lim_{t \to \infty} y(t) = \lim_{s \to 0} sT_w(s)W(s) = \lim_{s \to 0} \tilde{T}_w(s) \frac{s^m}{s^k}$
  - Now we can see that $m > k \implies y_s s \to 0$, $m < k \implies y_{ss} \to \infty$ and $m = k \implies y_s s = \frac{1}{K_{m,w}}$
- Generally, the type of a system with respect to tracking can be different than the type with respect to regulation, so we must specify when stating the type
- We can also define a transfer function in terms of the noise, $\frac{Y(s)}{V(s)} = -H(s)\mathcal{T}(s) = T_v(s)$, assuming $R = W = 0$

- For the noise however the use of a polynomial input is less realistic, since noise is usually very high in frequency

Lecture 15, Mar 4, 2024

PID Controllers

- For the following analyses we will assume unity feedback, but this is easily extended to other kinds of feedback

Proportional Control (P)

- The simplest controller simply applies a gain to the error feedback: $u(t) = k_P e_a(t) \implies D_{cl} = k_P$
- Consider a second order plant $G(s) = \dfrac{K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \implies GD_{cl}(s) = \dfrac{k_P K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
  - No poles at the origin in the open-loop transfer function, therefore this is a type 0 system
- Closed loop transfer function: $\dfrac{Y(s)}{R(s)} = \dfrac{GD_{cl}(s)}{1 + GD_{cl}(s)} = \dfrac{k_P K\omega_n^2}{s^2 + 2\zeta\omega_n s + (1 + k_P K)\omega_n^2}$
  - Notice that the new natural frequency is $\omega_n' = \sqrt{1 + k_P K\omega_n^2}$, which is increased
  - The new damping ratio is $\zeta' = \dfrac{\zeta}{\sqrt{1 + k_P K}}$, which is decreased (obtained by comparing $2\zeta'\omega_n'$ with $2\zeta\omega_n$)
  - Increased natural frequency leads to shorter rise time (faster system), but decreased damping leads to more overshoot
- For $R(s) = \dfrac{1}{s}$, we have $e_{ss} = \lim_{s \to 0} s\dfrac{1}{1 + GD_{cl}}\dfrac{1}{s} = \dfrac{1}{1 + k_P K}$
  - The steady-state error in the step response is reduced, but not eliminated entirely
- The same analysis can be made for the disturbance regulation

Integral Control (I)

- The integral controller applies $u(t) = k_I \displaystyle\int_0^t e_a(\tau)\,\mathrm{d}\tau$
  - Instead of the error itself, the control signal is proportional to the area underneath the error curve
- The controller transfer function is $\dfrac{U(s)}{E_a(s)} = D_{cl}(s) = \dfrac{k_I}{s}$
- Consider the same second-order plant $G(s) = \dfrac{K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$:
- Closed loop transfer function: $\dfrac{Y(s)}{R(s)} = \dfrac{k_I G(s)}{s + k_I G(s)} = \dfrac{k_I K\omega_n^2}{s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + k_I K\omega_n^2} = \mathcal{T}(s)$
  - Notice that this system is now third order; the increased order of the system makes it more sluggish, so the rise time increases and overshoot decreases
  - Taking $s \to 0$ we see that the DC gain is 1, so there is no more steady-state error
  - Unlike the second-order system, we can no longer conclude that the system is always stable, since this is a third-order system
  - Using the Routh criterion, we find that $k_I < \dfrac{2\zeta\omega_n}{K}$ is the maximum value of k_I for the system to be stable
    * The integral controller can destabilize the system!
  - The removal of steady-state error is a robust property, holding regardless of the value of k_I and plant parameters
    * We can find the sensitivity transfer function and show that this always goes to 0 as $s \to 0$
- $GD_{cl}(s) = \dfrac{k_I K\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$; $E_{cl}(s) = \dfrac{1}{1 + GD_{cl}}R(s)$
  - There is one pole at the origin, so this is a type 1 system

  - This can now follow a position setpoint with no error, and a velocity setpoint with constant error
  - Note we only get zero error for position setpoints when we have unity feedback!
  - The velocity constant is $K_v = \lim_{s \to 0} sGD_{cl}(s) = k_I K \implies e_{ss} = \frac{1}{k_I K}$

Derivative Control (D)

- The derivative controller applies $u(t) = k_D \dot{e}_a(t) \implies \frac{U(s)}{E_a(s)} = D_{cl}(s) = k_D s$
  - Derivative control tends to speed up the system, since it anticipates future behaviour of the system
- The closed-loop transfer function is $\frac{k_D K \omega_n^2 s}{s^2 + (2\zeta + k_D K \omega_n)\omega_n s + \omega_n^2}$ for the second-order plant
  - The additional zero speeds up the system and increases the overshoot
  - However, the damping ratio is increased to $\zeta' = \zeta + \frac{1}{2} k_D K \omega_n$, which decreases the overshoot
  - Overall, the combination leads to increased system speed and decreased overshoot
  - Furthermore, increased damping ratio and constant natural frequency moves the poles away from the imaginary axis, enhancing stability
- $GD_{cl}(s) = \frac{k_D K \omega_n^2 s}{s^2 + 2\zeta\omega_n s + \omega_n^2}$
  - No poles at zero, therefore the system is type 0 and maintains constant error for a step input
  - The position constant is $K_p = \lim_{s \to 0} GD_{cl}(s) = 0$
  - The steady state error is $e_{ss} = \frac{1}{1 + K_p} = 1$
    * This means that the output ultimately converges to zero, i.e. the derivative controller can't do anything about the steady state error
- Generally, derivative control enhances the transient behaviour of the system but does nothing to its long-term behaviour
- The transfer function for the controller is not causal
  - This means we can't implement it with analog controllers
  - We can still implement this digitally, but taking numerical derivatives highly amplifies noise
  - Therefore, in reality derivative controllers may not be practical
  - Practically, we use another technique called lead functions instead of derivatives
- Derivative control can be used to damp the control response, so we don't get sharp reactions to suddenly changing signals
  - If there is a sudden jump in the output due to transient effects, there will be a jump in error and also $u(t)$, which is not desirable
  - A derivative feedback path will correct for this

Proportional-Integral Control (PI)

- $u(t) = k_P e_a(t) + k_I \int_0^t e_a(\tau)\,\mathrm{d}\tau \implies D_{cl}(s) = k_P + \frac{k_I}{s}$
- The closed-loop transfer function is $\frac{(k_P s + k_I)K\omega_n^2}{s^3 + 2\zeta\omega_n s^2 + (1 + k_P K)\omega_n^2 s + k_I K \omega_n^2}$
  - We still increase the system order, but also added a zero, which counteracts the slowdown effect
    * The final system can be faster than the initial plant
  - There is a zero that we can use to cancel a stable pole, which would make the system behave like second-order, making it easier to analyze and control
- $GD_{cl}(s) = \frac{(k_P s + k_I)K\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$
  - The system is type 1, with $K_v = k_I K$ and steady-state error $e_{ss} = \frac{1}{k_I K}$
  - Also type 1 in regulation
- Stability criterion: $k_I < \frac{2\zeta\omega_n(1 + k_P K)}{K}$

- Note that we only have two adjustable parameters k_P and k_I, but there are 3 roots, so our ability to control the characteristic equation is limit
- Generally used to allow for a faster response compared to a pure integral controller

Proportional-Derivative-Integral Control (PID)

- $u(t) = k_P e_a(t) + k_I \int_0^t e_a(\tau)\,\mathrm{d}\tau + k_D \dot{e}_a(t) \implies D_{cl}(s) = k_P + \frac{k_I}{s} + k_D s$
- Second-order closed loop: $\dfrac{(k_D s^2 + k_P s + k_I)K\omega_n^2}{s^3 + (2\zeta + k_D K\omega_n)\omega_n s^2 + (1 + k_P K)\omega_n^2 s + k_I K\omega_n^2}$
  - We can fully control the location of the poles since there are 3 poles and we have 3 parameters
- $GD_{cl}(s) = \dfrac{(k_D s^2 + k_P s + k_I)K\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$
  - The system is type 1, and has $k_v = k_I K \implies e_{ss} = \dfrac{1}{k_I K}$
- Stability criterion: $k_I < \dfrac{(2\zeta + k_D K\omega_n)(1 + k_P K)\omega_n}{K}$

Example System: DC Servo Motor

- Consider the DC motor system derived earlier
- $l_a \dfrac{\mathrm{d}i_a}{\mathrm{d}t} + r_a i_a = v_a - K_e\dot{\theta}_m \implies (l_a s + r_a)I_a(s) = V_a(s) - K_e s\theta_m(s)$
  - This models the back EMF and inductive/resistive effects of the motor coil
- $J_m\ddot{\theta}_m + b_m\dot{\theta}_m = K_t i_a - \eta T \implies (J_m s + b_m)s\Theta_m(s) = K_t I_a(s) - \eta T(s)$
  - This models torque on the shaft, including friction and an external resisting force
- $V_a(s)$ is the input to the system, $T(s)$ is a disturbance, and $\Theta_m(s)$ is the final output
  - $V_a(s)$ first passes through a transfer function to get $I_a(s)$, then this is multiplied by K_t to get a torque
  - This is summed with the torque from the disturbance and passes through the mechanical transfer function to get $\dot{\theta}_m$
  - A final integrator gets us $\Theta_m(s)$
  - The back EMF introduces a feedback path with constant gain K_e

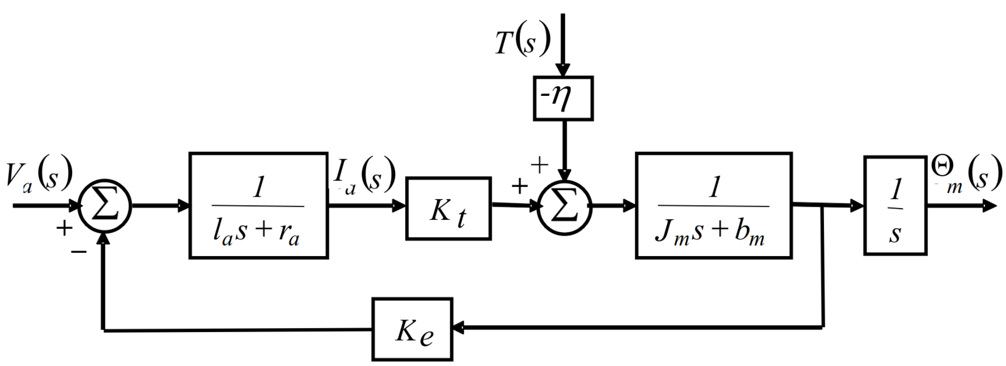


Figure 39: Block diagram for the DC motor system.

- The system has two inputs (V_a and T), since it is linear we can consider one at a time to get transfer functions
  - $\dfrac{\Theta_m(s)}{V_a(s)} = \dfrac{K_t}{s((l_a s + r_a)(J_m s + b_m) + K_c K_t)}$
  - $\dfrac{\Theta_m(s)}{T(s)} = \dfrac{-\eta(l_a s + r_a)}{s((l_a s + r_a)(J_m s + b_m) + K_e K_t)}$
- We can again make the simplifying assumption that the electrical part of the system operates on a much faster time scale than the mechanical part, so the inductance l_a can be taken to 0

- $\dfrac{\Theta_m(s)}{V_a(s)} = \dfrac{\frac{K_t}{r_a}}{s\left(J_m s + \left(b_m + \frac{K_e K_t}{r_a}\right)\right)} = \dfrac{K}{s(\tau s + 1)}$
- $\dfrac{\Theta_m(s)}{T(s)} = \dfrac{-\eta}{s\left(J_m s + \left(b_m + \frac{K_e K_t}{r_a}\right)\right)} = \dfrac{C}{s(\tau s + 1)}$
- $\tau = \dfrac{J_m r_a}{b_m r_a + K_e K_t}$
- $K = \dfrac{K_t}{b_m r_a + K_e K_t}$
- $C = \dfrac{-\eta r_a}{b_m r_a + K_e K_t}$
- We can now build a much simpler block diagram

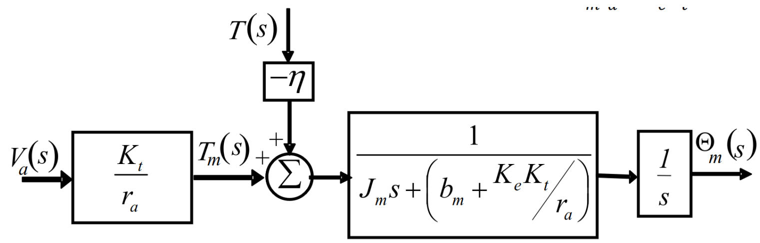


Figure 40: Simplified block diagram for the DC motor system.

- Now we close the loop, with a feedback transfer function $H(s) = hs$ and controller $D_{cl}(s)$
  - We can do this for either a position or velocity controller
  - Since we have non-unity feedback, we can no longer only look at the poles to tell the system type and must use brute force
- For position control: $\dfrac{\Theta_m(s)}{\Theta_{mr}(s)} = \dfrac{KD_{cl}(s)}{s(\tau s+1) + KhD_c(s)}$, $\dfrac{\Theta_m(s)}{T(s)} = \dfrac{-K\eta}{s(\tau s+1) + KhD_c(s)}$
  - Consider a PID controller: $D_{cl}(s) = k_P + \dfrac{k_I}{s} + k_D s$
  - $\mathcal{T}(s) = \dfrac{\Theta_m(s)}{\Theta_{mr}(s)} = \dfrac{Kk_P + \frac{Kk_I}{s} + Kk_D s}{\tau s^2 + s + Khk_P + \frac{Khk_I}{s} + Khk_D s} = \dfrac{\frac{1}{\tau}(Kk_D s^2 + Kk_P s + Kk_I)}{s^3 + \frac{1}{\tau}(Khk_D + 1)s^2 + \frac{1}{\tau}Khk_P s + \frac{1}{\tau}Khk_I}$
  - $T_w(s) = \dfrac{\Theta_m(s)}{T_s(s)} = \dfrac{-K\eta}{\tau s^2 + s + Khk_P + \frac{Khk_I}{s} + Khk_D s} = \dfrac{-\frac{1}{\tau}K\eta s}{s^3 + \frac{1}{\tau}(Khk_D + 1)s^2 + \frac{1}{\tau}Khk_P s + \frac{1}{\tau}Khk_I}$
- For tracking: $E(s) = \Theta_{mr}(s) - \Theta_m(s)$

$$= (1 - \mathcal{T}(s))\Theta_{mr}(s)$$

$$= \left(\frac{s^3 \frac{1}{\tau}(Khk_D + 1)s^2 + \frac{1}{\tau}Khk_P s + \frac{1}{\tau}Khk_I - \frac{1}{\tau}Kk_D s^2 - \frac{1}{\tau}Kk_P s - \frac{1}{\tau}Kk_I}{s^3 + \frac{1}{\tau}(Khk_D + 1)s^2 + \frac{1}{\tau}Khk_P s + \frac{1}{\tau}Khk_I}\right)\Theta_{mr}(s)$$

  - For a step $\Theta_{mr}(s) = \dfrac{1}{s}$, so $e_{ss} = \lim_{s\to 0} sE(s) = \dfrac{\frac{1}{\tau}Kk_I(h-1)}{\frac{1}{\tau}Khk_I} = \dfrac{h-1}{h}$
  - We have a constant error, so this system is only type 0
  - Even though we have an integral term, the error was not reduced to 0 because the system is not unity feedback
    * If the system was unity feedback, then $h = 0$ and we would have a type 1 system
- For regulation: $E(s) = \Theta_{mr}(s) - \Theta_m(s)$

$$= -\Theta_m(s)$$

$$= -T_w T(s)$$

  - For a step disturbance $T(s) = \dfrac{1}{s}$
    * $e_{ss} = \lim_{s\to 0} sE(s) = \lim_{s\to 0} -T_w(s) = 0$
    * This system is higher than type 0

- For a ramp disturbance $T(s) = \frac{1}{s^2}$
  * $e_{ss} = \lim_{s \to 0} -sT_w(s) = \frac{\frac{1}{\tau}K\eta}{\frac{1}{\tau}Khk_I} = \frac{\eta}{hk_I}$
  * Therefore this system is type 1 with respect to regulation
  * The velocity constant is $\frac{1}{e_{ss}} = \frac{hK_I}{\eta}$

- In summary, the system is type 0 with respect to tracking and type 1 with respect to regulation for PID
  - For PI, we have the same type 0 in tracking and type 1 in regulation
  - For P, we have type 0 in both regulation and tracking
- The same analysis can be applied for velocity control, where our feedback will be taken from $\Omega_m(s)$, the speed of the shaft
  - Construct the same transfer functions for regulation and tracking
  - For velocity control, we also have the same types with the controllers
- In general, for PI and PID control the system type is usually the same

Lecture 16, Mar 7, 2024

PID Controllers, Continued

Ziegler-Nichols Tuning Method

- While we can find gain values through theoretical analysis of a system, we don't often know the transfer functions perfectly, so fine-tuning on top of theoretical gains is often needed
- For PID tuning, we rely on mostly heuristic methods (instead of rigorous theoretical methods)
- For a PID controller, do the following in order:
  - Use k_P to decrease the rise time
  - Use k_D to reduce the overshoot and settling time
  - Use k_I to eliminate the steady-state error (while keeping the system stable)

Response	Rise Time	Overshoot	Settling Time	S-S Error
k_P	Decrease	Increase	NT	Decrease
k_I	Increase	Decrease	Increase	Eliminate
k_D	NT	Decrease	Decrease	NT

Figure 41: Effect of increasing each of the PID gains.

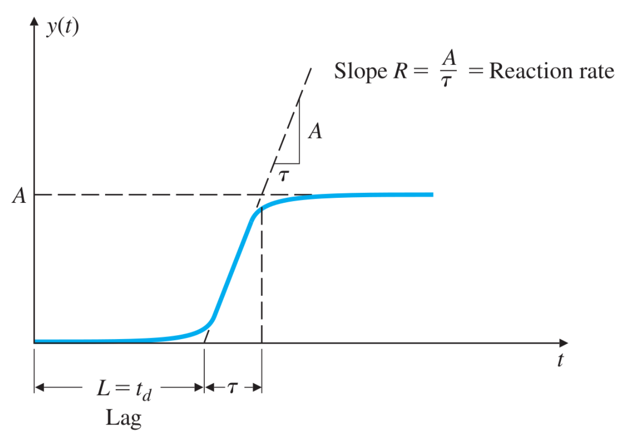


Figure 42: Process reaction curve.

Type of Controller	Optimum Gain
P	$k_P = 1/RL$
PI	$\begin{cases} k_P = 0.9/RL \\ T_I = L/0.3 \end{cases}$
PID	$\begin{cases} k_P = 1.2/RL \\ T_I = 2L \\ T_D = 0.5L \end{cases}$

Ziegler–Nichols Tuning for the Regulator $D_c(s) = k_P(1 + 1/T_I s + T_D s)$, for a Decay Ratio of 0.25

Figure 43: Ziegler-Nichols table.

- The *Ziegler-Nichols* method is an empirical tuning method that gives a set of gains from empirical observations of the system only
  - This works well for plants that don't have poles at the origin, or dominant complex poles near the origin
    * This is because these plants are stable, and oscillatory components of the response are not dominant
  - The plant's behaviour should be well approximated by $\frac{Y(s)}{U(s)} = \frac{Ae^{-t_d s}}{\tau s + 1}$
    * The $e^{-t_d s}$ is a delay by t_d
    * This is saying that the step response first has some delay $L = t_d$, and then rises with an approximate slope $R = \frac{A}{\tau}$ until it reaches the DC gain of A
    * This is known as a *process reaction curve* and is characterized by L and R
- Given a plant, we can inject it with a step input and measure its response and derive L and R
  - The Ziegler-Nichols method gives a set of PID gains based on L and R only
  - The gains of a PID controller $D_{cl}(s) = k_P\left(1 + \frac{1}{T_I s} + T_D s\right)$ can be looked up in the table
- This method doesn't apply for all plants, especially not those that are unstable
- Theoretically, we can show that Ziegler-Nichols creates a system response with 25% decay ratio (ratio of the first overshoot to the second overshoot), about equivalent to $\zeta \approx 0.21$ for a second-order system
  - This damping ratio leads to around 50% overshoot
  - We can usually reduce k_P by 50% after tuning to reduce overshoot/oscillations without affecting the other properties much
- The method was first derived purely empirically, but it can be shown that the resulting gain values are close to those derived from optimal control, where we minimize the energy of the controller
- If it's impractical to observe the system's step response (e.g. unstable system), we can instead use the *ultimate sensitivity* method
  1. Close the loop with only a proportional controller with gain k_P, so the system is stable
  2. Increase k_P until the system enters a steady oscillation in response to a step input
     - The gain at which this happens is the *ultimate gain* K_u, and the oscillation period is the *ultimate period* P_u
  3. Look up values for the system gains based on the ultimate gain and ultimate period from the table
     - Again we can often reduce k_P by half to reduce oscillations

- Example: heat exchanger; we control a valve which varies the amount of steam into the tank, which adjusts the temperature of the water at the tank exit
  - Typical fluids systems are similar to underdamped second-order systems
  - Assume $T_m = T_w(t - t_d)$ (a delay) so $\frac{T_m(s)}{A_s(s)} = \frac{Ke^{-t_d s}}{(\tau_1 s + 1)(\tau_2 s + 1)}$
    * $a_s(t)$ is the amount that we open the valve by

Ziegler-Nichols Tuning for the Regulator $D_c(s) = k_P(1 + 1/T_I s + T_D s)$, Based on the Ultimate Sensitivity Method

Type of Controller	Optimum Gain
P	$k_P = 0.5K_u$
PI	$\begin{cases} k_P = 0.45K_u \\ T_I = \frac{P_u}{1.2} \end{cases}$
PID	$\begin{cases} k_P = 1.6K_u \\ T_I = 0.5P_u \\ T_D = 0.125P_u \end{cases}$

Figure 44: Ziegler-Nichols table for ultimate sensitivity.

- Assume that we give a step input to the plant and its output is shown in the figure below
  * $L \approx 13$ (very short delay)
  * $R \approx \dfrac{1}{90}$
    • If we take the tangent when the response is increasing, it takes about 90 seconds to hit 1
  * For P control we have $k_P = \dfrac{1}{RL} = 6.92$
  * For PI control $k_P = \dfrac{0.9}{RL} = 6.22, T_I = \dfrac{L}{0.3} = 43.3$
- Assume that we use a P controller and increased the gain until we saw steady oscillations in the figure below
  * $K_u \approx 15.3, P_u \approx 42$
  * For P control $k_P = 0.5K_u = 7.65$
  * For PI control $k_P = 0.45K_u, T_I = \dfrac{P_u}{1.2} = 35.0$
  * Notice that the PI controller gains derived from this method resulted in a response with more oscillation

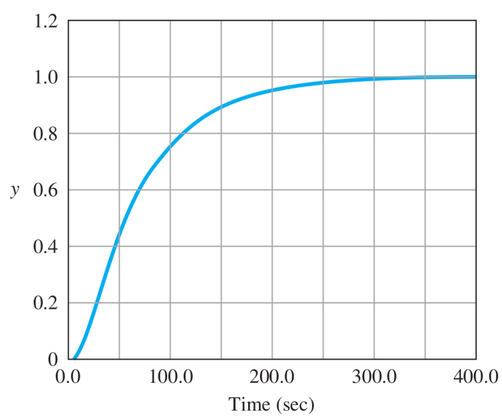


Figure 45: Step response of the example plant.

Feedforward Control

- Using only P doesn't eliminate steady-state error, but using PI to eliminate the error makes the system sluggish, decreases damping and degrades stability
- Another way to eliminate steady-state error is to use a *feedforward controller*, where we first multiply the reference by the inverse DC gain of the plant

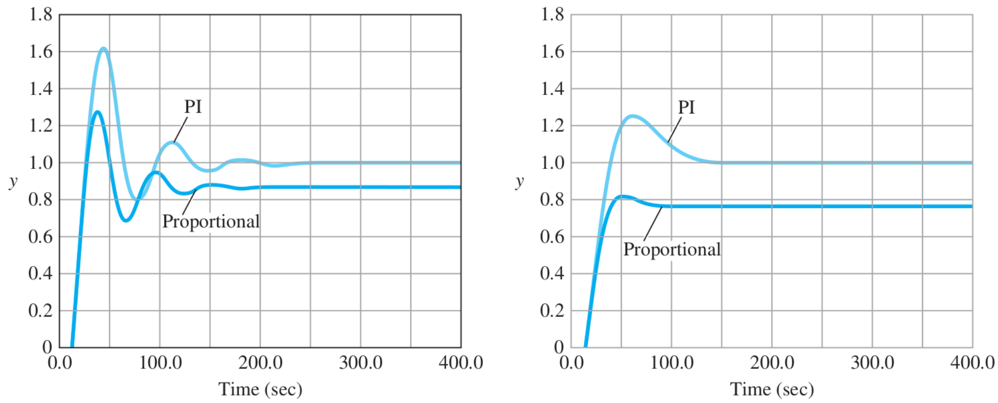


Figure 46: Closed-loop step responses from the controller using the step response method, before and after reducing k_P.

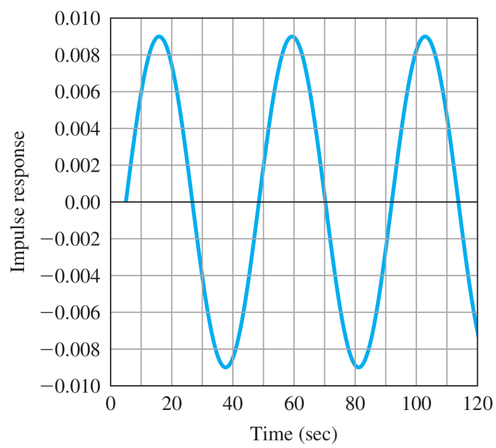


Figure 47: Steady oscillation of the example plant from a P controller.

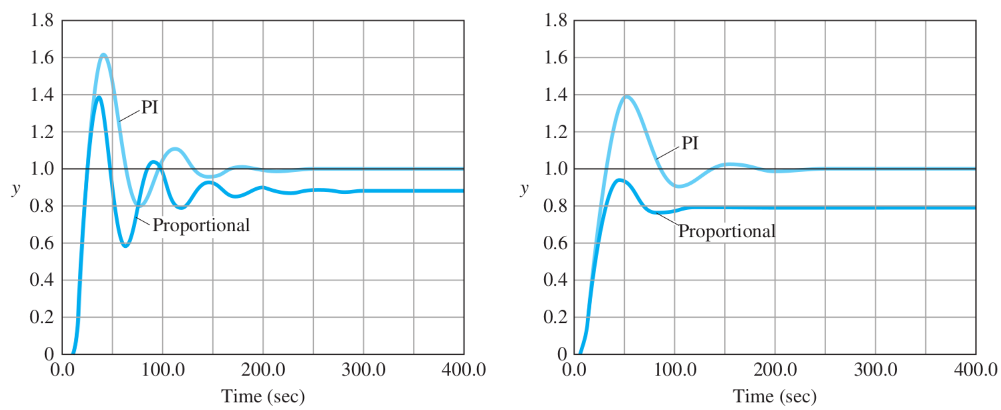


Figure 48: Closed-loop step responses from the controller using ultimate sensitivity, before and after reducing k_P.

- Equivalent to having $(G^{-1}(s) + D_c(s))E_a(s)$ instead of just $D_c(s)E_a(s)$ to the plant
- $Y(s) = G(s)(D_c(s)E(s)G^{-1}(0)R(s)) \implies \dfrac{Y(s)}{R(s)} = \dfrac{(D_c(s) + G^{-1}(0))G(s)}{1 + D_c(s)G(s)}$
- Now when we take $s \to 0$ we get a DC gain of $\dfrac{D_c(0)G(s) + G^{-1}(0)G(0)}{1 + D_c(0)G(0)} = 1$, so there is no steady-state error

* Practically we don't always know $G^{-1}(0)$ exactly, which is why we still need a P/PI controller; the system with just a feedforward is not robust

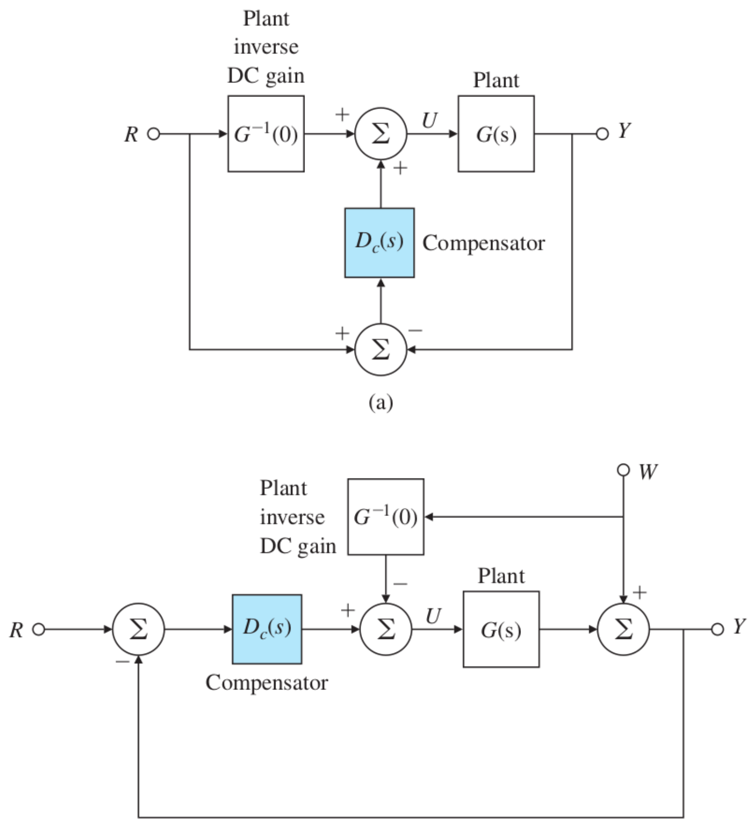


Figure 49: Feedforward controllers for tracking and disturbance rejection.

Lecture 17, Mar 11, 2024

Root-Locus Design Method

* A graphical method (set of rules) for finding the *locus* (set of locations on a line) of the roots of a system's characteristic equation, as a result of changing parameters
  - Allows us to find how the roots of a system move as a result of variation in some system parameter
  - e.g. we can find how the poles move as a result of changing the gain, so we can assess the system's stability, speed, etc
  - The *root locus* is the set of all locations that a root can take as a result of changing some parameter
  - Note the parameter must affect the characteristic equation linearly
* In controls, we use this to find how the roots of the characteristic equation (i.e. the poles) are affected by changing system gains
* Consider the closed-loop transfer function $\dfrac{Y(s)}{R(s)} = \mathcal{T}(s) = \dfrac{D_c(s)G(s)}{1 + D_c(s)G(s)H(s)}$
  - Rewrite the characteristic equation into the form of $1 + D_c(s)G(s)H(s) = a(s) + Kb(s) = 0$

- Then we have $1 + K\frac{b(s)}{a(s)} = 0 \implies 1 + KL(s) = 0$ where $L(s) = \frac{b(s)}{a(s)} = -1\frac{1}{K}$
    - * Writing $L(s) = \frac{b(s)}{a(s)}$ is known as the root-locus or *Evans form*
  - Now our poles are locations where $L(s) = -\frac{1}{K}$, which is often a negative real number
  - Since the original poles are at $D_c(s)G(s)H(s) = -1$, $KL(s) = D_c(s)G(s)H(s)$, the open-loop transfer function
    - * Sometimes we will just refer to the open-loop transfer function as $L(s)$ and ignore the K
  - Most often K is a positive real number since it is a gain, but in rare cases we can also deal with $K < 0$
- The roots of the characteristic equation are located where the open-loop transfer function of the system becomes a real negative value
  - Therefore we can plot the location of all possible roots s of the characteristic equation by varying K; this is the root locus
  - The root locus allows us to select the best controller gains and study the effect of potentially adding additional poles and zeros
- Let $b(s) = s^m + b_1 s^{m-1} + \cdots + b_m = \prod_{i=1}^{m}(s - z_i), a(s) = s^n + a_1 s^{n-1} + \cdots + a_n = \prod_{i=1}^{n}(s - p_i)$
  - z_i are the open-loop zeroes, p_i are the open-loop poles
  - Note $n \geq m$ because $L(s) \propto D_c(s)G(s)H(s)$ is causal
- Let $a(s) + Kb(s) = \prod_{i=1}^{n}(s - r_i)$ (note $n \geq m$ so the summation ends at n)
  - The r_i are the closed-loop poles; note this is not the same as the open-loop poles
  - Our goal is to draw all the possible locations of r_i for different values of K
- Example: $D_c(s) = K$, $G(s) = \frac{1}{s(s+c)}$ and consider $c = 1$; plot the root locus with respect to K
  - $G(s)D_c(s) = \frac{K}{s^2 + s} \implies \mathcal{T}(s) = \frac{K}{(s^2 + s) + K} = \frac{K}{a(s) + Kb(s)}$
  - We have $b(s) = 1, a(s) = s^2 + s \implies m = 0, n = 2, z_i = \varnothing, p_i = \{0, -1\}, r_i = -\frac{1}{2} \pm \frac{\sqrt{1 - 4K}}{2}$
  - $L(s) = \frac{b(s)}{a(s)} = \frac{1}{s(s+1)}$
  - For $K = 0$, we have two real roots $r_1 = -1, r_2 = 0$
    - * Notice that these are the same as the open-loop poles, since $K = 0 \implies a(s) + Kb(s) = a(s)$
  - For $K = \frac{1}{4}, r_1 = r_2 = -\frac{1}{2}$
  - For $K > \frac{1}{4}$ the roots will be imaginary, and the pair of poles will move up further from the real axis
  - The two directions that the poles move in are the 2 *branches*
    - * The branches start at the open loop poles, which are the *start points*
    - * The locus has one *breakaway point*, where the two poles join and separate
      - Note breakaway points are when the poles move in from the real axis
  - Suppose we want $\zeta = 0.5$, geometrically we can draw out a line at an angle $\sin^{-1}\zeta = 30°$ from the origin, and find where it intersects with the root locus

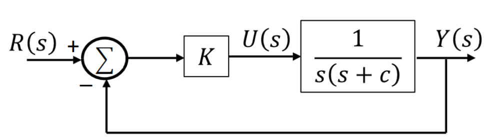


Figure 50: Example feedback control system.

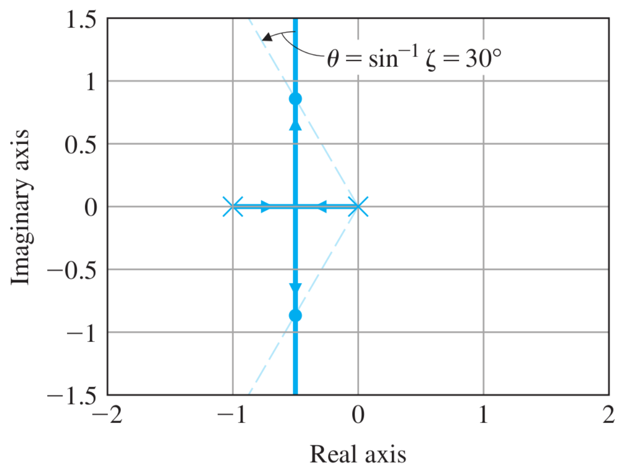


Figure 51: Root locus plot of the example system.

- Example: root locus of the previous system with respect to c
  - $\mathcal{T}(s) = \dfrac{1}{s^2+1+cs} = \dfrac{1}{a(s)+cb(s)} \implies \begin{cases} b(s) = s \\ a(s) = s^2+1 \end{cases}, L(s) = \dfrac{s}{s^2+1}$
  - The roots are $z_i = 0, p_i = \pm j$
  - $a(s)+cb(s) = s^2+cs+1 = 0 \implies r_1, r_2 = -\dfrac{c}{2} \pm \dfrac{\sqrt{c^2-4}}{2}$
  - For $c = 0$ we have $r_1, r_2 = \pm j$, giving the start of the plot
  - For $c = 2$ the two roots meet at $r_1 = r_2 = -1$
  - As $c \to \infty$, one of the poles moves to $-\infty$ while the other converges to 0
  - The circle on the diagram indicates the location of $z_1 = 0$
  - This root locus has 2 start points, 2 branches, and 1 *break-in point* (where the poles meet and separate, but they come from the imaginary axis)

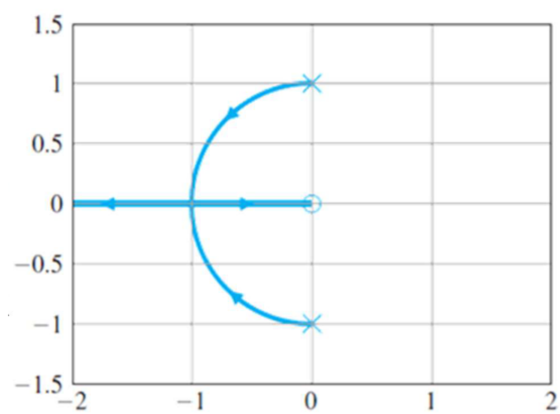


Figure 52: Root locus plot with respect to c for the example system.

Root Locus Determination

Definition

A *root locus* is the set of all possible values of s for which the characteristic equation $1 + KL(s) = 0$ holds, as the real parameter K varies from 0 to ∞ (sometimes $-\infty$). In controls, the characteristic equation is typically for a closed-loop system, so the roots of the locus are the system poles.

- If K is real and positive, then $L(s)$ must be real and negative, so its phase must be $+180°$ (*positive locus*)
  - In rare cases K is negative, then $L(s)$ has a phase of 0 (*negative locus*)
- We can alternatively define the root locus as the set of points in the s-plane where the phase of $L(s)$ equals 180° for positive loci, or 0° for negative loci
  - This will help us plot the locus
- Recall that for $L(s) = \dfrac{b(s)}{a(s)}$ the phase of $L(s)$ is equal to the phase of $b(s)$ minus the phase of $a(s)$
- Consider a test point s_0
  - To find the phase of $L(s_0)$, we need to find the phase of $b(s_0)$ and $a(s_0)$
  - $\angle b(s) = \sum_{i=1}^{m} \angle(s_0 - z_i)$ and $\angle a(s) = \sum_{i=1}^{n} \angle(s_0 - p_i)$
  - We need to check that $\sum_{i=1}^{m} \angle(s_0 - z_i) - \sum_{i=1}^{n} \angle(s_0 - p_i) = 180° + 360°(l - 1)$
  - The phase of each $s_0 - z_i$ is the angle from each open-loop zero to s_0; the phase of each $s_0 - p_i$ is the angle from each open-loop pole to s_0
  - Therefore we take the sum of the angles of s_0 from the open-loop zeros, denoted ϕ_i, and subtract the sums of the angles of s_0 from the open-loop poles, denoted ϕ_i

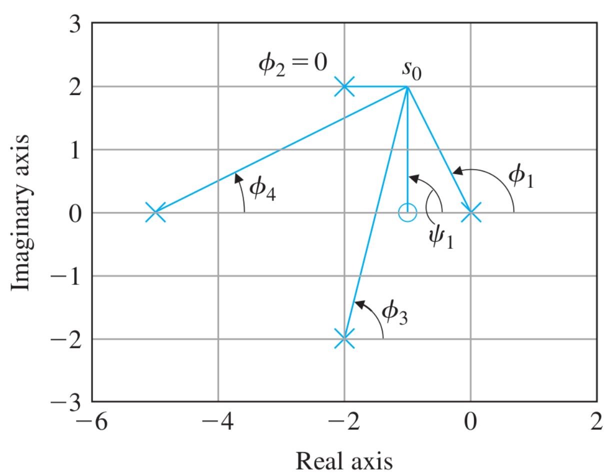


Figure 53: Testing whether a point s_0 is part of the root locus.

Lecture 18, Mar 14, 2024

Root Locus Determination

- Note due to complex conjugate roots, the locus is always symmetric about the real axis
- Given $L(s) = \dfrac{b(s)}{a(s)} = \dfrac{\prod_{i=1}^{n}(s - z_i)}{\prod_{i=1}^{m}(s - p_i)}$, a positive root locus follows the following rules:
  1. There are n branches each starting from the open-loop poles; m of these branches will end at the open-loop zeros of $L(s)$, while the rest go to infinity

- $K \to 0$ means $a(s) + Kb(s) = 0$ is satisfied for $a(s) = 0$, hence the poles start at the open-loop poles
- $K \to \infty$ means $L(s) = \dfrac{b(s)}{a(s)} = -\dfrac{1}{K}$ is satisfied for $b(s) \to 0$ or $a(s) \to \infty$
  * m poles go to the open-loop zeros where $b(s) \to 0$
  * For the other $n - m$ poles, $b(s)$ does not have a zero, so we need $a(s) \to \infty$ to have $L(s) \to 0$
    - $a(s)$ will always outgrow $b(s)$ since the degree $n > m$ for causal systems
- When poles and zeros are repeated, there are multiple branches departing from or arriving at these poles/zeros, one for each degree of multiplicity

2. The segments of the locus on the real axis are always to the left of an odd number of real poles and zeros (on the real axis)
   - For any point on the real axis, the phase angles of conjugate poles or zeros cancel each other, so we need not consider them
   - For poles and zeros on the real axis, having a point to the left of an odd number of them gives a total phase of 180°
     * Having one pole to the right gives a phase from that pole to the point of 180° as required
     * A pole and a zero on the right of the point cancel each other out in phase
     * Two poles or two zeros add to a phase of ±360° and doesn't matter
   - This gives us all segments of the real axis included in the root locus
3. For the $n - m$ poles that must go to infinity, their asymptotes are lines radiating from the real axis at $s = \alpha$ at angles ϕ_l, where:
   - $\alpha = \dfrac{\sum_i p_i - \sum_i z_i}{n - m}$
   - $\phi_l = \dfrac{180° + 360°(l-1)}{n - m}$
   - $l = 1, 2, \dots, n - m$ is the branch number
   - Geometrically this means that the asymptotes evenly divide the 360° and are always symmetric about the real axis; for an odd number of branches, there is always an asymptote towards the negative real axis
4. Each branch departs at an angle of $\phi_{l,d} = \sum_i \psi_i - \sum_{i \neq l} \phi_i - 180°$ from an open-loop pole, where ψ_i are the angles from zeros to the pole, and ϕ_i are angles from the other poles to the pole
   - Note this is exactly the phase condition we need for a point to be on the root locus
   - If the pole is repeated q times, $\phi_{l,d} = \sum_i \psi_i - \sum_{i \neq l} \phi_i - 180° - 360°(l-1)$ for $l = 1, 2, \dots, q$
     * The directions are again spaced evenly apart
   - Similarly, the angles of arrival at a zero are $\psi_{l,a} = \sum \phi_i - \sum_{i \neq l} \psi + 180° + 360°(l-1)$
5. At points where branches intersect (where the characteristic polynomial has repeated roots), if q branches intersect at the point, then their departure angles are $\dfrac{180° + 360°(l-1)}{q}$ plus an offset; together the q branches arriving and q branches departing should form an array of $2q$ evenly spaced rays
   - If the intersection is on the real axis, use Rule 2 to determine the orientation, otherwise use Rule 4
   - Note that it doesn't matter which branch breaks out at which angle
6. The breakaway/break-in points of the locus (i.e. intersection points) are among points where $\dfrac{\mathrm{d}L(s)}{\mathrm{d}s} = 0$
   - Note that some of the solutions are not actually the breakaway/break-in points, so we need to test
   - To determine which of the solutions are actually intersection points, we can use geometry or check with the phase angle method for whether the point is on the locus
   - We can also substitute into $L(s)$ and check that we have a negative real result

– If the multiplicity of the root of $\frac{\mathrm{d}L(s)}{\mathrm{d}s} = 0$ is r, then the multiplicity of the corresponding root in the closed-loop characteristic equation is $q = r + 1$ (i.e. $r + 1$ branches meet)

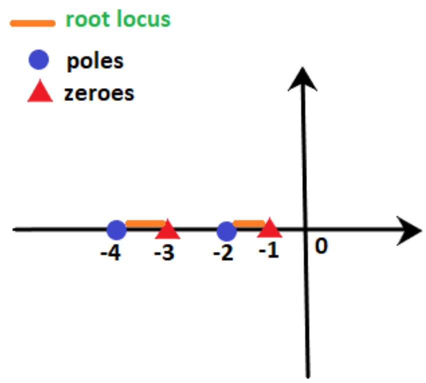


Figure 54: Illustration of Rule 2.

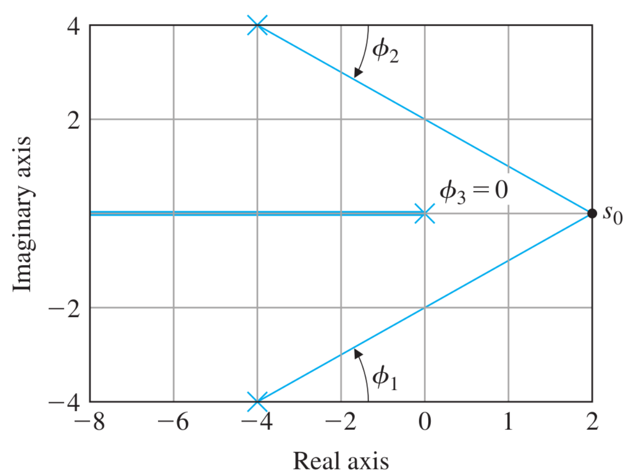


Figure 55: Justification of Rule 2.

- Example: characteristic equation $1 + K\frac{s+1}{s(s+2)(s+3)} = 0$
  - $b(s) = s + 1, m = 1, z_1 = -1$
  - $a(s) = s(s+2)(s+3), n = 3, p_1 = 0, p_2 = -2, p_3 = -3$
  - From rule 1, there are 3 branches, starting from $s = 0, s = -2, s = -3$; one of the branches ends at $s = -1$ while the others go to infinity
  - From rule 2, the segments of the locus on the real axis are at $[-1, 0]$ and $[-3, -2]$
    * Note that the segment $[-1, 0]$ starts at a pole and ends at a zero, so we've found an entire branch
  - From rule 3:
    * Asymptotes radiate from $\alpha = \frac{\sum_i p_i - \sum_i z_i}{n - m} = \frac{0 - 2 - 3 + 1}{3 - 1} = -2$
    * Angles are $\phi_l = \frac{180° + 360°(l-1)}{n - m} = 90° + 180°(l - 1) = 90°, 270°$
    * We have two asymptotes, one pointing vertically upward and one downward, intersecting the real axis at $s = -2$

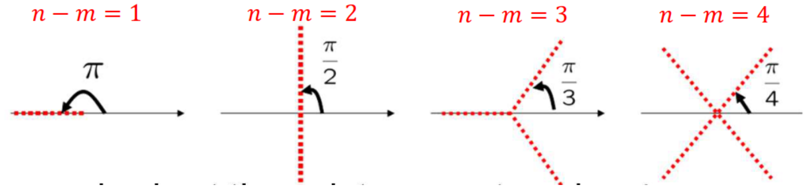


Figure 56: Illustration of Rule 3.

- From rule 5: departure angles are $\dfrac{180° + 360°(l-1)}{2} = 90°, 270°$
- From rule 6: $\dfrac{\mathrm{d}L}{\mathrm{d}s} = \dfrac{-2s^3 - 8s^2 - 10s - 6}{(s(s+2)(s+3))^2} = 0 \implies s = -2.46, -0.77 \pm j0.79$
  * From simple geometric intuition we see that $s = -2.46$ is the real breakaway point, but we can also check the other points and find that K is not real

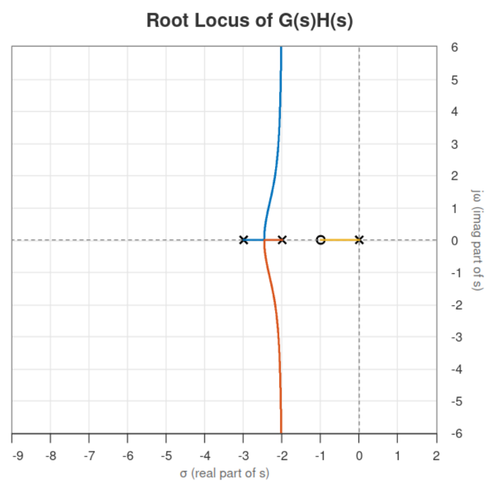


Figure 57: Root locus plot of $1 + K\dfrac{s+1}{s(s+2)(s+3)} = 0$.

Example: Control Gain Selection

- Consider the open-loop transfer function $L(s) = \dfrac{1}{s((s+4)^2+16)}$
- $b(s) = 1, m = 0, z_i = \varnothing$ and $a(s) = s^3 + 8s^2 + 32s, n = 3, p_i = 0, -4 \pm j4$
- In root locus form the characteristic equation is $1 + K\dfrac{1}{s((s+4)^2+16)} = 0$
- Rule 1:
  - We can now mark out the start points of the root locus at $s = 0, s = -4 \pm j4$
  - All 3 branches go to infinity, since we have no zeros
- Rule 2:
  - The segment $(-\infty, 0]$ on the real axis is on the root locus since $p_1 = 0$; this is the complete branch for p_1
- Rule 3:
  - $\alpha = \dfrac{0 - 4 + j4 - 4 - j4}{3 - 0} = -2.67$
  - $\phi_l = \dfrac{180° + 360°(l-1)}{n-m} = 60°, 180°, 300°$
- Rule 4:

- $\phi_{1,d} = \sum_i \psi_i - \sum_{i \neq 1} \phi_i - 180° = 0 - (-45° + 45°) - 180° = -180°$
    * This matches what we had earlier; the entire branch of p_1 consists of the segment going left to minus infinity on the real axis
  - Similarly $\phi_{2,d} = -45°, \phi_{3,d} = +45°$
- Rule 6: omitted here, but if we take $\frac{\mathrm{d}L}{\mathrm{d}s} = 0$ we will find that none of the solutions are points on the locus, so there are no intersections
- Note that these 6 rules don't give us the complete shape, but we gain enough of an intuition about the behaviour of the roots for design

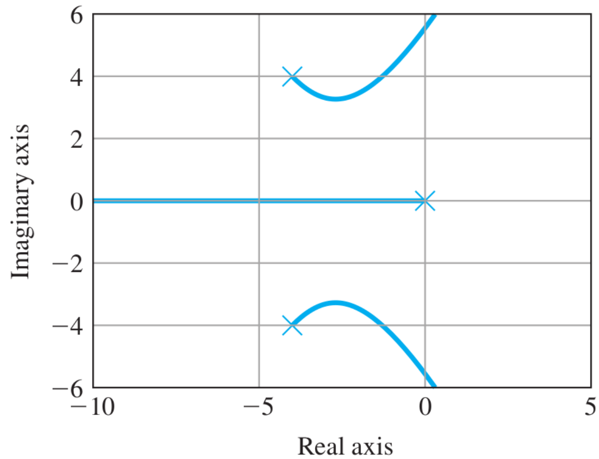


Figure 58: Root locus of $L(s) = \frac{1}{s((s+4)^2+16)}$.

- Now we want to select K such that the system behaves like having $\zeta = 0.5$
  - This means the phase angle of the closed-loop poles should be $\sin^{-1}\zeta = 30°$ (or $\phi_{s_0} = 90° + 30° = 12°$)
  - Using this, we find the intersection with the root locus to find s_0
  - Now we can find K as $K = \frac{1}{|L(s_0)|} = |s_0 - s_1||s_0 - s_2||s_0 - s_3|$

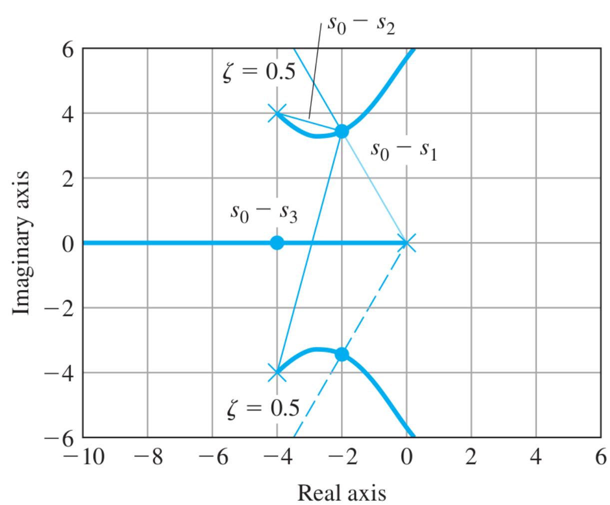


Figure 59: Determination of s_0 and its associated K value to match $\zeta = 0.5$.

- Note that this is a third-order system; the additional pole will increase the system's rise time and decrease its overshoot, since it makes the system more sluggish
  - When we select $\zeta = 0.5$, we are designing for the worst case of the overshoot

Lecture 19, Mar 18, 2024

Gain Selection from Root Locus

- Once we found a point on the root locus, s_0, that meets our requirements, we can find its value of K
- Since $L(s) = -\frac{1}{K}$ is the condition for the locus, $K = \frac{1}{|L(s)|} = \frac{|\prod_{i=1}^{n}(s_0 - p_i)|}{|\prod_{i=1}^{m}(s_0 - z_i)|} = \frac{\prod_{i=1}^{n}|s_0 - p_i|}{\prod_{i=1}^{m}|s_0 - z_i|}$
  - These magnitudes of the difference of s_0 from the poles and zeros can be obtained geometrically by measuring the distance of s from the roots and zeroes
- Once we have K, we can now solve for the values of s that make $L(s) = -\frac{1}{K}$ to find all the roots of the closed-loop system (since we only get one root initially)
- To identify $s_0 = -\sigma + j\omega$ given ζ:
  - We know $\frac{\omega}{\sigma} = \tan(\sin^{-1}\zeta)$
  - Substitute s_0 into $L(s) = -\frac{1}{K}$ and solve for the value of K
  - This will give us two equations, one for the real part (containing K), and another one for the imaginary part (which should equal 0)
  - Using the relation between σ and ω we can solve for their values using the imaginary equation
  - Substitute these values back into the real equation to solve for K

Example: 1-DoF Satellite Attitude Control

- Consider planar angular control of a satellite with a thruster generating a force F_c, and a disturbance M_D causing an unwanted moment
- $T_C + M_D = F_C d + M_D = I\ddot{\theta}$ where d is the distance from the centre of mass to the thruster and I is the satellite's moment of inertia
- Transfer function: assume $M_D = 0$, so $\frac{\Theta(s)}{T_C(s)} = G(s) = \frac{1}{Is^2} = \frac{A}{s^2}$
  - This a double-integrator
- Now consider an instrument attached to the satellite via a flexible boom, which can bend and vibrate
  - The total system has two degrees of freedom, the rotation of the satellite and the rotation of the instrument boom
  - The boom is modelled as a (rotational) spring-dashpot system between two discs
  - Bottom disc (attached to satellite): $T_C = I_1\ddot{\theta}_1 + b(\dot{\theta}_1 - \dot{\theta}_2) + k(\theta_1 - \theta_2)$
  - Top disc (attached to instrument): $0 = I_2\ddot{\theta}_2 + b(\dot{\theta}_1 - \dot{\theta}_1) + k(\theta_2 - \theta_1)$
  - We will simplify the system and assume $b = 0$
- Laplace transform:
  - $T_C = (I_1 s^2 + k)\Theta_1(s) - k\Theta_2(s)$
  - $0 = -k\Theta_1(s) + (I_2 s^2 + k)\Theta_2(s)$
- For this system, we can have two cases: either we want to control the attitude of the satellite, or the attitude of the instrument
  - $\frac{\Theta_1(s)}{T_C(s)} = \frac{I_2 s^2 + k}{I_1 I_2 s^2 \left(s^2 + \frac{k}{I_1} + \frac{k}{I_2}\right)}$
    * Here we are controlling the side attached to the satellite
    * This is the case of *collocated control*: both the actuator and the sensor dynamics are on one body
  - $\frac{\Theta_2(s)}{T_C(s)} = \frac{k}{I_1 I_2 s^2 \left(s^2 + \frac{k}{I_1} + \frac{k}{I_2}\right)}$

* Here we are controlling the instrument boom
    * This is the *non-collocated* case: the actuator and sensor are not on the body we want to control
  - Notice that the collocated case has 2 zeros, which the non-collocated case misses – we will later see that the zeros in the first case make the control a lot simpler
- Consider a proportional controller $D_c(s) = k_P$ to control only the satellite without the boom, $\frac{\Theta}{T_C} = \frac{A}{s^2}$
  - Closed loop TF: $\frac{k_P \frac{1}{s^2}}{1 + k_P \frac{1}{s^2}}$ with characteristic equation $1 + k_P \frac{1}{s^2} = 0$
  - This is already in root locus form; $L(s) = \frac{1}{s^2} \implies b(s) = 1, a(s) = s^2$
  - Root locus determination:
    1. Two branches, both starting at $s = 0$, both going to infinity since there are no open-loop zeros
    2. No segments on the real axis; since both open-loop poles are at $s = 0$, for $s < 0$ we are on the left of 2 poles, and for $s > 0$ we are on the left of none
    3. Two asymptotes, intersecting at $\alpha = 0$ and at angles $\pm 90°$
    4. Branches have departure angles from $s = 0$ of $\pm 90°$ (one goes up, one goes down)
  - Notice that now all poles are on the imaginary axis – no matter what we do, we get oscillations with no damping

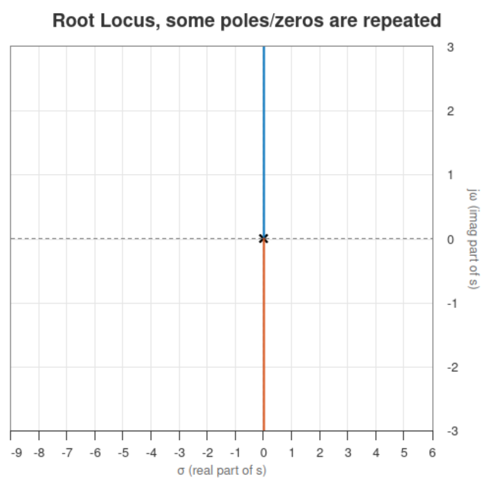


Figure 60: Root locus when using a proportional controller.

- Now consider using a PD controller $D_c(s) = k_P + k_D s$
  - Closed loop TF: $\frac{(k_P + k_D s)\frac{1}{s^2}}{1 + (k_P + k_D s)\frac{1}{s^2}}$ with characteristic equation $1 + (k_P + k_D s)\frac{1}{s^2} = 0$
  - Assume $k_D = K$ and $\frac{k_P}{k_D} = 1$, the characteristic equation is $1 + K\frac{s+1}{s^2} = 0$
    * The derivative gain introduced an open-loop zero to the system
  - Root locus:
    1. Two branches, both starting at $s = 0$, one of them going to the zero at $s = -1$, and the other going to ∞
    2. On the real axis, everywhere to the left of $s = -1$ is a part of the root locus, since that is to the left of 2 poles and 1 zero
    3. One asymptote along the negative real axis
    4. Departure angles from double pole at $s = 0$ are $\pm 90°$
    5. Two branches on the real axis meet at $\pm 90°$
    6. Break-in point at $s = -2$
  - Notice that the additional zeros has "pulled" the root locus to the left, adding damping and

allowing us to have a response that does not oscillate forever

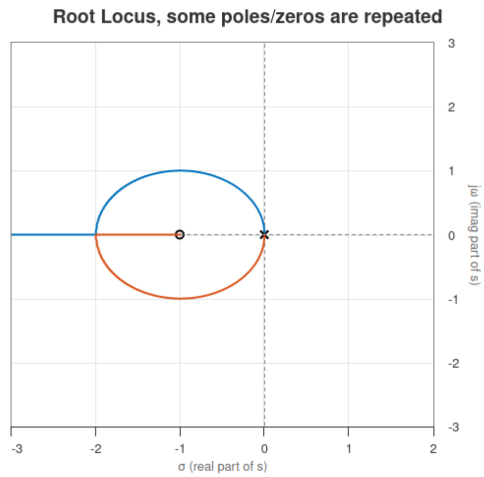


Figure 61: Root locus for the PD controller.

- However, in the real world any controller using a derivative gain is non-casual; implementing it in software will greatly amplify the noise in the system
  - To remedy this, we can try to add a denominator to the controller to make it casual
  - We add a factor in the denominator of $\frac{s}{p} + 1$
    * If we choose p to be large, this will have little effect on the system response, but we can make the system causal and practically workable
    * We make the order of the denominator as small as possible to reduce sluggishness
- PD controller with *lead compensator*: $D_c(s) = k_P + \frac{k_D s}{\frac{s}{p} + 1}$
  - $D_c(s) = k_P + \frac{pk_P s}{s+p} = \frac{(k_P + pk_D)s + k_P p}{s+p} = \frac{(k_P + pk_D)\left(s + \frac{k_P p}{k_P + pk_D}\right)}{s+p}$
  - Let $k_P + pk_D = K$ and $\frac{k_P p}{k_P + pk_D}$ so $D_c(s) = K\frac{s+z}{s+p}$
    * With the large p, the pole it introduces is very far in the negative real axis, so it has a very small effect on the overall system
  - Characteristic equation: $1 + D_c(s)G(s) = 1 + K\frac{s+z}{s^2(s+p)} = 0$
  - Consider the following cases of p and z:
    * $z = 1$ and $p = 12$:
      - Root locus determination:
        1. 3 branches, two starting at $s = 0$, one starting at $s = -12$, one branch ends at $s = -1$, two at infinity
        2. Real axis $-12 \le s \le -1$ is on the locus
        3. 2 asymptotes centered at $-\frac{11}{2}$ at angles $\pm 90°$
        4. Departure angles at $s = 0$ are $\pm 90°$, at $s = -12$ is $0°$
        5. Break-in point at angle of $\pm 90°$
        6. Break-in point at $s = -2.3$ for the two branches starting at $s = 0$; two other branches depart at $s = -5.2$
      - We see that the root locus is close to that of just a PD controller
    * $z = 1$ and $p = 4$:
      - Now the root locus branches are pushed to the right, causing oscillatory responses
      - The pole being much closer means that it now starts to matter
    * $z = 1$ and $p = 9$:

    - For this in-between value we see that the new pole does impact the root locus, but the impact is smaller
  * As the pole gets closer to the zero, the branches begin to merge together
  * The pole should always be placed as far away as possible from the zero, but this has tradeoffs

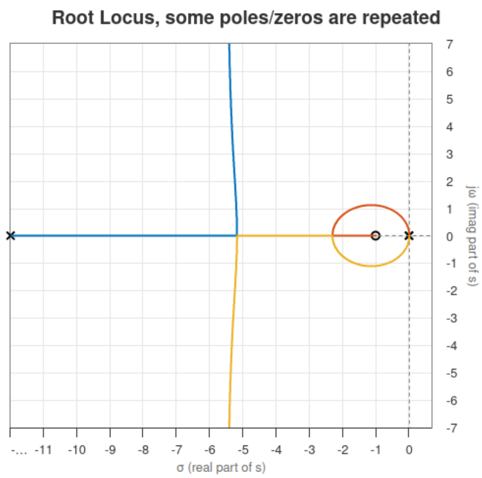


Figure 62: Root locus for the lead compensator, for $z = 1, p = 12$.

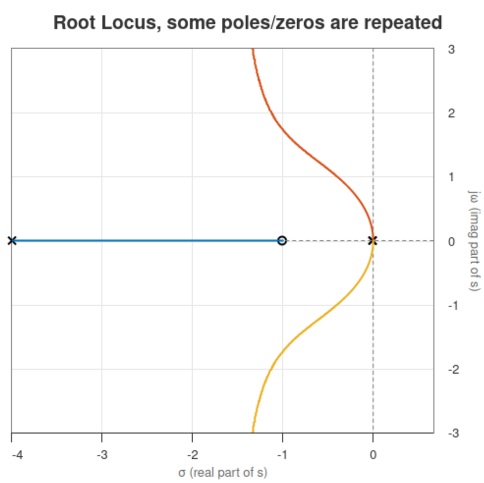


Figure 63: Root locus for the lead compensator, for $z = 1, p = 4$.

Lecture 20, Mar 21, 2024

Example: 1-DoF Satellite Attitude Control (Continued)

- Consider the case of collocated control (Θ_1), with the previous lead compensator at $z = 1, p = 12$
  - The characteristic equation is $1 + K\frac{s+1}{s+12}\frac{(s+0.1)^2+6^2}{s^2((s+0.1)^2+6.6^2)} = 0$
  - The flexible mode adds two additional branches, but since it also has two zeros, the two new branches go to the new zeros
  - Even though the 2 new poles are closer to the imaginary axis and have less damping, because they are very close to zeros, they are mostly cancelled out

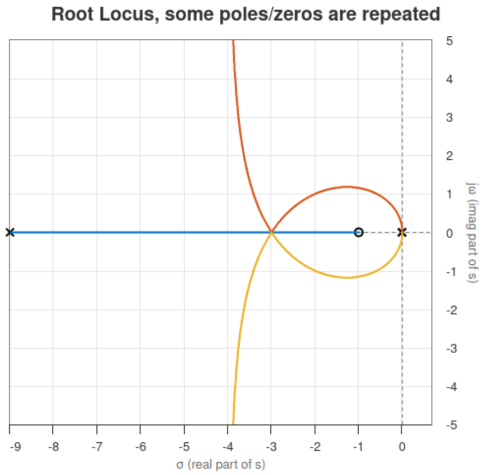


Figure 64: Root locus for the lead compensator, for $z = 1, p = 9$.

  - Therefore the response of the system is still mostly dominated by the same two poles as in the double-integrator case
    * The actual response will exhibit very small oscillations (added to the normal response) caused by the flexible modes
      - Since these are almost undamped, they will stay for a very long time
    * If the gain is very large, the dominating poles are now on the asymptote
  - Overall, the single flexible mode brings lightly damped roots
- Note that in the above we assumed that the open-loop zeroes are the same as the closed-loop zeros, which is only true when we have a unity feedback system

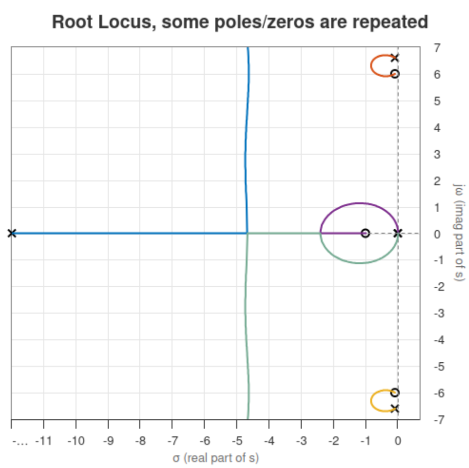


Figure 65: Root locus plot of the collocated case.

- In the non-collocated case (Θ_2), we are missing the two zeros
  - Because we don't have the zeros, the new branches now go to infinity instead of their zeros; the asymptotes make the poles go into the RHP, introducing instability
  - These poles are still barely in the LHP, so the system can still be stable for some gain values, but it is now unstable for larger gains
  - Furthermore these poles are no longer cancelled out by zeros, so they will dominate the system and introduce very high overshoot
  - This is why the non-collocated system is much harder to control

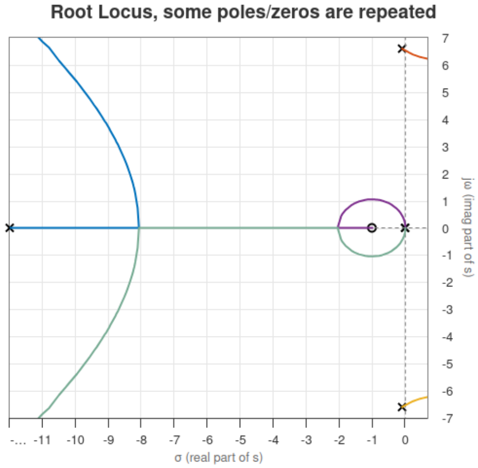


Figure 66: Root locus plot of the non-collocated case.

Design for Dynamic Compensation

- Lead compensator: $D_c(s) = K\dfrac{s+z}{s+p}$ where $z < p$
  - For a sinusoidal input, its output leads the input (output phase shift is positive)
  - Note that due to causality, the output doesn't start earlier than the input; but with a sustained sinusoidal input, the phase shift gradually approaches positive
  - This comes at a cost of some amplitude
  - Approximates PD control; speeds up response (lowering rise time) and decreases overshoot
- Lag compensator: $D_c(s) = K\dfrac{s+z}{s+p}$ where $z > p$
  - For a sinusoidal input, the output lags the input (negative phase shift)
  - The amplitude of the output is now larger than the input
  - Approximates PI control, decreasing steady-state error
- Notch compensator: $D_c(s) = K\dfrac{s^2 + 2\zeta\omega_0 s + \omega_0^2}{(s+\omega_0)^2}$
  - Attenuates the input around some unwanted frequency, acting as a band-stop filter
  - Enhances stability for plants with lightly damped flexible modes (cancels them out)
  - Typically has two complex zeros, which can capture problematic poles
    * Also has two real poles, but typically ω_0 is large, so they are far out in the LHP and usually has little effect
- Note that all 3 compensator do not have any poles at the origin, so the type of the plant is unchanged by adding a compensator
- Consider the example plant $G(s) = \dfrac{1}{s(s+1)}$, e.g. a servo mechanism
- Example: lead compensation
  - We typically start with the simplest possible controller first
  - Consider P control: $D_c(s) = K$
    * Since the asymptote is close to the imaginary axis, the damping is very low
    * If we want a certain ω_n for a certain rise time, we will have a large overshoot
    * e.g. for $\omega_n = 2 \implies \zeta = 0.25$
  - Now consider PD control: $D_c(s) = K(s+2)$
    * Now for the same value of ω_n, our poles will be on the circle, and ζ is significantly larger, improving damping without sacrificing speed
    * e.g. for $\omega_n = 2 \implies \zeta = 0.75$
  - Now the lead compensator $D_c(s) = K\dfrac{s+2}{s+p}$

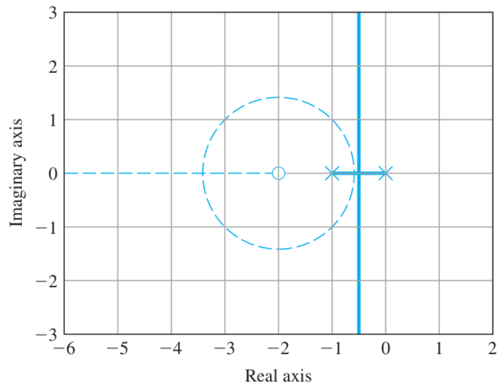


Figure 67: Root locus plots for a P (solid line) and PD (dashed line) controller.

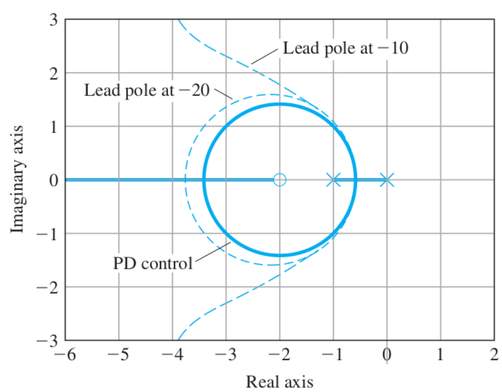


Figure 68: Root locus plots for different lead compensator gains.

      * As we've seen previously, depending on the location of the pole relative to the zero, we can get very different behaviour
      * For small K, the lead compensator approximates PD control well, regardless of where the pole is
      * For large p, the lead compensator also behaves like PD control
      * The additional pole slightly lowers damping (for the same ω_n we see that ζ is smaller)
        - This effect is negligible for low K and large p
      * Typically, we place the zero near the desired closed-loop ω_n (0.25 to 1 times ω_n) and the pole 5 to 25 times the value of the zero
        - The further p is, the closer we get to PD; we get slightly better performance, but noise will increase

Lecture 21, Mar 25, 2024

Design for Dynamic Compensation (Continued)

Lead Compensator

- Example: $G(s) = \dfrac{1}{s(s+1)}$; design a lead compensator for the position control system to provide an overshoot of no more than 20% and rise time of no more than 0.3 seconds
  - This gives us a required damping ratio of $\zeta \geq 0.5$ and $\omega_n \geq 6\,\text{rad/s}$; we will choose $\omega_n \geq 7$ for some margin
  - Initial trial with $D_c(s) = K\dfrac{s+2}{s+10}$
    * We start with a zero at $s = -2$, since this is in the range of 1/4 to 1 times the natural frequency we want
    * Start with a pole of 10, at 5 times the location of the zero (recall rule of thumb was 5 to 25 times)
    * By drawing out the circle corresponding to $\omega_n = 7$ and the angle for $\zeta = 0.5$, we find a small segment on the root locus that gives the desired response
    * Note that the additional pole on the real axis is very close to a closed-loop zero (which are the same as the open-loop zeros due to unity feedback), so its effects are small
    * However, when we plot the response for $K = 70$ we see an overshoot of 22%
      - From here, we can try to lower K, but this is not the best option
      - We can increase the pole slightly, so the response is closer to that of a PD controller
      - We could also try to increase the zero, but we chose to increase the pole first since we have more range on it
  - Second trial with $D_c(s) = K\dfrac{s+2}{s+13}$
    * Now with a gain of $K = 91$ we have a controller with rise time of 0.19 seconds and overshoot of 17%
- In general when designing a closed-loop system, we typically start with a lead compensator:
  1. Determine where the closed-loop roots need to be to meet the desired physical response characteristics
  2. Create a root locus with only a proportional controller
  3. If more damping is needed, choose z to be 1/4 to 1 times the desired ω_n and pick p to be 5 to 25 times z
  4. If less damping is needed, decrease p; if more damping is needed, increase p and/or decrease z
     - The ratio p/z should be as low as possible (less than 25) in order to minimize the effects of noise from a derivative controller
  5. When values of z and p are found so that the root locus passes through the desired region, select the value of K and check the step response
  6. Determine if the value of K meets the steady-state error requirements; if a value of K that meets

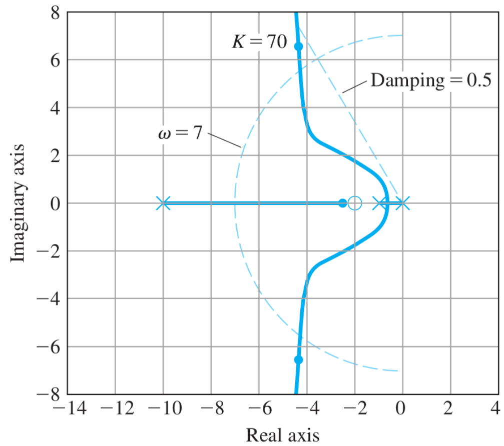


Figure 69: Root locus for $z = -2, p = -10$.

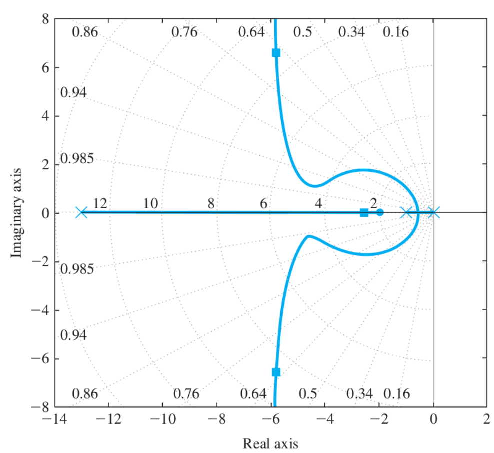


Figure 70: Root locus for $z = -2, p = -13$.

the requirements cannot be found, add integral control or a lag compensator

- The lead compensator will make the steady-state error worse (for the same value of K)
  - The position constant is $K_p = \lim_{s\to 0} K\frac{s+z}{s+p}G(s) = K\frac{z}{p}\lim_{s\to 0} G(s)$
  - Since $p > z$, overall this makes K_p smaller, making e_{ss} larger
  - In order to reduce the steady-state error again, we want to introduce another term $\frac{s+z_2}{s+p_2}$ where $z_2 > p_2$, so the position constant is increased
    * This is the idea behind the lag compensator

Lag Compensator

- Lag compensation has a similar effect as an integrator in decreasing the steady-state error at low frequencies, without affecting the transient response created by the lead compensator
  - The position/velocity/acceleration constant is increased by a factor equal to z/p per the above discussion
  - The ratio z/p is typically between 3 to 10; anything more than this could affect the transient response
  - We choose the value of p and z to be extremely small (100-200 times smaller than the closed-loop ω_n), so $\frac{s+z}{s+p} \approx 1$ for any nonzero s, therefore it won't affect the transient response
  - Note that we need to be mindful of the resolution of our controller; if z and p are too small, it may not be practically implementable
- Example: Increase K_v for the previous system to decrease the steady-state error, without changing its transient response
  - Lag compensator $D_{c2}(s) = \frac{s+z}{s+p}$ where $z > p$
  - Uncompensated $K_v = \lim_{s\to 0} sD_{c1}(s)G(s) = 14$ so $e_{ss} = \frac{1}{14}$
  - Suppose we want to increase K_v to 70, so we need $\frac{z}{p} = \frac{70}{14} = 5$
  - Choose $z = 0.05, p = 0.01$
  - On the root locus, this adds a very small circle near the origin; the overall root locus is almost unchanged

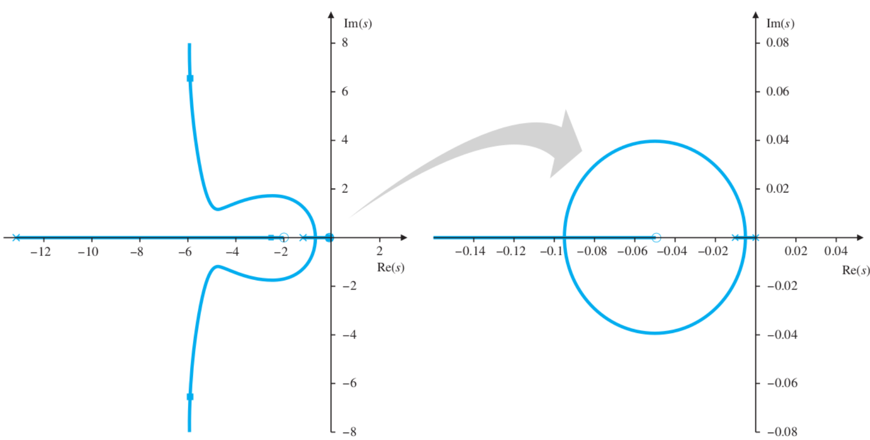


Figure 71: Root locus with lead and lag compensation.

- Lag compensator design process:
  1. Determine the amount of gain amplification we want to achieve the desired error constant, and determine the ratio z/p
  2. Select the value of z to be approximately 100 to 200 times smaller than the system's dominant natural frequency

3. Plot the resulting root locus and verify that it is still satisfactory and adjust z and p as necessary
4. Plot the step input to verify that the time domain response is still satisfactory
   - If the slow root of the lag compensator is too slow, increase z and p while keeping their ratio constant
   - Note that the closer z and p are to the dominant poles, the more effect they will have on the transient response

Notch Compensator

- A notch compensator is used to dampen the oscillation at some specific resonant frequency, e.g. due to a flexible mode in non-collocated control
  - The overall system response will have been handled by the other controllers; the notch compensator acts like a filter
  - Has form $\dfrac{s^2 + 2\zeta\omega_0 s + \omega_0^2}{(s + \omega_0)^2}$
  - The two real zeros cancel out the undesirable oscillatory poles in the system
  - The real poles are introduced so that the controller is causal and has a DC gain of 1, so the steady-state response is unaffected
  - Choose ω_0 to be very large as to not affect the transient response
- The position of the zero relative to the undesirable pole needs to be chosen to ensure that the resulting root locus is entirely in the LHP
  - Whether the zero should above or below the pole depends on the system
- Example: assume that the system has a flexible mode, so $G(s) = \dfrac{1}{s(s+1)} \cdot \dfrac{2500}{(s^2 + s + 2500)}$
  - The poles that were added are approximately $-0.5 \pm j50$; they are dominant and very lightly damped
  - Assume that we have the same lead-lag compensator from before
  - Add notch compensation $D_{c3}(s) = \dfrac{s^2 + 2\zeta\omega_0 s + \omega_0^2}{(s + \omega_0)^2} = \dfrac{s^2 + 0.8s + 3600}{(s + 60)^2}$
    * The zeros are at approximately $-0.4 \pm j60$
      - Notice that zero is close to the pole but not exactly on it
      - The imaginary part is above the undesirable pole so that the root locus is entirely in the LHP
      - Typically the zero is chosen to be a little bit closer to the imaginary axis than the undesirable pole
    * The new poles we introduced at $s = -60$ are very far so they do not have any effect
- In practice, a notch compensator will often increase the overshoot, so we may need to iterate on the design
- Note practically, we design lead first, then notch, and finally lag, because the notch compensator affects the design of the lag compensator

Example: Quadrotor Drone Control (Pitch Axis)

- $G(s) = \dfrac{1}{s^2(s+2)}$
  - The double integrator represents a delay
- From the root locus we can see that with a proportional controller, the system is unstable for any value of K, since we have two branches going into the RHP
  - Adding a lead compensator $D_c(s) = K\dfrac{s+1}{s+10}$ pushes the root locus to the left, making the system stable
- Consider non-collocated behaviour where there is flexibility between the actuators and the body, so we introduce a flexible mode
- $G(s) = \dfrac{1}{s^2(s+5)} \cdot \dfrac{225}{((s+0.1)^2 + 15^2)}$
- Goal: $t_r \leq 1\,\text{s}, M_p \leq 40\%, t_s \leq 10\,\text{s}, K_a \geq 12\,\text{rad}$ without high frequency oscillations in response

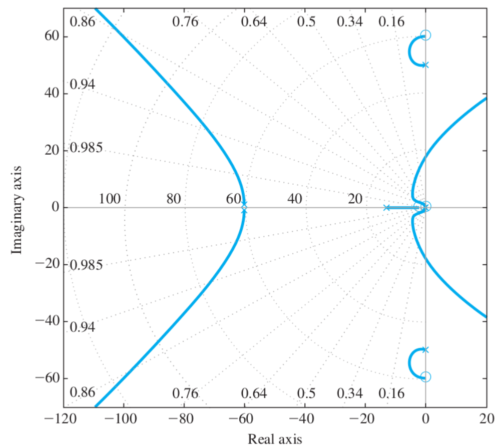


Figure 72: Root locus with flexible mode, lead, lag and notch compensation.

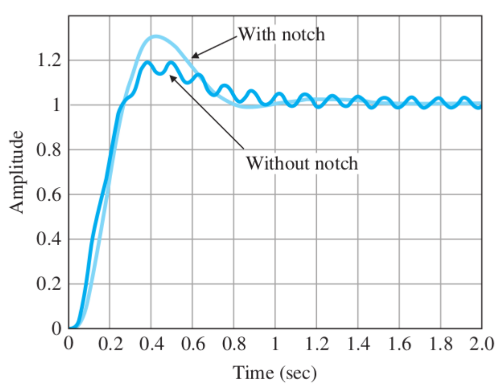


Figure 73: Step response with and without notch compensator.

- We can recognize that the system is type 2 due to the s^2 in the denominator, and adding compensators does not change the system type

- This translate to $\omega_n \geq 1.8\,\text{rad/s}, \zeta \geq 0.3, \sigma \geq 0.46$; we also need a lag and notch filter
- Proportional controller is again unstable with any gain
- Lead compensator: choose $z = 0.5$ (approximately $0.3\omega_n$), $p = 10$ (20 times the zero) so $D_{c1}(s) = \dfrac{s + 0.5}{s + 10}$
  - From the root locus we see that $K = 80$ is appropriate
  - Plotting the step response gives us a satisfactory overshoot and rise time
- Notch compensator: $D_{c3}(s) = \dfrac{(s + 0.05)^2 + 16^2}{(s + 16)^2}$
  - This cancels out the unwanted oscillations but slightly affects the transient response
  - Modify the lead compensator slightly to compensate
  - $K_a = \lim_{s \to 0} s^2 D_{c1}(s) D_{c3}(s) G(s) = 0.58$
- Lag compensator: need a ratio $\dfrac{z}{p} \geq \dfrac{12}{0.58} = 20.7$ so choose $D_{c2}(s) = \dfrac{s + 0.02}{s + 0.001}$
  - Modify the control gain as necessary

Lecture 22, Mar 28, 2024

Frequency Response Design Method

- For a (stable) LTI system $G(s)$, the steady-state response to an input $u(t) = A\sin(\omega_0 t)1(t)$ is given by $y_{ss}(t) = A|G(j\omega_0)|\sin(\omega_0 t + \angle G(j\omega_0))$
  - The response is a sinusoid of the same frequency, scaled by a factor of $|G(j\omega_0)|$ (the magnitude of the transfer function, known as the *gain* or *amplitude/magnitude ratio*), with a phase shift of $\angle G(j\omega_0)$ (the phase of the transfer function)
  - Knowing the magnitude $M(\omega)$ and phase $\phi(\omega)$ of $G(j\omega)$ for all possible frequencies ω fully specifies the transfer function
- In general, the complete response is the sum of a number of exponentials and a sinusoid; since the system is stable, all the exponentials decay to 0 as $t \to \infty$ and we are only left with the sinusoid
- Example: RC circuit, output $y(t)$ is the voltage across the capacitor, input $\mathcal{K}u(t)$ is an input voltage that is sinusoidal
  - $RC\dfrac{\mathrm{d}y}{\mathrm{d}t} + y(t) + \mathcal{K}u(t) \implies \dfrac{\mathrm{d}y}{\mathrm{d}t} + ky(t) = u(t)$ where $k = \dfrac{1}{RC}$, assuming $\mathcal{K} = RC$
  - $G(s) = \dfrac{1}{s + k}$
  - Given $u(t) = \sin(10t)1(t)$, $U(s) = \dfrac{10}{s^2 + 100}$
  - At $s = j10$, $|G(j10)| = \dfrac{1}{\sqrt{1^2 + 10^2}}$ and $\angle G(j10) = -\tan^{-1}\left(\dfrac{10}{1}\right)$
  - Therefore the response is $y(t) = \dfrac{1}{\sqrt{101}}\sin(10t - \tan^{-1}(10))$
- Example: lead network $D_c(s) = K\dfrac{T_s + 1}{\alpha T_s + 1}$ for $\alpha < 1$
  - Note that this is mathematically identical to the form of the lead compensator we had before, but this form is more common and convenient for frequency response design
    - The zero is at $\dfrac{1}{T}$, the pole at $\dfrac{1}{\alpha T}$ and the gain is $\dfrac{K}{\alpha}$
  - Frequency response: $D_c(j\omega) = K\dfrac{Tj\omega + 1}{\alpha Tj\omega + 1}$
  - Gain: $M = |K|\dfrac{\sqrt{1 + \omega^2 T^2}}{\sqrt{1 + \alpha^2\omega^2 T^2}}$
  - Phase: $\phi = \tan^{-1}(\omega T) - \tan^{-1}(\alpha\omega T)$
  - For $\omega \to 0$, we have $M \to |K|$ and $\phi \to 0$

– For $\omega \to \infty$ we have $M \to \left|\frac{K}{\alpha}\right|$ and $\phi \to 0$

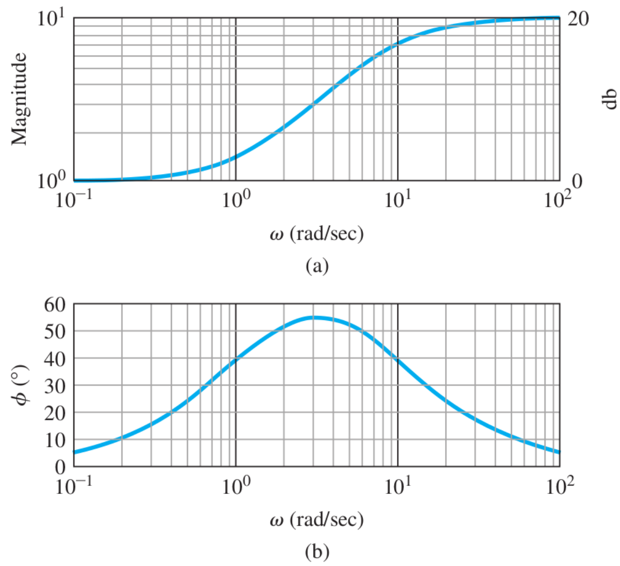


Figure 74: Bode magnitude and phase plots for the lead compensator, for $K = 1, \alpha = 0.1, T = 1$.

- The gain and phase for a range of values of ω can be summarized in a *Bode plot*
  – The top plot is the magnitude plot; the bottom plot is the phase plot
  – The bode plot is log-log for magnitude and semi-log for phase
    * Using a log-log plot for gain allows us to cover a wide range of ω and gain, and also allows us to simply add up the magnitude plots of transfer functions to get the final plot, since multiplication of gains is just addition of logs
  – The vertical axis of the magnitude plot often uses decibels, $\text{dB} = 20 \log|G(j\omega)|$
- Note in MATLAB, `bode(sys, w)` gives `[mag, phase]`, which we can plot to get the Bode plot
  – Use `logspace()` to get the points for `w`

System Behaviour From Frequency Response

- The gain and phase of the system's frequency response completely determines the behaviour of the system; we design using it just like we design using the root locus
  – The root locus is to the root locus design method as the Bode plot is to the frequency design method
- Typical closed-loop systems exhibit a low-pass filter behaviour
  – The gain is close to 1 at lower frequencies, i.e. the output follows the input well
  – Beyond a certain frequency, the gain deviates from 1; for most systems, it increases first before decreasing
  – For most systems when the frequency gets very large the gain approaches 0, i.e. the output stops following the input at all
- The *bandwidth* ω_{BW} is defined as the highest frequency ω where the output still tracks the (sinusoidal) input in a satisfactory manner
  – Traditionally we define this to be when the gain hits $\sqrt{2}/2 = 0.707$
  – This is known as the *half-power point*; if the gain is a voltage gain, then at this point, the power of the response will be only half
  – A higher bandwidth means a faster response – the larger ω_{BW} is, the larger ω_n is and the shorter our rise and peak times
- The *resonant peak* M_r is the maximum value of the amplitude ratio
  – M_r has a direct relationship with ζ, so we can estimate the damping and overshoot of the system

from M_r

- When we design a controller, we examine its bode plot and tune the gains to get the desired bandwidth and resonant peak, just like we identify pole locations on a root locus

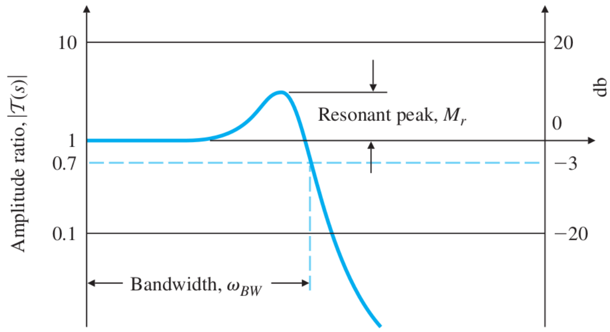


Figure 75: Definitions of bandwidth and resonant peak.

Lecture 23, Apr 1, 2024

Plotting Bode Plots

- Consider a general transfer function $G(s) = K\dfrac{(s+z_1)\dots(s^2+2\zeta_1\omega_{n1}s+\omega_{n1}^2)\dots}{(s+p_1)\dots(s^2+2\zeta_a\omega_{na}s+\omega_{na}^2)}$
  - z_i and p_i are real; the complex poles and zeros are in the quadratic factors, represented by their natural frequencies and damping ratios
- Rearrange as $G(s) = K_0 s^n \dfrac{(\tau_1 s+1)(\tau_2 s+1)\dots\left(\left(\frac{s}{\omega_{n1}}\right)^2+2\zeta_1\left(\frac{s}{\omega_{n1}}\right)+1\right)\dots}{(\tau_a s+1)(\tau_b s+1)\dots\left(\left(\frac{s}{\omega_{na}}\right)^2+2\zeta_a\left(\frac{s}{\omega_{na}}\right)+1\right)\dots}$
  - We factor out the poles and zeros at the origin to s^n, where n could be positive or negative
  - $\tau_1, \tau_2, \dots$ correspond to the real zeros, $\tau_a, \tau_b, \dots$ correspond to the real poles
  - $\omega_{n1}, \omega_{n2}, \dots$ and ζ_1, ζ_2 correspond to the complex zeros; $\omega_{na}, \omega_{nb}, \dots$ and ζ_a, ζ_b correspond to the complex poles
  - Substitute $s = j\omega$: $G(s) = K_0(j\omega)^n \dfrac{(j\omega\tau_1+1)(j\omega\tau_2+1)\dots\left(\left(\frac{j\omega}{\omega_{n1}}\right)^2+2\zeta_1\left(\frac{j\omega}{\omega_{n1}}\right)+1\right)\dots}{(j\omega\tau_a+1)(j\omega\tau_b+1)\dots\left(\left(\frac{j\omega}{\omega_{na}}\right)^2+2\zeta_a\left(\frac{j\omega}{\omega_{na}}\right)+1\right)\dots}$
  - This is the *Bode form* of the transfer function
- The Bode form is a composite of simpler transfer functions of the 3 classes:
  1. $K_0(j\omega)^n$ where $n \in \mathbb{Z}$
  2. $(j\omega\tau+1)^{\pm 1}$ (if the power is 1, then it is numerator class 2, while power -1 is denominator class 2)
  3. $\left(\left(\dfrac{j\omega}{\omega_n}\right)^2+2\zeta\left(\dfrac{j\omega}{\omega_n}\right)+1\right)^{\pm 1}$ (power 1 → numerator class 3; -1 → denominator class 3)
- To find the bode plot of a composite transfer function, we plot the Bode plots of each of the individual classes, and sum them up, since multiplication is addition of logs
- Class 1: $K_0(j\omega)^n$
  - Magnitude: $\log K_0|(j\omega)^n| = \log K_0 + n\log\omega$
    * The magnitude plot is a straight line with slope n (or n times 20 decibels per decade)
      - For $\omega = 1$, the value of the gain is $\log K_0$
    * For very low values of ω we will see that this is the only class that affects the slope of the Bode plot
  - Phase: $\angle K_0(j\omega)^n = n \cdot 90°$

* The phase plot is a constant value, determined by n

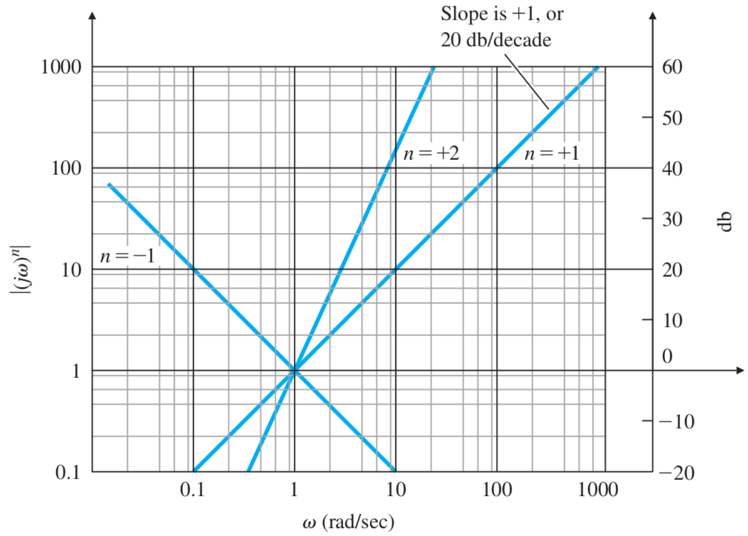


Figure 76: Magnitude plot of $(j\omega)^n$.

- Class 2: $(j\omega\tau + 1)^{\pm 1}$
  - For $\omega\tau \ll 1$, $(j\omega\tau + 1)^{\pm 1} \approx 1$
  - For $\omega\tau \gg 1$, $(j\omega\tau + 1)^{\pm 1} \approx (j\omega\tau)^{\pm 1}$
  - The *break point* is defined as $\omega = \dfrac{1}{\tau}$
  - Magnitude:
    * Below the break point, the gain is approximately a constant 1
    * Above the break point, the gain behaves like a class 1 term of $\tau^{\pm 1}(j\omega)^{\pm 1}$
      - The slope is a constant 1 or -1 (or ±20 decibels per decade) for this asymptote
      - The intercept is at $\tau^{\pm 1}$
    * At the break point, the gain is a factor of 1.41 (or 3 decibels) above for numerator class 2, or 0.707 (or -3 decibels) below for denominator class 2
      - At the break point, $|j\omega\tau + 1|^{\pm 1} = |j + 1|^{\pm 1} = \sqrt{2}^{\pm 1}$
  - Phase:
    * Below the break point, the phase is $\angle 1 = 0°$
    * Above the break point, the phase is $\angle (j\omega\tau)^{\pm 1} = \pm 90°$
    * At the break point, the phase is $\angle (j + 1)^{\pm 1} = \pm 45°$
    * The middle asymptote intersects the lower and upper asymptotes at 5 times above and below the break point
    * At the intersection of asymptotes, the actual phase deviates from the asymptotes by about $\angle (j/5 + 1)^{\pm} 1 = \pm 11°$
  - For very low frequencies, class 2 gives a gain of 1 and phase of 0, so it has no effect on the Bode plot of the composite function
    * Rule of thumb is to ignore for ω a factor of 10 or more below the break point
- Class 3: $\left(\left(\dfrac{j\omega}{\omega_n}\right)^2 + 2\zeta\left(\dfrac{j\omega}{\omega_n}\right) + 1\right)^{\pm 1}$
  - The break point is $\omega = \omega_n$
  - For $\omega \ll \omega_n$, $\left(\left(\dfrac{j\omega}{\omega_n}\right)^2 + 2\zeta\left(\dfrac{j\omega}{\omega_n}\right) + 1\right)^{\pm 1} \approx 1$

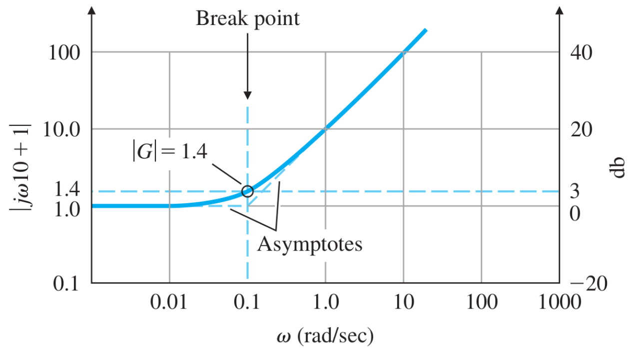


Figure 77: Magnitude plot of $(j\omega\tau + 1)$ for $\tau = 10$.

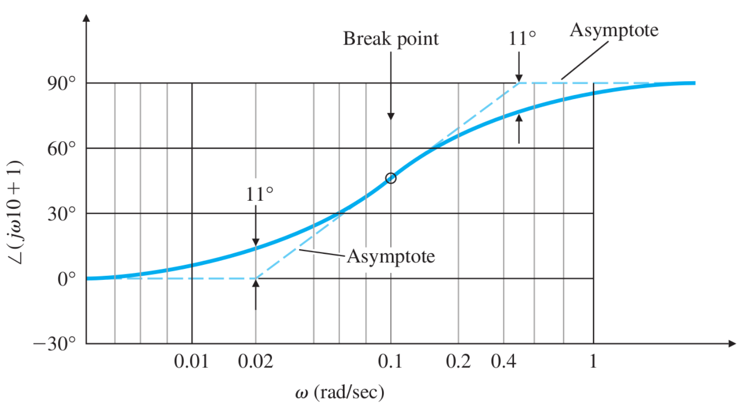


Figure 78: Phase plot of $(j\omega\tau + 1)$ for $\tau = 10$.

- For $\omega \gg \omega_n$, $\left(\left(\frac{j\omega}{\omega_n}\right)^2 + 2\zeta\left(\frac{j\omega}{\omega_n}\right) + 1\right)^{\pm 1} \approx \left(\frac{j\omega}{\omega_n}\right)^{\pm 2}$
- Magnitude:
  * For $\omega \ll \omega_n$ the gain is again approximately 1
  * For $\omega \gg \omega_n$ the gain behaves like a class 1 term of $\frac{1}{\omega_n^{\pm 2}}(j\omega)^{\pm 2}$
    - The slope is a constant ± 2 (or ± 40 decibels per decade)
  * The transition between the two asymptotes depends on ζ
    - At the break point, the magnitude is a factor of $(2\zeta)^{\pm 1}$ above/below a gain of 1
      - For $\omega = \omega_n$, $(j^2 + 2\zeta + 1)^{\pm 1} = (2\zeta)^{\pm 1}$
    - For a power of $+1$ the magnitude goes down at the break point, while for -1 the magnitude goes up
  * The peak has a magnitude of $\frac{1}{2\zeta\sqrt{1-\zeta^2}}$ and occurs at $\omega_r = \omega_n\sqrt{1-2\zeta^2}$
    - This can be obtained by differentiating the expression for the magnitude
    - For values of $\zeta > \frac{1}{\sqrt{2}}$, the resonant peak does not exist
    - The smaller ζ is, the closer the peak is to ω_n and the larger the magnitude of the peak
- Phase:
  * For $\omega \ll \omega_n$, $\angle 1 = 0°$
  * For $\omega \gg \omega_n$, $\angle (j\omega)^{\pm 2} = \pm 180°$
  * For $\omega \approx \omega_n$, $\angle(\pm j2\zeta) = \pm 90°$
  * The smaller the ζ, the faster the phase will transition between $0°$ and $\pm 180°$
    - For $\zeta = 0$, the transition is essentially a step function and the change is an instantaneous jump
    - For $\zeta = 1$, we just have a multiplication of two class 2 terms with the same break point

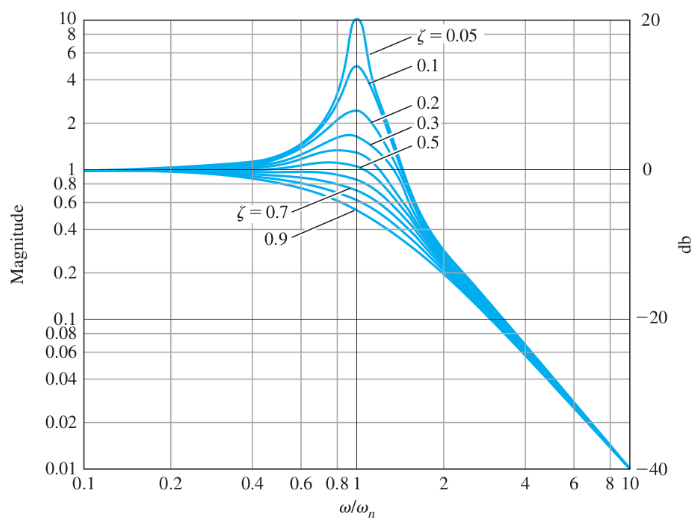


Figure 79: Magnitude plot of $\left(\left(\frac{j\omega}{\omega_n}\right)^2 + 2\zeta\left(\frac{j\omega}{\omega_n}\right) + 1\right)^{-1}$.

- Process for plotting a composite Bode plot:
  1. Manipulate the transfer function into Bode form to identify all break point frequencies
  2. Plot the low-frequency asymptote: Determine the value of n for the class 1 term and plot its magnitude as a line with slope of n passing through K_0 at $\omega = 1$
  3. Draw the asymptotes for the magnitude plot: Extend the low-frequency asymptote until the next break point, then change the slope by ± 1 or ± 2 depending on the class of the break point and

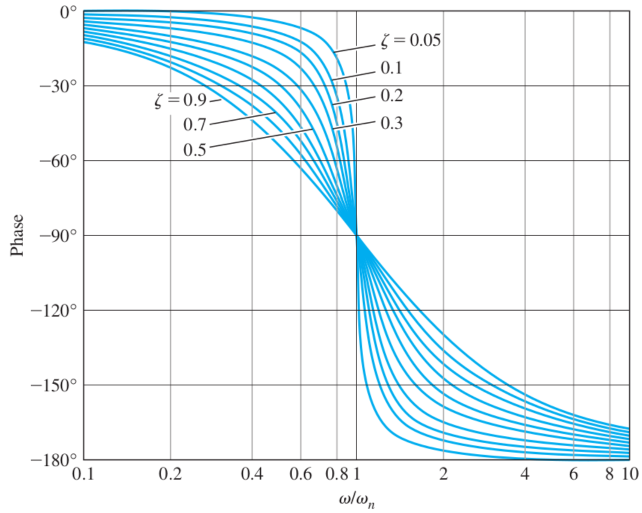


Figure 80: Phase plot of $\left(\left(\frac{j\omega}{\omega_n}\right)^2 + 2\zeta\left(\frac{j\omega}{\omega_n}\right) + 1\right)^{-1}$.

whether it is numerator or denominator; repeat until all break points are accounted for

4. Correct the magnitude values at break points:
   - For class 2, increase the magnitude by a factor of 1.41 (numerator) or decrease by a factor of 0.707 (denominator)
   - For class 3, change by a factor of (2ζ) (numerator) or a factor of $\frac{1}{2\zeta}$ (denominator)
   - Note these values may change of break points are close together; if break points are less than a factor of 10 away, the break point offsets are inaccurate
5. Plot the low-frequency asymptote of the phase curve: $\phi = n \cdot 90°$
6. Draw the horizontal asymptotes for phase: Change the value of the phase asymptote by $\pm 90°$ for class 2 break points and $\pm 180°$ for class 3 break points for each break point in ascending order
7. Determine intermediate asymptotes for each break point
8. Add each phase curve together graphically

- Example: $G(s) = \frac{2000(s+0.5)}{s(s+10)(s+50)}$
  - $G(s) = 2s^{-1}\frac{\left(\frac{s}{0.5}+1\right)}{\left(\frac{s}{10}+1\right)\left(\frac{s}{50}+1\right)}$
  - Class 1: $2(j\omega)^{-1}$
  - Class 2: $\left(\frac{j\omega}{0.5}+1\right)$ with break point 0.5, $\left(\frac{j\omega}{10}+1\right)^{-1}$ with break point 10, $\left(\frac{j\omega}{50}+1\right)^{-1}$ with break point 50
  - Steps:
    1. Bode form: $2(j\omega)^{-1}\frac{\left(\frac{j\omega}{0.5}+1\right)}{\left(\frac{j\omega}{10}+1\right)\left(\frac{j\omega}{50}+1\right)}$
    2. From the class 1 term: At $\omega = 1$, the gain is 2; the slope is -1
    3. Continue the slope of -1 until the first break point 0.5, then increase slope by 1 (to 0); next break point is at 10, decrease slope by 1 (to -1); next break point is at 50, decrease slope by 1 (to -2)
    4. Increase magnitude by a factor of 1.41 at break point 0.5; decrease by a factor of 0.707 at break point 10; decrease by a factor of 0.707 at break point 50
    5. Low-frequency phase asymptote: $\phi = -90°$
    6. Increase phase by 90° at $\omega = 0.5$ (to 0°), decrease by 90° at 10 (to $-90°$), decrease by another

90° at 50 (to −180°)

7. Draw the phase curves for the individual terms
8. Graphically add the individual phase curves to obtain the final phase plot

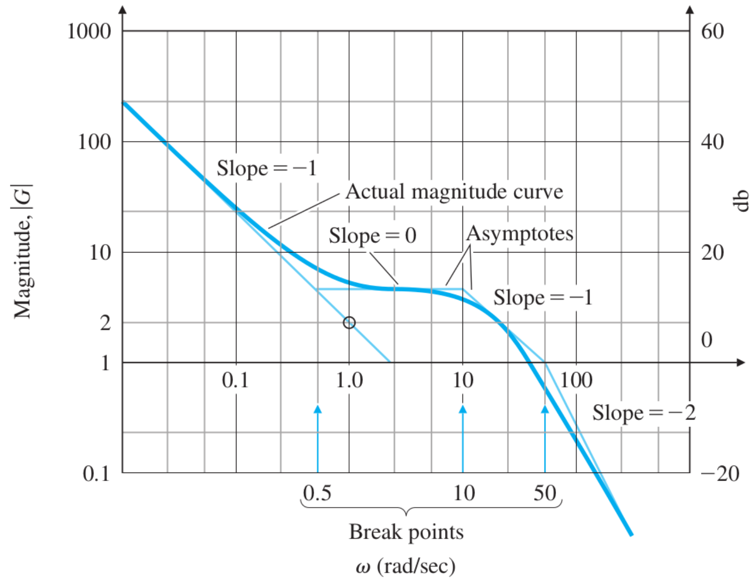


Figure 81: Magnitude plot of $G(s) = \dfrac{2000(s+0.5)}{s(s+10)(s+50)}$.

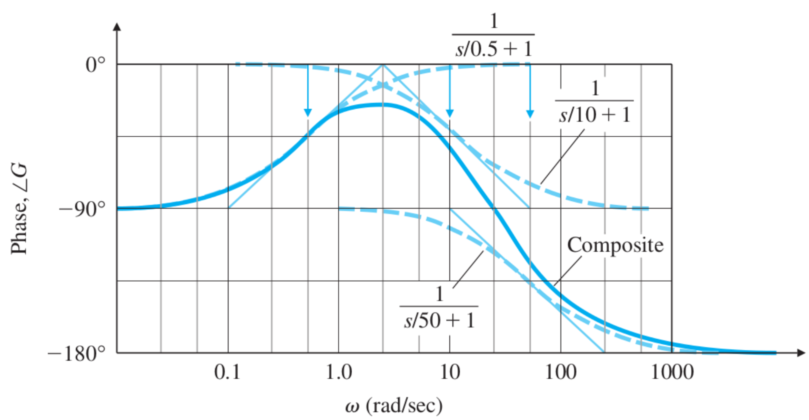


Figure 82: Phase plot of $G(s) = \dfrac{2000(s+0.5)}{s(s+10)(s+50)}$.

- Example: $G(s) = \dfrac{10}{s(s^2 + 0.4s + 4)}$
  1. Bode form: $G(j\omega) = 2.5(j\omega)^{-1} \dfrac{1}{\left(\left(\frac{j\omega}{2}\right)^2 + 2(0.1)\left(\frac{j\omega}{2}\right) + 1\right)}$
  2. Class 1 term: $2.5(j\omega)^{-1}$
     - First asymptote with slope of -1 having a value of 2.5 at $\omega = 1$
  3. Class 3 term: $\omega_n = 2$ and $\zeta = 0.1$, denominator
     - Decrease the asymptote slope by 2 at $\omega = 2$
  4. Increase magnitude by a factor of $\dfrac{1}{2(0.1)} = 5$ at the break point, and plot the magnitude
  5. Low-frequency asymptote at $\phi = -90°$

6. Decrease phase by $-180°$ at $\omega = 2$ (to $-270°$)
7. Draw the phase plot

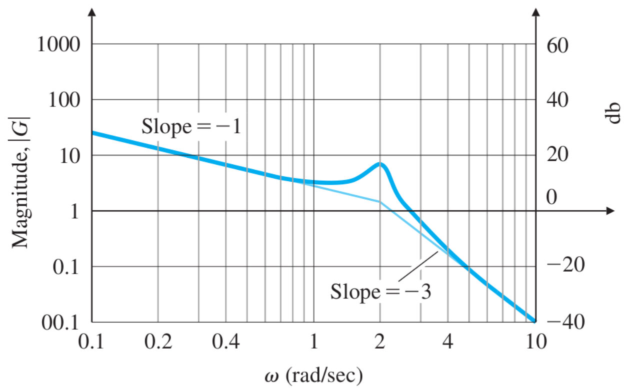


Figure 83: Magnitude plot of $G(s) = \dfrac{10}{s(s^2 + 0.4s + 4)}$.

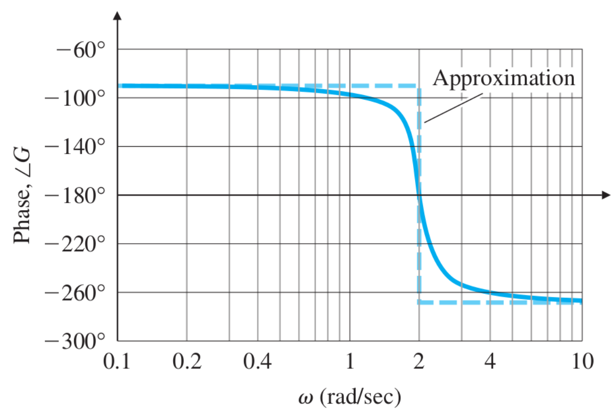


Figure 84: Phase plot of $G(s) = \dfrac{10}{s(s^2 + 0.4s + 4)}$.

Lecture 24, Apr 4, 2024

System Response from Frequency Response

- Consider a unity feedback system with open-loop transfer function $L(s) = KG(s)$
- A typical root locus starts with all poles on the left hand side, and as K increases, the locus crosses the imaginary axis at some point and the system becomes unstable
- The Bode plot of $KG(j\omega_c)$ has a magnitude plot that is simply shifted vertically, and a phase plot that is identical as $G(j\omega_c)$
  - Multiplying by K increases the magnitude by a constant factor at all frequencies and has a phase of 0
- The conditions for marginal/neutral stability are $|KG(j\omega_c)| = 1$ and $\angle G(j\omega_c) = -180°$
  - These are the same conditions as having the closed-loop poles being on the imaginary axis for a root locus
  - We can look at the phase plot to see the ω_c that gives a phase of $-180°$, and then look at the value of K that gives magnitude 1 at ω_c
- For most systems, decreasing K from the neural stability value will make the system stable, while increasing it will make the system unstable

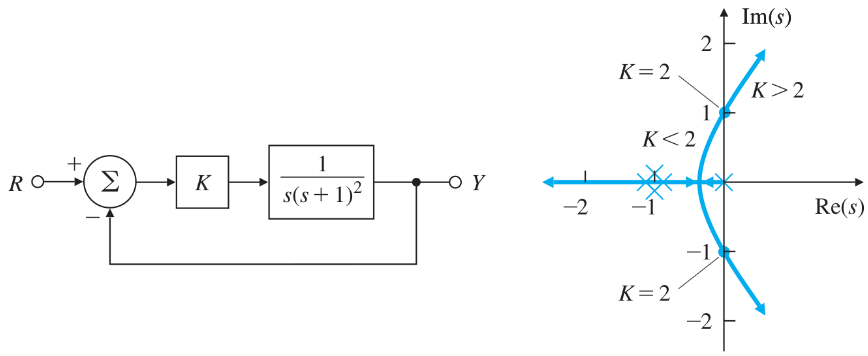


Figure 85: Typical closed-loop system and root locus.

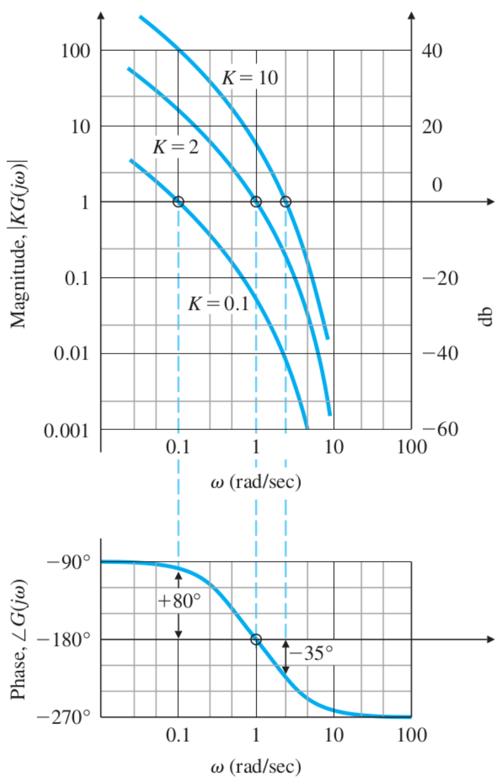


Figure 86: Open-loop Bode plot for the example system.

- Therefore if $|KG(j\omega) < 1|$ at $\angle G(j\omega) = -180°$ then the system is stable; otherwise it is unstable
- Note this does not apply if the open loop Bode plot crosses $|KG(j\omega)| = 1$ more than once
  * For such systems we need to use techniques to shift the plot so it crosses unity only once

- The degree of stability is how far we are from the value of K that gives marginal stability; we measure this through two quantities:
  - *Gain margin* (GM): the factor by which K can be increased before the system becomes unstable
    * On a Bode plot, this is how much we can move the magnitude plot up before we reach $|KG(j\omega)| = 1$
    * This is the value of $\dfrac{1}{|KG(j\omega)|}$ where $\angle G(j\omega) = -180°$
      - On a decibel scale this is the vertical distance between the value of the magnitude plot and the 0 decibel line
    * On a root locus, this is the ratio of the K value that puts the closed-loop poles on the imaginary axis and the K value that gives the poles given
    * $GM < 1$ (or negative in decibels) indicates an unstable system
  - *Phase margin* (PM): the amount by which the phase $G(j\omega)$ exceeds $-180°$ (less negative) when $|KG(j\omega)| = 1$
    * On a Bode plot, find the value of ω that gives a magnitude of 1, and the phase margin is the value of the phase at this point minus $-180°$
    * $PM < 0$ indicates an unstable system
      - A value of $PM = 30°$ is typically regarded as the lowest value for a safe stability margin
      - In design we try to go for an ideal value of $PM = 90°$ but usually we have to compromise
    * The PM for any value of K can be obtained directly from the Bode plot for $G(j\omega)$ (i.e. $K = 1$), by finding the ω that gives $|G(j\omega)| = 1/K$ and taking the phase at this frequency, subtracting $-180°$
      - This is because $|G(j\omega)| = 1/K \implies |KG(j\omega)| = 1$
      - We can also go backwards; for a value of PM, note the required ω, find the value of $|G(j\omega)|$ and take $K = 1/|G(j\omega)|$
- The (gain) *crossover frequency* ω_c is the frequency at which the open-loop magnitude is unity
  - This is highly correlated with the closed-loop bandwidth and hence the system response speed
  - $PM = \angle L(j\omega_c) - (-180°)$
- PM is more commonly used than GM in practice:
  - For a typical second order system $GM = \infty$ since phase reaches $-180°$ only at $\omega \to \infty$, at which point $|G(j\omega)| \to 0$
  - PM is also closely related to the system damping ratio
- Consider $G(s) = \dfrac{\omega_n^2}{s(s + 2\zeta\omega_n)} \implies \mathcal{T}(s) = \dfrac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$, a typical closed-loop system
  - We can derive $PM = \tan^{-1}\left(\dfrac{2\zeta}{\sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}}\right)$
    * $G(j\omega) = \dfrac{\omega_n^2}{(j\omega)(j\omega + 2\zeta\omega_n)}$
    * $|G(j\omega_c)| = 1 \implies \dfrac{\omega_n^2}{\omega_c\sqrt{\omega_c^2 + 4\zeta^2\omega_n^2}} = 1 \implies \omega_c^2 = -2\zeta^2\omega_n^2 \pm \sqrt{4\zeta^4\omega_n^4 + \omega_n^4} \implies \omega_c = \omega_n\sqrt{\sqrt{1 + 4\zeta^4} - 2\zeta^2}$

* $PM = \angle G(j\omega_c) - (-180°)$

$$= \angle\omega_n^2 - \angle(j\omega_c) - \angle(j\omega_c + 2\zeta\omega_n) + 180°$$

$$= 0 - 90° - \tan^{-1}\left(\frac{\omega_c}{2\zeta\omega_n}\right) + 180°$$

$$= 90° - \tan^{-1}\left(\frac{\omega_c}{2\zeta\omega_n}\right)$$

$$= 90° - \tan^{-1}\left(\frac{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}{2\zeta}\right)$$

$$= \tan^{-1}\left(\frac{2\zeta}{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}\right)$$

- For $PM < 65°$, we can use a linear approximation $\zeta \approx \frac{PM°}{100}$
  * This is used as a rule of thumb for other systems as well
- The resonant peak M_r and overshoot M_p can be obtained from PM as well since both are related to ζ
  * This can also serve as a rough estimate for systems other than the second-order closed-loop system we have

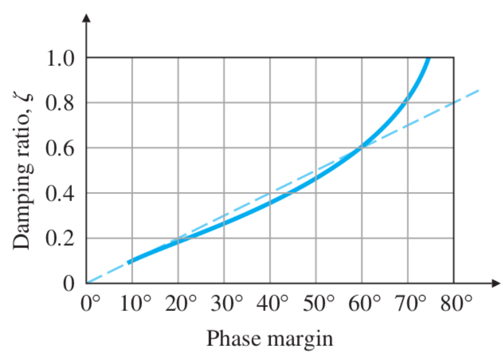


Figure 87: Relationship between ζ and PM.

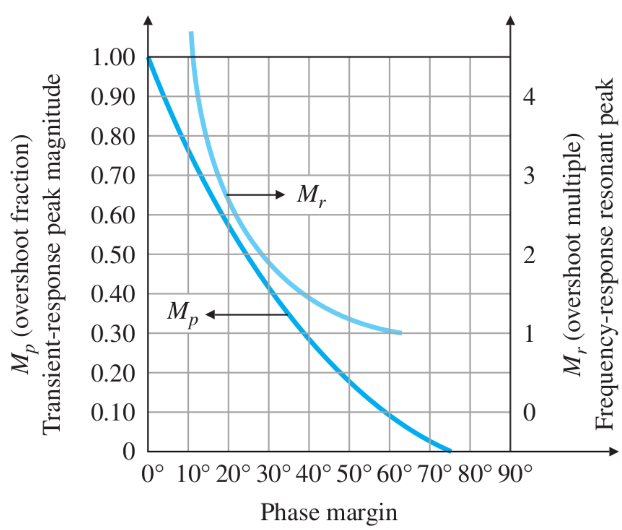


Figure 88: Relationship between M_p and M_r and PM.

- For any stable minimum phase system (i.e. no poles or zeros in the RHP), the phase of $G(j\omega)$ is uniquely related to the magnitude of $G(j\omega)$
  - $\angle G(j\omega_0) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{d}M}{\mathrm{d}u} W(u)\,\mathrm{d}u$ where $\begin{cases} M = \log|G(j\omega)| \\ u = \log(\omega/\omega_0) \\ W \approx \frac{\pi^2}{2}\delta(u) \end{cases}$
    * The phase is related to the slope of the magnitude plot on a log-log scale, near the frequency ω_0 we want to study
    * $\delta(u)$ is a weighting function (plot shown below)
      - This applies a much higher weight to values near $u = 0$
      - Even though the integral goes to infinity on both sides, the weighting makes it insignificant
  - If the slope of the gain is nearly constant around ω_0, we can take out $\frac{\mathrm{d}M}{\mathrm{d}u}$
  - $\angle G(j\omega_0) \approx \frac{\pi}{2}\frac{\mathrm{d}M}{\mathrm{d}u} = n \cdot 90°$ if $\frac{\mathrm{d}M}{\mathrm{d}u}$ is constant for a decade around ω_0

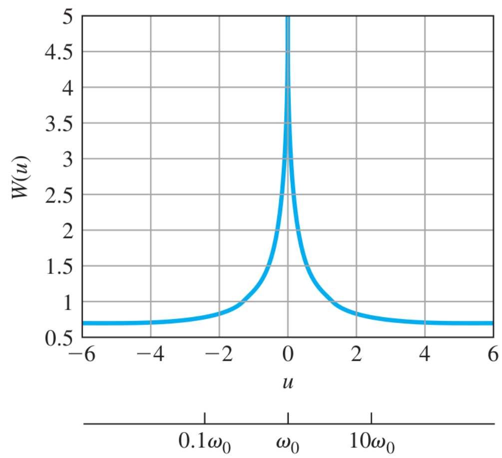


Figure 89: Plot of the weighting function.

- This means that if we can manage $|KG(j\omega)|$ to have a constant slope of -1 for a decade around the crossover frequency ω_c (i.e. where $|KG(j\omega)| = 1$), we will get a phase of $-90°$ at ω_c, which gives a PM of 90°, guaranteeing good stability of the system and a high ζ to reduce overshoot
  - This is the rule of thumb for design
  - We can adjust the value of K to shift the plot so that the slope is -1 at unity gain, or we can add compensators to change the slope for the same value of K

Example: Spacecraft Attitude Control

- Find a suitable $KD_c(s)$ to provide $M_p < 15\%$ and a bandwidth of $0.2\,\mathrm{rad/s}$ for the plant $G(s) = \frac{1}{s^2}$ and determine the frequency where the sensitivity function $|\mathcal{S}| = 0.7$
  - $\frac{1}{s^2}$ is class 1, so the phase plot is a constant $-180°$, and the system is always unstable; the slope of the magnitude plot is -2 which is also not good
- We want to increase the slope, so we want to add a numerator class 2 term
  - Use a PD controller: $KD_c(s) = K(T_D s + 1)$
- Start with the bandwidth of $0.2\,\mathrm{rad/s}$ which gives us a hint for ω_c; we choose $\omega_c = 0.2$
- The break point of the controller is $\frac{1}{T_D}$
  - We need to put this break point sufficiently before ω_c, so we have a sufficiently constant slope around ω_c
  - Choose the break point to be 1/4 of ω_c, so have $\omega_1 = 0.05$ and $T_D = 20$

- Plot $|D_cG(j\omega)|$ for $K = 1$, and notice the magnitude at 0.2 – in this case we have 100
- Therefore we choose $K = \dfrac{1}{|D_cG(j\omega_c|} = 0.01$
- Validate our assumption that the bandwidth is around 0.2:
  - $|\mathcal{T}(j\omega)| = \dfrac{|KD_cG|}{|1 + KD_cG|}$
  - From the plot we can see that the bandwidth is around 0.25 (when magnitude reaches around 0.7), which is close to ω_c
- For a unity feedback system, $\mathcal{S}(s) = \dfrac{E(s)}{\Theta(s)}$ (in general $1 - \mathcal{T}(s)$)
  - We want the sensitivity function to be low at the frequencies we work with, so the system is insensitive to an error in the reference
- The *disturbance rejection bandwidth*, ω_{DRB}, is the max frequency at which the disturbance rejection (i.e. sensitivity $\mathcal{S}$) is below a certain amount, usually -3 decibels
  - We always want to maximize this

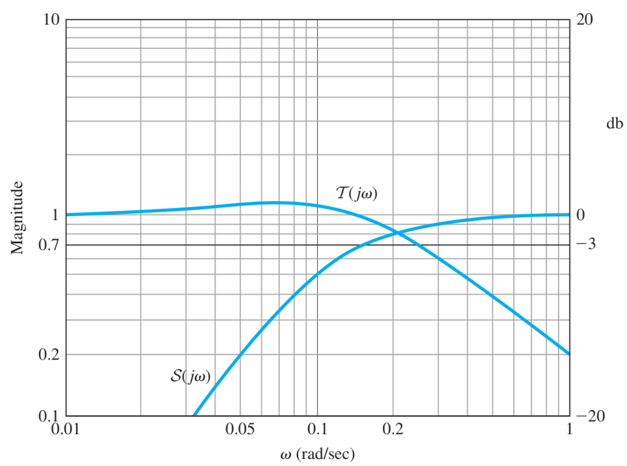


Figure 90: Bode magnitude plots of the closed-loop transfer function and sensitivity transfer function.

Lecture 25, Apr 8, 2024

Dynamic Response from Frequency Response

- For common systems, typically the open-loop transfer function has $|KG(j\omega)| \gg 1$ for $\omega \ll \omega_c$ and $|KG(j\omega)| \ll 1$ for $\omega \gg \omega_c$
  - Therefore at $\omega \ll \omega_c$, $|\mathcal{T}(j\omega)| \approx 1$, and at $\omega \gg \omega_c$, $|\mathcal{T}(j\omega)| \approx |KG(j\omega)|$
  - The magnitude of the closed-loop gain near ω_c is closely related to the phase margin
    * Note again that this peak is not exactly at ω_c
  - e.g. for $K = 1$, if $PM = 45°$, then $\angle G(j\omega_c) = -180° + PM = -135°$, and $|G(j\omega_c)| = 1$ by definition, then $|\mathcal{T}(j\omega_c)| = \left|\dfrac{G(j\omega_c)}{1 + G(j\omega_c)}\right| = \dfrac{1}{|\sqrt{(1 + \cos(-135°))^2 + \sin^2(-135°)}|} = 1.31$
- By the above calculation, $PM = 90°$, then $\omega_c = \omega_{BW}$ exactly; if $PM < 90°$, then $\omega_c \leq \omega_{BW} \leq 2\omega_c$
  - ω_{BW} is always within 1 *octave* of ω_c
  - Bandwidth is roughly equal to the natural frequency of the system, again within 1 octave
  - We typically define a closed-loop system by its bandwith and phase margin
    * $\zeta \approx \dfrac{PM°}{100}$ for $PM < 65°$ and $\omega_n \approx \omega_{BW}$
- We can find system type in the frequency response; for a unity feedback system:
  - A type 0 system's open-loop magnitude plot starts with a slope of 0 at low frequencies

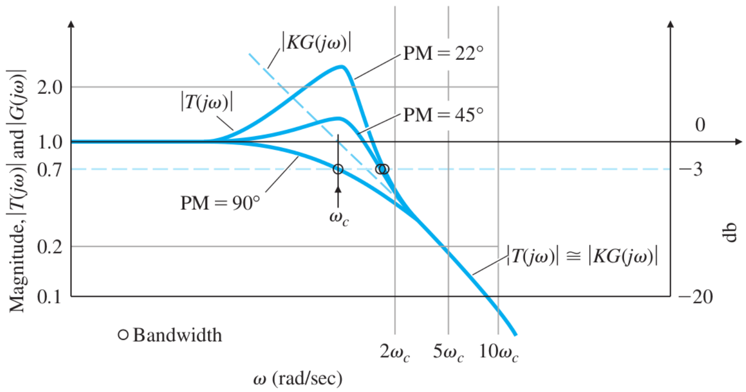


Figure 91: Closed-loop gain at ω_c for different phase margins.

* To have a slope of 0 at low frequencies means our class 1 term has a power of $n = 0$, so it does not contribute an initial slope, so this means no poles at the origin and thus type 0
* The low-frequency gain, K_0, is equal to the position constant K_p, since $K_p = \lim_{s\to 0} KD_cG(s) = \lim_{\omega\to 0} |KD_cG(j\omega)|$
* We can control the steady-state error by controlling the gain K – we are shifting the entire plot up or down, which changes the low-frequency gain

– A type 1 system's open-loop magnitude plot starts with a slope of -1

* Now $K_v = \lim_{s\to 0} sKD_cG(s) = \lim_{\omega\to 0} \omega|KD_cG(j\omega)|$ so at low frequencies, $|KD_cG(j\omega)| \approx \dfrac{K_v}{\omega}$
* We can find K_v by going to $\omega = 1$ and finding the intersection of the initial asymptote with the vertical line $\omega = 1$

– A type 2 system's open-loop magnitude plot starts with a slope of -2

* At low frequencies, $|KD_cG(j\omega)| \approx \dfrac{K_a}{\omega^2}$
* Similarly, we can find K_a by finding the intersection of the initial slope -2 asymptote with the vertical line $\omega = 1$

Lead Compensator Design

- Consider a PD controller, $D_c(s) = (T_Ds + 1)$, which is added to improve stability and dynamic response
  – This is a numerator class 2 term, which steps up the slope of the magnitude plot at the break point $\dfrac{1}{T_D}$
  – This essentially increases ω_c, which increases ω_{BW} and therefore ω_n which speeds up the system
  – This also increases the phase (since it's a denominator term), which increases the phase margin, which increases damping
  – Shortcomings:
    * At low frequencies the gain is 1, so this doesn't do much to the steady-state response
    * At high frequencies (i.e. noise), the gain is very high, so the noise is amplified
      - Instead, we often use a lead compensator, which has a gain that flattens at higher frequencies, to avoid noise amplification
  – Specify the break point so that the amount of increased phase desired happens near the crossover, so we can increase the PM
    * From the design requirements and the Bode plot of the uncompensated system, we can see how much additional PM we need
- Practically, we use a lead compensator, $D_c(s) = \dfrac{T_Ds + 1}{\alpha T_Ds + 1}$ where $\alpha < 1$, with corner frequencies $\omega_l = \dfrac{1}{T_D}$ (low) and $\omega_h = \dfrac{1}{\alpha T_D}$ (high)

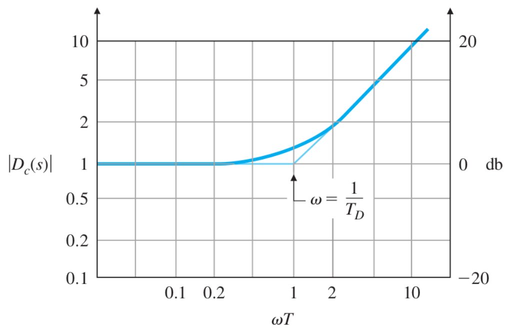


Figure 92: Bode magnitude plot for PD control.

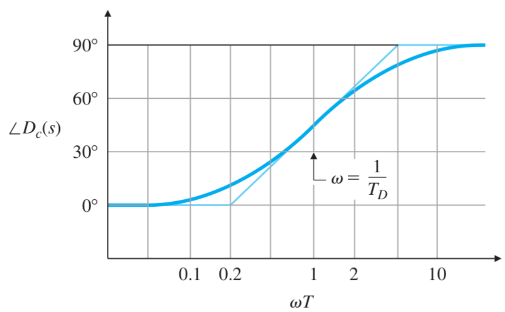


Figure 93: Bode phase plot for PD control.

- The additional denominator class 2 term steps the slope down at higher frequencies (so the magnitude plot becomes flat), so we avoid amplifying high frequency noise
- This comes at the cost of having the phase going up and then back down (instead of staying at +90° like the PD controller); the corner frequencies need to be chosen carefully so we get the maximum amount of increase to the PM
  * We typically choose $\omega_h \gg \omega_l$, typically $\omega_h > 5\omega_l$
- The phase increase is $\phi = \angle D_c(j\omega) = \tan^{-1}(T_D\omega) - \tan^{-1}(\alpha T_D\omega)$
  * This gives $\phi_{max} = \tan^{-1}\left(\frac{1}{\sqrt{\alpha}}\right) - \tan^{-1}(\sqrt{\alpha})$ occurring at $\omega_{max} = \frac{1}{T_D\sqrt{\alpha}}$ (by differentiation)
  * $\sin\phi_{max} = \frac{1-\alpha}{1+\alpha} \implies \alpha = \frac{1-\sin\phi_{max}}{1+\sin\phi_{max}}$
    - This gives us a simpler form to find α from ϕ_{max}
    - In design, we decide how much ϕ_{max} to use, and then we obtain α
  * $\frac{1}{\alpha}$ is the *lead ratio*; the higher the lead ratio, the more we approach a PD compensator
    - Selecting this is a tradeoff between a desired PM (for good damping) and an acceptable level of high-frequency noise amplification
    - Rule of thumb is to have have a lead compensator contribute no more than 70° to the phase; if we need even more, a double lead compensator can be used

- Both PD controller and lead compensator have no poles at the origin, so the system type is not changed
- Example: for the plant $\frac{1}{s(s+1)}$, design a lead compensator to obtain a response to a unit-ramp input with an overshoot $M_P < 25\%$ and steady-state error of no more than 0.1
  - The open-loop transfer function is type 1 (we couldn't have changed it with a lead compensator anyway)
  - Open loop TF: $L(s) = K\frac{T_Ds+1}{\alpha T_Ds+1} \cdot \frac{1}{s(s+1)}$

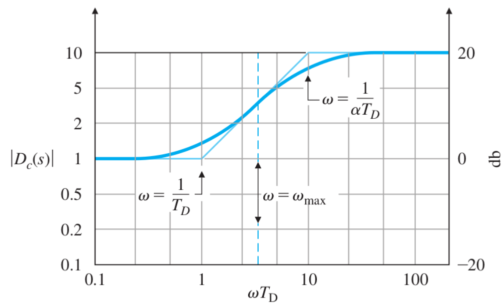


Figure 94: Bode magnitude plot for lead compensator.

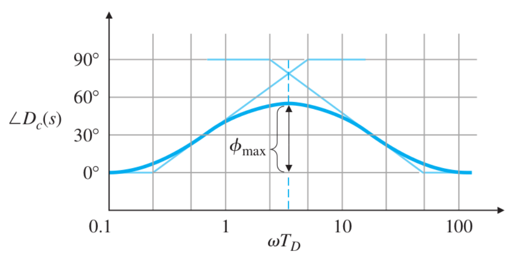


Figure 95: Bode phase plot for lead compensator.

- For $R(s) = \frac{1}{s^2}$, $e_{ss} = \lim_{s \to 0} \frac{1}{s + KD_c(s)\frac{1}{(s+1)}} = \frac{1}{KD_c(0)}$, therefore we need $K_v = KD_c(0) \geq 10$ from the steady-state error requirement
  * This yields a value of $K = 10$, since the lead compensator always has $D_c(0) = 1$
  * Since we don't have a lag compensator we have to use K for the steady-state response; if we had one we could save K to optimize the dynamic response
- For $M_P < 25\%$, we use the direct relation to get $PM = 45°$
- PM of the uncompensated system is only 20°, so we need to add more than 25°
  * The phase increase needs to be more than 25°, since the compensator zero increases ω_c due to the increase in slope, and the overall trend in phase is decreasing
  * We need to add a safety margin
- For $\phi_{max} = 40°$ of lead, $\frac{1}{\alpha} = 5$
- To get T_D we normally look at the desired ω_c (which influences system speed)
  * For this question we don't have a restriction on speed
  * $\frac{1}{T_D} = \omega_{max}\sqrt{\alpha}$
  * By trial and error selecting ω_{max}, we find $T_D = 0.5$
- The final controller is $D_c(s) = 10\frac{\frac{s}{2} + 1}{\frac{s}{10} + 1}$

- For a lead compensator, we specify the parameters from design requirements as follows:
  - The crossover frequency ω_c, which determines the bandwidth hence and speed of response
  - The phase margin PM, which determines the damping ratio and overshoot
  - The low-frequency gain K, which determines the steady-state error
  - In general, lead compensation increases the ratio $\frac{\omega_c}{KD_cG(0)}$
- Design procedure for lead compensator:
  1. Determine K to satisfy error or bandwidth requirements
     - For error, pick K to satisfy the error constant
     - For bandwidth, pick K so that ω_c is within a factor of two below the desired closed-loop bandwidth

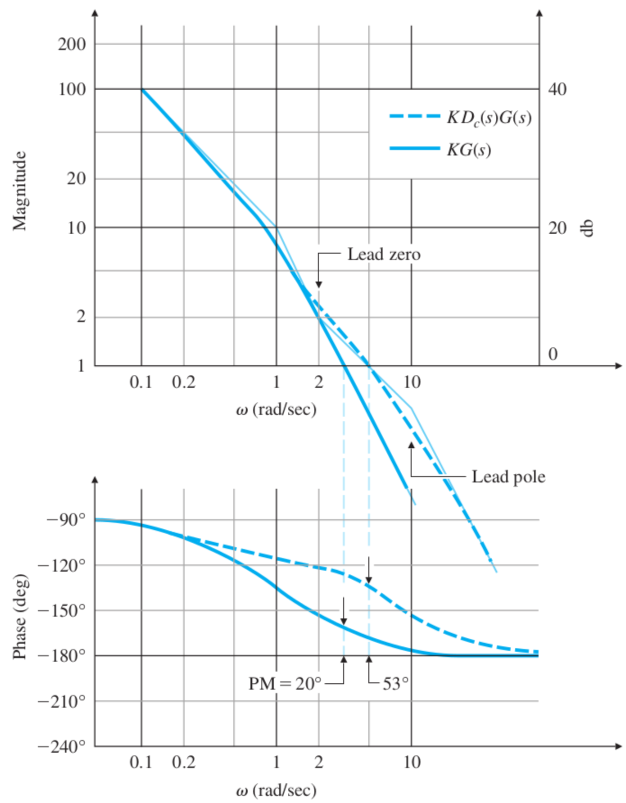


Figure 96: Bode plot of the example lead compensated system.

2. Evaluate the PM of the uncompensated system using this K
3. Find the amount of PM increase we need (add a safety margin, usually 5° or more)
4. Determine $\alpha = \dfrac{1-\sin\phi_{max}}{1+\sin\phi_{max}}$
5. Pick the desired crossover frequency and make ω_{max} there, and determine T_D using $\dfrac{1}{T_D} = \omega_{max}\sqrt{\alpha}$
6. Draw the compensated frequency response and check that the PM requirement is satisfied; iterate if not

- Example: type 1 servo mechanism, $KG(s) = K\dfrac{10}{s(s/2.5+1)(s/6+1)}$; design a lead compensator to obtain $PM = 45°$ and $K_v = 10$
  1. $\dfrac{1}{K_v} = \dfrac{1}{10} = \lim_{s\to 0} s\dfrac{1}{1+KD_c(s)G(s)}\dfrac{1}{s^2} \implies K = 1$
  2. Uncompensated PM is $-4°$ at $\omega_c \approx 4$
  3. We want the lead to add $\phi_{max} = 54°$ (with a safety margin of 5°)
  4. Use formulas to get $\alpha = 0.1$
  5. Choose a desired ω_c, e.g. 6 (in this case we have no hard speed requirement), giving $T_D = \dfrac{1}{\omega_c\sqrt{\alpha}} \approx 0.5$
  6. Draw the new Bode plot for $D_{c1}(s) = \dfrac{s/2+1}{s/20+1} = 10\dfrac{s+2}{s+20}$
     - We see that the PM requirement is not satisfied!
     - More iterations show that a single lead compensator cannot meet this PM requirement due to the high-frequency slope of -3
  7. Double the lead compensator; on examination this gives $PM = 46°$, meeting the requirements

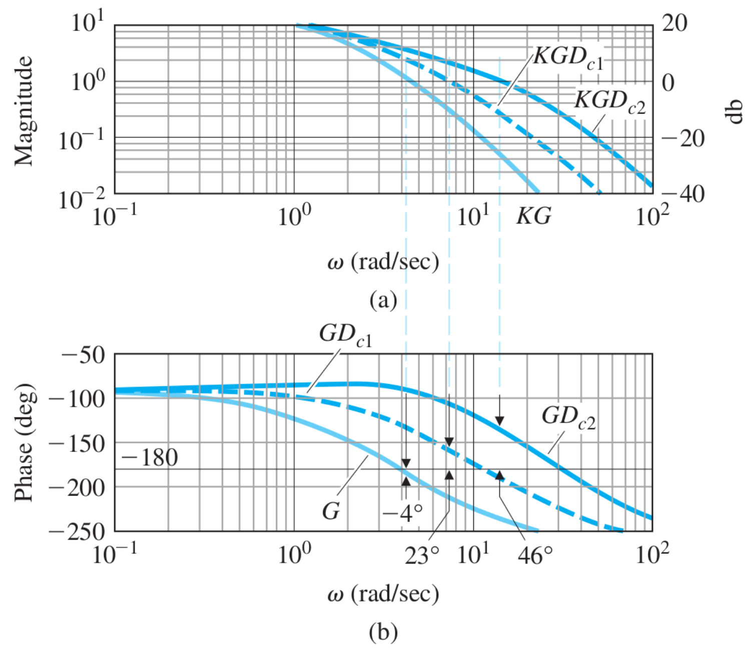


Figure 97: Bode plots for the uncompensated system, and the two iterations of lead compensators.

Lecture 26, Apr 12, 2024

Design for Dynamic Compensation

- For PI control, $D_c(s) = 1 + \frac{1}{T_I s}$, the steady-state error of the system is reduced with minimal impact on the bandwidth
  - The gain is high at low frequencies, which reduces the steady-state error
    * This increases the system type
  - However, it causes the reduction of phase margin at frequencies lower than the breakpoint $\frac{1}{T_I}$, which degrades stability
    * This makes sense since we know that an integral controller may destabilize the system
  - We usually place the break point $\frac{1}{T_I}$ at a frequency substantially less (one octave to multiple decades) than the crossover frequency, so that the impact on PM is minimal
  - The main practical problem of PI control is *integral windup* (aka *overflow*), leading to saturation of the system
    * A sudden change in the reference causes the integral term to accumulate too much
    * This leads to a very sluggish controller

- We typically use a lag compensator instead, $D_c(s) = \alpha \frac{T_I s + 1}{\alpha T_I s + 1}$, where $\alpha > 1$, so the pole has a lower break point than the zero
  - This can decrease the steady-state error without lowering the crossover frequency
  - The magnitude no longer increases to infinity at lower frequencies and instead converges to α, however the phase at low frequencies now converges to 0, instead of the $-90°$ before
    * This allows us to still reduce the steady-state error, without sacrificing too much phase margin
  - We choose the poles and zero relatively close together, and well below (one octave to multiple decades) the crossover frequency (i.e. choose a large T_I)
    * Having the corner frequencies far from the break point minimizes the reduction in phase margin

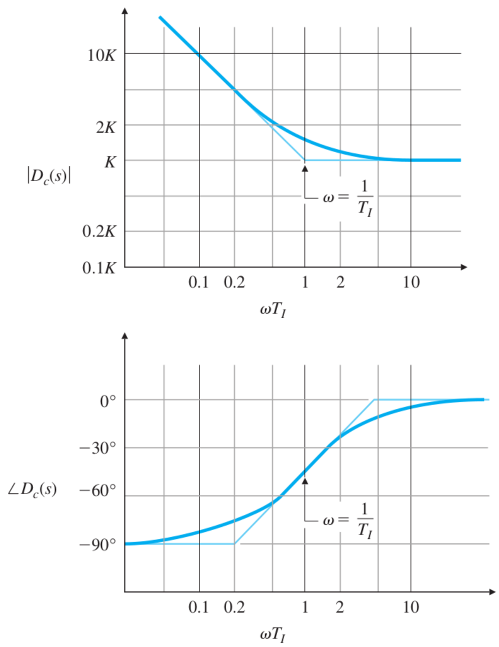


Figure 98: Bode plots of the integral controller.

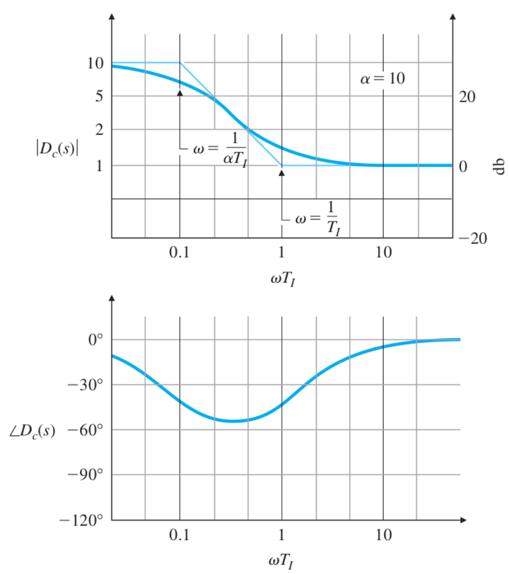


Figure 99: Bode plots of the lag compensator.

- Lag compensator design procedure:
  1. Determine the gain K required to get the desired PM without compensation, with a 5° to 10° margin to account for the PM reduction of the compensator
  2. Draw the Bode plot of the uncompensated open-loop TF and check the low-frequency gain, which gives the steady-state error
  3. Determine the value of α to meet the steady-state error requirement – α is how much more we need to multiply the low-frequency gain by in order to meet the steady-state error requirement
  4. Choose the upper corner frequency $\frac{1}{T_I}$ (the zero) to be one octave to multiple decades below the uncompensated ω_c
  5. Iterate on the design and verify that it meets requirements
- Example: $G(s) = \frac{115}{(s+1)(s+3)(s+28)}$, design a lag compensator to get an overshoot of less than 15% and a steady-state error of less than 2%
  - $M_P < 15\% \implies PM = 56°$ from the plots; having a margin gives $PM = 66°$
  - We have no restrictions on ω_c so pick it so that we get $PM = 66°$
    - Plot the Bode plot and find that $\omega_c = 2.5$ gives us the desired PM; at this value, the gain is currently 0.38
    - We get $K = 2.63$ to get us the desired ω_c
  - Using this value of K, the uncompensated K_p is 3.6 (same as the value of the magnitude plot at $\omega = 0$)
  - We want $e_{ss} = \frac{1}{1+K_p} < 0.02 \implies K_p > 49$ instead; choose $K_p = 50$ for some margin
    - Therefore $\alpha = \frac{50}{3.6} = 14$
  - Choose $\frac{1}{T_I}$ one decade below the crossover frequency, and double check that requirements are satisfied
- A PID compensator $D_c(s) = K(T_D s + 1)\left(1 + \frac{1}{T_I s}\right)$ can be used to improve both transient and steady-state responses
  - Roughly equivalent to combining lead and lag compensators
    - $D_c(s) = \gamma\left(\frac{T_D s + 1}{\frac{T_D}{\gamma}s + 1}\right)\left(\frac{T_I s + 1}{\gamma T_I s + 1}\right)$ for $\gamma > 1$
- For some systems, the Bode plot will cross over the real axis multiple times; this results from the natural modes of vibration of the system
  - *Gain stabilization* is the simple approach of modifying K to bring the entire plot down
  - *Phase stabilization* is the use of notch compensators that remove the system's response at the problematic frequencies
- A lead-lag compensator, $D_c(s) = \beta\left(\frac{T_I s + 1}{\beta T_I s + 1}\right)\left(\frac{T_D s + 1}{\alpha T_D s + 1}\right)$ for $\alpha < 1, \beta > 1$, combines both
- Example: given $G(s) = \frac{1}{s^2(s+2)}$, design a lead-lag compensator to get $t_r \leq 1$, $M_p \leq 40\%$, $t_s \leq 10$ (for 2%), and $e_{ss} \leq 10\%$
  - Convert: $\omega_n \geq 1.8$, $\zeta \geq 0.3 \implies PM \geq 30°$, $\sigma = \zeta\omega_n \geq 0.46$, $e_{ss} = \frac{1}{K_a} \leq 0.1$
    - The requirement for e_{ss} suggests that a lead-lag compensator is likely needed
  - Initially, choose $\omega_n = 2$ and so $\omega_{BW} \approx 2$; start with crossover frequency at half bandwidth, $\omega_c = 1$, and phase margin of 40° (with margin added)
  - At ω_c, the magnitude $|G(j\omega_c)| = 0.447$ for the uncompensated system, so choose $K = \frac{1}{0.447}$ to make this the crossover frequency
  - The phase is $-207°$ at ω_c, so the initial phase margin is $-27°$ – we need to add $\phi_{max} = 67°$ of phase margin
  - Using the formula, $\alpha = 0.042$

- Choose $\omega_{max} = \omega_c = 1.0$, so that $D_{c1}(s) = \frac{4.88s + 1}{0.21s + 1}$
  * Now at ω_c the magnitude is 4.86, so reduce K further by this factor to get $K = 0.46$
  * This gives a PM of 40°
- Plotting the step response shows that the overshoot and settling time meet requirements, but not rise time (by a very small amount)
  * Increase K by a small amount to 0.5, which increases overshoot and allows meeting the rise time requirement
- The existing steady-state error is 0.25; we need $e_{ss} = \frac{1}{K_a} \leq 0.1$ so $K_a \geq 10$
  * The open-loop gain at small frequencies needs to be increased by a factor of 40, so $\beta = 40$
- Choose the upper corner frequency at a tenth of ω_c, so $\frac{1}{T_I} = 0.1$, giving $T_I = 10 \implies D_{c2}(s) = 40\frac{10s + 1}{400s + 1}$
  * As expected, the open-loop response is faster with worse overshoot
- In time domain this now has an overshoot that is slightly over the limit, so we need to iterate:
  * Try $\frac{1}{T_I} = 0.05 \implies T_I = 20 \implies D_{c2}(s) = 40\frac{20s + 1}{800s + 1}$
    - This now doesn't meet the requirements