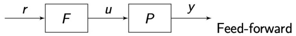
Control Systems 1

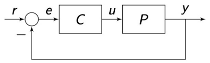
Loris Frey

Version: September 11, 2025

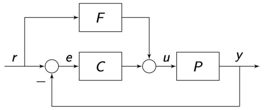
Systems and Control architecture



Feed-forward



Feedback



Two degrees of freedom

Feed-forward/Open-loop: Requires precise knowledge of the system and the environment as the output can't be controlled.
Feedback: Stabilizes unstable system, but can also cause instability in otherwise stable systems.
Two degrees of freedom: Combines the benefits of feed-forward and feedback control. Allows independent tuning of the tracking behavior (how well the output follows the reference) and the disturbance rejection or stability properties.
Cascaded control: In the case, where we have a system with two outputs and a single input, the control architecture is split into an inner and outer loop. The bandwidth of the inner loop must be much faster than the bandwidth of the outer loop. As the inner loop is closer to the disturbance it can better react to it.

System Classification

Linear vs Nonlinear: An input-output system Σ is linear if, for all input signals u_a, u_b and scalars $\alpha,\ \beta \in \mathbb{R}$

$$\Sigma(\alpha u_a + \beta u_b) = \alpha(\Sigma u_a) + \beta(\Sigma u_b) = \alpha y_a + \beta y_b$$

Examples:

- Linear: $y(t) = u(t^2)$, $y(t) = \int_{-\infty}^{t} u(\tau)d\tau$, $y(t) = \frac{d}{dt}u$
- Non-linear: $y(t) = u(t) + 1$, $y(t) = \int_0^t u^2(\tau)d\tau$, $y(t) = u(t) + t$, $y(t) = \cos(u(t))$

Causal vs Non-causal: An input-output system Σ is causal if future inputs do not affect the present output. Examples:

- Causal: $y(t) = u(t-\tau)\,\forall\tau > 0$, $y(t) = \cos(3t)u(t-1)$, $y(t) = \int_{-\infty}^{t} u(\tau)d\tau$, $y(t) = \int_{-\infty}^{t-a} u(\tau + a)d\tau$
- Non-causal: $y(t) = \int_{-\infty}^{t+1} u(\tau)d\tau$, $y(t) = u(bt),\ \forall b > 1$, $y(t) = u(t-a)\,\forall a < 0$, $y(t) = \frac{d}{dt}u$, $y(t) = u(t+1)$

Memoryless (or Static) vs Dynamic: An input-output system Σ is memoryless if the output at the present time depends only on the input at the present time.
Time invariant vs Time-varying: A time invariant system is a map between input and output signals that is the same at any point in time. More formally a system is time-variant if

$$y(t-\tau) = \Sigma\tilde{u}(t), \quad \forall\tau, t \in \mathbb{T}$$

where $\tilde{u} = u(t-\tau)$.
Remark: For an integral, we have to align the integration bounds to be able to compare the two terms. We do this using the change of variables $l = \rho - \tau \Leftrightarrow \rho = l + \tau$. This results in

$$\Sigma\tilde{u} = \int_0^t u(\rho - 4 - \tau)d\rho = \int_0^{t-\tau} u(l-4)dl$$

We can then check whether this mathches the term for $y(t-\tau)$.
Examples:

- Time-invariant: $y(t) = \int_0^t u^2(\tau)d\tau$, $y(t) = \frac{d}{dt}u$
- Time-variant: $y(t) = tu(t)$, $y(t) = u(\sin(t))$, $y(t) = u(t^2)$, $y(t) = \cos(u(t))$, $y(t) = u(3t+1)$, $y(t) = u(-t)$

State-Space Form

Linear, time-invariant systems can be fully described using the state-space form

$$\begin{cases} \dot{x}(t) = f(t) = Ax(t) + Bu(t) \\ y(t) = g(t) = Cx(t) + Du(t) \end{cases}$$

The response of an LTI system consists of the superposition of the initial condition response and forced response:

$$\begin{cases} x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \\ y(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t) \end{cases}$$

The first term in y is the initial condition response y_{IC}, the rest is the forced response y_{F}. The last term is also called feed-through.
Careful: 1) Make sure not to forget u. 2) Multiply the feed-through term D with u. 3) Do they want the forced response or time response?
Remarks:

- The time response is always a real valued function.
- The time response for distinct real eigenvalues of A is a linear combination of exponentials $e^{\lambda_i t}$.
- The time response for complex eigenvalues of A is a linear combination of sinusoids.

Linearization

A system is at equilibrium when for a given input u_e, $x(t)$ remains constant over time, i.e. $\dot{x}(t) = f(x_e, u_e) = 0$. Nonlinear systems can be linearly approximated near their equilibrium positions (x_e, u_e):

$$\begin{cases} \dot{x} = f(x(t), u(t)) \\ y = g(x(t), u(t)) \end{cases} \rightarrow \begin{cases} \dot{x} = Ax + Bx \\ y = Cx + Dx \end{cases}$$

with

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{(x_e, u_e)}, \quad B = \begin{bmatrix} \frac{\partial f_1}{\partial u} \\ \vdots \\ \frac{\partial f_n}{\partial u} \end{bmatrix}_{(x_e, u_e)}$$

and

$$C = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n} \end{bmatrix}_{(x_e, u_e)}, \quad D = \begin{bmatrix} \frac{\partial g(x,u)}{\partial u} \end{bmatrix}_{(x_e, u_e)}$$

Hartman-Grobman theorem: If the linearization is asymptotically stable, then the non-linear system is locally asymptotically stable.

Stability

The eigenvalues of A determine the stability of the system:

- Asymptotically stable: State converges to zero for bounded initial conditions and zero input. $\mathrm{Re}(s) < 0$ for all λ_i.
- Lyapunov stable: State will remain bounded for bounded initial conditions and zero input. $\mathrm{Re}(s) \leq 0$ for all λ_i.
- BIBO stable: Output remains bounded for every bounded input.

We have the following relations

- Asymptotically stable $\rightarrow$ BIBO and Lyapunov stable
- Asymptotically stable $\leftrightarrow$ BIBO stable (for minimal LTI systems)

Transfer Functions

The transfer function describes how a stable system G transforms an input $u = e^{st}$ into the output $y_{ss} = G(s)e^{st}$ at steady state. It is defined as

$$G(s) = C(sI - A)^{-1}B + D \ \in \mathbb{C}$$

with $A \in n \times n$, $B \in n \times 1$, $C \in 1 \times n$ and $D \in \mathbb{R}$.
Careful: This only holds, when s is not an eigenvalue of A.
Remarks: Performing a pole-zero cancellation is only possible if the zero that is canceled is minimum phase (and hence the poles stable).

Partial fraction form

$$G(s) = \frac{r_1}{s - p_1} + \frac{r_2}{s - p_2} + \ldots + \frac{r_n}{s - p_n} + r_0$$

where $r_0 = d$, p_i are the poles and r_i are the residues:

- p_i is a single pole:

$$\boxed{r_i = \lim_{s \to p_i} [(s - p_i) \cdot G(s)]}$$

- p_i is repeated m-times:

$$\boxed{r_i = \frac{1}{(m-1)!} \lim_{s \to p_i} [\frac{d^{m-1}}{ds^{m-1}}((s - p_i)^m \cdot G(s))]}$$

Root Locus form

$$G(s) = \frac{k_{\mathrm{RL}}}{s^q} \frac{(s - z_1)(s - z_2)\ldots(s - z_m)}{(s - p_1)(s - p_2)\ldots(s - p_{n-q})}$$

where z_i are the zeros and p_i the poles of $G(s)$. The poles are the eigenvalues of A. A system is minimum phase if $\Re(z_i) < 0,\ \forall i$.

Bode form

$$G(s) = \frac{k_{\mathrm{Bode}}}{s^q} \frac{(\frac{s}{-z_2} + 1)(\frac{s}{-z_1} + 1)\ldots(\frac{s}{-z_m} + 1)}{(\frac{s}{-p_2} + 1)(\frac{s}{-p_1} + 1)\ldots(\frac{s}{-p_{n-q}} + 1)}$$

Transfer Function to State-space

If the Transfer function is written as a partial fraction expansion of the form then a realization is

$$A = \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix}, \quad B = \begin{bmatrix} \sqrt{r_1} \\ \vdots \\ \sqrt{r_n} \end{bmatrix}$$

$$C = \begin{bmatrix} \sqrt{r_1} & \cdots & \sqrt{r_n} \end{bmatrix}, \quad D = [d]$$

Careful: The sign of p_i is inverted with respect to the transfer function.
In the general case

$$G(s) = \frac{b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \ldots + b_0}{s^n + a_{n-1}s^{n-1} + \ldots + a_0} + d$$

the minimal realization of $G(s)$ is given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$C = \begin{bmatrix} b_0 & b_1 & \cdots & b_{n-1} \end{bmatrix}, \quad D = [d]$$

which is called **controllable canonical form**.
Careful: The order is reversed with respect to the transfer function.
Remarks:

- When there are no pole-zero cancellations the order of the denominator corresponds to the dimension of the matrix A.
- If the orders of the numerator and denominator are the same, $D \neq [0]$.

Initial- and Final Value Theorem

The short- and long-term output of a system are calculated using:

- Initial value theorem:

$$\boxed{\lim_{t \to 0}[y(t)] = \lim_{s \to \infty}[sY(s)]}$$

- Final value theorem:

$$\boxed{\lim_{t \to \infty}[y(t)] = \lim_{s \to 0}[sY(s)]}$$

The latter only applies for at most one integrator.

Magnitude and Phase

A sinusoidal input to an asymptotically stable system causes a sinusoidal steady-state response with the same frequency but a different amplitude and phase-shift:

$$\boxed{y_{ss} = |G(j\omega)| \cdot \sin(t + \angle G(j\omega))}$$

where ω is the input frequency, $|G(j\omega)|$ the amplitude and $\angle G(j\omega)$ the phase shift:

$$\boxed{|G(j\omega)| = |k| \frac{\prod_i |j\omega - z_i|}{\prod_j |j\omega - p_j|}},$$

$$\boxed{\angle G(j\omega) = \angle k + \sum_i \angle(j\omega - z_i) - \sum_j \angle(j\omega - p_j)}$$

Careful: Always check if there are any double zeros or poles.

Pole-zero map and Step response pairings

The poles and zeros of the transfer function influence the initial condition response to a step input:

- Minimum-phase zeros (LHP): Cause fast initial response and overshoot, which increases closer to the imaginary axis.
- Non-minimum-phase zeros (RHP): Cause undershoot, which increases closer to the imaginary axis.
- Real part of the pole: Determines the decay rate and rise time. The more negative the real part, the faster the decay and the rise.
- Imaginary part of the pole: Determines the oscillation frequency and overshoot. The larger the imaginary part, the higher the frequency and greater the overshoot.

Example:

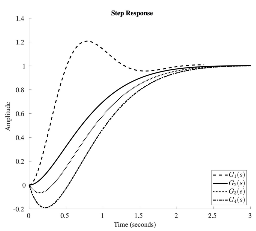


Figure 8: Step responses of G_i, $i = 1, 2, 3, 4$.

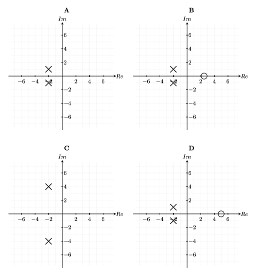


Figure 9: Pole-zero maps of G_i, $i = 1, 2, 3, 4$. Poles ($\times$) and zeros ($\circ$)

Non-minimum phase zeros in B) and D) which cause undershoot. As the zero in B is closer to the imaginary axis it has greater undershoot. $B \rightarrow G_4$ and $D \rightarrow G_3$. The imaginary part of the poles in C) greater than in A) which means a higher oscillation frequency. $C \rightarrow G_1$ and $A \rightarrow G_2$.

Dominant Poles approximation

A system of higher order can be approximated by a system of lower order by only keeping the dominant poles. Dominant poles are those closest to the imaginary axis and which are far away from the zeros.

- If $p_i \approx z_i$ with negative real part, cancel p_i and z_i
- Cancel the poles that are furthest away from the imaginary axis (which have the most negative real part).

The Root Locus

The Root Locus shows how the poles of a closed-loop transfer function change with the proportional gain k. The following rules help us sketch the root-locus:

1. The Root Locus is symmetric with respect to the real axis.
2. The number of closed-loop poles is the same as the number of open-loop poles.
3. As k increases from 0 to ∞ the closed-loop poles move from the open-loop poles to the open-loop zeros. If there are more closed-loop poles than zeros, the excess poles approach infinity along certain asymptotes. These asymptotes meet in the **center of mass**

$$\boxed{S = \frac{\sum p - \sum z}{m - n}}$$

with respective **angles**

$$\boxed{\alpha_n = \frac{2k+1}{m-n} \cdot 180^\circ}$$

where m is the number of poles, n the number of zeros and $k = \{0, 1, \ldots, m-n-1\}$.

4. All points on the real axis to the left of an odd number of poles/zeros are on the (positive) k root locus. The Root Locus is always sketched from right to left.
5. All points on the real axis to the left of an even number of poles/zeros (or none) are on the negative k root locus.
6. When two branches come together on the real axis, there will be "breakaway" or "break-in" points. Hence, the two poles that come together leave the real axis for the same value of k.
7. The zeros of the closed-loop system are equal to the zeros of the open-loop system. The zeros don't move as k changes.

Careful: Rules 4 and 5 assume that the numerator and denominator of $G(s)$ have positive leading coefficients. If this is not the case, the positive Root Locus of $G(s)$ corresponds to the negative Root Locus of $-G(s)$.

Magnitude- and Angle rule

The magnitude rule:

$$\frac{1}{|k|} = |L(s)| = \frac{|s-z_1| \cdot |s-z_2| \cdot \ldots \cdot |s-z_m|}{|s-p_1| \cdot |s-p_2| \cdot \ldots \cdot |s-p_n|}$$

We can use this to find the critical gain for which a closed-loop system becomes unstable:

1. Determine the open-loop $L(s)$ and write down

$$\frac{1}{|k_{\text{crit}}|} = |L(0)|$$

2. Solve for k_{crit}.

The angle rule:

$$\begin{aligned}\angle(s-z_1) + \angle(s-z_2) + \ldots + \angle(s-z_m) \\ - \angle(s-p_1) - \angle(s-p_2) - \ldots - \angle(s-p_n) \\ = \begin{cases} 180^\circ(\pm 360^\circ) & \text{if } k > 0 \\ 0^\circ(\pm 360^\circ) & \text{if } k < 0 \end{cases}\end{aligned}$$

Time-domain specifications

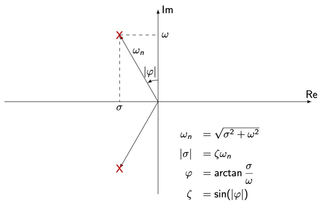
Time-domain specifications quantify a system's transient response and its steady-state error, defining how fast and how accurately the output reaches its target.

Transient response

We consider a second-order system

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

with poles $p_{1,2} = \sigma \pm j\omega$ and $\omega_n = \sqrt{\sigma^2 + \omega^2}$.



We then have the following specifications:

- **Settling time:**

$$\boxed{T_d = \tau \log(100/d) = \frac{1}{|\sigma|}\log(100/d)}$$

The more negative the real part of the pole, the faster the decay and the smaller the settling time.

- **Time to peak:**

$$\boxed{T_p = \frac{\pi}{\omega}}$$

The larger the imaginary part, the shorter the time to peak.

- **Peak overshoot (and damping ratio):**

$$\boxed{M_p = e^{-\frac{|\sigma|\pi}{\omega}}}$$

The larger the imaginary part, the larger the Peak overshoot. Careful: For the peak overshoot to be smaller than $\frac{1}{e}$, $\frac{|\sigma|\pi}{\omega}$ has to be larger than 1. M_p is directly related to the damping ratio of the system:

$$\boxed{\zeta^2 = \frac{\ln(M_p)^2}{\pi^2 + \ln(M_p)^2}}$$

- **Rise time:**

$$\boxed{T_{100\%} = \frac{\pi/2 - \varphi}{\omega} \approx \frac{\pi}{2\omega_n}}$$

- **(Underdamped systems) Peak frequency:**

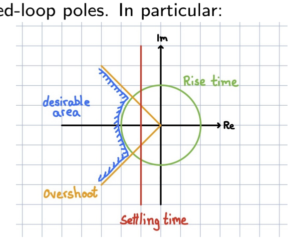
$$\boxed{\omega_p = \omega_n\sqrt{1 - 2\zeta^2} \approx \omega_n}$$

and

$$\boxed{G(j\omega_p)| = -20\log_{10}(2\zeta\sqrt{1-\zeta^2})}$$

where ζ can be read off from the transfer function.

These specifications impose constraints on the locations of the dominant closed-loop poles. In particular:



- Poles with short settling time lie on the left of the red line.
- Poles between the orange lines have small overshoots.
- Poles outside the green circle have small rise times.

Steady-state error

The steady-state error e_{ss} of the closed-loop system is the difference between the reference value and the output for $t \to \infty$:

$$\boxed{e_{ss} = \lim_{t\to\infty}[r(t) - y(t)] = \lim_{s\to 0}\left[\frac{sR(s)}{1 + L(s)}\right]}$$

Alternatively, e_{ss} can be derived as follows

e_{ss}	$q=0$	$q=1$	$q=2$
Type 0	$\frac{1}{1+k_{\text{Bode}}}$	∞	∞
Type 1	0	$\frac{1}{k_{\text{Bode}}}$	∞
Type 2	0	0	$\frac{1}{k_{\text{Bode}}}$

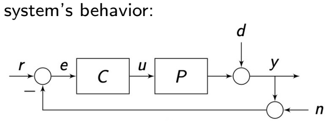
where q is the order of the ramp input and the Type corresponds to the number of poles at the origin. Written out, this means:

- The steady-state error gets smaller, as the Bode gain increases.
- In order to have zero steady-state errors to unit input ramps of order q, it is necessary to have at least $q+1$ integrators on the path from the error e to the (reference/disturbance) input.

Careful: Always check the given input before determining whether the steady-state error can be zero.

Controllers

Controllers are components that can be designed to manipulate and adjust a system's behavior:



where $C(s)$ is the controller, $P(s)$ the plant and $L(s) = C(s)P(s)$ the open-loop transfer function. r is the reference input, e the error, u the input to the plant, d the output disturbance and n the sensor noise.
The **Sensitivity** function

$$\boxed{S(s) = \frac{1}{1 + L(s)}}$$

maps the input reference r or the output disturbance d to the error e. For small errors relative to the reference inputs (error tracking) and disturbance rejection, $S(s)$ must be small (≈ 0).
The **Complementary Sensitivity** function

$$\boxed{T(s) = \frac{L(s)}{1 + L(s)} \approx L(s)}$$

maps the reference input r and the noise n to the output y. To reject noise $T(s)$ must be small (≈ 0) at high frequencies. For good command tracking it must be large (≈ 1) at low frequencies.

Design requirements

We can achieve desired control objectives as follows

- No steady state error to a step reference: Pole at $s = 0$ if there if the plant doesn't already have an integrator.
- Tracking error to sinusoidal inputs of frequencies up to ω_r should not exceed $d\%$ in magnitude:

$$|S(j\omega)| < \frac{d}{100} \quad \forall\omega < \omega_r$$

$$\frac{100}{d} < 1 + |L(j\omega_r)| \approx |L(j\omega)| = P(j\omega)C(j\omega)$$

- (Noise) frequencies higher than ω_r should be suppressed at least by a factor k:

$$|T(j\omega)| < \frac{1}{k}|P(j\omega)| \quad \forall\omega > \omega_r$$

- Phase margin of at least φ_r: Determine the gain cross-over frequency

$$L(j\omega_{gc}) = P(j\omega_{gc})C(j\omega_{gc})$$

and read off the corresponding phase margin. Careful: Use the Bode form when reconstructing the transfer function from the plant and don't forget the Bode gain.

PID Controllers

A (PID) controller creates an input signal to the plant that depends on the negative error e, the integral of e and the derivative of e:

$$u(t) = k_p \cdot e(t) + k_i \cdot \int_0^t e(\tau)d\tau + k_d \cdot \frac{de(t)}{dt}$$

The transfer function of a PID controller is:

$$C_{\text{PID}}(s) = k_d s + k_p + \frac{k_i}{s} = \frac{k_d s^2 + k_p s + k_i}{s}$$

With less poles than zeros, PID controllers are non-causal and hence not realizable. To make it causal, an additional element with $C(s) = \frac{1}{s-b}$ is added to the system. Increasing k_p, k_d and k_i changes the behavior of the system as follows:

	Advantages	Disadvantages
k_p	e_{ss} decreases, Faster response, Increases Bandwidth	More sensitive to noise, More oscillations, Phase margin decreases
k_d	Less oscillations, Phase margin increases, Reduces overshoot	More sensitive to noise, Slower response
k_i	e_{ss} decreases, Faster response	More oscillations, Phase margin decreases

Careful: Before choosing our controller we have to consider what characteristics the plant has for which we apply the controller.

Frequency Response

Bode Plots

The Bode plots visualizes the frequency dependency of the magnitude and phase of the steady state response using two separate plots. It can be used to determine the stability of open-loop stable and minimum phase systems.

Magnitude / Phase	-20 dB/dec	$+20$ dB/dec
-90°	stable pole	non-minimum phase zero
$+90^\circ$	unstable pole	minimum phase zero

Additionally, we have to consider the effects of the following terms

Term	Magnitude	Phase
Constant K	$20\log_{10}(\lvert K\rvert)$	$\begin{cases} 0^\circ & K > 0 \\ \pm 180^\circ & K < 0 \end{cases}$
Pole at Origin $\frac{1}{s}$	-20dB/dec	-90° for all ω
Zero at Origin s	$+20$dB/dec	$+90^\circ$ for all ω

Bode plot and Step response pairings

When matching Bode plots to step responses for a second order system we can consider the following points:

- If the step response to a Transfer function oscillates, the system must have complex conjugate poles, which are indicated by the presence of a peak in the Bode plot. The higher the frequency at which the peak appears, the faster the oscillations.
- If the step response to a Transfer function does not oscillate but instead approaches the steady-state, the system only has real poles. To differentiate between different rise times we have to consider the frequency in the Bode plot, at which the magnitude starts decreasing: The lower this frequency is, the longer the rise time.

Nyquist Plots

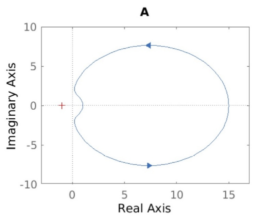
The Nyquist plot is a way to visualize the frequency response of a system as a parametric curve showing $G(j\omega)$ in the complex plane, in which ω is implicit. It is the most general tool for analysis, as it can be used to assess system stability of any LTI system.

1. Magnitude in the limit cases:
$$\lim_{\omega\to 0}|G(j\omega)| \quad \text{and} \quad \lim_{\omega\to\infty}|G(j\omega)|$$
2. Phase in the limit cases:
$$\lim_{\omega\to 0}\angle G(j\omega) \quad \text{and} \quad \lim_{\omega\to\infty}\angle G(j\omega)$$
Measured from the positive real axis in CCW direction. Tells us at what angle the Nyquist plot starts and ends or rather the angle of the asymptote at which it approaches the origin.
3. If this isn't unambiguous, consider the net phase-change of the system:
    - Determine the net phase change of the transfer function by looking at the effects of the poles and zeros.
    - Follow the Nyquist plot $\omega = 0$ to $\omega \to \infty$ and sum up the changes in angle (positive in CCW ↺ direction and negative in CW ↻ direction).
4. Match and the transfer function and the Nyquist plot.

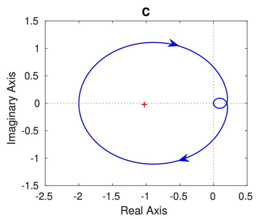
Remarks:

- The phase of the Nyquist plot is always measured from the origin.
- We trace the effects of the poles and zeros on the phase from the smallest absolute value to the largest one. This gives us the correct sequence of phase changes.
- When a transfer function has more poles than zeros $G(j\omega) \to 0$ for $\omega \to \infty$.
- If two transfer functions differ by a negative sign, then their Nyquist plots are related by a point reflection through the origin by 180° in the complex plane.
- If two Nyquist plots differ only in orientation, we simply need to determine which transfer function has a positive net phase change and match it to the plot that winds CCW ↺.

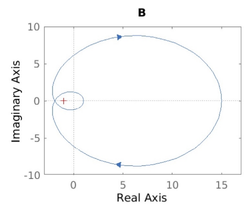
Nyquist plot net phase changes: Examples



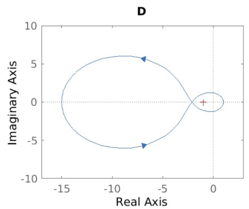
0°



-270°



-360°



-180°

The Nyquist criterion

We can use the Nyquist criterion to determine the number of unstable closed-loop poles:
$$Z = N + P$$
where P is the number of unstable open-loop poles, N is the number of CW ↻ encirclements of $-\frac{1}{k}$ and Z is the number of unstable closed-loop poles.
Careful: CCW ↺ encirclements have to be counted as -1.
How do we treat poles and zeros of $L(s)$ on the imaginary axis?

- We consider the pole on the imaginary axis to be in the LHP, so it does not contribute to P. We do this by drawing the D-contour to the right of it.
- Consequently, we have to close the Nyquist plot CW ↻ at infinity.

Note that we can also do this differently: We can count the pole on the imaginary axis as a RHP pole, which contributes to P and close the Nyquist plot CCW ↺ at infinity.

Nyquist Criterion: Example 1

Problem: Consider the closed-loop system T shown in Figure 17, where L is a linear time-invariant system.

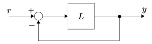


Figure 17: Closed-loop system T.

Assume that $L(s)$ is given by
$$L(s) = k\frac{s+a}{s(s+b)},$$
with $k, a, b \in \mathbb{R}\setminus\{0\}$ unknown. Through two experiments you are able to derive the Nyquist plot of the system which is shown in Figure 18, and you are able to determine that $k > 0$.

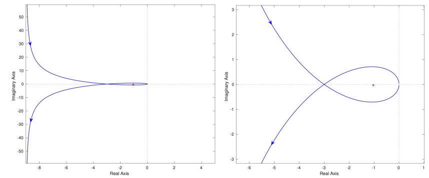


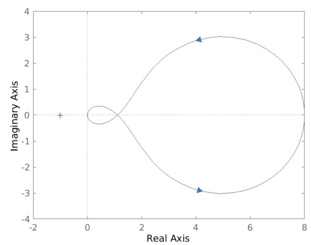
Figure 18: Nyquist plot of $L(s)$. [Left] Zoom-out into asymptotic behavior. [Right] Zoom-in around the (-1,0) point.

As there is one more pole than zero in the open loop transfer function, $|L(j\omega \to 0)| = \infty$ and $|L(j\omega \to \infty)| = 0$ which gives us the direction of the polar plot. Next, we note that $\angle L(j\omega \to 0) = -270^\circ$ and $\angle L(j\omega \to \infty) = -90^\circ$ which means that there is a net phase change of $+180^\circ$. This is only possible if a is a minimum phase zero and b an unstable pole. Note that due to the pole at $s = 0$ we have to close the Nyquist plot in Clockwise direction.

Nyquist Criterion: Example 2

The closed-loop system T is defined as above.

Assume that $L(s)$ is given by
$$L(s) = k\frac{s+a}{(s+b)(s+c)},$$
with $k, a, b, c \in \mathbb{R}$ unknown. Through two experiments you are able to derive the Nyquist plot of the system which is shown in Figure 10, and you are able to determine that the zero a is non-minimum phase, i.e. $a < 0$.



Careful: $a < 0$ means that the zero is non-minimum phase.
As $a < 0$, $L(s)$ has a non-minimum phase zero which contributes -90° to the angle. As the net phase change is -90° we must have one stable and one unstable pole. Hence, $P = 1$.

Phase and Gain margins

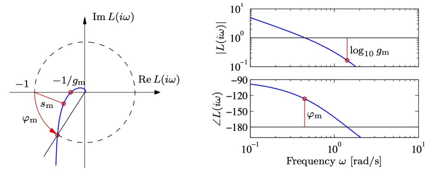
The gain margin and phase margin measure how close the system is to closed-loop instability.
Phase margin: Indicates how much the phase can be changed at the cross-over frequency ω_c before encircling -1.

1. Find the cross-over frequency ω_c:
$$|L(j\omega_c)| = 1 \quad \text{or} \quad |L(j\omega_c)|_{\text{dB}} = 0$$
2. Determine the phase from the Bode plot and calculate the phase margin:
$$\boxed{\varphi_m = \varphi + 180^\circ}$$

Gain margin: Indicates how much the magnitude can be changed at the frequency where the phase is -180° before encircling -1.

1. Find the frequency ω at which $\angle L(j\omega) = -180^\circ$. Then determine $|L(j\omega)|_{\text{dB}}$ for the same ω from the Bode plot.
2. Compute the gain margin
$$\boxed{g_{m,\text{dB}} = 0 - |L(j\omega)|_{\text{dB}} \quad \text{and} \quad g_m = 10^{\frac{g_{m,\text{dB}}}{20}}}$$

Careful: $|L(j\omega)|_{\text{dB}} = 20\log_{10}(|L(j\omega)|)$ can be both positive or negative. We have to read it off from the Bode plot.



In the graph both φ_m and $g_{m,\text{dB}}$ are positive.
The system is stable for

- $g_{m,\text{dB}} > 0 \quad \text{or} \quad |L(j\omega)|_{\text{dB}} < 0$
- $\varphi_m > 0$

Careful: This condition is only true, if the open-loop $L(s)$ is stable and minimum phase.
When the magnitude or phase of the system are changed, we have to make sure that the phase and gain margin are still positive in order for the system to be stable.

- Magnitude is increased: Subtract the increase in magnitude from g_m to get the new gain margin. Compute the phase margin at the new gain cross-over frequency.
- Phase is decreased: Subtract the decrease in phase from φ_m to get the new phase margin. Compute the gain margin at the new phase cross-over frequency.

An unstable system can only be stabilized by decreasing the magnitude and / or increasing the phase.

Bandwidth

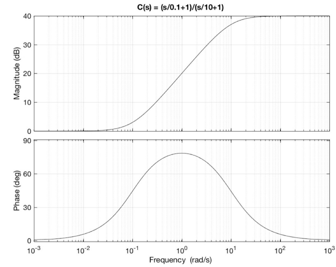
The Bandwidth is defined as the frequency for which
$$|T(j\omega)| > \frac{1}{\sqrt{2}} \quad \text{or} \quad |T(j\omega)|_{\text{dB}} > -3.01$$
respectively. The bandwidth is approximately the cross-over frequency of the open-loop system.
A higher gain and poles close to the origin increase the bandwidth. However, the stability might be lost, making the bandwidth irrelevant. Zeros close to the origin decrease the bandwidth
Careful: To go from $[\frac{\text{rad}}{\text{s}}]$ to [Hz] we have to divide by 2π.
Example: Maximum achievable Bandwidth of a system with proportional control and given phase margin condition: (1) Plug the phase margin φ_m into
$$\varphi_m = \varphi + 180^\circ$$
and solve for φ. (2) The frequency at the given phase, which can be read from the Bode plot is the maximum possible Bandwidth.

Compensators

Lead Compensators

A lead compensator is a loop element that increases the phase margin. However, it also makes a system more sensitive to noise.
$$\boxed{C_{\text{lag}} = \frac{s/a+1}{s/b+1} = \frac{b}{a}\cdot\frac{s+a}{s+b}, \quad 0 < a < b}$$
$|L(j\omega)|$ changes by $+20$dB/dec between a and b. The maximum increase of φ is:
$$\boxed{\Delta\varphi \approx 2\cdot\arctan(\sqrt{b/a}) - 90^\circ}$$
Design procedure:

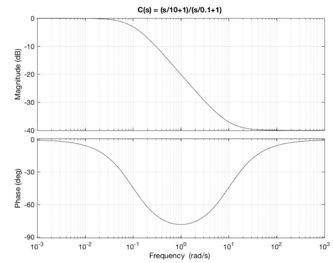
1. Pick the desired crossover frequency $\omega_c = \sqrt{ab}$
2. Pick b/a depending on the desired phase increase
3. Possibly add a proportional gain k to set ω_c back to the desired frequency ($|kL(jw)| = 1$)



Lag Compensators

A lag compensator is a loop element that improves disturbance rejection and command tracking by decreasing the sensitivity to noise. However, it also decreases the phase margin of the system and can hence make it unstable:
$$\boxed{C_{\text{lag}} = \frac{s/a+1}{s/b+1} = \frac{b}{a}\cdot\frac{s+a}{s+b}, \quad 0 < b < a}$$
$|L(j\omega)|$ changes by -20dB/dec between a and b. The maximum decrease of φ is:
$$\boxed{\Delta\varphi \approx 2\cdot\arctan(\sqrt{b/a}) - 90^\circ}$$
Careful: As the phase decreases, $\Delta\varphi$ is negative.
Design procedure:

1. Choose a/b as the desired increase in magnitude at low ω.
2. Pick $\sqrt{ab}$ as far as possible from the desired crossover frequency ω_c to not risk instability.

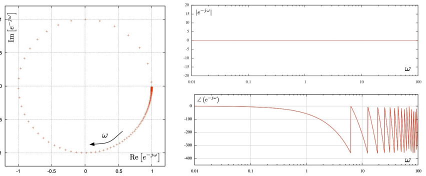


Remarks:

- Always check whether $0 < b < a$ is true.
- The maximum phase decrease is at $\omega = \sqrt{ab}$.

Time-Delays

A time delay is a linear operator with the transfer function $C(s) = e^{-\tau s}$ where τ is the time-delay. It causes the phase to oscillate but doesn't change the magnitude of the open-loop system.



A time-delay reduces the phase margin

$$\boxed{\varphi_{m,\text{time-delay}} = \varphi_m - \omega_c \cdot \tau}$$

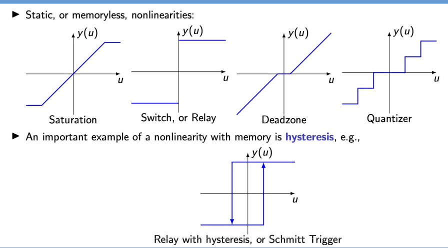
where ω_c is the cross-over frequency. It can hence destabilize the closed-loop system.
Since the transfer function of a pure time delay isn't a rational function, it's stability can't be analyzed using the Root Locus Method. However, we can approximate it with a rational function using the Padé approximation

$$\boxed{e^{-\tau s} \approx \frac{2/\tau - s}{2/\tau + s}}$$

When a time delay is introduced to a closed-loop system, the achievable bandwidth decreases (performance limitation).

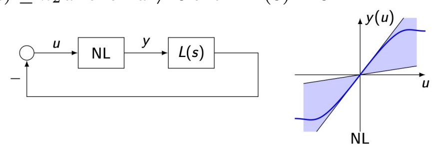
Non-Linear Systems

Common nonlinear elements



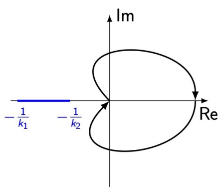
Necessary and sufficient condition

We consider the feedback interconnection of a linear system $L(s)$ with a static nonlinear gain element NL, such that $k_1 u \leq NL(u) \leq k_2 u$ for all $u \neq 0$ and $NL(0) = 0$.

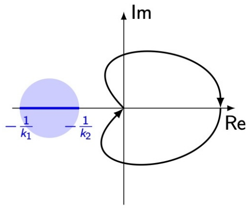


The system is absolutely stable if, for any choice of the nonlinearity NL, $u = 0$ is a globally asymptotically stable equilibrium point for the closed loop system.
We can check for absolute stability using the following conditions:

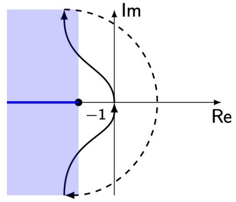
- Necessary condition: The Nyquist plot encircles the **segment** $[-1/k_1, -1/k_2]$ counterclockwise a number of times equal to the number unstable poles of $L(s)$.



- Sufficient condition (Circle criterion): The Nyquist plot encircles the **circle** with diameter $[-1/k_1, -1/k_2]$ a number of times equal to the number unstable poles of $L(s)$.



The necessary condition must hold if the system is absolutely stable, but it doesn't guarantee stability. The sufficient condition guarantees absolute stability if satisfied and is therefore stronger. If the Nyquist plot touches the circle, no conclusions can be made about the absolute stability.



Describing functions

When applying a sinusoidal input signal $u(t) = A\sin(\omega t)$ to a static nonlinearity NL, the output will be of the form

$$y(t) = f(A\sin(\omega t))$$

which is some periodic function with the same frequency as the input. Due to its periodicity, we can compute its Fourier series

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)]$$

where

$$a_n = \frac{1}{\pi}\int_{-\infty}^{\infty} y(t)\cos(n\omega t)d(\omega t)$$

and

$$b_n = \frac{1}{\pi}\int_{-\infty}^{\infty} y(t)\sin(n\omega t)d(\omega t)$$

Note that if $y(t)$ is odd, then $a_n = 0 \; \forall n \in \mathbb{N}$. As physical systems generally act as low-pass filters we can approximate the output by its first harmonic

$$y(t) \approx b_1 \sin(\omega t)$$

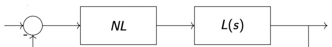
The describing function is then defined as follows

$$N(A) = \frac{b_1(A)}{A} = \frac{1}{\pi A}\int_{-\pi}^{\pi} y(t)\sin(n\omega t)d(\omega t)$$

and is used to approximate the nonlinearity as an amplitude-dependent gain. Note that this only holds for odd, static nonlinearities. More in general, for an input $u(t) = Ae^{j\omega t}$, the describing function is a complex number defined as an approximate transfer function

$$N(A, w) = \frac{c_1(A,\omega)}{A} e^{j\phi_1(A,\omega)}$$

We can use the describing function to analyze the stability of the feedback interconnection of a linear system with a non-linear system.

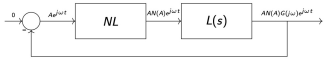


The new transfer function is given by

$$L'(A, s) \approx N(A)L(s)$$

Limit Cycles

For an input to the nonlinearity of the form $Ae^{j\omega t}$



we can observe that the oscillation is self sustained if $A = -AN(A)L(j\omega)$, i.e. if

$$-\frac{1}{N(A)} = G(j\omega)$$

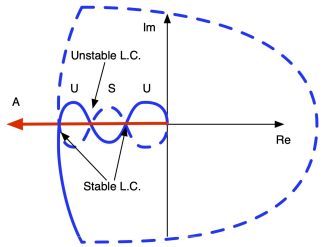
as the input $Ae^{j\omega t}$ produces an output $-Ae^{j\omega t}$ that is negatively fed back. This causes the system to oscillate indefinitely, causing a so-called limit cycle. When checking for Limit cycles, we can apply the following procedure:

1. Sketch the polar plot of the frequency response $L(j\omega)$.
2. Sketch the polar plot of $-\frac{1}{N(A)}$.
3. Limit cycles exist at the intersection between the two curves.

To analyze stability we use the $-\frac{1}{N(A)}$ point as the -1 (or $-\frac{1}{k}$) point in the Nyquist analysis:

- If the $-\frac{1}{N(A)}$ point is in an unstable region in the Nyquist plot, the amplitude of the oscillations will tend to increase.
- If the $-\frac{1}{N(A)}$ point is in a stable region in the Nyquist plot, the amplitude of the oscillations will tend to decrease.

If small amplitude decreases (from left to right) will make the Nyquist plot increase, the limit cycle is stable. Otherwise it is unstable.



In other words: 1) We determine the number of unstable poles and compute the number of (CCW) encirclements needed for the system to become stable. 2) We then apply the following rule to asses each limit cycle: If after an increase of the amplitude A the point $-\frac{1}{N(A)}$ remains in a stable region and after a decrease it moves into the unstable region, then the limit cycle is stable.

Appendix

Geometry

The phase of an imaginary number (measured CCW ↺ from the real axis) is measured as follows:

$$\angle(a+jb) = \arctan\left(\frac{b}{a}\right) + \begin{cases} \pi & a<0,\; b\geq 0 \\ -\pi & a<0,\; b<0 \\ 0 & \end{cases}$$

where we can use the following table to evaluate the arctangent:

θ	0	30° $\frac{\pi}{6}$	45° $\frac{\pi}{4}$	60° $\frac{\pi}{3}$	90° $\frac{\pi}{2}$
$\tan(\theta)$	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	undef.
$\sin(\theta)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

Furthermore the following properties can be helpful when computing the angle:

$$\arctan(-x) = -\arctan(x)$$

$$\arctan(x) + \arctan\left(\frac{1}{x}\right) = \frac{\pi}{2}$$

Linear Algebra

The inverse of a 2×2 matrix A is computed as follows:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The exponential of a real square matrix A is defined via $\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!}A^k$ where A^0 is the identity matrix.
If A is in diagonal form, e^{At} can be simplified to

$$\exp\left(\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} t\right) = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}$$

If A is not diagonal but diagonalizable, we determine the eigenvalues of A and compute e^{At} as above.
If A is in Jordan form, e^{At} can be simplified to

$$\exp\left(\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} t\right) = e^{\lambda t}\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

$$\exp\left(\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} t\right) = e^{\lambda t}\begin{bmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

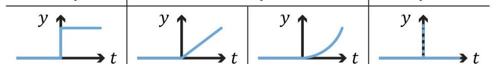
dB-Tabelle

Decibels $dB = 20\log_{10}(x)$ are converted to decimals as follows:

Decimal	Decibel	Decimal	Decibel
100	40	$\frac{1}{\sqrt{2}}$	-3.01
10	20	0.1	-20
5	13.97	0.01	-40
2	6.02	0	$-\infty$
1	0	-	-

Signals and Responses

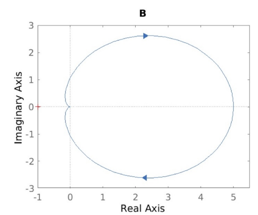
The most commonly used reference signals are



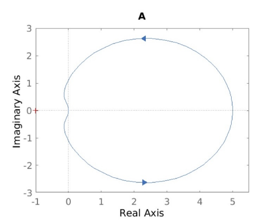
Step	Ramp		Impulse
$r(t) = h(t)$	$t \cdot h(t)$	$t^2/_2 \cdot h(t)$	$\delta(t)$
$R(s) = 1/_s$	$1/_{s^2}$	$1/_{s^3}$	1

where $R(s)$ is the Laplace transform of $r(t)$.

Nyquist plot net phase changes: Further examples



-180°



$+90^\circ$