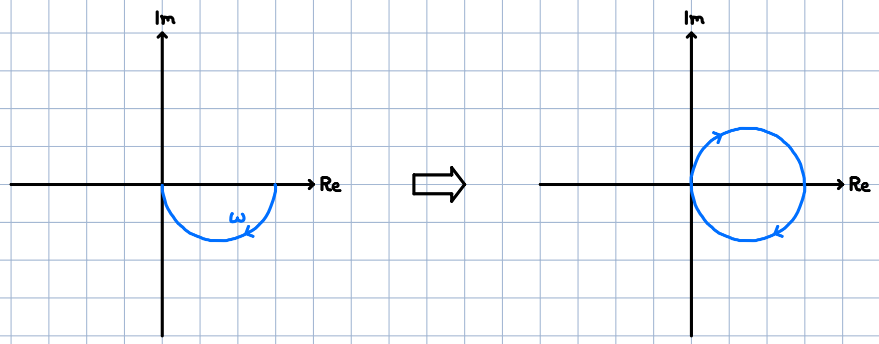
Control Systems I nbartzsch

11. Recital 29.11.24

Recap

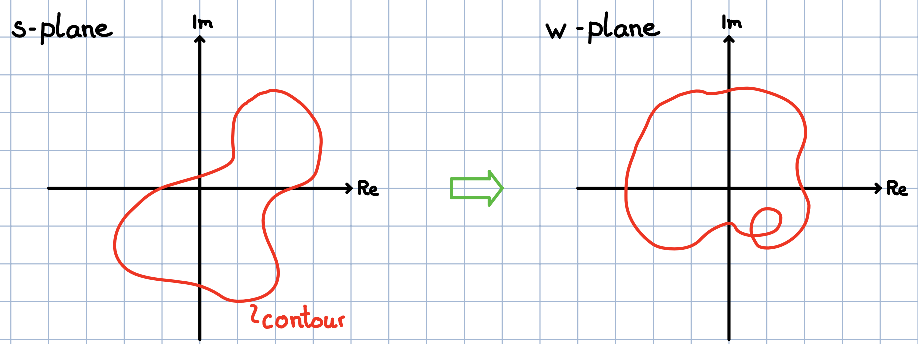
Nyquist Plot:

To draw this plot, just take the polar plot and mirror it along the real axis!



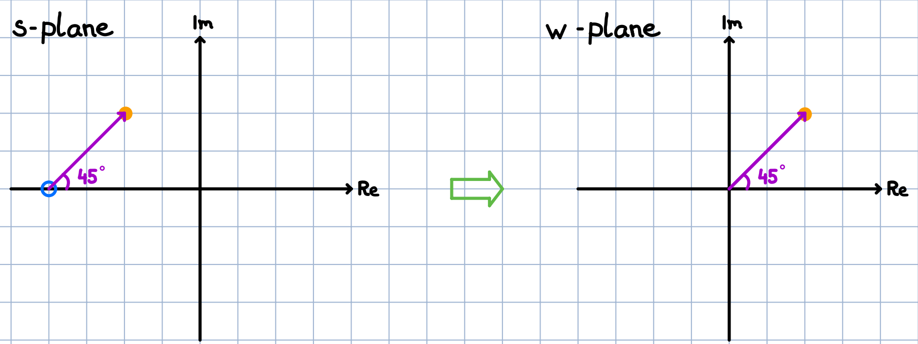
Principle of variation of the argument (Cauchy's Argument Principle)

Given a TF $G(s)$, we can plug in different values of s and get another complex number. We can plot this new complex number in a plane and call it the w-plane. We can now say that our TF $G(s)$ maps a point from the s-plane to the w-plane. We can extend this mapping and map entire lines from s to w. We can also look at how closed curves, so called contours, map from s to w.

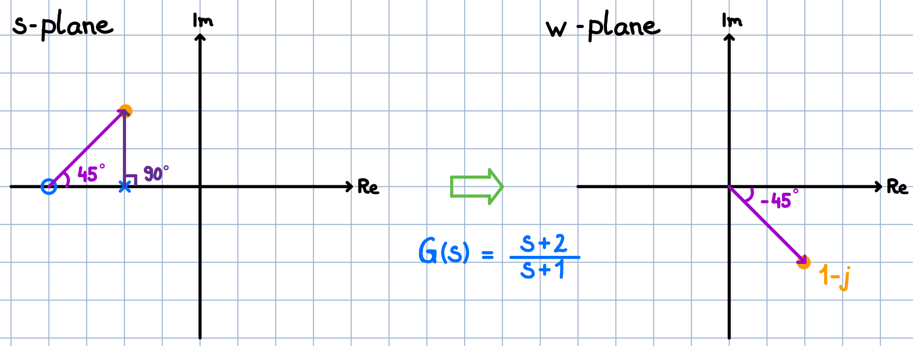


a closed curve in the s-plane will also result in a closed curve in the w-plane. This closed curve in the w-plane now also includes informations about the TF we used to map from s to w.

Consider a simple example

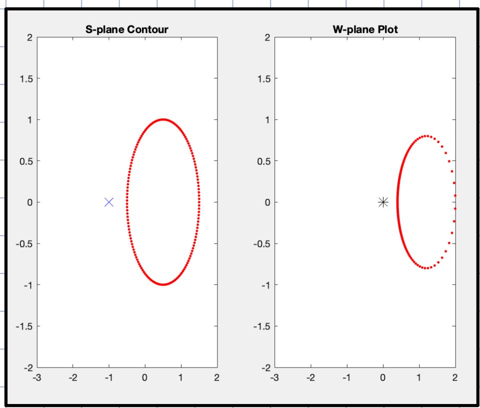


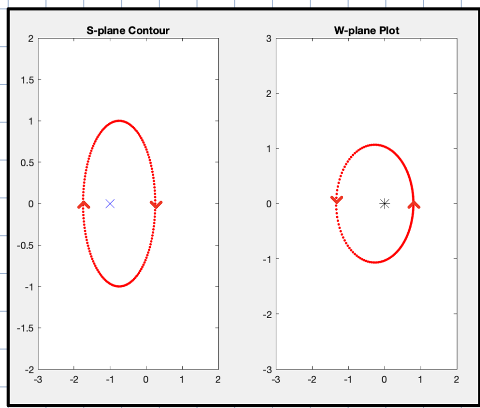
We observe that the phasor of the point in the w-plane is the same as the one from the zero to the original point in the s-plane. Let's add in a pole and see what happens:

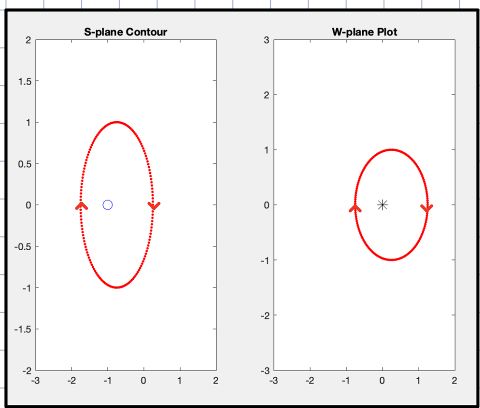


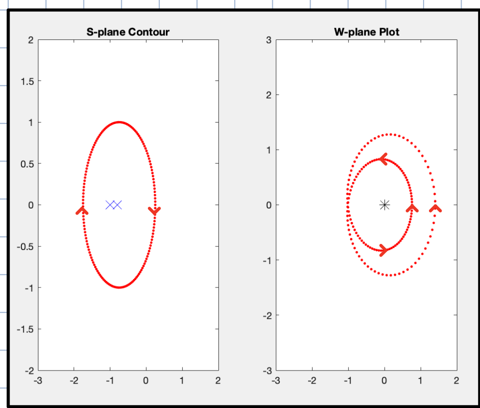
The general rule is: Add phases of zeros subtract phases of poles.

Let's look at some different contours.





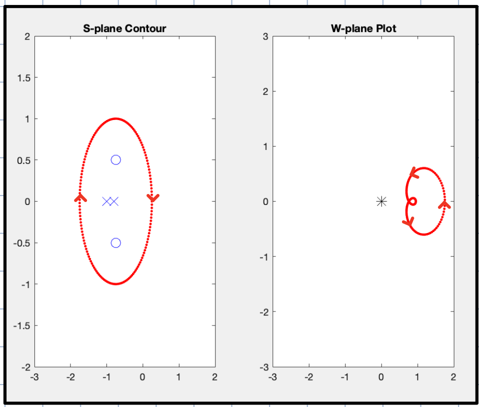


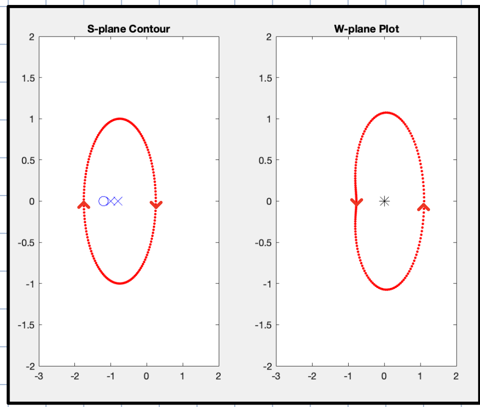


What we can see is that for every time we encircle a pole / zero in the s-plane we also encircle the origin in the w-plane. For each clockwise encirclement of zeros we get one CW encirclement of the origin, and for each CW encirclement of a pole we get one CCW encirclement of the origin.

Think of a zero as adding 360° and a pole as subtracting 360°.

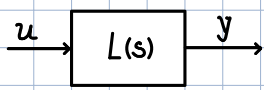
If we have as many poles as zeros we end up with no encirclements, and if we have one more pole than zeros, we end up with one CCW encirclement of the origin.





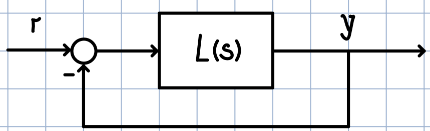
This means that you can tell the relative difference of poles and zeros inside a contour by how many times the plot circles the origin and in which direction.

Let's get back to the Nyquist plot. Recall that for an open loop system,



we can check wether the system is stable by looking at the poles of $L(s)$. If there are any poles in the RHP the system is unstable.

In closed-loop systems:

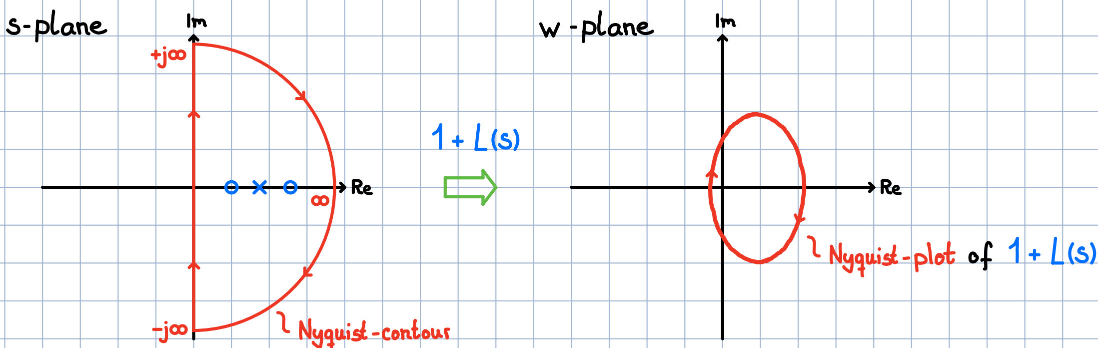


we now have to look at the poles of:

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

i.e. all points where $1 + L(s) = 0$. To asses the closed loop stability we have to check if any of the the zeros of $1 + L(s)$ are in the RHP.

To do this we can use the principle of variation of the argument introduced above. We can choose our contour to encircle the entire RHP:



we now know:

$$\boxed{\#\text{ Encirclements of } 0 \text{ using } 1+L(s) = \#\text{ Zeros of } 1+L(s) \text{ in RHP} - \#\text{ Poles of } 1+L(s) \text{ in RHP}}$$

We can make use of these characteristics:

→ Instead of counting the encirclements of $1+L(s)$ arround zero, we can shift the coordinate system by -1 and count the encirclements of $L(s)$ arround -1.

→ If $L(s)=\frac{N(s)}{D(s)}$, then the poles of $L(s)$ are given by D(s). We can also re-write $1+L(s)$:

$$1+L(s) = 1+\frac{N(s)}{D(s)} = \frac{D(s)+N(s)}{D(s)}$$, the poles are also given by D(s)

Thus, $L(s)$ and $1+L(s)$ have the same poles

We can combine both and get:

$$\boxed{\#\text{ Encirclements of } -1 \text{ using } L(s) = \#\text{ Zeros of } 1+L(s) \text{ in RHP} - \#\text{ Poles of } L(s) \text{ in RHP}}$$

Which finally yields the Nyquist stability theorem:

$$\boxed{Z = N + P}$$

Zeros of $1+L(s)$ in RHP

= # unstable closed loop poles

#CW Encirclements of -1 using $L(s)$

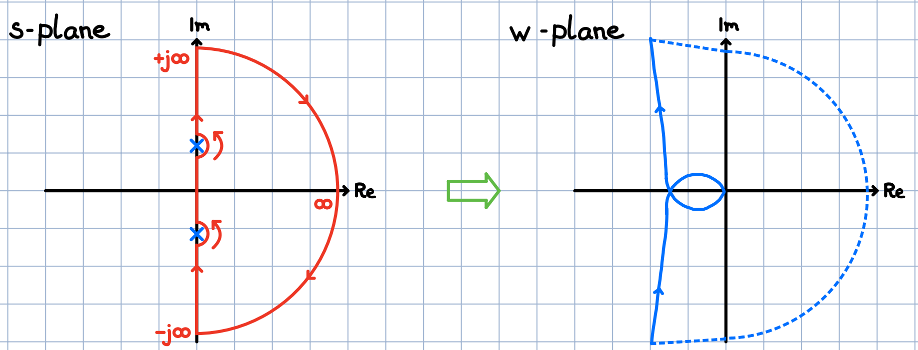
Poles of $L(s)$ in RHP

= # unstable open loop poles

We can now asses wether a closed loop system is stable, by only looking at the OL poles and the Nyquist-plot

<u>Special case:</u>

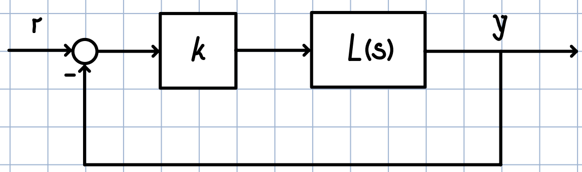
How do we treat poles and zeros of $L(s)$ on the imaginary axis?



We make little indents on the imaginary axis. If you move around the poles CCW, then you have to close the Nyquist-plot CW at infinity.

Note:

If your CL has some gain k:



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

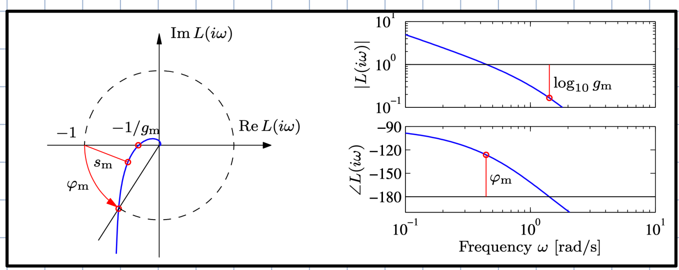
you have to count the #CW Encirclements of $-\frac{1}{k}$.

Stability Margins:

Next to stability of the CL system, the Nyquist-plot can also tell us how far away we are from being unstable.

Assume our OL to be stable, i.e. $P = 0$. For our CL system to be stable we now need $N = 0$. Then $Z = N + P = 0$.

We can now define a phase- and gain margin that tell us how "close" we are to encircling -1.



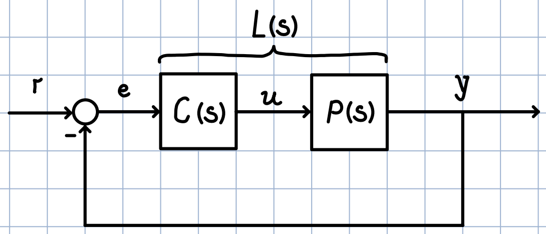
What does each margin mean?

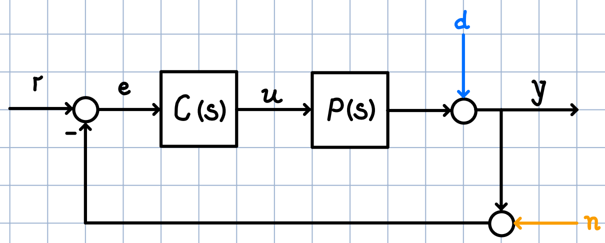
→ Gain Margin g_m: The point at 180°. It tells us how much we can scale until reaching -1.

→ Phase Margin φ_m: Point at magnitude 1. It tells us how much we can change the phase until reaching -1.

Frequency Domain Specifications

Similar to the time-domain specs. that resulted in feasible areas in the s-plane, we can define frequency domain specs, that dictate/shape how our Bode plot should look like. Recall the sensitivity/CL TFs we introduced a while back.





→ Open loop TF:

$$L(s) = C(s)P(s)$$

→ Complementary Sensitivity:

maps $r \to y$, $n \to y$

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

→ Sensitivity:

maps $r \to e$, $d \to y$

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

If we have disturbances d and/or noise n entering our CL system, we can use T and S to map the noise n and disturbances d to the output y. Usually disturbances have low frequencies and noise has high frequencies. The commands we input to our system usually also have a relativley low frequency. Knowing this we can constrain the magnitudes of $S(j\omega)$ and $T(j\omega)$ in the following way:

→ $|S(j\omega)| \ll 1$ at low frequencies for disturbance rejection and good command tracking.

→ $|T(j\omega)| \ll 1$ at high frequencies for noise rejection.

Remember: $S + T = 1 \;\; \forall \omega$. We can't make the sensitivity functions arbitrarily small over all frequencies.

As always, we would like to have these constraints as a function of the OL TF $L(s)$. In this case:

→ $|S(j\omega)| = \left|\frac{1}{1+L(j\omega)}\right| \ll 1 \iff L(s)$ has to be large at low frequencies

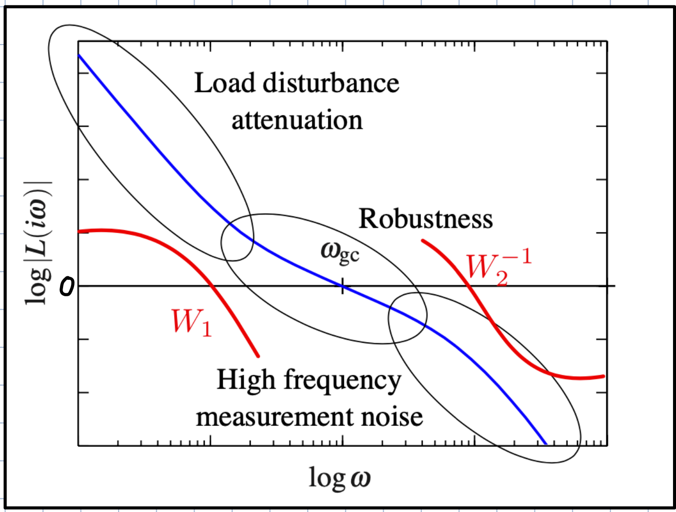
→ $|T(j\omega)| = \left|\frac{L(j\omega)}{1+L(j\omega)}\right| \ll 1 \iff L(s)$ has to be small at high frequencies

We can quantify how large or small we want $L(s)$ to be, with some function $W(j\omega)$. So we can write:

$$|S(j\omega)| \cdot |W_1(j\omega)| < 1 \iff |1+L(j\omega)| > |W_1(j\omega)| \overset{\text{approx.}}{\iff} |L(j\omega)| > |W_1(j\omega)|$$

$$|T(j\omega)| \cdot |W_2(j\omega)| < 1 \underset{T(j\omega)\approx L(j\omega)}{\overset{\text{approx.}}{\iff}} |L(j\omega)| \cdot |W_2(j\omega)| < 1 \iff |L(j\omega)| < |W_2(j\omega)|^{-1}$$

This results in the following "obstacle course" for the Bode plot of $L(j\omega)$.



Next to high- and low-frequency behavior we can also constrain the bandwidth of CL system. The bandwidth tells us the maximum frequency for which the output can track commands within a factor ≈ 0.7. In other words the bandwidth tells us for which max freq. we get satisfactory operation. ($|T(j\omega)| > 1/\sqrt{2}$)

We can usually approximate the CL bandwidth with the OL crossover frequency ω_{gc}.

But how do we apply these constraints to our $L(j\omega)$?

Loop Shaping

Now we get to design our controler. We know how the desired $L(s)$ should look like, and how to model the system or plant $P(s)$. Our controler should then be something like $C(s) = L(s)\,P(s)^{-1}$.

Usually we approach this problem with some basic building blocks to steer $L(s)$ through the Bode obstacle course.

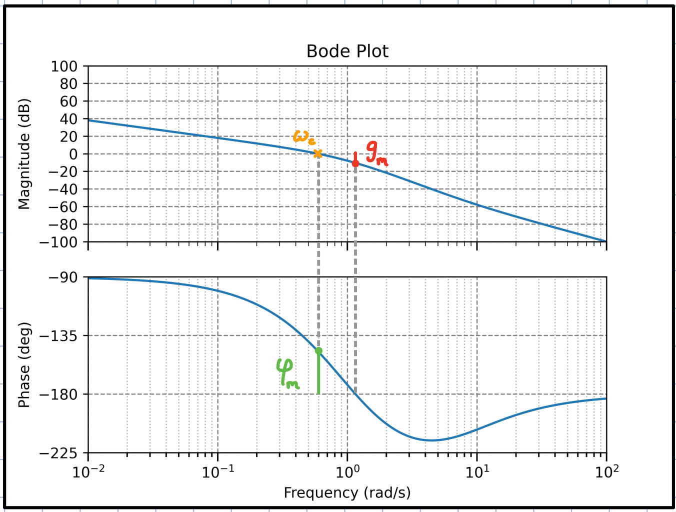
That way we construct a dynamic compensator $C(s)$ that fullfills our requirements.

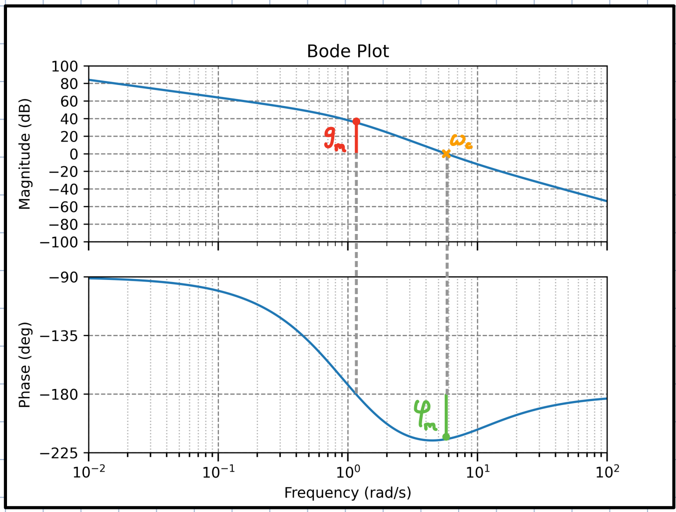
Proportional compensation:

In this case $C(s) = k$, where k is a simple gain.

- → Shifts the magnitude, while phase is unaffected.
- → For stable OL systems, small k $(k \to 0)$ yield stable CLs.
- → Improves command tracking (higher magnitude at low ω) and CL bandwidth (moves crossover freq. to right)
- → Stability can be compromized!!

$L(s)$ $\qquad\qquad$ $40 \cdot L(s)$



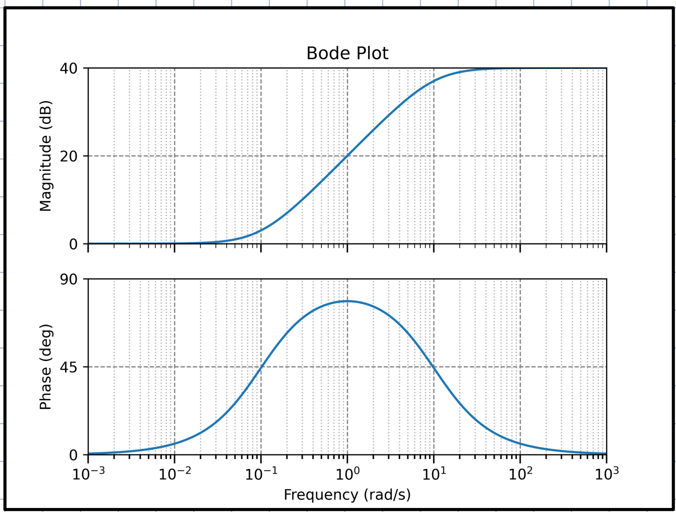


Lead compensator

A lead compensator is a pole zero pair. We can write it in a general form as:

$$C_{lead}(s) = \frac{\frac{s}{a}+1}{\frac{s}{b}+1} = \frac{b}{a}\,\frac{s+a}{s+b} \qquad 0<a<b$$

In a lead compensator the zero always comes before the pole. Graphically it can look like this:



Example for $a=0.1$, $b=10$, i.e. $C_{lead}(s) = \frac{10s+1}{0.1s+1} = 100\,\frac{s+0.1}{s+10}$

The main effects are:

→ Increase magnitude at high freq. by $\frac{b}{a}$, while low freq. are unafected.

→ Increase slope of magnitude between a and b by $20\ ^{dB}/_{dec}$

→ Increase phase around $\sqrt{ab}$ by up to 90°. The max phase increase is of

$$\varphi_{max} = 2\arctan\left[\sqrt{\frac{b}{a}}\right] - 90^\circ$$

Main use: increase phase margin. $C_{lead}(s) = k\,\frac{b}{a}\,\frac{s+a}{s+b}$ optional to adjust ω_c

i. Pick $\sqrt{ab}$ at desired ω_c

ii. Pick $\frac{b}{a}$ depending on desired phase increase.

iii. Adjust k to put ω_c at desired freq.

Side effects: increase magnitude at high freq.

Exam problems: HS22

Problem: Consider a lead compensator

$$C_{\text{lead}}(s) = \frac{s/a+1}{s/b+1}, \quad 0 < a < b, \quad (1)$$

used in a standard feedback architecture to control a given plant P.

Q37 (0.5 Points) Mark the correct answer for each statement.

Statement	True	False
A lead compensator is typically used to reduce the phase margin.		x
A lead compensator can increase the sensitivity to high-frequency noise.	x	
The slope of the magnitude at frequencies between a and b is approximately +20 dB/decade.	x	

Solution: See list above.

Q38 (1.5 Points) A lead compensator as given in Equation (1) is to be used in order to increase the phase of the open-loop system $P(s) \cdot C_{\text{lead}}(s)$ by at most 30° at a frequency of 10 rad/s.

Which choice of the parameters a and b satisfies the aforementioned requirement? Mark the correct answer.

- ☐ $a = \frac{10}{\sqrt{3}}$, $b = 10\sqrt{3}$
- ☐ $a = 1\sqrt{3}$, $b = 100\sqrt{3}$
- ☐ $a = 10\sqrt{3}$, $b = \frac{10}{\sqrt{3}}$
- ☐ $a = 100\sqrt{3}$, $b = 1\sqrt{3}$

Solution:

The max phase increase at $\sqrt{ab}$ is: $\varphi_{max} = 2\arctan\left[\sqrt{\frac{b}{a}}\right] - 90°$.

From this we get:

$\sqrt{ab} = 10\,\frac{rad}{s} \iff a = \frac{100}{b} \iff a^2 = \frac{100}{3} \iff \boxed{a = \frac{10}{\sqrt{3}}}$

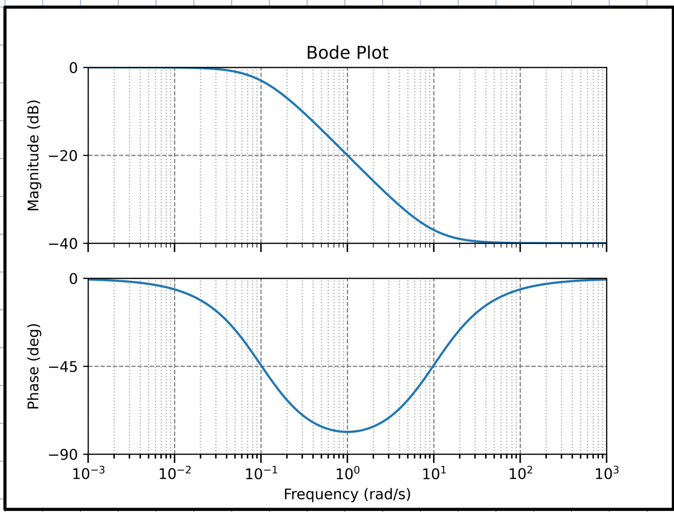
$\varphi_{max} = 30° = 2\arctan\left[\sqrt{\frac{b}{a}}\right] - 90° \iff b = 3a \quad \boxed{b = 3\,\frac{10}{\sqrt{3}} = 10\sqrt{3}}$

Lag compensator:

A lag compensator is also a pole zero pair. We can write it in a general form as:

$$C_{lag}(s) = \frac{\frac{s}{a}+1}{\frac{s}{b}+1} = \frac{b}{a}\,\frac{s+a}{s+b} \qquad 0<b<a$$

In a lag compensator the pole always comes before the zero. Graphically it can look like this:



Example for $a=10$, $b=0.1$, i.e. $C_{lag}(s) = \frac{0.1s+1}{10s+1} = \frac{1}{100}\frac{s+10}{s+0.1}$

The main effects are:

→ Decrease magnitude at high freq. by $\frac{b}{a}$, while low freq. are unafected.

→ Decrease slope of magnitude between a and b by $20\ ^{dB}/_{dec}$

→ Decrease phase around $\sqrt{ab}$ by up to $90°$. The max phase increase is of

$$\varphi_{max} = 2\arctan\left[\sqrt{\frac{b}{a}}\right] - 90°$$

Main use: improve command tracking / disturbance rejection. $C_{lag}(s) = k\,\frac{\frac{s}{a}+1}{\frac{s}{b}+1}$

i. Pick $\frac{a}{b}$ as the desired increase in magnitude of low freq.

ii. Multiply k by $\frac{a}{b}$ (high freq. not affected)

iii. Pick a to be sufficiently small not to affect ω_c

Side effects: reduction of phase margin.

Exam problems: HS22

Problem: Consider a lag compensator

$$C_{\text{lag}}(s) = \frac{a}{b} \cdot \frac{s/a + 1}{s/b + 1}, \quad 0 < b < a,$$

used in a standard feedback architecture to control a given plant P.

Q39 **(0.5 Points)** Mark the correct answer for each statement.

Statement	**True**	**False**
A lag compensator is frequently used to improve command tracking/disturbance rejection.	x	
A lag compensator cannot cause the system to become unstable.		x
The slope of the magnitude at frequencies between a and b is approximately -20 dB/decade.	x	

Solution: See list above.

Q40 **(1.5 Points)** A lag compensator is to be used in order to improve the steady-state error to a step response to less than 2%. In order to achieve this, the low-frequency magnitude of the open-loop transfer function needs to be increased by a factor of 5. The current control design already features a satisfying phase margin and a satisfying open-loop gain crossover frequency ω_{gc} currently at 100 rad/s.

Which choice of the parameters a and b below satisfy aforementioned steady-state error requirement and have least impact on the crossover frequency and phase margin? Mark the correct answer.

- ☐ $a = 15,\ b = 75$
- ☐ $a = 1,\ b = 5$
- ☐ $a = 1,\ b = 0.2$
- ☐ $a = 75,\ b = 15$

Solution:

Lag compensator: $0 < b < a$! Rule out i, ii

- choose a sufficiently small not to affect ω_c.

  $a = 75$ closer to 100. Option iii the best

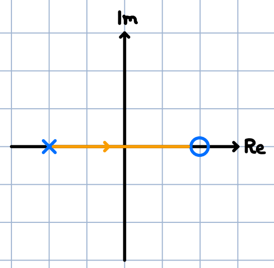
- $\frac{a}{b} = 5$, fullfilled by iii.

Limitations:

The tools from above work well for stable, minimum phase systems. But what if we have poles and zeros in the RHP?

Loop shaping for non-minimum-phase systems:

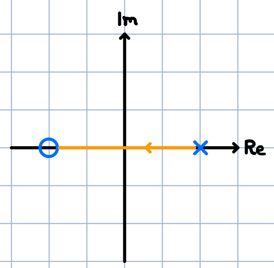
Remember from the root locus that closed loop poles approach open loop zeros. If one of the zeros is non-minimum-phase, the closed loop system might become unstable.



This means for large enough gains the closed loop will become unstable. A n.m.p. zero also limits the maximum crossover freq.. The consequence is that the CL system becomes slow.

Loop shaping for open-loop unstable systems:

We can look at the root locus again to see what happens when we have a CL pole in the RHP.



We need a high gain to stabalize the system. In real Life a high gain means fast and strong actuators. In this case the crossover freq. becomes large.

This should give you a sense of how, for many systems, there are clear performance limitations. Certain requirements can thus never be satisfied. E.g. have good noise and disturbance rejection whilst also having good command tracking.

Time Delays

After choosing our controler $C(s)$, we have to implement it. Usually we use computers for this task. Unfortunatley, computers have a finite compute time, which means that the control input to a certain error has some delay. Some physical systems themselves also have delays. An extreme example would be communication with e.g. a mars rover.

In this example the delay can be several minutes. How can we take this into account?

Mathematically we can express a time delay as:

$$y(t) = u(t-T)$$

The time delay is a linear operator that transforms an input $u(t)$ into a delayed output $y(t)$. T is the ammount of delay.

We can also compute the TF of a time delay. Assume $u(t) = e^{st}$.

$$y(t) = e^{s(t-T)} = e^{-sT} u(t)$$

The TF of a time delay is thus given by: e^{-sT}

This not a rational function! We have no poles or zeros! (Root locus doesn't work)