Introduction to Control (00340040)
lecture no. 10

Leonid Mirkin

Faculty of Mechanical Engineering
Technion—IIT

Outline

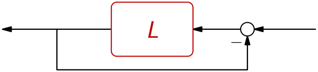
Stability margins

Time-delay systems

Delay margin

Outline

Loop shaping (contd)



- . . .
- having appropriate crossover frequency, ω_c
- high loop gain, $|L(j\omega)| \gg 1$, at low frequencies ($\omega \ll \omega_c$)
- low loop gain, $|L(j\omega)| \ll 1$, at high frequencies ($\omega \gg \omega_c$)
- keeping $L(j\omega)$ "far" from the critical point in the crossover region

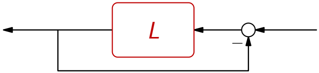
There is a need to quantify / measure this "far" requirement . . .

Such a measure should

1. reflect motivations for "far from the critical point" requirement and
2. be easily computable

Toward this end, the notions of stability margins introduced

Loop shaping (contd)



- plot of $L(\mathrm{j}\omega)$ agrees with the Nyquist stability criterion
- having appropriate crossover frequency, ω_c
- high loop gain, $|L(\mathrm{j}\omega)| \gg 1$, at low frequencies ($\omega \ll \omega_c$)
- low loop gain, $|L(\mathrm{j}\omega)| \ll 1$, at high frequencies ($\omega \gg \omega_c$)
- keeping $L(\mathrm{j}\omega)$ "far" from the critical point in the crossover region

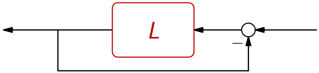
There is a need to quantify / measure this "far" requirement . . .

Such a measure should

1. reflect motivations for "far from the critical point" requirement and
2. be easily computable

Toward this end, the notions of stability margins introduced

Loop shaping (contd)



- plot of $L(j\omega)$ agrees with the Nyquist stability criterion
- having appropriate crossover frequency, ω_c
- high loop gain, $|L(j\omega)| \gg 1$, at low frequencies ($\omega \ll \omega_c$)
- low loop gain, $|L(j\omega)| \ll 1$, at high frequencies ($\omega \gg \omega_c$)
- keeping $L(j\omega)$ "far" from the critical point in the crossover region

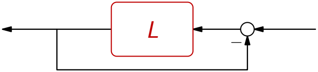
There is a need to quantify / measure this "far" requirement . . .

Such a measure should

1. reflect motivations for "far from the critical point" requirement and
2. be easily computable

Toward this end, the notions of stability margins introduced

Loop shaping (contd)



- plot of $L(j\omega)$ agrees with the Nyquist stability criterion
- having appropriate crossover frequency, ω_c
- high loop gain, $|L(j\omega)| \gg 1$, at low frequencies ($\omega \ll \omega_c$)
- low loop gain, $|L(j\omega)| \ll 1$, at high frequencies ($\omega \gg \omega_c$)
- keeping $L(j\omega)$ “far” from the critical point in the crossover region

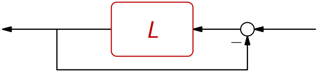
There is a need to quantify / measure this “far” requirement . . .

Such a measure should

1. reflect motivations for “far from the critical point” requirement and
2. be easily computable

Toward this end, the notions of stability margins introduced.

Gain and phase margins: definitions



Provided the closed-loop system is stable:

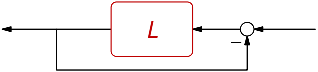
Gain margin (μ_{g}) is the *minimal* factor by which the loop gain should be changed to render the closed-loop system unstable
(typically, $\mu_{\mathrm{g}} > 1$, although it might also be contractive)

Phase margin (μ_{ph}) is the *minimal* amount by which the loop phase should be changed to render the closed-loop system unstable
($\mu_{\mathrm{ph}} > 0$ if a phase lag leads to instability and $\mu_{\mathrm{ph}} < 0$ if a phase lead)

In both cases, the minimal gain / phase change should result in

– Nyquist plot crossing the critical point.

Gain and phase margins: definitions



Provided the closed-loop system is stable:

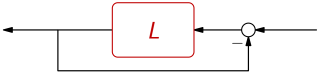
Gain margin (μ_{g}) is the *minimal* factor by which the loop gain should be changed to render the closed-loop system unstable
(typically, $\mu_{\mathrm{g}} > 1$, although it might also be contractive)

Phase margin (μ_{ph}) is the *minimal* amount by which the loop phase should be changed to render the closed-loop system unstable
($\mu_{\mathrm{ph}} > 0$ if a phase lag leads to instability and $\mu_{\mathrm{ph}} < 0$ if a phase lead)

In both cases, the minimal gain / phase change should result in

– Nyquist plot crossing the critical point.

Gain and phase margins: definitions



Provided the closed-loop system is stable:

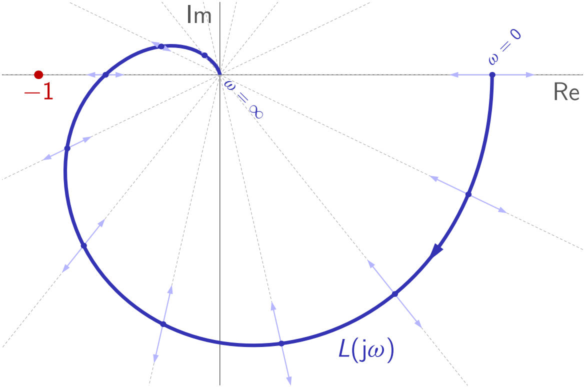
Gain margin (μ_{g}) is the *minimal* factor by which the loop gain should be changed to render the closed-loop system unstable
(typically, $\mu_{\mathrm{g}} > 1$, although it might also be contractive)

Phase margin (μ_{ph}) is the *minimal* amount by which the loop phase should be changed to render the closed-loop system unstable
($\mu_{\mathrm{ph}} > 0$ if a phase lag leads to instability and $\mu_{\mathrm{ph}} < 0$ if a phase lead)

In both cases, the minimal gain / phase change should result in

– Nyquist plot crossing the critical point.

What happens when only the magnitude changes?



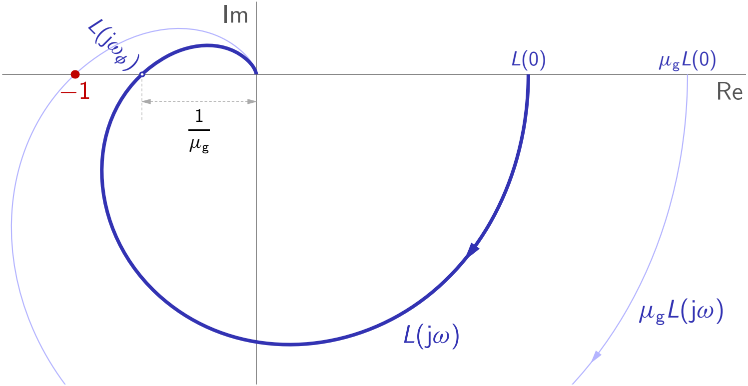
- when the loop static gain increases, the polar plot is inflated
- when the loop static gain decreases, the polar plot is deflated

Gain margin on Nyquist plot

From μ_g viewpoint, we are only interested in

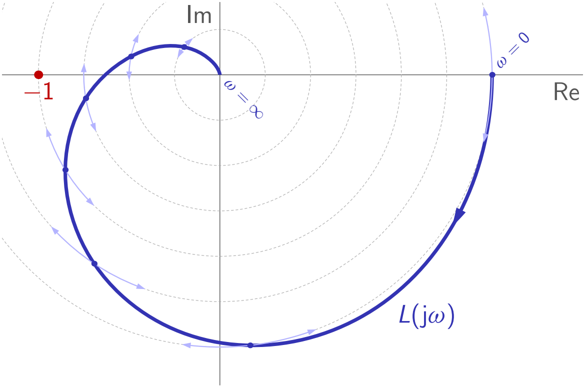
– points where the Nyquist plot crosses the negative real semi-axis

i.e. where[1] $\arg L(j\omega) \equiv -\pi \pmod{2\pi}$:



[1] The frequencies at which this happens called phase crossover frequencies, ω_ϕ.

What happens when only the phase changes ?



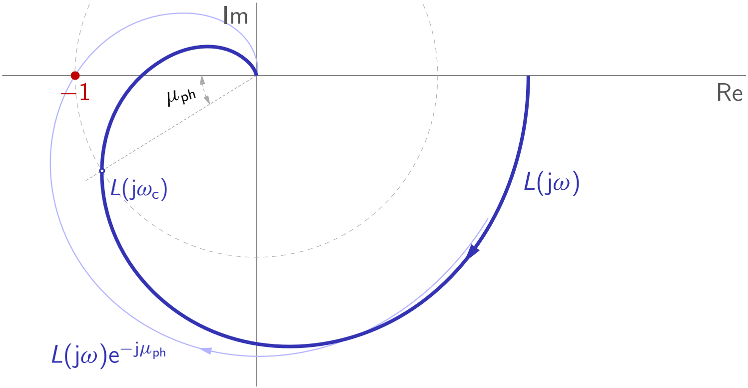
- when the phase increases, the polar plot pivots counterclockwise
- when the phase decreases, the polar plot pivots clockwise

Phase margin on Nyquist plot

From μ_{ph} viewpoint, we are only interested in

- points where the Nyquist plot crosses the unit circle

i.e. where[2] $|L(\mathrm{j}\omega)| = 1$:



[2] As we know, the frequencies at which this happens called crossover frequencies, ω_{c}.

Gain and phase margins on Bode diagram

- μ_{g} calculated from the magnitude plot at ω_ϕ

  if the closed-loop system stable, then μ_{g} equals the absolute value of the magnitude of $L(\mathrm{j}\omega_\phi)$ in dB, i.e. the distance from $|L(\mathrm{j}\omega_\phi)|$ to 0 dB
- μ_{ph} calculated from the phase plot at ω_{c}

  if the closed-loop system is stable, then μ_{ph} equals the distance between $\arg L(\mathrm{j}\omega_{\mathrm{c}})$ and the closest $-\pi + 2\pi k$, $k \in \mathbb{Z}$, point

For example (here $\mu_{\mathrm{g}} = 2 \approx 6.021\,\mathrm{dB}$ and $\mu_{\mathrm{ph}} = 57.9^\circ \approx 0.32\,\pi$ (rad)):

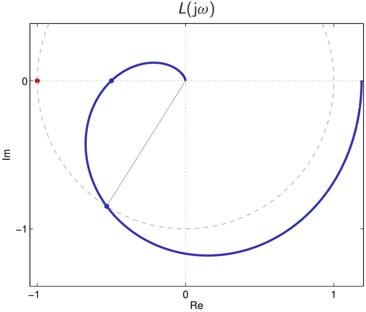
Gain and phase margins on Bode diagram

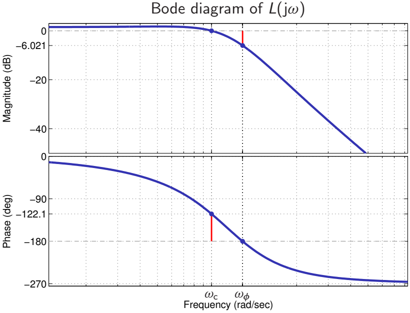
- μ_g calculated from the magnitude plot at ω_ϕ

  if the closed-loop system stable, then μ_g equals the absolute value of the magnitude of $L(j\omega_\phi)$ in dB, i.e. the distance from $|L(j\omega_\phi)|$ to 0 dB
- μ_{ph} calculated from the phase plot at ω_c

  if the closed-loop system is stable, then μ_{ph} equals the distance between $\arg L(j\omega_c)$ and the closest $-\pi + 2\pi k$, $k \in \mathbb{Z}$, point

For example (here $\mu_g = 2 \approx 6.021\,\text{dB}$ and $\mu_{ph} = 57.9^\circ \approx 0.32\,\pi$ (rad)):





μ_{g} and μ_{ph}: distance from the critical point

From what we learned,

- μ_{g} is the distance from the critical point along the real axis
- μ_{ph} is the angular distance from the critical point

Thus, the "far from the critical point" requirement may be translated to the "sufficiently large stability margins" requirement.

Rules of thumb for adequate gain and phase margins:

- $\mu_{\mathrm{g}} \approx 6 \div 12\,\mathrm{dB}$ and $\mu_{\mathrm{ph}} \approx 45 \div 60^\circ$.

μ_g and μ_{ph}: distance from the critical point

From what we learned,

- μ_g is the distance from the critical point along the real axis
- μ_{ph} is the angular distance from the critical point

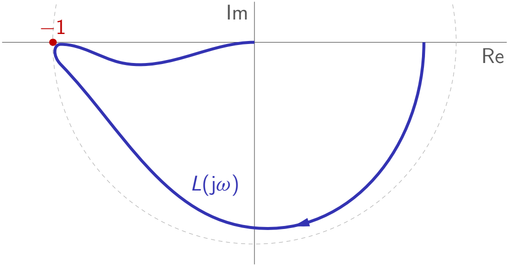
Thus, the "far from the critical point" requirement may be translated to the "sufficiently large stability margins" requirement.

Rules of thumb for adequate gain and phase margins:

- $\mu_g \approx 6 \div 12\,\text{dB}$ and $\mu_{ph} \approx 45 \div 60^\circ$.

Stability margins *cum grano salis*

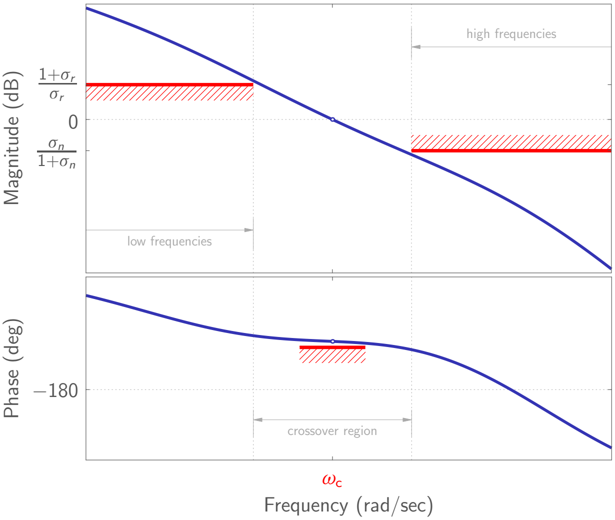
One should not take stability margins too seriously though. Neither μ_g nor μ_{ph}, nor even both of them, reflects the distance from the Nyquist plot of L to the critical point comprehensively. For example, if $L(j\omega)$ looks like



then

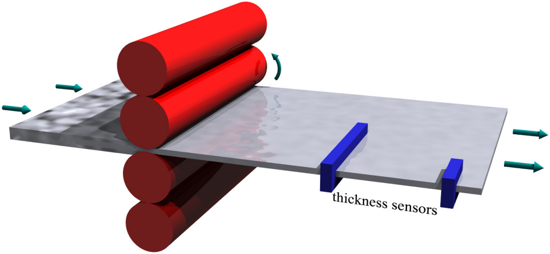
- $\mu_g = \infty$ and $\mu_{ph} = \infty$, yet $L(j\omega)$ is very close to the critical point.

Loop shaping: big picture



Outline

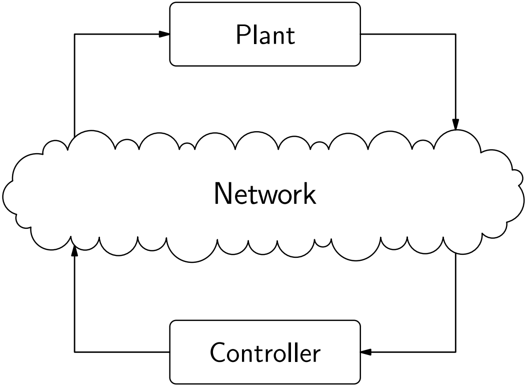
Hot strip mill profile control



Thickness can only be measured at some distance from rolls, leading to

- measurement delays

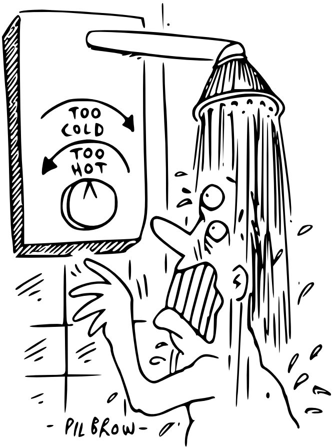
Networked control



Sampling, encoding, transmission, decoding need time. This gives rise to

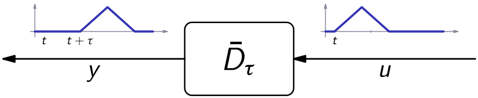
- measurement delays
- actuation delays

Temperature control



Everybody experienced this, I guess. . .

Delay element in time domain



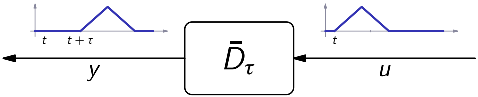
I/O relation:

$$y = \bar{D}_\tau u \iff y(t) = u(t - \tau)$$

This system is

- linear
  $\bar{D}_\tau(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 u_1(t-\tau) + \alpha_2 u_2(t-\tau) = \alpha_1(\bar{D}_\tau u_1) + \alpha_2(\bar{D}_\tau u_2)$
- time invariant
  $\bar{D}_{\tau_1}(\mathbb{S}_{\tau_2} u) = u(t - \tau_2 - \tau_1) = \mathbb{S}_{\tau_2}(\bar{D}_{\tau_1} u)$
- BIBO stable
  $\|y\|_\infty = \|u\|_\infty$ for all $u \in L_\infty$

Delay element in time domain



I/O relation:

$$y = \bar{D}_\tau u \iff y(t) = u(t - \tau)$$

This system is

- linear

  $\bar{D}_\tau(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 u_1(t - \tau) + \alpha_2 u_2(t - \tau) = \alpha_1(\bar{D}_\tau u_1) + \alpha_2(\bar{D}_\tau u_2)$

- time invariant

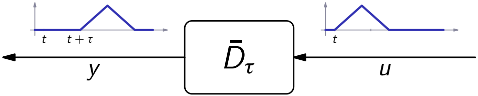
  $\bar{D}_{\tau_1}(\mathbb{S}_{\tau_2} u) = u(t - \tau_2 - \tau_1) = \mathbb{S}_{\tau_2}(\bar{D}_{\tau_1} u)$

- BIBO stable[3]

  $\|y\|_\infty = \|u\|_\infty$ for all $u \in L_\infty$

[3] $L_\infty := \{x : \mathbb{R} \to \mathbb{R} \mid \|x\|_\infty < \infty\}$, where $\|x\|_\infty := \sup_{t \in \mathbb{R}} |x(t)|$

Delay element in s-domain



By the time shift property of the Laplace transform:

$$y(t) = u(t-\tau) \iff Y(s) = \mathrm{e}^{-\tau s} U(s)$$

Therefore, the delay element has the transfer function

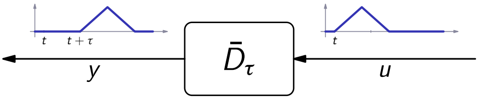
$$\bar{D}_\tau(s) = \mathrm{e}^{-\tau s}.$$

This transfer function is

– irrational,

so $\bar{D}_\tau$ is an infinite-dimensional system.

Delay element in s-domain



By the time shift property of the Laplace transform:

$$y(t) = u(t - \tau) \iff Y(s) = \mathrm{e}^{-\tau s} U(s)$$

Therefore, the delay element has the transfer function

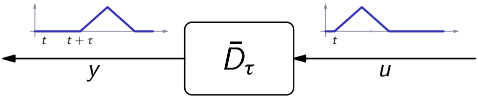
$$\bar{D}_\tau(s) = \mathrm{e}^{-\tau s}.$$

This transfer function is

– irrational,

so $\bar{D}_\tau$ is an infinite-dimensional system.

Delay element in s-domain



By the time shift property of the Laplace transform:

$$y(t) = u(t - \tau) \iff Y(s) = \mathrm{e}^{-\tau s} U(s)$$

Therefore, the delay element has the transfer function

$$\bar{D}_\tau(s) = \mathrm{e}^{-\tau s}.$$

This transfer function is

- irrational,

so $\bar{D}_\tau$ is an infinite-dimensional system.

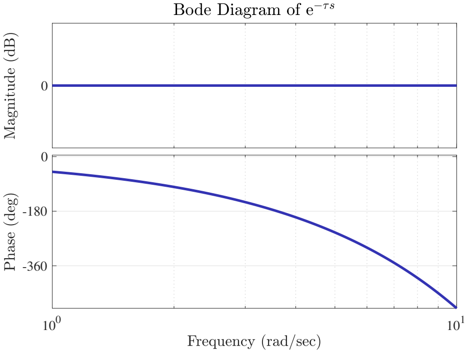
Frequency response

Obtained as

$$\mathrm{e}^{-\tau s}\big|_{s=\mathrm{j}\omega} = \mathrm{e}^{-\mathrm{j}\tau\omega} = \cos(\tau\omega) - \mathrm{j}\sin(\tau\omega)$$

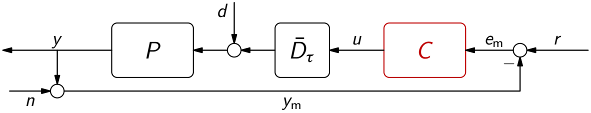
It has

- unit magnitude ($|\mathrm{e}^{-\mathrm{j}\tau\omega}| \equiv 1$) and
- linearly decaying phase ($\arg \mathrm{e}^{-\mathrm{j}\tau\omega} = -\tau\omega$, in radians if ω is in rad/sec)



Dead-time systems

Systems with loop delays:

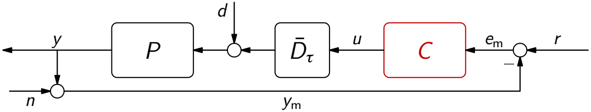


where

P is a plant with a rational transfer function $P(s)$

C is a controller with a (rational) transfer function $C(s)$

Effect of loop delay on characteristic polynomial



Let $P(s) = \frac{N_P(s)}{D_P(s)}$ and $C(s) = \frac{N_C(s)}{D_C(s)}$. Then

$$\chi_\mathrm{cl}(s) = \mathrm{e}^{-\tau s} N_P(s) N_C(s) + D_P(s) D_C(s)$$

has infinitely many roots (known as quasi-polynomial).

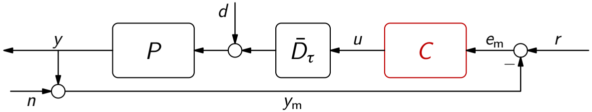
Example

If $P(s) = 1$ and $C(s) = k_\mathrm{p}$, then

$$\chi_\mathrm{cl}(s) = k_\mathrm{p} \mathrm{e}^{-\tau s} + 1 \quad \text{has roots at } s = \frac{\ln k_\mathrm{p}}{\tau} + \mathrm{j}\frac{\pi + 2\pi i}{\tau}$$

for all $i \in \mathbb{Z}$.

Effect of loop delay on characteristic polynomial



Let $P(s) = \frac{N_P(s)}{D_P(s)}$ and $C(s) = \frac{N_C(s)}{D_C(s)}$. Then

$$\chi_{cl}(s) = e^{-\tau s} N_P(s) N_C(s) + D_P(s) D_C(s)$$

has infinitely many roots (known as quasi-polynomial).

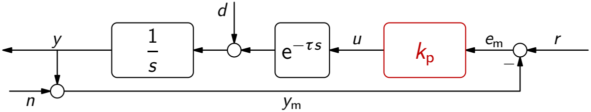
Example

If $P(s) = 1$ and $C(s) = k_p$, then

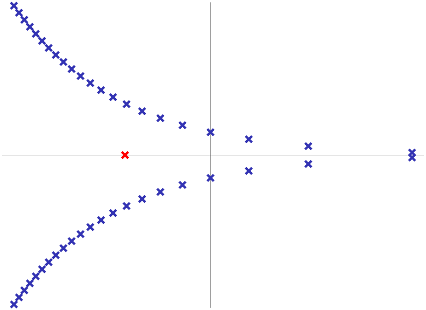
$$\chi_{cl}(s) = k_p e^{-\tau s} + 1 \quad \text{has roots at } s = \frac{\ln k_p}{\tau} + j\frac{\pi + 2\pi i}{\tau}$$

for all $i \in \mathbb{Z}$.

Another scary example

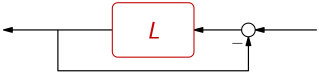


Comparing delay-free ($\tau = 0$) and delayed ($\tau > 0$) closed-loop poles,



we can see that the addition of a delay considerably complicates matters.

Effect of loop delay on $L(j\omega)$



Let $L(s) = L_r(s)e^{-\tau s}$ for some rational $L_r(s)$. In this case

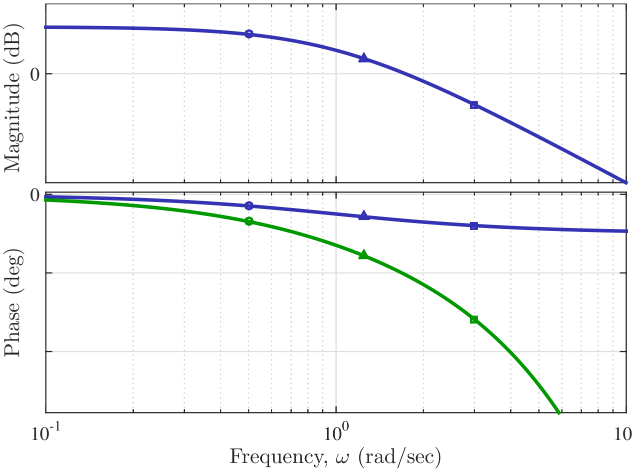
$$L(j\omega) = L_r(j\omega)e^{-j\tau\omega}$$
$$\Downarrow$$
$$|L(j\omega)| = |L_r(j\omega)| \quad \text{and} \quad \arg L(j\omega) = \arg L_r(j\omega) - \tau\omega.$$

In other words, delay in this case

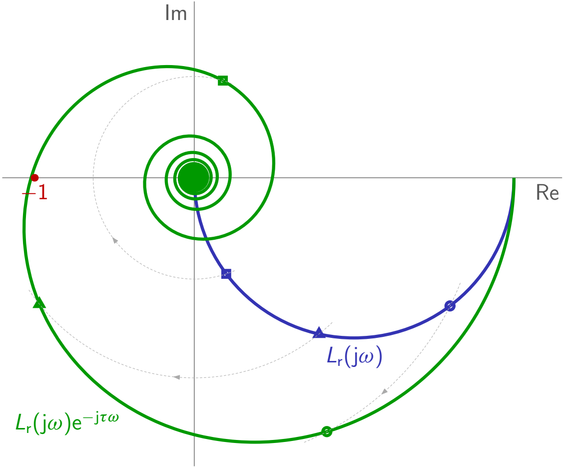
- does not change the magnitude of $L_r(j\omega)$ and
- adds phase lag proportional to ω,

which is not hard to account for.

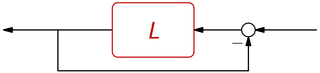
Effect of loop delay on $L(j\omega)$: Bode diagram



Effect of loop delay on $L(j\omega)$: polar plot



Nyquist criterion: what changes for dead-time systems ?



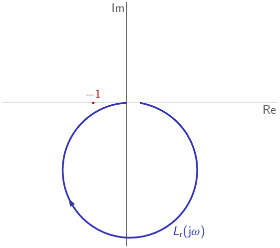
Addition of loop delay changes practically nothing, because

- both $1 + L(s)e^{-\tau s}$ and $S(s) = 1/(1 + L(s)e^{-\tau s})$ are still merpmorphic
  so Cauchy's argument principle applies
- if $|L(\infty)| < 1$, then $S(s)$ has at most a finite number of poles in $\mathbb{C}_0$
  so all of them are inside the Nyquist contour

The option to use the Nyquist stability criterion for dead-time systems is a

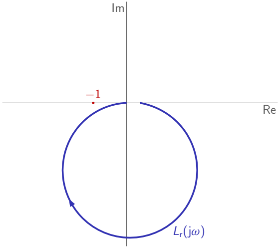
- great advantage of the loop-shaping philosophy.

Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$



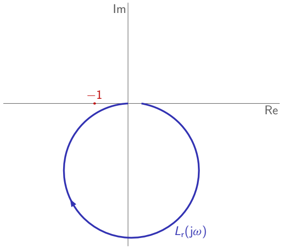
$\tau = 0$

Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$

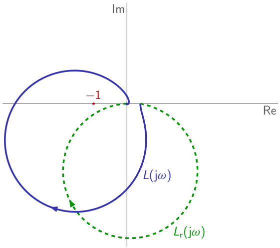


$\tau = 0$, closed-loop system is stable

Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$

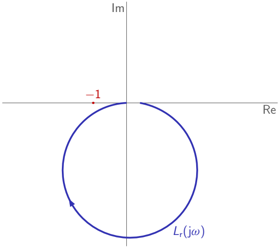


$\tau = 0$, closed-loop system is stable

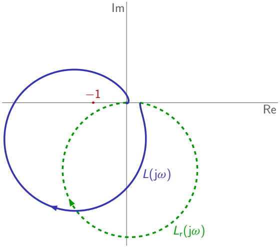


$\tau = 1$

Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$

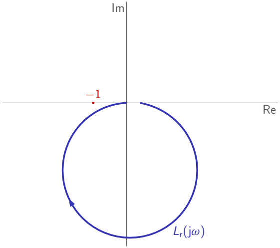


$\tau = 0$, closed-loop system is stable

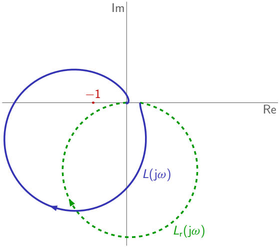


$\tau = 1$, closed-loop system is unstable

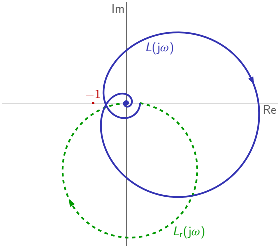
Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$



$\tau = 0$, closed-loop system is stable

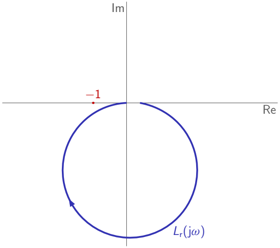


$\tau = 1$, closed-loop system is unstable

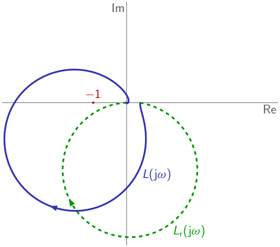


$\tau = 5$

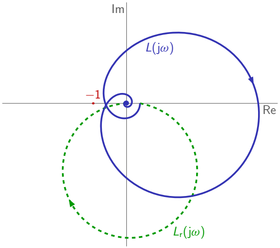
Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}e^{-\tau s}$



$\tau = 0$, closed-loop system is stable

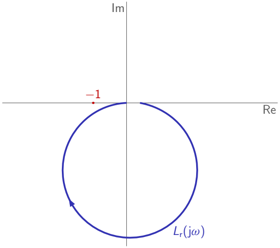


$\tau = 1$, closed-loop system is unstable

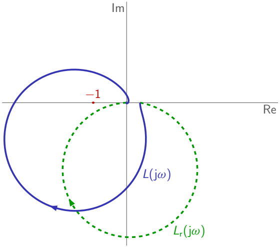


$\tau = 5$, closed-loop system is stable

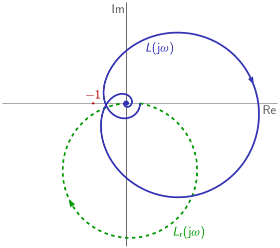
Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1}\mathrm{e}^{-\tau s}$



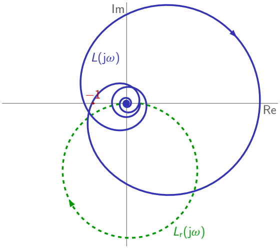
$\tau = 0$, closed-loop system is stable



$\tau = 1$, closed-loop system is unstable

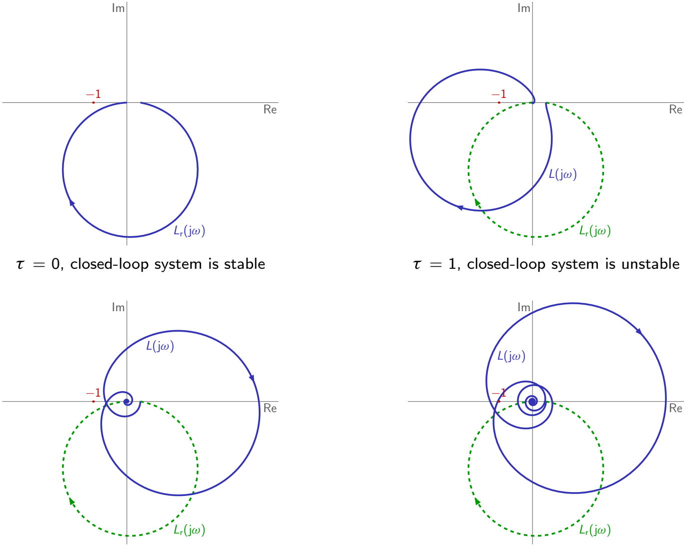


$\tau = 5$, closed-loop system is stable



$\tau = 11$

Nyquist criterion: example, $L(s) = \frac{0.4}{s^2+0.1s+1} e^{-\tau s}$



$\tau = 0$, closed-loop system is stable

$\tau = 1$, closed-loop system is unstable

$\tau = 5$, closed-loop system is stable

$\tau = 11$, closed-loop system is unstable

Outline

Motivation

A possible reason for adding a phase lag w/o affecting the magnitude is the presence of loop delay $\bar{D}_\tau$. One might be tempted to think that

- μ_{ph} can be used as a measure of tolerance to loop delay variations

To see if this is true, let $L_1(s) = \sqrt{2}/(10s + 1)$ and $L_2(s) = \sqrt{2}/(0.1s + 1)$:

so that

	μ_g	μ_{ph}
L_1	∞	0.75π
L_2	∞	0.75π

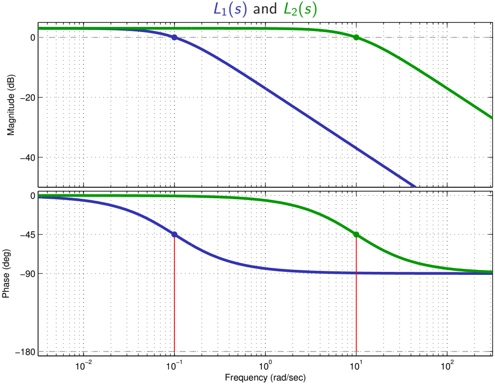
(i.e. L_1 and L_2 have the same μ_{ph})

Motivation

A possible reason for adding a phase lag w/o affecting the magnitude is the presence of loop delay $\bar{D}_\tau$. One might be tempted to think that

- μ_{ph} can be used as a measure of tolerance to loop delay variations

To see if this is true, let $L_1(s) = \sqrt{2}/(10s+1)$ and $L_2(s) = \sqrt{2}/(0.1s+1)$:



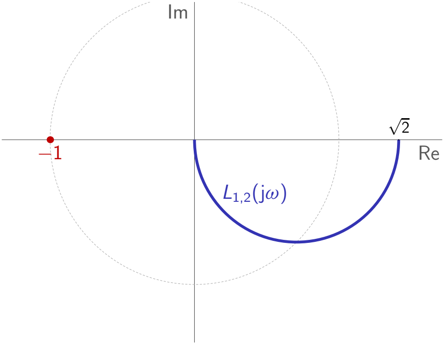
so that

	μ_{g}	μ_{ph}
L_1	∞	0.75π
L_2	∞	0.75π

(i.e. L_1 and L_2 have the same μ_{ph})

Motivation (contd)

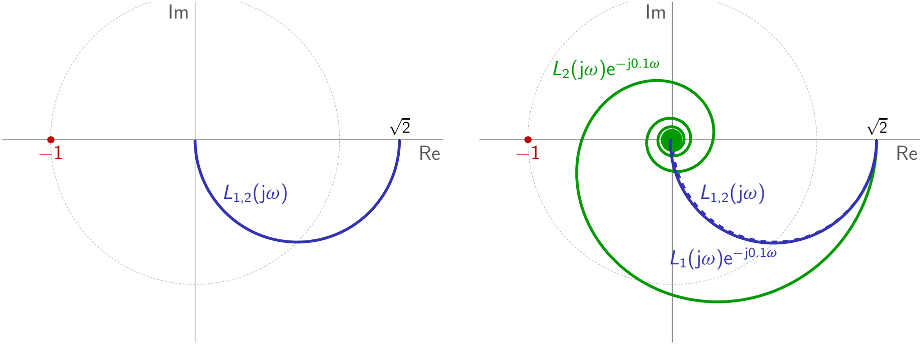
Whereas the polar plots of $L_1(s)$ and $L_2(s)$ coincide,



… the polar plots of $L_1(s)e^{-\tau s}$ and $L_2(s)e^{-\tau s}$ do not!

Motivation (contd)

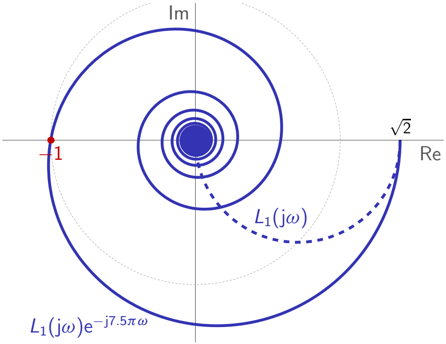
Whereas the polar plots of $L_1(s)$ and $L_2(s)$ coincide,



…the polar plots of $L_1(s)\mathrm{e}^{-\tau s}$ and $L_2(s)\mathrm{e}^{-\tau s}$ do not!

Motivation (contd)

Minimal destabilizing delays:



Hence, L_1 and L_2 are two systems having

- the same phase margins, whereas
- remarkably different tolerances to loop delays

What was wrong

Thus, the phase margin, μ_{ph}, might not reflect the sensitivity of the system to loop delays. Underlying reason is that

- μ_{ph} does not take into account the crossover frequency,

which is an important factor in analyzing the effect of τ on stability. Indeed, the phase lag due to loop delay is proportional to the frequency, hence the destabilizing phase lag due to the delay increases as ω_{c} grows.

This leads us to the need to introduce yet another stability margin:

Delay margin (μ_{d} or dead-time tolerance) is the the smallest destabilizing delay that may be introduced in the loop
(typically, $\mu_{\mathsf{d}} > 0$, although it might also be negative)

What was wrong

Thus, the phase margin, μ_{ph}, might not reflect the sensitivity of the system to loop delays. Underlying reason is that

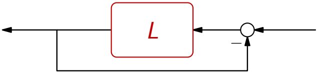
- μ_{ph} does not take into account the crossover frequency,

which is an important factor in analyzing the effect of τ on stability. Indeed, the phase lag due to loop delay is proportional to the frequency, hence the destabilizing phase lag due to the delay increases as ω_c grows.

This leads us to the need to introduce yet another stability margin:

Delay margin (μ_d or dead-time tolerance) is the the smallest destabilizing delay that may be introduced in the loop
(typically, $\mu_d > 0$, although it might also be negative)

Delay margin computation



Assume that

1. the closed-loop system is stable,
2. L has only one crossover frequency, ω_c, and
3. $\lim_{\omega\to\infty}|L(j\omega)| < 1$.

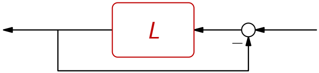
If a delay, say τ, is added to the loop, then the closed-loop system becomes unstable when $\arg\bigl(L(j\omega_c)e^{-j\tau\omega_c}\bigr) = -\pi$. Since

$$\arg\bigl(L(j\omega_c)e^{-j\tau\omega_c}\bigr) = \arg L(j\omega_c) - \tau\omega_c = -\pi + \mu_{ph} - \tau\omega_c,$$

the systems becomes unstable for $\tau = \mu_{ph}/\omega_c$. Thus,

$$\mu_d = \frac{\mu_{ph}}{\omega_c} \quad \text{(where } \mu_{ph} \text{ must be in } \textit{radians}\text{)}.$$

Delay margin computation



Assume that

1. the closed-loop system is stable,
2. L has only one crossover frequency, ω_c, and
3. $\lim_{\omega\to\infty}|L(j\omega)| < 1$.

If a delay, say τ, is added to the loop, then the closed-loop system becomes unstable when $\arg\bigl(L(j\omega_c)e^{-j\tau\omega_c}\bigr) = -\pi$. Since

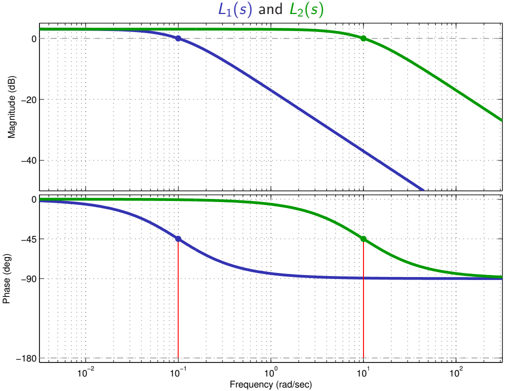
$$\arg\bigl(L(j\omega_c)e^{-j\tau\omega_c}\bigr) = \arg L(j\omega_c) - \tau\omega_c = -\pi + \mu_{ph} - \tau\omega_c,$$

the systems becomes unstable for $\tau = \mu_{ph}/\omega_c$. Thus,

– $\mu_d = \dfrac{\mu_{ph}}{\omega_c}$ (where μ_{ph} must be in *radians*).

Example

Consider again $L_1(s) = \sqrt{2}/(10s+1)$ and $L_2(s) = \sqrt{2}/(0.1s+1)$:



$L_1(s)$: $\mu_{ph} = 0.75\pi$ and $\omega_c = 0.1$, thus $\mu_d = 7.5\pi \approx 23.562$

$L_2(s)$: $\mu_{ph} = 0.75\pi$ and $\omega_c = 10$, thus $\mu_d = 0.075\pi \approx 0.23562$

Delay margin: design implication

Equality

$$\mu_{\mathsf{d}} = \frac{\mu_{\mathsf{ph}}}{\omega_{\mathsf{c}}}$$

implies that

– the larger ω_{c} is, the more sensitive the closed loop to delays is.

i.e. that the increase of ω_{c} renders the system more sensitive to (inevitable) loop delays. This

– imposes yet another limitation on ω_{c} and, therefore, on the achievable closed-loop bandwidth ω_{b}

For curious: if condition 2. fails (multiple crossovers)

In this case we should check the phase margins at all crossover frequencies. Important to realize that

- μ_{d} might not correspond to the crossover with the largest μ_{ph},

like in the example below (with $L_{\mathrm{r}}(s) = \frac{0.1(-2s+1)}{s(s+1)(s^2+0.13s+1)}$):