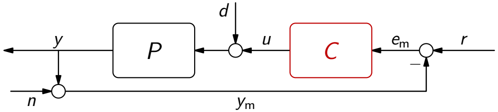
Introduction to Control (00340040)
lecture no. 8

Leonid Mirkin

Faculty of Mechanical Engineering
Technion—IIT

Previously on steady-state performance. . .

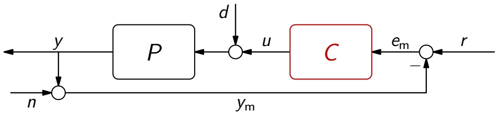


Zero steady-state errors to

- $r(t) = \mathbb{1}(t)$ requires an integrator in PC
- $d(t) = \mathbb{1}(t)$ requires an integrator in C
- $r(t) = \text{ramp}(t)$ requires a double integrator in PC
- $r(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires poles at $\pm j\omega$ in $P(s)C(s)$
- $d(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires poles at $\pm j\omega$ in $C(s)$
- $n(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires zeros at $\pm j\omega$ in $P(s)C(s)$

What if $e_{ss} = 0$ need not be attained? Or if r and / or d have their spectra spread over some frequency range and (uniformly) high-gain feedback is not feasible?

Previously on steady-state performance...



Zero steady-state errors to

- $r(t) = \mathbb{1}(t)$ requires an integrator in PC
- $d(t) = \mathbb{1}(t)$ requires an integrator in C
- $r(t) = \text{ramp}(t)$ requires a double integrator in PC
- $r(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires poles at $\pm j\omega$ in $P(s)C(s)$
- $d(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires poles at $\pm j\omega$ in $C(s)$
- $n(t) = \sin(\omega t + \phi)\mathbb{1}(t)$ requires zeros at $\pm j\omega$ in $P(s)C(s)$

What if $e_{ss} = 0$ need not be attained? Or if r and / or d have their spectra spread over some frequency range and (uniformly) high-gain feedback is not feasible?

Outline

Steady-state performance of closed-loop systems and loop shaping

Transient performance of closed-loop systems

Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

Outline

Steady-state performance of closed-loop systems and loop shaping

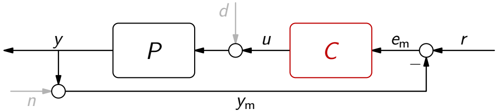
Transient performance of closed-loop systems

Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

Command following



Denote by Ω_r the frequency range where the spectrum of r is concentrated. By good steady-state command response we understand that

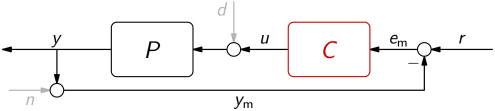
$$|E(j\omega)| \ll |R(j\omega)|, \quad \forall\omega \in \Omega_r$$

$$\Downarrow$$

$$|S(j\omega)| \ll 1, \quad \forall\omega \in \Omega_r.$$

(remember, $e = Sr$ if $d = n = 0$).

Good command following and loop gain



Denote the loop transfer function as $L(s) := P(s)C(s)$, so that

$$S(s) = \frac{1}{1 + L(s)} \quad \text{and } |S(\mathrm{j}\omega)| \leq \frac{1}{|L(\mathrm{j}\omega)| - 1} \text{ whenever } |L(\mathrm{j}\omega)| > 1$$

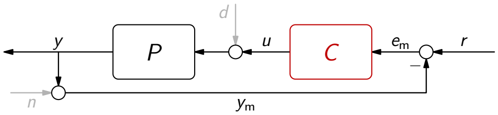
(by the triangle inequality, $|L| = |1 + L - 1| \leq |1 + L| + 1$). Hence,

$$|S(\mathrm{j}\omega)| \leq \sigma_r < 1 \quad \Longleftarrow \quad |L(\mathrm{j}\omega)| > \frac{1 + \sigma_r}{\sigma_r} = 1 + \frac{1}{\sigma_r} > 2$$

for every $\omega \in \Omega_r$. Qualitatively,

– *high loop gain* in the whole frequency range $\omega \in \Omega_r$ guarantees good steady-state command response.

Good command following and loop gain



Denote the loop transfer function as $L(s) := P(s)C(s)$, so that

$$S(s) = \frac{1}{1 + L(s)} \quad \text{and } |S(j\omega)| \le \frac{1}{|L(j\omega)| - 1} \text{ whenever } |L(j\omega)| > 1$$

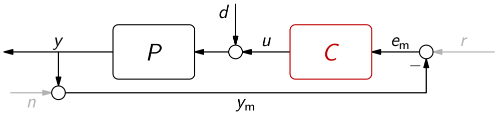
(by the triangle inequality, $|L| = |1 + L - 1| \le |1 + L| + 1$). Hence,

$$|S(j\omega)| \le \sigma_r < 1 \quad \Longleftarrow \quad |L(j\omega)| \ge \frac{1 + \sigma_r}{\sigma_r} = 1 + \frac{1}{\sigma_r} > 2$$

for every $\omega \in \Omega_r$. Qualitatively,

– *high loop gain* in the whole frequency range $\omega \in \Omega_r$ guarantees good steady-state command response.

Disturbance attenuation



Denote by Ω_d the frequency range where the spectrum of d is concentrated. By good steady-state disturbance attenuation we understand that

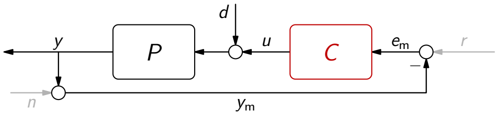
$$|E(\mathrm{j}\omega)| = |Y(\mathrm{j}\omega)| \ll |D(\mathrm{j}\omega)|, \quad \forall\omega \in \Omega_d$$

$$\Downarrow$$

$$|T_\mathrm{d}(\mathrm{j}\omega)| = |P(\mathrm{j}\omega)|\,|S(\mathrm{j}\omega)| \ll 1, \quad \forall\omega \in \Omega_d$$

(remember, $y = -e = T_\mathrm{d} d$ if $r = n = 0$).

Good disturbance attenuation and loop gain



Now,

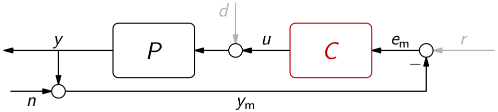
$$|P(j\omega)|\,|S(j\omega)| \le \sigma_d < 1 \quad \Longleftarrow \quad |L(j\omega)| \ge 1 + \frac{|P(j\omega)|}{\sigma_d}$$

for every $\omega \in \Omega_d$ (think of bounding S for $\sigma_r = \sigma_d/|P(j\omega)|$). Qualitatively,

– *high loop gain* in the whole frequency range $\omega \in \Omega_d$ guarantees good steady-state disturbance attenuation.

Remark: Note that a low plant gain could also help, but this is independent of the choice of the controller C.

Measurement noise sensitivity



Denote by Ω_n the frequency range where the spectrum of n is concentrated. By low steady-state sensitivity to measurement noise we understand that

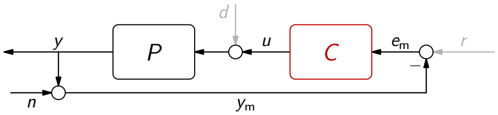
$$|E(j\omega)| = |Y(j\omega)| \ll |N(j\omega)|, \quad \forall\omega \in \Omega_n$$

$$\Downarrow$$

$$|T(j\omega)| \ll 1, \quad \forall\omega \in \Omega_n$$

(remember, $y = -e = Tn$ if $r = d = 0$).

Measurement noise sensitivity and high loop gain



Because

$$|T(j\omega)| = \frac{|L(j\omega)|}{|1 + L(j\omega)|} \geq \frac{|L(j\omega)|}{1 + |L(j\omega)|},$$

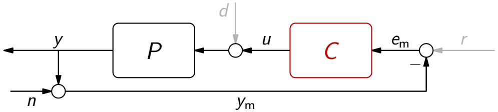
we have that

$$|L(j\omega)| > 1 \quad \Longrightarrow \quad |T(j\omega)| > \frac{1}{2}$$

and as $|L(j\omega)|$ increases, $|T(j\omega)| \to 1$. This means that

– high loop gain does *not* lead to low measurement noise sensitivity.

Low measurement noise sensitivity and loop gain



Because $T(s) = L(s)/(1 + L(s)) = 1/(1 + 1/L(s))$, we have that

$$|T(j\omega)| \le \frac{1}{1/|L(j\omega)| - 1} \quad \text{whenever } |L(j\omega)| < 1.$$

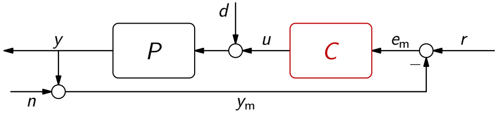
Hence,

$$|T(j\omega)| \le \sigma_n < 1 \quad \Longleftarrow \quad |L(j\omega)| \le \frac{\sigma_n}{1 + \sigma_n} \in \left(0, \frac{1}{2}\right)$$

for every $\omega \in \Omega_n$. Qualitatively,

– *low loop gain* in the whole frequency range $\omega \in \Omega_n$ guarantees low steady-state noise sensitivity.

Catch-22 situation ?



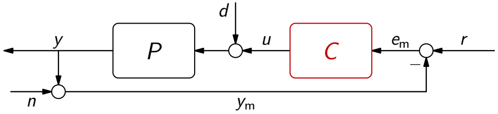
On the one hand,

– we need high loop gain (in $\omega \in \Omega_r$ and $\omega \in \Omega_d \setminus \{\omega \mid |P(j\omega)| \ll 1\}$).

On the other hand,

– we need low loop gain (in $\omega \in \Omega_n$).

"Typical" spectra of r, d, and n



In many cases[1],

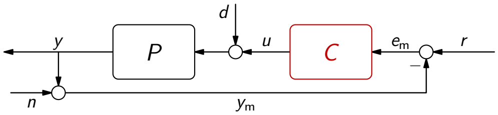
– command signals are "slow"
  (i.e. Ω_r mostly includes low frequencies)
– measurement noise is "fast"
  (i.e. Ω_n mostly includes high frequencies)

Moreover, since most physical processes are low-pass,

– only "slow" components of d should be worried about
  ("fast" part of d doesn't show up in y anyway as $|P(j\omega)| \ll 1$ at high frequencies)

[1] *Oi va voi* if this is not true!

"Typical" spectra of r, d, and n



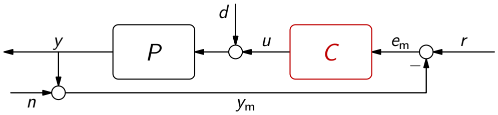
In many cases,

- command signals are "slow"
  (i.e. Ω_r mostly includes low frequencies)
- measurement noise is "fast"
  (i.e. Ω_n mostly includes high frequencies)

Moreover, since most physical processes are low-pass,

- only "slow" components of d should be worried about
  ("fast" part of d doesn't show up in y anyway as $|P(j\omega)| \ll 1$ at high frequencies)

The first acquaintance with *loop shaping*



Thus, we may endeavor to design loops with

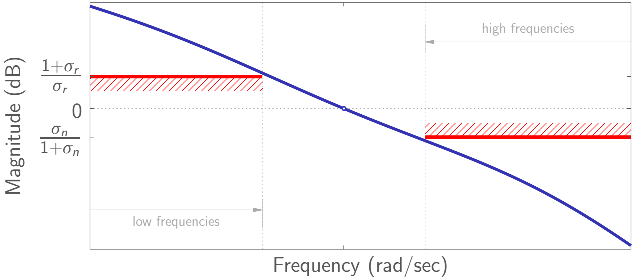
- high loop gain, $|L(j\omega)| \gg 1$, at "low" frequencies
- low loop gain, $|L(j\omega)| \ll 1$, at "high" frequencies

where "high" and "low" frequency ranges depend on spectral properties of exogenous signals in the application.

This control design philosophy is called loop shaping.

Loop shaping: big picture (magnitude)

What we shall try to do is to shape $|L(j\omega)|$ like this:

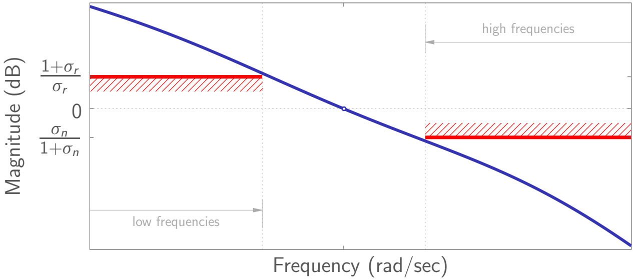


Note that

– there is always a region where the loop gain is neither high nor low

Loop shaping: big picture (magnitude)

What we shall try to do is to shape $|L(j\omega)|$ like this:



Note that

- there is always a region where the loop gain is neither high nor low

Outline

Steady-state performance of closed-loop systems and loop shaping

Transient performance of closed-loop systems

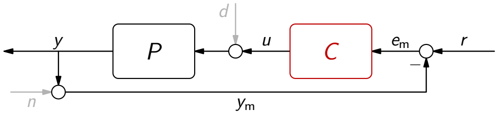
Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

Closed-loop transient response

We're mostly concerned with transient performance of command response:



and measure it on the basis of the step response (its speed and smoothness).

We know (from Lecture 4) that transient properties in time and frequency domains are related as follows:

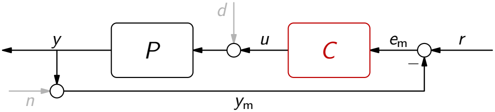
- the wider the bandwidth of $T(j\omega)$ is, the faster its step response is
- the higher resonant peaks of $T(j\omega)$ are, the larger over / undershoot is

The question:

- could these requirements be expressed in terms of $L(j\omega)$?

Closed-loop transient response

We're mostly concerned with transient performance of command response:



and measure it on the basis of the step response (its speed and smoothness).

We know (from Lecture 4) that transient properties in time and frequency domains are related as follows:

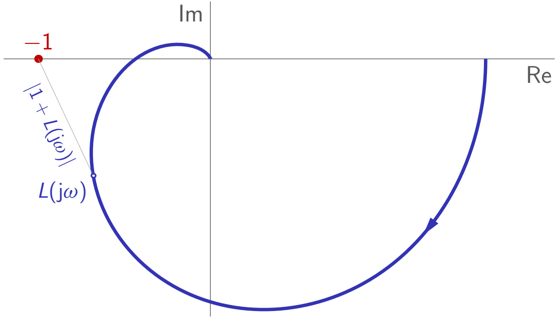
- the wider the bandwidth of $T(j\omega)$ is, the faster its step response is
- the higher resonant peaks of $T(j\omega)$ are, the larger over / undershoot is

The question:

- could these requirements be expressed in terms of $L(j\omega)$?

Closed vs. open loop: resonant peak of T

Given ω, $|1 + L(\mathrm{j}\omega)|$ is the distance between the points $L(\mathrm{j}\omega)$ and $-1 + \mathrm{j}0$ in the complex plane of $L(\mathrm{j}\omega)$:



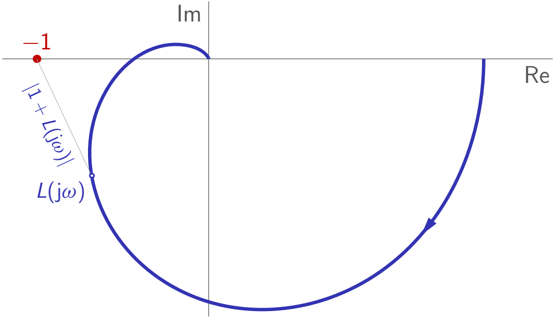
Thus,

– the closer $L(\mathrm{j}\omega)$ to $-1 + \mathrm{j}0$ is, the larger $|T(\mathrm{j}\omega)| = \dfrac{|L(\mathrm{j}\omega)|}{|1 + L(\mathrm{j}\omega)|}$ is

(as $L(\mathrm{j}\omega)$ approaches $-1 + \mathrm{j}0$, magnitude $|L(\mathrm{j}\omega)| \to 1$).

Closed vs. open loop: resonant peak of T

Given ω, $|1 + L(j\omega)|$ is the distance between the points $L(j\omega)$ and $-1 + j0$ in the complex plane of $L(j\omega)$:



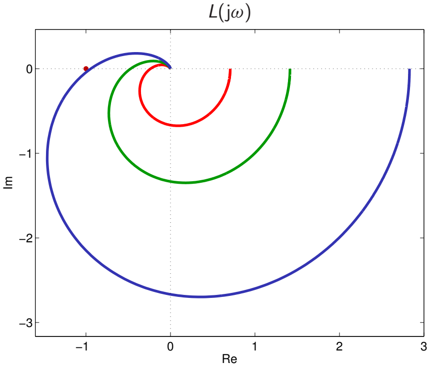
Thus,

– the closer $L(j\omega)$ to $-1 + j0$ is, the larger $|T(j\omega)| = \dfrac{|L(j\omega)|}{|1 + L(j\omega)|}$ is

(as $L(j\omega)$ approaches $-1 + j0$, magnitude $|L(j\omega)| \to 1$).

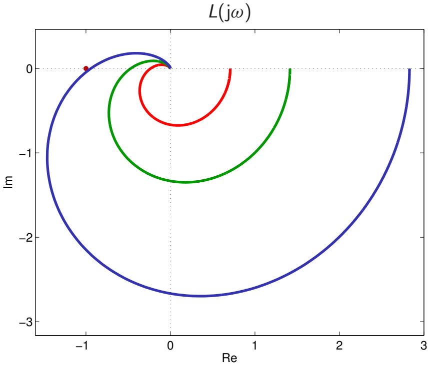
Closed vs. open loop: example

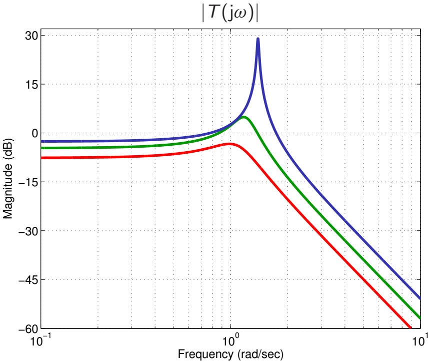
Let $L(s) = \dfrac{k\sqrt{2}}{(s+1)(s^2+s+1)}$. Then for $k \in \{\, 0.5, 1, 2 \,\}$ we have:



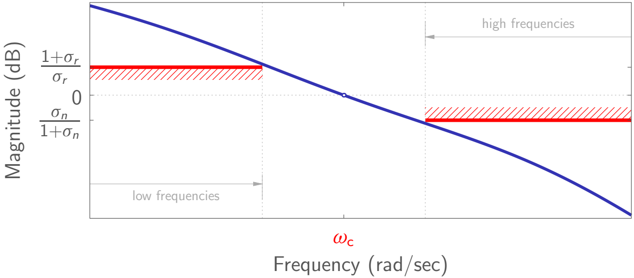
Closed vs. open loop: example

Let $L(s) = \dfrac{k\sqrt{2}}{(s+1)(s^2+s+1)}$. Then for $k \in \{\, 0.5, 1, 2 \,\}$ we have:





Crossover frequency and crossover region



When ω increases, $L(\mathrm{j}\omega)$ passes from the low- to high-frequency range. On its way it necessarily passes the area with $|L(\mathrm{j}\omega)| \approx 1$. This frequency range is called the crossover region and the frequency ω at which $|L(\mathrm{j}\omega)| = 1$ is called the crossover frequency and denoted ω_c, i.e.

$$|L(\mathrm{j}\omega_c)| = 1.$$

There may be more than one crossover frequencies.

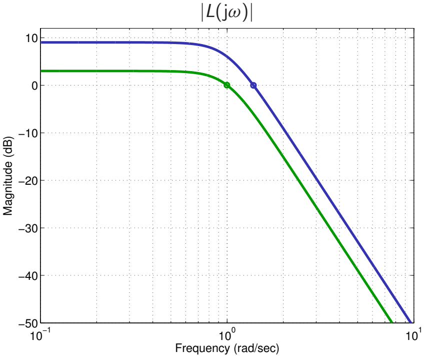
Closed vs. open loop: bandwidth of T

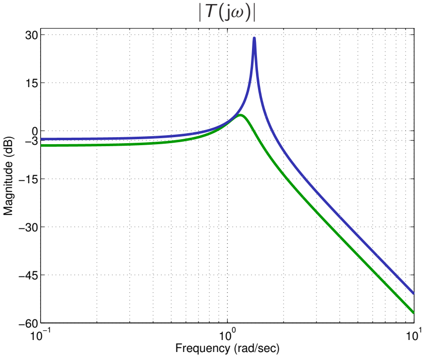
The closed-loop bandwidth ω_{b} is typically close to the crossover frequency ω_{c}. A rule of thumb is that $\omega_{\mathsf{b}} \approx 1.2 \div 1.5\omega_{\mathsf{c}}$.

Closed vs. open loop: bandwidth of T

The closed-loop bandwidth ω_b is typically close to the crossover frequency ω_c. A rule of thumb is that $\omega_\mathrm{b} \approx 1.2 \div 1.5\omega_\mathrm{c}$.

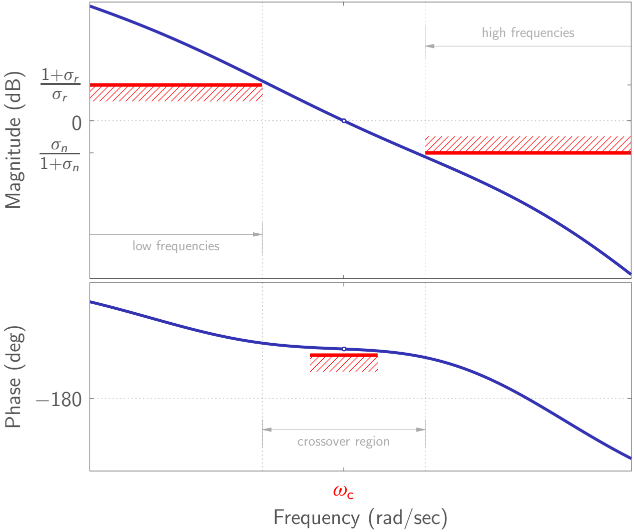
For example, for $L(s) = \dfrac{k\sqrt{2}}{(s+1)(s^2+s+1)}$ and $k \in \{1, 2\}$ we have:





Loop shaping: big picture (more details will follow)

What we shall try to do is to shape $|L(j\omega)|$ like this:



Outline

Steady-state performance of closed-loop systems and loop shaping

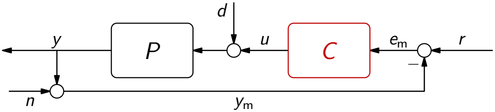
Transient performance of closed-loop systems

Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

Closed-loop control signal



Remember, from Lecture 5, that

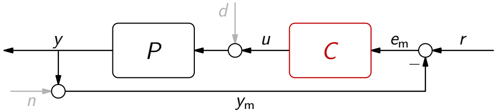
$$u = T_c r - T d - T_c n.$$

Thus, properties of the control signal in closed-loop control are

- shaped by properties of T_c and T

(rather than by properties of C_{ol} as in the open-loop case).

Steady-state control effort: command response



Since

$$|T_c(j\omega)| = \frac{|T(j\omega)|}{|P(j\omega)|},$$

properties of u are again

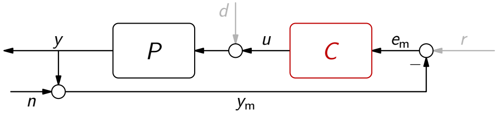
- determined by relations between controlled & uncontrolled bandwidths.

For example, if both P and T are low-pass filters and $\omega_{b,T} > \omega_{b,P}$, we may expect that $|T_c(j\omega)|$ grows in $\omega \in (\omega_{b,P}, \omega_{b,T})$. Hence,

- $\omega_{b,T} \gg \omega_{b,P}$ would lead to $|T_c(j\omega)| \gg 1$ around $\omega_{b,T}$,

which might not be affordable.

Steady-state control effort: measurement noise response



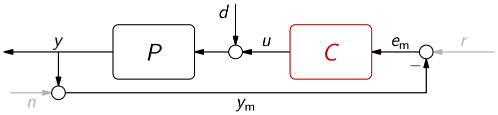
The growth of $|T_c(j\omega)|$ at high frequencies is

- even more dangerous from the noise response perspective

as this might be where the spectrum of the noise n is concentrated. Having high-magnitude high-frequency oscillations of u is highly undesirable as it

- might harm controlled process / actuators
- might excite poorly modeled high-frequency modes of the plant

Steady-state control effort: disturbance response



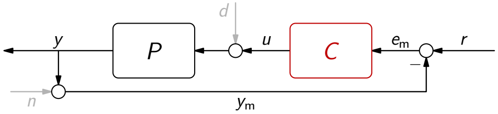
As we anyway should

- aim at avoiding high magnitude of $T(j\omega)$,

the effect of the input disturbances on the control signal

- needs no special attention.

Control effort during transients



An additional side effect of reaching $\omega_{b,T} \gg \omega_{b,P}$

– $T_c(j\omega)$ has high-frequency resonant peak(s),

which, in turn, leads to high-amplitude peaks in the step response of u (like in open-loop control, cf. discussion in Lecture 5).

Outline

Steady-state performance of closed-loop systems and loop shaping

Transient performance of closed-loop systems

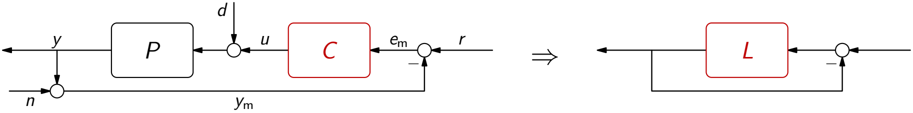
Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

What did we learn by now

We may replace



and aim at

- having appropriate crossover frequency, ω_c

  high enough: to cover spectra of reference / disturbances and have sufficiently fast transients

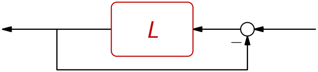
  not too high: to avoid the amplification of measurement noise and excessive grows of the control signal

- 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 ($\omega \gg \omega_c$)

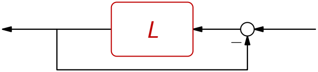
- keeping $L(j\omega)$ "far" from the point $-1 + j0$ in the crossover region

Don't we miss something important ?



- 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

Don't we miss something important ?

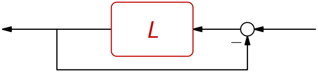


- ...
- having appropriate crossover frequency, ω_c
- high loop gain, $|L(\mathrm{j}\omega)| \gg 1$, at low frequencies ($\omega \ll \omega_\mathsf{c}$)
- low loop gain, $|L(\mathrm{j}\omega)| \ll 1$, at high frequencies ($\omega \gg \omega_\mathsf{c}$)
- keeping $L(\mathrm{j}\omega)$ "far" from the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

Of course,

stability

Stability analysis



We can analyze the stability of this system by

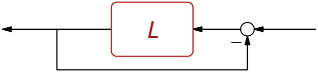
- algebraic analysis of the closed-loop characteristic polynomial
- graphical root-locus analysis

But

- ☹ neither of them does it in terms of the frequency response of L,

which is what loop shaping needs.

Nyquist stability criterion: what does it offer



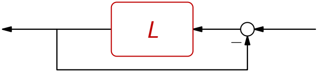
It analyzes the stability of the closed-loop system on the basis of

- number of unstable pole of $L(s)$ and
- position of the polar plot of $L(j\omega)$ with respect to the point $-1 + j0$

For example, if L is stable, then

closed-loop system is stable closed-loop system is unstable

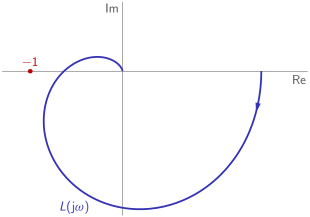
Nyquist stability criterion: what does it offer



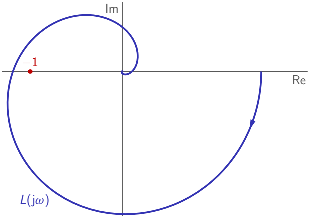
It analyzes the stability of the closed-loop system on the basis of

- number of unstable pole of $L(s)$ and
- position of the polar plot of $L(j\omega)$ with respect to the point $-1 + j0$

For example, if L is stable, then



closed-loop system is stable



closed-loop system is unstable

Outline

Steady-state performance of closed-loop systems and loop shaping

Transient performance of closed-loop systems

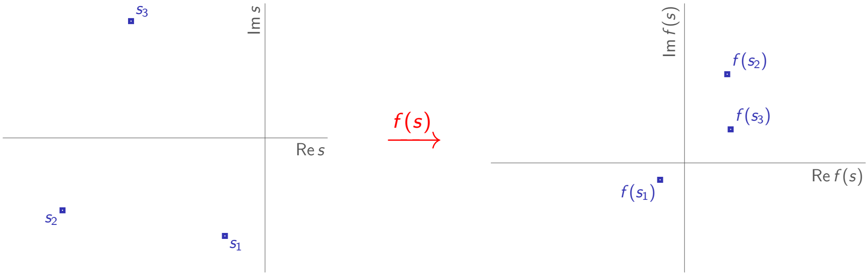
Control effort

Loop shaping: what we have and what we miss

Mathematical preliminaries: the (Cauchy's) argument principle

Mapping points in s-plane

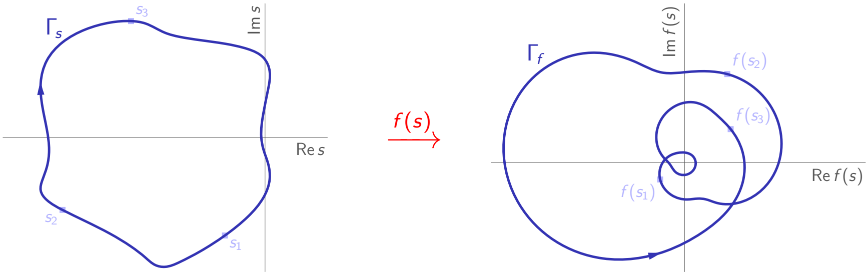
Consider a complex function $f(s)$. For any $s \in \mathbb{C}$ from its domain, $f(s) \in \mathbb{C}$ too. We say that s is mapped by f from the s-plane to the $f(s)$-plane:



(here $f(s) = \dfrac{0.273(-s+0.2)}{(s+0.6)(s+1)(s+1.3)}$, if you're curious).

Mapping contours in s-plane

Similarly, an s-plane contour Γ_s lying in the domain of $f(s)$ is mapped to a contour Γ_f in the $f(s)$-plane:



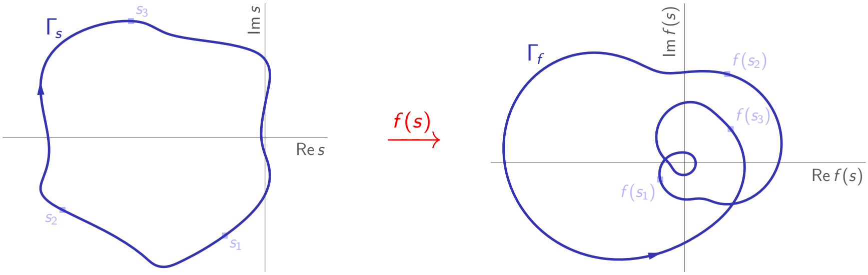
A contour is said to be

- simple if it does not intersect itself and
- closed if it starts and ends at the same point

(Γ_s above is simple closed, whereas Γ_f is closed yet not simple).

Mapping contours in s-plane

Similarly, an s-plane contour Γ_s lying in the domain of $f(s)$ is mapped to a contour Γ_f in the $f(s)$-plane:



A contour is said to be

- simple if it does not intersect itself and
- closed if it starts and ends at the same point

(Γ_s above is simple closed, whereas Γ_f is closed yet not simple).

The argument principle

Let

- Γ_s be a simple closed contour,
- $f(s)$ be meromorphic (i.e. only poles as singularities) inside and on Γ_s,
- Z_f be the number of zeros of $f(s)$ inside Γ_s,
- P_f be the number of poles of $f(s)$ inside Γ_s.

Theorem (Cauchy)

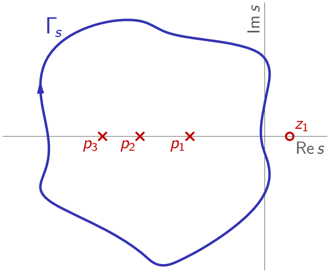
If Γ_s passes through neither poles nor zeros of $f(s)$, then Γ_f encircles the origin $Z_f - P_f$ times in the clockwise direction as s traverses Γ_s in the clockwise direction.

The argument principle: example 1

Consider

$$f(s) = \frac{0.273(-s + 0.2)}{(s + 0.6)(s + 1)(s + 1.3)},$$

which has $P_f = 3$ poles and $Z_f = 0$ zeros inside Γ_s.



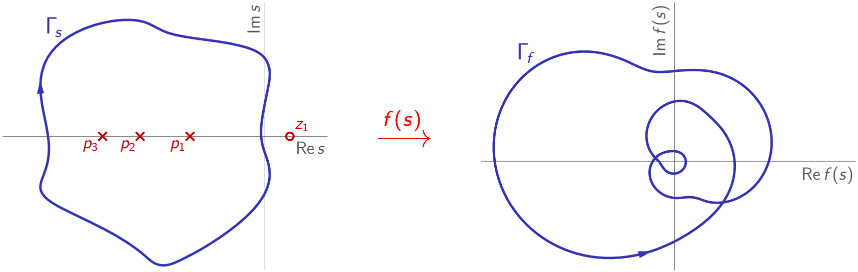
Hence, Γ_f encircles the origin -3 times in the clockwise direction.

The argument principle: example 1

Consider

$$f(s) = \frac{0.273(-s+0.2)}{(s+0.6)(s+1)(s+1.3)},$$

which has $P_f = 3$ poles and $Z_f = 0$ zeros inside Γ_s.



Hence, Γ_f encircles the origin -3 times[1] in the clockwise direction.

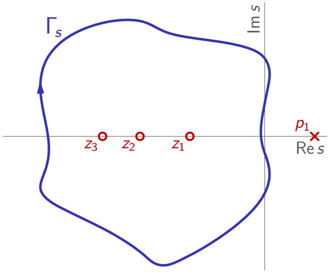
[1]That is, 3 times in the counterclockwise direction.

The argument principle: example 2

Consider

$$f(s) = \frac{0.13(s+0.6)(s+1)(s+1.3)}{-s+0.4},$$

which has $P_f = 0$ poles and $Z_f = 3$ zeros inside Γ_s.



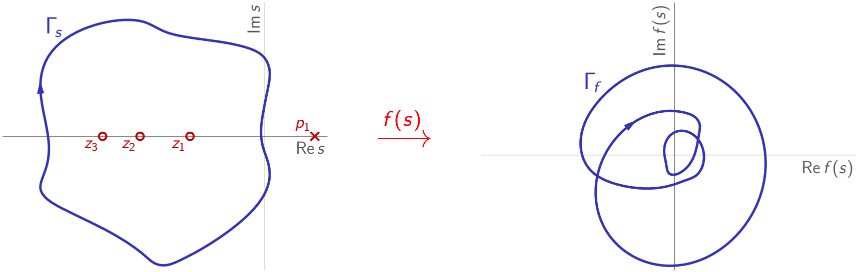
Hence, Γ_f encircles the origin 3 times in the clockwise direction.

The argument principle: example 2

Consider

$$f(s) = \frac{0.13(s+0.6)(s+1)(s+1.3)}{-s+0.4},$$

which has $P_f = 0$ poles and $Z_f = 3$ zeros inside Γ_s.



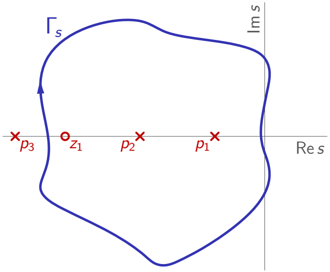
Hence, Γ_f encircles the origin 3 times in the clockwise direction.

The argument principle: example 3

Consider

$$f(s) = -\frac{0.7(s+1.6)}{(s+0.4)(s+1)(s+2)},$$

which has $P_f = 2$ poles and $Z_f = 1$ zero inside Γ_s.



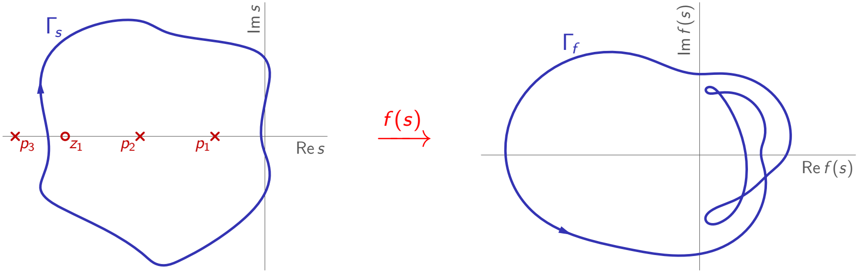
Hence, Γ_f encircles the origin -1 time in the clockwise direction.

The argument principle: example 3

Consider

$$f(s) = -\frac{0.7(s+1.6)}{(s+0.4)(s+1)(s+2)},$$

which has $P_f = 2$ poles and $Z_f = 1$ zero inside Γ_s.



Hence, Γ_f encircles the origin -1 time[2] in the clockwise direction.

[2] That is, once in the counterclockwise direction.

Shift by a constant

Let $f(s) = \alpha + g(s)$ for some $\alpha \in \mathbb{R}$. Then