Control Theory (035188)
lecture no. 0

Leonid Mirkin

Faculty of Mechanical Engineering
Technion—IIT

Outline

Signals and systems in the frequency domain

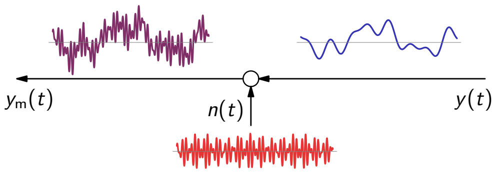
A crash review of control principles

Goals and tradeoffs in the frequency domain

Loop shaping fundamentals

Example

Let y_m be the following measurement of a signal y:



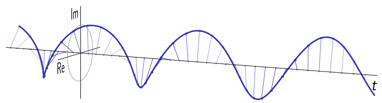
where n is a measurement noise. Important question then is

– how information (y) can be recovered from measurements (y_m)?

Harmonic signal

Signal

$$f(t) = ce^{j\omega t} =$$



for $c \in \mathbb{C}$ and $\omega \in \mathbb{R}$ is called harmonic signal with frequency ω, amplitude $|c|$, and initial phase $\phi = \arg(c)$. It is $2\pi/|\omega|$-periodic (ω may be negative).

Euler's formula $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$ connects real and complex sines. Also, we have that

$$\cos(\omega t) = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \qquad \text{and} \qquad \sin(\omega t) = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}$$

so normally harmonics with ω and $-\omega$ come together.

We say that $c_1 e^{j\omega_1 t}$ is faster / slower than $c_2 e^{j\omega_2 t}$ if $|\omega_1| > |\omega_2|$ / $|\omega_1| < |\omega_2|$.

Frequency-domain representation

Given $f : \mathbb{R} \to \mathbb{F}^n$, its Fourier transform

$$\mathfrak{F}\{f\} = F(\mathrm{j}\omega) := \int_{\mathbb{R}} f(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t,$$

where $\omega \in \mathbb{R}$ is frequency (in radians per time unit). Inverse transform

$$\mathfrak{F}^{-1}\{F\} = f(t) = \frac{1}{2\pi}\int_{\mathbb{R}} F(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}\omega.$$

$F = \mathfrak{F}\{f\}$ is called the frequency-domain representation (or spectrum) of f.

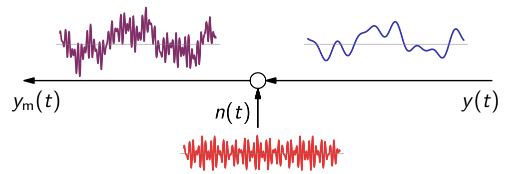
- f is a superposition of elementary harmonics $\mathrm{e}^{\mathrm{j}\omega t}$
- $F(\mathrm{j}\omega_0)$ quantifies the contribution of $\mathrm{e}^{\mathrm{j}\omega_0 t}$ to f

Harmonic signals can be categorized on the "fast–slow" principle, namely

- $\mathrm{e}^{\mathrm{j}\omega_1 t}$ is faster (slower) than $\mathrm{e}^{\mathrm{j}\omega_2 t}$ if $|\omega_1| > |\omega_2|$ ($|\omega_1| < |\omega_2|$)

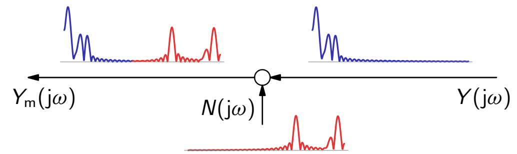
Thus, the spectrum of f offers a different, informative, viewpoint on it and facilitates the separation of fast and slow components.

Fourier transform: interpretation (contd)



Signal y and noise n

- cannot be separated in the time domain,
- but can be separated in the frequency domain:



Separation can be done via filtering.

LTI systems in the frequency domain

If $P : u \mapsto y$ is LTI (linear time-invariant), then

$$y(t) = \int_{-\infty}^{\infty} g(t-s)u(s)\mathrm{d}s =: (g * u)(t),$$

where g is the impulse response of P.

In the frequency domain,

$$y = g * u \implies Y(\mathrm{j}\omega) = G(\mathrm{j}\omega)U(\mathrm{j}\omega),$$

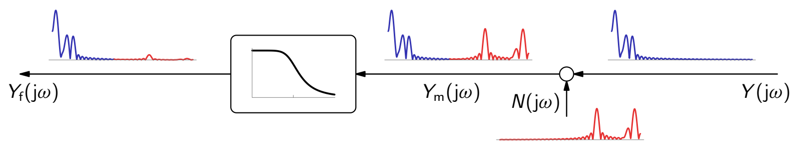
with the Fourier transform of the impulse response,

- $G(\mathrm{j}\omega) = G(s)|_{s=\mathrm{j}\omega}$, is known as the frequency response of G.

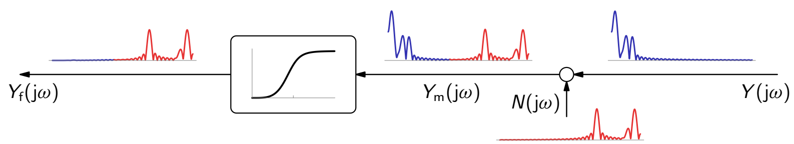
Systems as filters

Systems can be used to shape the spectrum of outputs.

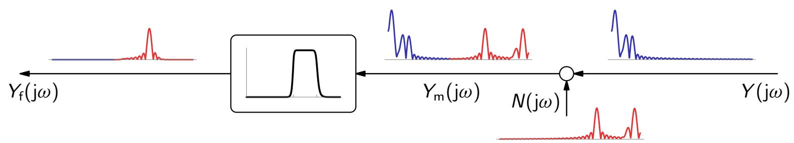
- low-pass



- hi-pass



- band-pass



Outline

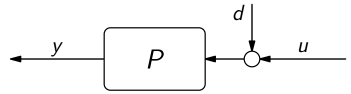
Signals and systems in the frequency domain

A crash review of control principles

Goals and tradeoffs in the frequency domain

Loop shaping fundamentals

Prototype control problem



y: regulated signal

u: control signal (means)

d: load disturbance

P: plant

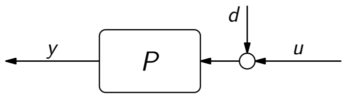
Goal

$$u \longrightarrow y = r$$

where

r: reference signal (desired behavior)

Ultimate methodology: plant inversion



$$y = P(d + u) \;\wedge\; y = r$$

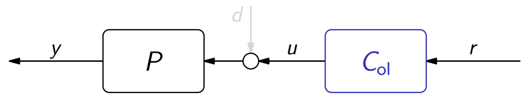
$$\Downarrow$$

$$r = P(d + u)$$

$$\Downarrow$$

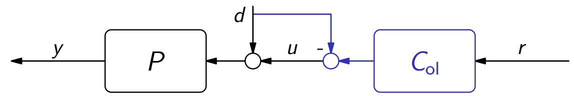
$$u = \frac{1}{P} r - d$$

Open-loop plant inversion



with

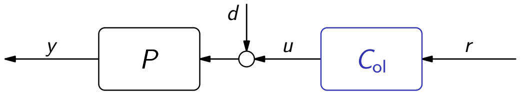
$$C_{ol} = \frac{1}{P}$$



with

$$C_{ol} = \frac{1}{P}$$

Limitations of open-loop plant inversion: internal stability



Signals

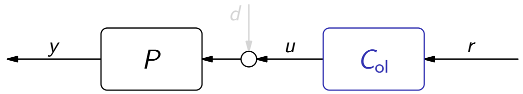
$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} PC_{\text{ol}} & P \\ C_{\text{ol}} & 0 \end{bmatrix} \begin{bmatrix} r \\ d \end{bmatrix},$$

must bounded.

Hence, must have

- stable P
- bounded $C_{\text{ol}} r$
  - if r is unknown a priori $\implies$ $C_{\text{ol}} = 1/P$ must be stable
  - if r is known $\implies$ stability of C_{ol} may be relaxed, e.g. non-proper $C_{\text{ol}}(s)$

Approximate open-loop plant inversion



Pragmatic alternative if P is not stably invertible:

$$C_{\text{ol}} \approx \frac{1}{P} \quad \Longrightarrow \quad C_{\text{ol}} = \frac{T_{\text{r}}}{P} \quad \to \quad y = T_{\text{r}}\, r$$

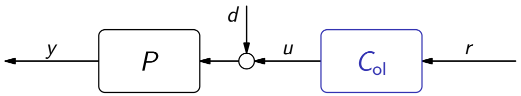
where the purpose of the reference model $T_{\text{r}} : r \mapsto y$ is twofold:

1. render C_{ol} feasible
2. keep $T_{\text{r}} r \approx r$ e.g. $T_{\text{r}}(\mathrm{j}\omega) \approx 1$ at dominant frequencies of $r(\mathrm{j}\omega)$

Technically,

- T_{r} must be stable itself
- T_{r} must result in bounded $u = C_{\text{ol}} r = (T_{\text{r}}/P) r$
  - nonminimum-phase zeros of $P(s)$ must be zeros of $T_{\text{r}}(s)$
  - large enough pole excess of $T_{\text{r}}(s)$
    - $\geq$ pole excess of $P(s)$, if r is unknown a priori $\implies$ proper $C_{\text{ol}}(s)$
    - no "overly" high-order derivatives in C_{ol} if r is known

Limitations of open-loop plant inversion: other



- unmeasured d nothing to do
- uncertain P nothing to do

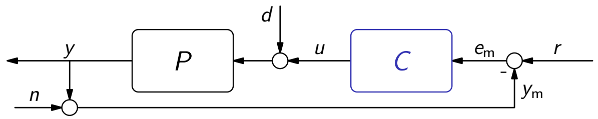
Namely, if

$$C_{\text{ol}} = \frac{T_{\text{r}}}{P},$$

while $P_{\text{true}} \neq P$ and $d \neq 0$, then

$$\begin{aligned} y = P_{\text{true}}(d + u) &= P_{\text{true}} d + \frac{P_{\text{true}}}{P} T_{\text{r}}\, r \\ &= T_{\text{r}}\, r + P_{\text{true}} d - \left(1 - \frac{P_{\text{true}}}{P}\right) T_{\text{r}}\, r \end{aligned}$$

Unity-feedback closed-loop setup



Gang of four

$$\begin{bmatrix} S(s) & T_{\text{c}}(s) \\ T_{\text{d}}(s) & T(s) \end{bmatrix} := \frac{1}{1 + P(s)C(s)} \begin{bmatrix} 1 & C(s) \\ P(s) & P(s)C(s) \end{bmatrix},$$

must be all stable (internal stability requirement, no unstable cancellations).

Signals

$$\begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} T & T_{\text{d}} & -T \\ T_{\text{c}} & -T & -T_{\text{c}} \\ S & -T_{\text{d}} & T \end{bmatrix} \begin{bmatrix} r \\ d \\ n \end{bmatrix},$$

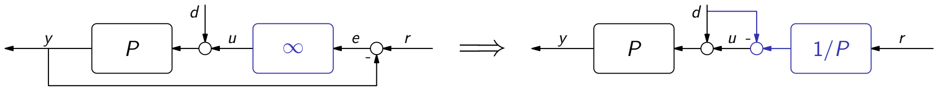
where $e := r - y = e_{\text{m}} + n$.

The marvel of feedback: closed-loop plant inversion

Because

$$T_c = \frac{1}{1/C + P} \xrightarrow{C\to\infty} \frac{1}{P} \quad \text{and} \quad -T = -\frac{P}{1/C+P} \xrightarrow{C\to\infty} -1,$$

we have



Thus,

$$T_d = \frac{P}{1+PC} \xrightarrow{C\to\infty} 0 \quad \text{and} \quad S = \frac{1}{1+PC} \xrightarrow{C\to\infty} 0,$$

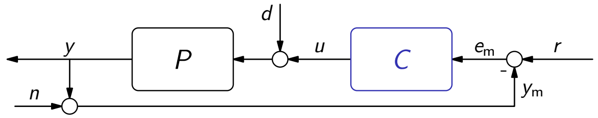
independently of the plant and w/o an explicit measurement of d. In other words,

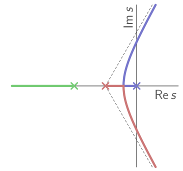
- feedback is capable of handling uncertainty.

Also,

- feedback can stabilize unstable systems. but might also destabilize

Limitations of closed-loop plant inversion





- closed-loop stability
- closed-loop stability
- closed-loop stability
- measurement noise sensitivity ($T \to 1$ and $T_c \to 1/P$)
- limited u
- ...

Hence,

- (nontrivial) tradeoffs, conveniently over different frequencies

Outline

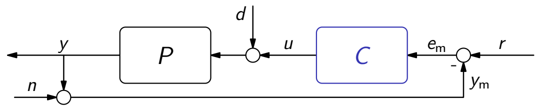
Signals and systems in the frequency domain

A crash review of control principles

Goals and tradeoffs in the frequency domain

Loop shaping fundamentals

Requirements to control systems



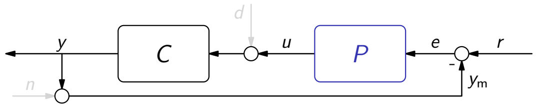
We want:

1. internal stability (i.e. stability of all closed-loop treansfer functions),
2. good command following (i.e. small tracking error e),
3. good disturbance attenuation (i.e. attenuation of the effect of d on y),
4. low sensitivity to measurement noise,

and all this

5. with "reasonably small" control efforts.

Command response: steady-state performance



Remember that

$$y = Tr \quad \text{and} \quad e = Sr.$$

By good steady-state command response we understand that

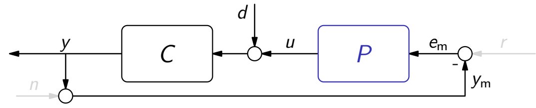
$$Y(j\omega) \approx R(j\omega) \quad \text{or} \quad |E(j\omega)| \ll |R(j\omega)|$$

in the frequency range where the spectrum of r concentrated. Hence, good command response requires

$$T(j\omega) \approx 1 \quad \text{or, equivalently,} \quad |S(j\omega)| \ll 1$$

in the frequency range where the spectrum of r concentrated.

Disturbance response: steady-state performance



In this case

$$y = T_d d = PSd \,(= -e).$$

By good steady-state disturbance attenuation we understand that

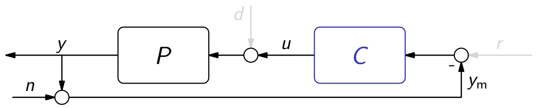
$$|Y(j\omega)| \ll |D(j\omega)|$$

in the frequency range where the spectrum of d concentrated. Hence, good disturbance response requires $|P(j\omega)S(j\omega)| \ll 1$ which often reduces to

$$|S(j\omega)| \ll 1 \quad \text{or} \quad |P(j\omega)| \ll 1$$

(the latter condition does *not* depend on the controller).

Noise sensitivity



In this case

$$y = -Tn.$$

By low sensitivity to measurement noise we understand that

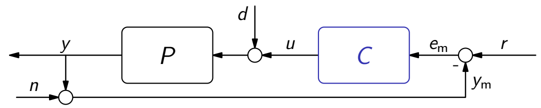
$$|Y(j\omega)| \ll |N(j\omega)|$$

in the frequency range where the spectrum of n concentrated. This requires

$$|T(j\omega)| \ll 1$$

in the frequency range where the spectrum of n concentrated.

Steady-state requirements



Thus (although this is a bit simplistic), we want

1. internal stability
2. $|S(j\omega)| \ll 1$ (good command and disturbance responses)
3. $|T(j\omega)| \ll 1$ (low noise sensitivity)

A (small) problem is that

– 2. and 3. cannot be achieved simultaneously,

S-T tradeoff

Fundamental constraint:

$$S(j\omega) + T(j\omega) \equiv 1$$

This means that $S(j\omega)$ and $T(j\omega)$ cannot be made small simultaneously:

- if $|S(j\omega)| \ll 1$, then $T(j\omega) \approx 1$;
- if $|T(j\omega)| \ll 1$, then $S(j\omega) \approx 1$;
- it might even happen that $|S(j\omega)| \gg 1$ and $|T(j\omega)| \gg 1$ (why?);
- yet never that $|S(j\omega)| \ll 1$ and $|T(j\omega)| \ll 1$

The trick is that $|S(j\omega)|$ and $|T(j\omega)|$ required to be

- small at different frequencies.

Frequency properties of r, d, and n

In many cases[1],

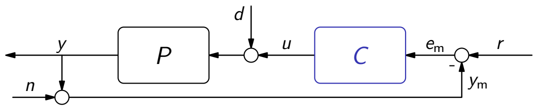
- command signals are "slow"
  (i.e. spectrum of r concentrated in the low-frequency range)
- measurement noise is "fast"
  (i.e. spectrum of n concentrated in the high-frequency range)

Moreover, since most physical processes are low-pass filters,

- 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!

Steady-state requirements (contd)



We thus have the following requirements:

1. internal stability
2. $|S(j\omega)| \ll 1$ at "low" frequencies
3. $|T(j\omega)| \ll 1$ at "high" frequencies

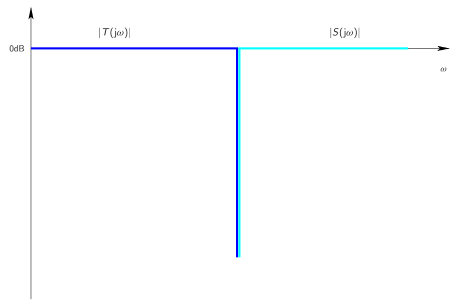
where "low" and "high" depend upon properties of exogenous signals ($r(t)$, $d(t)$, and $n(t)$) and the plant $P(s)$ (e.g. of its bandwidth).

Steady-state requirements (contd)

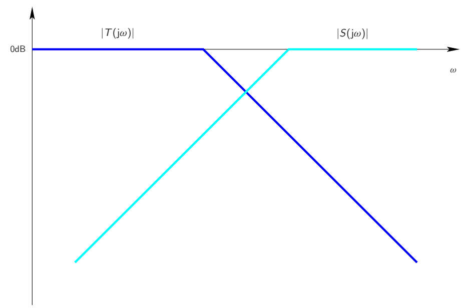
In other words,

- $S(s)$ should be high-pass filter,
- $T(s)$ should be low-pass filter.

Ideally, something like this:

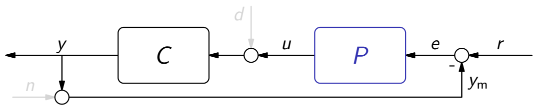


More realistically, something like this:



Transient performance of command response

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

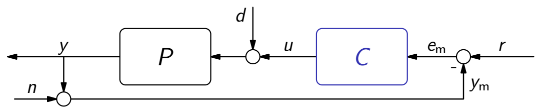


and measure it on the basis of the step response (its speed and smoothness).

We know that from the transient performance viewpoint time-domain and frequency-domain properties 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.

Control effort



Since

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

properties of the step response of u

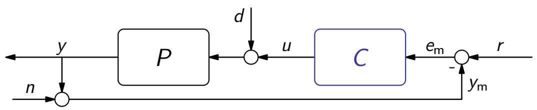
- determined by the ratio of the closed- and open-loop bandwidths.

For example, if $\omega_{b,T(s)} \gg \omega_{b,P(s)}$,

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

which, in turn, leads to high-amplitude peaks in the step response of u.

Requirements to control systems



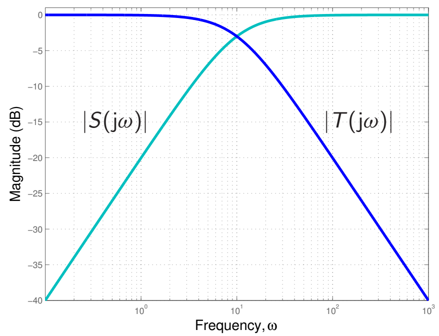
We thus end up with the following requirements:

- internal stability
- $|S(j\omega)| \ll 1$ at "low" frequencies
- $|T(j\omega)| \ll 1$ at "high" frequencies
- sufficiently "wide" (but not too "wide") bandwidth ω_b of $T(j\omega)$
- no high resonant peaks in $T(j\omega)$

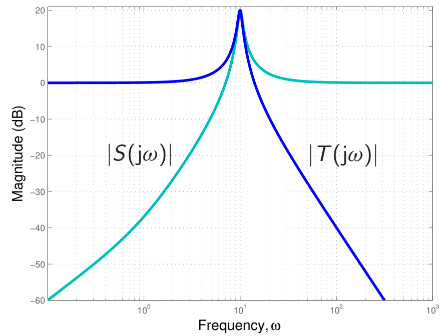
where "low" and "high" depend upon properties of exogenous signals (r, d, and n) and the plant $P(s)$ (e.g. its bandwidth) and ω_b is limited from above by control effort constraints and the spectrum of n.

The good, the bad and ...

This is what we can get:



"Good" S and T



"Bad" S and T

Outline

Signals and systems in the frequency domain

A crash review of control principles

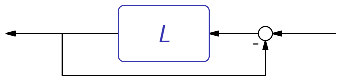
Goals and tradeoffs in the frequency domain

Loop shaping fundamentals

Loop shaping

Design of $C(s)$ to obtain desired $S(s)$ and $T(s)$ complicated by the fact that

– $S = 1/(1 + PC)$ and $T = PC/(1 + PC)$ are nonlinear as functions of the controller.



Loop shaping is design method attempting to produce desirable $S(s)$ and $T(s)$ by shaping the frequency response of $L(s)$. In other words, it aims at

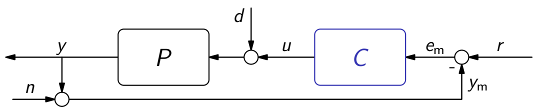
– producing required closed-loop f.r. by shaping open-loop f.r.

Advantages:

– $L(s) = P(s)C(s)$ is linear as a function of the controller
(this facilitates modular design in which simple controller blocks added at each step)

– we have only one transfer function to take care of

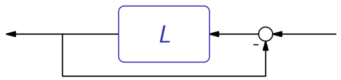
Closed vs. open loop: stability



The closed-loop system internally stable iff

1. no unstable pole-zero cancellations occur in $P(s)C(s)$,
2. the Nyquist plot of $L(j\omega)$ agrees with the Nyquist stability criterion

Closed vs. open loop: S and T



As

$$|S(j\omega)| = \frac{1}{|1 + L(j\omega)|} \quad \text{and} \quad |T(j\omega)| = \frac{|L(j\omega)|}{|1 + L(j\omega)|},$$

we have:

– If $|L(j\omega)| \gg 1 \implies |S(j\omega)| \ll 1$ (independently of $\arg L(j\omega)$)

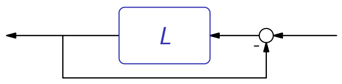
– If $|L(j\omega)| \ll 1 \implies |T(j\omega)| \ll 1$ (independently of $\arg L(j\omega)$)

– If $|L(j\omega)| \approx 1$, the situation is more delicate. In this case

$$|T(j\omega)| \approx |S(j\omega)| \approx \tfrac{1}{|1+L(j\omega)|}$$

might be in $[\approx \frac{1}{2}, \infty)$, depending on $\arg L(j\omega)$ (e.g. check $L = \pm 1$).

Crossover region



By crossover frequency, ω_c, we understand the frequency at which

$$|L(j\omega_c)| = 1\ (= 0\,\text{dB}).$$

Frequency range (around ω_c) where $|L(j\omega_c)| \approx 1$ called crossover region.

Shaping $L(j\omega)$ in the crossover region is

- most delicate and, arguably, most important

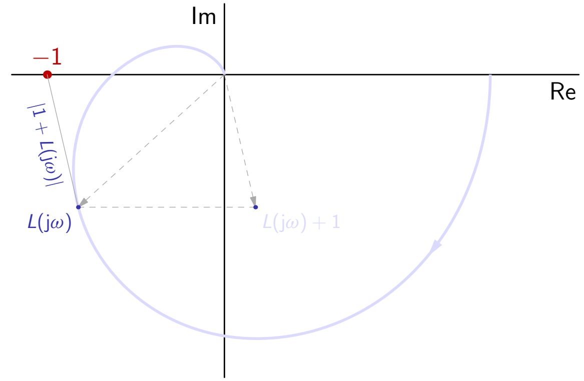
part of loop shaping because

- stability, transient performance, sensitivity to modeling inaccuracies

all heavily depend upon it.

Closed vs. open loop: resonant peak of T

$|1 + L(j\omega)|$ is the distance between $L(j\omega)$ and the critical point:

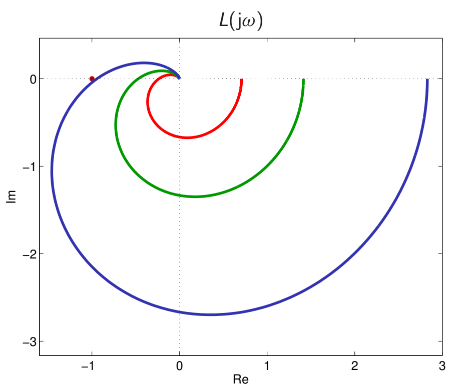


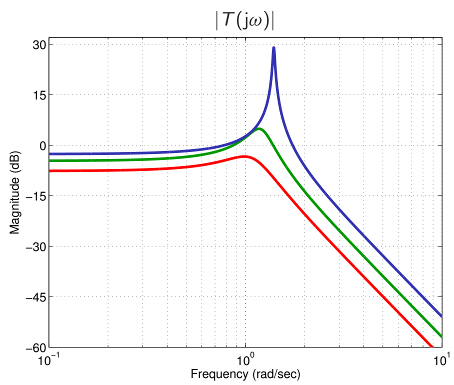
Thus,

- the closer $L(j\omega)$ to the critical point is, the larger $|T(j\omega)|$ is.

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:

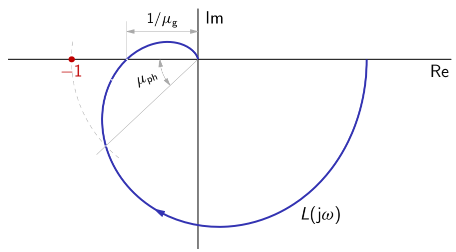


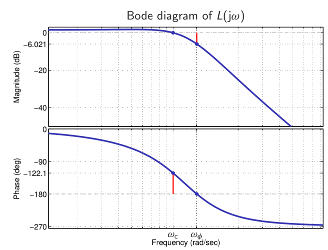


Closeness to the critical point

Is normally (and simplistically) measured by

- stability margins, like gain μ_g, phase μ_{ph}, ...
  (all are measured around the crossover frequency ω_c, at which $|L(j\omega_c)| = 1$)

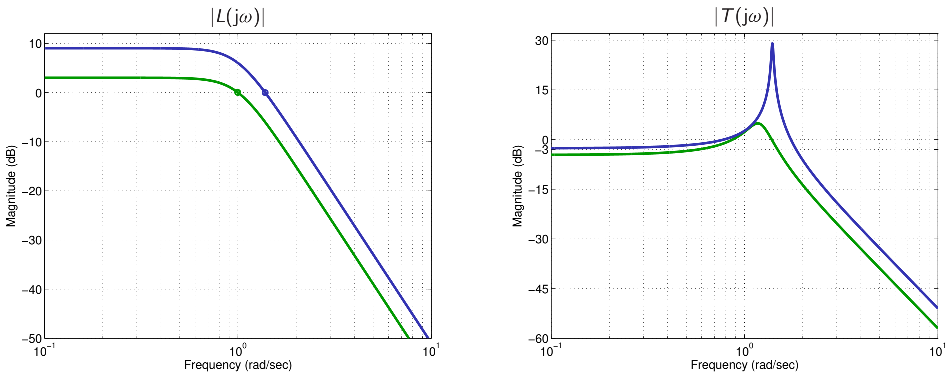




Closed vs. open loop: bandwidth of T

The closed-loop bandwidth ω_b is typically close to the crossover frequency ω_c. A rule of thumb is that $\omega_b \approx 1.2 \div 1.5\,\omega_c$.

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

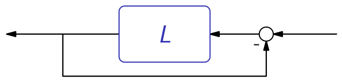


Closed vs. open loop: the bottom line

Roughly, the following relationships hold:

- Closed-loop system stable $\iff$ $L(j\omega)$ verifies the Nyquist criterion
- $|S(j\omega)| \ll 1 \iff |L(j\omega)| \gg 1$
- $|T(j\omega)| \ll 1 \iff |L(j\omega)| \ll 1$
- $T(j\omega)$ has "sufficiently wide" bandwidth ω_b $\Updownarrow$ $L(j\omega)$ has "sufficiently large" crossover frequency ω_c
- $T(j\omega)$ does not have high resonant peaks $\Updownarrow$ $L(j\omega)$ is "far" from the critical point in the crossover region

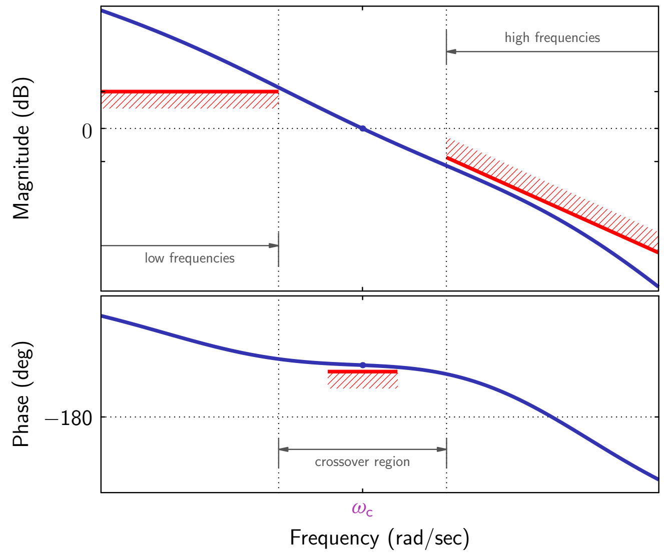
Requirements to $L(j\omega)$



We finally get:

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

Big picture of loop shaping



. . . although could be different in some cases, like lightly-damped systems