Chapter 11: Design in the frequency domain
Nyquist stability criterion

August 19, 2015

Outline

Introduction

The root-locus method dealt with the s- and z-domains and the location of the poles and zeros formed the basis for that method.

For the **frequency domain**, we use the following substitution: $s = j\omega$, which means we will only regard perfect oscillations.

Sines, cosines and exponential signals are eigenfunctions of Linear Time Invariant systems (LTI).
Perfect oscillations form the natural decomposition of each signal when you are dealing with LTIs.

Introduction

We need to translate our **design criteria** to something that fits the discussed method. For the root-locus method, we had to express the criteria in positions of poles and zeros.
For the frequency domain, typical design criteria are:

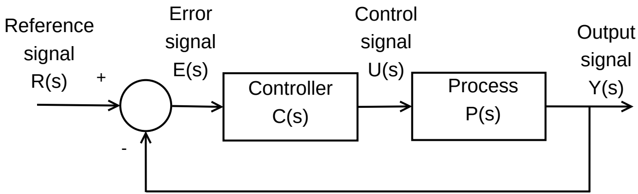
- Phase and gain margin
- Bandwidth
- Zero-frequency magnitude (= DC gain)

We will discuss two different **graphical representations** to design compensators in the frequency domain:

- Nyquist plots
- Bode plots (for the design of lead, lag and lead-lag compensators: next lecture)

Outline

Stability of the closed loop system



We write the output signal as a function of the input signal:

$$Y(s) = \frac{P(s)C(s)}{1 + P(s)C(s)} R(s).$$

The closed loop system stability is determined by the poles of $\frac{P(s)C(s)}{1+P(s)C(s)}$, or the roots of $1 + P(s)C(s) = 0$.

Stability of the closed loop system

In the root-locus method, we determined the positions of the poles by plotting the roots of $1 + P(s)C(s) = 0$.

The system is stable if all the roots remain in the left half plane.

We are however not interested in the positions of the poles.
We only need to know whether there *are* poles in the right half plane.

There is a cheaper alternative: **the Nyquist stability criterion**.

Stability: Nyquist criterion

The Nyquist stability criterion avoids determining the roots of $1 + P(s)C(s)$ exactly. It only determines the *number* of roots in the right half plane.

It uses a theorem from complex calculus that finds the difference between the number of poles and the number of zeros within a **contour** (a closed curve).

We will apply this theorem to $1 + P(s)C(s)$ (which can be seen as a complex function) and the contour will encircle the entire right half plane (= **the Nyquist contour**).

Outline

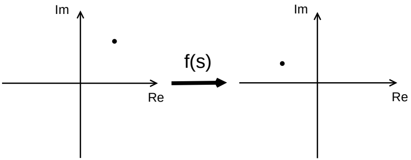
Outline

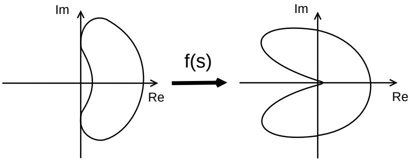
Complex function

Before we get to the theorem, we discuss the concept of a complex function.

A complex function $f(s) = u(x, y) + jv(x, y)$ maps the complex number $s = x + jy$ onto the complex number $w = u + jv$.

For a complex number: or for a contour:





The function $1 + P(s)C(s)$ can also be regarded as a complex function, so we can use it as a mapping.

Cauchy's argument principle

This is the engine behind the Nyquist stability criterion.

If a contour Γ in the s-plane encircles Z zeros and P poles of $f(s)$ in clockwise direction, the contour Γ', which is the image of Γ as mapped by $f(s)$, encircles the origin (in the w-plane) $Z - P$ times in the clockwise direction.

So the only thing we are looking at is the **number of encirclements** of the origin.
On the next slides, we will prove this.

Cauchy's argument principle

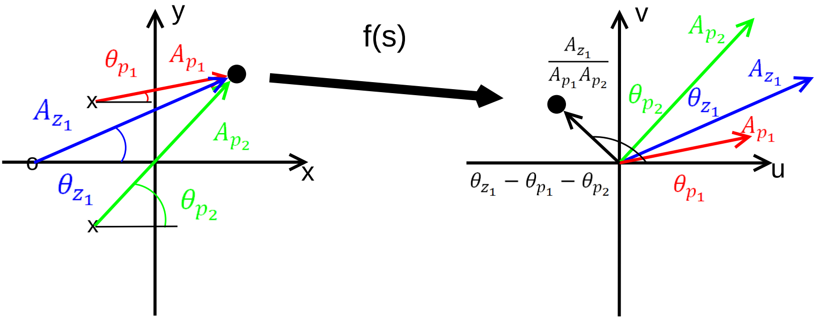
Let's take a complex function $f(s) = \frac{(s-z_1)(s-z_2)(s-z_3)\ldots}{(s-p_1)(s-p_2)(s-p_3)\ldots}$.

If we would apply this function to a complex number c, this comes down to multiplying factors $c - z_i = A_{z_i} e^{j\theta_{z_i}}$ and $\frac{1}{c-p_i} = \frac{1}{A_{p_i}} e^{-j\theta_{p_i}}$.

So the **modulus** of $f(c)$ can be easily found by evaluating $\frac{A_{z_1} A_{z_2} A_{z_3} \ldots}{A_{p_1} A_{p_2} A_{p_3} \ldots}$. This might help if you want to map a point, but it is not important for us.

The evaluation of the **argument** of $f(c)$ is what will be interesting: $\angle f(c) = \theta_{z_1} + \theta_{z_2} + \theta_{z_3} + \ldots - \theta_{p_1} - \theta_{p_2} - \theta_{p_3} - \ldots$

Cauchy's argument principle: graphically

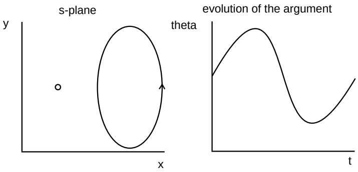


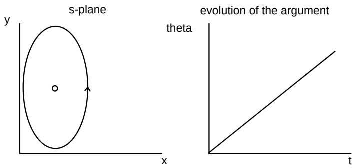
Cauchy's argument principle

When does the image of a contour in the w-plane encircle the origin?

That happens when the argument of the image of the contour at the beginning and the end differ by 2π.

A pole or zero outside the contour will never have that effect. One inside the contour has the following effect:





Cauchy's argument principle

A pole results in -2π (counterclockwise rotation), if the contour is followed clockwise.

A zero results in $+2\pi$ (clockwise rotation).

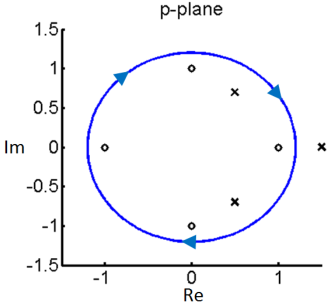
This follows from the sign of their effect:
$\angle f(c) = \theta_{z_1} + \theta_{z_2} + \theta_{z_3} + ... - \theta_{p_1} - \theta_{p_2} - \theta_{p_3} - ...$

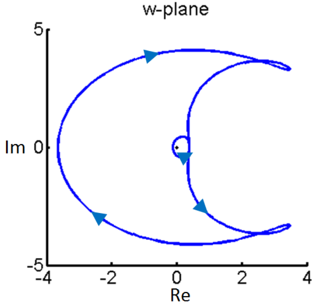
It is also possible that the origin is encircled when there are no poles or zeros in the contour (in the s-plane). But then the amount of clockwise encirclements equals the amount of counterclockwise encirclements, hence no net encirclements.

Outline

Example 1

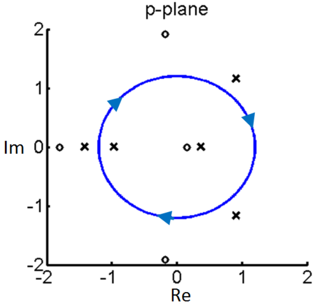
- Two encircled poles (the x's): $P = 2$
- Four encircled zeros (the o's): $Z = 4$
- Hence: $N = Z - P = 2$
- Indeed, the image of the contour encircles the origin twice in the w-plane (in clockwise direction).

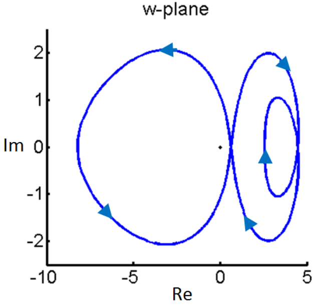




Example 2

- Two encircled poles: $P = 2$
- One encircled zero: $Z = 1$
- Hence: $N = Z - P = -1$
- Indeed, the image of the contour encircles the origin once in the w-plane (in counterclockwise direction).

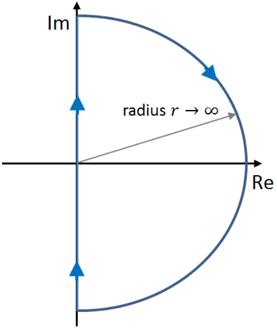




Outline

Nyquist stability criterion

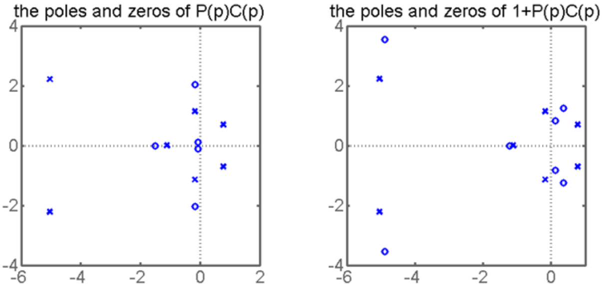
If we apply Cauchy's argument principle to the following contour (the **Nyquist contour**), the amount of clockwise encirclements around the origin of the mapping of this contour (the **Nyquist plot**) in the w-plane equals $Z - P$ of the right half plane (RHP).



Nyquist stability criterion

That way, we will find the difference between the number of poles and zeros of $1 + P(s)C(s)$ in the RHP: $N = Z - P$.

So if we know P (the number of poles in the RHP of $1 + P(s)C(s)$), we know how many roots of $1 + P(s)C(s)$ there are in the RHP.



Nyquist stability criterion

Luckily this last aspect is simple, since the roots of $1 + P(s)C(s)$ equal those of $P(s)C(s)$. Hence the amount of RHP poles is equal (the connection between the zeros is not as clear).

So if the number of RHP poles of $P(s)C(s)$ is known (which is assumed), we know whether the system with unity (negative) feedback is stable.

Nyquist stability criterion

If we apply this to $1 + P(s)C(s)$, we need to count the number of encirclements of the origin.
However, the Nyquist stability criterion uses $P(s)C(s)$.

- The zeros of $1 + P(s)C(s)$ and the poles and zeros of $P(s)C(s)$ are hard to relate.
- This is in sharp contrast with how easily the Nyquist plots relate: the Nyquist plot of $P(s)C(s)$ equals the one of $(1 + P(s)C(s)$, after it has been moved to the right over a distance 1.
- To find $Z - P$, one has to count the number of clockwise encirclements of the image of $P(s)C(s)$ around the point $(-1, 0)$, since this equals the number of clockwise encirclements of the image of $1 + P(s)C(s)$ around the origin.

Nyquist stability criterion

Nyquist stability criterion

If the open loop system $P(s)C(s)$ has ℓ poles in the right half plane (of the s-plane), then the system with unity feedback is stable if and only if the Nyquist plot encircles the point (-1,0) ℓ times in the counter clockwise direction.

Outline

Outline

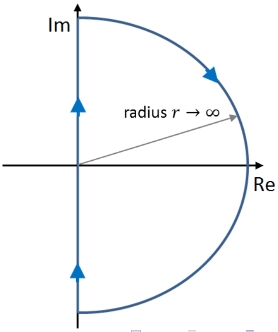
1. Introduction
2. Stability
3. Cauchy's argument principle
   - Cauchy's argument principle
   - Examples
4. Nyquist stability criterion
5. Nyquist plot
   - Nyquist plot
   - Poles on the imaginary axis
   - Discrete-time
6. Margins
   - Gain margin
   - Phase margin
   - Design example

Nyquist plot

We know how to deduce stability of the closed loop system from the Nyquist plot.
Now we will discuss how these plots can be found.

First some simplifications:

1. For physically realizable systems (i.e. relevant systems to engineers) the circle bow will be mapped on a point.
2. The image of the positive imaginary axis is the mirror image of the negative imaginary axis



Nyquist plot: physically realizable systems

1. **Every physically realizable system is causal**
   This is logical: you cannot build a system that knows the future
2. **Every causal system has a transfer function with a degree of the denominator that is larger than or equal to the degree of the numerator**
   Take for example the following transfer function:
   $$H(z) = \frac{b_2 z^2 + b_1 z + b_0}{a_0 z + a_1}$$
   The corresponding difference equation is:
   $$a_0 y_{n-1} + a_1 y_n = b_0 u_{n-1} + b_1 u_n + b_2 u_{n+2}$$
   It is non-causal, because the output depends on future input.

Nyquist plot: physically realizable systems

3. **From the transfer function, we can easily show that the circle bow maps onto a point.**
   - If the degree of the denominator is strictly higher than the degree of the numerator:
     If $s \to \infty$, then $P(s)C(s) \to 0$
   - If the degree of the denominator is equal to the degree of the numerator:
     If $s \to \infty$, then $P(s)C(s) \to c$, a real number

Nyquist plot: symmetry

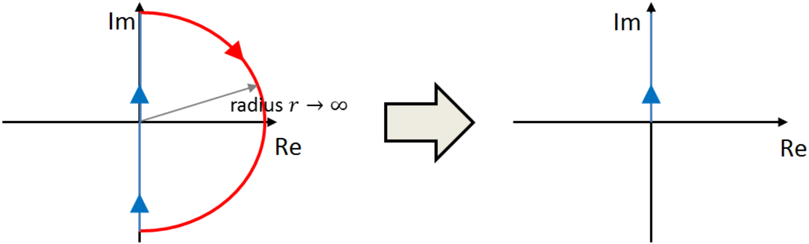
The symmetry follows directly from the way $f(c)$ can be evaluated.

Since the position of $f(c)$ only depends on the location of the poles and zeros, and those poles and zeros only occur symmetrically round the real axis, the Nyquist plot will be symmetrical round the real axis.

Nyquist plot

So we will only have to study the positive (or the negative) imaginary axis!

- The circular part maps onto one point, which is the same as where $j\omega$ maps onto
- The other half of the imaginary axis will give the mirror image of the studied half



Nyquist plot

Extracting the number of times $(-1, 0)$ is encircled isn't that difficult anymore:

- You search for the image of $j0^+$
- You search for the image of $j\infty$
- You search for the positive real y for which $f(jy)$'s imaginary part changes sign

This information allows you to determine if you encircle $(-1, 0)$

Let's visualize this with a simple example, but of course you can use software to do this (e.g. nyquist in Matlab)

Nyquist plot: a simple example

Let's take the following open loop system:

$$P(s)C(s) = \frac{1}{s^2 - 2s + 2}$$

Substitute s with $j\omega$

$$\frac{1}{-\omega^2 - 2j\omega + 2} = \frac{1}{-\omega^2 + 2 - 2j\omega} \frac{-\omega^2 + 2 + 2j\omega}{-\omega^2 + 2 + 2j\omega} = \frac{-\omega^2 + 2 + 2j\omega}{(-\omega^2 + 2)^2 + 4\omega^2}$$

- $f(j0^+) = \frac{1}{2}$
- $f(j\infty) = 0$
- Imaginary part: $\frac{2\omega}{(-\omega^2+2)^2+4\omega^2} = 0 \Rightarrow \omega = 0$

The real axis gets crossed 2 times: first at $\frac{1}{2}(\omega = 0^+)$ and then at $0(\omega = \infty) \to (-1, 0)$ is not encircled

Nyquist plot: a simple example

Remember our open loop system:

$$P(s)C(s) = \frac{1}{s^2 - 2s + 2}.$$

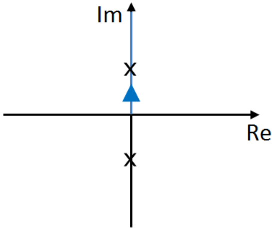
Z and P are respectively the number of zeros and poles of $1 + P(s)C(s)$.
The poles of $P(s)C(s)$ and $1 + P(s)C(s)$ are the same ($P = 2$ in our example).
Since $(-1, 0)$ is not encircled: $Z - P = 0$, hence there are 2 zeros in the right half plane.
Remember that the zeros of $1 + P(s)C(s)$ are the poles of the closed loop system and thus, the unity feedback controller is unstable.

Outline

Nyquist plot: poles on the imaginary axis

Poles on the imaginary axis: why are they a problem?

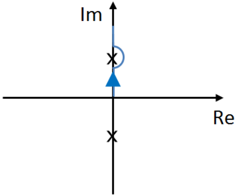
- Take for instance the case with one pair of imaginary poles at jc and $-jc$
- When coming close to jc, the argument will remain 0 and the gain will increase to infinity
- At jc itself, the gain will be infinite, but the argument is undetermined, hence we cannot map this point



Nyquist plot: poles on the imaginary axis

How do we solve this?

- Instead of going through the poles, we will evade them by an infinitesimally small amount (see figure)
- That way we do not have the problem of an undetermined mapping at the pole
- Since we avoid them by an infinitesimally small amount, we also know we will not wrongly avoid a pole that lies in the right half plane



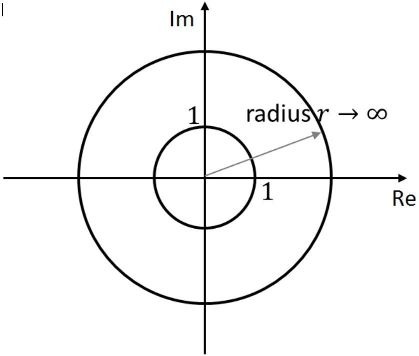
Now the Nyquist plot will go to infinity as the pole is approached, then the argument will change from 0 to π as the semi-circle is traversed and then the Nyquist plot will return from infinity.

Outline

Discrete-time case

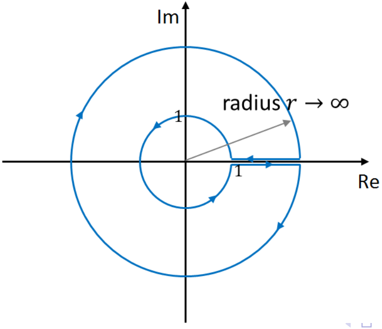
The Cauchy argument principle still applies for discrete-time systems, since $P(z)C(z)$ also has the shape of a rational polynomial.

The contour will have to encircle the entire complex plane except for the unity ci



Discrete-time case

However, this is not a contour, but it can be solved with the following trick:
the two horizontal pieces are both infinitely close to the real axis, that way they are identical but with opposite signs. They will cancel each other out



Outline

Margins

The Nyquist plot allows us to determine the stability of the system We know the stability changes when $1 + P(s)C(s)$ has an imaginary root (then the system is marginally stable).

We can see such a root in the Nyquist plot of $P(s)C(s)$. After all, the Nyquist plot is the image of the imaginary axis, so if there is a root on the imaginary axis, the Nyquist plot of $1 + P(s)C(s)$ would go through 0 and the Nyquist plot of $P(s)C(s)$ would go through -1.

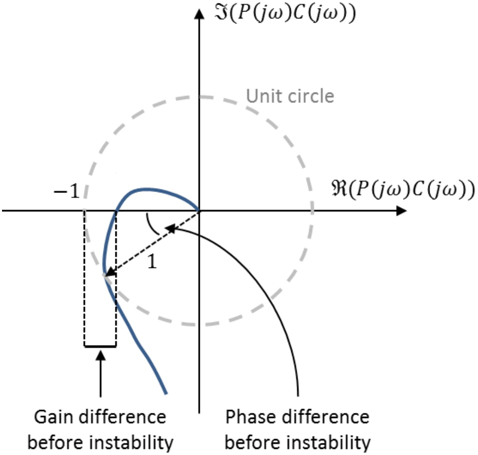
Margins

This explains why we can use the 'distance to -1' as a measure of stability.

We will see two different stability margins, which can be easily read off the Nyquist diagram:

- **The gain margin**: the amount of extra gain you can allow before instability occurs (in dB)
- **The phase margin**: the amount of phase you can add before instability occurs

Margins

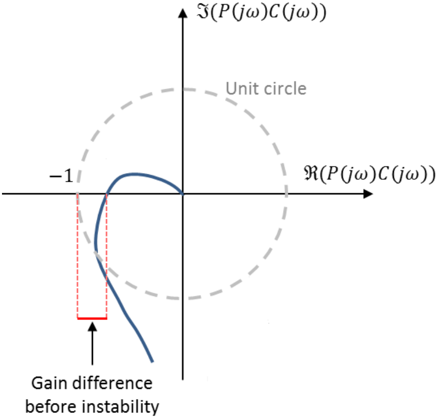

$\Im(P(j\omega)C(j\omega))$
Unit circle
-1
$\Re(P(j\omega)C(j\omega))$
1
Gain difference before instability
Phase difference before instability

Outline

Gain margin

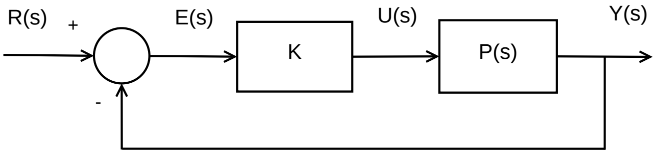
The gain margin is the amount of extra gain allowed before the system becomes unstable (or how much larger the gain has to be, before the system becomes unstable).

The gain margin is **multiplicative**, so it is the factor you have to multiply the gain with, so the Nyquist plot goes through -1 in the w-plane.
It will be expressed in dB ($20log_{10}K$).



Gain margin

Let's look at it the following way:



The stability margin of $P(s)$ with unity feedback is the K for which the system above is marginally stable.

Gain margin

So $KP(s)$ should equal -1 for an imaginary $s = j\omega$.

This requires $\angle(KP(j\omega)) = \angle P(j\omega)$ to equal -180°. This ω_π is called the **Gain Crossover Frequency (GCF)**.
K then has to be set such that $\left|KP(j\omega_\pi)\right| = 1$.

So a large gain can lead to instability and this risk only exists when there exists a ω_π for which $\angle P(j\omega) = -180^\circ$.

We will illustrate this with an example.

Gain margin: example

Consider the process

$$P(s) = \frac{1}{s(s+2)}.$$

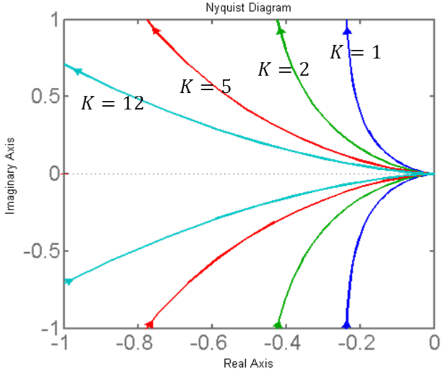
We derive the argument as a function of ω:

$$\angle P(j\omega) = -tan^{-1}(\frac{\omega}{0}) - tan^{-1}(\frac{\omega}{2}) = -90^\circ - tan^{-1}(\frac{\omega}{2}).$$

For this to equal to -180°, it requires $\omega_\pi \to \infty$.
This makes $P(j\omega) = 0$ and this means the gain margin is infinite.

Gain margin: example

This solution is shown in the Nyquist plot below:

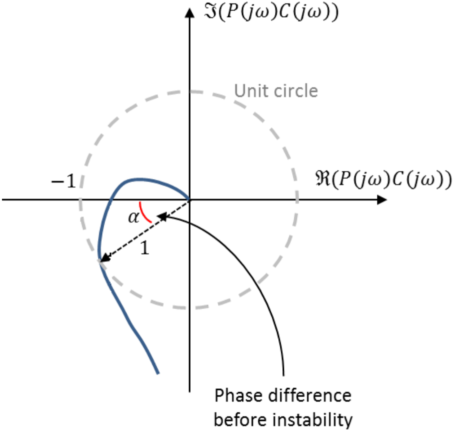


The only crossover with the real axis occurs at $\omega = 0$ and that does not change with increasing K.

Outline

Phase margin

The phase margin is that amount of additional phase lag to bring the system to the verge of instability.



Phase margin

This can be interpreted as multiplying $P(s)$ with $e^{j\theta}$ until -1 is crossed, so $P(s)e^{j\theta} = -1$ for an imaginary $s = j\omega$.

This requires $\left|P(j\omega)e^{j\theta}\right| = \left|P(j\omega)\right|$ to equal 1. This ω_0 is called the **Phase Crossover Frequency (PCF)**.
θ then has to be set such that

$$\angle(P(s)e^{j\theta}) = \angle P(s) + \theta = -180^\circ.$$

The gain margin is defined as positive, but that doesn't matter, because of the symmetry with respect to the real axis.
If a rotation of θ degrees results in a crossing of -1, then a rotation of $-\theta$ does the same.

Phase margin: example

Consider the process

$$P(s) = \frac{1}{s(s+2)}.$$

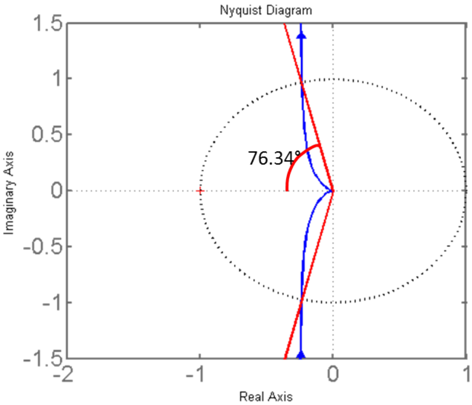
We derive the modulus as a function of ω:

$$\left|P(j\omega)\right| = \frac{1}{\omega}\frac{1}{\sqrt{\omega^2+4}}.$$

For this to equal to 1, we find $\sqrt{\omega^4+4\omega^2} = 1$ or $\omega_0 = \sqrt{\sqrt{5}-2} = 0.486$.
Now we need to find θ so $\angle(P(j\omega_0)e^{j\theta}) = -180^\circ$.
This results in $\theta = -180^\circ + tan^{-1}(\frac{\omega_0}{0}) + tan^{-1}(\frac{\omega_0}{2}) = -76.34^\circ$.
This means the phase margin is 76.34°.

Phase margin: example

This solution is shown graphically in the Nyquist plot below:



What should the margins be?

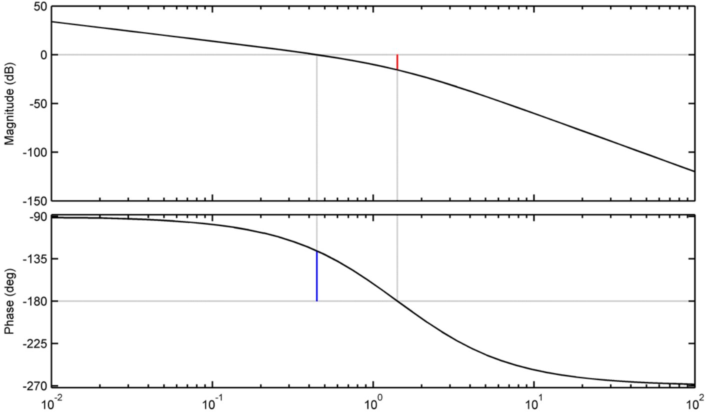
Phase margin:

- This is more subtle than the gain margin
- If it is too small, instability might occur due to practicalities, the model is not perfect
- If it is too small, we get large overshoots and large oscillations that fade away very slowly
- Sometimes a good value is 60°, but it is highly case-dependent

A good margin does not offer certainty about the stability, whereas a bad phase margin (very large or very small) does give certainty about instability.

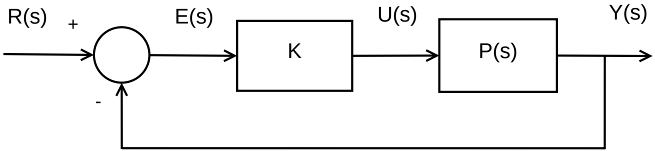
Margins using Bode plots

We can also easily derive the gain margin and phase margin from the Bode plot of P(s)C(s):



Outline

Margins: design example



Consider the following process

$$P(s) = \frac{1-s}{(1+s)^2}.$$

Design a simple proportional controller K with a phase margin of 45°.

Margins: design example

We find ω in the following way:

$$\begin{aligned}
\angle(P(j\omega)e^{\frac{j\pi}{4}}) &= -180^\circ \\
45^\circ + tan^{-1}(-\omega) - tan^{-1}(\omega) - tan^{-1}(\omega) &= -180^\circ \\
-3tan^{-1}(\omega) &= -225^\circ \\
\omega &= tan(75^\circ) = 3.73
\end{aligned}$$

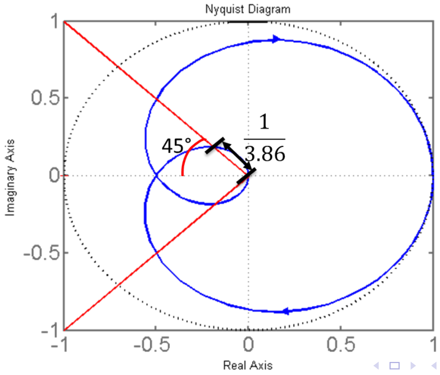
Now we can find K (which doesn't influence the argument) by setting $\left|KP(j\omega)\right| = 1$:

$$K\frac{\sqrt{\omega^2+1}}{\sqrt{\omega^2+1}^2} = \frac{K}{\sqrt{\omega^2+1}} = \frac{K}{3.86} = 1 \rightarrow K = 3.86 = 5.87dB$$

Margins: design example

We can also do this graphically with the Nyuist plot.

First determine the point that corresponds to $\theta = 45°$.
Then determine the modulus, K is the inverse of that modulus.



Summary

- The Nyquist stability criterion came to existence as a cheap alternative to determine stability of a closed loop system with unity feedback
- It also shows the phase margin and the gain margin, which are used to measure the stability of the system
- It is relevant as a design tool, as shown in the last example