Design in the frequency domain Nyquist stability criterion

Van Assche Jonah

July 8, 2015

Lay-out

- Moving to the frequency domain
- Nyquist stability criterion: what and why?
- Stability of the closed loop system
- Complex function
- Cauchy's argument principle
- Nyquist stability criterion and Nyquist plots
- The Nyquist criterion in the discrete time case
- Stability margins
- Let's recapitulate

Moving to the frequency domain

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

- Now we go to the frequency domain: $s = j\omega$ which means we will only regard perfect oscillations

Moving to the frequency domain

- This cannot be seen apart from the fact that sines, cosines and exponential signals are eigenfunctions of Linear Time Invariant (LTI) systems
- Perfect oscillations form the natural decomposition of each signal when you are dealing with LTI's

Design criteria

- Like before, we need to translate our design criteria into a language that fits the discussed method
- For the root-locus method that meant that we had to express our criteria in positions of poles and zeros
- Typical design criteria in the frequency domain are:
  - Phase and gain margin: their meaning will be discussed thoroughly in the section about the Nyquist stability criterion
  - Bandwidth: this will be discussed when we tackle Bode plots
  - Zero-frequency magnitude ($=$ DC gain)

Design tools

- We will use two different graphical representations to design compensators in the frequency domain:
  - First we will introduce the Nyquist stability diagram and Nyquist plots
  - When we discuss the design of lead, lag and lead-lag compensators we will use Bode plots (frequency domain design tool)

Nyquist stability criterion: what and why?

- What?

  The Nyquist stability criterion is a way of determining the stability of a closed loop linear time-invariant system based on the open loop system

- Why is it useful?

  It is cheap to compute It allows to determine the phase and gain margins

- Do we have what it takes?

  In order to use it, you need to know the number of right half plane poles of the open loop system; which you usually know The calculations simplify significantly for a physically realizable system

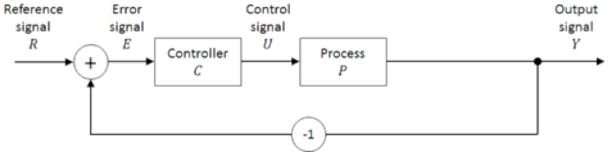
Nyquist stability criterion: what and why?

- It was discovered in 1932 in order to have a cheaper method of determining stability (back then finding the roots of $1 + P(s)C(s)$ was a very expensive endeavor)

- Now it is still relevant as a design tool

Stability of the closed loop system

- 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 $1 + P(s)C(s) = 0$



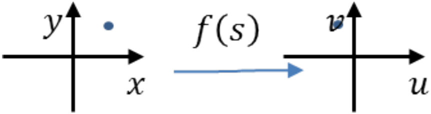
- In the root locus method we determined the position of the poles by plotting the roots of $1 + P(s)C(s) = 0$ thus finding whether it is stable by checking whether all the roots remain in the left half plane
- Solving $1 + P(s)C(s) = 0$ was very expensive 50 years ago
- We are not interested in the position of the poles. The only thing we need to know is whether there are poles in the right half plane or not. We can find a cheaper alternative, the Nyquist stability criterion

Stability of the closed loop system: Nyquistcriterion

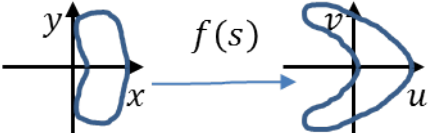
- The Nyquist stability criterion avoids determining the roots of $1 + P(s)C(s)$ exactly, it only determines the number of roots in the right half plane
- It uses a theorem from complex calculus that finds the difference between the number of poles and the number of zeros within a contour (a closed curve)
- 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 Nyquistcontour)

Complex function

- Before we get to the theorem that allows us to determine $Z - P$ (the number of zeros minus the number of poles within a contour), we have to show what a complex function is
- A complex function $f(s) = u(x, y) + jv(x, y)$ maps the complex number $s = x + jy$ onto the complex number $w = u + jv$



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

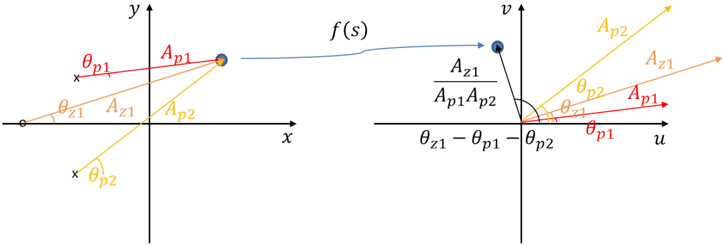
- This is the engine behind the Nyquist stability criterion
- If a contour Γ in the s-plane circles Z zeros and P poles of $f(s)$ in clockwise (CW) rotation, then the contour Γ', which is the image of Γ as mapped by $f(s)$, circles the origin (in the w-plane) $Z - P$ times in the clockwise (CW) direction
- So the only thing we are looking at is the encirclements of the origin, all the other wiggles dont matter to us
- On the next slides we will prove this
- If you want to know more about this and related topics, we refer you to the course 'Complexefunctieleer1stmaster WIT'

Cauchys argument principle

- Lets take a complex function $f(s) = \frac{(s-z_1)(s-z_2)(s-z_3)...}{(s-p_1)(s-p_2)(s-p_3)...}$
- So if we would apply this function to a complex number c then this comes down to multiplying factors $c - z_1 = A_{z_1} e^{j\theta_{z_1}}$ and $\frac{1}{c-p_1} = \frac{1}{A_{p_1}} e^{-j\theta_{p_1}}$
- So the modulus of $f(c)$ can be easily found by evaluating $\frac{A_{z_1} A_{z_2} A_{z_3} \cdots}{A_{p_1} A_{p_2} A_{p_3} \cdots}$;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; it is also very easy to find: $\angle f(c) = \theta_{z_1} + \theta_{z_2} + \theta_{z_3} + ... - \theta_{p_1} - \theta_{p_2} - \theta_{p_3} - ...$

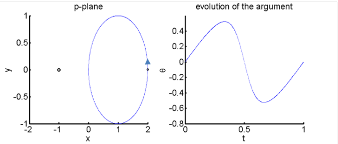
Cauchys argument principle

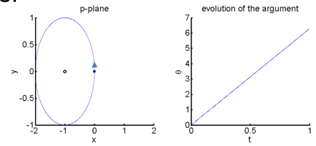
Lets first show this graphically:



Cauchys argument principle

- When does the image of a contour in the w-plane encircle the origin?
- That happens when the image of the beginning and the end of the contour their argument differs by 2π
- A pole/zero outside the contour will never have that effect and one inside the contour does:

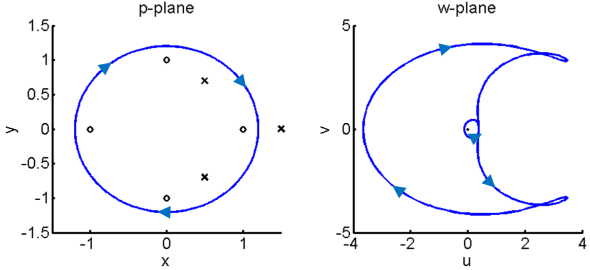




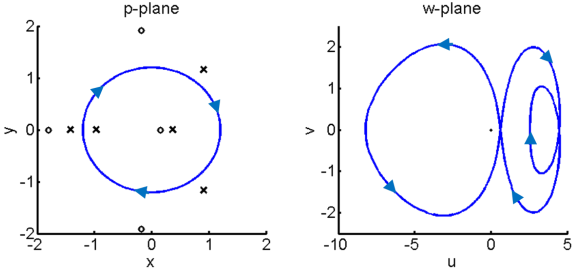
Cauchys argument principle

- A pole results in -2π (CCW rotation), if the contour is followed CW, and a zero results in $+2\pi$ (CW rotation) follows from the sign of their effect:
  $\angle f(c) = \theta_{z_1} + \theta_{z_2} + \theta_{z_3} + ... - \theta_{p_1} - \theta_{p_1} - \theta_{p_1} - ...$
- Watch out: 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 CW encirclements equals the amount of CCW encirclements hence no net encirclements

- 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 the CW direction)

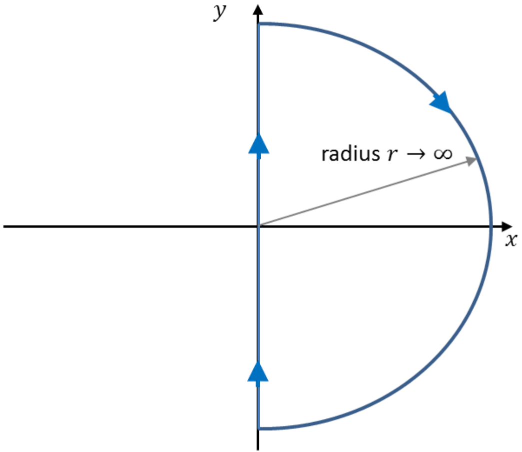


- Two encircled poles (the x' s): $P = 2$
- Four encircled zeros (the o' s): $Z = 1$
- Hence: $N = Z - P = -1$
- Indeed, the image of the contour encircles the origin once in the w-plane (in the CCW direction)



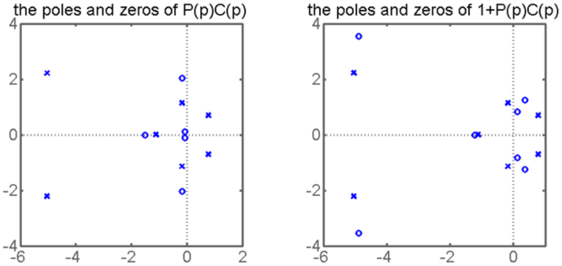
Nyquist stability criterion

If we apply Cauchys argument principle to the following contour (the Nyquistcontour), the amount of CW encirclements around the origin of the image of this contour (called the Nyquist plot) in the w-plane equals $Z - P$ of the 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 now know P (the number of poles in the RHP of $1 + P(s)C(s)$), we can know whether (and how many) zeros of $1 + P(s)C(s)$ are in the RHP



- Luckily this last aspect is simple; since the poles 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 clear)
- So if we assume the number of RHP poles of $P(s)C(s)$ is known (which we'll assume), we can know whether the system with unity feedback is stable

Nyquist stability criterion

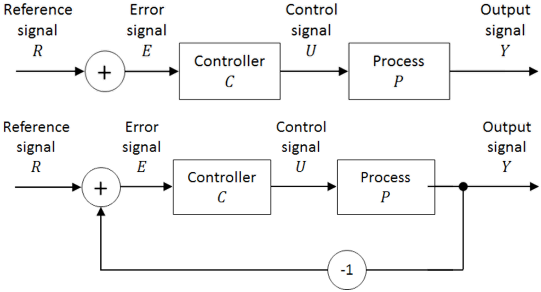
So we could apply this to $1 + P(s)C(s)$, then the encirclements of the origin counts indeed

But the Nyquist stability criterion applies this to $P(s)C(s)$

- We saw that 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 plots of $P(s)C(s)$ equals the one of $1 + P(s)C(s)$, after he has been moved 1 to the right
- So to find $Z - P$ one has to count the number of CW encirclements of the image of $P(s)C(s)$ around $(-1,0)$ since this equals the number of CW encirclements of the image of $1 + P(s)C(s)$ around the origin.

Nyquist stability criterion:

If the open loop system $P(s)C(s)$ has l-poles in the open 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



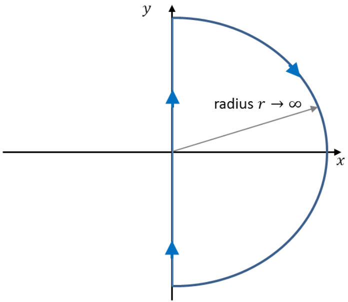
Nyquist plot

We now found how to deduce stability of the closed loop system from the Nyquist plot

Now we will discuss how these plots can be found

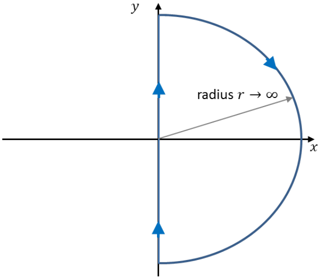
But first some simplifications:

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



Nyquist plot: physically realizable systems

- Where does the very convenient property of physically realizable systems come from?
- The reason is:
  1. Every physically realizable system is causal
  2. Every causal system has a transfer function with an order of the denominator that is larger than or equal to the order of the numerator
  3. From this shape of the transfer function we can easily show that the circle bow maps onto a point



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 an order of the denominator that is larger than or equal to the order of the numerator
   This can be easily understood in the discrete time case:
   $a_0 y_n + a_1 y_{n-1} + a_2 y_{n-2} + \ldots = b_o u_n + b_1 u_{n-1} + b_2 u_{n-2} + \ldots$
   Apply the z-transformation:
   $\frac{Y(z)}{U(z)} = H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots}$

Nyquist plot: physically realizable systems

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

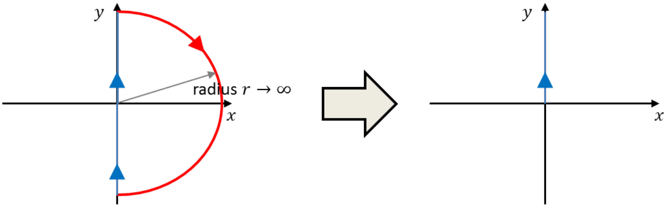
Nyquist plot: symmetry

- The symmetry follows directly from the way $f(c)$ can be evaluated (see previously)
- 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 circled now isn't that difficult anymore:
  - You search for the image of $j0^+$
  - You search for the image $j\infty$ you serach for the positive real y for which f(jy)'s imaginary part changes sign
- This information allows you to determine if you encircle $(-1{,}0)$
- Lets show 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

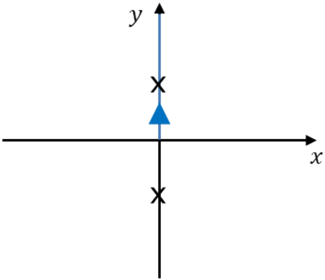
- Lets take the following open loop system:
  $P(s)C(s) = \frac{1}{s^2-2s+2}$
- Fill in
  $s = j\omega \rightarrow \frac{1}{-\omega^2-2j\omega+2} = \frac{1}{-\omega^2+2-2j\omega}\frac{-\omega^2+2+2j\omega}{-\omega^2+2+2j\omega} = \frac{-\omega^2+2+2j\omega}{(-\omega^2+2)^2+4\omega^2}$
  - $f(j0^+) = \frac{1}{2}$
  - $f(j\infty) = 0$
  - Imaginary part: $\frac{2\omega}{(-\omega^2+2)^2+4\omega^2} = 0 \Rightarrow \omega = 0$
- The real axis get's crossed 2 times; first at $\frac{1}{2}$ $(\omega = 0^+)$ an then at 0 $(\omega = \infty)\rightarrow (-1{,}0)$ 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 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

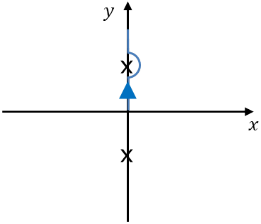
Nyquist plot: poles on the imaginary axis

- There is one more thing to know about Nyquist plots: how to deal with 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

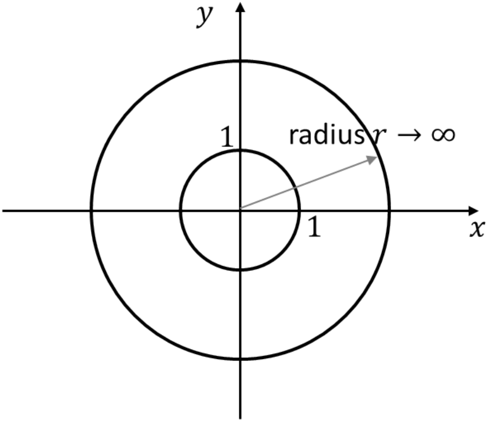
- How do you solve this?
  - Instead of going through the poles, we will evade them by an infinitesimally small amount (see figure)
  - That way we do not have the problem of an undetermined mapping at the pole
  - Since we avoid them by an infinitesimally small amount we also know we will not wrongly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



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

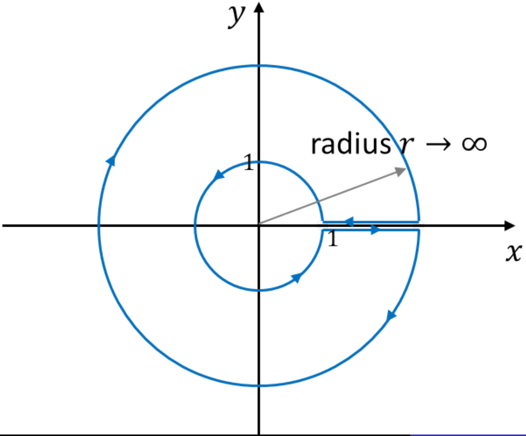
The Nyquist criterion in the discrete time case

- The Cauchy argument principle still applies, since $P(z)C(z)$ also has the shape of a rational polynomial
- The contour now will have to encircle the entire complex plane except for the unity circle



The Nyquist criterion in the discrete time case

This is not a contour however, but this 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.

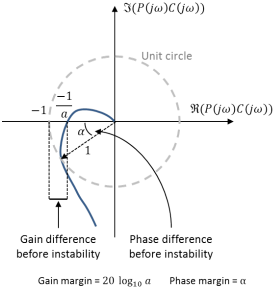


Margins

- The Nyquist plot allows us to determine how stable the system will be
- We know that the stability changes when $1 + P(s)C(s)$ has an imaginary zero (then the system is marginally stable)
- Can we see such a zero in the Nyquist plot of $P(s)C(s)$?
- Of course: the Nyquist plot is the image of the imaginary axis, so if there is a zero on the imaginary axis, the Nyquist plot of $1 + P(s)C(s)$ would go through 0 and the Nyquist plot of $P(s)C(s)$ will go through -1
- Hence the system is marginally stable if the Nyquist plot goes through -1

Margins

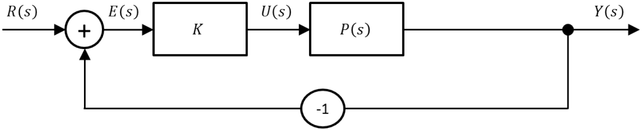
- This explains why we can use the 'distance to -1' as a measure of stability
- We will see two different stability margins, which can be easily read off the Nyquist diagram:
  - The gain margin: the amount of extra gain you can allow before instability occurs (in dB)
  - The phase margin: the amount of phase you can add before instability occurs



Margins: 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 with which you have to multiply the gain so the Nyquist plot goes through -1 in the
  w-plane
- It will be expressed in dB; take K a certain factor, then this can be expressed in dB: $20.log_{10}K$ dB

- Lets look at this the following way:



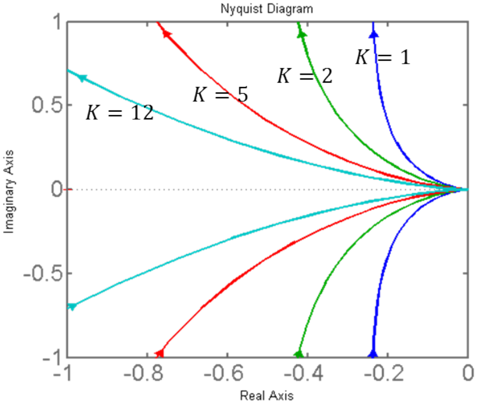
- The stability margin of $P(s)$ with unity feedback is the K for which the system above is marginally stable

Margins: 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 $\omega_\pi$is called the gain crossover frequency (GCF)
- K then has to be set such that $|KP(j\omega_\pi)| = 1$
- So now you should be able to understand that a large gain can lead to instability; and that this risk only exists when there exists a ω_π for which $\angle P(j\omega) = -180$
- We will illustrate this with a few examples

- Take $P(s) = \frac{1}{s(s+2)}$
- $\angle P(j\omega) = -tan^{-1}(\frac{\omega}{0}) - tan^{-1}(\frac{\omega}{2}) = -90^\circ - tan^{-1}(\frac{\omega}{2})$
- So when is this equal to -180° ?
- This would require $\omega \to \infty$
  which makes $P(j\omega) = 0$
- So the gain margin is infinite in this case

- We will show this again in the Nyquist plot:



- The only crossover with the real axis is at $\omega = 0$ and that doesn't change with increasing K

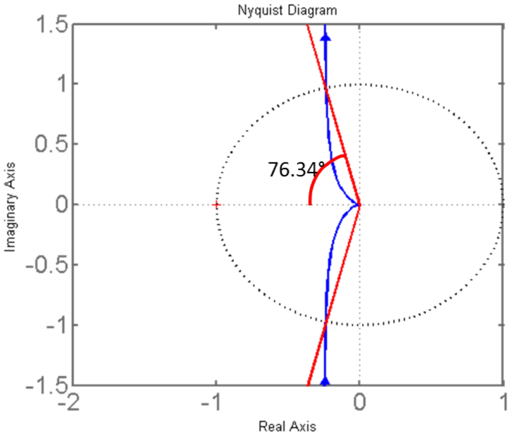
Margins: phase margin

- The phase margin is how much you can rotate the Nyquist diagram before -1 is crossed
- 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 $|P(j\omega)e^{j\theta}| = |P(j\omega)|$ to equal 1; this ω_0 is called the Phase Crossover Frequency (PCF)
- $|\theta|$ 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 doesnt 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

Margins: phase margin

- Lets take $P(s) = \frac{1}{s(s+2)}$ again
- When is $|P(j\omega)| = 1$?
  $\frac{1}{\omega} \cdot \frac{1}{\sqrt{\omega^2+4}} = 1$, hence $\sqrt{\omega^4 + 4\omega^2} = 1$, or
  $\omega = \sqrt{\sqrt{5} - 2} = 0.486$
- Find θ so $\angle(P(j\omega)e^{j\theta}) = -180^\circ$
- $\theta = -180^\circ + tan^{-1}(\frac{\omega}{0}) + tan^{-1}(\frac{\omega}{2}) = -76.34^\circ$
  The phase margin is 76.34°

Well show this graphically with the Nyquist plot:

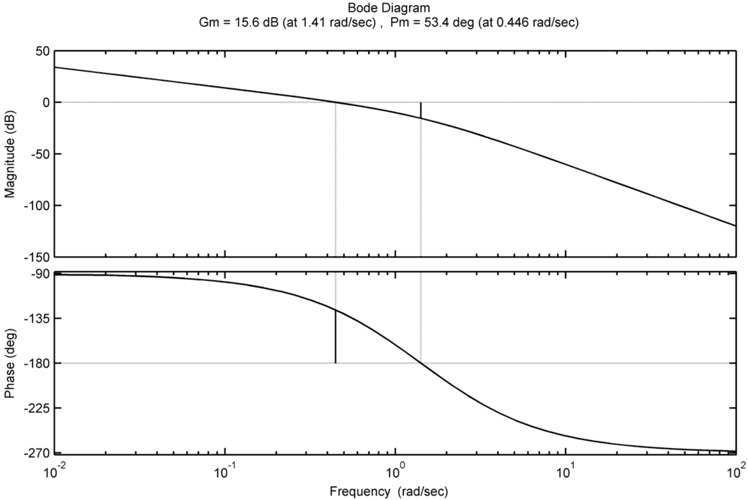


Phase margin:

- This is more subtle than the gain margin
- If it is too small instability might occur due to practicalities
- If it is too small we get large overshoots and large oscillations that fade away very slowly(which is not something we associate with a good form of stability)
- Sometimes a good value is 60°, but it is highly case dependent

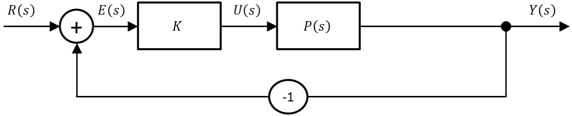
A good margin does not offer certainty about the stability; whereas a bad phase margin (very large or very small) does give certainty about instability

We can also easily derive the gain and phase margins from the Bode plots of $P(s)C(s)$:



Margins: design example

- Well design a simple proportional controller for the system $P(s) = \frac{1-s}{(1+s)^2}$



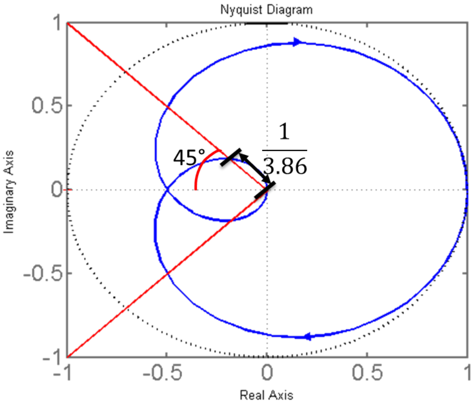
- What will K have to be in order to have a phase margin of 45° ?

Margins: design example

- We find ω by setting $\angle(P(j\omega)e^{\frac{j\pi}{4}}) = -180^\circ$
  - $45^\circ + tan^{-1}(-\omega) - tan^{-1}(\omega) - tan^{-1}(\omega) = -180^\circ$
  - $-3tan^{-1}(\omega) = -225^\circ$
  - $\omega = tan(75^\circ) = 3.73$
- Now we can find K (which doesn't influence the argument) by setting $|KP(j\omega)| = 1$:
  $K\frac{\sqrt{\omega^2+1}}{\sqrt{\omega^2+1}^2} = \frac{K}{\sqrt{\omega^2+1}} = \frac{K}{3.86} = 1 \rightarrow K = 3.86 = 5.87\text{dB}$

Margins: design example

- We can do this graphically with the Nyquist plot
- First determine the point that corresponds to $\theta = 45^\circ$
- Then determine the modulus;
  K is then equal to the inverse of that modulus



Let's recapitulate

- The Nyquist stability criterion came to existence as a cheap alternative to determine stability of a closed loop system with unity feedback
- It also allows to see the phase margin and the gain margin; which are used to measure how stable the system actually is
- Its relevance today is as a design tool, as shown in the final example
- In the next section we will discuss new kinds of classical control components
- Bode plots make up a similar tool, yet the Nyquist stability criterion is (slightly) more generally applicable

Interesting links

- About the Nyquist argument principle:
  - https://www.youtube.com/watch?v=sof3meN96MA&index=3&list=PLEJyHgf7y0ie3nwSzMDVzAdxf4dK9m89a
  - https://www.youtube.com/watch?v=tsg0stfoNhk
- About margins:
  - https://www.youtube.com/watch?v=Hw_hrxsX4_M
  - https://www.youtube.com/watch?v=BTNZ8SRs7Y8
  - https://www.youtube.com/watch?v=V1mfy7VhJNQ