Introduction to control theory

WEEK 3

Bode and Nyquist plots

The first part of this course (from the middle of WEEK 1 up to WEEK 7) is devoted to classical transfer function methods in control, that were developed in the period 1930 – 1960.

Recall that if Σ is a stable system with transfer function G and its input function is

$u(t) = \cos \omega t$, then its output function

is $y(t) = A_\omega \cos(\omega t + \varphi_\omega) + e(t)$, where the transient response $e(t) \to 0$ and

$$A_\omega = |G(i\omega)| \text{ is the gain at } \omega,$$

$$\varphi_\omega = \arg G(i\omega) \text{ is the phase shift at } \omega.$$

Gains and phase shifts are important in control engineering, and hence the need for a graphical representation of the two functions A_ω and φ_ω. When this is done with a logarithmic ω-axis, with the gain represented logarithmically by $20 \log A_\omega$ and the phase represented linearly in degrees, the two plots together are called the Bode plots of G. The logarithmic scales are suitable for looking at several frequency ranges in one glance, and they also make the drawing of the plots easier, because of the appearance of linear asymptotes, as we shall see.

Any rational SISO transfer function G can be factored as

$$G(s) = F_1(s) \dots F_N(s)$$

where

$$F_j(s) = \begin{cases} k & k \in \mathbb{R} \\ s^{\pm 1} & \\ \left(1 + \frac{s}{\omega_c}\right)^{\pm 1} & \omega_c \in \mathbb{R} \\ \left(1 + \frac{2\xi}{\omega_n} s + \frac{1}{\omega_n^2} s^2\right)^{\pm 1} & \omega_n > 0 \\ & -1 < \xi < 1 \end{cases}$$

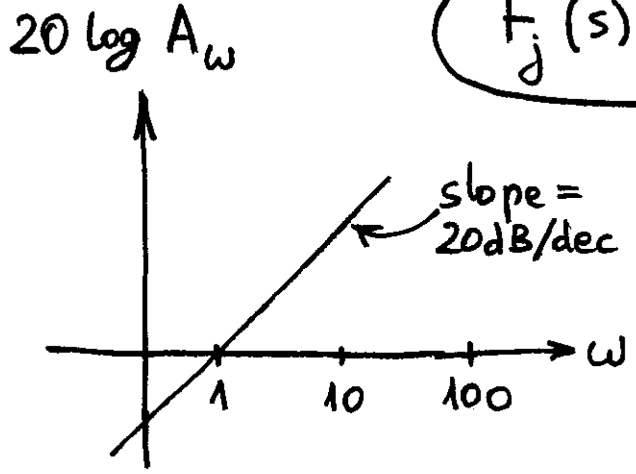
If we know how to draw the Bode plots for these four types of factors, we can draw the plots for G by adding the individual plots. Indeed,

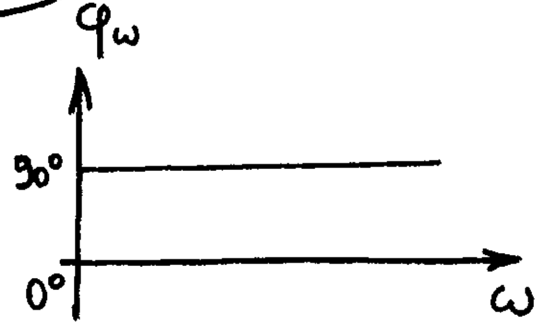
$$\log |G(i\omega)| = \log |F_1(i\omega)| + \dots + \log |F_N(i\omega)|,$$

$$\arg G(i\omega) = \arg F_1(i\omega) + \dots + \arg F_N(i\omega).$$

The plots for the factor k are trivial. We look at the other three types of factors.

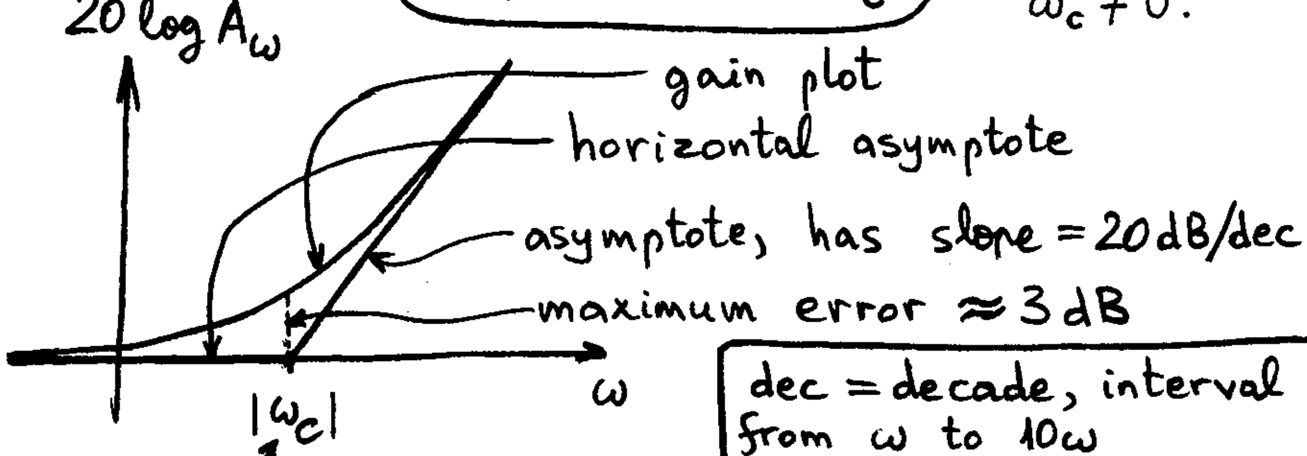
$$F_j(s) = s$$





$$F_j(s) = 1 + \frac{s}{\omega_c}$$

$\omega_c \in \mathbb{R}$, $\omega_c \neq 0$.

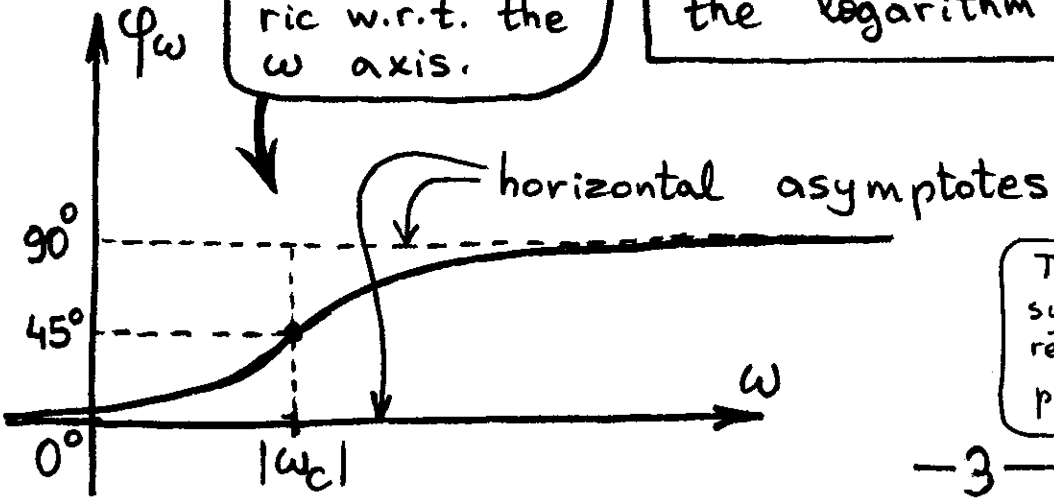


dec = decade, interval from ω to 10ω

"corner frequency", meeting point of the two asymptotes

dB ("decibel") is not a physical unit, it just indicates that we are dealing with 20 times the logarithm (in basis 10)

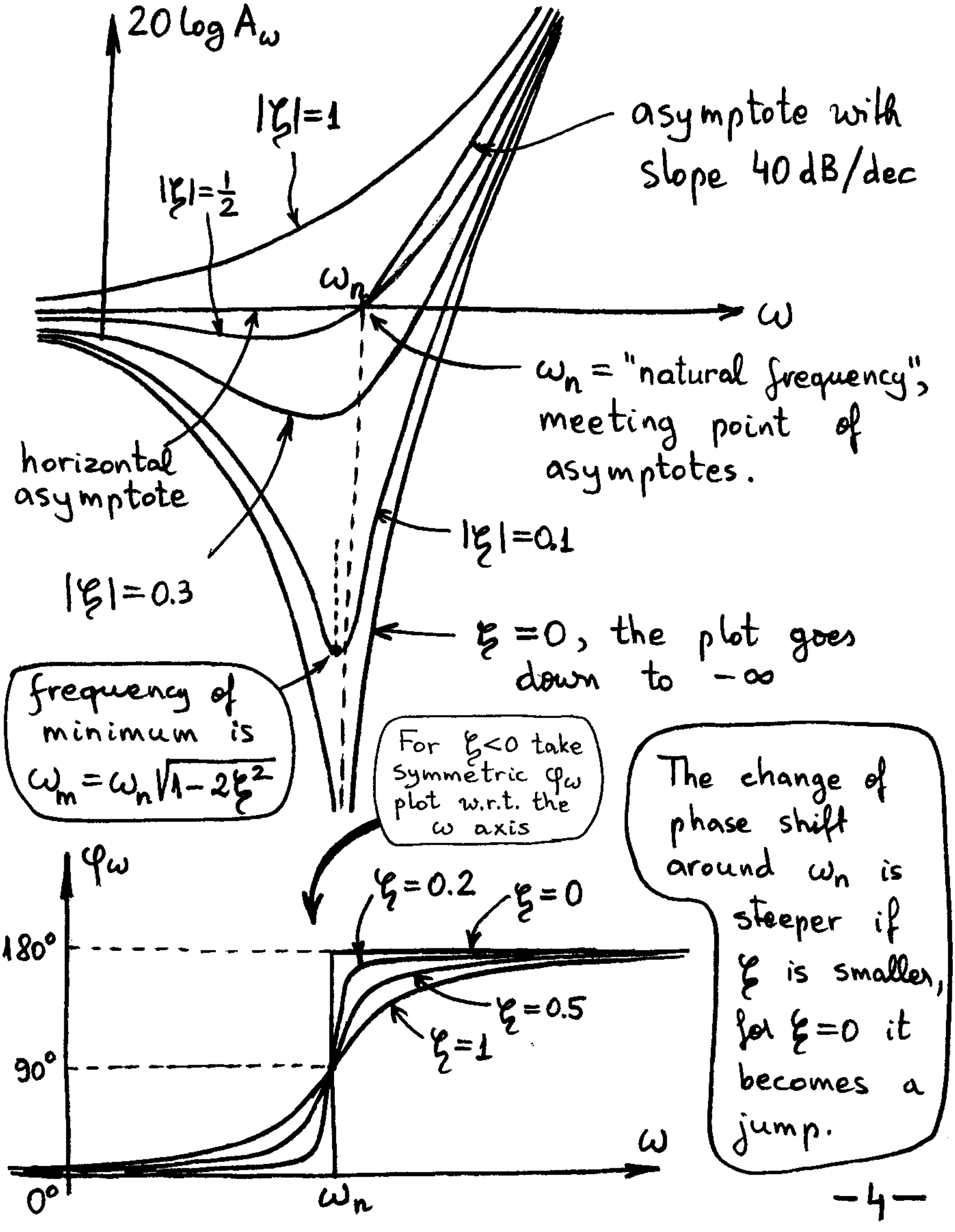
This plot corresponds to $\omega_c > 0$. For $\omega_c < 0$ it is symmetric w.r.t. the ω axis.



This plot is symmetric with respect to the point $(|\omega_c|, 45°)$.

$$\boxed{F_j(s) = 1 + \frac{2\xi}{\omega_n}s + \frac{1}{\omega_n^2}s^2}$$

This corresponds to two complex conjugate poles $p, \bar{p}$ such that $|p| = |\bar{p}| = \omega_n$ and $\xi = -\mathrm{Re}\,p/\omega_n$.

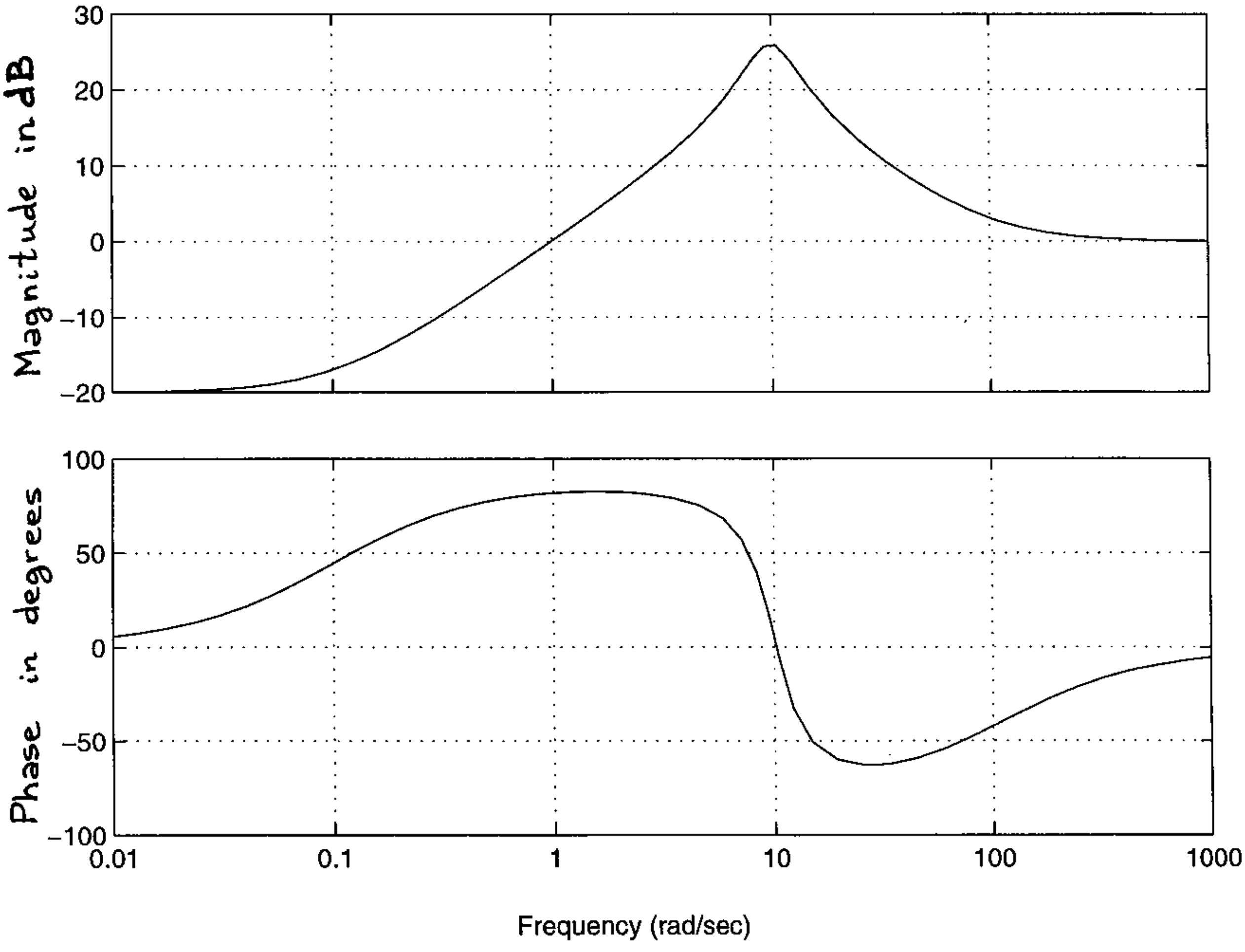


Example: Below we see the Bode plots of

$$G(s) = \frac{s^2 + 100.1\, s + 10}{s^2 + 5s + 100} = \frac{(s+0.1)(s+100)}{s^2 + 2\frac{1}{4}10s + 10^2}$$

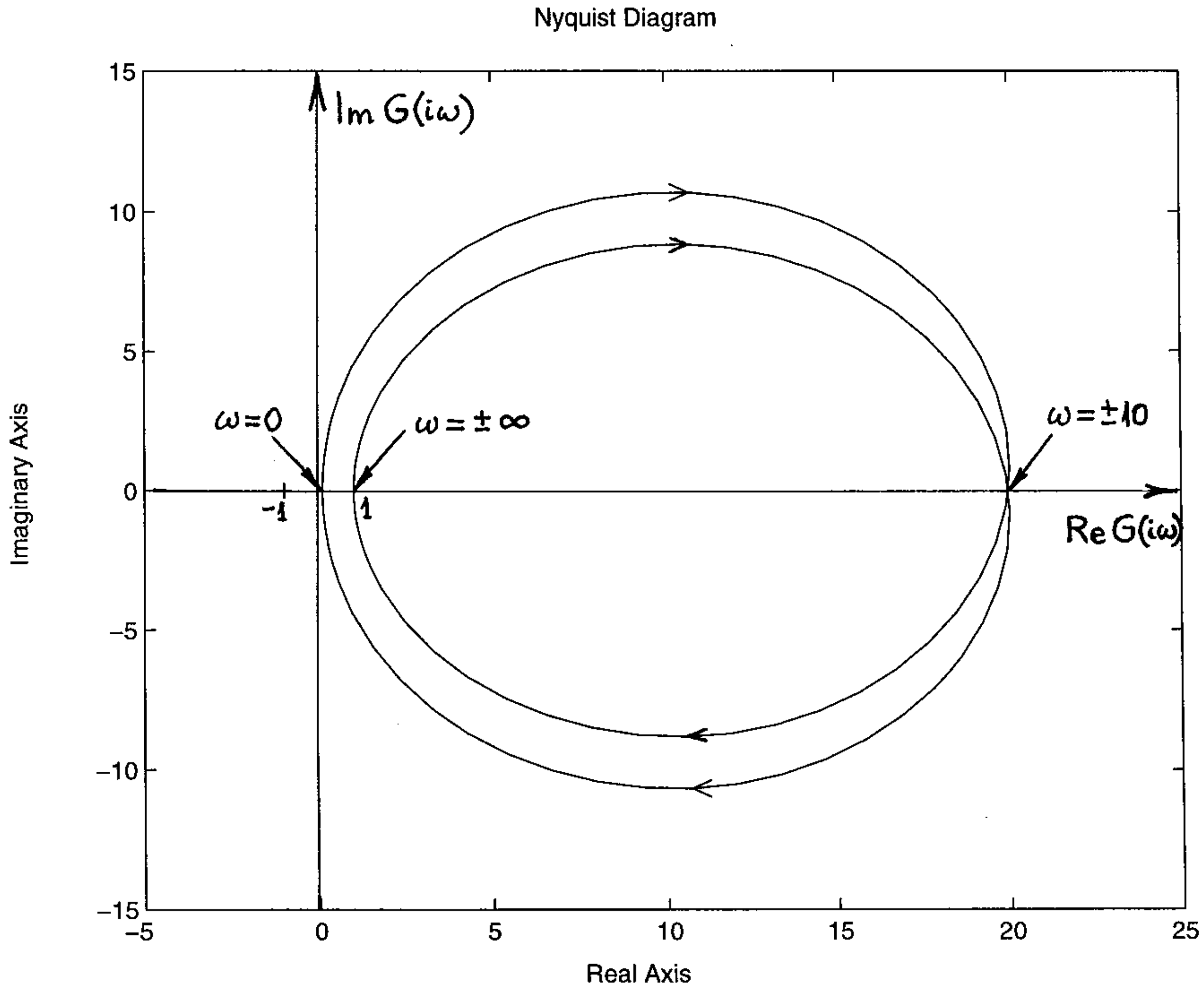
$$= 0.1 \cdot \frac{\left(1+\frac{s}{0.1}\right)\left(1+\frac{s}{100}\right)}{1 + \frac{2\cdot 0.25}{10}s + \frac{1}{10^2}s^2},$$ as drawn by MATLAB.

We can recognize in the plots the corner frequencies $\omega_1 = 0.1$, $\omega_2 = 100$ and the natural



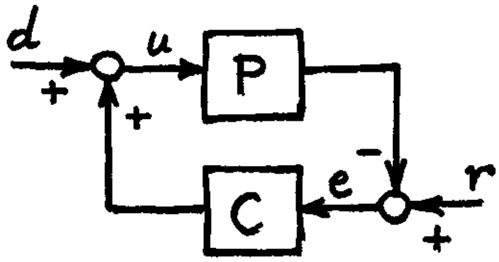
frequency $\omega_n = 10$, with damping ratio $\zeta = 0.25$. Note that for very small ω, $G(i\omega) \approx 0.1$, which corresponds to -20 dB, while for very large ω, $G(i\omega) \approx 1$, which corresponds to 0 dB.

Another important graphic representation of the function $G(i\omega)$ is its <u>Nyquist plot</u>. This is the curve described by $G(i\omega)$ as ω moves from $-\infty$ to $+\infty$. There is no ω-axis, but the direction of increasing ω is indicated by arrows and various points may be marked with their corresponding ω. The Nyquist plot of G from the previous page, as drawn by MATLAB, is shown below:



The Nyquist (and Bode) plots of a plant can be drawn based on measurements (with sinusoidal signals) even if the mathematical model (the transfer function) of the system is not known. It is always a closed curve, symmetric about the real axis.

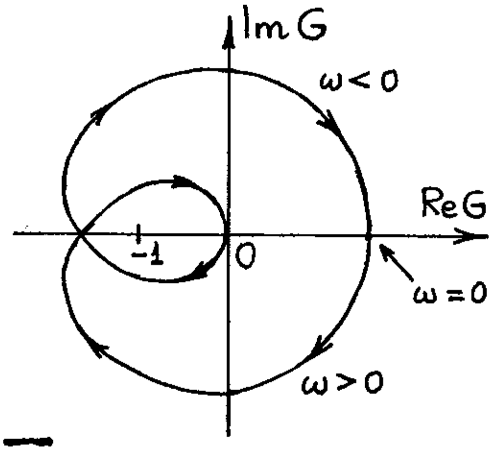
We shall see that the Nyquist plot is very useful for determining the stability of feedback systems. Recall that a feedback system built of two blocks, as shown, is stable if the four transfer functions from the inputs d, r to the signals u and e are all stable. This happens iff the <u>sensitivity</u> $S = \frac{1}{1+PC}$ is stable and there are no unstable pole-zero cancellations in the product PC. S is the transfer function from r to e (or from d to u). If both P and C are stable (as is often the case) then all we have to check is that S is stable. Denote for simplicity $G = PC$. G is called the <u>loop gain</u> of the feedback system.



FOR STABLE LOOP GAIN

> <u>Nyquist stability criterion</u>. The transfer function $S = (1+G)^{-1}$ (with G stable) is stable if and only if the Nyquist plot of G does not encircle or hit the point -1.

For example, the feedback system with G as on the previous page is stable. However, if the plot looks like here (on the right), then S is unstable.



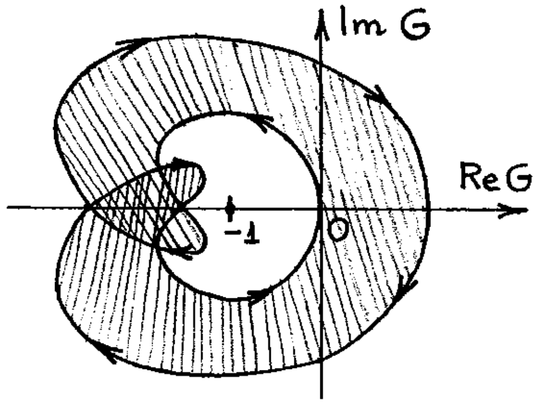
The precise definition of when a closed curve Γ encircles a point $z_0 \in \mathbb{C}$ is the following: the <u>winding number</u> of Γ around a point z_0 which is not on Γ is

$$n = \frac{1}{2\pi i} \int_{\Gamma} \frac{dz}{z - z_0} .$$

The intuitive meaning of n is: how many times does Γ go counterclockwise around z_0. n is an integer (positive or negative), and it can be guessed by looking at the curve, without computing the integral. Γ does not encircle z_0 if $n = 0$. The set of points z_0 where $n \neq 0$ is called <u>the area encircled by Γ</u>. In particular, for Γ being the Nyquist plot of G and $z_0 = -1$, we have

$$n = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{G'(i\omega)\, d\omega}{G(i\omega) + 1} .$$

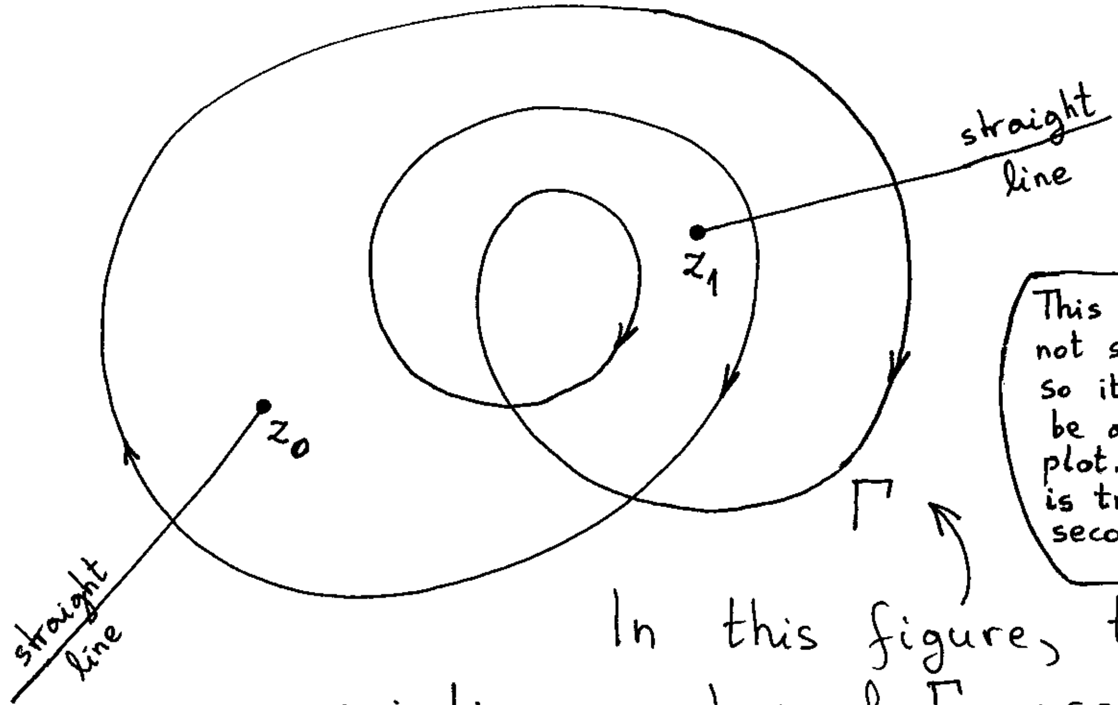
For a stable G, this n is either zero or negative. For example, for the Nyquist plot on the previous page, $n = -2$. The Nyquist criterion follows from the fact that the area encircled by the Nyquist plot is the image of the right half-plane $\mathbb{C}_+$ through G – think about it!



In this figure, the area encircled by the Nyquist plot is shaded, and it does not include -1, so S is stable.

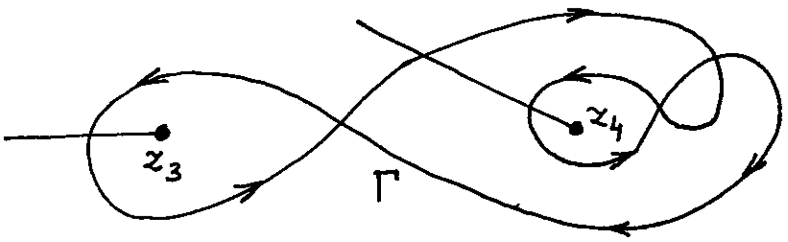
Another method for determining the winding number of a curve Γ around a point z_0 which is not on Γ: construct a straight line that starts at z_0 and extends to ∞ (in any direction). On this line, count the number of times Γ cuts the straight line (in counter-clockwise direction, clockwise cuts count as -1 cut). This number n is the winding number of Γ around z_0.

Examples:



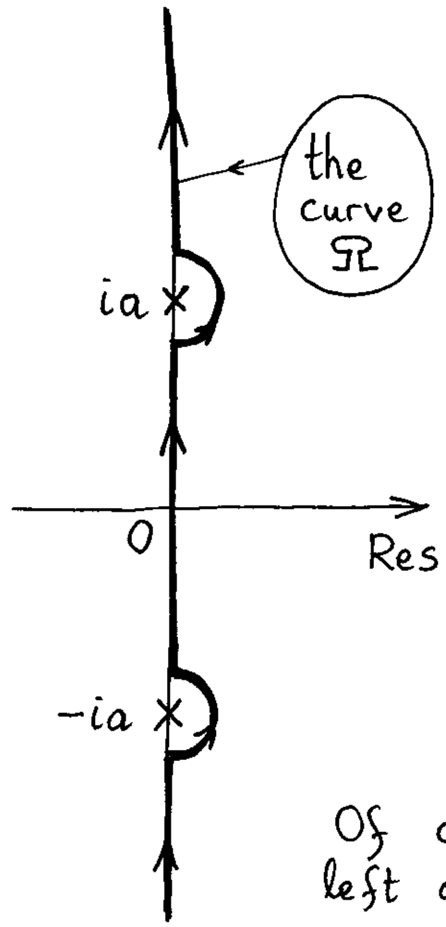
This curve is not symmetric, so it cannot be a Nyquist plot. The same is true for the second curve below.

In this figure, the winding number of Γ around z_0 is -1, around z_1 it is -2



In the second figure, the winding numbers are 1 for z_3 and 0 for z_4.

Now we state a more general version of the Nyquist stability criterion. We allow $G = PC$ to be unstable, i.e., it may have poles on the imaginary axis $i\mathbb{R}$ and/or in the open right half-plane $\mathbb{C}_+$ (where $\mathrm{Re}\, s > 0$). Because of the possible poles on $i\mathbb{R}$, we have to redefine what we mean by the Nyquist plot: we replace the imaginary axis by a curve Ω which avoids the imaginary poles on very small half-circles centered in the imaginary poles and lying in $\mathbb{C}_+$, see the figure:



In this figure, $\pm ia$ are poles of G. The curve Ω is almost equal to the imaginary axis, but it avoids (circumvents) the poles of G.

> The Nyquist plot of G is the image of the curve Ω through G.

Of course, G may have poles both left and right of the imaginary axis $i\mathbb{R}$.

The radius of the half-circles should be very small. If the half-circles are sufficiently small, then their precise size does not matter for the Nyquist criterion, given below:

<u>Nyquist stability criterion (general version).</u>

Let Γ be the Nyquist plot of G, and let n be the winding number of Γ around -1. Let p denote the number of poles of G in $\mathbb{C}_+$. Then $S=(1+G)^{-1}$ has $p-n$ poles in $\mathbb{C}_+$. In particular, it follows that for S to be stable, we need to have $n=p$.

<u>Remark</u>: If Γ runs through the point -1 then the winding number n is not defined. In this case we know that S has poles on the imaginary axis, so it is not stable.

YOU MAY IGNORE THIS

<u>Remark</u>. The proof of the Nyquist criterion (general version) is based on the following result about analytic functions: If f is a meromorphic function on a simply connected open set D and Ω is a closed curve in D, then $n(f(\Omega);0)$, the winding number of $f(\Omega)$ around 0, is given by

multiplicity of the zero

multiplicity of the pole

$$n(f(\Omega);0)=\sum_{f(z)=0} n(\Omega;z)\,m(z)-\sum_{f(p)=\infty} n(\Omega;p)\,m(p).$$

YOU MAY IGNORE THIS

The terminology means the following: a function is meromorphic on an open set if it is analytic except for a finite number of singular points, which are poles. An open set is called simply connected if any two points in it can be joined by a curve lying in this set, and the set does not have holes. To apply the mathematical result, we take $\mathfrak{R}$ to be as on p. 10, but we transform it into a closed curve by adding a half-circle of very large radius lying in $\mathbb{C}_+$. We take $f = 1 + G$, so that $f(\mathfrak{R})$ is the Nyquist plot shifted to the right by 1. Thus, $n(f(\mathfrak{R}); 0)$ is in fact n from the generalized Nyquist criterion. We have $n(\mathfrak{R}; s) = -1$ for all relevant points $s \in \mathbb{C}_+$. Hence, $n = p - z$, where z is the number of zeros of $f = 1 + G$ in $\mathbb{C}_+$. Clearly, z is the number of poles of S in $\mathbb{C}_+$. This whole remark is intended for math enthusiasts only, and it may be ignored without consequences.

The formula on the bottom of p. 11 is called the <u>Principle of the Argument</u> – look up its proof in any book about complex analysis.