Analyzing frequency response

We have previously seen how to analyze systems subject to basic input functions. We have even learned how to design them. Now we'll examine systems subject to periodic functions.

1 Bode diagrams

The behaviour of a system subject to periodic input is called the **frequency response**. We want to examine this. A very nice way to do this, is by making Bode diagrams. In this part we'll examine how to draw these diagrams.

1.1 Subjecting systems to a sine function input

Let's suppose we have a system with transfer function $F(s)$. We give it as input a sine function. So,

$$u(t) = X \sin \omega t, \qquad \text{or, equivalently} \qquad U(s) = \frac{\omega X}{s^2 + \omega^2}. \tag{1.1}$$

Here X is the **input amplitude** and ω is the **frequency**. We want to know how this input effects the output $y(t)$. To find that, we first examine $Y(s)$. Using a partial-fraction expansion, we can show that

$$Y(s) = F(s)U(s) = \frac{a}{s + \omega i} + \frac{\bar{a}}{s - \omega i} + F(s), \qquad \text{where} \quad a = -\frac{XF(-\omega i)}{2i} \quad \text{and} \quad \bar{a} = \frac{XF(\omega i)}{2i}. \tag{1.2}$$

We assume that our system is a stable system. In this case the effects due to initial conditions eventually fade out. The so-called **steady-state solution** for $y(t)$ (noted as $y_{ss}(t)$) is then given by

$$y_{ss}(t) = X|F(\omega i)| \sin(\omega t + \phi) = Y \sin(\omega t + \phi), \qquad \text{where} \qquad \phi = \arg(F(\omega i)). \tag{1.3}$$

Here Y is the **output amplitude**. Also ϕ is the **phase angle**.

A few important things can be noted. We note that the frequency of the output is the same as the frequency of the input. However, the amplitude is not the same. It is magnified by an **amplitude gain** $K = Y/X = |F(\omega i)|$. There is also a phase difference, having magnitude $\phi = \arg(F(\omega i))$. By the way, if $\phi < 0$, then there is so-called **phase lag**. The system/network is then called a **lag network**. Similarly, if $\phi > 0$, then there is **phase lead**. The network is then called a **lead network**.

1.2 Introduction to Bode diagrams

We want to know how systems respond to periodic inputs. So it would be nice to know how the amplitude gain K and the phase angle ϕ depend on the frequency ω. We display that in so-called **Bode diagrams**. These diagrams actually consist of two diagrams.

In the first diagram (the **gain Bode diagram**) we display the gain K versus the frequency ω. However, both values can differ rather much. So to get some meaningful graphs, we don't just display the gain K. Instead, we first transform this gain K to the unit dB. We do this using

$$K_{dB} = 20 \log K. \tag{1.4}$$

This **decibel gain** is then plotted versus $\log \omega$. The plot is therefore a **log-log plot**. (By the way, the base of the logarithm in both cases is 10.)

In the second diagram (the **phase angle Bode diagram**) we display the phase angle ϕ versus the logarithm of the frequency. So we plot ϕ versus $\log \omega$.

Since both parts of the Bode diagram have the same horizontal axis, they are usually placed above each other. This causes the decibel gain and the phase angle, corresponding to the same frequency, to be above each other as well.

One important question remains. How do we find these diagrams? Well, we simply insert $s = \omega i$ into the transfer function $F(s)$. This gives us a complex number. We then determine the length and argument of this number. These are then the gain K and phase angle ϕ. To investigate how this works, we will now examine some basic transfer functions $F(s)$ and describe their Bode diagrams.

1.3 A constant transfer function

If $F(s)$ is constant ($F(s) = K$), then the gain is simply K. In other words, the gain Bode diagram is simply a horizontal line, situated at $K_{dB} = 20 \log K$.

Usually this gain K is a positive number. The argument of a positive number is simply $\phi = 0$. So the phase angle Bode diagram of $F(s)$ is simply a horizontal line at $\phi = 0$.

1.4 A rather simple transfer function

Now let's suppose $F(s) = 1/s$. In this case $F(\omega i) = 1/\omega i = -i/\omega$. The decibel gain now is

$$K_{dB} = 20 \log |-i/\omega| = 20 \log 1/\omega = -20 \log \omega. \tag{1.5}$$

Remember that we plotted K_{dB} versus $\log \omega$. This means that the gain Bode diagram of $F(s)$ is simply a straight line. It has a slope of -20. (So the line is directed downward.) Also, if $\omega = 1$, then $\log \omega = 0$ and thus also $K_{dB} = 0$.

We also examine the phase angle Bode diagram. This one is rather easy. We can find that $\phi = \arg(-i/\omega) = -90°$. The graph is thus a horizontal line.

1.5 A first order transfer function

Now we examine $F(s) = 1/(1 + Ts)$. This gives us $F(\omega i) = 1/(1 + T\omega i)$. The decibel gain now becomes

$$K_{dB} = 20 \log \left| \frac{1}{1 + T\omega i} \right| = -20 \log \sqrt{1 + \omega^2 T^2}. \tag{1.6}$$

Let's investigate the factor ωT more closely. If ωT is very small, then $K_{dB} \approx 0$. So for small ωT, we simply have a horizontal line positioned at $K_{dB} = 0$. If, however, ωT is very big, then $K_{dB} \approx -20 \log \omega T$. It can be shown that this is a straight line with slope -20. (It thus goes downward.) It crosses the $0dB$ axis at $\omega = 1/T$. This frequency is called the **corner frequency** or **break frequency**.

We can actually apply a trick now. If we want to draw the gain Bode diagram of $F(s)$, then we only have to draw the two straight lines we just described. This gives a good approximation of the actual Bode diagram already. The only place where it is seriously off is near the corner frequency $\omega = 1/T$. (It is off by about $3dB$ there.) If we take into account for this deviation, then we have actually drawn a rather accurate curve.

But what about the phase angle Bode diagram? Well, we can find that the phase angle is

$$\phi = -\arctan \omega T. \tag{1.7}$$

But what does this mean? For small ωT, the value of ϕ converges to $0°$. For big ωT, we have $\phi \approx -90°$. Finally at the corner frequency $\omega = 1/T$ we have $\phi = -45°$. Just draw a smooth curve through these points and you've drawn the phase angle Bode diagram.

1.6 A second order transfer function

Finally, we examine $F(\omega i) = 1/(1 + 2\zeta i\omega/\omega_n + (i\omega/\omega_n)^2)$. The decibel gain is now given by

$$-20 \log \sqrt{\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta \frac{\omega}{\omega_n}\right)^2}. \tag{1.8}$$

We will examine ω/ω_n. If this quantity is very small, then we have $K_{dB} \approx 0$. So this is a horizontal line at K_{dB}. (Does it sound familiar?) If ω/ω_n is very big, then $K_{dB} \approx -40 \log \omega/\omega_n$. In other words, we find a straight line again. However, the slope of this straight line now is -40.

But what happens if $\omega \approx \omega_n$? Now something interesting occurs. The gain Bode diagram actually goes up. If the input frequency is close to the natural frequency, the amplitude gain increases. There is thus a peak in the gain Bode diagram. This phenomenon is called **resonance**. The **resonant frequency** ω_r is defined as the frequency at which the gain K is maximal. This frequency is given by

$$\omega_r = \omega_n \sqrt{1 - 2\zeta^2}. \tag{1.9}$$

So resonance only occurs if $\zeta < \sqrt{2}/2$. By the way, the corresponding value of the gain K is called the **resonant peak value** M_r. It is given by

$$M_r = |G(\omega i)|_{max} = |G(\omega_r i)| = \frac{1}{2\zeta\sqrt{1 - \zeta^2}}. \tag{1.10}$$

So as $\zeta \to 0$, then $M_r \to \infty$.

Well, so much for the gain Bode diagram. Now let's examine the phase angle Bode diagram. This time ϕ is given by

$$\phi = -\arctan\left(\frac{2\zeta\omega/\omega_n}{1 - (\omega/\omega_n)^2}\right). \tag{1.11}$$

What can we derive from this? For small ω/ω_n, we have $\phi \approx 0°$. For big ω/ω_n we have $\phi \approx -180°$. If $\omega = \omega_n$, then $\phi = -90°$.

Surprisingly, none of these points depend on ζ. You may wonder, doesn't the shape of this graph depend on ζ at all? Well, ζ does have effect. It determines the 'smoothness' of the graph. If $\zeta \approx 1$, then the graph of ϕ goes from $0°$ to $-180°$ rather slowly. If, however, $\zeta \approx 0.1$, then this transition occurs rather sudden.

1.7 Powers of basic functions

We now know what the Bode diagrams of basic transfer function $F(s)$ looks like. But what if we get a power of such a basic function? For example, let's suppose that $F(s) = 1/(1 + Ts)$. What should we do if we need to make a Bode diagram of $F(s)^n = 1/(1 + Ts)^n$?

Well, there is a trick we can use. We know that the decibel gain of $F(s)$ is $K_{dB} = 20 \log |F(\omega i)|$. The decibel gain of $F(s)^n$ then becomes

$$K_{dB} = 20 \log |F(\omega i)^n| = 20 \log |F(\omega i)|^n = 20n \log |F(\omega i)|. \tag{1.12}$$

So the decibel gain simply becomes n times as big! But what happens to the phase angle? Well, the phase angle of $F(s)^n$ is equal to the argument of $F(\omega i)^n$. ($\phi = \arg(F(\omega i)^n)$.) And there is a rule in complex number theory that states that $\arg(\lambda^n) = n \arg(\lambda)$. This means that ϕ also becomes n times as big!

What can we conclude from this? If we know the Bode diagram of a certain function $F(s)$, then finding the diagram of $F(s)^n$ is just a matter of multiplication.

By the way, this also works if n is negative. For example, let's suppose we want to find the Bode diagram of $F(s) = 1 + \omega T$. Then we just multiply the Bode diagram of $F(s) = 1/(s + \omega T)$ by -1. (In other words, mirror it about the horizontal axis.)

1.8 Bode diagrams of general functions

Now let's suppose we have some transfer function $F(s)$ that can be written as

$$F(s) = K\frac{(1 + T_{z1}s)(1 + T_{z2}s)\dots(1 + T_{zm}s)}{(1 + T_{p1}s)(1 + T_{p2}s)\dots(1 + T_{pn}s)}, \tag{1.13}$$

how can we draw the Bode diagram? Again, we can use a trick. And it is similar to the trick of the previous paragraph.

There is a rule stating $\log ab = \log a + \log b$. There is another rule stating that $\arg(ab) = \arg(a) + \arg(b)$. What can we derive from this? Well, if we want to find the Bode diagram of $F(s)$, we simply look at all the separate Bode diagrams. So we examine the bode diagrams of $K, (1 + T_{z1}s), \dots (1 + T_{zm}s), (1 + T_{p1}s), \dots, (1+T_{pn}s)$. Then (according to the rules above) we can simply add up all these Bode diagrams. This then gives us the Bode diagram of $F(s)$, which is exactly what we want.

1.9 Cutoff parameters

So what have we learned thus far? We have seen that a system can pass on certain frequencies to its output. However, other frequencies are damped out. A Bode diagram clearly displays which frequencies are damped out. However, we can also express this in certain parameters.

First we define the **zero-frequency gain** K_{dB_0}. This is the decibel gain K_{dB} as $\omega \to 0$. The **cutoff frequency** ω_b now satisfies

$$K_{dB}(\omega) < K_{dB_0} - 3dB \qquad \text{for every } \omega > \omega_b. \tag{1.14}$$

In other words, the system can't really transmit any frequency higher than the cutoff frequency. However, the system can transmit frequencies between 0 and ω_b. This frequency range is therefore called the **bandwidth** of the system.

Finally, we are also interested in how fast the decibel gain K_{dB} drops below $K_{dB_0} - 3dB$. In other words, we want to know the slope of the gain Bode diagram at $\omega = \omega_b$. This slope is called the **cutoff rate**. A high cutoff rate usually indicates a high resonance peak, and therefore a less stable system.

2 Nyquist plots and Nyquist stability

Another way of examining stability is by using the Nyquist method. We will delve into that method in this part.

2.1 Nyquist plots

Nyquist plots (also known as **polar plots**) are rather different from Bode diagrams. In a Nyquist plot the gain K and the phase angle ϕ are plotted in polar coordinates. The gain K then is the distance from the origin, and ϕ is the angle.

Let's take a closer look at what we're actually plotting. We have as radius $K = |F(\omega i)|$ and as angle $\phi = \arg(F(\omega i))$. So, in fact, what we plot is the function $F(\omega i)$ itself! If we let ω vary from 0 to ∞ and plot $F(\omega i)$ in the complex plane, then we will find the Nyquist plot!

Nyquist plots are often very fun. There is, however, one slight problem with them. The trick of adding up basic solutions (which worked so well for Bode diagrams) doesn't apply here. So, instead, we just have to draw these diagrams in the old fashioned way: calculate a few points (for certain ω) and draw a nice line through them.

2.2 The mapping theorem

Soon we will delve into the Nyquist stability criterion. But before we do that, we will look at the **mapping theorem**. (We will need it for the Nyquist stability criterion.)

Suppose we have some complex function $L(s)$, which is the ratio of two polynomials. We now take a closed contour in the complex plane. We say that there are P poles within this contour and Z zeroes. (There may not be any poles or zeroes on the contour itself.)

For every value s on this contour, we will now map the value of $L(s)$. This gives as a second closed curve. We now walk in a clockwise direction along this second curve. We call N the amount of times we have walked around the origin of the complex plane. The mapping theorem now states that $N = Z - P$.

2.3 The Nyquist stability criterion

It's finally time to examine the Nyquist stability criterion. This criterion states something about the stability of a basic closed-loop system.

Let's suppose we have a basic closed-loop system. The transfer function of this system is $F(s) = G(s)/(1+G(s)H(s))$. If the system is stable, then all the poles of $F(s)$ must lie in the left half of the complex plane. We can also see that the poles of $F(s)$ are the roots of the denominator $1 + G(s)H(s)$. So the roots of $L(s) = 1 + G(s)H(s)$ must also lie in the left half of the complex plan.

Let's examine this function $L(s)$ more closely. In fact, let's apply the mapping theorem to $L(s)$. First we will define the first curve. And now comes the brilliant step: we simply take the curve which encloses the right half of the complex plane. Then P is the number of zeroes in the right half of the complex plane. Similarly, Z is the number of zeroes.

Let's examine this curve more closely. One part of this curve consists of the imaginary axis. However, the other part has a distance infinitely far away from the origin. That seems complicated. However, if $\lim_{s\to\infty} L(s) = \text{constant}$, then it can be shown that this second part isn't interesting at all. So, we actually only need to concern ourselves with the imaginary axis!

Now let's draw the second curve. Our second curve is formed by $L(s)$, where s can be any point on the imaginary axis. In other words, our second curve is formed by $L(\omega i)$, where ω can be any real value. So, this second curve is simply the Nyquist plot of $L(s)$! Well, it almost is. In a Nyquist plot ω ranged from 0 to ∞. But now it goes from $-\infty$ to ∞. This isn't really a problem. Just take the Nyquist plot of $L(s)$, and mirror it with respect to the x-axis. That will give you the complete plot for $-\infty < \omega < \infty$.

Let's suppose we have found the entire second curve. To find N, we can simply count the number of times which the curve goes around the origin. But what can we conclude from this? Well, we know that $N = Z - P$. To ensure stability, there may not be any zeroes of $L(s)$ in the right half of the complex plane. We therefore must have $Z = 0$. So, the system is stable if, and only if, $N = -P$. Usually the poles of $L(s)$ (or equivalently, the poles of $G(s)H(s)$) are known. So we can readily check whether the condition $N = -P$ holds.

2.4 Sidenotes to the Nyquist stability criterion

There is only one slight problem with the Nyquist stability criterion. We have applied the mapping theorem. And the mapping theorem demanded that no poles/zeroes lie on the first curve. In other words, no curves/zeroes of $L(s)$ may lie on the imaginary axis. This may sound like a problem, but luckily there is a solution. We simply change our curve slightly. If we encounter a pole/zero on the imaginary axis, we simply go around it. In other words, we take an infinitely small half circle, and place it around the pole/zero. This then solves our problem.

By the way, we also may alter the above criterion slightly. Previously we have plotted $L(\omega i) = 1 + G(\omega i)H(\omega i)$ and counted how many times the curve went around the origin. Now let's do something which is only slightly different. Instead, we plot $G(\omega i)H(\omega i)$. This time we count how many times the curve goes around the point -1. This small trick doesn't change the number N, so we may readily apply it.

2.5 Crossover frequencies and phase/gain margins

Let's suppose we have the Nyquist plot of $G(s)$. According to the Nyquist stability criterion, we can determine whether it is stable. We can also deduce something else from this criterion. Namely that, if the Nyquist plot goes through the point -1, then the system is on the verge of instability. In this case there is some frequency ω for which $|G(\omega i)| = 1$ and $\arg(G(\omega i)) = 180°$.

How can we check whether this is the case? For that, we first find the gain crossover frequency and the phase crossover frequency.

- The **gain crossover frequency** ω_{gc} is the frequency for which the gain is equal to $|G(\omega i)| = 1$.
- The **phase crossover frequency** ω_{pc} is the frequency for which the phase angle is $\arg(G(\omega i)) = 180°$.

(Note that there can be multiple gain/phase crossover frequencies.) If any of the gain crossover frequencies is equal to any of the phase crossover frequencies, then we know that the system is on the verge of instability.

Of course $G(\omega i)$ doesn't always pass through the point -1. But even if it doesn't, we would still like to know how close $G(\omega i)$ is to this point. (Since, if the curve goes closer to -1, the system becomes less stable.) To examine this 'closeness', we define the phase and gain margin.

- Let's examine the angle $\phi = \arg(G(\omega_{gc} i))$ at the gain crossover frequency. The **phase margin** γ now is the difference between this angle and the angle of -1 (which is $180°$). So $\gamma = \phi - 180°$.
- Let's examine the gain $|G(\omega_{pc} i)|$ at the phase crossover frequency. The **gain margin** K_g is the reciprocal of this magnitude. So, $K_g = 1/|G(\omega_{pc} i)|$. (In decibels this becomes $K_{g_{dB}} = -20 \log |G(\omega_{pc} i)|$.)

For most systems, both margins should be positive to ensure stability. Negative margins are often signs of instability. Although this does depend on the system itself.

3 Nichols plots

Let's suppose we have a simple closed loop system. The closed-loop transfer function is given by $F(s) = G(s)/(1 + G(s))$. We also know the frequency response of the open-loop transfer function $G(s)$. Can we derive from this the frequency response of $F(s)$ itself? It turns out that we can. We will find out how in this part.

3.1 M circles and N circles

First we make some definitions. Let's define the closed loop gain $M = |F(\omega i)|$ and the closed loop phase angle $\alpha = \arg(F(\omega i))$. We also define $N = \tan \alpha$. And finally we write $G(\omega j)$ as $G(\omega j) = X + Yi$.

Now let's look at the magnitude of $F(\omega i)$. This gives us an expression for M, being

$$M = |F(\omega i)| = \frac{|G(s)|}{1+|G(s)|} = \frac{|X+Yi|}{|1+X+Yi|} = \frac{\sqrt{X^2+Y^2}}{\sqrt{(1+X)^2+Y^2}}. \tag{3.1}$$

If $M = 1$, then we find that $X = -1/2$ (for every Y). So if we make a plot of Y versus X, we will get a vertical line. If $M \neq 1$, then we can rewrite the above equation to

$$\left(X + \frac{M^2}{M^2-1}\right)^2 + Y^2 = \left(\frac{M}{M^2-1}\right)^2. \tag{3.2}$$

But this is the equation of a circle! The circle has a radius $|M/(M^2-1)|$! The center of the circle is located at $X = -M^2/(M^2-1)$ and $Y = 0$. The circles we can find with these equation are called M **circles** or **constant magnitude loci**. Along these circles, the magnitude of $F(\omega i)$ is constant.

Similarly, we can find an expression for α (or equivalently, for N), by looking at the argument of $F(\omega i)$. This time we can derive that

$$\left(X + \frac{1}{2}\right)^2 + \left(Y - \frac{1}{2N}\right)^2 = \frac{1}{4} + \left(\frac{1}{2N}\right)^2. \tag{3.3}$$

This is also the expression for a circle! This time the radius is $\sqrt{1/4 + 1/(2N)^2}$. The center of the circle lies at $X = -1/2$ and $Y = 1/(2N)$. The circles we can find with this equation are called N **circles** or **constant phase angle loci**. Along these circles, the argument of $F(\omega i)$ is constant.

So how do we derive the frequency response of $F(s)$ from the frequency response of $G(s)$? Well, we simply make a Nyquist plot of $G(s)$. We then lay this plot over the plots of the M circles and N circles. From this we can derive the value of M and N (and thus α) for every frequency ω.

3.2 Nichols charts

Working with multiple graphs (as was done in the previous paragraph) may be a bit inconvenient. That's why often **Nichols charts** are used. The horizontal axis of such a Nichols chart is the phase angle ϕ of $G(s)$. The vertical axis is the decibel gain K_{dB} of $G(s)$. In this graph, both the M circles and N circles are plotted. (Although they aren't exactly circles anymore.)

So how do we use such a Nichols chart? Once more we plot the frequency reponse properties of $G(s)$: We plot the decibel gain K_{dB} versus the phase angle ϕ. We then lay this curve over a Nichols chart. We mark the points where the curve intersects the M and N circles. These points then indicate the corresponding values of M and N (and thus of α). By taking enough intersection points, we can actually derive the frequency response of $F(s)$ quite accurately.