# Adjusting system properties

Making an aircraft stable is one thing. But giving it a satisfactory behaviour and being able to control it is another story. In this chapter, we're going to look at some parameters which a system can have. After that, we'll examine how we can influence these parameters.

## 1 Important system parameters

For every system, we can find several parameters that mention something about the system. Some parameters give us hints about the stability of the system. And other parameters are nice to know for other reasons. We will now examine quite some parameters.

### 1.1 Phase and gain margins

Let's again examine the system with transfer function $F(s) = G(s)/(1 + G(s)H(s))$. We'll examine this function in the frequency domain, and thus substitute $s$ by $j\omega$. If the term $G(j\omega)H(j\omega)$ ever becomes $-1$, then the system becomes unstable. We are thus interested in the points where $|G(j\omega)H(j\omega)| = 1$ and $\arg(G(j\omega)H(j\omega)) = -180°$. The frequency at which $\phi = \arg(G(j\omega)H(j\omega)) = -180°$ is called the **phase crossover frequency** $\omega_{\phi=-180°}$. Similarly, the frequency at which $K = |G(j\omega)H(j\omega)| = 1$ (or $K_{dB} = 0$) is called the **gain crossover frequency** $\omega_{K=1}$.

We would like to know how close we are to instability. So, let's suppose that we already have a phase angle of $\phi = -180°$. (We thus have a frequency equal to the phase crossover frequency $\omega_{\phi=-180°}$.) The **gain margin** $GM$ is now defined as

$$GM = \frac{1}{|G(j\omega_{\phi=-180°})|} = \frac{1}{K_{\phi=-180°}}. \tag{1.1}$$

A gain margin of $GM < 1$ (or similarly, $GM_{dB} < 0$) indicates instability. As a rule of thumb, we would like to have $GM_{dB} > 6$ dB.

Similarly, we can suppose we already have a gain of $K = 1$. (We thus have a frequency equal to the gain crossover frequency $\omega_{K=1}$.) The **phase margin** $PM$ is now defined as

$$PM = 180° + \arg\left(G(j\omega_{K=1})\right) = 180° + \phi_{K=1}. \tag{1.2}$$

A phase margin of $PM < 0°$ indicates instability. As a rule of thumb, we would like to have $30° < PM < 60°$.

The phase and gain margins can also be found in the various plots that were discussed. To find them, you first have to find the point where $\phi = -180°$ (for the gain margin) or $K = 1$ (for the phase margin). You then have to find the gain/phase angle, and by using the definition for the gain margin/phase margin you can find the corresponding value. In this way, you can also find the **ultimate gain** $K_{ult}$, which is defined as the gain $K$ at the phase crossover frequency. It can be found from the gain margin, using

$$K_{ult} = GM = 10^{\left(\frac{GM_{dB}}{20}\right)}. \tag{1.3}$$

The phase and gain margins can, however, be misleading. It may happen that the phase and gain margins appear safe, but there still is a value of $\omega$ for which $G(j\omega)H(j\omega)$ comes close to $-1$. Therefore, instead of looking at phase and gain margins, it is often wise to simply look at the Nyquist plot of $G(j\omega)H(j\omega)$ and see if it comes close to $-1$. If not, then the system appears to be quite alright.

## 1.2 Other frequency domain parameters

There are more parameters that are related to the frequency domain. Most of these parameters can easily be derived from the Bode plot. We'll discuss a couple of them now.

Let's examine a Bode diagram. In this Bode diagram is a frequency region in which the system performs satisfactory. This region is usually a region with a more or less constant gain $K_0$. The point(s) where the gain drops below 3 dB less than this constant gain $K_0$ is called the **cutoff frequency**. The slope of the Bode plot at this point is called the **cutoff rate**. Also, the frequency range in which the system performs satisfactory (being the frequency range between the cutoff frequencies) is called the **bandwidth** $\omega_b$.

In a Bode diagram, you can often find a peak at which the gain $K$ is at a maximum. This phenomenon is called **resonance**. The corresponding maximum value of the gain $K$ is denoted by the **resonance peak** $M_p$. The frequency at which this resonance occurs is called the **resonance frequency** $\omega_p$.

The last important parameter for the frequency domain is the delay time. The **delay time** $t_d(\omega)$ for a given frequency $\omega$ is given by

$$t_d(\omega) = -\frac{d\phi}{d\omega} = -\frac{d\arg\left(G(j\omega)\right)}{d\omega}. \tag{1.4}$$

## 1.3 Time domain parameters

In the time domain, there are also several parameters that are important. Let's suppose that we have a system in which the output $y(t)$ needs to follow the input $u(t)$. Also, suppose that we put a step function of size $k$ on the input. (Though usually $k = 1$ is selected.) So, for $t < 0$ we have $u(t) = 0$ and for $t > 0$ we have $u(t) = k$.

Of course, in the time domain, time matters. So, let's examine some characteristic times. First, the **delay time** $t_d$ is defined such that $y(t_d) = 0.5y_{ss}$. In other words, at the delay time the system is halfway with adjusting itself to the new input value. We also have the **rise time** $t_r$. But before we can define it, we first need to define $t_{r_{initial}}$ and $t_{r_{final}}$. These parameters are defined such that $y(t_{r_{initial}}) = 0.1$ and $y(t_{r_{final}}) = 0.9$. The rise time is now given by $t_r = t_{r_{final}} - t_{r_{initial}}$. Thirdly, the **settling time** $t_s$ is the time it takes for the system to come and stay close to the steady state output. So, for all $t > t_s$ we must have $|y(t) - y_{ss}| < 0.02y_{ss}$. (Of course, the parameter 0.02 can be varied. A value of 0.05 is often used as well.)

Next to these important time parameters, there are also parameters not related to time. For example, there is the **(maximum) overshoot** $M_p$. This is the difference between the maximum value of $y(t)$ and its steady state value $y_{ss}$. (So, $M_p = \max(y(t)) - y_{ss}$.) And finally, there is the **apparent time constant** $\tau$. To grasp its meaning, we have to suppose that the output is given by a function of the form $y(t) = y_{ss} - Ae^{-\alpha t}\cos(\omega t + \phi)$. The parameter $\tau$ is now defined as $\tau = 1/\alpha$. In other words, it is the time it takes until the amplitude of the oscillation has reduced to 37% of its value.

## 1.4 Error specifications

When designing a system, there usually are requirements. These requirements can also concern the error which the system has. To examine the error, we first simply assume that $H(s) = 1$. Thus $F(s) = G(s)/\left(1 + G(s)\right)$. We then rewrite the open loop transfer function $G(s)$ of the system as

$$G(s) = K\frac{s^a \prod_{i=1}^{i=m-a}(s + z_i)}{s^b \prod_{j=1}^{j=n-b}(s + p_j)} = K_{mod}\frac{\prod_{i=1}^{i=m-a}(\tau_{z,i}s + 1)}{s^l \prod_{j=1}^{j=n-b}(\tau_{p,j}s + 1)}. \tag{1.5}$$

In other words, the open loop transfer function has $m$ zeroes and $n$ poles. A number $a$ of these zeroes is equal to zero. Similarly, a number $b$ of the poles is zero as well. We also have $l = b - a$. We will soon see that this parameter $l$ is very important. In fact, it denotes the **type** of the system. If $l = 0$ then we have a **type 0** system, if $l = 1$ then we have a **type 1** system, and so on.

The system output $Y(s)$ should follow the system input $U(s)$. So, we define the error $E(s)$ as the difference. It is thus equal to

$$E(s) = U(s) - Y(s) = U(s) - U(s)F(s) = U(s) - U(s)\frac{G(s)}{1+G(s)} = \frac{U(s)}{1+G(s)}. \quad (1.6)$$

To find the eventual error $e(\infty)$ of the system, we can use the **final value theorem**. It implies that

$$e(\infty) = \lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{sU(s)}{1+G(s)} = \frac{sU(s)}{1+\frac{K_{mod}}{s^l}}. \quad (1.7)$$

Now we can put various inputs into this system and find the error. This gives us the following results.

- First, we insert a **step input**. Thus, $u(t) = 1$ (for $t > 0$) and $U(s) = 1/s$. We now find that for type 0 systems, there is a **steady state error** of $e(\infty) = 1/(1 + K_{mod})$. However, for type 1 and beyond, the error is zero. (By the way, this error is called a **position error**.)
- Second, we insert a **ramp input**. So, $u(t) = t$ (for $t > 0$) and $U(s) = 1/s^2$. This time type 0 systems give an infinite error: it diverges. Type 1 systems give a steady state error of $e(\infty) = \frac{1}{K_{mod}}$. Type 2 systems and beyond give a zero error. (This error is called a **velocity error**.)
- Third, we insert a **parabolic input**. So, $u(t) = \frac{1}{2}t^2$ (for $t > 0$) and $U(s) = 1/s^3$. This time type 0 and type 1 systems give an infinite error. Type 2 systems give a steady state error of $e(\infty) = \frac{1}{K_{mod}}$. Type 3 systems and beyond give a zero error. (This error is called a **acceleration error**.)

I think you can understand the general trend of the above experiments now. So remember, the type of the system determines which kind of position, velocity and acceleration errors the system has.

# 2 Controllers - time domain

By varying the (proportional) open-loop gain $K$ of the system, we can already vary its properties by quite a bit. But, sometimes varying this gain is not enough. In that case, we need a compensator or a controller. First, we'll examine controllers.

## 2.1 PID Control

Let's examine a basic feedback loop with $H(s) = 1$. In this feedback loop, the output signal $Y(s)$ is fed back to the system. Usually, the signal that is fed back is proportional to the output. We thus have a **proportional controller**: $K(s) = K_p$. (Here, $K_p$ is the **proportional gain**. $K(s)$ is the **controller function**.) A proportional controller generally reduces the rise time $t_r$, increases the overshoot $M_p$ and reduces the steady state error $e_{ss}$.

Sometimes, however, it may be convenient to get the derivative of the output as feedback signal. In this case, we use a **derivative controller**: $K(s) = K_D s$. ($K_D$ is the **derivative gain**.) A derivative controller reduces the overshoot $M_p$ and the settling time $t_s$.

Finally, we can also use an **integral controller**: $K(s) = \frac{1}{s}K_I$. ($K_I$ is the **integral gain**.) An integral controller reduces the rise time $t_r$ and sets the steady state error $e_{ss}$ to zero. However, it increases the overshoot $M_p$ and the settling time $t_s$.

Of course, we can also combine all these controllers. This gives us the **PID controller**:

$$K(s) = K_p + \frac{K_I}{s} + K_D s = \frac{K_D s^2 + K_p s + K_I}{s}. \tag{2.1}$$

By using the PID controller, we can influence the parameters $t_r$, $t_s$, $M_p$ and $e_{ss}$ in many ways. Just vary the gains $K_p$, $K_D$ and $K_I$. But which gains do we choose? For that, we can use tuning rules.

### 2.2 The Ziegler-Nichols tuning rules

We will now examine the **Ziegler-Nichols tuning rules**. There are two variants: the **quarter decay ratio** method and the **ultimate sensitivity** method. For both methods, we first write $K(s)$ as

$$K(s) = K_p \left(1 + \frac{1}{T_I s} + T_D s\right). \tag{2.2}$$

Now let's examine the quarter decay ratio method. These tuning rules should give a decay ratio of 0.25. (The **decay ratio** is the ratio of the magnitudes of two consecutive peaks of an oscillation.) First, we examine the response of the original system to a unit step input. From this we determine the **lag** $L$, which is the time until the system really starts moving. (We have $L \approx t_d$.) We also find the **slope** $R$, which is the average slope of the system response during its rise time. (We have $R \approx y_{ss}/t_r$.)

Based on the values of $L$ and $R$, we can choose our gains. If we only use proportional gain, then $K_p = \frac{1}{RL}$. If we use a PI controller, then $K_p = \frac{0.9}{RL}$ and $T_I = \frac{L}{0.3}$. Finally, if we use a PID controller, then $K_p = \frac{1.2}{RL}$, $T_I = 2L$ and $T_D = 0.5L$. These rules should then roughly give a decay ratio of 0.25. Although some additional tuning is often necessary/recommended.

Now let's examine the ultimate sensitivity method. First, we examine the original system with a gain equal to the ultimate gain $K_p = K_{ult}$. In other words, we choose $K_p$ such that the system has continuous oscillations without any damping. The corresponding **ultimate period** of these oscillations is now denoted by $P_u$. (This does mean that the ultimate sensitivity method can only be used when continuous oscillations can be achieved. In other words, the root locus plot has to cross the imaginary axis at a point other than zero.)

Based on the values of $K_{ult}$ and $P_u$, we can choose our gains. For proportional control, we use $K_p = 0.5K_{ult}$. For PI control, we use $K_p = 0.45K_{ult}$ and $T_I = \frac{P_u}{1.2}$. For PID control, we use $K_p = 0.6K_{ult}$, $T_I = \frac{1}{2}P_u$ and $T_D = \frac{1}{8}P_u$. Again, additional tuning is often necessary/recommended.

## 3 Compensators - frequency domain

### 3.1 Three kinds of compensators

There are three important kinds of compensators. These are the lead compensator, the lag compensator and the lead-lag compensator, respectively given by the transfer functions

$$D_1(s) = K(s + z), \qquad D_2(s) = \frac{K}{s + p} \qquad \text{and} \qquad D_3(s) = K\frac{s + z}{s + p}. \tag{3.1}$$

Let's look at these compensators individually.

The **lead compensator** offers PD control. This causes it to speed up the response of a system. In other words, the rise time $t_r$ goes down. Also, the overshoot $M_p$ becomes less. The lead compensator does have a problem though. It increases the gain of the system at high frequencies. In other words, with a lead compensator high frequencies are amplified. This is generally not very positive.

The **lag compensator** offers PI control. This means that it improves the steady state accuracy. (If you need to have $e_{ss} \approx 0$, then a lag compensator comes in handy.) The PI controller reduces high-frequency noise. As such, it can be used as a **low-pass filter**. (This is a filter that only lets low frequencies pass.)

The **lead-lag compensator** combines the lead and the lag compensator. In this way, the negative effects of the lead compensator can be compensated for. First, a lead compensator can be used to speed up the response of the system. Then a lag compensator is also added, such that the high frequency effects are limited. This lag compensator is made such that its effects on the biggest part of the system are negligible.

In the lead-lag compensator, the lead compensator is the most important part. However, we can also put it together such that the lag compensator is the most important part. In this case, we often call the compensator a **lag-lead compensator**.

## 3.2 Tuning the compensators

Using lead and lag compensators is like adding zeros and poles to the system. But when doing this, an important question arises: where do we put the zeros and poles? For this, we can use the root locus plot. We now have a nice rule of thumb: poles push the locus away, whereas zeros attract the locus. But we also have more precise rules to place the zeros and poles.

Let's suppose we're setting up a lead compensator. We thus need to choose its zero. It is often wise to put this zero in the neighbourhood of the **natural frequency** $\omega_N$ which you want the system to have. This natural frequency can roughly be determined from the parameters $t_r$ and/or $t_s$ using the approximate equations

$$t_r \approx \frac{1.8}{\omega_N} \qquad \text{and} \qquad t_s \approx \frac{4.6}{\zeta\omega_N}. \tag{3.2}$$

In this equation, the value of $\zeta$ can often be determined from the required value of the overshoot peak $M_p$, according to

$$\zeta \approx 0.7 \text{ when } M_p \approx 5\%, \qquad \zeta \approx 0.5 \text{ when } M_p \approx 15\% \quad \text{and} \quad \zeta \approx 0.3 \text{ when } M_p \approx 35\%. \tag{3.3}$$

To compensate for high frequency effects, we also add a pole (as a lag compensator). This pole, however, should be relatively far away from the zero. A rule of thumb is to place the pole 5 to 20 times further from the origin as the zero. Thus, $p \approx (5 \text{ to } 20) \cdot z$.