Introduction and theory recapitulation

When flying an aircraft, you need to control it. This controlling can be done by humans. But often, it is much easier, safer, faster, more efficient and more reliable to do so by computer. But, how do we automatically control a flight? That is what this summary is all about.

1 Basic flight control concepts

1.1 Introduction automatic flight control

The system that is used to control the flight is called the **flight control system** (FCS). In the early days of flying, the FCS was mechanical. By means of cables and pulleys, the control surfaces of the aircraft were given the necessary deflections to control the aircraft. However, new technologies brought with it the fly-by-wire FCS. In this system electrical signals are sent to the control surfaces. The signals are sent by the **flight (control) computer** (FC/FCC). In this way, the aircraft is controlled.

But what is the advantage of automatic flight control? Why would we use an FC instead of a pilot? There are several reasons for this. First of all, a computer has a much higher reaction velocity than a pilot. Also, it isn't subject to concentration losses and fatigue. Finally, a computer can more accurately know the state the aircraft is in. (Computers can handle huge amounts of data better and also don't need to read a small indicator to know, for example, the velocity or the height of the aircraft.) However, there also is a downside to FCs. They are only designed for a certain flight envelope. When the aircraft is outside of the flight envelope, the system can't really operate the aircraft anymore. For these situations, we still need pilots.

1.2 Set-up of the flight control system

The FCS of an aircraft generally consists of three important parts.

- The **stability augmentation system** (SAS) augments to the stability of the aircraft. It mostly does this by using the control surfaces to make the aircraft more stable. A good example of a part of the SAS is the **phugoid damper** (or similarly, the **yaw damper**). A phugoid damper uses the elevator to reduce the effects of the phugoid: it damps it. The SAS is always on when the aircraft is flying. Without it, the aircraft is less stable or possibly even unstable.
- The **control augmentation system** (CAS) is a helpful tool for the pilot to control the aircraft. For example, the pilot can tell the CAS to 'keep the current heading'. The CAS then follows this command. In this way, the pilot doesn't continuously have to compensate for heading changes himself.
- Finally, the **automatic control** system takes things one step further. It automatically controls the aircraft. It does this by calculating (for example) the roll angles of the aircraft that are required to stay on a given flight path. It then makes sure that these roll angles are achieved. In this way, the airplane is controlled automatically.

There are important differences between the above three systems. First of all, the SAS is always on, while the other two systems are only on when the pilot needs them. Second, there is the matter of **reversibility**. In the CAS and automatic control, the pilot feels the actions that are performed by the computer. In other words, when the computer decides to move a control panel, also the stick/pedals of the pilot move along. This makes these systems reversible. The SAS, on the other hand, is not reversible: the pilot doesn't receive feedback. The reason for this is simple. If the pilot would receive feedback, the only things he would feel are annoying vibrations. This is of course undesirable.

2 Flight dynamics recap

2.1 Reference frames

To be able to express the state of the aircraft, we need a reference frame. Several reference frames are around. We'll discuss the four most important ones here. (You've probably seen them before quite some times, but for completeness, we do mention them.)

- The **Earth-fixed frame of reference** F_E is a right-handed orthogonal system. The Z_E axis points to the Earth's center, the X_E axis points North and the Y_E axis points East. The origin of the system is initially positioned at the aircraft center of gravity (cog). However, the F_E reference frame is fixed to Earth. So, when the aircraft cog moves, the origin of the F_E system stays fixed with respect to the Earth.
- The **body-fixed frame of reference** F_B also is a right-handed orthogonal system. Its origin lies at the aircraft cog and is fixed to it. (So, if the aircraft moves, the frame of reference moves along.) The X_B axis is parallel to the aircraft longitudinal axis and points forward. The Y_B axis is parallel to the lateral axis and points to the right. Finally, the Z_B axis points downward.
- The **stability frame of reference** F_S also is a right-handed orthogonal system. Its origin is fixed to the aircraft cog, just like with F_B. Also, the Y_S axis coincides with the Y_B axis. However, this time the X_S axis is rotated downward by the angle of attack α. To be more precise, the X_S axis is parallel to the projection of the velocity vector on the plane of symmetry of the aircraft. The Z_S axis still points downward, but it is of course also rotated by an angle α.
- Finally, there is the **aircraft frame of reference** F_r. Contrary to the other systems, this is a left-handed orthogonal system. Its origin is a certain fixed point on the aircraft (though not the cog). The X_r axis points to the rear of the aircraft, the Y_r axis points to the left and the Z_r axis points upward.

2.2 The equations of motion

By using the reference frames, we can derive the equations of motion. (There are force equations, moment equations and kinematic relations.) These equations are, however, nonlinear. So to be able to work with them more easily, they are linearized about an equilibrium position of the aircraft. (The derivations for this are not given here, since this summary is not about that subject. For the derivations, see the summary of the third-year Flight Dynamics course.) After we have linearized the equations of motion, we can put them in a matrix form. For the symmetric equations of motion, we get

$$\begin{bmatrix} C_{X_u} - 2\mu_c D_c & C_{X_\alpha} & C_{Z_0} & C_{X_q} \\ C_{Z_u} & C_{Z_\alpha} + (C_{Z_{\dot{\alpha}}} - 2\mu_c) D_c & -C_{X_0} & 2\mu_c + C_{Z_q} \\ 0 & 0 & -D_c & 1 \\ C_{m_u} & C_{m_\alpha} + C_{m_{\dot{\alpha}}} D_c & 0 & C_{m_q} - 2\mu_c K_Y^2 D_c \end{bmatrix} \begin{bmatrix} \hat{u} \\ \alpha \\ \theta \\ \frac{q\bar{c}}{V} \end{bmatrix} = \begin{bmatrix} -C_{X_{\delta_e}} & -C_{X_{\delta_t}} \\ -C_{Z_{\delta_e}} & -C_{Z_{\delta_t}} \\ 0 & 0 \\ -C_{m_{\delta_e}} & -C_{m_{\delta_t}} \end{bmatrix} \begin{bmatrix} \delta_e \\ \delta_t \end{bmatrix}. \tag{2.1}$$

Similarly, for the asymmetric equations of motion, we get

$$\begin{bmatrix} C_{Y_\beta} + (C_{Y_{\dot{\beta}}} - 2\mu_b) D_b & C_L & C_{Y_p} & C_{Y_r} - 4\mu_b \\ 0 & -\frac{1}{2} D_b & 1 & 0 \\ C_{l_\beta} & 0 & C_{l_p} - 4\mu_b K_X^2 D_b & C_{l_r} + 4\mu_b K_{XZ} D_b \\ C_{n_\beta} + C_{n_{\dot{\beta}}} D_b & 0 & C_{n_p} + 4\mu_b K_{XZ} D_b & C_{n_r} - 4\mu_b K_Z^2 D_b \end{bmatrix} \begin{bmatrix} \beta \\ \varphi \\ \frac{pb}{2V} \\ \frac{rb}{2V} \end{bmatrix} = \begin{bmatrix} -C_{Y_{\delta_a}} & -C_{Y_{\delta_r}} \\ 0 & 0 \\ -C_{l_{\delta_a}} & -C_{l_{\delta_r}} \\ -C_{n_{\delta_a}} & -C_{n_{\delta_r}} \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_r \end{bmatrix}. \tag{2.2}$$

To use these equations for computations, we often have to transform them into state space form. To put an equation in state space form, we first have to isolate all terms with one of the **differential operators** D_c and D_b on one side of the equation. After this, we apply the definition of these operators, being

$$D_c = \frac{\bar{c}}{V}\frac{d}{dt} \qquad \text{and} \qquad D_b = \frac{b}{V}\frac{d}{dt}. \tag{2.3}$$

If we then also move the term $\frac{\bar{c}}{V}$ (for the symmetric equations) or the term $\frac{b}{V}$ (for the asymmetric equations) to the other side of the equation, we have put the equation in its state space form.

3 Control theory recap: frequency domain and diagrams

3.1 The frequency domain

Let's suppose we have the state space representation of a system, but we want to examine the system in the **frequency domain**. To do this, we have to put the system in the frequency domain first. To accomplish this, we need to follow several steps. First, we have to rewrite the state space form in the Laplace domain. (Take the Laplace transform of the equation.) We can assume zero initial conditions here. (This simplifies matters a bit.) Second, we eliminate the **state vector** $X(s)$ and rewrite the system of equations as $Y(s) = F(s)U(s)$. (Here, $Y(s)$ is the **output vector** and $U(s)$ is the **input vector**.) Thirdly, we substitute the Laplace variable s by $j\omega$, with $j = \sqrt{-1}$ the complex variable. If everything has gone well, then we should have found

$$F(j\omega) = C\left(j\omega I - A\right)^{-1} B + D. \tag{3.1}$$

We can now examine the frequency domain of the system. Let's suppose that we give our system a sinusoidal input $U(s)$. This input has unit magnitude and frequency ω. The result is that the output $Y(s)$ will start to oscillate as well. However, it doesn't do that in exactly the same way. Instead, the amplitude is multiplied by the **amplitude gain** K. Next to this, there is also a **phase angle** ϕ. Both parameters follow from the transfer function $F(j\omega)$ and can be found using

$$K = |F(j\omega)| \qquad \text{and} \qquad \phi = \arg\left(F(j\omega)\right). \tag{3.2}$$

So, if $K > 1$, the oscillation is amplified. Otherwise, it is reduced in strength. Similarly, if $\phi > 0$, the system has phase lead. Otherwise, it has phase lag.

3.2 Different kinds of diagram

We would like to know how the gain K and the phase angle ϕ vary with the frequency ω. This is displayed in a **Bode diagram**. In fact, a Bode diagram consists of two plots. Both plots have on the horizontal axis the frequency ω, on a logarithmic scale. The first plot shows the gain K in decibel (linearly). To put the gain K in decibel, you can use the equation

$$K_{dB} = 20 \cdot \log_{10} K \qquad \Leftrightarrow \qquad K = 10^{\left(\frac{K_{dB}}{20}\right)}. \tag{3.3}$$

The second plot shows the phase angle ϕ (also linearly).

Next to the Bode diagram, there is also the **Nyquist diagram** (also known as the **polar plot**). To make it, we make a complex plot of $F(j\omega)$ with respect to ω. In other words, we plot the real part of $F(j\omega)$ on the x-axis and its imaginary part on the y-axis for varying ω. The distance from the origin now indicates the gain, while the (CCW) angle with respect to the x-axis indicates the phase angle.

Finally, there is the **Nichols diagram**. In this diagram, we plot the decibal gain K_{dB} (vertically) against the phase angle ϕ (horizontally).

3.3 The Nyquist stability criterion

Let's suppose we have a basic feedback system, with transfer function $F(s) = G(s)/(1 + G(s)H(s))$. $F(s)$ now is the **closed loop transfer function** (CL). Also, $G(s)$ is the **feed forward transfer function** (FF) and $G(s)H(s)$ is the **open loop transfer function** (OL). We can make a Nyquist diagram of the open loop transfer function $G(s)H(s)$. The **Nyquist stability criterion** now tells us something about the stability of the entire closed-loop transfer function $F(s)$.

First, we need to count the number of poles k of the transfer function $G(s)H(s)$ with real part bigger than zero. (So, the number of poles in the right half plane.) Second, we need to count the number of net counterclockwise encirclements of the point -1 of the Nyquist diagram of $G(s)H(s)$. If this number is equal to the number k, then the closed loop system is stable. Otherwise, it is unstable.

4 Control theory recap: system properties

4.1 Controllability, observability, stabilizability and detectability

Let's examine a system in state space form. In other words, the system can be described by $\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}$ and $\mathbf{y} = C\mathbf{x} + D\mathbf{u}$. We can now make several definitions concerning this system.

The system, or equivalently the pair (A, B), is said to be **state controllable** if, for any initial state $\mathbf{x}(0) = \mathbf{x_0}$, any time $t_1 > 0$ and any final state $\mathbf{x_1}$, there exists an input $\mathbf{u}(t)$ such that $\mathbf{x}(t_1) = \mathbf{x_1}$. Otherwise the system is said to be state **uncontrollable**. To find out whether a system is state controllable, we can examine the controllability matrix R, defined as

$$R = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}. \tag{4.1}$$

The system is controllable if, and only if, the matrix R is of full rank. In other words, all its rows are linearly independent.

The system, or equivalently the pair (A, C), is said to be **state observable** if, for any initial time $t_1 > 0$, the initial state $\mathbf{x}(0) = \mathbf{x_0}$ can be determined from the time history of the input $\mathbf{u}(t)$ and the output $\mathbf{y}(t)$ in the interval $[0, t_1]$. Otherwise the system is said to be **state unobservable**. To find out whether a system is state observable, we can examine the observability matrix W, defined as

$$W = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{4.2}$$

The system is observable if, and only if, the matrix W is of full rank. In other words, all its columns are linearly independent.

The system, or equivalently the pair (A, B), is said to be **state stabilizable** if all unstable modes are state controllable. This is the case if there exists a **feedback matrix** F which stabilizes the system (thus causing $A + BF$ to be stable). To test for stabilizability, we can use the **Hautus test**. It says that (A, B) is stabilizable if and only if the matrix

$$\begin{bmatrix} A - \lambda I & B \end{bmatrix} \tag{4.3}$$

has full rank for all unstable eigenvalues λ. (In other words, for all λ with positive real parts, the above matrix has linearly independent rows.)

The system, or equivalently the pair (A, C), is said to be **state detectable** if all unstable modes are state observable. This is the case if there exists a matrix L such that $A + LC$ is stable (and thus has all its eigenvalues in the left half part of the complex plane). To test for detectability, we can again use the Hautus test. This time, it says that (A, C) is detectable if and only if the matrix

$$\begin{bmatrix} A - \lambda I \\ C \end{bmatrix} \tag{4.4}$$

has full rank for all unstable eigenvalues λ. (In other words, for all λ with positive real parts, the above matrix has linearly independent columns.)

4.2 Varying the poles and zeroes of the open loop transfer function

Let's examine a basic closed loop system with transfer function $F(s)$. We can add a constant gain K into this system. This turns the transfer function from

$$F(s) = \frac{G(s)}{1 + G(s)H(s)} \qquad \text{into} \qquad F(s) = \frac{KG(s)}{1 + KG(s)H(s)}. \tag{4.5}$$

By varying this gain K, we will vary the properties of the system. This is displayed by a **root locus plot**. (As you remember from Control Theory, a root locus plot shows how the poles of the closed loop transfer function vary with K.)

Sometimes, however, we can also wind up in a situation, where we are allowed to choose one pole p_{var} of the open loop transfer function $G(s)H(s)$. This situation is very similar to the case where we can choose the gain K. To show this, we define

$$G(s)H(s) = \left(\frac{1}{s - p_{var}}\right) Q(s). \tag{4.6}$$

This will turn the closed loop transfer function into

$$F(s) = \frac{G(s)}{1 + G(s)H(s)} = \frac{\frac{G(s)}{s+Q(s)}\left(s - p_{var}\right)}{1 - \frac{p_{var}}{s+Q(s)}} = \frac{G_1(s)}{1 + G_1(s)H_1(s)}. \tag{4.7}$$

In this equation, we have defined the **modified transfer functions** $G_1(s)$ and $H_1(s)$ as

$$G_1(s) = \frac{G(s)}{s + Q(s)}\left(s - p_{var}\right) \qquad \text{and} \qquad H_1(s) = -\frac{p_{var}}{G(s)\left(s - p_{var}\right)}. \tag{4.8}$$

We can now see something interesting. Previously, the denominator of the closed loop transfer function was $1 + KG(s)H(s)$. By varying K, we varied the poles. However, this time the denominator is $1 - p_{var}\left(s + Q(s)\right)^{-1}$. It is of the same form! So, varying p_{var} is just like varying a gain K. And we can again make a root locus plot.

We can do the same with varying zeroes. If we can choose a zero z_{var} of the open loop transfer function, then we write $G(s)H(s) = (s - z_{var})P(s)$. The closed loop transfer function now turns into

$$F(s) = \frac{\frac{G(s)}{1+sP(s)}}{1 - \frac{z_{var}P(s)}{1+sP(s)}} = \frac{G_2(s)}{1 + G_2(s)H_2(s)}. \tag{4.9}$$

Again, we see that varying z_{var} is like varying the gain K. We can therefore again use it to influence the properties of the closed loop system.